

Spectrum™ Technology Platform

Version 0-SNAPSHOT

Guide Dataflow Designer



Table des matières

1 - Premiers pas

Configuration de SAP	5
Démarrage d'Enterprise Designer	6
Découverte d'Enterprise Designer	6
Mon premier flux de données (Job)	10
Mon premier flux de données (Service)	13
Modèles de flux de données	16
Importation et exportation de dataflows	17

2 - Conception d'un flux

Types de flux	19
Entrée d'un flux	21
Champs	27
Stages de contrôle	43
Sortie d'un flux	112
Flux de données intégrés	117
Rapports	122
Considérations en matière de performances	126
Versions de flux de données	141

3 - Inspection et tests

Recherche d'erreurs dans un flux	145
Inspection d'un flux de données	146
Test d'un service avec Management Console	151

4 - Exécution d'un flux

Exécution d'un job ou d'un process flow	153
Exposition d'un service	176
Options de Runtime	178

Configuration des notifications par courrier électronique pour un flux de données	182
---	-----

5 - Combinaison de flux en un process flow

Introduction aux processflow	184
Conception des processflow	184

6 - Création de composants de flux réutilisables

Introduction aux sous-flux	225
Utilisation d'un sous-flux comme source	225
Utilisation d'un sous-flux au milieu d'un flux de données	227
Utilisation d'un sous-flux comme collecteur de données	228
Modification d'un sous-flux	229
Suppression d'un sous-flux	229
Exposition et annulation de l'exposition d'un sous-flux	230
Conversion d'un stage en un sous-flux	230

7 - Exemples de flux

Introduction	233
Intégration entre SugarCRM OnPremises et Microsoft Dynamics 365 Online	234
Intégration entre Salesforce et Oracle Eloqua	237

8 - A propos de Spectrum™ Technology Platform

Qu'est-ce que Spectrum™ Technology Platform ?	240
Architecture de gestion des données d'entreprise	241
Architecture Spectrum™ Technology Platform	245
Modules et composants	249

1 - Premiers pas

In this section

Configuration de SAP	5
Démarrage d'Enterprise Designer	6
Découverte d'Enterprise Designer	6
Mon premier flux de données (Job)	10
Mon premier flux de données (Service)	13
Modèles de flux de données	16
Importation et exportation de dataflows	17

Configuration de SAP

Les outils client de Spectrum™ Technology Platform sont des applications que vous pouvez utiliser afin d'administrer votre serveur, et de concevoir et d'exécuter des dataflows et des process flows. Vous devez installer votre serveur Spectrum™ Technology Platform avant d'installer les outils client.

Avant l'installation, assurez-vous de lire les notes de publication. Les notes de publication contiennent une liste de problèmes connus, des informations de compatibilité importantes et des notes d'installation propres à chaque version.

Cette procédure explique comment installer les outils client suivants :

- **Enterprise Designer**— Utilisez Enterprise Designer pour créer, modifier et exécuter des dataflows.
- **Job Executor**—Job Executor est un outil de ligne de commande qui vous permet d'exécuter une tâche depuis une ligne de commande ou un script. Le travail doit au préalable avoir été créé et enregistré sur Spectrum™ Technology Platform via Enterprise Designer.
- **Process Flow Executor**—Process Flow Executor est un outil de ligne de commande qui permet l'exécution d'un process flow depuis une ligne de commande ou un script. Le process flow doit au préalable avoir été créé et enregistré sur Spectrum™ Technology Platform via Enterprise Designer.
- **Utilitaire Administration**—L'utilitaire Administration fournit un accès de ligne de commande à plusieurs fonctions administratives. Vous pouvez l'utiliser dans un script, ce qui vous permet d'automatiser certaines tâches administratives. Vous pouvez également l'utiliser de manière interactive.

Remarque : À compter de Spectrum version 11.0, Management Console est un outil Web et non plus un client installable comme il l'était dans les précédentes versions.

Pour installer les outils client :

1. Ouvrez un navigateur internet et allez sur la page d'accueil Spectrum™ Technology Platform à :

`http://<nomduserveur>:<port>`

Par exemple, si vous avez installé Spectrum™ Technology Platform sur un ordinateur appelé « monspectrumplatform » et qu'il utilise le port HTTP par défaut, 8080, vous devrez aller sur :

`http://monspectrumplatform:8080`

2. Cliquez sur **Platform Client Tools**.
3. Téléchargez l'outil client que vous souhaitez installer.

Démarrage d'Enterprise Designer

Enterprise Designer est une application Windows permettant de créer des flux de données. Pour démarrer Enterprise Designer :

1. Sélectionnez **Démarrer > Tous les programmes > Pitney Bowes > Spectrum™ Technology Platform > Outils clients > Enterprise Designer**.
2. Saisissez le nom de serveur ou l'adresse IP ou sélectionnez-le/la dans la liste déroulante. Si vous utilisez Spectrum™ Technology Platform dans un cluster, saisissez le nom de l'adresse IP de l'équilibreur de charge du cluster.
3. Saisissez vos nom d'utilisateur et votre mot de passe.
4. Dans le champ Port, saisissez le port du réseau que le serveur doit utiliser pour la communication de Spectrum™ Technology Platform, selon sa configuration. Le numéro de port par défaut est 8080.
5. Cliquez sur **Utiliser une connexion sécurisée** si vous voulez que la communication entre le client et le serveur ait lieu par l'intermédiaire d'une connexion HTTPS.

Remarque : Une connexion sécurisée n'est disponible que si la communication HTTPS a été configurée sur le serveur. Si vous exécutez Enterprise Designer sous Windows 7, l'utilisation de l'adresse IP du champ **Nom de serveur** peut ne pas fonctionner, suivant le type de certificat utilisé pour sécuriser la communication entre Enterprise Designer et le serveur. Si l'adresse IP ne fonctionne pas, utilisez le nom d'hôte à la place.

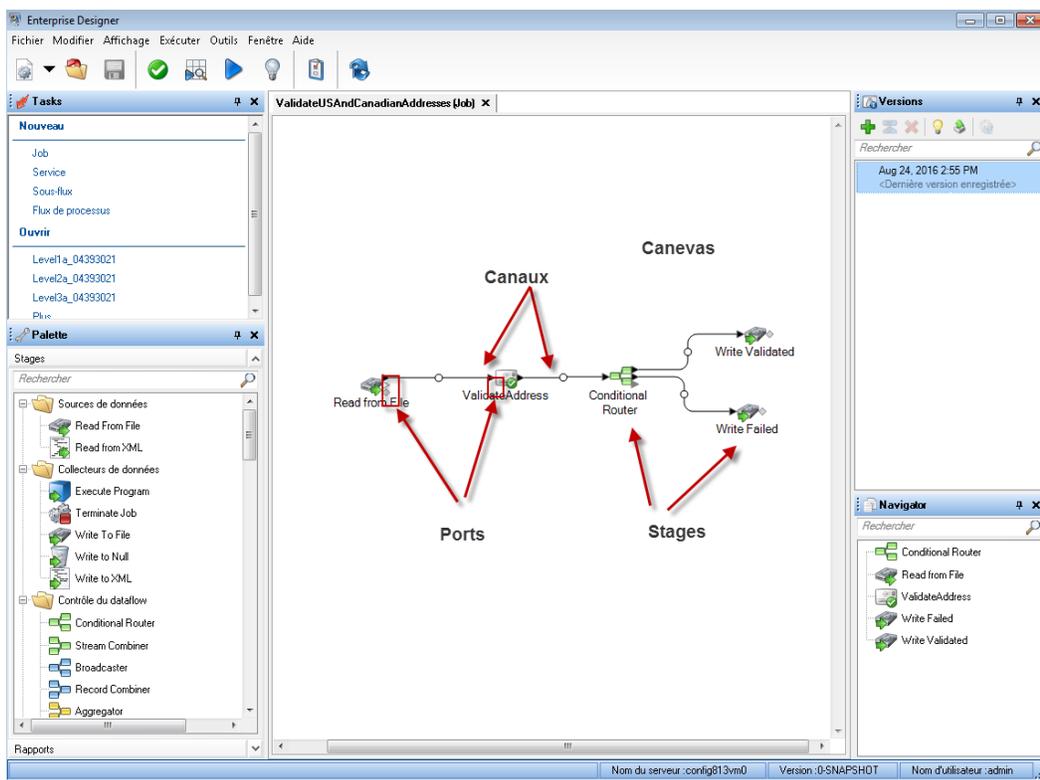
6. Cliquez sur **Connexion**.

Découverte d'Enterprise Designer

Enterprise Designer est un outil visuel permettant de créer des flux de données. À l'aide de ce client, vous pouvez effectuer les opérations suivantes :

- Créer et modifier les jobs, services, sous-flux et processflow
- Tester les flux de données pour détecter des problèmes
- Exposer et masquer des services
- Générer des rapports

La fenêtre Enterprise Designer ressemble à cela :



Pour utiliser des flux de données, vous devez comprendre quelques termes importants :

Canevas Le canevas est la principale zone de travail. L'image ci-dessus présente le canevas ouvert avec un flux de données nommé ValidateUSAndCanadianAddresses. Il s'agit d'un flux de données de job, ce qui signifie qu'il effectue un traitement par lots en lisant et écrivant les données dans un fichier. Dans ce cas, le flux de données écrit la sortie dans deux fichiers.

Stage Les stages, représentés par des icônes sur le canevas, effectuent un type d'activité spécifique, tel que le tri des enregistrements, la validation des adresses, la mise en correspondance d'enregistrements similaires, etc. Pour ajouter un stage, faites glisser le stage de la palette (sur le côté gauche de la fenêtre) sur le canevas.

Si un stage requiert votre attention, un cercle bleu apparaît sur l'icône :



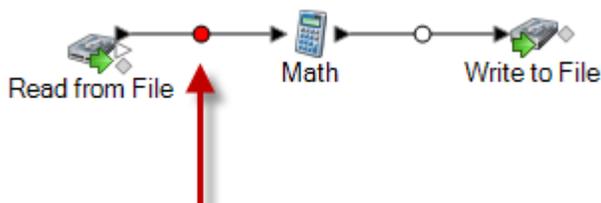
il est impossible d'exécuter correctement un flux de données si celui-ci comporte des stages nécessitant une attention. Par conséquent, double-cliquez sur le stage pour configurer les paramètres requis. Une fois que vous avez configuré tous les paramètres requis, le cercle bleu n'apparaît plus :



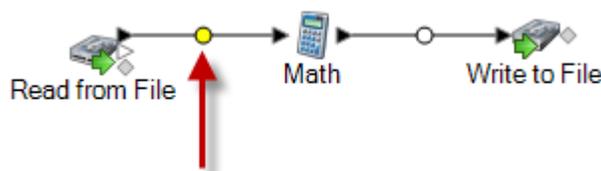
Canal Un canal est une connexion entre deux stages ou plus, à travers lequel les enregistrements sont transmis d'un stage à un autre. Dans l'exemple ci-dessus, vous pouvez remarquer que le stage Read from File est connecté au stage ValidateAddress avec un canal. Les enregistrements sont lus dans le flux de données dans Read from File, puis envoyés à ValidateAddress via ce canal. ValidateAddress est ensuite connecté à Conditional Router via un canal. Conditional Router, qui analyse des enregistrements et les envoie sur différents chemins dans un flux de données, selon les conditions définies par le concepteur de flux de données, présente deux canaux sortants, l'un se dirigeant vers un stage Write Validated et l'un vers un stage Write Failed.

Le point au milieu d'un canal peut changer de couleur pour indiquer différentes conditions :

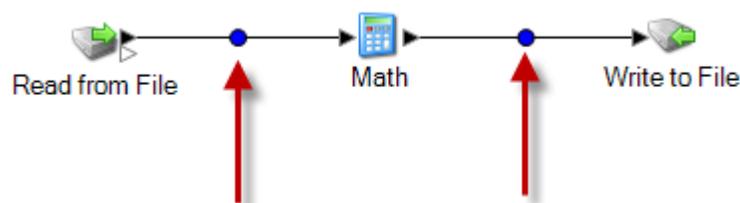
Rouge Indique une erreur, comme un échec de conversion de type qui rend un champ inutilisable par le stage en aval.



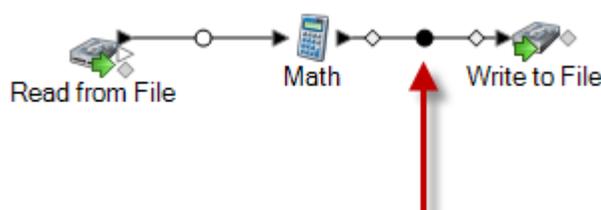
Jaune Vous avez supprimé un champ nécessaire à un stage en aval.



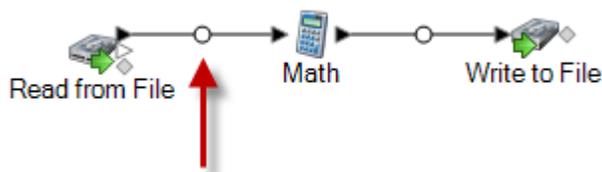
Bleu La conversion de type automatique a correctement converti un champ en type de données nécessaire au stage en aval.



Noir Un champ est renommé dans le canal.



Blanc Aucune action n'est effectuée sur les champs.



Port Si vous examinez les icônes de stage, vous remarquerez de petits ports en forme de triangle ou de losange, sur les côtés de chaque stage. Un port représente le mécanisme à travers lequel un stage envoie ou lit des données dans un canal. Les stages qui lisent des données dans le flux de données (appelés « sources ») ne disposent que de ports de sortie car ils se trouvent toujours au début d'un flux de données. Les stages qui envoient des données depuis le flux de données (appelés « collecteurs de données ») ne disposent que de ports d'entrée car ils se trouvent toujours à la fin d'un flux de données. Tous les autres stages disposent à la fois de ports d'entrée et de sortie. En outre, certains stages présentent des ports d'erreur, utilisés pour sortir des enregistrements causant des erreurs lors du traitement du stage, et certains stages disposent de ports de rapport, utilisés pour générer des rapports sur la sortie du stage.

Par ailleurs, la fenêtre Enterprise Designer présente ces fonctionnalités :

Fonction	Description
Tâches	Permet de créer rapidement un job, service, sous-flux ou processflow. Vous permet également d'ouvrir des flux de données récemment ouverts.
Explorateur Serveur	Affiche tous les flux enregistrés sur le serveur Spectrum™ Technology Platform. Si cela n'est pas visible dans l'Explorateur Serveur, sélectionnez Vue > Explorateur Serveur . Vous pouvez organiser les flux dans des dossiers. Pour créer un dossier, cliquez avec le bouton droit sur le nom du serveur et sélectionnez Nouveau dossier . Les noms de flux doivent être uniques dans tous les dossiers. Vous ne pouvez pas avoir deux flux portant le même nom, même s'ils se trouvent dans des dossiers différents.
Palette	Contient tous les stages et rapports que vous pouvez ajouter à votre flux de données. Les stages qui sont disponibles dans la palette dépendent des modules compris dans votre licence.
Canevas	Espace de travail graphique vers lequel vous faites glisser les stages et les reliez aux canaux pour la création des flux de données. Vous pouvez travailler avec plusieurs canevas de flux de données ouverts simultanément.

Fonction	Description
Versions	La fonctionnalité Versions dans Enterprise Designer vous permet de conserver un historique de révision de vos flux de données. Vous pouvez afficher les versions précédentes d'un flux de données, exposer des versions plus anciennes pour leur exécution et conserver un historique de vos modifications au cas où vous auriez besoin de revenir à une version précédente d'un flux de données.
Navigateur	Répertorie les stages et les rapports dans le flux. Vous pouvez faire un clic droit sur un élément dans le volet Navigateur pour modifier ses options.

Mon premier flux de données (Job)

Dans cette rubrique, vous allez créer un simple flux de données qui lit les données d'un fichier, les trie, puis les écrit dans un fichier. Sachant que ce flux de données lit les données d'un fichier et écrit leur sortie dans un fichier, il s'agit d'un flux de données de type « job » (tâche), à savoir, un flux de données qui exécute un traitement par lots. (L'autre principal type de flux de données, dit « service », exécute un traitement interactif via un appel à une API ou un service Web auprès du serveur.)

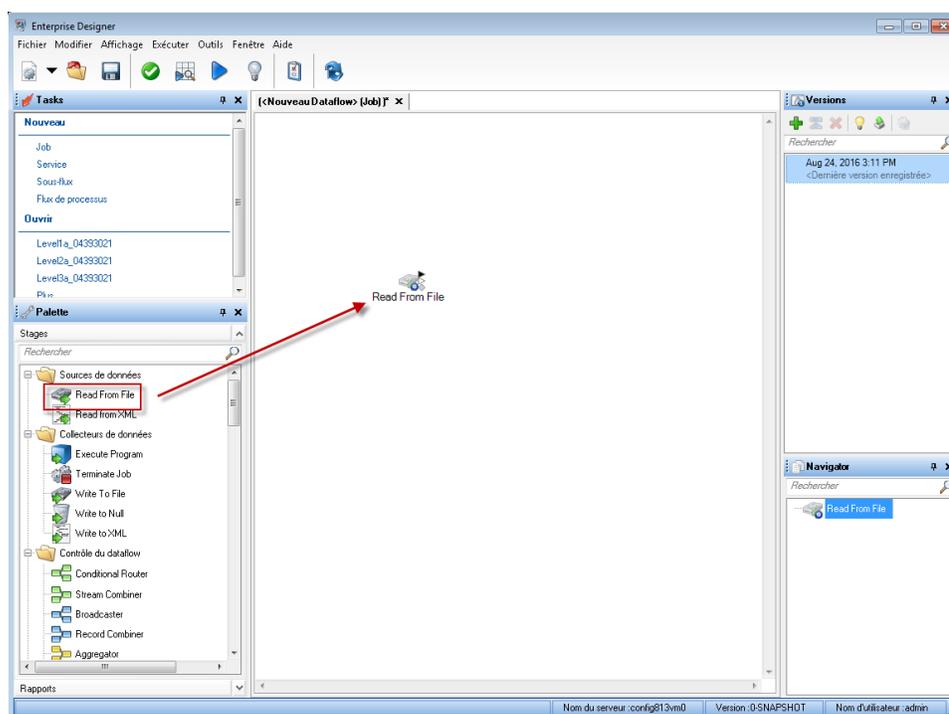
1. La toute première étape consiste à créer un échantillon de données à utiliser comme entrée dans votre flux de données. À l'aide d'un éditeur de texte, créez un fichier du type suivant :

```

FirstName, LastName, Region, Amount
Alan, Smith, East, 18.23
Jeannie, Wagner, North, 45.43
Joe, Simmons, East, 10.87
Pam, Hiznay, Central, 98.78

```

2. Enregistrez le fichier à un emplacement facile à retrouver.
3. Sélectionnez **Démarrer > Programmes > Pitney Bowes > Spectrum™ Technology Platform > Outils clients > Enterprise Designer**.
4. Sélectionnez **Fichier > Nouveau > Dataflow > Job**.
5. Vous pouvez maintenant commencer à créer votre flux de données. La première étape consiste à définir l'entrée du flux de données. Pour ce faire, procédez comme suit :
 - a) Glissez un stage Read from File sur le canevas :



- b) Double-cliquez sur le stage Read from File du canevas.
- c) Dans le champ **File name**, indiquez le fichier créé à l'étape 1 à la page 10.
- d) Dans le champ **Type d'enregistrement**, choisissez **Délimité**.
- e) Dans le champ **Séparateur de champs**, sélectionnez **Virgule (,)**.
- f) Cochez la case **Le premier enregistrement correspond aux en-têtes**.
- g) Cliquez sur l'onglet **Champs**.
- h) Cliquez sur **Régénérer**, puis sur **Oui**.

Le stage est automatiquement configuré pour les champs de votre fichier d'entrée.

- i) Cliquez sur **Détecter le type**. Cette opération balaie le champ d'entrée et détermine le type de données approprié pour chaque champ. Notez que le type chaîne du champ **Amount** devient un type double.
 - j) La configuration de Read from File est terminée. Cliquez sur **OK**.
6. Ensuite, ajoutez un stage qui trie les enregistrements par région. Pour ce faire, procédez comme suit :
- a) Glissez le stage Sorter sur le canevas.
 - b) Cliquez sur le triangle noir plein à droite du stage Read from File (port de sortie) et glissez-le à gauche du stage Sorter du canevas pour créer un canal connectant Read from File et Sorter.

Votre dataflow prend la forme suivante :



- c) Double-cliquez sur le stage Sorter du canevas.
 - d) Cliquez sur **Ajouter**.
 - e) Dans le champ **Field Name**, sélectionnez **Region**.
 - f) La configuration de Sorter est terminée. Cliquez sur **OK**.
7. Pour finir, définissez le fichier de sortie dans lequel le flux de données va écrire sa sortie. Pour ce faire, procédez comme suit :
- a) Glissez un stage Write to File sur le canevas.
 - b) Cliquez sur le triangle noir plein à droite du stage Sorter et glissez-le à gauche du stage Write to File du canevas.

Votre dataflow prend la forme suivante :



- c) Double-cliquez sur le stage Write to File.
- d) Dans le champ **File name**, indiquez un fichier de sortie. Il peut s'agir de tout fichier de votre choix.
- e) Dans le champ **Séparateur de champs**, sélectionnez **Virgule (,)**.
- f) Cochez la case **Le premier enregistrement correspond aux en-têtes**.
- g) Cliquez sur l'onglet **Champs**.
- h) Cliquez sur **Ajout rapide**.
- i) Cliquez sur **Tout sélectionner**, puis sur **OK**.
- j) À l'aide des boutons **Déplacer vers le haut** et **Déplacer vers le bas**, réorganisez les champs de sorte qu'ils se trouvent dans l'ordre suivant :

FirstName
 LastName
 Région
 Montant

Ainsi, l'ordre des champs des enregistrements de votre fichier de sortie est identique à celui des enregistrements de votre fichier d'entrée.

- k) La configuration de Write to File est terminée. Cliquez sur **OK**.
8. Dans Enterprise Designer, sélectionnez **Fichier > Enregistrer**.
 9. Nommez votre flux de données, puis cliquez sur **OK**.
 10. Votre flux de données est maintenant opérationnel. Sélectionnez **Exécuter > Exécuter le flux actif**.

11. La fenêtre **Détails d'exécution** apparaît, affichant le statut du job. Cliquez sur **Actualiser**. Une fois que le statut affiche **Succès**, cliquez sur **Fermer**.

Ouvrez le fichier de sortie indiqué dans le stage Write to File. Vous constatez que vos enregistrements ont été triés par région, comme indiqué dans le stage Sorter.

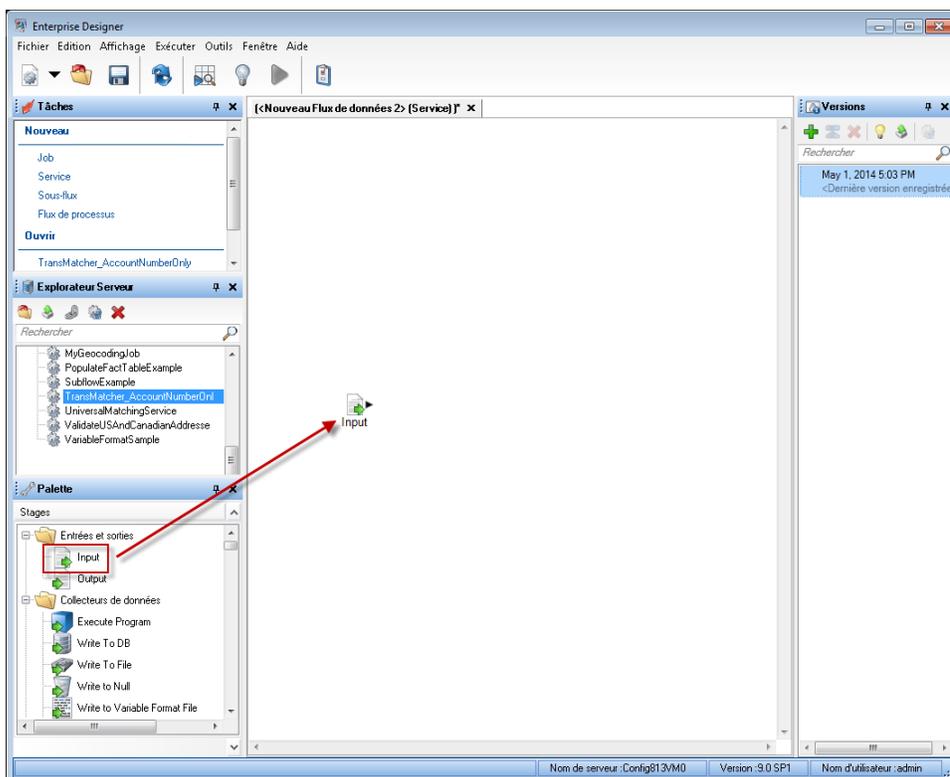
```
FirstName,LastName,Region,Amount
Pam,Hiznay,Central,98.78
Alan,Smith,East,18.23
Joe,Simmons,East,10.87
Jeannie,Wagner,North,45.43
```

Félicitations ! Vous avez conçu et exécuté votre premier flux de données de type job.

Mon premier flux de données (Service)

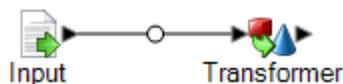
Dans cette rubrique, vous allez créer un simple flux de données qui accepte les données d'un appel à une API ou un service Web, les traite, puis renvoie une réponse via l'API ou le service Web. Sachant que ce flux de données doit être exposé sous forme de service sur le serveur Spectrum™ Technology Platform, il s'agit d'un flux de données de type « service ». (L'autre principal type de flux de données, dit « job », exécute un traitement par lots en lisant les données d'un fichier ou d'une base de données, en les traitant, puis en écrivant la sortie dans un fichier ou une base de données.)

1. Sélectionnez **Démarrer > Tous les programmes > Pitney Bowes > Spectrum™ Technology Platform > Outils clients > Enterprise Designer**.
2. Sélectionnez **Fichier > Nouveau > Flux de données > Service**.
3. Vous pouvez maintenant commencer à créer votre flux de données. La première étape consiste à définir l'entrée du flux de données. Votre flux de données accepte deux champs en entrée : FirstName et LastName.
 - a) Faites glisser un stage Input de la palette jusqu'au canevas.



- b) Double-cliquez sur le stage Input du canevas.
 - c) Cliquez sur **Ajouter**, puis de nouveau sur **Ajouter**.
 - d) Dans le champ **Field Name**, saisissez `FirstName`.
 - e) Cliquez sur **OK**, puis de nouveau sur **OK**.
 - f) Cliquez sur **Ajouter**, puis de nouveau sur **Ajouter**.
 - g) Dans le champ **Field Name**, saisissez `LastName`.
 - h) Cliquez sur **OK**, puis de nouveau sur **OK**.
 - i) La définition de l'entrée du flux de données est terminée. Cliquez sur **OK**.
4. Ensuite, ajoutez un stage modifiant la casse des données des champs `FirstName` et `LastName` en majuscules.
- a) Glissez un stage Transformer de la palette jusqu'au canevas.
 - b) Cliquez sur le triangle noir plein à droite du stage Input (port de sortie) et glissez-le à gauche du stage Transformer du canevas pour créer un canal connectant Input et Transformer.

Votre flux de données prend la forme suivante :



- c) Double-cliquez sur le stage Transformer.
- d) Cliquez sur **Ajouter**.
- e) Dans l'arborescence à gauche, sous **Formatage**, cliquez sur **Casse**.

- f) Dans le champ **Field**, sélectionnez **FirstName**. Laissez **Majuscule** sélectionné.
 - g) Cliquez sur **Ajouter**.
 - h) Dans le champ **Field**, sélectionnez **LastName**. Laissez **Majuscule** sélectionné.
 - i) Cliquez sur **Ajouter**.
 - j) Cliquez sur **Fermer**.
 - k) La configuration de Transformer, qui transforme la valeur des champs FirstName et LastName en majuscules, est terminée. Cliquez sur **OK**.
5. Pour finir, définissez la sortie du flux de données. Votre flux de données renverra les champs FirstName et LastName en sortie.
- a) Glissez un stage Output sur le canevas.
 - b) Cliquez sur le triangle noir plein à droite du stage Transformer et glissez-le à gauche du stage Output du canevas.

Votre flux de données prend la forme suivante :

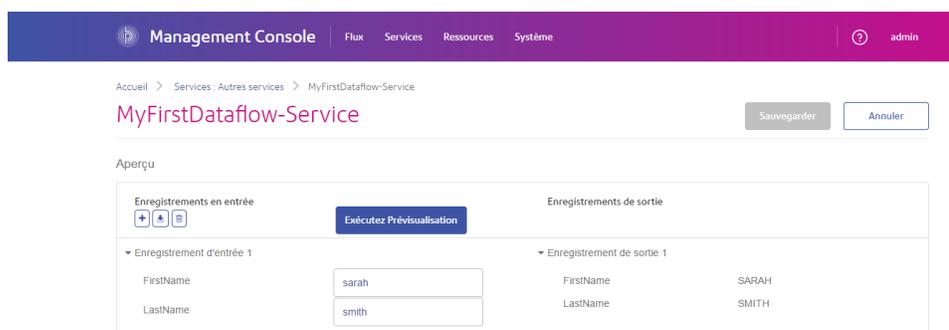


- c) Double-cliquez sur l'icône Output du canevas.
 - d) Cochez la case **Exposer**. Les cases à cocher en regard de FirstName et LastName doivent être cochées.
 - e) Cliquez sur **OK**.
6. Dans Enterprise Designer, sélectionnez **Fichier > Enregistrer**.
7. Nommez votre flux de données `MyFirstDataflow-Service`, puis cliquez sur **OK**.
8. Sélectionnez **Fichier > Exposer/Ne plus exposer et Enregistrer**. Cette opération expose votre flux de données, le rendant disponible sous forme de service sur le serveur.
9. Pour tester votre flux de données, procédez comme suit :
- a) Ouvrez Management Console en accédant à cette URL dans un navigateur web :

`http://server.port/managementconsole`

Où : *server* est l'adresse IP ou le nom de serveur de votre serveur Spectrum™ Technology Platform et *port* le port HTTP utilisé par Spectrum™ Technology Platform. Par défaut, le port HTTP est 8080.
 - b) Accédez à **Services > Other Services**.
 - c) Dans la liste des services, cochez la case en regard de **MyFirstDataflow-Service**, puis cliquez sur le bouton Modifier .
 - d) Saisissez un nom tout en minuscules dans le champ FirstName.
 - e) Saisissez un nom tout en minuscules dans le champ LastName.
 - f) Cliquez sur **Exécuter l'aperçu**.

Vous constatez que le service a transformé les noms de champs tout en majuscules, comme indiqué dans le stage Transformer de votre flux de données.



Félicitations ! Vous avez conçu et exécuté votre premier flux de données de type service. Le service est maintenant disponible sur le serveur et accessible via un appel vers une API ou des services Web. L'URL de la ressource de la destination SOAP de ce service est la suivante :

`http://<ServerName>:<Port>/soap/MyFirstDataflow-Service`

L'URL de la ressource de la destination REST de ce service est la suivante :

`http://<ServerName>:<Port>/rest/MyFirstDataflow-Service`

Modèles de flux de données

Les modèles de flux de données illustrent les différentes façons d'utiliser Spectrum™ Technology Platform et ses modules pour répondre à vos besoins commerciaux. Ils montrent comment certains modules résolvent diverses spécificités, comme le parsing, la normalisation et la validation des noms et adresses, le géocodage des adresses et ainsi de suite.

Les modèles de flux de données sont livrés avec chaque module pour lesquels vous prenez une licence. Par exemple, si vous avez une licence pour le module Data Normalization, vous recevez un modèle de flux de données de Standardisation des noms de personnes. Si vous avez une licence pour le module Universal Addressing, vous recevez les modèles de flux de données de Validation des adresses pour les États-Unis et le Canada.

Selon le but à atteindre pour chaque modèle, celui-ci peut être une tâche avec un échantillon de données, ou il peut être un service, sans échantillon de données. Vous pouvez utiliser les flux de données dans leur état d'origine et exécuter celles qui sont livrées comme tâches pour voir comment elles se comportent. Autrement, vous pouvez manipuler les flux de données en modifiant les fichiers d'entrée ou en apportant des services à vos propres tâches et en ajoutant des fichiers de sortie.

Remarque : Ces exemples ont pour objectif d'illustrer les différentes fonctionnalités Spectrum™ Technology Platform. Ils ne sont pas prévus pour être des solutions complètes pour votre environnement commercial particulier.

Importation et exportation de dataflows

Vous pouvez échanger des dataflows avec d'autres utilisateurs d'Enterprise Designer grâce aux fonctionnalités d'importation et d'exportation.

Remarque : Les dataflows ne peuvent être échangés qu'entre des versions identiques de Spectrum™ Technology Platform.

- Pour exporter un dataflow, sélectionnez **Fichier > Exporter**. Si vous avez utilisé la fonctionnalité Versions pour enregistrer les versions du dataflow, la version que vous avez sélectionnée est exportée.

Remarque : N'utilisez pas de caractères spéciaux dans le nom des services et des jobs que vous définissez. Ceci pourrait provoquer une erreur lors de l'exportation.

- Pour importer un process flow, sélectionnez **Fichier > Importer > Process flow**.
- Pour importer un dataflow, sélectionnez **Fichier > Importer > Dataflow**. Les stages du dataflow doivent être disponibles sur votre système avant que vous n'importiez le dataflow. Si le dataflow que vous importez contient des stages qui ne sont pas disponibles, vous verrez apparaître une erreur.
- Si vous utilisez Explorateur Serveur pour organiser vos dataflows, vous pouvez également exporter un dataflow en cliquant dessus avec le bouton droit de la souris et en sélectionnant **Exporter**. Pour importer un dataflow à l'aide d'Explorateur Serveur, cliquez avec le bouton droit de la souris dans l'emplacement d'Explorateur Serveur dans lequel vous souhaitez importer le dataflow et sélectionnez **Importer**.

2 - Conception d'un flux

In this section

Types de flux	19
Entrée d'un flux	21
Champs	27
Stages de contrôle	43
Sortie d'un flux	112
Flux de données intégrés	117
Rapports	122
Considérations en matière de performances	126
Versions de flux de données	141

Types de flux

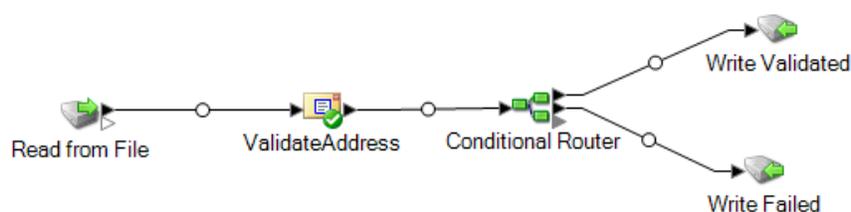
Un dataflow constitue une série d'opérations qui prennent des données d'une source, les traitent, puis écrivent la sortie vers une destination. Le traitement des données peut aller d'un simple tri à des actions d'enrichissement et de qualité des données plus complexes. Le concept de dataflow en lui-même est simple, mais vous pouvez concevoir des dataflows très complexes avec des chemins d'accès à embranchements, différentes sources d'entrée et diverses destinations de sortie.

Il existe quatre types de dataflows : jobs, services, sous-flux et processflow.

Job

Un job est un dataflow effectuant du traitement de lots. Un job lit les données d'un(e) ou plusieurs fichiers ou bases de données, traite ces données et écrit les résultats dans un(e) ou plusieurs fichiers ou bases de données. Les jobs peuvent être exécutés de manière manuelle dans Enterprise Designer ou à partir de la ligne de commande via l'exécuteur de job.

Le dataflow suivant est un job. Notez bien qu'il utilise le stage Read from File pour l'entrée et deux stages Write to File pour la sortie.

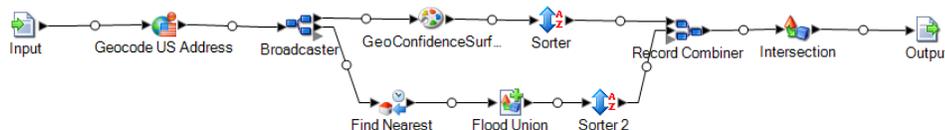


Service

Un service est un dataflow auquel vous pouvez accéder sous forme de services Web ou via l'API Spectrum™ Technology Platform. Vous transmettez un enregistrement au service et vous avez la possibilité de définir des options à utiliser lors du traitement de l'enregistrement. Le service traite les données et les renvoie.

Certains services deviennent disponibles dès lors que vous installez un module. Par exemple, quand vous installez le module Universal Addressing Module, le service ValidateAddress devient disponible sur votre système. Dans d'autres cas, vous devez créer un service dans Enterprise Designer, puis exposer ce service sur votre système comme service défini par l'utilisateur. Par exemple, les stages du module Location Intelligence ne sont pas disponibles en tant que services à moins que vous ne créiez d'abord un service qui utilise les stages de ce module.

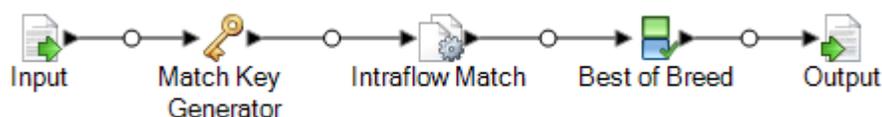
Vous pouvez également concevoir vos propres services personnalisés dans Enterprise Designer. Par exemple, le dataflow suivant détermine si une adresse présente un risque d'inondation :



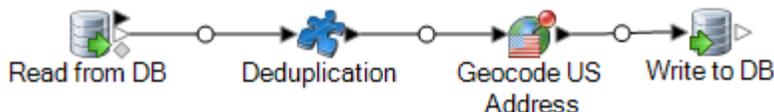
Remarque : Les noms du service, de l'option et du champ sont destinés à devenir des éléments XML, ils ne peuvent pas contenir de caractères considérés comme des éléments XML invalides (par exemple, les espaces sont interdits). Les services qui ne répondent pas à ces exigences fonctionneront toujours, mais ils ne seront pas exposés en tant que services Web.

Sous-flux

Un sous-flux est un dataflow qui peut être réutilisé par d'autres dataflows. Les sous-flux s'avèrent utiles quand vous souhaitez créer un processus réutilisable, facile à intégrer à des dataflows. Par exemple, vous pouvez souhaiter créer un sous-flux qui effectue une déduplication en utilisant certains paramètres de chaque stage, pour pouvoir utiliser le même processus de déduplication dans différents dataflows. Pour ce faire, vous pouvez créer un sous-flux comme suit :



Vous pouvez ensuite utiliser ce sous-flux dans un dataflow. Par exemple, vous pouvez utiliser le sous-flux de déduplication dans un dataflow qui effectue un géocodage de sorte que les données soient dédupliquées avant l'opération de géocodage :



Dans cet exemple, les données d'une base de données sont lues, puis transmises au sous-flux de déduplication, où elles sont traitées par Match Key Generator, Intraflow Match, puis Best of Breed. Ensuite, pour finir, elles quittent le sous-flux et sont transmises au stage suivant du dataflow parent, dans ce cas, Geocode US Address. Les sous-flux sont représentés par une icône morceau de puzzle, comme illustré ci-dessus.

Les sous-flux enregistrés et exposés s'affichent dans le dossier **Stages définis par l'utilisateur** d'Enterprise Designer.

Process flow

Un processflow exécute une série d'activités, comme des jobs et des applications externes. Chaque activité du flux de processus s'exécute à la fin de l'activité précédente. Les processflow s'avèrent utiles si vous souhaitez exécuter plusieurs flux de données en séquence ou un programme externe. Par exemple, un processflow peut exécuter un job de pour normaliser des noms, valider des adresses, puis utiliser une application externe pour trier les enregistrements dans le bon ordre pour l'obtention de réductions sur les tarifs postaux. Un tel processflow ressemble à ceci :



Dans cet exemple, les jobs Standardize Names et Validate Addresses sont des jobs exposés sur le serveur Spectrum™ Technology Platform. Exécuter le programme appelle une application externe, et une activité Réussi indique la fin du processflow.

Entrée d'un flux

Pour définir l'entrée d'un flux de données, utilisez un stage « source ». La définition de la source des données est la première étape d'un dataflow. Elle définit les données d'entrée à traiter.

Entrée d'un job

Les données d'entrée d'un job peuvent provenir d'un fichier ou d'une base de données. Spectrum™ Technology Platform peut lire des données dans de nombreux formats de fichier et types de base de données. Les types de sources de données que vous pouvez lire dépendent des modules pour lesquels vous disposez d'une licence. Le module Enterprise Data Integration fournit un accès à la plupart des sources de données de chaque module.

Remarque : Lorsque vous concevez une tâche, il serait préférable de prendre en compte la possibilité qu'il pourrait y avoir des enregistrements d'entrée incorrects. Un enregistrement incorrect est un enregistrement ne pouvant pas être analysé via l'une des classes d'analyseur syntaxique fourni par Spectrum™ Technology Platform. Pour en savoir plus sur le traitement des enregistrements d'entrée incorrects, reportez-vous à la section [Gérer les enregistrements d'entrée non conformes](#) à la page 22.

Entrée d'un service

Les données d'entrée d'un service sont définies dans un stage Input. Ce stage définit les champs que le service accepte d'une requête de service Web ou d'un appel d'API.

Définition de l'entrée d'un job

Les données d'entrée d'un job peuvent provenir d'un fichier, d'une base de données ou d'un service Cloud, suivant les modules que vous avez sous licence. Chaque module prend en charge l'entrée provenant de différentes sources, et la procédure de configuration de chaque type de source varie considérablement. Reportez-vous au Guide sur les solutions de vos modules, disponibles sur support.pb.com/spectrum.

Gérer les enregistrements d'entrée non conformes

Spectrum™ Technology Platform ne peut pas analyser un enregistrement non conforme. Quand Spectrum™ Technology Platform rencontre un enregistrement non conforme, les actions de la plate-forme peuvent être de :

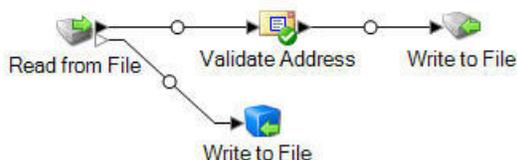
- Terminer le job
- Poursuivre le traitement
- Poursuivre le traitement jusqu'à ce qu'un certain nombre de mauvais enregistrements soit rencontré.
- Poursuivre le traitement mais en écrivant les mauvais enregistrements dans un fichier journal (via un stage récepteur facultatif).

Remarque : La fonctionnalité des enregistrements non conformes est limitée aux sources configurées pour la lecture de fichiers locaux situés sur le serveur et pour lesquels le tri n'a pas été configuré. Quand une source est configurée avec un fichier distant ou des champs de tri et que la source rencontre un enregistrement non conforme, le job est annulé, indépendamment de la manière dont les enregistrements non conformes sont configurés.

Pour gérer les enregistrements non conformes,

1. Ouvrez le job dans Enterprise Designer.
2. Ajouter un récepteur d'enregistrements non conformes dans votre flux de données.
 - a) Créez votre job en définissant votre fichier d'entrée et votre stage source et en ajoutant des services et sous-flux à votre flux de données.
 - b) Réalisez l'une des actions suivantes :
 - Connecter un stage récepteur au port de sortie facultatif du stage source dans votre flux de données. Le port facultatif est le port de sortie clair juste en dessous du port de sortie noir sur votre stage source. Si vous passez la souris sur ce port, vous verrez apparaître une astuce indiquant « error_port ». Les enregistrements non conformes sont écrits dans ce récepteur.
 - Ne rien connecter au port de sortie facultatif du stage source de votre flux de données, Spectrum™ Technology Platform va ignorer les enregistrements incorrects.

Le flux de données terminé doit ressembler à ceci :



Lorsque vous exécutez un job, l'Historique d'exécution contiendra une colonne qui montrera le nombre d'enregistrements non conformes rencontrés au cours de ce job.

3. Par défaut, Spectrum™ Technology Platform annule un job s'il rencontre un enregistrement incorrect. Ce comportement par défaut peut être modifié dans Management Console. Quel que

soit le comportement par défaut de votre système, vous pouvez le remplacer pour un job en procédant comme suit :

- a) Ouvrez le job dans Enterprise Designer.
- b) À l'intérieur d'un job ouvert, allez dans **Édition > Options de job**.
- c) Sélectionnez **Ne pas arrêter le job à cause d'un enregistrement non conforme** ou **Arrêter le job après avoir rencontré...** et entrez le nombre d'enregistrements incorrects autorisés qu'un job peut rencontrer avant qu'il ne se termine.

Définition de l'entrée d'un service

Le Stage Input définit les champs d'entrée pour un service ou sous-flux. Ce stage définit aussi les données de test devant être utilisées au cours de l'inspection des données.

Onglet Champs d'entrée

Cet onglet répertorie les champs que le flux de données accepte en entrée. Si le stage Input est connecté à un autre stage du flux de données, une liste des champs utilisés par les stages du flux de données s'affiche. Pour plus d'informations, reportez-vous à la section **Définition des champs d'entrée d'un service ou d'un sous-flux** à la page 23.

Onglet Entrée à inspecter

Cet onglet vous permet de spécifier les enregistrements d'entrée test à utiliser avec l'outil Inspection de données. Pour plus d'informations sur l'inspection de données, reportez-vous à la section **Inspection d'un flux de données** à la page 146.

Définition des champs d'entrée d'un service ou d'un sous-flux

Pour définir les champs d'entrée d'un service ou d'un sous-flux, vous devez utiliser le stage Input.

Remarque : Si vous définissez des données hiérarchiques dans les champs d'entrée, vous ne pourrez pas importer de données ou afficher les données verticalement.

1. Glissez un stage Input jusqu'au canevas.
2. Connectez le stage Input au stage suivant du flux de données.
3. Double-cliquez sur le stage Input.
4. Sélectionnez les champs que vous voulez utiliser comme entrée. La liste des champs affichés dépend du stage auquel le stage Input est connecté.
5. Pour ajouter un nouveau champ à la liste de champs, cliquez sur **Ajouter**. La fenêtre **Ajouter un champ personnalisé** s'affiche.
6. Cliquez encore une fois sur **Ajouter**.
7. Dans le champ **Nom de champ**, indiquez le nom à utiliser pour ce champ.
8. Sélectionnez le type de données.

Les types de données suivants sont pris en charge :

bigdecimal Type de données numérique prenant en charge 38 points de précision décimaux. Utilisez ce type de données pour les données utilisées dans les calculs mathématiques requérant un haut niveau de précision, en particulier ceux impliquant des données financières. Le type de données bigdecimal prend en charge des calculs plus précis que le type de données double.

booléen Type logique doté de deux valeurs : True et False.

bytearray Tableau (liste) d'octets.

Remarque : ByteArray n'est pas pris en charge en entrée pour un service REST.

date Type de données contenant un mois, un jour et une année. Par exemple, 2012-01-30 ou January 30, 2012. Vous pouvez indiquer un format de date par défaut dans Management Console.

datetime Type de données contenant un mois, un jour, une année et des heures, des minutes et des secondes. Par exemple, 2012/01/30 6:15 PM.

double Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre 2^{-1074} et $(2-2^{-52}) \times 2^{1023}$. En notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.

flottant Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre 2^{-149} et $(2-2^{23}) \times 2^{127}$. En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.

integer Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre -2^{31} (-2,147,483,648) et $2^{31}-1$ (2,147,483,647).

liste Au sens strict du terme, une liste n'est pas un type de données. Cependant, lorsqu'un champ contient des données hiérarchiques, il est traité comme un champ de « liste ». Dans Spectrum™ Technology Platform, une liste représente un ensemble de données composées de plusieurs valeurs. Par exemple, un champ Noms peut contenir une liste de valeurs de noms. Cela peut être représenté sous forme d'une structure XML, comme suit :

```
<Names>
  <Name>John Smith</Name>
  <Name>Ann Fowler</Name>
</Names>
```

Il est important de souligner que le type de données de liste Spectrum™ Technology Platform est différent du type de données de liste de schéma XML dans le sens où ce dernier est un type de données simple composé de plusieurs valeurs, tandis que le type de données de liste Spectrum™ Technology Platform est semblable à un type de données complexe XML.

long	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre -2^{63} (-9,223,372,036,854,775,808) et $2^{63}-1$ (9,223,372,036,854,775,807).
chaîne	Séquence de caractères.
time	Type de données contenant l'heure du jour. Par exemple, 21:15:59 ou 9:15:59 PM.

Vous pouvez également ajouter un nouveau type de données défini par l'utilisateur si nécessaire, et ce nouveau type peut provenir d'une liste de tout type de données défini. Par exemple, vous pouvez définir une liste de noms (chaîne) ou un nouveau type de données d'adresses qui inclut AddressLine1 (chaîne), City (chaîne), StateProvince (chaîne) et PostalCode (chaîne). Après avoir créé le champ, vous pouvez afficher le type de données en accédant à la boîte de dialogue Options d'entrée et en appuyant sur le bouton dans la colonne Type de donnée. La boîte de dialogue **Détails de type de données** apparaîtra, affichant la structure du champ.

9. Cliquez de nouveau sur **OK**.
10. Cochez la case de la colonne **Exposer** pour rendre le champ accessible aux opérations du stage. Décochez la case et cliquez sur **OK** pour supprimer le champ de la liste de champs.
11. Le champ **Nom de type de données** affiche le nom d'élément par défaut à utiliser pour les enregistrements d'entrée dans les requêtes de service Web SOAP et REST vers ce service. La valeur par défaut est `Row`. Si vous souhaitez utiliser un autre nom d'élément pour les enregistrements d'entrée, saisissez-le ici.

Par exemple, avec la valeur par défaut `Row`, une requête de service Web JSON utilise `Row` comme nom d'élément pour l'enregistrement d'entrée, comme indiqué ici :

```
{
  "Input":
  {
    "Row": [
      {
        "AddressLine1": "1825 Kramer Ln",
        "City": "Austin",
        "StateProvince": "TX"
      }
    ]
  }
}
```

Si vous remplacez la valeur du champ **Nom de type de données** par `Address`, la requête JSON doit utiliser `Address` au lieu de `Row` comme nom d'élément pour l'enregistrement, comme indiqué ici :

```
{
  "Input":
  {
```

```

    "Address": [
      {
        "AddressLine1": "1825 Kramer Ln",
        "City": "Austin",
        "StateProvince": "TX"
      }
    ]
  }
}

```

Définir un type de données de service Web

Le champ **Data type name** vous permet de contrôler les interfaces WSDL (SOAP) et WADL (REST) pour le service que vous êtes en train de créer. Le nom de l'élément Lignes est déterminé par le nom que vous donnez à ce stage dans le service, et le nom de l'élément Ligne est déterminé par le texte que vous saisissez ici.

Remarque : Pour WSDL, à la fois les requêtes et les réponses sont affectées, mais pour WADL, seules les réponses le sont.

Avant de nommer ce stage et d'entrer un texte dans ce champ, votre code ressemble peut-être à ceci :

```

<Rows>
  <Row>
    <FirstName>John</FirstName>
    <LastName>Doe</LastName>
  </Row>
  <Row>
    <FirstName>Jane</FirstName>
    <LastName>Doe</LastName>
  </Row>
</Rows>

```

Après avoir nommé ce stage et entré un texte dans ce champ, votre code ressemble peut-être à ceci :

```

<Names>
  <Name>
    <FirstName>John</FirstName>
    <LastName>Doe</LastName>
  </Name>
  <Name>
    <FirstName>Jane</FirstName>
    <LastName>Doe</LastName>
  </Name>
</Names>

```

Champs

Données plates et hiérarchiques

Spectrum™ Technology Platform prend en charge les données plates et les données hiérarchiques. En général, vous pouvez utiliser des données plates ou hiérarchiques en tant qu'entrée et sortie pour un flux de données. Certains stages dans le module Enterprise Routing exigent des données dans un format hiérarchique.

Données plates

Les données plates se composent d'enregistrements, un sur chaque ligne, et de champs dans chaque enregistrement. Les champs sont séparés par un caractère spécifique ou positionnés dans un emplacement défini sur la ligne. Par exemple, voici des données plates dotées de champs séparés par une virgule :

```
Sam,43,United States
Jeff,32,Canada
Mary,61,Ireland
```

Pour lire les données plates dans un flux de données, vous pouvez utiliser Read from file, Read from DB ou les stages d'entrée. Pour écrire une sortie de données plates à partir d'un flux de données, vous pouvez utiliser Write to File, Write to DB, ou les stages de sortie.

Données hiérarchiques

Les données hiérarchiques constituent une structure hiérarchique avec des éléments de données qui disposent de relations nommées parent/enfant. Spectrum™ Technology Platform peut lire et écrire des données hiérarchiques au format XML et au format de fichier de format variable. Par exemple, cela présente les données hiérarchiques au format XML :

```
<customers>
  <customer>
    <name>Sam</name>
    <age>43</age>
    <country>United States</country>
  </customer>
  <customer>
    <name>Jeff</name>
    <age>32</age>
    <country>Canada</country>
  </customer>
</customers>
```

```

<name>Mary</name>
<age>61</age>
<country>Ireland</country>
</customer>
</customers>

```

Cet exemple illustre une structure dans laquelle `<customer>` représente un enregistrement et où chaque enregistrement est constitué d'éléments XML simples (`<name>`, `<age>` et `<country>`).

Conversion de données

Il existe de nombreux cas où vous devrez convertir des données plates au format hiérarchique, ou inversement. Par exemple, vous disposez peut-être d'une entrée de flux de données au format hiérarchique mais souhaitez que le flux de données renvoie des données plates en sortie. Vous pouvez également avoir besoin de convertir des données d'entrée plates en données hiérarchiques pour certains stages (en particulier les stages dans le module Location Intelligence), puis de reconverter les données en données plates pour la sortie.

Pour convertir des données plates en données hiérarchiques, vous pouvez utiliser les éléments suivants :

- L'outil Liste des processus
- Le stage Aggregator dans un flux de données

Pour convertir les données hiérarchiques en données plates, utilisez le stage Splitter.

Conversion de données plates en liste

Liste des processus est un outil que vous pouvez utiliser dans un service ou un sous-flux pour renvoyer les données dans une liste. Ceci s'avère utile si votre flux de données inclut des stages nécessitant une entrée sous forme de liste, comme celle du module Location Intelligence.

1. Avec un flux de données déjà en place, cliquez avec le bouton droit sur le stage dont vous désirez convertir la sortie en liste. Il peut s'agir de n'importe quel stage excepté Entrée ou Sortie.
2. Sélectionnez **Liste de processus**. Le stage apparaît sur un arrière-plan bleu de forme carrée.
3. Pour déplacer un stage dans et hors de la liste des processus, appuyez sur la touche **Maj** tout en faisant glisser le stage supplémentaire.

Remarque : Si vous avez plusieurs stages et que vous désirez que la Liste des processus manipule les données, envisagez de créer un sous-flux, en l'amenant dans votre flux de données ; et en appliquant la fonctionnalité Liste des processus au sous-flux dans son ensemble.

4. Les champs d'entrée et de sortie d'une liste de processus sont appelés « ListField ». À l'aide de la fonction Renommer les champs, vous devez mapper votre champ de stage d'entrée en « ListField » dans le canal d'entrée, et mapper « ListField » dans le champ de votre étape de sortie. Pour plus d'informations, reportez-vous à la section [Modification du nom d'un champ](#) à la page 43.

5. Si vous désirez que la liste conserve les données dans le même ordre que celui dans lequel elles ont été entrées, cliquez avec le bouton droit sur la case Liste des processus et sélectionnez Options. Puis cochez la case Garder l'ordre de tri.
6. Pour confirmer que l'entrée des données dans le stage suivant sera répertoriée sous forme de liste, validez ou inspectez le flux de données. Pour plus d'informations sur l'inspection des données, voir [Inspection d'un flux de données](#) à la page 146.

types de données

Spectrum™ Technology Platform prend en charge différents types de données complexes, numériques et de chaîne. En fonction du type de traitement que vous voulez réaliser, vous devrez peut-être recourir à un ou de plusieurs de ces types. Pour un flux de données de validation d'adresse, vous ne pouvez utiliser que des données de chaînes. Pour un flux de données qui implique des calculs mathématiques, vous pouvez utiliser des types de données numériques ou booléennes. Pour un flux de données qui réalise des traitements spatiaux, vous pouvez utiliser un type de données complexe. Pour un flux de données qui combine les éléments pré-cités, vous pouvez utiliser plusieurs types de données.

Spectrum™ Technology Platform prend en charge les types de données suivants.

bigdecimal	Type de données numérique prenant en charge 38 points de précision décimaux. Utilisez ce type de données pour les données utilisées dans les calculs mathématiques requérant un haut niveau de précision, en particulier ceux impliquant des données financières. Le type de données bigdecimal prend en charge des calculs plus précis que le type de données double.
booléen	Type logique doté de deux valeurs : True et False.
bytearray	Tableau (liste) d'octets. Remarque : ByteArray n'est pas pris en charge en entrée pour un service REST.
date	Type de données contenant un mois, un jour et une année. Par exemple, 2012-01-30 ou January 30, 2012. Vous pouvez indiquer un format de date par défaut dans Management Console.
datetime	Type de données contenant un mois, un jour, une année et des heures, des minutes et des secondes. Par exemple, 2012/01/30 6:15 PM.
double	Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre 2^{-1074} et $(2-2^{-52}) \times 2^{1023}$. En notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.
flottant	Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre 2^{-149} et $(2-2^{23}) \times 2^{127}$. En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.

integer	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre -2^{31} (-2,147,483,648) et $2^{31}-1$ (2,147,483,647).
liste	Au sens strict du terme, une liste n'est pas un type de données. Cependant, lorsqu'un champ contient des données hiérarchiques, il est traité comme un champ de « liste ». Dans Spectrum™ Technology Platform, une liste représente un ensemble de données composées de plusieurs valeurs. Par exemple, un champ Noms peut contenir une liste de valeurs de noms. Cela peut être représenté sous forme d'une structure XML, comme suit :
	<pre><Names> <Name>John Smith</Name> <Name>Ann Fowler</Name> </Names></pre>
	Il est important de souligner que le type de données de liste Spectrum™ Technology Platform est différent du type de données de liste de schéma XML dans le sens où ce dernier est un type de données simple composé de plusieurs valeurs, tandis que le type de données de liste Spectrum™ Technology Platform est semblable à un type de données complexe XML.
long	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre -2^{63} (-9,223,372,036,854,775,808) et $2^{63}-1$ (9,223,372,036,854,775,807).
chaîne	Séquence de caractères.
time	Type de données contenant l'heure du jour. Par exemple, 21:15:59 ou 9:15:59 PM.

Spécification du type de données d'un champ

Vous pouvez spécifier le type de données d'un champ dans ces situations :

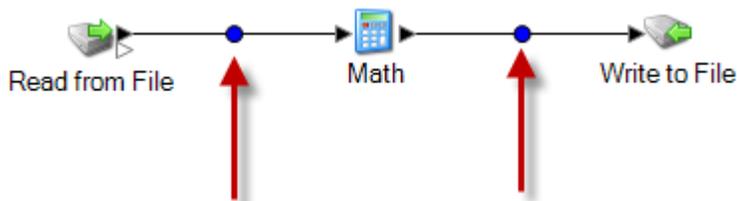
- **Stages Source** : la spécification des types de données vous permet de définir le type de données au début d'un flux de données, ce qui vous évite d'avoir recours à une conversion de type de données ultérieure dans le flux de données. Remarquez que pour le stage Read from DB, le type de données est sélectionné automatiquement et ne peut pas être modifié.
- **Stages Sink** : la spécification des types de données vous permet de contrôler le format des données renvoyées par le flux de données. Remarquez que pour Write to DB, le type de données est sélectionné automatiquement et ne peut pas être modifié.
- **Stage Transformer** : vous pouvez spécifier des types de données dans ce stage si vous utilisez un script personnalisé.
- **Stage Math et stage Group Statistics** : étant donné que ces stages réalisent des calculs mathématiques, choisir d'utiliser un type de données numérique particulier peut affecter les résultats de ces calculs, par exemple au niveau de la précision de l'opération de division. Si vous spécifiez un type de données pour un champ qui est différent du type de données du champ entrant dans le stage, le canal aval convertit automatiquement le champ au type de données que vous spécifiez, comme décrit à la section [Conversion de type de données automatique](#) à la page 31.

Remarque : Chaque stage prend en charge différents types de données. Pour une description des types de données supportés par chaque stage, voir la documentation du stage en question.

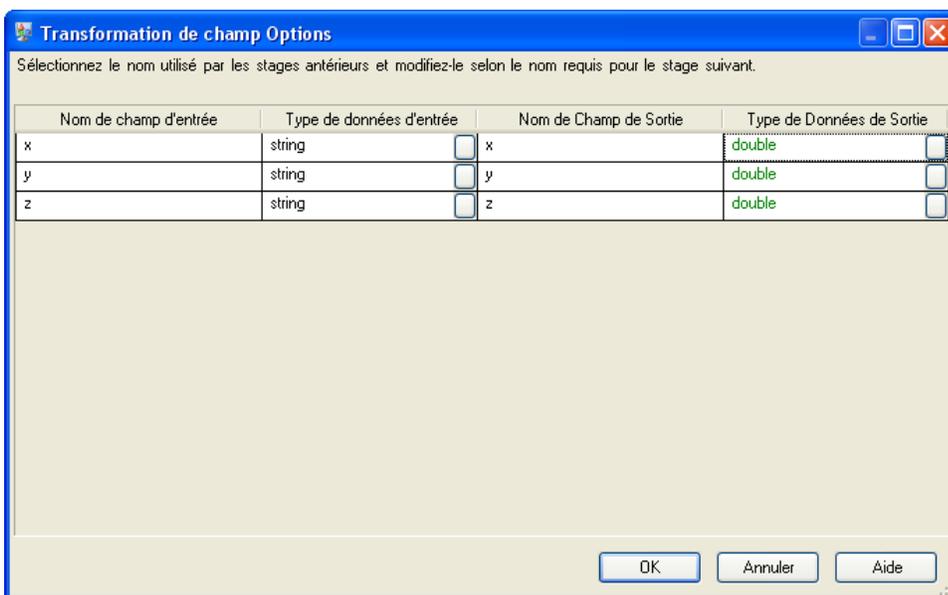
Conversion de type de données automatique

Lorsque les données présentées à un stage ne sont pas d'un type approprié, Spectrum™ Technology Platform peut, dans certains cas, convertir de façon automatique les données au type approprié. Par exemple, Validate Address accepte uniquement les données chaîne comme entrée. Si le champ d'entrée PostalCode est de type entier, Spectrum™ Technology Platform peut convertir le champ automatiquement en chaîne et ainsi traiter avec succès le champ PostalCode. De même, le stage Math ne prend que des données de type numérique. Si les données entrantes sont de type string, Spectrum™ Technology Platform peut convertir ces données en type de données indiqué dans l'onglet Champs du stage Math.

Les conversions de type de données automatiques ont lieu dans les canaux d'un flux de données. Si un canal est en train de convertir correctement un type de données, un point bleu apparaît au milieu du canal :

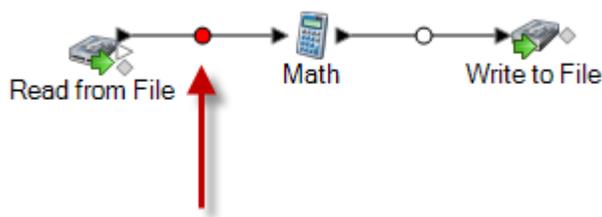


Si vous double-cliquez sur le canal, vous verrez la conversion du type de données qui est en train d'avoir lieu. Dans ce cas, des données de chaîne (string) sont en train d'être converties en données de type entier (integer) :



Remarquez que vous ne pouvez pas modifier les types de données dans cette boîte de dialogue en vue de conversions de types de données automatiques. Le type de données de sortie est déterminé par les réglages définis dans le stage en aval.

Les champs qui ne contiennent pas de valeurs valides ou qui ne peuvent pas être convertis sont encerclés en rouge dans le canal.



Dans les options de conversion de type, vous pouvez indiquer comment le flux de données doit réagir si la conversion de type échoue.

Définition des options de conversion de type de données pour un dataflow

La conversion de type de données se produit quand un dataflow convertit automatiquement un champ de type de données requis par un stage. La conversion de type de données se produit également au sein de certains stages. Par exemple, dans Read From DB, vous pouvez choisir d'avoir un champ utilisant des données de type chaîne, même si les données sources sont de type numérique. Les données sont converties en type de données chaîne lorsqu'elles sont lues dans le dataflow.

Il existe deux paramètres que vous pouvez utiliser pour contrôler les conversions de type de données. Tout d'abord, il existe des paramètres qui déterminent le mode de mise en forme des données numériques, date et heure converties en une chaîne. Par exemple, vous pouvez souhaiter représenter des données date converties en chaîne au format mm/jj/aaaa plutôt que jj/mm/aaaa. L'autre paramètre contrôle ce qui doit se produire si le système ne peut pas convertir un champ d'un type de données à un autre.

Les paramètres de conversion de type de données par défaut de votre système sont définis dans Management Console. Vous pouvez remplacer les formats par défaut des dataflows individuels dans Enterprise Designer.

Cette procédure explique comment remplacer les options de conversion de type de données par défaut d'un dataflow.

Remarque : Les sous-flux héritent des réglages de conversion du type du dataflow dans lequel ils se trouvent. Vous ne pouvez pas spécifier les réglages de conversion de type pour les sous-flux.

1. Ouvrez le dataflow dans Enterprise Designer.
2. Sélectionnez **Éditer > Options de conversion de type**.
3. Cochez la case **Neutralisation des options système par défaut avec les valeurs suivantes**.
4. Dans le champ **Gestion des échecs**, indiquez comment réagir lorsque la valeur d'un champ ne peut pas être automatiquement convertie dans le type de données requis par un stage.

Invalider le dataflow	Si un champ ne peut pas être converti, le dataflow sera invalidé.
Faire échouer l'enregistrement	Si un champ ne peut pas être converti, l'enregistrement échouera mais le dataflow continuera de s'exécuter.
Initialiser le champ en utilisant les valeurs par défaut	Si un champ ne peut pas être converti, la valeur du champ est remplacée par la valeur que vous indiquez ici. Cette option est utile si vous savez que certains enregistrements contiennent des données erronées et que vous souhaitez les remplacer par une valeur par défaut. Spécifiez une valeur pour chaque type de données.

5. Indiquez les formats à utiliser pour les données de date et d'heure qui sont converties en une chaîne. Lorsque la date ou l'heure est convertie en une chaîne, la chaîne sera au format que vous indiquez ici.

- a) Dans le champ **Locale**, sélectionnez le pays dont vous souhaitez utiliser le format pour les dates converties en une chaîne. Votre sélection déterminera les valeurs par défaut dans les champs **Date**, **Heure** et **Date et heure**. Votre sélection déterminera également la langue utilisée lorsqu'un mois est épilé. Par exemple, si vous indiquez Anglais, le premier mois de l'année sera « January », alors que si vous indiquez Français, ce sera « Janvier ».
- b) Dans le champ **Date**, sélectionnez le format à utiliser pour les données de date lorsqu'elles sont converties en une chaîne. Une liste des formats les plus couramment utilisés pour les paramètres régionaux sélectionnés est fournie.

Par exemple, si vous sélectionnez le format **M/J/AA** et qu'un champ de date contient 2012-3-2, ces données de date seront converties en la chaîne 3/2/12.

- c) Dans le champ **Time**, sélectionnez le format à utiliser pour les données d'heure lorsqu'elles sont converties en une chaîne. Une liste des formats les plus couramment utilisés pour les paramètres régionaux sélectionnés est fournie.

Par exemple, si vous sélectionnez le format **h:mm:a** et qu'un champ d'heure contient 23:00, ces données d'heure seront converties en la chaîne 11:00 PM.

- d) Dans le champ **DateTime**, sélectionnez le format à utiliser pour les champs contenant le type de données Date et heure lorsqu'ils sont convertis en une chaîne. Une liste des formats les plus couramment utilisés pour les paramètres régionaux sélectionnés est fournie.

Par exemple, si vous sélectionnez le format **M/j/aa h:mm:a** et qu'un champ DateTime contient 2012-3-2 23:00, les données DateTime seront converties en la chaîne 3/2/12 11:00 PM.

- e) Dans le champ **Whole numbers**, sélectionnez le formatage à utiliser pour les nombres entiers (types de données flottants et doubles).

Par exemple, si vous sélectionnez le format **#,###**, le nombre 4324 sera formaté en 4 324.

Remarque : Si vous laissez ce champ vide, les nombres seront formatés de la même manière que dans Spectrum™ Technology Platform 8.0 et version antérieure. En particulier, aucun séparateur de milliers n'est utilisé, le point (« . ») est utilisé

comme séparateur décimal, les nombres inférieurs à 10^{-3} ou supérieurs ou égaux à 10^7 apparaissent dans la notation scientifique et les nombre négatifs présentent un signe moins (« - ») devant ceux-ci. Si vous laissez ce champ vide, les nombres utilisant le type de données bigdecimal seront toujours au format `#,###.000`.

- f) Dans le champ **Decimal numbers**, sélectionnez le formatage à utiliser pour les nombres contenant une valeur décimale (types de données entier et long).

Par exemple, si vous sélectionnez le format `#,###0.0#`, le nombre 4324.25 sera formaté en `4 324,25`.

Remarque : Si vous laissez ce champ vide, les nombres seront formatés de la même manière que dans Spectrum™ Technology Platform 8.0 et version antérieure. En particulier, aucun séparateur de milliers n'est utilisé, le point (« . ») est utilisé comme séparateur décimal, les nombres inférieurs à 10^{-3} ou supérieurs ou égaux à 10^7 apparaissent dans la notation scientifique et les nombre négatifs présentent un signe moins (« - ») devant ceux-ci. Si vous laissez ce champ vide, les nombres utilisant le type de données bigdecimal seront toujours au format `#,###.000`.

Vous pouvez également spécifier vos propres formats de date, heure et nombre si ceux disponibles ne répondent pas à vos besoins. Pour indiquer votre propre format de date ou heure, tapez le format dans le champ à l'aide de la notation décrite dans [Modèles de date et d'heure](#) à la page 35. Pour indiquer votre propre format numérique, tapez le format dans le fichier à l'aide de la notation décrite dans [Modèles numériques](#) à la page 37.

6. Sous **Gestion Null**, choisissez quoi faire si un champ qui a besoin de conversion de type contient une valeur null. Si vous sélectionnez une des options suivantes, soit le dataflow ou l'enregistrement contenant la valeur null échoue selon que vous avez sélectionné **Échec du dataflow** ou **Échec de l'enregistrement** sous **Échecs de type de conversion**.

Échec chaîne nulle Faire échouer le dataflow ou l'enregistrement si la conversion de type est nécessaire sur un champ de type chaîne qui contient une valeur nulle.

Échec valeur booléenne nulle Faire échouer le dataflow ou l'enregistrement si la conversion de type est nécessaire sur un champ booléen qui contient une valeur null.

Échec valeur numérique nulle Faire échouer le dataflow ou l'enregistrement si la conversion de type est nécessaire sur un champ numérique qui contient une valeur null. Les champs numériques incluent double, flottant, long, entier et champs Big Decimal.

Échec date nulle Faire échouer le dataflow ou l'enregistrement si la conversion de type est nécessaire sur un champ de date qui contient une valeur nulle. Cela inclut les champs date, heure et DateHeure.

Modèles de date et d'heure

Lors de la définition des options de type de données pour les données de date et d'heure, vous pouvez créer votre propre modèle de date ou d'heure personnalisé si les modèles prédéfinis ne répondent pas à vos besoins. Pour créer un modèle de date ou d'heure, utilisez la notation décrite dans la table suivante. Par exemple, ce modèle :

dd MMMM yyyy

Produirait une date comme suit :

14 décembre 2012

Lettre	Description	Exemple
G	Code de l'ère	AD
yy	Année à deux chiffres	96
yyyy	Année à quatre chiffres	1996
M	Mois numérique de l'année.	7
MM	Mois numérique de l'année. Si le nombre est inférieur à 10, un zéro est ajouté pour qu'il s'agisse d'un nombre à deux chiffres.	07
MMM	Nom abrégé du mois	Jul
MMMM	Version longue du nom du mois	Juillet
w	Semaine de l'année	27
ww	Semaine de l'année à deux chiffres. Si la semaine est inférieure à 10, un zéro supplémentaire est ajouté.	06
L	Semaine du mois	2
D	Jour de l'année	189
DDD	Jour de l'année à trois chiffres. Si le numéro contient moins de trois chiffres, des zéros sont ajoutés.	006
d	Jour du mois	10

Lettre	Description	Exemple
dd	Jour du mois à deux chiffres. Les nombres inférieurs à 10 ont un zéro ajouté.	09
F	Jours de la semaine dans le mois	2
E	Nom abrégé du jour de la semaine	Ma
EEEE	Version longue du nom du jour de la semaine	Mardi
a	Marque AM/PM	PM
H	Heure du jour, avec la première heure étant 0 et la dernière heure étant 23.	0
HH	Heure du jour à deux chiffres, la première heure étant 0 et la dernière heure étant 23. Les nombres inférieurs à 10 ont un zéro ajouté.	08
k	Heure du jour, avec la première heure étant 1 et la dernière heure étant 24.	24
kk	Heure du jour à deux chiffres, la première heure étant 1 et la dernière heure étant 24. Les nombres inférieurs à 10 ont un zéro ajouté.	02
k	Heure du matin (AM) ou de l'après-midi (PM), avec 0 étant la première heure et 11 étant la dernière heure.	0
KK	Heure du jour à deux chiffres, la première heure étant 1 et la dernière heure étant 24. Les nombres inférieurs à 10 ont un zéro ajouté.	02
h	Heure du matin (AM) ou de l'après-midi (PM), 1 étant la première heure et 12 étant la dernière heure.	12
hh	Heure à deux chiffres du matin (AM) ou de l'après-midi (PM), 1 étant la première heure et 12 étant la dernière heure. Les nombres inférieurs à 10 ont un zéro ajouté.	09
m	Minute de l'heure	30

Lettre	Description	Exemple
mm	Minutes de l'heure à deux chiffres. Les nombres inférieurs à 10 ont un zéro ajouté.	05
s	Seconde de la minute	55
ss	Secondes de la minute à deux chiffres. Les nombres inférieurs à 10 ont un zéro ajouté.	02
S	Milliseconde de la seconde	978
SSS	Millisecondes de la seconde à trois chiffres. Les nombres contenant moins de trois chiffres ont un ou deux zéros ajoutés pour qu'il s'agisse de nombres à trois chiffres.	978078008
z	Abréviation de l'heure du nom de fuseau horaire. Si le fuseau horaire ne porte aucun nom, décalage GMT.	PSTGMT-08:00
zzzz	Nom complet du fuseau horaire. Si le fuseau horaire ne porte aucun nom, décalage GMT.	Pacific Standard TimeGMT-08:00
Z	Fuseau horaire RFC 822.	-0800
X	Fuseau horaire ISO 8601.	-08Z
XX	Fuseau horaire ISO 8601 avec les minutes.	-0800Z
XXX	Fuseau horaire ISO 8601 avec les minutes et deux-points comme séparateur entre les heures et les minutes.	-08:00Z

Modèles numériques

Lors de la définition des options de type de données pour les données numériques, vous pouvez créer votre propre modèle numérique personnalisé si les modèles prédéfinis ne répondent pas à vos besoins. Un modèle numérique de base se compose des éléments suivants :

- Un préfixe tel qu'un symbole de devise (facultatif)
- Un modèle de nombres contenant un caractère de regroupement facultatif (par exemple, une virgule comme séparateur de milliers)
- Un suffixe (facultatif)

Par exemple, ce modèle :

\$ ###,###.00

Produirait un nombre formaté comme suit (remarquez l'utilisation d'un séparateur de milliers après les trois premiers chiffres) :

232 998,60 \$

Modèles pour les nombres négatifs

Par défaut, les nombres négatifs sont formatés de la même manière que les nombres positifs mais présentent le signe négatif en préfixe. Le caractère utilisé pour le signe numérique dépend des paramètres régionaux. Dans la plupart des paramètres régionaux, le signe négatif est « - ». Par exemple, si vous spécifiez ce modèle numérique :

0.00

Le nombre négatif dix serait formaté comme suit dans la plupart des paramètres régionaux :

-10.00

Néanmoins, si vous souhaitez définir un autre préfixe ou suffixe à utiliser pour les nombres négatifs, indiquez un second modèle, en le séparant du premier modèle à l'aide d'un point virgule (« ; »). Par exemple :

0.00; (0.00)

Dans ce modèle, les nombres négatifs seraient contenus entre parenthèses :

(10.00)

Notation scientifique

Pour formater un nombre dans la notation scientifique, utilisez le caractère `E` suivi d'un nombre minimal de chiffres à inclure dans l'exposant. Par exemple, dans le modèle suivant :

0.###E0

Le nombre 1234 serait formaté comme suit :

1.234E3

En d'autres termes, 1.234×10^3 .

Remarques :

- Le nombre de chiffres après le caractère d'exposant indique le nombre minimal de chiffres exponentiels. Il n'existe pas de valeur maximale.
- Les exposants négatifs sont formatés à l'aide du signe moins localisé, non du préfixe ou du suffixe du modèle.
- Les modèles de notation scientifique ne peuvent pas contenir de séparateurs de regroupement (par exemple, un séparateur de milliers).

Caractères de modèle numérique spéciaux

Les caractères suivants permettent de produire d'autres caractères, plutôt que d'être reproduits littéralement dans le nombre résultant. Si vous souhaitez utiliser un de ces caractères spéciaux comme caractères littéraux dans le préfixe ou suffixe de votre modèle numérique, entourez le caractère spécial de guillemets.

Symbole	Description
0	Représente un chiffre dans le modèle incluant des zéros où cela est nécessaire pour remplir le modèle. Par exemple, lorsqu'il est appliqué au modèle, le nombre vingt-sept 0000 serait : 0027
#	Représente un chiffre mais les zéros sont omis. Par exemple, lorsqu'il est appliqué au modèle, le nombre vingt-sept #### serait : 27
.	Le séparateur décimal ou le séparateur décimal monétaire utilisé dans les paramètres régionaux sélectionnés. Par exemple, aux États Unis, le point (.) est utilisé comme séparateur décimal mais, en France, la virgule (,) est utilisée comme séparateur décimal.
-	Le signe négatif utilisé dans les paramètres régionaux sélectionnés. Pour la plupart des paramètres régionaux, il s'agit du signe moins (-).

Symbole	Description
,	<p>Le caractère de regroupement utilisé dans les paramètres régionaux sélectionnés. Le caractère approprié pour les paramètres régionaux sélectionnés sera utilisé. Par exemple, aux États-Unis, la virgule (,) est utilisée comme séparateur.</p> <p>Le séparateur de regroupement est généralement utilisé pour les milliers mais, dans certains pays, il sépare les dizaines de milliers. La taille de regroupement est un nombre constant de chiffres entre les caractères de regroupement, tel que 3 pour 100,000,000 ou 4 pour 1,0000,0000. Si vous fournissez un modèle avec plusieurs caractères de regroupement, l'intervalle entre le dernier et la fin de l'entier est celui utilisé. Par exemple, tous les modèles suivants produisent le même résultat :</p> <pre># , ## , ### , #### ##### , ##### ## , #### , #####</pre>
E	<p>Sépare la mantisse et l'exposant dans la notation scientifique. Vous ne devez pas entourer le E de guillemets dans votre modèle. Reportez-vous à la section Notation scientifique à la page 38.</p>
;	<p>Sépare les sous-modèles positifs et négatifs. Reportez-vous à la section Modèles pour les nombres négatifs à la page 38.</p>
%	<p>Multipliez le nombre par 100 et affichez le nombre sous forme de pourcentage. Par exemple, le nombre .35, lorsqu'il est appliqué à ce modèle,</p> <pre>##%</pre> <p>produirait le résultat suivant :</p> <pre>35%</pre>
¤	<p>Le symbole de devise pour les paramètres régionaux sélectionnés. S'il est doublé, le symbole de devise international est utilisé. S'il est présent dans un modèle, le séparateur décimal monétaire est utilisé au lieu du séparateur décimal.</p>
'	<p>Utilisé pour mettre des caractères spéciaux entre guillemets dans un préfixe ou un suffixe. Par exemple :</p> <pre>" ' # ' # "</pre> <p>Formate 123 en :</p> <pre>" #123 "</pre> <p>Pour créer un guillemet simple, utilisez-en deux à la suite :</p> <pre>"# o' 'clock"</pre>

Modification du type de données d'un champ

Spectrum™ Technology Platform modifie automatiquement les types de données de champ si nécessaire, à l'aide des paramètres de conversion du type spécifiés dans Management Console ou des options de conversion du type de flux de données spécifiées dans Enterprise Designer. Dans la plupart des situations, vous ne devez pas modifier manuellement les types de données de champ car toute conversion de type de données nécessaire est gérée automatiquement. Néanmoins, dans les cas où un stage ne peut pas convertir les données entrantes dans le type de données nécessaire, vous devrez peut-être modifier manuellement le type de données dans le canal en amont.

Il n'existe que quelques conversions de type possibles que vous pouvez effectuer manuellement. Les voici :

- Les types Polygone et Multi-polygone peuvent être convertis dans et à partir d'un type de géométrie.
- Les types de données de date, d'heure et d'horodatage peuvent être convertis dans et à partir d'un type de chaîne.

Pour modifier manuellement le type de données d'un champ, suivez cette procédure.

1. Dans Enterprise Designer, double-cliquez sur le canal dans lequel vous souhaitez modifier le type de données du champ. Un canal représente la ligne qui relie deux stages sur le canevas.
2. Cliquez sur le bouton en forme de petit carré en regard du type de données à modifier.

Remarque : Si un bouton en forme de petit carré n'est pas visible en regard du type de données, la conversion de type de données manuelle n'est pas disponible pour votre situation.

3. Pour les types de données de date, d'heure et d'horodatage, procédez comme suit :

Remarque : Seules les options appropriées seront affichées en fonction du type de données choisi.

- a) Dans le champ **Locale**, sélectionnez le pays dont vous souhaitez utiliser le format pour les dates converties en une chaîne. Votre sélection déterminera les valeurs par défaut dans les champs **Date**, **Heure** et **Date et heure**. Votre sélection déterminera également la langue utilisée lorsqu'un mois est épelé. Par exemple, si vous indiquez Anglais, le premier mois de l'année sera « January », alors que si vous indiquez Français, ce sera « Janvier ».
- b) Dans le champ **Date**, sélectionnez le format à utiliser pour les données de date lorsqu'elles sont converties en une chaîne. Une liste des formats les plus couramment utilisés pour les paramètres régionaux sélectionnés est fournie.

Par exemple, si vous sélectionnez le format **M/J/AA** et qu'un champ de date contient 2012-3-2, ces données de date seront converties en la chaîne 3/2/12.

- c) Dans le champ **Time**, sélectionnez le format à utiliser pour les données d'heure lorsqu'elles sont converties en une chaîne. Une liste des formats les plus couramment utilisés pour les paramètres régionaux sélectionnés est fournie.

Par exemple, si vous sélectionnez le format **h:mm:a** et qu'un champ d'heure contient 23:00, ces données d'heure seront converties en la chaîne 11:00 PM.

- d) Dans le champ **DateTime**, sélectionnez le format à utiliser pour les champs contenant le type de données Date et heure lorsqu'ils sont convertis en une chaîne. Une liste des formats les plus couramment utilisés pour les paramètres régionaux sélectionnés est fournie.

Par exemple, si vous sélectionnez le format **M/j/aa h:mm:a** et qu'un champ DateTime contient 2012-3-2 23:00, les données DateTime seront converties en la chaîne 3/2/12 11:00 PM.

- e) Dans le champ **Whole numbers**, sélectionnez le formatage à utiliser pour les nombres entiers (types de données flottants et doubles).

Par exemple, si vous sélectionnez le format **#,###**, le nombre 4324 sera formaté en 4 324.

Remarque : Si vous laissez ce champ vide, les nombres seront formatés de la même manière que dans Spectrum™ Technology Platform 8.0 et version antérieure. En particulier, aucun séparateur de milliers n'est utilisé, le point (« . ») est utilisé comme séparateur décimal, les nombres inférieurs à 10^{-3} ou supérieurs ou égaux à 10^7 apparaissent dans la notation scientifique et les nombres négatifs présentent un signe moins (« - ») devant ceux-ci. Si vous laissez ce champ vide, les nombres utilisant le type de données bigdecimal seront toujours au format **#,###.000**.

- f) Dans le champ **Decimal numbers**, sélectionnez le formatage à utiliser pour les nombres contenant une valeur décimale (types de données entier et long).

Par exemple, si vous sélectionnez le format **#,###.0#**, le nombre 4324.25 sera formaté en 4 324,25.

Remarque : Si vous laissez ce champ vide, les nombres seront formatés de la même manière que dans Spectrum™ Technology Platform 8.0 et version antérieure. En particulier, aucun séparateur de milliers n'est utilisé, le point (« . ») est utilisé comme séparateur décimal, les nombres inférieurs à 10^{-3} ou supérieurs ou égaux à 10^7 apparaissent dans la notation scientifique et les nombres négatifs présentent un signe moins (« - ») devant ceux-ci. Si vous laissez ce champ vide, les nombres utilisant le type de données bigdecimal seront toujours au format **#,###.000**.

4. Cliquez sur **OK**.

La couleur du nom de type de données devient verte.

5. Cliquez de nouveau sur **OK** pour enregistrer la modification.

Modification du nom d'un champ

Il existe une variété de situations où vous devrez peut-être renommer un champ dans un flux de données. Par exemple :

- Une entrée de stage exige certains noms de champs, mais la sortie du stage précédent utilise d'autres noms de champ
- Vous souhaitez conserver des données dans un champ quand un stage en aval écrit des données dans un champ au nom identique

Remarque : Une fois un champ renommé, celui-ci n'est plus disponible dans les stages suivants sous son ancien nom.

1. Dans un flux de données, double-cliquez sur le canal entre deux stages. La boîte de dialogue **Option de transformation de champ** apparaît.
2. Modifiez le ou les nom(s) de champ(s) comme vous le voulez.

Par exemple, le dernier stage peut nécessiter « AddressLine3 », mais le stage antérieur utilise « FirmName » à la place. Dans ce cas, cliquez sur la flèche de déroulement dans Nom de champ d'entrée qui correspond à AddressLine3 comme Nom de Champ de Sortie, puis sélectionnez « FirmName ».

La couleur du nom de champ de sortie devient verte.

3. Cliquez sur **OK**.

Noms de champ réservés

Les noms de champ suivants sont utilisés par le système. Vous ne devez donc pas les utiliser dans un flux de données.

Status
Status.Code
Status.Description

Stages de contrôle

Utilisez des stages de contrôle pour déplacer des données sur différents chemins d'accès au sein d'un flux, scinder ou regrouper des enregistrements et réaliser des transformations et des opérations mathématiques de base sur les données.

Aggregator

Aggregator convertit les données texte en données hiérarchiques. Il prend des données d'entrée d'une source unique, crée un schéma (une hiérarchie de données structurée) en regroupant les données en fonction des champs que vous indiquez, puis construit les groupes dans le schéma.

Remarque : Si vos données contiennent un champ avec lequel vous grouperez vos données, par exemple un champ ID, vous devez trier vos données avant de l'exécuter via Aggregator. Vous pouvez effectuer cette opération en triant vos données avant de les introduire dans le flux de données, en triant le fichier d'entrée dans Enterprise Designer (pour les jobs et sous-flux, mais pas les services) ou en ajoutant un stage Sorter (trieur) à votre flux de données (pour les jobs, les services ou les sous-flux).

Grouper par

Choisissez le champ à utiliser comme base pour l'agrégation dans une hiérarchie en sélectionnant **Grouper par** dans l'arborescence et en cliquant sur **Ajouter**. Les enregistrements disposant de la même valeur dans le champ de votre choix verront leurs données agrégées dans une seule hiérarchie. Si vous sélectionnez plusieurs champs, les données de tous les champs doivent correspondre afin de regrouper les enregistrements dans une hiérarchie.

Par exemple, si vous souhaitez regrouper les données par numéro de compte, vous devez sélectionner le champ du numéro de compte. Tous les enregistrements entrants disposant de la même valeur dans le champ du numéro de compte verront leurs données regroupées dans un seul enregistrement hiérarchique.

Remarque : Vous devez connecter un stage au port d'entrée de Aggregator pour qu'une liste de champs soit disponible pour la sélection.

Listes de résultats

Les champs que vous sélectionnez sous **Listes de résultats** déterminent les champs qui sont inclus dans chaque enregistrement créé par Aggregator. Pour ajouter un champ, sélectionnez **Listes de résultats** puis cliquez sur **Ajouter** et choisissez l'une des options suivantes :

Champ existant	Sélectionnez cette option si vous souhaitez ajouter un champ du flux de données à la hiérarchie.
Nouveau type de données	Sélectionnez cette option si vous souhaitez créer un champ parent dans lequel vous pouvez ensuite ajouter des champs enfant.
Modèle	Cette option vous permet d'ajouter un champ en fonction des données du stage connecté au port de sortie de Aggregator.

Si vous souhaitez que le champ dispose de champs enfant, cochez la case **Liste**.

Saisissez le nom du champ dans la zone de texte **Nom**, ou ne la modifiez pas si celle-ci s'est remplie automatiquement et si le nom vous convient. Gardez à l'esprit que le stage Aggregator n'accepte

pas les caractères XML invalides dans les noms de champ ; mais les caractères alphanumériques sont permis, tels que le point (.), le trait de soulignement (_), et le tiret (-).

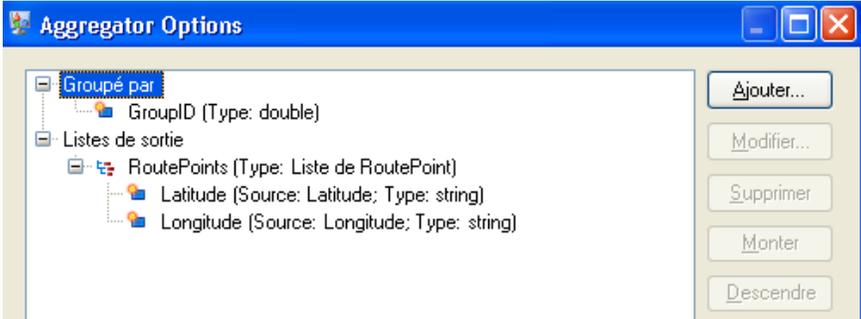
Cliquez sur **Ajouter** pour ajouter le champ. Vous pouvez spécifier un autre champ à ajouter au même niveau dans la hiérarchie ou vous pouvez cliquer sur **Fermer**.

Pour ajouter des champs enfant à un champ existant, sélectionnez le champ parent, puis cliquez sur **Ajouter**.

Remarque : Vous pouvez modifier le groupe de champs en surlignant une ligne et en cliquant sur **Modifier**, et vous pouvez supprimer un groupe de champs en surlignant une ligne et en cliquant sur **Supprimer**. Vous pouvez également changer l'ordre des champs en cliquant sur un champ puis en cliquant sur **Déplacer vers le haut** ou **Déplacer vers le bas**.

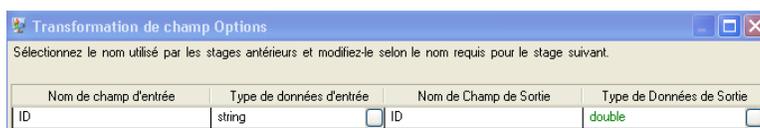
Exemple de Aggregator

Un exemple de la fonction de Aggregator consiste à prendre un groupe d'adresses postales et à les transformer en itinéraires routiers. Vous pouvez réaliser ceci à l'aide de deux points : un point de départ et un point d'arrivée, ou bien vous pouvez placer plusieurs points le long d'un itinéraire. Le flux de données correspondant à ce type de fonction pourrait ressembler à ceci :

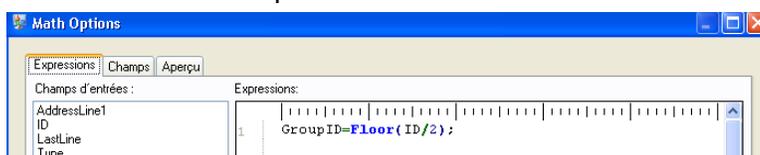


Le flux de données exécute la fonction de la manière suivante :

1. Le stage **Read from File** contient les adresses postales dans un fichier plat. Les champs de ce fichier sont les suivants :
 - un **ID** (Identifiant), qui identifie une adresse particulière dans le fichier.
 - un **Type**, qui indique si l'adresse est une adresse « Destinataire » ou « Expéditeur ».
 - un champ **AddressLine1**, qui fournit la rue de l'adresse.
 - un champ **LastLine**, qui comprend des informations telles que la ville, l'État, et/ou le code postal.
2. La **Transformation de champ** entre le stage Read from File et le stage Math modifie le format du champ ID de chaîne en double parce que le stage Math n'accepte pas les données chaîne.



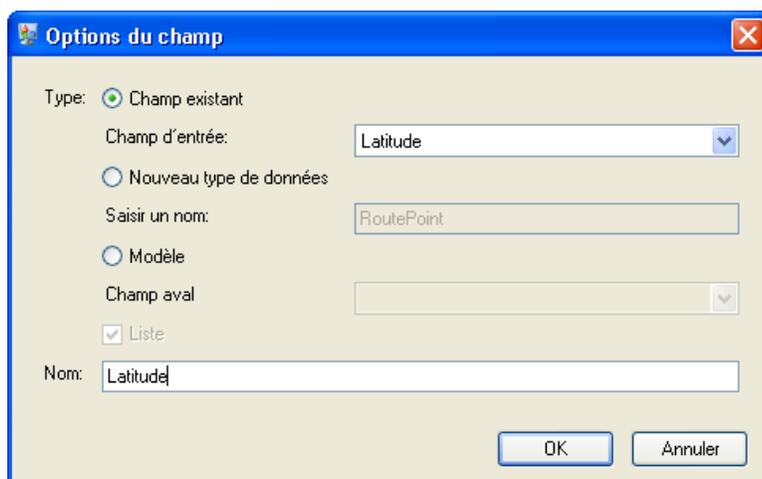
3. Le stage **Math** génère une expression qui établit un champ Group ID (Identifiant du groupe) à utiliser en aval dans le flux de données. Dans cet exemple, celui-ci calcule le Group ID comme un plancher (floor) de, ou un arrondi au chiffre inférieur de, la valeur du champ ID divisé par 2. Par conséquent, pour un ID de 3, l'expression sera $3/2$, qui est égal à 1,5. Lorsque vous arrondissez 1,5 au chiffre inférieur, vous obtenez 1. Si l'identifiant est 2, l'expression sera de $2/2$, égal à 1, et il n'y aura donc nul besoin d'arrondir au chiffre inférieur. Donc, les ID de 2 et 3 ont un même Group ID de 1.



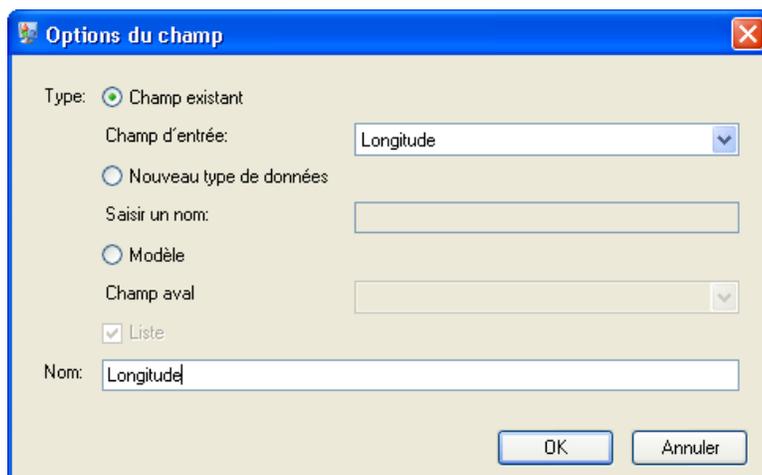
4. **Geocode US Address** obtient les latitudes et les longitudes de chaque adresse.
5. Le stage **Aggregator** établit le regroupement des données par le champ GroupID (identifiant du groupe) et les listes de sortie doivent inclure les points de l'itinéraire composés de leurs latitudes et longitudes.

Les instructions ci-dessous montrent comment configurer **manuellement** le stage Aggregator pour ce flux de données.

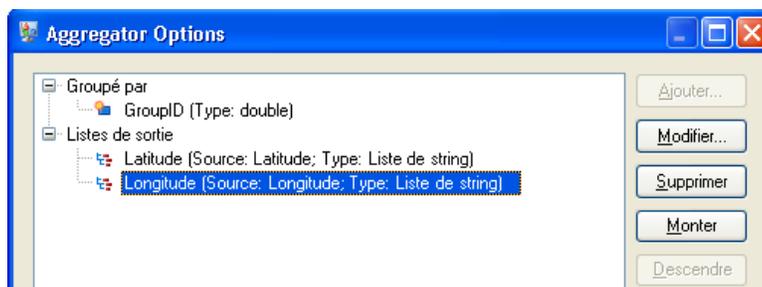
- Double-cliquez sur le stage Aggregator puis sur **Groupé par**.
- Sélectionnez le champ **GroupID** et cliquez sur **OK**. Ce champ vous permet d'inclure des points d'itinéraire (Route Points) pour les prochains stages du flux de données. Les points d'itinéraire sont des éléments essentiels pour un flux de données qui produit des itinéraires.
- Double-cliquez sur **Listes de sortie**. La boîte de dialogue **Options du champ** s'affiche.
- Sélectionnez **Nouveau type de données**. Dans le champ **Type name**, entrez `RoutePoint`. Dans le champ **Name**, entrez `RoutePoints`. Par défaut, il s'agit d'une liste qui ne peut être modifiée, donc la case à cocher est grisée.
- Appuyez sur **OK**.
- Cliquez sur **RoutePoints** (points d'itinéraire) et cliquez sur **Ajouter**. La boîte de dialogue **Options du champ** s'affiche à nouveau.
- Les points d'itinéraires sont composés de latitudes et longitudes, c'est pourquoi nous devons d'abord ajouter un **Champ existant** à partir du champ d'entrée existant `Latitude`. Le champ **Name** se remplit de manière automatique.



Répétez cette étape pour la Longitude.



Le stage Aggregator une fois terminé ressemblera à ceci :



6. **Get Travel Directions** fournit des itinéraires voitures à partir des ID de points 0, 2, et 4 jusqu'aux ID de points 1, 3, et 5, respectivement.
7. Le stage **Splitter** établit la division des données du champ Itinéraire, et inclut tous les champs possibles du stage Get Travel Directions dans les listes de sortie.

8. Le stage **Write to File** écrit les directions dans un fichier de sortie.

Broadcaster

Un Broadcaster prend un flux d'enregistrements et le divise en plusieurs flux, ce qui vous permet d'enregistrer plusieurs stages pour un traitement simultané.

Il n'y a pas de paramètres à modifier pour le Broadcaster.

Conditional Router

Le stage **Conditional Router** envoie les enregistrements à différents chemins d'accès du flux de données, en fonction des critères que vous indiquez. Suivant les critères définis, le stage peut comporter un ou plusieurs ports de sortie. Les ports de sortie sont numérotés de manière consécutive, en commençant par 1 (qui s'affiche comme « port »).

Les ports de sortie sont connectés à différents stages auxquels les données sont envoyées, suivant les conditions définies. Par exemple, vous pouvez envoyer un ensemble d'enregistrements au port 1 en cas de correspondance réussie, tandis qu'un autre ensemble d'enregistrements peut être envoyé au port 2 en cas d'échec de correspondance.

Un enregistrement d'entrée est écrit sur le port de sortie de Conditional Router uniquement si l'expression toute entière est évaluée sur true.

Configuration d'un Conditional Router

1. Sous **Stages de contrôle**, cliquez sur **Conditional Router** et faites-le glisser jusqu'au canevas en le plaçant à l'emplacement de votre choix dans le flux de données.
2. Connectez le routeur à d'autres stages du canevas.

Remarque : Cette étape est obligatoire avant la définition des paramètres des ports. Sinon, les ports ne peuvent pas être modifiés.

3. Double-cliquez sur le stage **Conditional Router** du canevas. La fenêtre **Options du Conditional Router** apparaît.
4. Cliquez sur le bouton carré de la colonne **Condition/Expression** en regard de la ligne de **port**. La fenêtre **Éditeur d'expressions** apparaît.
5. Dans la section **Choisir le type d'expression**, sélectionnez l'une des options suivantes :
 - **Expression créée à l'aide du générateur d'expressions** : sélectionnez cette option pour créer une expression de base dans laquelle vous pouvez ajouter des Groupes et des

Expressions qui peuvent être combinés à l'aide de différents opérateurs logiques. Pour plus d'informations, reportez-vous à la section **Utilisation d'Expression Builder** à la page 49.

- **Expression personnalisée** : sélectionnez cette option pour écrire une expression à l'aide du langage de script Groovy. Pour plus d'informations, reportez-vous à la section **Écriture d'une expression personnalisée** à la page 53.
- **Expression par défaut** : sélectionnez cette option pour acheminer les enregistrements vers ce port par défaut. Les enregistrements ne correspondant à aucune des expressions des autres ports seront routés vers ce port. Vous devriez toujours disposer d'un port de sortie avec « défaut » comme expression, afin d'être sûr qu'il ne manque aucune ligne en cas de port non concordant et que toutes les lignes sont écrites depuis le routeur.

6. Cliquez sur **OK**. La fenêtre **Éditeur d'expressions** se ferme.
7. Cliquez sur **OK** dans la fenêtre **Options du Conditional Router**.

Utilisation d'Expression Builder

Expression Builder, du stage **Conditional Router**, vous permet de créer une expression qui doit être évaluée sur true pour permettre l'acheminement d'un enregistrement d'entrée jusqu'au port de sortie du stage.

1. Chaque groupe parent se compose d'une combinaison conditionnelle souhaitée d'expressions enfants et de groupes enfants.
2. Chaque expression se compose d'un opérande gauche, d'un opérande droit et d'un opérateur logique.
3. Chaque groupe doit spécifier s'il faut que *l'ensemble* ou *l'une* de ses conditions constituantes soit définies sur true pour que le groupe tout entier soit évalué sur true.

Pour créer une expression à l'aide d'Expression Builder :

1. Dans **Expression Editor**, sélectionnez l'option **Expression créée avec Expression Builder**. Par défaut, l'option Expression Builder est sélectionnée et un groupe parent est affiché dans l'arborescence de hiérarchie des expressions à gauche de la section **Expression Builder**.
2. Pour ajouter un groupe enfant au sein du groupe sélectionné, cliquez sur **Ajouter un groupe**. Ce nouveau groupe est ajouté comme enfant du groupe parent et il est sélectionné par défaut dans l'arborescence. Au sein de chaque groupe, vous pouvez ajouter des expressions enfants et des groupes enfants.
3. Pour chaque groupe, sélectionnez soit **Toutes vraies** ou **Au moins une** sous l'en-tête **Combiner méthode d'expression**.
 - **Toutes vraies** : le groupe est évalué sur true uniquement si tous les critères enfants du groupe sont définis sur true.
 - **Au moins une** : le groupe est évalué sur true si ne serait-ce qu'un de ses critères enfants est défini sur true.
4. Pour ajouter une expression au groupe sélectionné, cliquez sur **Ajouter une expression**. La nouvelle expression est ajoutée comme enfant du groupe parent et elle est sélectionnée par défaut dans l'arborescence.

Pour définir cette expression enfant :

- a) Spécifiez l'opérande gauche de l'expression sélectionnée à l'aide de la liste déroulante **Champ** pour sélectionner l'une des colonnes du fichier d'entrée.
- b) Spécifiez l'opérateur logique connectant les deux composants de l'expression sélectionnée en sélectionnant l'opérateur approprié dans le champ **Opérateur**, comme indiqué ci-dessous :

Tableau 1 : Opérateurs d'Expression Builder

Opérateur	Description
Egal à	Vérifie si la valeur du champ correspond à la valeur ou au champ indiqué.
n'est pas égal à	Vérifie si la valeur du champ ne correspond pas à la valeur ou au champ indiqué.
Est nul	Vérifie si le champ est une valeur nulle.
N'est pas nul	Vérifie si le champ n'est pas une valeur nulle.
Est vide	Vérifie si le champ est null ou s'il s'agit d'une chaîne avec une longueur de 0. Remarque : Cette opération est disponible uniquement pour les champs comportant des données de type chaîne.
N'est pas vide	Vérifie si le champ n'est pas renseigné, et n'est pas une chaîne avec une longueur de 0. Remarque : Cette opération est disponible uniquement pour les champs comportant des données de type chaîne.
Est inférieur à	Vérifie si le champ a une valeur numérique inférieure à la valeur indiquée. Cet opérateur fonctionne sur des types de données numériques ainsi que sur des champs de chaînes contenant des nombres. Remarque : Cette opération n'est pas disponible pour les champs comportant des données de type booléen.

Opérateur	Description
est inférieur ou égal à	<p>Vérifie si le champ a une valeur numérique inférieure ou égale à la valeur indiquée. Cet opérateur fonctionne sur des types de données numériques ainsi que sur des champs de chaînes contenant des nombres.</p> <p>Remarque : Cette opération n'est pas disponible pour les champs comportant des données de type booléen.</p>
Est supérieur à	<p>Vérifie si le champ a une valeur numérique supérieure à la valeur indiquée. Cet opérateur fonctionne sur des types de données numériques ainsi que sur des champs de chaînes contenant des nombres.</p> <p>Remarque : Cette opération n'est pas disponible pour les champs comportant des données de type booléen.</p>
est supérieur ou égal à	<p>Vérifie si le champ a une valeur numérique supérieure ou égale à la valeur indiquée. Cet opérateur fonctionne sur des types de données numériques ainsi que sur des champs de chaînes contenant des nombres.</p> <p>Remarque : Cette opération n'est pas disponible pour les champs comportant des données de type booléen.</p>
Commence par	<p>Vérifie si le champ commence par les caractères indiqués.</p> <p>Remarque : Cette opération est disponible uniquement pour les champs comportant des données de type chaîne.</p>
Ne commence pas par	<p>Vérifie si le champ ne commence pas par les caractères indiqués.</p> <p>Remarque : Cette opération est disponible uniquement pour les champs comportant des données de type chaîne.</p>
Contient	<p>Vérifie si le champ contient la chaîne spécifiée.</p> <p>Remarque : Cette opération est disponible uniquement pour les champs comportant des données de type chaîne.</p>
Ne contient pas	<p>Vérifie si le champ ne contient pas la chaîne indiquée.</p> <p>Remarque : Cette opération est disponible uniquement pour les champs comportant des données de type chaîne.</p>

Opérateur	Description
Se termine par	Vérifie si le champ se termine par les caractères indiqués. Remarque : Cette opération est disponible uniquement pour les champs comportant des données de type chaîne.
Ne se termine pas par	Vérifie si le champ se termine par les caractères indiqués. Remarque : Cette opération est disponible uniquement pour les champs comportant des données de type chaîne.
correspond à l'expression régulière	Met le champ en correspondance avec une expression régulière afin d'identifier des chaînes de texte pertinentes, telles que des caractères spécifiques, des mots ou des schémas de caractères. Le champ de valeur doit contenir un schéma d'expression régulière valide. Remarque : Cette opération est disponible uniquement pour les champs comportant des données de type chaîne.

c) Spécifiez l'opérande droit de l'expression sélectionnée en sélectionnant `Valeur` ou `Champ`.

- `Valeur` : l'opérande gauche de l'expression sélectionnée est comparé à cette valeur.
- `Champ` : l'opérande gauche de l'expression sélectionnée est comparé à cette colonne du même fichier d'entrée. Dans la liste déroulante, sélectionnez la colonne de l'opérande droit.

5. Pour ajouter une expression enfant ou un groupe enfant à une entité, sélectionnez l'entité dans l'arborescence, puis cliquez sur **Ajouter une expression** ou **Ajouter un groupe**, respectivement.
6. Pour déplacer une expression enfant ou un groupe enfant d'un groupe parent vers un autre groupe parent, faites-le glisser vers l'en-tête de groupe parent de votre choix dans l'arborescence de critères à gauche.
7. Répétez les étapes ci-dessus pour ajouter autant d'expressions enfants et de groupes enfants que nécessaire pour créer les critères d'expressions finaux souhaités.
8. Cliquez sur **OK**.

La colonne **Condition/Expression** de la fenêtre **Options du Conditional Router** affiche les critères d'expressions définis, qui doivent être évalués sur true pour qu'un enregistrement soit écrit sur le port de sortie correspondant du stage.

Écriture d'une expression personnalisée

Vous pouvez écrire vos propres expressions personnalisées pour contrôler la manière dont Conditional Router achemine les enregistrements à l'aide du langage de script Groovy pour créer une expression.

Utilisation de la méthode de script Groovy

Pour obtenir des informations sur Groovy, consultez l'adresse suivante : groovy-lang.org.

Les expressions Groovy utilisées dans le stage Conditional Router doivent être évaluées par rapport à une valeur booléenne (true ou false indiquant si l'enregistrement doit être écrit vers le port. L'enregistrement est routé vers le premier port de sortie dont l'expression est évaluée comme étant vraie.

Si, par exemple, vous devez router des enregistrements avec un niveau de confiance de validation ≥ 85 vers un stage et des enregistrements avec un niveau de confiance de validation < 85 vers un autre stage, votre script ressemblera à cela :

```
data['Confidence']>=85
```

Le script de l'autre port ressemblerait à cela :

```
data['Confidence']<85
```

Le routeur évaluerait la valeur du champ Confidence par rapport selon votre critère pour déterminer sur quel port de sortie il doit envoyer.

Vérification qu'un champ contient une seule valeur

Cet exemple est évalué sur true si le champ Status contient la lettre « F ». Il faut que ce soit une correspondance exacte, par conséquent « f » ne serait pas évalué sur true.

```
return data['Status'] == 'F';
```

Vérification qu'un champ contient plusieurs valeurs

Cet exemple est évalué sur true si le champ Status contient la lettre « F » ou « f ».

```
boolean returnValue = false;
if (data['Status'] == 'F' || data['Status'] == 'f')
{
    returnValue = true;
}
return returnValue;
```

Évaluation de la longueur d'un champ

Cet exemple est évalué sur true si le champ PostalCode contient plus de 5 caractères.

```
return data['PostalCode'].length() > 5;
```

Vérification qu'un caractère se trouve dans une valeur de champ

Cet exemple est évalué sur true si le champ PostalCode contient un tiret.

```
boolean returnValue = false;
if (data['PostalCode'].indexOf('-') != -1)
{
    returnValue = true;
}
return returnValue;
```

Directives de script

1. Les noms de colonne doivent être encadrés par des guillemets simples ou doubles.

Par exemple, cette syntaxe est incorrecte, car le nom de colonne `PostalCode` n'est pas encadré par des guillemets simples ou doubles.

```
return data[PostalCode];
```

2. Le nom de colonne est obligatoire.

Par exemple, cette syntaxe est incorrecte, car aucune colonne n'est indiquée.

```
return data[];
```

3. Une instruction `return` doit renvoyer une valeur `boolean`.

Par exemple, ce script est incorrect, car `row.set('PostalCode', '88989')` ne renvoie pas de valeur `boolean`. Il définit juste la valeur du champ `PostalCode` sur `88989`.

```
return row.set('PostalCode', '88989');
```

4. Employez un seul signe égal (=) pour définir la valeur d'un champ, et deux signes égal (==) pour vérifier la valeur d'un champ.

Group Statistics

Le stage Group Statistics vous permet d'effectuer des opérations de statistique sur plusieurs lignes de données divisées dans des groupes que vous souhaitez analyser. Si aucun groupe n'est défini, toutes les lignes seront considérées comme appartenant à un groupe.

Nous appelons un groupe un ou plusieurs champs partageant la même valeur sur de multiples rangées de données.

Par exemple, les données de la table suivante pourraient être regroupées par région, état, ou les deux.

Région	État
Est	MD
Est	MD
Est	CT
Ouest	CA
Ouest	CA

Un regroupement par Région produirait les groupes Est et Ouest. Un regroupement par État donnerait Californie (CA), Connecticut (CT) et Maryland (MD). Un regroupement par Région et par État donnerait Est/Maryland, Est/Connecticut, et Ouest/Californie.

Entrée

Le stage Group Statistics stage prend tous les champs comme entrée. Les regroupements peuvent être réalisés sur des données numériques ou des chaînes.

Options

Tableau 2 : Onglet Opérations

Option	Description
Champs d'entrée	Répertorie les champs dans le dataflow que vous pouvez utiliser pour regrouper les enregistrements et effectuer des calculs.

Option	Description
Row	<p>Indique le ou les champs que vous souhaitez utiliser en tant que catégories pour les calculs. Par exemple, si vous disposez de données qui comprennent un champ Région et que vous voulez calculer la population totale par région, vous devez grouper le champ Région.</p> <p>Pour ajouter un champ, sélectionnez le champ dans la liste Champs d'entrée puis cliquez sur >>.</p>
Colonne	<p>Facultatif. Pour créer un tableau croisé dynamique, indiquez le ou les champs des valeurs que vous souhaitez ajouter dans les colonnes des tableaux croisés.</p> <p>Pour ajouter un champ, sélectionnez le champ dans la liste Champs d'entrée puis cliquez sur >>.</p> <p>Par exemple, si vous disposez de données qui comprennent des régions et des dates d'expéditions, et que vous voulez calculer le nombre d'expéditions par jour pour chaque état, vous devez spécifier le champ d'état sous forme de ligne et le champ de date d'expédition sous forme de colonne.</p>
Les lignes et les colonnes sont pré-triées dans l'ordre configuré	<p>Indique que les données d'entrée sont déjà triées.</p> <p>Si cette case est cochée, le stage ne trie pas les données et effectue l'opération spécifiée directement sur les données d'entrée.</p>
Opération	<p>Indique le calcul à réaliser sur chaque groupe. Pour ajouter une opération, sélectionnez le champ dans la liste Champs d'entrée que vous souhaitez utiliser pour l'opération puis cliquez sur >>.</p> <p>Pour plus d'informations sur les opérations de Group Statistics prises en charge, reportez-vous à la section Opérations à la page 58.</p>

Option	Description
Type	Pour les champs d'entrée et de sortie, indique le type de données. <ul style="list-style-type: none"> Entier Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre -2^{31} (-2,147,483,648) et $2^{31}-1$ (2,147,483,647) Long Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre -2^{63} (-9,223,372,036,854,775,808) et $2^{63}-1$ (9,223,372,036,854,775,807) Flottant Un type de données numérique contenant des nombres de précision uniques positifs entre 2^{-149} (1.4E-45) et $(2-2^{23})\times 2^{127}$ (3.4028235E38) Double Un type de données numérique contenant des nombres de précision doubles positifs et négatifs entre 2^{-1074} (4.9E-324) et $(2-2^{-52})\times 2^{1023}$ (1.7976931348623157E308) <p>Remarque : Lors de l'utilisation du types entier et long, des données peuvent être perdues si le numéro d'entrée ou le numéro calculé depuis une opération contient des données décimales.</p>
Obtenir le nombre d'enregistrements calculés	Indique le nombre réel d'enregistrements d'un groupe sur lequel l'opération est réalisée. Cette colonne <i>Computational Count</i> exclut les enregistrements d'entrée dont la colonne sur laquelle l'opération est réalisée contient des valeurs <code>null</code> .

Onglet Champs

L'onglet Champs est utilisé lors de la création d'un tableau croisé dynamique. Pour plus d'informations, reportez-vous à la section [Création d'un tableau croisé dynamique](#) à la page 65.

Onglet Sortie

Option	Description
Affiche une ligne par groupe	Pour chaque groupe de lignes, renvoie une seule ligne qui contient les données agrégées pour toutes les lignes dans le groupe. Les lignes individuelles seront ignorées. Si cette option n'est pas sélectionnée, toutes les lignes seront renvoyées. Aucune donnée ne sera ignorée. Cette option n'est pas disponible si vous utilisez les opérations Rang de Pourcentages ou Zscore.
Renvoyer un nombre de lignes dans chaque groupe	Renvoie le nombre de lignes dans chaque groupe. Le nom du champ de sortie par défaut de sortie qui contiendra le compte est GroupCount.

Option	Description
Renvoie un ID unique pour chaque groupe	Renvoie un ID unique pour chaque groupe de lignes. L'ID commence à partir de 1 et augmente de 1 pour chaque groupe supplémentaire trouvé. Le nom de champ par défaut est GroupID.

Opérations

Les calculs disponibles sont les suivants :

Moyenne	Pour chaque groupe, calcule la valeur moyenne d'un champ donné. Par exemple, si vous disposez d'un groupe d'enregistrements avec des valeurs 10, 12, 1, et 600 dans un champ donné, la valeur moyenne de ce champ pour ce groupe est 155,75, calculé comme suit $(10 + 12 + 1 + 600) \div 4$.
Maximum	Pour chaque groupe, renvoie la plus grande valeur dans un champ donné. Par exemple, si vous disposez d'un groupe d'enregistrements avec des valeurs 10, 12, 1, et 600 dans un champ donné, la valeur maximale de ce champ pour ce groupe est 600.
Minimum	Pour chaque groupe, renvoie la plus petite valeur dans un champ donné. Par exemple, si vous disposez d'un groupe d'enregistrements avec des valeurs 10, 12, 1, et 600 dans un champ donné, la valeur minimale de ce champ pour ce groupe est 1.
Rang de Pourcentages	Pour chaque enregistrement dans un groupe, calcule le rang de centile d'une valeur dans un champ donné par rapport à d'autres enregistrements dans le groupe. Le rang de centile représente le pourcentage d'enregistrements dans le groupe avec des valeurs inférieures dans le champ.
Centile	Pour chaque groupe, calcule la valeur qui représente le centile spécifié (0 - 100) pour un champ donné. Un centile représente le pourcentage d'enregistrements qui disposent d'un score inférieur. Par exemple, si vous disposez d'un groupe d'enregistrements avec les valeurs 22, 26, et 74, et que vous effectuez le calcul des centiles en précisant le 60e centile, l'opération renvoie 35,6. Cela signifie qu'un enregistrement avec une valeur de 35,6 dans le champ donné est dans le 60e centile d'enregistrements dans le groupe.
Écart type	Pour chaque groupe, calcule l'écart-type pour un champ donné. L'écart-type mesure la quantité de dispersion au sein du groupe. Plus l'écart-type est faible, plus les valeurs sont centrées autour de la valeur moyenne, et donc la moins dispersées des valeurs. Plus la valeur est élevée, plus les valeurs sont largement dispersées. L'écart-type est exprimé dans les mêmes unités que les données. L'écart-type est la racine carrée de la variance.
Somme	Pour chaque groupe, calcule la somme des valeurs d'un champ donné.

Variance	Pour chaque groupe, calcule la variance pour un champ donné. La variance mesure la quantité de dispersion au sein du groupe. Il s'agit du carré de l'écart-type.
ZScore	Pour chaque enregistrement dans un groupe, retourne le Zscore. Le Zscore indique la quantité d'écarts-types d'une valeur supérieure ou inférieure à la moyenne du groupe.
Alphabétique d'abord	Pour chaque groupe, renvoie la première valeur du dictionnaire. Si plusieurs valeurs de champ ont la même longueur ou figurent au même endroit du dictionnaire, renvoie la première occurrence de cette valeur. Par exemple, si un groupe comporte un enregistrement avec les valeurs Joel et Joey dans un champ, la première valeur alphabétique du groupe est Joel, car l est avant y dans l'alphabet.
Alphabétique en dernier	Pour chaque groupe, renvoie la dernière valeur du dictionnaire. Si plusieurs valeurs de champ ont la même longueur ou figurent au même endroit du dictionnaire, renvoie la dernière occurrence de cette valeur. Par exemple, si un groupe comporte un enregistrement avec les valeurs Joel et Joey dans un champ, la dernière valeur alphabétique du groupe est Joey, car y est après l dans l'alphabet.
Le plus long	Pour chaque groupe, renvoie la valeur la plus longue. Par exemple, si un groupe comporte un enregistrement avec les valeurs Joel et Jacob dans un champ, la valeur la plus longue du groupe est Jacob, car elle compte cinq lettres, alors que Joel n'en compte que quatre.
Le plus court	Pour chaque groupe, renvoie la valeur la plus courte. Par exemple, si un groupe comporte un enregistrement avec les valeurs Joel et Jacob dans un champ, la valeur la plus courte du groupe est Joel, car elle ne compte que quatre lettres, alors que Jacob en compte cinq.
Le plus récent	Pour chaque groupe, renvoie la valeur date/datetime la plus récente. Par exemple, si un groupe comporte un enregistrement avec les valeurs 15-12-2014 et 24-12-2014 dans un champ, la valeur la plus récente du groupe est 24-12-2014.
Le plus tôt	Pour chaque groupe, renvoie la valeur date/datetime la plus ancienne. Par exemple, si un groupe comporte un enregistrement avec les valeurs 15-12-2014 et 24-12-2014 dans un champ, la valeur la plus ancienne du groupe est 15-12-2014.

Colonnes de sortie

Nom du champ	Description/Valeurs valides				
<Operation>Of<InputFieldName>	contient le résultat d'un calcul. Group Statistics crée un champ de sortie par opération et nomme le champ en fonction de l'opération et du champ. Par exemple, le nom de champ par défaut d'une opération <code>Sum</code> effectuée sur un champ nommé <code>Population</code> est <code>SumOfPopulation</code> .				
<Value>_<Operation>	Contient le résultat d'un tableau croisé dynamique, où <Value> est l'une des valeurs d'une colonne du tableau et <Operation> est l'opération effectuée sur la colonne. Pour plus d'informations, reportez-vous à la section Création d'un tableau croisé dynamique à la page 65.				
Compter groupe	Indique le nombre d'enregistrements dans le groupe.				
GroupID	Un numéro unique attribué à chaque groupe de manière séquentielle. Le premier groupe a une valeur GroupID de 1, le second a une valeur de 2, et ainsi de suite.				
ComputationalCount<Operation> Of<InputFieldName>	Indique le nombre réel d'enregistrements d'un groupe sur lequel l'opération est réalisée. Par exemple, pour l'opération <code>Average</code> réalisée sur la colonne <code>Salary</code> , la colonne <code>ComputationalCountAverageOfSalary</code> est générée.				
État	Il signale la réussite ou l'échec des calculs de Group Statistics. <table border="0"> <tr> <td>null</td> <td>Réussie</td> </tr> <tr> <td>F</td> <td>Échec</td> </tr> </table>	null	Réussie	F	Échec
null	Réussie				
F	Échec				
Status.Code	Raison de l'échec, le cas échéant. Les codes d'état disponibles sont les suivants : <table border="0"> <tr> <td>UnableToDoGroupStatistics</td> <td>Le stage Group Statistics n'a pas été en mesure d'effectuer les calculs.</td> </tr> <tr> <td>Erreur de calcul de la valeur centile</td> <td>La valeur centile n'a pas pu être calculée en utilisant les données d'entrée fournies.</td> </tr> </table>	UnableToDoGroupStatistics	Le stage Group Statistics n'a pas été en mesure d'effectuer les calculs.	Erreur de calcul de la valeur centile	La valeur centile n'a pas pu être calculée en utilisant les données d'entrée fournies.
UnableToDoGroupStatistics	Le stage Group Statistics n'a pas été en mesure d'effectuer les calculs.				
Erreur de calcul de la valeur centile	La valeur centile n'a pas pu être calculée en utilisant les données d'entrée fournies.				

Nom du champ	Description/Valeurs valides
Status.Description	<p>Une description textuelle de l'erreur.</p> <p>La valeur du champ d'entrée n'a pas pu être convertie en type de champ. Elle doit être en dépassement !</p> <p>Un nombre dans un champ d'entrée est plus grand que le nombre autorisé par le type de données. Essayez la conversion vers un type de données qui prend en charge de plus grands nombres, comme le double.</p>

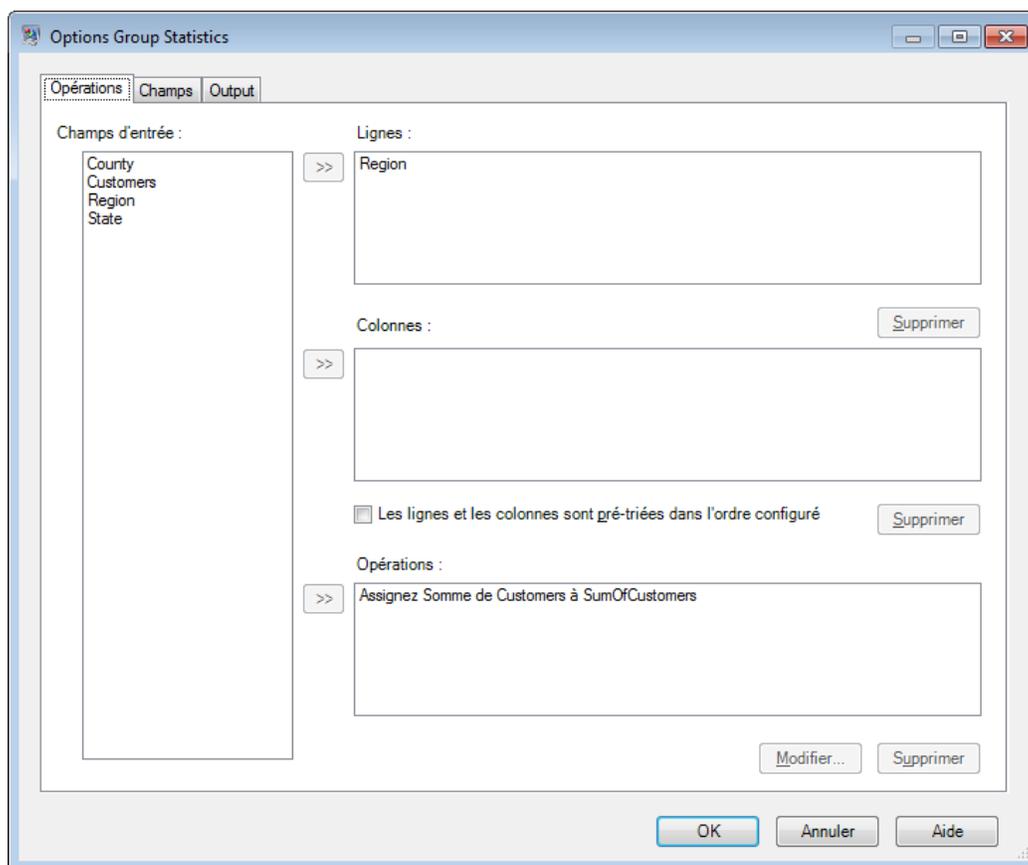
Exemple de Group Statistics

Ces données d'entrée illustrent le nombre de vos clients dans certains comtés. Les données illustrent également l'État (aux États-Unis) dans lequel le comté se situe (MD, VA, CA et NV), ainsi que la région (est ou ouest). La première ligne est un enregistrement d'en-tête.

```

Region|State|County|Customers
East|MD|Calvert|25
East|MD|Calvert|30
East|MD|Prince Georges|30
East|MD|Montgomery|20
East|MD|Baltimore|25
East|VA|Fairfax|45
East|VA|Clarke|35
West|CA|Alameda|74
West|CA|Los Angeles|26
West|NV|Washoe|22
    
```

Si vous souhaitez calculer le nombre total de clients pour chaque région, vous devez définir le champ Région sous forme de ligne dans l'onglet **Opérations** Pour l'opération, vous pouvez effectuer une opération de somme dans le champ Clients.



Résultat :

```
Region | SumOfCustomers
East | 210.0
West | 122.0
```

Remarque : Cet exemple affiche une opération de Group Statistics de base en utilisant uniquement les lignes pour agréger les données. Vous pouvez également créer un tableau croisé dynamique, qui regroupe à la fois les lignes et les colonnes, en indiquant une colonne à regrouper dans l'onglet **Opérations**.

Pour plus d'informations sur la création d'un tableau croisé dynamique, reportez-vous à la section [Création d'un tableau croisé dynamique](#) à la page 65.

Tableaux croisés dynamiques

Un tableau croisé dynamique regroupe et transpose les valeurs de colonne dans le flux de données pour faciliter l'analyse visuelle des données. Avec un tableau croisé dynamique, vous pouvez

organiser les colonnes d'entrée dans un format de tableau à double entrée (connu également comme croisé) qui produit des lignes, des colonnes et des valeurs comprimées. Vous pouvez également utiliser les champs en entrée et ne pas les afficher. Vous pouvez utiliser un tableau croisé dynamique pour ajouter un tableau croisé dynamique en deux dimensions ou pour regrouper des données en une seule dimension

Cet exemple illustre des données de ventes pour des chemises.

Tableau 3 : Données d'entrée

Région	Sexe	Style	Date d'expédition	Unités	Prix	Coût
Est	Garçon	Tee	1/31/2010	12	11,04	10,42
Est	Garçon	Golf	31/06/10	12	13,00	10,60
Est	Garçon	Fancy	2/25/2010	12	11,96	11,74
Est	Fille	Tee	1/31/2010	10	11,27	10,56
Est	Fille	Golf	31/06/10	10	12,12	11,95
Est	Fille	Fancy	1/31/2010	10	13,74	13,33
Ouest	Garçon	Tee	1/31/2010	11	11,44	10,94
Ouest	Garçon	Golf	2/25/2010	11	12,63	11,73
Ouest	Garçon	Fancy	2/25/2010	11	12,06	10,51
Ouest	Fille	Tee	2/25/2010	15	13,42	13,29
Ouest	Fille	Golf	31/06/10	15	11,48	10,67
Nord	Garçon	Tee	2/25/2010	17	16,04	10,42
Nord	Garçon	Fancy	2/25/2010	12	11,56	12,42
Nord	Fille	Tee	2/25/2010	16	12,32	18,42

Région	Sexe	Style	Date d'expédition	Unités	Prix	Coût
Nord	Garçon	Golf	1/31/2010	18	11,78	13,23
Nord	Fille	Tee	2/25/2010	12	18,45	11,64
Nord	Fille	Golf	2/25/2010	14	11,23	19,85
Nord	Garçon	Fancy	1/31/2010	16	12,54	13,42
Nord	Fille	Tee	2/25/2010	17	181,73	15,83
Sud	Garçon	Fancy	1/31/2010	19	14,15	13,42
Sud	Fille	Tee	2/25/2010	11	11,85	12,92
Sud	Fille	Fancy	1/31/2010	13	11,54	14,35
Sud	Garçon	Tee	2/25/2010	15	14,14	14,73
Sud	Garçon	Golf	2/25/2010	16	17,83	17,83
Sud	Fille	Fancy	31/06/10	11	18,24	12,35
Sud	Fille	Tee	1/31/2010	20	19,94	12,95
Sud	Garçon	Golf	2/25/2010	12	21,25	19,56

Nous voulons déterminer la quantité d'unités vendues dans chaque région pour chaque date d'expédition. Pour ce faire, nous utilisons la fonction de tableau croisé pour générer ce tableau :

Tableau 4 : Tableau croisé dynamique

Région	1/31/2010_ShipDate	2/25/2010_ShipDate	6/31/2010_ShipDate
Est	32	12	22
Nord	34	88	

Région	1/31/2010_ShipDate	2/25/2010_ShipDate	6/31/2010_ShipDate
Sud	52	54	11
Ouest	11	37	15

Dans ce cas, la colonne est Date d'expédition, la ligne est Région et les données que vous souhaitez visualiser sont Unités. Le nombre total d'unités expédiées s'affiche ici en utilisant somme d'agrégation.

Création d'un tableau croisé dynamique

Un tableau croisé dynamique synthétise les données pour faciliter les analyses en créant des catégories de ligne et de colonne de tableau basées sur les données d'entrée. Pour plus d'informations, reportez-vous à la section [Tableaux croisés dynamiques](#) à la page 62.

Dans les options du stage **Group Statistics** :

1. Dans l'onglet **Opérations**, sélectionnez un champ, parmi les **Champs d'entrée**, contenant les données que vous souhaitez utiliser comme libellés de ligne dans votre tableau croisé dynamique. Puis cliquez sur le bouton **>>** situé en regard du champ **Lignes**.
2. Sélectionnez un champ contenant les données que vous souhaitez utiliser comme colonnes dans votre tableau croisé dynamique, puis cliquez sur le bouton **>>** en regard du champ **Colonnes**.

Conseil : À ce stade, exécutez l'inspection pour voir les résultats de vos sélections. Cela vous aidera à visualiser les résultats des tableaux croisés basés sur les colonnes et les lignes sélectionnées.

3. Pour ignorer le tri des enregistrements d'entrée, cochez la case **Les lignes et les colonnes sont pré-triées dans l'ordre configuré**.

Si ce champ est coché, le stage traite les enregistrements d'entrée sans les trier.

Remarque : Cochez cette case si les enregistrements sont déjà triés.

4. Pour définir l'opération à effectuer, cliquez sur le bouton **>>** situé en regard du champ **Opérations**.

Dans la fenêtre **Ajouter une opération** :

- a) Sélectionnez l'**Opération** à effectuer.
- b) Dans la section **Champ d'entrée**, sélectionnez le **Nom** et le **Type** du champ d'entrée sur lequel l'opération doit être réalisée.
- c) Dans la section **Champ de sortie**, saisissez le **Nom** et sélectionnez le **Type** du champ de sortie à générer une fois que l'opération est réalisée.

- d) Pour récupérer le nombre réel des enregistrements d'entrée sur lesquels l'opération est réalisée sous forme de colonne de sortie distincte, cochez la case **Obtenir le nombre d'enregistrements calculés**.

Les enregistrements avec des valeurs `null` ne sont pas inclus dans le nombre

`ComputationalCount<Operation>Of<InputFieldName>`.

Fonctions prenant en charge le calcul du nombre :

- Moyenne
- Variance
- ZScore
- Écart type
- Centile
- Rang de Pourcentages
- Somme

Pour toute autre opération, cette case reste désactivée.

5. Pour définir les champs de sortie de chaque colonne du tableau croisé dynamique, cliquez sur l'onglet **Champs** des options de stage.

Conseil : Pour définir des champs avec précision, exécutez une fois un flux d'inspection, avant cette étape, pour afficher les noms de colonne générés par vos données.

- a) Cliquez sur **Ajouter**.
La fenêtre **Ajouter un Champ** s'affiche.
- b) Dans la fenêtre **Ajouter un Champ**, les colonnes de la grille sont basées sur les champs **Colonnes** que vous avez sélectionnés dans l'onglet **Opérations**. Dans ces colonnes de grille, saisissez les valeurs que vous voyez sous forme d'en-têtes de colonne lors de l'exécution d'un flux d'inspection.

Les enregistrements de la colonne Données peuvent également être renseignés en une seule fois à l'aide de la fonction Importer. Pour importer des données à partir d'un fichier CSV ou TXT :

1. Cliquez sur **Importer**.
2. Accédez au fichier source à l'aide du champ **Nom du fichier**.
3. Saisissez les valeurs **Séparateur de champ et d'enregistrement**
4. Cliquez sur OK.

Tous les enregistrements du fichier sont renseignés dans la table Données de colonne.

Remarque : Le fichier source ne doit pas comporter de ligne d'en-tête.

Par exemple, si vous avez sélectionné un champ d'entrée appelé `ShipDate` dans **Colonnes** de l'onglet **Opérations**, la grille de la fenêtre **Ajouter un champ** comprend une colonne appelée `ShipDate`. Dans cette colonne de grille, saisissez les valeurs `ShipDate` exactes présentes dans les données d'entrée de votre flux de données, par exemple, `2/25/2010`, `1/31/2010`.

- c) Dans le champ **Opération**, sélectionnez une ou plusieurs opérations pour lesquelles les colonnes de sortie sont générées pour chaque valeur de champ de colonne saisie. Notez que l'opération que vous sélectionnez affecte uniquement le nom du champ et ne contrôle pas le calcul réel.

Pour modifier les opérations répertoriées dans le champ **Opération** de l'onglet **Champs**, modifiez les valeurs du champ **Opérations** de l'onglet **Opérations**.

Avertissement : L'option d'opération de calcul du nombre

`ComputationalCount<Operation>Of<InputFieldName>` est répertoriée uniquement si la case **Obtenir le nombre d'enregistrements calculés** est cochée lors de la définition de la valeur **Opération** de l'onglet **Opérations**.

- d) Cliquez sur **Ajouter**.

6. Cliquez sur **OK**.

Pour chaque valeur d'entrée que vous avez saisie dans la grille ci-dessus, les colonnes de sortie sont automatiquement créées via leur mappage avec chacune des valeurs **Opération** sélectionnées. Un *produit cartésien* des valeurs de colonne d'entrée saisies dans la grille et des Opérations sélectionnées est utilisé pour générer automatiquement les colonnes de sortie finales.

Les noms de ces colonnes de sortie suivent la convention de nommage

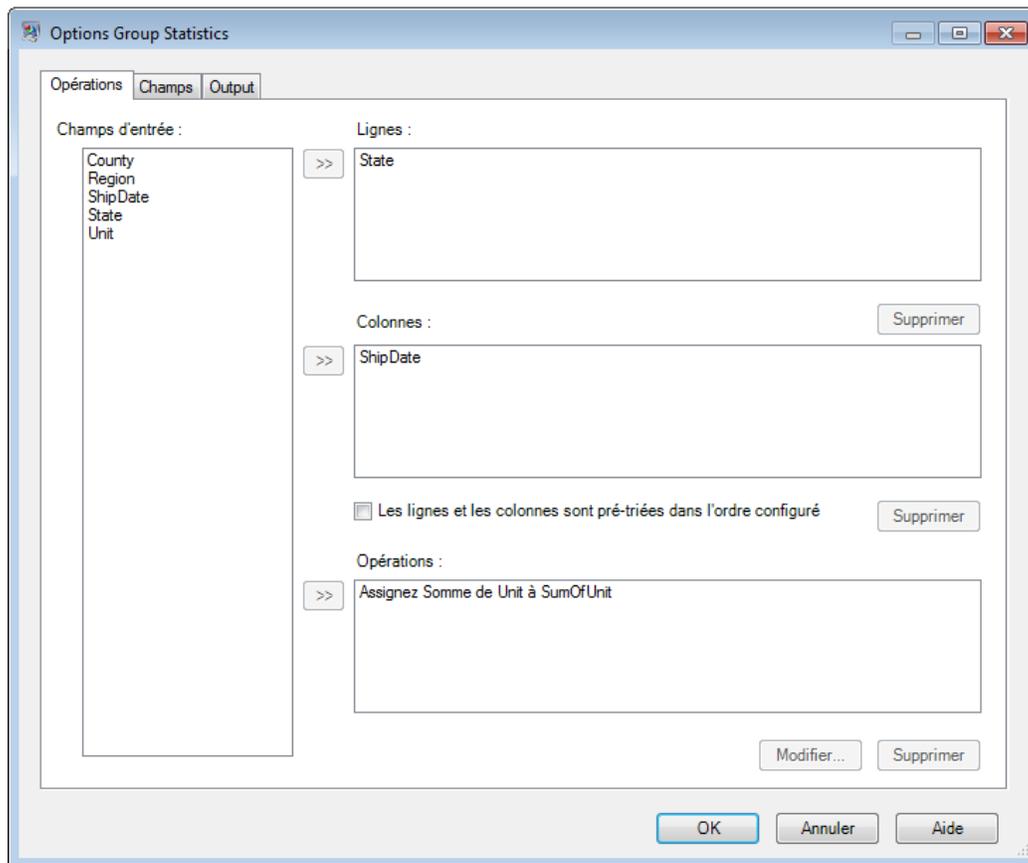
`<Data>_<Operation>Of<InputFieldName>`, où `<Data>` est la valeur que vous avez indiquée dans le premier champ, `<Operation>` est l'opération que vous avez sélectionnée dans le champ **Opération** et `<InputFieldName>` est la colonne d'entrée sur laquelle l'opération est réalisée.

Exemple de tableau croisé dynamique

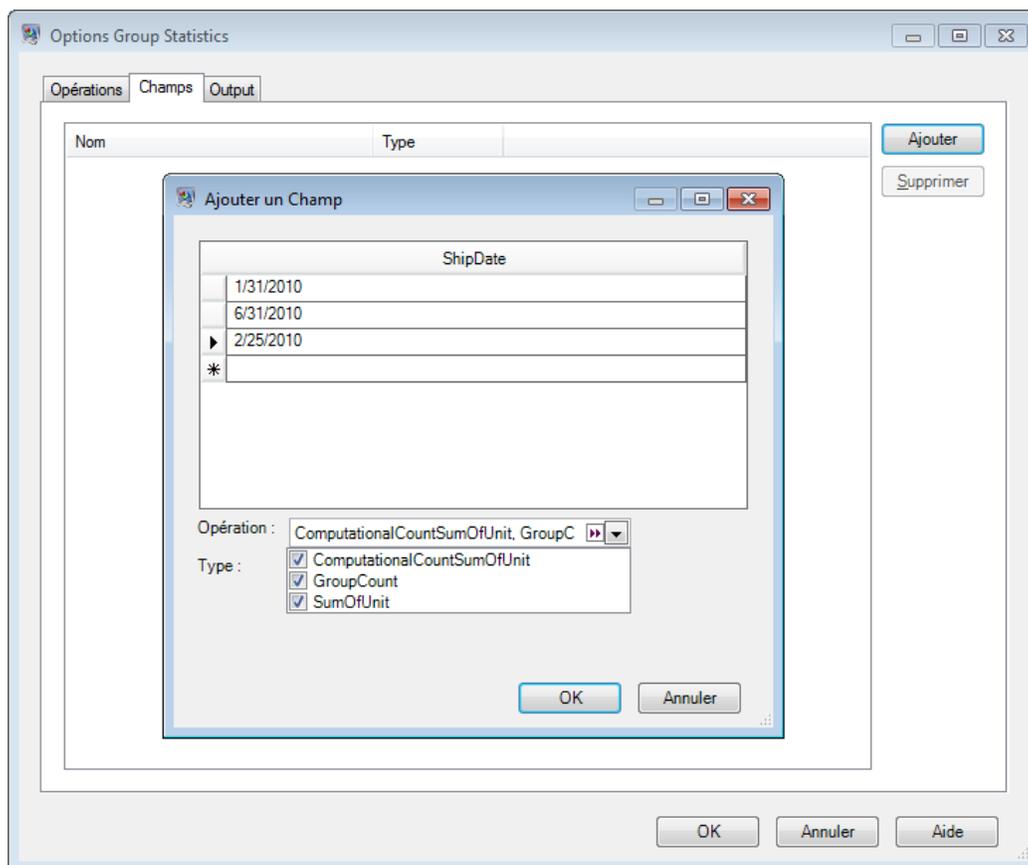
Données d'entrée qui affichent les informations d'expédition du service de gestion des commandes .

```
Region,State,County,ShipDate,Unit
East,MD,Calvert,1/31/2010,
East,MD,Calvert,6/31/2010,212
East,MD,Calvert,1/31/2010,633
East,MD,Calvert,6/31/2010,234
East,MD,Prince Georges,2/25/2010,112
East,MD,Montgomery,1/31/2010,120
East,MD,Baltimore,6/31/2010,210
East,VA,Fairfax,1/31/2010,710
West,CA,SanJose,1/31/2010,191
West,CA,Alameda,2/25/2010,411
West,CA,Los Angeles,2/25/2010,
West,CA,Los Angeles,2/25/2010,215
West,CA,Los Angeles,6/31/2010,615
West,CA,Los Angeles,6/31/2010,727
```

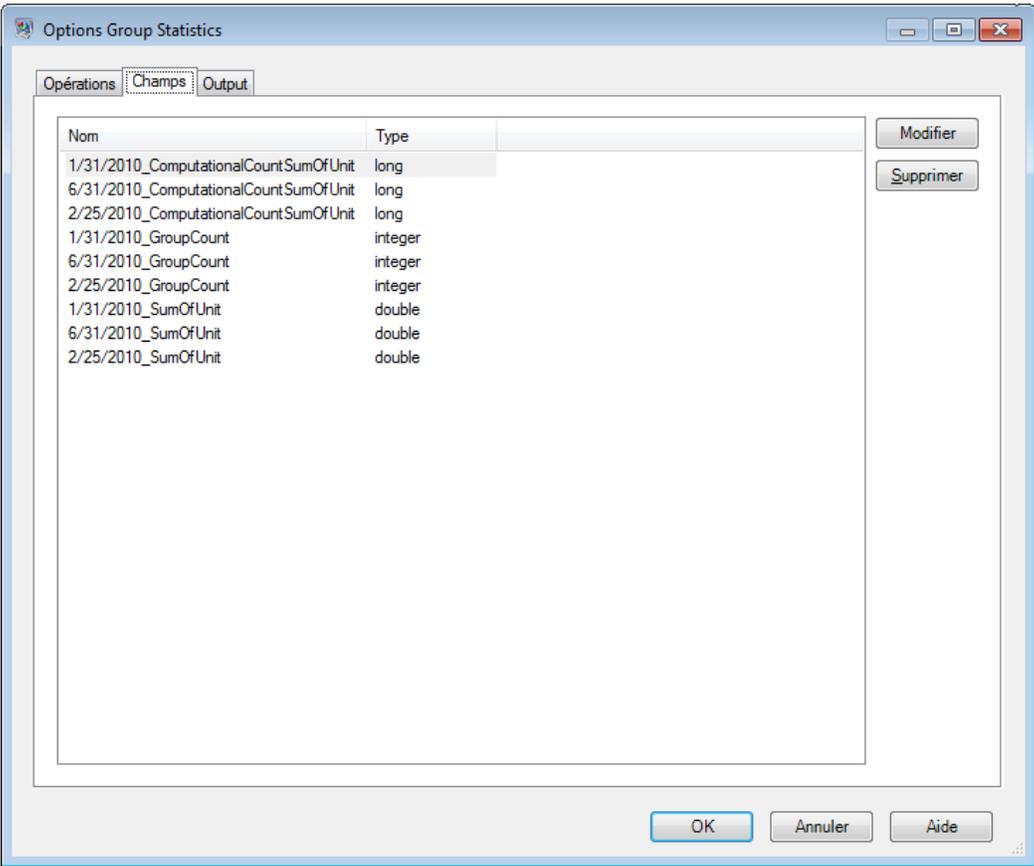
Pour déterminer le nombre d'expéditions effectuées à chaque date d'expédition pour chaque état, créez un tableau croisé dynamique en configurant le stage Group Statistics comme suit :



Dans l'onglet **Champs** des options du stage, ajoutez les dates exactes de la grille qui apparaissent dans le champ `ShipDate` des données d'entrée du flux de données, puis sélectionnez les valeurs **Opération** à afficher pour chacune de ces valeurs de colonne.



Si vous cliquez sur **OK** dans la fenêtre **Ajouter un champ**, les colonnes de sortie à créer sont automatiquement répertoriées dans l'onglet **Champs**. Ces colonnes de sortie sont un *produit cartésien* des valeurs d'entrée exactes et des opérations que vous avez sélectionnées dans la fenêtre **Ajouter un champ**.



Nom	Type
1/31/2010_ComputationalCountSumOfUnit	long
6/31/2010_ComputationalCountSumOfUnit	long
2/25/2010_ComputationalCountSumOfUnit	long
1/31/2010_GroupCount	integer
6/31/2010_GroupCount	integer
2/25/2010_GroupCount	integer
1/31/2010_SumOfUnit	double
6/31/2010_SumOfUnit	double
2/25/2010_SumOfUnit	double

Sortie

```
State,1/31/2010_GroupCount,1/31/2010_ComputationalCountSumOfUnit,
1/31/2010_SumOfUnit,2/25/2010_GroupCount,2/25/2010_ComputationalCountSumOfUnit,
2/25/2010_SumOfUnit,6/31/2010_GroupCount,6/31/2010_ComputationalCountSumOfUnit,
6/31/2010_SumOfUnit
VA,1,1,710,,,,,
CA,1,1,191,3,2,626,2,2,1342
MD,3,2,753,1,1,112,3,3,656
```

Math

Le stage Math s'occupe de calculs mathématiques sur une seule ligne et vous permet d'effectuer plusieurs fonctions mathématiques en utilisant une ou plusieurs expressions. Les données sont entrées comme des chaînes mais, en fonction du type d'opération étant effectué sur les données, les valeurs doivent être numériques ou booléennes.

1. Sous stages de contrôle, cliquez sur le stage Math et faites-le glisser sur le canevas, en le plaçant à l'endroit que vous voulez sur le flux de données.

2. Connectez le stage à d'autres stages sur le canevas.
3. Double-cliquez sur le stage Math. La boîte de dialogue **Options de Math** apparaît, avec l'onglet Expressions ouvert. Cet aperçu affiche les champs d'entrée, la Calculatrice et le canevas Expressions. Vous pouvez aussi cliquer sur l'onglet Fonctions pour utiliser les fonctions au lieu de la Calculatrice.

Le contrôle Champs d'entrée répertorie les champs valides trouvés sur le port d'entrée. La syntaxe de nom de champ est très flexible, mais elle est néanmoins soumise à des restrictions dues aux règles des scripts Groovy. Si vous ne connaissez pas bien les scripts Groovy, veuillez consulter ce site Web qui fournit des informations complètes concernant Groovy : groovy-lang.org.

Utilisation de la calculatrice

Le contrôle Calculatrice contient des boutons servant à saisir des constantes numériques et des opérateurs dans une expression. En double-cliquant sur les champs, les constantes, les opérateurs et les fonctions, vous les insérez dans une expression.

Tableau 5 : Opérateurs de la calculatrice

Opérateur	Description
Retour arrière	Sert à retourner en arrière d'un espace dans une expression
pi	Pi, une constante mathématique qui correspond au rapport entre la circonférence d'un cercle et son diamètre.
e	Nombre d'Euler, une constante mathématique qui constitue la base du logarithme naturel.
/	Division
*	Multiplication
+	Addition
-	Soustraction
x^y	Puissance de (par exemple, x^2 signifie x puissance 2)
Mod	Modulo, le reste de l'opération

Opérateur	Description
;	Point-virgule, employé en fin d'expression
=	Opérateur d'assignation
()	Parenthèses, pour représenter une hiérarchie dans une expression
.	Point décimal
ifelse	Instruction conditionnelle pour réaliser une action si la condition est true, sinon, réaliser une action différente
ifelse ifelse	Instruction conditionnelle multiple pour réaliser une action si la condition est true, sinon, réaliser une action différente
==	Égal à, dans une fonction mathématique
!=	N'est pas égal à
&&	Et logique
	Ou logique
>	Est supérieur à
>=	Supérieur ou égal à
<	Inférieur à
<=	Inférieur ou égal à

Utilisation des Fonctions et des Constantes

Le stage Math propose plusieurs fonctions à utiliser dans une expression. Les fonctions prennent la forme générale fonction(paramètre) ; fonction(paramètre,paramètre) ; fonction(paramètre,...), où « paramètre » est une constante numérique, une variable, ou une expression mathématique. Les fonctions peuvent être employées avec d'autres expressions mathématiques (par exemple, $x = \text{Sin}(y) * \text{Cos}(z)$).

Constantes, Conversion, Mathématiques, et Trigonométrie. Chacune des fonctions prises en charge sont répertoriées ci-dessous dans la catégorie correspondante.

Tableau 6 : Fonctions prises en charge

Fonction	Description
Constantes	
e	Une constante mathématique qui est la base du logarithme naturel.
faux	Une constante booléenne qui représente la valeur fausse.
Infini	Une constante mathématique qui représente l'infini.
NaN	Une constante mathématique qui représente une valeur qui n'est pas un nombre.
Pi	Une constante mathématique qui est le rapport entre la circonférence d'un cercle et son diamètre.
vrai	Une constante booléenne qui représente la valeur vraie.
Conversion	
Abs (valeur)	Prend un paramètre. Renvoie la valeur absolue de la valeur donnée.
Ceil (valeur)	Prend un paramètre. Renvoie une valeur arrondie au nombre supérieur (par exemple, Ceil(5,5) renvoie 6).
DegToRad (valeur)	Prend un paramètre. Convertit la valeur donnée des degrés aux radians.
Floor (valeur)	Prend un paramètre. Renvoie une valeur arrondie au nombre inférieur (par exemple, Floor(5,5) renvoie 5).
RadToDeg (valeur)	Prend un paramètre. Convertit la valeur donnée des radians aux degrés.
Round (valeur)	Prend un paramètre. Renvoie une valeur arrondie.

Fonction	Description
Math	
Avg (valeur, valeur, ...)	Prend un ou plusieurs paramètres. Renvoie la moyenne de toutes les valeurs données.
Exp (valeur)	Prend un paramètre. Renvoie le nombre d'Euler élevé à la puissance de la valeur donnée.
Fac (valeur)	Prend un paramètre. Renvoie la factorielle de la valeur donnée (par exemple, Fac(6) est calculée $6*5*4*3*2*1$ et renvoie 720).
Ln (valeur)	Prend un paramètre. Renvoie le logarithme naturel (base e) d'une valeur donnée.
Log (valeur)	Prend un paramètre. Renvoie le logarithme naturel (de base 10) d'une valeur donnée.
Max (valeur, valeur,...)	Prend un ou plusieurs paramètres. Renvoie la valeur maximale transmise.
Min (valeur, valeur,...)	Prend un ou plusieurs paramètres. Renvoie la valeur minimale transmise.
Sqrt (valeur)	Prend un ou plusieurs paramètres. Renvoie la racine carrée de la valeur transmise.
Sum (valeur)	Prend un paramètre. Renvoie la somme des valeurs données.
Trigonométrie	
ArcCos (valeur)	Prend un paramètre. Renvoie l'arc cosinus d'une valeur.

Fonction	Description
ArcSin (valeur)	Prend un paramètre. Renvoie l'arc sinus d'une valeur.
ArcTan (valeur)	Prend un paramètre. Renvoie l'arc tangente d'une valeur.
Cos (valeur)	Prend un paramètre. Renvoie le cosinus d'une valeur.
Ln (valeur)	Prend un paramètre. Renvoie le logarithme naturel (base e) d'une valeur donnée.
Sin (valeur)	Prend un paramètre. Renvoie le sinus d'une valeur.
Tan (valeur)	Prend un paramètre. Renvoie la tangente d'une valeur.

Utilisation des instructions conditionnelles

Des instructions conditionnelles peuvent être employées pour réaliser des actions dans le cas où diverses conditions ont la valeur true ou false. Les groupes qui utilisent des parenthèses (et) peuvent être employés pour des conditions plus complexes.

Tableau 7 : Conditions

Condition	Description
Est égal à	expression = expression
N'est pas égal à	expression != expression
Est supérieur à	expression > expression
Supérieur ou égal à	expression >= expression

Condition	Description
-----------	-------------

Inférieur à	expression < expression
-------------	-------------------------

Inférieur ou égal à	expression <= expression
---------------------	--------------------------

N'est pas une condition	!condition
-------------------------	------------

Et	condition && condition
----	------------------------

Ou	condition condition
----	------------------------

Instruction If

```
if(condition)
{
    actions to take if condition is true
}
```

Les crochets ne sont nécessaires que si plusieurs instructions sont exécutées après le « if ».

Instructions If-Else If

```
if(condition)
{
    actions to take if condition is true
}
else if(condition)
{
    actions to take if condition is true
}
else if...
```

```
if(SideLength != NaN)
{
    AreaOfPolygon=
    ((SideLength^2)*NumberOfSides)/
    (4*Tan(pi/NumberOfSides));
}
else if(Radius != NaN)
{
    AreaOfPolygon=
    (Radius^2)*NumberOfSides*Sin((2*pi)/NumberOfSides)/2;
}
```

Une ou plusieurs instructions else if peuvent être indiquées. Les crochets ne sont nécessaires que si plusieurs instructions sont exécutées après « ifelse ifelse ».

Instruction Else-If

```
if(condition)
{
  actions to take if condition is true
}
else if(condition)
{
  actions to take if condition is true
}
else if...
else
{
  actions to take if no conditions are met
}
```

Utilisation de la Console des expressions

La Console des expressions sert à saisir des expressions mathématiques devant être évaluées par le stage Math. Les contrôles Entrée, Calculatrice et Fonctions servent à saisir des valeurs dans cette console. Vous pouvez aussi entrer les expressions dans la console manuellement. Les expressions prennent la forme d'une constante, d'une variable ou d'une opération mathématique, et sont composées de constantes et de variables. Les constantes numériques sont des nombres entiers et décimaux, qui peuvent prendre des signes. Les variables représentent les données en provenance de la rangée entrante ; par exemple, si les champs x, y et z sont définis dans l'entrée, alors x, y et z peuvent être employés dans une expression. Les variables sont remplacées par les valeurs de champ au moment de l'exécution.

Le stage Math autorise également le regroupement d'expressions, mettant en jeu des parenthèses pour regrouper les expressions et ignorer la priorité de l'opérateur. Par exemple, $2*5^2$ est égal à 50 tandis que $(2*5)^2$ est égal à 100.

Remarque : Chaque expression saisie doit se terminer par un point-virgule.

De plus, des conditions peuvent être employées dans la console Expressions pour agir en conséquence selon si plusieurs conditions sont évaluées comme étant vraies ou fausses. Voir [Utilisation des instructions conditionnelles](#) à la page 75 pour en savoir plus sur les conditions.

Le stage Math s'occupe principalement d'expressions d'assignation dans lesquelles les résultats d'une expression sont attribués à une variable. Ce stage prend en charge les opérations d'assignations multiples et peuvent utiliser le résultat d'une opération d'assignation précédente.

Exemples d'expression d'assignation

Dans la situation suivante, $x=10$ et $z=1000$:

```
x=5+5
z=x*100
```

Dans la situation suivante, l'aire d'un polygone est calculée en fonction de la longueur d'un côté et le nombre de côtés.

```
AreaOfPolygon=
  ((SideLength^2) *NumberOfSides) /
  (4*Tan(pi/NumberOfSides));
```

Utilisation du contrôle Champs

Le contrôle Champs vous permet de modifier les types de champs d'entrée et de sortie. Vous pouvez modifier les types de champ depuis ce contrôle en cliquant sur la flèche déroulante dans la colonne Type et en le sélectionnant depuis la liste comportant les options suivantes :

booléen Type logique doté de deux valeurs : True et False. Des variables booléennes peuvent être utilisées dans des instructions conditionnelles afin de contrôler un flux. L'exemple de code suivant illustre une expression booléenne :

```
if(x && y)
  z=1;
else if(x)
  z=2;
else if(y)
  z=3;
else
  z=4;
```

double Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre 2^{-1074} et $(2-2^{-52}) \times 2^{1023}$. En notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.

flottant Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre 2^{-149} et $(2-2^{-23}) \times 2^{127}$. En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.

integer Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre -2^{31} (-2,147,483,648) et $2^{31}-1$ (2,147,483,647).

long Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre -2^{63} (-9,223,372,036,854,775,808) et $2^{63}-1$ (9,223,372,036,854,775,807).

Utilisation du Contrôle d'Aperçu

Le contrôle d'Aperçu vous permet de tester les expressions mathématiques. Les champs figurent dans la zone Données d'entrée ; vous pouvez fournir des valeurs spécifiques pour valider l'expression et consulter les résultats dans la zone Résultats en dessous de Données d'entrée

Les champs numériques sont initialisés sur 0 (0.000 pour double) et les champs booléens sont initialisés sur False. Les champs doubles et flottants sont limités à quatre décimales tandis que les champs entiers et longs n'ont pas de décimales

Record Combiner

Un Record Combiner combine deux enregistrements, ou plus, provenant de plusieurs flux, en un seul enregistrement. Record Combiner peut disposer d'un ou plusieurs ports d'entrée de stage. Par exemple, vous pouvez avoir un groupe d'enregistrements provenant d'une entrée de stage (port) et un autre groupe d'enregistrements provenant d'une seconde entrée de stage (port 2), et les enregistrements seront fusionnés pour ne former qu'un seul enregistrement. Si vous supprimez un stage intermédiaire, les ports ne seront pas renumérotés de manière consécutive.

Remarque : Record Combiner ne délivre pas d'enregistrement en sortie tant que chacun de ses ports d'entrée n'a pas reçu un enregistrement. Celui-ci doit combiner autant d'enregistrements qu'il compte de ports d'entrée avant de sortir un enregistrement.

Vous pouvez identifier le port à préserver dans le cas où les flux d'entrée disposent de champs du même nom. Par exemple, si vous combinez des enregistrements de deux flux et les deux flux contiennent un champ nommé AccountNumber, vous pouvez identifier le flux dont vous souhaitez préserver le champ AccountNumber en choisissant le port d'entrée Record Combiner qui correspond au flux à préserver. Les données du champ AccountNumber dans l'autre flux seront rejetées.

Record Joiner

Record Joiner effectue une opération de style SQL `JOIN` pour combiner les enregistrements de différents flux en fonction d'une relation entre les champs des flux. Vous pouvez utiliser Record Joiner pour joindre des enregistrements de plusieurs fichiers, plusieurs bases de données ou tout canal en amont dans le flux de données. Vous devez connecter au moins deux canaux d'entrée à Record Joiner. Les résultats de l'opération `JOIN` sont ensuite écrits vers un canal de sortie. Éventuellement, les enregistrements qui ne correspondent pas à la condition join peuvent être écrits dans un canal de sortie distinct.

Remarque : Avant d'utiliser Record Joiner, vous devez bien comprendre l'opération SQL `JOIN`. Pour plus d'informations, reportez-vous à [wikipedia.org/wiki/Join_\(SQL\)](https://wikipedia.org/wiki/Join_(SQL)).

Définition de jointure

Option	Description
Port gauche	<p>Port dont vous souhaitez utiliser les enregistrements comme table gauche dans l'opération JOIN. Tous les autres ports d'entrée sont alors utilisés comme tables droites dans l'opération JOIN.</p> <p>Remarque : La table « gauche » et la table « droite » sont des concepts SQL JOIN. Avant d'utiliser Record Joiner, vous devez bien comprendre l'opération SQL JOIN. Pour plus d'informations, reportez-vous à wikipedia.org/wiki/Join_(SQL).</p>
Type de jointure	<p>Type de l'opération JOIN à effectuer. L'un des éléments suivants :</p> <p>Externe Gauche Renvoie tous les enregistrements du port gauche même si aucune correspondance n'existe entre le port gauche et les autres ports. Cette option renvoie tous les enregistrements du port gauche plus tout enregistrement correspondant dans l'un des autres ports.</p> <p>Complet Renvoie tous les enregistrements de tous les ports.</p> <p>Interne Renvoie uniquement les enregistrements présentant une correspondance entre le port gauche et l'autre port. Par exemple, si vous avez quatre sources d'entrée et que le port 1 est le port gauche, une jointure interne renverra les enregistrements ayant des champs correspondants entre port 1 et port 2, port 1 et port 3, et port 1 et port 4.</p>
Joindre les champs	<p>Le ou les champs du port gauche doivent correspondre aux données contenues dans un champ d'un autre port afin de joindre les enregistrements.</p> <p>Remarque : Seuls les champs dotés d'un type de données chaîne, entier, date ou datetime peuvent être utilisés comme champs de jointure.</p>

Option	Description
Données du port gauche triées	<p>Détermine si les enregistrements dans le port gauche sont déjà triés par le champ indiqué dans Joindre les champs. Si les enregistrements sont déjà triés, vous pouvez améliorer les performances en cochant cette case. Si vous ne cochez pas cette case, Record Joiner triera les enregistrements en fonction du champ indiqué dans Joindre les champs avant d'effectuer l'opération de jointure.</p> <p>Si vous avez indiqué plusieurs champs de jointure, les enregistrements doivent être triés dans l'ordre des champs répertoriés dans Joindre les champs. Par exemple, si vous disposez de deux champs de jointure :</p> <p style="padding-left: 20px;">Montant Région</p> <p>Les enregistrements doivent être triés tout d'abord par le champ Montant, puis par le champ Région.</p> <p>Important : Si vous sélectionnez cette option mais que les enregistrements ne sont pas triés, vous obtiendrez des résultats incorrects de Record Joiner. Ne sélectionnez cette option que si vous êtes sûr que les enregistrements dans le port gauche sont déjà triés.</p>
Définitions de jointure	<p>Décrit les conditions de jointure qui seront utilisées pour déterminer si un enregistrement du port gauche doit être joint à un enregistrement d'un des autres ports. Par exemple :</p> <pre>port1.Name = port2.Name</pre> <p>Cela indique que si la valeur dans le champ Name d'un enregistrement de port1 correspond à la valeur dans le champ Nom d'un enregistrement de port2, les deux enregistrements seront joints.</p> <p>Pour modifier une condition de jointure, cliquez sur Modifier. Sélectionnez un champ du port droit dont les données doivent correspondre aux données dans le champ Join du port gauche pour joindre les enregistrements. Pour modifier le champ du port gauche, cliquez sur Cancel et modifiez-le dans le champ Join Fields. Si les enregistrements dans le port droit sont triés par le champ de jointure, cochez la case Données du port droit triées. Vous pouvez améliorer les performances en cochant cette case.</p> <p>Important : Si vous sélectionnez Données du port droit triées mais que les enregistrements ne sont pas triés, vous obtiendrez des résultats incorrects de Record Joiner. Ne sélectionnez cette option que si vous êtes sûr que les enregistrements dans le port droit sont déjà triés.</p>

Résolution des champs

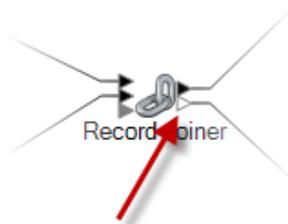
Cet onglet indique le port dont vous souhaitez utiliser les données dans l'enregistrement joint, dans le cas où le même nom de champ existe dans plusieurs ports d'entrée. Par exemple, si vous effectuez une jointure de deux sources de données et que chaque source contient un champ nommé DateOfBirth, vous pouvez déterminer le port dont vous souhaitez utiliser les données dans le champ DateOfBirth dans l'enregistrement joint.

Si vous disposez de champs portant le même nom mais des données différentes, et que vous souhaitez préserver les données des deux champs dans l'enregistrement joint, vous devez renommer un des champs pour envoyer les données à Record Joiner. Vous pouvez utiliser le stage Transformer pour renommer les champs.

Manipulation des enregistrements qui ne sont pas joints

Pour qu'un enregistrement soit inclus dans la sortie Record Joiner, il doit remplir la condition de jointure, ou un type de jointure renvoyant les enregistrements joints et les enregistrements ne répondant pas à la condition de jointure doit être sélectionné. Par exemple, une jointure complète renverra tous les enregistrements de tous les ports d'entrée, qu'un enregistrement réponde ou non à la condition de jointure. Dans le cas d'un type de jointure qui ne renvoie pas tous les enregistrements de tous les ports, tel qu'une jointure externe gauche ou une jointure interne, seuls les enregistrements correspondant à la condition de jointure sont inclus dans la sortie Record Joiner.

Pour capturer les enregistrements qui ne sont pas inclus dans le résultat de l'opération de jointure, utilisez le port de sortie **not_joined**. La sortie de ce port contient tous les enregistrements non inclus dans le port de sortie régulier. Le port de sortie **not_joined** correspond au triangle blanc à droite du stage Record Joiner comme illustré ici :



Le champ **InputPortIndex** est ajouté aux enregistrements sortant de ce port. Ce champ contient le numéro du port d'entrée Record Joiner dont l'enregistrement provient. Cela vous permet d'identifier la source de l'enregistrement.

Remarque :

- Pour obtenir des performances optimales de ce stage, assurez-vous que deux flux d'enregistrements indépendants sont joints afin de générer une sortie consolidée.
- Si un seul chemin d'accès est tout d'abord branché à l'aide d'un stage Broadcaster ou Conditional Router, puis branché de nouveau à l'aide d'un stage Record Joiner, le flux peut se bloquer. Si plusieurs stages sont utilisés entre un embranchement et une jointure, utilisez Sorter aussi près que possible de Record Joiner.

Sorter

Sorter trie les enregistrements à l'aide des champs que vous indiquez. Par exemple, vous pouvez faire trier vos enregistrements par nom, ville ou n'importe quel autre champ de votre flux de données.

Tri des enregistrements avec Sorter

Le stage Sorter vous permet de trier les enregistrements à l'aide des champs que vous indiquez.

1. Sous Stages de contrôle, glissez Sorter sur le canevas afin de le placer là où vous le souhaitez sur le flux de données.
2. Double-cliquez sur Sorter.
3. Cliquez sur **Ajouter**.
4. Cliquez sur la flèche Bas dans la colonne **Nom du champ** et sélectionnez le champ que vous voulez trier.

Remarque : La liste des champs disponibles repose sur les champs utilisés dans les stages précédents dans le flux de données.

5. Dans la colonne **Classer**, déterminez si vous souhaitez effectuer un tri par ordre croissant ou décroissant.
6. Dans la colonne **Type**, sélectionnez le type de données du champ.

Remarque : Si vos données entrantes ne sont pas au format de chaîne, la colonne **Type** sera désactivée.

bigdecimal	Type de données numérique prenant en charge 38 points de précision décimaux. Utilisez ce type de données pour les données utilisées dans les calculs mathématiques requérant un haut niveau de précision, en particulier ceux impliquant des données financières. Le type de données bigdecimal prend en charge des calculs plus précis que le type de données double.
double	Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre 2^{-1074} et $(2-2^{-52}) \times 2^{1023}$. En notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.
flottant	Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre 2^{-149} et $(2-2^{-23}) \times 2^{127}$. En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.
integer	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre -2^{31} (-2,147,483,648) et $2^{31}-1$ (2,147,483,647).
long	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre -2^{63} (-9,223,372,036,854,775,808) et $2^{63}-1$ (9,223,372,036,854,775,807).
chaîne	Séquence de caractères.

7. Pour supprimer des espaces vides situés avant et après la valeur avant de trier, cochez la case de la colonne **Raccourcir**. L'option Raccourcir ne modifie pas la valeur du champ. Elle ne raccourcit la valeur que dans l'objectif du tri. Notez bien que si vos données entrantes ne sont pas au format de chaîne, la colonne **Raccourcir** sera désactivée.

8. Dans la colonne **Traiter nul comme**, sélectionnez **Le plus grand** ou **Le plus petit** pour indiquer l'emplacement des valeurs nulles dans la liste triée. Le placement dépend de la combinaison d'options sélectionnées dans les champs **Ordre** et **Traiter nul comme**, comme indiqué dans le tableau ci-dessous :

Ordre	Traiter nul comme	Placement des valeurs nulles dans la liste triée
Ascendant	Le plus grand	En bas de la liste
Ascendant	Le plus petit	En haut de la liste
Descendant	Le plus grand	En haut de la liste
Descendant	Le plus petit	En bas de la liste

9. Répétez ces étapes tant que vous n'avez pas ajouté tous les champs à trier.
10. Réorganisez l'ordre de tri de la manière que vous désirez en cliquant sur **Vers le haut** ou **Vers le bas**. Cela vous permet d'effectuer d'abord un tri par un champ, puis de trier le résultat par un autre champ.
11. Pour remplacer les options de performances de tri par défaut ayant été définies par votre administrateur, cliquez sur **Avancé**, cochez la case **Écraser les options de performance du tri**, puis spécifiez les options suivantes :

Limite du nombre d'enregistrement en mémoire Spécifie le nombre maximum de rangées de données qu'un trieur peut contenir en mémoire avant que celui-ci commence à pager sur le disque. Par défaut, un tri de 10 000 enregistrements ou moins sera effectué en mémoire et un tri de plus de 10 000 enregistrements sera effectué sur le disque. La limite maximale est de 100 000 enregistrements. En général, un tri en mémoire est beaucoup plus rapide qu'un tri sur le disque ; donc, il faut définir une valeur assez haute pour que la plupart des tris s'effectuent en mémoire et que seuls les groupes de grande taille soient écrits sur le disque.

Remarque : Soyez conscient du fait qu'au sein d'environnements où des jobs s'exécutent de manière simultanée, une augmentation du paramètre **Dans la limite d'enregistrement mémoire** augmente la probabilité de ne plus disposer de suffisamment de mémoire.

Nombre maximal de fichiers temporaires Spécifie le nombre maximal de fichiers temporaires pouvant être employés par un processus de tri. L'utilisation d'un plus grand nombre de fichiers temporaires peut améliorer les performances. Cependant, le nombre optimal dépend très largement de la configuration du serveur qui exécute Spectrum™ Technology Platform. Nous vous conseillons d'essayer

différents paramètres et d'observer l'effet de l'utilisation d'un plus ou moins grand nombre de fichiers temporaires sur les performances. Pour calculer le nombre approximatif de fichiers temporaires nécessaires, utilisez l'équation suivante :

$$\frac{(NumberOfRecords \times 2)}{InMemoryRecordLimit} = NumberOfTempFiles$$

Notez que le nombre maximal de fichiers temporaires ne peut pas être supérieure à 1 000.

Activer la compression

Spécifie si les fichiers temporaires sont compressés lors de leur écriture sur le disque.

Remarque : Les paramètres de performances de tri optimal dépendent de la configuration matérielle de votre serveur. Néanmoins, l'équation suivante produit généralement de bonnes performances de tri :

$$\frac{(InMemoryRecordLimit \times MaxNumberOfTempFiles \div 2)}{TotalNumberOfRecords} \geq$$

12. Cliquez sur **OK**.

Remarque : Vous pouvez supprimer les critères de tri à votre convenance en surlignant une rangée et en cliquant sur **Supprimer**.

Splitter

Splitter (Diviseur) convertit des données hiérarchiques en données à plat. Les Splitters disposent d'un port d'entrée et d'un port de sortie qui transmettent les données du Splitter au stage suivant. Une façon de se servir de la fonctionnalité de Splitter consiste à prendre une liste d'informations dans un fichier et extraire chaque élément d'informations discret dans sa propre ligne de données. Par exemple, votre entrée peut comprendre des éléments géographiques compris à une certaine distance d'un point latitudinal/longitudinal, et Splitter peut mettre chaque élément géographique dans une ligne de données distincte.

Utilisation du stage Splitter

1. Sous Stages de contrôle, cliquez sur **Splitter** et faites-le glisser sur le canevas, en le plaçant à l'endroit souhaité sur le flux de données, et connectez-le aux stages d'entrée et de sortie.
2. Double-cliquez sur Splitter. La boîte de dialogue **Options de Splitter** apparaît.
3. Cliquez sur l'élément déroulant **Diviser à** pour voir les autres types de liste disponibles pour ce stage. Cliquez sur le type de liste que vous voulez que Splitter génère. La boîte de dialogue

Options de Splitter s'ajuste en fonction de votre sélection, affichant les champs disponibles pour ce type de liste.

Vous pouvez aussi cliquer sur le bouton de sélection (...) en regard de l'élément déroulant **Diviser à**. La boîte de dialogue **Schéma de Champ** apparaît, affichant le schéma pour les données entrant dans Splitter. Les types de listes s'affichent en gras, suivis des listes individuelles de chaque type. Le format de ces champs s'affiche également (chaîne, double et ainsi de suite). Cliquez sur le type de liste que vous voulez que Splitter crée puis cliquez sur **OK**. La boîte de dialogue **Options de Splitter** s'ajuste en fonction de votre sélection, affichant les champs disponibles pour ce type de liste.

4. Cliquez sur **Enregistrement d'en-tête de sortie** pour renvoyer l'enregistrement original avec la liste divisée extraite.
5. Cliquez sur **Uniquement lorsque la liste d'entrée est vide** pour renvoyer l'enregistrement original uniquement en l'absence de liste divisée pour cet enregistrement.
6. Sélectionnez les champs que vous voulez que Splitter inclut à la sortie en cochant la case **Inclure** pour les champs correspondants.
7. Cliquez sur **OK**.

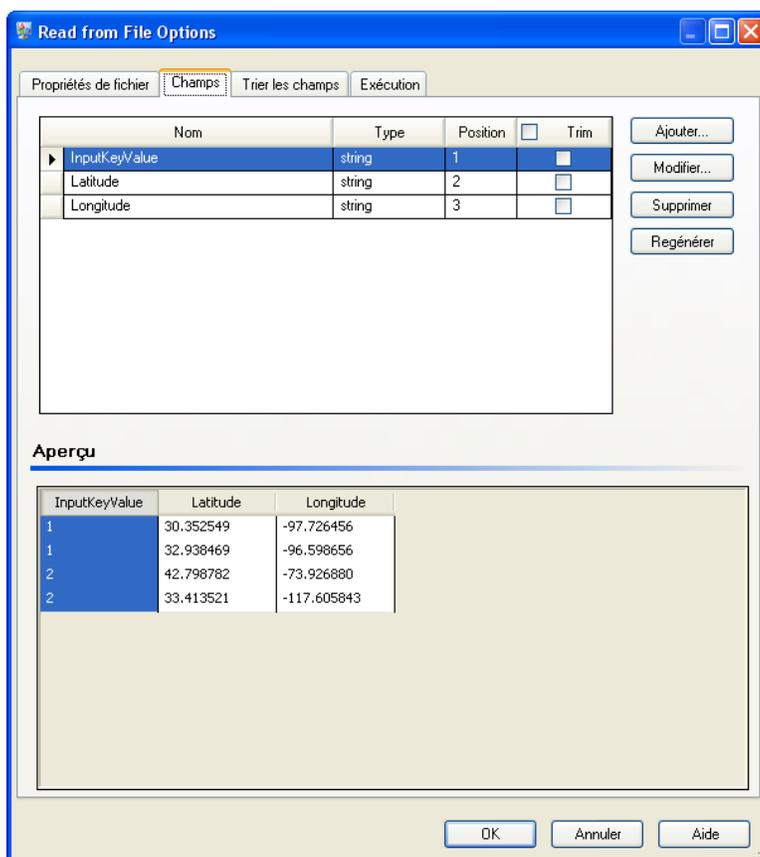
Exemple de Splitter

L'exemple suivant prend la sortie d'un stage d'itinéraire incluant des itinéraires pour la voiture et met chaque itinéraire (ou article de liste) dans une ligne de données. Le flux de données ressemble à ceci :

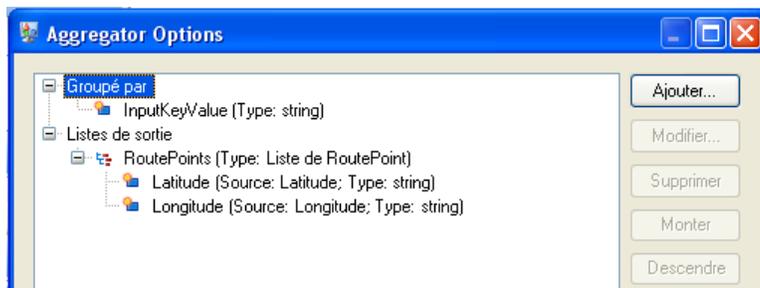


Le flux de données exécute la fonction de la manière suivante :

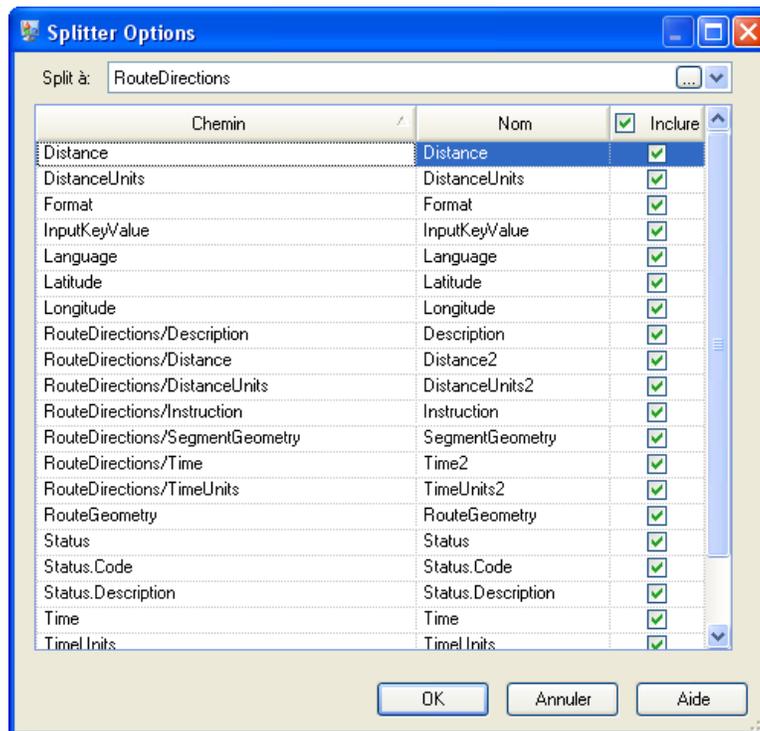
1. Le stage **Read from File** contient des latitudes, longitudes et valeurs clés d'entrée pour vous aider à identifier les points individuels.



2. Le stage **Aggregator** génère les données à partir du stage **Read from File** sous forme de schéma (une hiérarchie structurée de données) et identifie le groupe de latitudes et de longitudes en tant que liste de points d'itinéraires, ce qui constitue une étape nécessaire pour le bon fonctionnement du prochain stage.



3. Le stage **Get Travel Directions** du Module Location Intelligence crée un itinéraire à partir d'un emplacement jusqu'à un autre à l'aide des points d'itinéraire de l'étape 2.
4. Le stage **Splitter** établit la division des données du champ Itinéraire, et inclut tous les champs possibles du stage **Get Travel Directions** dans les listes de sortie.



Le schéma est structuré comme suit, avec des Itinéraires et des Points d'itinéraire comme listes disponibles pour ce job :



5. Le stage **Write to File** écrit la sortie dans un fichier.

Stream Combiner

Un Stream Combiner (générateur de flux) rassemble deux ou plusieurs flux d'enregistrements de multiples stages. Stream Combiner dispose d'un ou de plusieurs ports d'entrée de stage. Par exemple, vous pouvez avoir un groupe d'enregistrements provenant d'un stage et un autre groupe provenant d'un second stage, et les enregistrements seront fusionnés pour ne former qu'un seul flux.

Stream Combiner ne présente aucun paramètre.

Transformer

Le stage Transformer modifie les valeurs et le formatage des champs. Vous pouvez sélectionner plus d'un transform à exécuter sur un champ à partir du moment où les noms des champs d'entrée et de sortie sont identiques.

Transformations générales

Construct Field Utilise les valeurs des champs existants et/ou les valeurs de constante pour remplacer les valeurs de champ ou créer un champ. Par exemple, imaginons que vous disposiez d'un champ nommé City et que vous souhaitiez ajouter l'expression « City of » pour les valeurs du champ City. Vous pouvez créer un modèle ressemblant à ceci :

```
City of ${City}
```

Dans le champ **To field**, vous devez sélectionner le champ City. Cela a pour effet de remplacer les valeurs existantes du champ City par une valeur construite à l'aide du modèle. Par exemple, si la valeur du champ City est Chicago, la nouvelle valeur est City of Chicago.

Certains caractères doivent être précédés d'une barre oblique inverse (« \ ») afin de produire un modèle valide. Par exemple, le caractère de guillemet simple doit être précédé d'une barre oblique inverse comme suit : \ '. Consultez groovy-lang.org/syntax.html pour obtenir une liste des caractères qui doivent être précédés d'une barre oblique inverse.

Copy Copie la valeur d'un champ à un autre.

Personnalisé Vous permet de définir votre propre transformation à l'aide du langage Groovy. Pour plus d'informations, reportez-vous à la section [Création d'une transformation personnalisée](#) à la page 94.

Pour les utilisateurs du module Location Intelligence, les transformations personnalisées peuvent accéder aux jeux de données spatiales. Reportez-vous à la section Stages du *Guide Spectrum Spatial* sur support.pb.com.

Renommer	Modifie le nom d'un champ. Vous pouvez effectuer une sélection dans une liste de noms de champs figurant déjà dans le flux de données ou vous pouvez saisir le nom de votre choix.
État	Modifie le champ État pour lui donner une valeur Réussite ou Échec. Si la valeur est réglée sur Echec, des valeurs facultatives Description et Code peuvent également être définies.

Mise en forme des transformations

Case	Modifie la casse (majuscules ou minuscules).
Mask	Applique ou supprime des caractères d'un champ. Pour plus d'informations, reportez-vous à la section Utilisation d'une transformation Masquer à la page 100.
Pad	Ajoute des caractères à gauche ou à droite de la valeur du champ.

Transformations de chaîne

Minimize Whitespace	Supprime les espaces blancs au début et à la fin d'un champ. Remplace aussi une séquence d'espaces blancs (espaces multiples ou consécutifs, par exemple) par un seul caractère d'espace blanc.
Remove Substring	Supprime toutes les occurrences d'une chaîne dans un champ. Par exemple, vous pouvez supprimer « CA » du champ StateProvince.
Substring	Copie une séquence contiguë de caractères d'un champ dans un autre.
Trim	Supprime les caractères indiqués à gauche, droite ou sur les deux côtés d'un champ. Remarquez sur cette transformation est sensible à la casse.
Truncate	Supprime un nombre spécifié de caractères à gauche et à droite d'un champ.

Transformations de liste

Cette fonction vous permet de créer une transformation en conserve qui agit sur des listes, par exemple l'entrée de Read From XML.

Pour définir des transformations de liste, procédez comme suit :

1. Sélectionnez une opération de transformation de liste. Les champs d'entrée apparaissent dans une vue d'arborescence à droite.
2. Sélectionnez un champ valide dans l'arborescence auquel appliquer l'opération. Les propriétés de l'opération s'affichent sous la vue d'arborescence de champs d'entrée.
3. Indiquez les propriétés de l'opération et cliquez sur Ajouter. La transformation est ajoutée à la liste de la fenêtre parente, c'est-à-dire, la fenêtre « Options de Transformer ».

Créer un champ	Permet la création d'un champ sous le champ de type de liste sélectionné par l'utilisateur. Par exemple, si une liste appelée Football comporte deux clubs, à savoir, Knitters et Lambs, l'utilisateur peut ajouter un nouveau club appelé Irons ; la liste comporte alors trois clubs.
-----------------------	---

- Trier** Exécute un tri sur les valeurs présentes dans le champ sélectionné. Dans une liste complexe, l'utilisateur doit indiquer l'élément clé pour le tri, tandis qu'en cas de liste simple, le tri a lieu sur les éléments présents dans la liste. L'utilisateur peut sélectionner l'ordre de tri, croissant ou décroissant. Dans l'exemple de Football, lorsque la liste comporte trois clubs, l'utilisateur doit sélectionner le champ Nom sous Club pour trier les clubs en fonction du nom. Les entrées de club existantes sont triées dans l'ordre suivant : Irons, Knitters et Lambs, si l'ordre de tri est croissant, et inversement s'il est décroissant. À présent, si l'utilisateur souhaite trier la liste des joueurs, le champ Joueurs doit être sélectionné et l'ordre de tri défini pour ce champ.
- Somme** Exécute la sommation de toutes les valeurs présentes dans le champ sélectionné. La sortie est stockée dans un champ défini par l'utilisateur. Par exemple, si l'utilisateur souhaite afficher le total des points obtenus par chaque club de football, il doit sélectionner le champ Points sous Tournoi et indiquer le nom de champ de sortie.
- Copy** Effectue une opération de copie du champ sélectionné dans le champ spécifié par l'utilisateur. Lorsque l'utilisateur sélectionne un champ à copier, le champ et tous les champs dessous (le cas échéant) sont copiés dans le nouveau champ spécifié. Cette opération a lieu au même niveau de hiérarchie.
- Renommer** Effectue une opération de renommage du champ sélectionné en un nouveau nom spécifié par l'utilisateur.

Le code XML suivant fournit une référence à la fonction Transformation de liste :

```
<?xml version="1.0"?>
<sports_details>
  <sports name="football">
    <clubs>
      <club name="Knitters">
        <player>Samuel</player>
        <player>Messi</player>
        <player>kaka</player>
        <player>Alan</player>
        <coach>Stuart</coach>
        <Tournament name="Football League">
          <result>won</result>
          <points>4</points>
        </Tournament>
        <Tournament name="UEFA">
          <result>draw</result>
          <points>2</points>
        </Tournament>
      </club>
      <club name="Lambs">
        <player>Ronaldo</player>
        <player>Neymar</player>
        <player>Zlatan</player>
        <player>Mesut</player>
        <coach>Ivan</coach>
      </club>
    </clubs>
  </sports>
</sports_details>
```

```

    <Tournament name="Airtel League">
<result>draw</result>
<points>2</points>
    </Tournament>
    <Tournament name="Champions League">
<result>lost</result>
<points>0</points>
    </Tournament>
</club>
<club name="Irons">
    <player>Scott</player>
    <player>Paul</player>
    <player>John</player>
    <player>Andrew</player>
    <coach>Jeff</coach>
    <Tournament name="CAF">
<result>won</result>
<points>4</points>
    </Tournament>
    <Tournament name="Copa America">
<result>won</result>
<points>4</points>
    </Tournament>
</club>
</clubs>
</sports>
<sports name="badminton">
<clubs>
    <club name="Shuttlers">
    <player>Saina</player>
    <player>Viktor</player>
    <player>Chen</player>
    <player>Srikanth</player>
    <coach>Jan</coach>
    <Tournament name="Olympic Games">
<result>won</result>
<points>4</points>
    </Tournament>
    <Tournament name="Commonwealth Games">
<result>won</result>
<points>4</points>
    </Tournament>
</club>
    <club name="Choppers">
    <player>Wang</player>
    <player>Sindhu</player>
    <player>Carolina</player>
    <player>Li Xuerui</player>
    <coach>Ratchanok</coach>
    <Tournament name="World Junior">
<result>draw</result>
<points>2</points>
    </Tournament>

```

```

    <Tournament name="Uber Cup">
    <result>draw</result>
    <points>2</points>
    </Tournament>
  </club>
  <club name="Lobbers">
    <player>Nozomi</player>
    <player>Chou</player>
    <player>Marc</player>
    <player>Lin</player>
    <coach>Kevin</coach>
    <Tournament name="World Senior">
    <result>won</result>
    <points>4</points>
    </Tournament>
    <Tournament name="Thomas Cup">
    <result>won</result>
    <points>4</points>
    </Tournament>
  </club>
</clubs>
</sports>
</sports_details>

```

Modification de l'ordre des transformations

Si vous comptez exécuter plus d'une fonction transform sur un champ de sortie particulier, vous pouvez définir l'ordre selon lequel celles-ci vont s'exécuter.

Remarque : Si vous mappez un seul champ sur deux champs de sortie différents (par exemple, ValidateAddress.City sur Output.City1 et ValidateAddress.City sur Output.City2), et que vous ajoutez des transforms à chaque champ, le transform pour le champ secondaire doit être exécutée en premier. Vous devez changer l'ordre d'exécution des transforms pour exécuter le transform du deuxième champ (Output.City2) en premier.

1. Double-cliquez sur le stage Transformer. La boîte de dialogue Options de Transformer apparaît.
2. Sélectionnez une fonction Transform et utilisez les boutons Déplacer vers le haut et Déplacer vers le bas pour réarranger l'ordre des transforms. La transform de niveau supérieur sera exécutée en premier.

Remarque : Les transforms dépendantes ne peuvent pas être déplacées au-dessus des transforms primaires (c'est-à-dire les transforms dont elles dépendent).

3. Cliquez sur **OK**.

Création d'une transformation personnalisée

Le stage Transformer dispose de transformations prédéfinies qui effectuent un éventail de transformations de données communes. Si les transformations prédéfinies ne répondent pas à vos besoins, vous pouvez écrire un script de transformation personnalisé via Groovy. Cette procédure explique comment créer des transformations personnalisées de base à l'aide de Groovy. Pour obtenir une documentation complète sur Groovy, consultez l'adresse suivante : groovy-lang.org.

1. Dans Enterprise Designer, ajoutez un stage Transformer au dataflow.
2. Double-cliquez sur le stage Transformer.
3. Cliquez sur **Ajouter**.
4. Sous **Général**, cliquez sur **Personnalisé**.
5. Dans le champ **Custom transform name**, entrez un nom pour la transformation que vous créez. Ce nom doit être unique.
6. Cliquez sur **Éditeur de script**.

Cet éditeur fournit un certain nombre de fonctionnalités qui facilitent le développement de la transformation, comme les fonctions et les champs de listes de palettes et de réalisation du code.

Tâche	Instructions
Pour ajouter une fonction, procédez comme suit :	<p>Dans le volet Fonctions, double-cliquez sur la fonction à ajouter.</p> <p>Remarque : Les fonctions répertoriées dans l'éditeur sont conçues pour faciliter l'écriture de scripts de transformation personnalisée. Elles exécutent des fonctions qui nécessiteraient plusieurs lignes de code Groovy. Il ne s'agit pas de fonctions Groovy standard.</p>
Pour obtenir la valeur d'un champ de dataflow, procédez comme suit :	<p>Dans le volet Champs d'entrée, double-cliquez sur le champ d'entrée de votre choix. Ce qui suit sera ajouté à votre script :</p> <pre>data['FieldName']</pre> <p>Par exemple, si vous souhaitez obtenir la valeur du champ CurrentBalance, ce qui suit sera ajouté :</p> <pre>data['CurrentBalance']</pre>
Pour définir la valeur d'un champ de dataflow,	<p>Saisissez le code suivant dans l'éditeur de script :</p> <pre>data['FieldName']=NewValue</pre>

Tâche	Instructions
procédez comme suit :	<p>Par exemple, pour définir le champ Day sur le jour de la semaine contenu dans le champ PurchaseDate :</p> <pre>data['Day']=dayOfWeek(data['PurchaseDate'])</pre> <p>Dans cet exemple, la fonction <code>dayOfWeek()</code> permet d'obtenir le jour de la valeur de date du champ PurchaseDate et le résultat est écrit dans le champ Day.</p> <p>Conseil : Vous pouvez double-cliquer sur le nom du champ de sortie dans le volet Champ de sortie pour ajouter la référence de champ au script.</p>

Pour modifier la portée d'une variable de script dans un dataflow

Pour modifier la portée d'une variable de script d'un seul enregistrement d'entrée vers tous les enregistrements d'entrée d'un dataflow, utilisez l'annotation `@Field` dans votre script, comme indiqué :

```
import groovy.transform.Field;
@Field ['data type']['VariableName']= Value;
```

Par exemple, pour définir la portée d'une variable `RecordNumber` sur un seul enregistrement d'entrée, spécifiez ceci :

```
int recordNumber = 1;
data['Record_Number']= recordNumber;
recordNumber++;
```

La sortie sera :

Date	Record_Number
3/14/18	1
3/15/18	1
3/16/18	1
3/17/18	1
3/18/18	1
3/19/18	1
3/20/18	1
3/21/18	1
3/22/18	1
3/23/18	1

Pour modifier la portée de cette variable vers tous les enregistrements d'entrée, spécifiez ceci :

```
import groovy.transform.Field
@Field int recordNumber = 1;
data['Record_Number']= recordNumber;
recordNumber++;
```

La sortie sera :

Tâche Instructions

Date	Record_Number
3/14/18	1
3/15/18	2
3/16/18	3
3/17/18	4
3/18/18	5
3/19/18	6
3/20/18	7
3/21/18	8
3/22/18	9
3/23/18	10

Pour créer un nouveau champ à l'aide d'un type de données numériques

Saisissez le code suivant dans l'éditeur de script :

```
data['FieldName'] = new constructor;
```

Où *constructor* est l'un des éléments suivants :

java.lang.Double(*nombre*)

Crée un champ avec des données de type Double.

java.lang.Float(*nombre*)

Crée un champ avec des données de type Float.

java.lang.Integer(*nombre*)

Crée un champ avec des données de type Integer. Vous pouvez également créer un nouveau champ d'entier en indiquant un nombre complet. L'exemple suivant crée un champ d'entier de valeur 23 :

```
data['MyNewField'] = 23;
```

java.lang.Long(*nombre*)

Crée un champ avec des données de type Long.

Par exemple, pour créer un nouveau champ nommé « Transactions » avec un type de données Double et la valeur 23.10, vous pouvez indiquer comme suit :

```
data['Transactions'] = new com.java.lang.Double(23.10);
```

Pour créer un nouveau champ à l'aide d'un type de

Saisissez le code suivant dans l'éditeur de script :

```
data['FieldName'] = new constructor;
```

Où *constructor* est l'un des éléments suivants :

Tâche	Instructions
données Date ou Time :	<p>com.pb.spectrum.api.datetime.Date(<i>année,mois,jour</i>)</p> <p>Crée un champ avec des données de type Date. Par exemple, December 23, 2013 serait :</p> <pre>2013,12,23</pre> <p>com.pb.spectrum.api.datetime.Time(<i>heures,minutes,secondes</i>)</p> <p>Crée un champ avec des données de type Time. Par exemple, 4:15 PM serait :</p> <pre>16,15,0</pre> <p>.</p> <p>com.pb.spectrum.api.datetime.DateTime(<i>année,mois,jour,heures,minutes,secondes</i>)</p> <p>Crée un champ avec des données de type DateTime. Par exemple, 4:15 PM on December 23, 2013 serait :</p> <pre>2013,12,23,16,15,0</pre> <p>Par exemple, pour créer un nouveau champ nommé « TransactionDate » avec un type de données Date et la valeur December 23, 2013, vous devez indiquer ceci :</p> <pre>data['TransactionDate'] = new com.pb.spectrum.api.datetime.Date(2013,12,23);</pre>
Pour créer un nouveau champ avec des données de type Boolean	<p>Saisissez le code suivant dans l'éditeur de script :</p> <pre>data['FieldName'] = true or false;</pre> <p>Par exemple, pour créer un champ nommé IsValidated et le définir sur false, vous devez indiquer ceci :</p> <pre>data['IsValidated'] = false;</pre>
Pour créer un champ de liste	<p>Utilisez la méthode <code>factory.create()</code> pour créer de nouveaux champs dans un enregistrement, puis utilisez l'opérateur <code>leftShift <<</code> permettant d'ajouter le nouvel enregistrement dans le champ de liste.</p> <pre>NewListField = []</pre>

Tâche Instructions

```

NewRecord = factory.create()
NewRecord['NewField1'] = "Value"
NewRecord['NewField12'] = "Value"
...
NewListField << NewRecord

NewRecord = factory.create()
NewRecord['NewField1'] = "Value"
NewRecord['NewField12'] = "Value"
...
NewListField << NewRecord
data['ListOfRecords'] = NewListField

```

Par exemple, cela crée un nouveau champ de liste appelé « addresses », composé de deux enregistrements « address ».

```

addresses = []
address = factory.create()
address['AddressLine1'] = "123 Main St"
address['PostalCode'] = "12345"
addresses << address

address = factory.create()
address['AddressLine1'] = "PO Box 350"
address['PostalCode'] = "02134"
addresses << address
data['Addresses'] = addresses

```

Vous pouvez également créer un champ de liste contenant une liste de champs individuels plutôt qu'une liste d'enregistrements. Par exemple, cela crée un champ de liste appelé PhoneNumbers contenant des numéros de téléphone à domicile et professionnels :

```

phoneNumbers = []
phoneNumbers << data['HomePhone']
phoneNumbers << data['WorkPhone']
data['PhoneNumbers'] = phoneNumbers

```

Pour concaténer les champs, procédez comme suit :

Utilisez le symbole +. Cet exemple concatène le champ FirstName et le champ LastName en une valeur et la stocke dans le champ FullName

```

String fullname = data['FirstName'] + ' ' + data['LastName'];
data['FullName']=fullname;

```

Tâche	Instructions
	<p>Dans cet exemple, il existe deux champs d'entrée (AddressLine1 et AddressLine2) qui sont concaténés et écrits dans le champ de sortie Address.</p>
	<pre data-bbox="427 411 1549 506">address1 = data['AddressLine1']; address2 = data['AddressLine2']; data['Address']=address1+ ',' + address2;</pre>
<p>Pour analyser un champ, procédez comme suit :</p>	<p>Identifiez un caractère de séparation, puis utilisez <code>substring</code> pour analyser le champ. Dans l'exemple suivant, si le champ PostalCode est supérieur à cinq caractères, celui-ci séparera le code ZIP à cinq caractères et la portion +4, et les écrit sur des champs distincts dans l'enregistrement de sortie.</p>
	<pre data-bbox="427 787 1549 1110">if (data['PostalCode'].length() > 5) { String postalCode = data['PostalCode']; int separatorPosition = postalCode.indexOf('-'); String zip = postalCode.substring(0, separatorPosition); String plusFour = postalCode.substring(separatorPosition + 1, postalCode.length()); data['Zip']=zip; data['PlusFour']=plusFour; }</pre>
<p>Pour effectuer un traitement conditionnel, procédez comme suit :</p>	<p>Utilisez une instruction <code>if</code> ou <code>switch</code>. Il s'agit des constructions de traitement conditionnel les plus courantes. Pour plus d'informations, consultez le site Web groovy-lang.org.</p> <p>Cet exemple définit le champ AddressCity sur la première ligne d'adresse et nom de ville si la ville est Austin.</p>
	<pre data-bbox="427 1430 1549 1551">city = data['City']; address1 = data['AddressLine1'] if(city.equals('Austin')) data['AddressCity']=address1 +',' + city;</pre>
<p>Pour effectuer une mise en boucle, procédez</p>	<p>Utilisez la boucle <code>for</code>. Il s'agit de la seule construction de bouclage dont vous avez besoin. Pour plus d'informations sur le bouclage ou la syntaxe, consultez l'adresse suivante : groovy-lang.org.</p>

Tâche	Instructions
comme suit :	
Pour augmenter des données, procédez comme suit :	<p>Définissez une constante et utilisez le caractère de concaténation +. Par exemple, le script suivant ajoute le mot « Incorporated » à la fin du champ <code>FirmName</code>.</p> <pre>firmname = data['FirmName']; constant = 'Incorporated'; if(firmname.length() > 0) data['FirmName']=firmname + ' ' + constant;</pre>
Pour accéder à une option spécifiée lors de l'exécution, procédez comme suit :	<p>Si le dataflow présente des options d'exécution activées, vous pouvez accéder aux paramètres transmis au dataflow lors de l'exécution à l'aide de la syntaxe suivante :</p> <pre>options.get("optionName")</pre> <p>Par exemple, pour accéder à une option nommée <code>casing</code>, vous devez inclure cela dans votre script de transformation personnalisée :</p> <pre>options.get("casing")</pre>

- Après avoir terminé de saisir votre script, cliquez sur le bouton « X » dans la fenêtre pour fermer l'éditeur.
- Dans le champ **Input fields**, sélectionnez le ou les champs auxquels appliquer la transformation.
- Dans le champ **Output fields**, spécifiez le champ dans lequel vous souhaitez écrire la sortie de la transformation. Si nécessaire, vous pouvez définir un nouveau champ en cliquant sur le bouton **Ajouter** à droite du champ **Champs de sortie**.
- Lorsque vous avez terminé, cliquez sur le bouton **Ajouter** en bas de la fenêtre.
- Cliquez sur **OK**.

Utilisation d'une transformation Masquer

Vous pouvez utiliser le stage Transformer pour appliquer une transformation de masque à un champ. Une transformation de masque applique des caractères à un champ ou supprime les caractères d'un champ, à l'aide d'un modèle spécifié. Par exemple, à l'aide d'une transformation de masque, vous pouvez formater une chaîne de chiffres, telle que `8003685806`, en un numéro de téléphone comme suit : `(800) 368 5806`.

- Dans Enterprise Designer, faites glisser un stage Transformer sur le canevas et connectez-le à l'emplacement désiré.

2. Double-cliquez sur le stage Transformer.
3. Cliquez sur **Ajouter**.
4. Développez **Formatage** et sélectionnez **Masque**.
5. Sélectionnez le type de masque à utiliser.

Appliquer Ajoute des caractères à un champ pour former la chaîne en un nouveau modèle.

Supprimer Extrait un modèle de caractères d'une chaîne.

6. Dans le champ **Mask string**, indiquez le modèle à utiliser lors de l'ajout ou de la suppression de caractères.

Il existe deux types de caractères que vous utilisez lors de la spécification de la chaîne de masque : caractères littéraux et caractères de masque.

Les caractères littéraux représentent les caractères effectivement présents dans la chaîne. Lorsque qu'un masque Supprimer est employé, le caractère entrant doit correspondre au caractère littéral de manière exacte. Si c'est le cas, alors ils seront supprimés de l'entrée. De même, les caractères littéraux seront ajoutés à l'entrée à la position indiquée par la définition du masque lorsque le masque Appliquer est utilisé.

L'autre type de caractère pouvant être utilisé dans une chaîne de masque est un caractère de masque. Un caractère de masque indique le type de caractère qui peut se situer dans un emplacement particulier de la chaîne d'entrée. Par exemple, si vous possédez une entrée dont le premier caractère est un nombre, le premier caractère de masque doit être #. Tout élément d'entrée qui correspond à ce caractère de masque sera conservé dans la sortie.

La table suivante répertorie les caractères de masque que vous pouvez utiliser dans le champ **Mask string** :

Tableau 8 : Caractères de masque

Caractère	Définition
#	Tout nombre.
'	Caractère d'échappement, servant à échapper n'importe quel caractère de mise en forme spécial.
U	Tout caractère. Toutes les lettres en minuscules sont mappées aux majuscules.
L	Tout caractère. Toutes les lettres en majuscules sont mappées aux minuscules.

Caractère	Définition
A	Tout caractère ou nombre
?	Tout caractère.
*	N'importe quoi.
H	Tout caractère hexadécimal (0-9, a-f ou A-F).

7. Cliquez sur **Ajouter**.
8. Cliquez sur **OK**.

Exemples de transformation Masquer

Il s'agit d'un masque Appliquer qui applique le formatage à une chaîne. « (» et «) » et <espace> sont des littéraux, ils sont ajoutés à la sortie. Tous les nombres seront conservés, car # est un caractère MASK.

Entrée : 8003685806 **Chaîne de masque :** (###) ### ####

Sortie : (800) 368 5806

L'exemple suivant est un masque de suppression qui supprime le trait d'union dans le code postal.

Entrée : 60510-1135 **Chaîne de masque :** *****-*****

Sortie : 605101135

Unique ID Generator

Le stage Unique ID Generator crée une clé unique qui identifie un enregistrement spécifique. L'affectation d'un identifiant unique s'avère cruciale pour les initiatives d'entrepôt de données pour lesquelles les transactions ne peuvent pas porter toutes les données d'adresse et de nom, mais doivent néanmoins être attribuées au même enregistrement/contact. Un identifiant unique peut être appliqué au niveau de l'individu, du foyer, de la société et/ou de l'établissement. Unique ID Generator (générateur unique d'identifiant) propose plusieurs algorithmes pour créer des ID uniques.

L'identifiant unique s'appuie soit sur un numéro séquentiel ou un tampon Date/Heure. Par ailleurs, vous pouvez éventuellement utiliser une variété d'algorithmes pour générer les données à ajouter

à l'ID, augmentant ainsi la possibilité pour l'ID d'être unique. Les identifiants de type numéro séquentiel ou tampon Date/Heure sont obligatoires et ne peuvent pas être supprimés de l'identifiant généré.

Unique ID Generator permet de générer une clé non unique à l'aide d'un des algorithmes de génération de clé. Dans le mode non unique, vous pouvez créer des clés à utiliser pour la correspondance. Cela peut être utile dans un entrepôt de données dans lequel vous avez déjà ajouté des clés à une dimension et si vous souhaitez générer une clé pour les nouveaux enregistrements, afin de voir si les nouveaux enregistrements correspondent à un enregistrement existant.

L'exemple suivant montre que chaque enregistrement de l'entrée reçoit un identifiant d'enregistrement séquentiel en sortie.

Enregistrement	RecordID
John Smith	0
Mary Smith	1
Jane Doe	2
John Doe	3

Le stage Unique ID produit un champ nommé RecordID qui contient l'ID unique. Vous pouvez renommer le champ RecordID, le cas échéant.

Définition d'un identifiant unique

Par défaut, le stage Unique ID Generator crée un ID séquentiel, avec le premier enregistrement doté d'un ID de 0, le second enregistrement doté d'un ID de 1, le troisième enregistrement doté d'un ID de 2, etc. Pour modifier la manière dont l'ID unique est généré, suivez cette procédure.

1. Dans le stage Unique ID Generator, sous l'onglet **Règles**, cliquez sur **Modifier**.
2. Choisissez la méthode à utiliser pour générer l'ID unique.

Options	Description
Le tag numérique séquentiel débute à	Affecte une valeur numérique incrémentielle à chaque enregistrement en commençant par le nombre que vous indiquez. Si vous indiquez 0, le premier enregistrement aura un ID de 0, le deuxième aura un ID de 1, etc.

Options	Description
Balise numérique séquentielle commençant à la valeur dans un champ de base de données	

Options

Description

Affecte une valeur numérique incrémentielle à chaque enregistrement en commençant par le nombre maximal de lectures à partir du champ de base de données. Ce nombre est ensuite incrémenté de 1 et affecté au premier enregistrement. Par exemple, si le nombre de lectures du champ de base de données est 30, le premier enregistrement aura un ID de 31, le deuxième enregistrement aura un ID de 32, etc.

Connexion Sélectionnez la connexion de base de données à utiliser. Vos choix varient en fonction des connexions qui sont définies dans le Gestionnaire de connexion de Management Console. Si vous devez établir une nouvelle base de données, ou modifier ou supprimer une connexion existante, cliquez sur **Gérer**.

Si vous ajoutez ou modifiez une connexion de base de données, remplissez ces champs :

Nom de la connexion

Saisissez le nom de la connexion.: Ce nom est entièrement de votre choix.

Pilote de la base de données

Sélectionnez le type de base de données approprié.

Options de connexion

Indiquez l'hôte, le port, l'instance, le nom d'utilisateur et le mot de passe à utiliser pour vous connecter à la base de données.

Vue de la table Indiquez la table ou la vue de la base de données que vous souhaitez interroger.

Champ Database Sélectionnez une colonne à partir de la liste pour générer une clé unique.

Les types de données pris en charge pour la génération d'ID unique sont les suivants :

long Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre -2^{63} (-9,223,372,036,854,775,808) et $2^{63}-1$ (9,223,372,036,854,775,807).

integer Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre -2^{31} (-2,147,483,648) et $2^{31}-1$ (2,147,483,647).

bigdecimal Type de données numérique prenant en charge 38 points de précision décimaux. Utilisez ce type de données pour les données utilisées dans les calculs mathématiques requérant un haut niveau de précision, en particulier ceux impliquant des données financières. Le type de données bigdecimal prend en charge des calculs plus précis que le type de données double.

double Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre 2^{-1074} et $(2-2^{-52}) \times 2^{1023}$. En

Options	Description
	notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.
flottant	Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre 2^{-149} et $(2-2^{-23}) \times 2^{127}$. En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.
Horodatage	Crée une clé unique en fonction de l'horodatage au lieu de la numérotation séquentielle.
UUID	Crée une clé d'identifiant de 32 chiffres universellement unique pour chaque enregistrement. Les chiffres de la clé sont affichés en cinq groupes séparés par des traits d'union, sous la forme 8-4-4-4-12, pour un total de 36 caractères (32 caractères alphanumériques et quatre traits d'union). Exemple : 123e4567-e89b-12d3-a456-432255330000
Désactivé	Sélectionnez cette option uniquement si vous souhaitez générer une clé non unique à l'aide d'un algorithme.

3. Cliquez sur **OK**.

Utilisation d'algorithmes pour augmenter un ID unique

Unique ID Generator génère un ID unique pour chaque enregistrement en numérotant chaque enregistrement de manière séquentielle ou en générant un tampon date/heure pour chaque enregistrement. Vous pouvez éventuellement utiliser des algorithmes pour ajouter des informations supplémentaires à l'ID unique séquentiel ou d'horodatage, créant ainsi un ID unique plus complexe et qui est plus susceptible d'être réellement unique.

1. Dans le stage Unique ID Generator, cliquez sur **Ajouter**.
2. Dans le champ **Algorithm**, sélectionnez l'algorithme à utiliser pour générer d'autres informations dans l'ID. L'un des éléments suivants :

Consonne	Renvoie les champs indiqués, les consonnes étant supprimées.
Metaphone double	Renvoie un code basé sur la représentation phonétique de leurs caractères. Le double Metaphone est une version améliorée de l'algorithme Metaphone et tente de prendre en compte les nombreuses irrégularités de plusieurs langues.
Koeln	Noms d'index par son, tels qu'ils sont prononcés en allemand. Permet aux noms ayant la même prononciation d'être encodés avec la même représentation afin qu'ils puissent être mis en correspondance, en dépit de différences mineures au niveau de l'orthographe. Le résultat est toujours une séquence de nombres ;

les caractères spéciaux et les espaces blancs sont ignorés. Cette option a été développée en réponse aux limites du Soundex.

- MD5** Algorithme qui produit une valeur hash de 128 bits. Cet algorithme est généralement utilisé pour vérifier l'intégrité des données.
- Metaphone** Renvoie une clé codée Metaphone des champs sélectionnés. Metaphone est un algorithme qui code les mots à l'aide de leur sonorité lorsque prononcé en anglais.
- Metaphone (Espagnol)** Renvoie une clé codée Metaphone des champs sélectionnés pour la langue espagnole. Cet algorithme Metaphone code les mots à l'aide de leur sonorité lorsque prononcé en espagnol.
- Metaphone3** Procède à une amélioration en fonction des algorithmes Metaphone et Double Metaphone avec des paramètres de consonne et de voyelle interne exacts qui vous permet de produire des mots ou des noms mis en correspondance de manière plus ou moins proche pour rechercher des termes au niveau phonétique. Metaphone 3 augmente l'exactitude de l'encodage phonétique à 98 %. Cette option a été développée en réponse aux limites du Soundex.
- Nysiis** L'algorithme de code phonétique qui met en correspondance une prononciation approximative avec une orthographe exacte et indexe des mots prononcés de manière similaire. Fait partie du système New York State Identification and Intelligence System. Imaginons, par exemple, que vous recherchez des informations sur une personne dans une base de données de personnes. Vous pensez que le nom de la personne sonne comme « John Smith », mais il est en fait orthographié « Jon Smyth ». Si vous procédez à une recherche de la correspondance exacte de « John Smith », aucun résultat n'est renvoyé. Cependant, si vous indexez la base de données à l'aide de l'algorithme NYSIIS et procédez à une recherche en utilisant de nouveau l'algorithme NYSIIS, la correspondance correcte est renvoyée car « John Smith » et « Jon Smyth » sont indexés comme « JAN SNATH » par l'algorithme.
- Phonix** Pré-traite les chaînes de nom en appliquant plus de 100 règles de transformation à des caractères uniques ou à des séquences de plusieurs caractères. 19 de ces règles s'appliquent uniquement si les caractères figurent au début de la chaîne, tandis que 12 des règles s'appliquent uniquement si les caractères figurent au milieu de la chaîne et 28 des règles s'appliquent uniquement si les caractères figurent à la fin de la chaîne. La chaîne de nom transformée est cryptée en un code composé d'une lettre au début, suivie de trois chiffres (en enlevant les zéros et les nombres en double). Cette option a été développée pour répondre aux limites de Soundex ; elle est plus complexe et donc plus lente que Soundex.
- Soundex** Renvoie un code Soundex des champs sélectionnés. Soundex produit un code de longueur fixe en s'appuyant sur la sonorité du mot lorsque prononcé en anglais.
- Substring** Renvoie une partie spécifiée du champ sélectionné.

3. Dans le champ **Field name**, choisissez le champ auquel appliquer l'algorithme. Par exemple, si vous choisissez l'algorithme soundex et un champ nommé City, l'ID sera généré en appliquant l'algorithme soundex aux données contenues dans le champ City.
4. Si vous avez sélectionné l'algorithme de sous-chaîne, indiquez la partie du champ à utiliser dans la sous-chaîne :
 - a) Dans le champ **Start position**, indiquez la position dans le champ où la sous-chaîne doit commencer.
 - b) Dans le champ **Length**, sélectionnez le nombre de caractères à partir de la position de départ que vous souhaitez inclure dans la sous-chaîne.

Par exemple, imaginons que vous disposez des données suivantes dans un champ nommé Nom de famille :

Augustine

Si vous avez indiqué 3 comme position de départ et 6 comme position de fin, la sous-chaîne produirait ce qui suit :

gustin

5. Cochez la case **Supprimer les caractères de bruit** pour supprimer tous les caractères non numériques et non alpha, tels que les traits d'union, les espaces blancs et autres caractères spéciaux du champ avant d'appliquer l'algorithme.
6. Pour les algorithmes de consonne et de sous-chaîne, vous pouvez trier les données dans le champ avant d'appliquer l'algorithme en cochant la case **Sort input**. Vous pouvez alors choisir de trier les caractères dans le champ ou les termes dans le champ par ordre alphabétique.
7. Cliquez sur **OK** pour enregistrer vos paramètres.
8. Répétez les étapes autant de fois que nécessaire, si vous souhaitez ajouter d'autres algorithmes pour produire un ID plus complexe.

Remarque : La définition de clé unique est toujours affichée d'une couleur différente et ne peut pas être supprimée.

Définition d'un ID non unique

Unique ID Generator permet de générer une clé non unique à l'aide d'un des algorithmes de génération de clé. Dans le mode non unique, vous pouvez créer des clés à utiliser pour la correspondance. Cela peut être utile dans un entrepôt de données dans lequel vous avez déjà ajouté des clés à une dimension et si vous souhaitez générer une clé pour les nouveaux enregistrements, afin de voir si les nouveaux enregistrements correspondent à un enregistrement existant.

1. Dans le stage Unique ID Generator, sous l'onglet **Règles**, cliquez sur **Modifier**.
2. Sélectionnez **Désactivé**.

Cela désactive la partie ID unique des règles de génération d'ID. Avec cette option désactivée, seul l'algorithme que vous choisissez dans les étapes suivantes sera utilisé pour créer l'ID. Cela

signifie que tout enregistrement doté des mêmes données dans les champs que vous utilisez pour générer l'ID disposera du même ID. Vous pouvez ensuite utiliser l'ID pour la mise en correspondance.

3. Cliquez sur **OK**.
4. À l'invite d'avertissement, cliquez sur **Oui**.
5. Dans le stage Unique ID Generator, cliquez sur **Ajouter**.
6. Dans le champ **Algorithm**, sélectionnez l'algorithme à utiliser pour générer d'autres informations dans l'ID. L'un des éléments suivants :

Consonne	Revoie les champs indiqués, les consonnes étant supprimées.
Metaphone double	Revoie un code basé sur la représentation phonétique de leurs caractères. Le double Metaphone est une version améliorée de l'algorithme Metaphone et tente de prendre en compte les nombreuses irrégularités de plusieurs langues.
Koeln	Noms d'index par son, tels qu'ils sont prononcés en allemand. Permet aux noms ayant la même prononciation d'être encodés avec la même représentation afin qu'ils puissent être mis en correspondance, en dépit de différences mineures au niveau de l'orthographe. Le résultat est toujours une séquence de nombres ; les caractères spéciaux et les espaces blancs sont ignorés. Cette option a été développée en réponse aux limites du Soundex.
MD5	Algorithme qui produit une valeur hash de 128 bits. Cet algorithme est généralement utilisé pour vérifier l'intégrité des données.
Metaphone	Revoie une clé codée Metaphone des champs sélectionnés. Metaphone est un algorithme qui code les mots à l'aide de leur sonorité lorsque prononcé en anglais.
Metaphone (Espagnol)	Revoie une clé codée Metaphone des champs sélectionnés pour la langue espagnole. Cet algorithme Metaphone code les mots à l'aide de leur sonorité lorsque prononcé en espagnol.
Metaphone3	Procède à une amélioration en fonction des algorithmes Metaphone et Double Metaphone avec des paramètres de consonne et de voyelle interne exacts qui vous permet de produire des mots ou des noms mis en correspondance de manière plus ou moins proche pour rechercher des termes au niveau phonétique. Metaphone 3 augmente l'exactitude de l'encodage phonétique à 98 %. Cette option a été développée en réponse aux limites du Soundex.
Nysiis	L'algorithme de code phonétique qui met en correspondance une prononciation approximative avec une orthographe exacte et indexe des mots prononcés de manière similaire. Fait partie du système New York State Identification and Intelligence System. Imaginons, par exemple, que vous recherchez des informations sur une personne dans une base de données de personnes. Vous pensez que le nom de la personne sonne comme « John Smith », mais il est en fait orthographié « Jon Smyth ». Si vous procédez à une recherche de la correspondance exacte de « John Smith », aucun résultat n'est renvoyé. Cependant, si vous indexez la base de données à l'aide de l'algorithme NYSIIS

et procédez à une recherche en utilisant de nouveau l'algorithme NYSIIS, la correspondance correcte est renvoyée car « John Smith » et « Jon Smyth » sont indexés comme « JAN SNATH » par l'algorithme.

- Phonix** Pré-traite les chaînes de nom en appliquant plus de 100 règles de transformation à des caractères uniques ou à des séquences de plusieurs caractères. 19 de ces règles s'appliquent uniquement si les caractères figurent au début de la chaîne, tandis que 12 des règles s'appliquent uniquement si les caractères figurent au milieu de la chaîne et 28 des règles s'appliquent uniquement si les caractères figurent à la fin de la chaîne. La chaîne de nom transformée est cryptée en un code composé d'une lettre au début, suivie de trois chiffres (en enlevant les zéros et les nombres en double). Cette option a été développée pour répondre aux limites de Soundex ; elle est plus complexe et donc plus lente que Soundex.
- Soundex** Renvoie un code Soundex des champs sélectionnés. Soundex produit un code de longueur fixe en s'appuyant sur la sonorité du mot lorsque prononcé en anglais.
- Substring** Renvoie une partie spécifiée du champ sélectionné.

7. Dans le champ **Field name**, choisissez le champ auquel appliquer l'algorithme. Par exemple, si vous choisissez l'algorithme soundex et un champ nommé City, l'ID sera généré en appliquant l'algorithme soundex aux données contenues dans le champ City.
8. Si vous avez sélectionné l'algorithme de sous-chaîne, indiquez la partie du champ à utiliser dans la sous-chaîne :
 - a) Dans le champ **Start position**, indiquez la position dans le champ où la sous-chaîne doit commencer.
 - b) Dans le champ **Length**, sélectionnez le nombre de caractères à partir de la position de départ que vous souhaitez inclure dans la sous-chaîne.

Par exemple, imaginons que vous disposez des données suivantes dans un champ nommé Nom de famille :

Augustine

Si vous avez indiqué 3 comme position de départ et 6 comme position de fin, la sous-chaîne produirait ce qui suit :

gustin

9. Cochez la case **Supprimer les caractères de bruit** pour supprimer tous les caractères non numériques et non alpha, tels que les traits d'union, les espaces blancs et autres caractères spéciaux du champ avant d'appliquer l'algorithme.
10. Pour les algorithmes de consonne et de sous-chaîne, vous pouvez trier les données dans le champ avant d'appliquer l'algorithme en cochant la case **Sort input**. Vous pouvez alors choisir de trier les caractères dans le champ ou les termes dans le champ par ordre alphabétique.
11. Cliquez sur **OK** pour enregistrer vos paramètres.

12. Répétez les étapes autant de fois que nécessaire, si vous souhaitez ajouter d'autres algorithmes pour produire un ID plus complexe.

Remarque : La définition de clé unique est toujours affichée d'une couleur différente et ne peut pas être supprimée.

Sortie d'un flux

Pour définir la sortie d'un flux de données, utilisez un stage « récepteur ». Un récepteur est le dernier stage d'un flux de données. Il définit ce qu'il faut faire avec la sortie du flux de données. Un récepteur peut également effectuer d'autres actions à la fin ou un flux de données, par exemple, exécuter un programme.

Sortie d'un job

La sortie d'un job peut être écrite dans un fichier ou une base de données. Spectrum™ Technology Platform peut écrire des données dans de nombreux formats de fichier et types de base de données. Les types de récepteur dans lesquels vous pouvez écrire dépendent des modules pour lesquels vous disposez d'une licence. Reportez-vous au Guide sur les solutions de vos modules, disponibles sur support.pb.com/spectrum.

Sortie d'un service

Les données de sortie d'un service sont définies dans un stage Output. Ce stage définit les champs que le service renvoie en réponse à une requête de service Web ou un appel d'API.

Définition de la sortie d'un service

Le stage Output définit les champs de sortie que le service ou sous-flux renvoie. Suivez la procédure ci-dessous pour définir la sortie de service.

1. Double-cliquez sur l'icône Résultat du canevas. La boîte de dialogue **Options de résultat** apparaît. Quand vous ouvrez la boîte de dialogue **Options de résultat** pour la première fois, une liste des champs définis dans l'entrée s'affiche.
2. Pour ajouter un nouveau champ à la liste de champs, cliquez sur **Ajouter**. La boîte de dialogue **Ajouter un champ personnalisé** apparaît. Vous pouvez aussi modifier ou supprimer un champ personnalisé.
3. Cliquez encore une fois sur **Ajouter**.
4. Saisissez le nom du champ dans la zone de texte.
5. Sélectionnez le **Type de données** et appuyez sur **OK**. Les types de données suivants sont pris en charge :

- bigdecimal** Type de données numérique prenant en charge 38 points de précision décimaux. Utilisez ce type de données pour les données utilisées dans les calculs mathématiques requérant un haut niveau de précision, en particulier ceux impliquant des données financières. Le type de données bigdecimal prend en charge des calculs plus précis que le type de données double.
- booléen** Type logique doté de deux valeurs : True et False.
- bytearray** Tableau (liste) d'octets.
- Remarque :** ByteArray n'est pas pris en charge en entrée pour un service REST.
- date** Type de données contenant un mois, un jour et une année. Par exemple, 2012-01-30 ou January 30, 2012. Vous pouvez indiquer un format de date par défaut dans Management Console.
- datetime** Type de données contenant un mois, un jour, une année et des heures, des minutes et des secondes. Par exemple, 2012/01/30 6:15 PM.
- double** Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre 2^{-1074} et $(2-2^{-52}) \times 2^{1023}$. En notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.
- flottant** Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre 2^{-149} et $(2-2^{23}) \times 2^{127}$. En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.
- integer** Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre -2^{31} (-2,147,483,648) et $2^{31}-1$ (2,147,483,647).
- liste** Au sens strict du terme, une liste n'est pas un type de données. Cependant, lorsqu'un champ contient des données hiérarchiques, il est traité comme un champ de « liste ». Dans Spectrum™ Technology Platform, une liste représente un ensemble de données composées de plusieurs valeurs. Par exemple, un champ Noms peut contenir une liste de valeurs de noms. Cela peut être représenté sous forme d'une structure XML, comme suit :

```
<Names>
  <Name>John Smith</Name>
  <Name>Ann Fowler</Name>
</Names>
```

Il est important de souligner que le type de données de liste Spectrum™ Technology Platform est différent du type de données de liste de schéma XML dans le sens où ce dernier est un type de données simple composé de plusieurs valeurs, tandis que le type de données de liste Spectrum™ Technology Platform est semblable à un type de données complexe XML.

long	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre -2^{63} (-9,223,372,036,854,775,808) et $2^{63}-1$ (9,223,372,036,854,775,807).
chaîne	Séquence de caractères.
time	Type de données contenant l'heure du jour. Par exemple, 21:15:59 ou 9:15:59 PM.

Vous pouvez également ajouter un nouveau type de données défini par l'utilisateur si nécessaire, et ce nouveau type peut provenir d'une liste de tout type de données défini. Par exemple, vous pouvez définir une liste de noms (chaîne) ou un nouveau type de données d'adresses qui inclut AddressLine1 (chaîne), City (chaîne), StateProvince (chaîne) et PostalCode (chaîne). Après avoir créé le champ, vous pouvez afficher le type de données en accédant à la boîte de dialogue Options d'entrée et en appuyant sur le bouton dans la colonne Type de donnée. La boîte de dialogue **Détails de type de données** apparaîtra, affichant la structure du champ.

6. Cliquez de nouveau sur **OK**.
7. Cliquez sur la case à cocher en regard de **Exposer** pour sélectionner les cases de tous les champs de la liste. Quand vous sélectionnez un champ dans la liste des champs, vous l'exposez au flux de données pour les opérations de stages. Cliquez sur la case à cocher une nouvelle fois pour désactiver les cases de tous les champs de la liste. Désactivez la case à cocher d'un ou plusieurs champs de la liste des champs et cliquez sur **OK** pour supprimer le champ de la liste.

Remarque : Si vous définissez des données hiérarchiques dans les champs d'entrée, vous ne pourrez pas importer de données ou afficher les données verticalement.

8. Cliquez sur **OK** pour retourner au canevas.

Définir un type de données de service Web

Le champ **Data type name** vous permet de contrôler les interfaces WSDL (SOAP) et WADL (REST) pour le service que vous êtes en train de créer. Le nom de l'élément Lignes est déterminé par le nom que vous donnez à ce stage dans le service, et le nom de l'élément Ligne est déterminé par le texte que vous saisissez ici.

Remarque : Pour WSDL, à la fois les requêtes et les réponses sont affectées, mais pour WADL, seules les réponses le sont.

Avant de nommer ce stage et d'entrer un texte dans ce champ, votre code ressemble peut-être à ceci :

```
<Rows>
  <Row>
    <FirstName>John</FirstName>
    <LastName>Doe</LastName>
  </Row>
  <Row>
```

```

    <FirstName>Jane</FirstName>
  <LastName>Doe</LastName>
</Row>
</Rows>

```

Après avoir nommé ce stage et entré un texte dans ce champ, votre code ressemble peut-être à ceci :

```

<Names>
  <Name>
    <FirstName>John</FirstName>
    <LastName>Doe</LastName>
  </Name>
  <Name>
    <FirstName>Jane</FirstName>
    <LastName>Doe</LastName>
  </Name>
</Names>

```

Exécution d'un programme externe

Un stage *Execute Program* appelle un exécutable, tel qu'un programme ou une commande de ligne de commande, quand il reçoit un enregistrement. Pour utiliser un stage *Execute Program* dans votre flux de données :

Options

Option	Description				
Ligne de commande	Nom et arguments de l'exécutable (le cas échéant). Les arguments peuvent être des données disponibles dans le flux de données. Pour accéder à ces données, cliquez sur le bouton [...] (Parcourir). Vous pouvez sélectionner l'un des trois contextes suivants : ID du job actuel, Nom du job actuel ou Nom d'utilisateur actuel. Vous pouvez également faire une sélection parmi les champs disponibles. Par exemple, JobStatus et JobComment.				
Timeout	Détermine si l'exécution doit être annulée si la commande ne répond pas dans un délai donné. L'un des éléments suivants : <table border="0" data-bbox="552 1680 1425 1837"> <tr> <td>Pas de délai d'attente</td> <td>Ne pas annuler l'exécution si la commande ne parvient pas à répondre.</td> </tr> <tr> <td>Délai d'attente millisecondes</td> <td>Annule la tentative d'exécution si la commande ne répond pas sous le nombre de millisecondes spécifié.</td> </tr> </table>	Pas de délai d'attente	Ne pas annuler l'exécution si la commande ne parvient pas à répondre.	Délai d'attente millisecondes	Annule la tentative d'exécution si la commande ne répond pas sous le nombre de millisecondes spécifié.
Pas de délai d'attente	Ne pas annuler l'exécution si la commande ne parvient pas à répondre.				
Délai d'attente millisecondes	Annule la tentative d'exécution si la commande ne répond pas sous le nombre de millisecondes spécifié.				

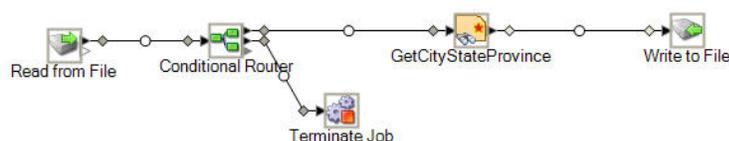
Option	Description
Variables d'environnement	<p>Facultatif. Indique les variables d'environnement à utiliser lors de l'exécution de la commande. Pour ajouter une variable d'environnement, cliquez sur Ajouter.</p> <p>Saisissez le mot clé approprié dans le champ Key. Par exemple, « JAVA_HOME ».</p> <p>Saisissez la valeur appropriée dans le champ Value. Par exemple, C:\Java\jre7. Vous pouvez aussi sélectionner un champ à partir de la boîte de dialogue Liste de champs en cliquant sur le bouton [...] (Parcourir). Vous pouvez sélectionner l'un des trois contextes suivants : ID du job actuel, Nom du job actuel ou Nom d'utilisateur actuel. Vous pouvez également faire une sélection parmi les champs disponibles. Par exemple, JobStatus et JobComment.</p>

Arrêt d'un job en fonction d'une condition

Le stage Terminate Job, associé à Conditional Router, permet de terminer un job si certains critères sont atteints au sein d'un enregistrement. Si un enregistrement est envoyé à Terminate Job, le job se termine.

Remarque : Terminate Job n'est pas disponible dans les services ou les sous-flux.

Pour utiliser Terminate Job, ajoutez un Conditional Router et un stage Terminate Job à votre flux de données. Connectez ensuite les stages et configurez Conditional Router. Le Router conditionnel doit être configuré pour contenir les critères que vous voulez pour déclencher la fin du job. Quand un enregistrement qui satisfait les critères est trouvé, celui-ci est transmis à Terminate Job et le job se termine, produisant un message qui indique « Job terminé par le stage : <libellé du stage> ». Le flux de données complété devrait ressembler à cela :



Ignorer des enregistrements

Le stage Write to Null annule les enregistrements. Les enregistrements sont comptés mais abandonnés. Utilisez ce stage si vous disposez d'enregistrements que vous ne souhaitez pas préserver une fois que le flux de données est terminé.

Flux de données intégrés

Un flux de données intégré réduit le nombre de stages affichés simultanément sur le canevas Enterprise Designer en regroupant les stages. Les stages regroupés apparaissent alors sous la forme d'un seul stage. Vous pouvez utiliser des flux de données intégrés aux fins suivantes :

- Pour simplifier la présentation de flux de données complexes en regroupant les stages et en les représentant sous la forme d'un seul stage sur le canevas.
- Pour traiter des enregistrements en groupes via la fonction d'itération.
- Pour utiliser une valeur dans un champ pour définir les options de stage du flux de données intégré.

Vous pouvez ajouter un nombre illimité de flux de données intégrés à un flux de données, et vous pouvez également placer des flux de données intégrés dans des flux de données intégrés.

Différences entre les flux de données intégrés et les sous-flux

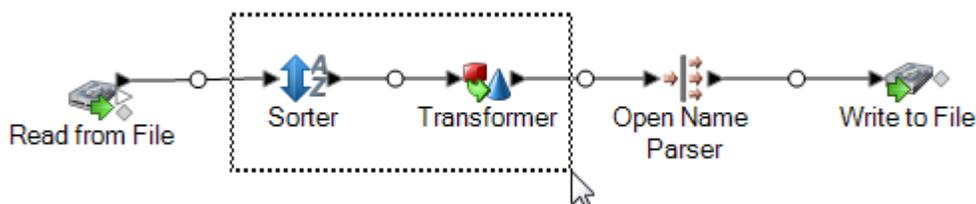
Il existe deux différences fondamentales entre un flux de données intégré et un sous-flux. La première différence réside dans le fait que l'itération est disponible uniquement dans les flux de données intégrés. L'itération vous permet de traiter ensemble des groupes d'enregistrements, par exemple, pour agréger des enregistrements à des fins de traitement ou pour définir des options de stage en fonction de valeurs de champ. La deuxième différence entre un flux de données intégré et un sous-flux réside dans le fait qu'un flux de données intégré ne peut pas être utilisé dans plus d'un flux de données. Pour réutiliser une portion d'un flux de données intégré dans plusieurs flux de données, vous devez créer un sous-flux au lieu d'un flux de données intégré. Si vous souhaitez réutiliser un flux de données intégré dans d'autres flux de données, vous pouvez le convertir en sous-flux, mais, dans ce cas, ses options d'itération sont supprimées.

Regroupement de stages dans un flux de données intégré

Un flux de données intégré regroupe des stages en un seul stage, vous permettant ainsi de simplifier la présentation de flux de données complexes et de définir des options de traitement à l'aide de valeurs de champ.

1. Dans votre flux de données, ajoutez les stages que vous souhaitez convertir en flux de données intégré.
2. Sélectionnez les stages que vous souhaitez convertir en flux de données intégré en cliquant sur un cadre autour des stages et en le faisant glisser.

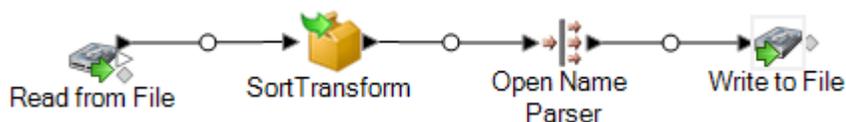
Par exemple, cette opération sélectionne les stages Sorter et Transformer pour les convertir en flux de données intégré.



Remarque : Les stages Report ne peuvent pas être inclus dans un flux de données intégré.

3. Faites un clic droit sur les stages sélectionnés, puis sélectionnez **Regrouper dans un flux de données intégré**
4. Saisissez un nom pour le flux de données intégré. Le nom est utilisé comme intitulé du stage du flux de données intégré sur le canevas.
5. Cliquez sur **OK**.

Les stages que vous avez sélectionnés sont maintenant regroupés dans un flux de données intégré. L'exemple suivant illustre un flux de données intégré nommé SortTransform.

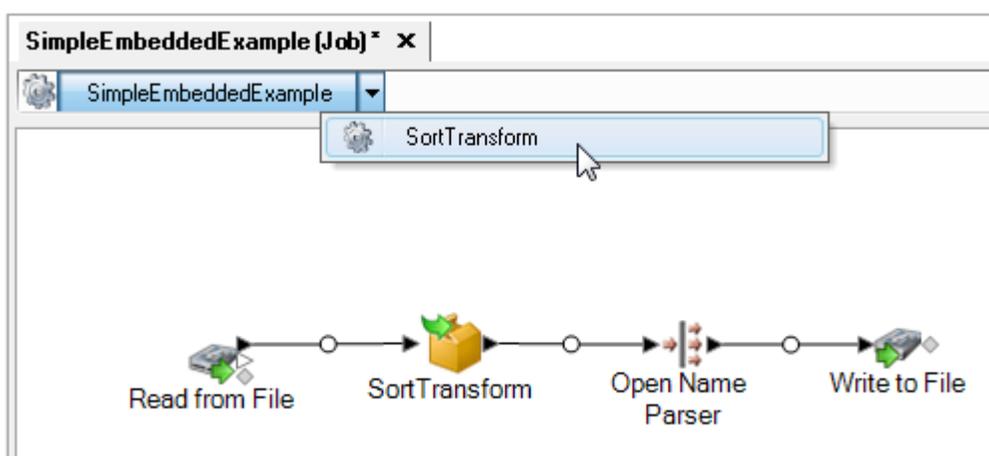


Modification d'un flux de données intégré

Un flux de données intégré regroupe des stages en un seul stage, vous permettant ainsi de simplifier la présentation de flux de données complexes et de définir des options de traitement à l'aide de valeurs de champ.

1. Faites un clic droit sur l'icône, puis sélectionnez **Modifier le flux de données intégré**.

Conseil : Sinon, vous pouvez utiliser les liens de navigation figurant tout en haut du flux de données pour ouvrir des flux de données intégrés. L'exemple suivant indique comment ouvrir un dataflow intégré nommé SortTransform :



Lorsque vous ouvrez un flux de données intégré, vous voyez apparaître un stage Input et un stage Output. Ces stages représentent l'entrée du flux de données intégré depuis le flux de données parent et la sortie du flux de données parent depuis le flux de données intégré.

2. Apportez les modifications nécessaires au flux de données intégré.

Toute modification que vous apportez à un flux de données intégré est enregistré lors de l'enregistrement suivant du flux de données parent.

Utilisation de l'itération avec un dataflow intégré

Les paramètres d'itération indiquent la manière dont un dataflow intégré doit traiter les enregistrements entrants. Par défaut, un dataflow intégré traite individuellement chaque enregistrement, exactement comme le ferait tout autre stage du dataflow. En revanche, si vous utilisez l'itération, vous pouvez traiter des groupes d'enregistrements ensemble, ce qui peut s'avérer utile lorsque, par exemple, vous effectuez des comparaisons ou des calculs en fonction de groupes d'enregistrements et non de l'ensemble des données d'entrée. Vous pouvez également utiliser l'itération pour appliquer des options de stage en fonction des données de chaque enregistrement.

Il existe deux types d'itération : l'itération par enregistrement et l'itération par groupe. Dans l'itération par enregistrement, un dataflow intégré traite un enregistrement à la fois et le résultat est envoyé au stage suivant du dataflow intégré. L'itération par enregistrement s'avère utile si vous souhaitez définir des options de stage par enregistrement à l'aide de valeurs de champ.

Dans l'itération par groupe, les enregistrements sont regroupés par un champ clé et le dataflow intégré traite chaque groupe. Tous les enregistrements d'un groupe sont traités en une itération, puis le groupe est écrit dans le stage suivant du dataflow intégré. Utilisez l'itération par groupe pour traiter des groupes d'enregistrements associés, ainsi que pour définir des options de stage à utiliser lors du traitement du groupe d'enregistrements. Par exemple, vous souhaitez peut-être regrouper des enregistrements par ID client pour pouvoir analyser les enregistrements de chaque client, pour déterminer, par exemple, le magasin dans lequel chaque client se rend le plus souvent.

Si vous utilisez l'itération, vous devez tenir compte de son impact sur les performances. À chaque lancement d'une nouvelle itération, l'initialisation du dataflow intégré requiert davantage de mémoire. Ce besoin accru en mémoire peut être considérable, en particulier si vous disposez de dataflows intégrés au sein d'autres dataflows intégrés. Par exemple, si un dataflow intégré effectue 1 000 itérations et qu'il contient un autre dataflow qui lui aussi effectue 1 000 itérations, le nombre total d'itérations est de 1 000 000. L'utilisation de l'itération par enregistrement a un impact plus important sur les performances, car chaque enregistrement déclenche une nouvelle itération.

1. Créez un dataflow intégré contenant le ou les stages que vous souhaitez utiliser pour l'itération.

Remarque : Il existe certaines limitations quant aux éléments susceptibles d'être inclus dans des dataflows intégrés dont l'option d'itération est activée :

- Le stage Stream Combiner ne peut pas être le premier stage d'un dataflow intégré dont l'option d'itération est activée.
- Le dataflow intégré ne peut pas contenir de récepteur écrivant dans un fichier figurant sur le client. Les récepteurs placés à l'intérieur d'un dataflow intégré doivent écrire dans un fichier figurant sur le serveur Spectrum™ Technology Platform ou sur un serveur de fichiers.

2. Double-cliquez sur l'icône de dataflow intégré.
3. Cochez la case **Activer l'itération**.
4. S'il existe plusieurs canaux d'entrée connectés au dataflow intégré, utilisez le champ **Port** pour sélectionner le port dont vous souhaitez utiliser les enregistrements pour l'itération.

Par exemple, imaginons que vous disposiez de deux ports d'entrée, A et B, et que vous décidez d'effectuer une itération à chaque modification d'un champ clé. Si vous décidez d'utiliser le port B pour l'itération, le dataflow intégré lance une nouvelle itération à chaque modification d'un champ clé des enregistrements du port B. Tous les enregistrements de l'autre port (le port A) sont lus dans le dataflow intégré, mis en cache et utilisés à chaque itération.

5. Sélectionnez le type d'itération à effectuer.

Répéter chaque fois qu'un champ clé change Dans ce type d'itération, le dataflow intégré traite des groupes d'enregistrements dont un ou plusieurs champs contiennent la même valeur. Lorsque le dataflow intégré termine le traitement du groupe d'enregistrements, il est réinitialisé et commence le traitement d'un nouveau groupe d'enregistrements. Utilisez ce type d'itération pour créer des dataflows intégrés qui traitent des groupes d'enregistrements, puis qui émettent chaque groupe d'enregistrements séparément.

Conseil : Si vous optez pour ce type d'itération, vous pouvez améliorer les performances en plaçant un stage Sorter devant le dataflow intégré et en triant les données en fonction du champ clé.

Répéter par enregistrement Dans ce type d'itération, le dataflow intégré traite un enregistrement à la fois. Chaque fois que le dataflow intégré termine le traitement d'un enregistrement, le résultat est envoyé à la sortie et un nouvel enregistrement est traité. Les dataflows intégrés qui effectuent des

itérations par enregistrement gèrent chaque enregistrement comme une nouvelle exécution de dataflows.

6. Si vous sélectionnez **Répéter chaque fois qu'un champ clé change**, cochez la case **Ignorer la casse lors de la comparaison de valeurs** pour ignorer les différences de casse lors de l'évaluation de valeurs de champ clé pour déterminer des groupes d'enregistrements.
7. Indiquez un ou plusieurs champs clés.
 - a) Cliquez sur **Ajouter**.
 - b) Sélectionnez le champ à utiliser comme champ clé.
 - c) Pour utiliser la valeur du champ pour définir une option de stage au sein du dataflow intégré, indiquez le nom de l'option de votre choix.
 - d) Cliquez sur **OK**.
 - e) Si nécessaire, ajoutez des champs clés supplémentaires.

Si vous sélectionnez plusieurs champs clés et l'option **Répéter chaque fois qu'un champ clé change**, pour que les enregistrements soient regroupés, ils doivent contenir la même valeur dans tous les champs clés.

Dissociation d'un flux de données intégré

Lorsque vous dissociez un flux de données intégré, ses stages sont placés dans le flux de données parent, à l'emplacement du flux de données intégré. Tous les paramètres d'itération configurés pour le flux de données intégré sont supprimés.

Pour dissocier les stages d'un flux de données intégré, faites un clic droit sur le flux de données intégré, puis sélectionnez **Dissocier ce flux de données intégré**.

Conversion d'un flux de données intégré en sous-flux

Pour réutiliser un flux de données intégré dans un autre flux de données, vous devez le convertir en sous-flux. En effet, les flux de données intégrés ne peuvent pas être utilisés dans d'autres flux de données. Une fois que vous avez converti un flux de données intégré en sous-flux, vous pouvez l'utiliser comme tout autre sous-flux.

Remarque : Si l'itération est activée pour le flux de données intégré, lors de la conversion de ce dernier en sous-flux, ses paramètres d'itération sont supprimés. Les sous-flux ne prennent pas l'itération en charge.

1. Dans Enterprise Designer, ouvrez le flux de données contenant le flux de données intégré que vous souhaitez convertir en sous-flux.

2. Faites un clic droit sur le flux de données intégré, puis sélectionnez **Convertir un stage en sous-flux**.
3. Saisissez un nom pour le nouveau sous-flux, puis cliquez sur **OK**.

Le flux de données intégré est converti en sous-flux et mis à disposition dans la palette Enterprise Designer, sous le dossier Stages définis par l'utilisateur.

Rapports

Spectrum™ Technology Platform fournit des fonctions de reporting pour les jobs. Vous pouvez utiliser les rapports standard fournis avec certains modules ou vous pouvez concevoir vos propres rapports. Lorsqu'un rapport est inclus dans un flux de données, le flux de données entier est exécuté et, après cela, les étapes de rapport dans le flux de données sont exécutées et les rapports sont enregistrés au format que vous choisissez, par exemple PDF.

Ajout d'un rapport standard à un job

Un rapport standard est un rapport préconfiguré, inclus avec un module Spectrum™ Technology Platform. Par exemple, le module Location Intelligence inclut le rapport Point In Polygon Summary, qui récapitule les résultats des calculs Point in Polygon, tels que le nombre d'inclusions spatiales, la base de données utilisée pour le job et d'autres informations.

La procédure suivante décrit comment ajouter un rapport standard à un job.

1. Dans Enterprise Designer, sur le côté gauche de la fenêtre sous Palette, cliquez sur **Rapports**. La liste des rapports disponibles apparaît.
2. Faites glisser le rapport de votre choix sur le canevas. Vous n'avez pas à connecter l'icône du rapport à quoi que ce soit.
3. Double-cliquez sur le rapport.
4. Sélectionnez les stages que vous souhaitez ajouter au rapport.
5. Cliquez sur l'onglet **Paramètres**.
6. Désélectionnez la case **Utiliser les options de rapport par défaut** et sélectionnez le format de sortie approprié si vous souhaitez indiquer un format qui ne soit pas PDF (par exemple html ou txt).

Définition d'options de rapport pour un job

Fournit des informations récapitulatives sur un job, telles que le nombre d'enregistrements traités, les paramètres utilisés pour le job, etc. Les options de rapport indiquent comment gérer les rapports générés par un job, par exemple, les options de format de sortie et d'archivage. Les valeurs par défaut des options de rapport sont configurées dans Management Console, mais vous pouvez remplacer les options par défaut d'un job dans Enterprise Designer.

La procédure suivante décrit comment indiquer les options de rapport d'un job.

1. Ouvrez le job dans Enterprise Designer et accédez à **Édition > Options du job**.
2. Cliquez sur l'onglet **Reporting**.
3. Désélectionnez la case **Utiliser les options de rapport globales**.
4. Sélectionnez le format à utiliser pour enregistrer les rapports. Les rapports peuvent être enregistrés au format HTML, PDF ou texte.
5. Sélectionnez l'emplacement d'enregistrement des rapports.

Enregistrer les rapports dans l'historique de tâches Enregistre les rapports sur le serveur dans le cadre de l'historique des jobs. Cela permet aux utilisateurs de Management Console et d'Enterprise Designer d'afficher facilement des rapports, dans la mesure où ces derniers sont disponibles dans l'historique d'exécution.

Enregistrer les rapports dans un fichier Enregistre les rapports dans un fichier à l'emplacement de votre choix. Ceci s'avère utile si vous souhaitez partager des rapports avec des personnes qui ne sont pas des utilisateurs Spectrum™ Technology Platform. Cette option s'avère également utile si vous souhaitez créer une archive de rapports à un emplacement différent. Pour afficher les rapports ainsi enregistrés, vous pouvez utiliser tout outil capable d'ouvrir le format de rapport, comme une visionneuse PDF pour les rapports PDF ou un navigateur Web pour les rapports HTML.

6. Si vous avez sélectionné **Enregistrer les rapports dans un fichier**, renseignez ces champs.

Emplacement des rapports Dossier dans lequel vous souhaitez enregistrer les rapports.

Ajouter au nom du rapport Spécifie les informations de variable à inclure dans le nom du fichier. Vous pouvez sélectionner une ou plusieurs des options suivantes :

ID de job ID unique attribué à l'exécution d'un job. La première fois que vous exécutez un job sur votre système, son ID est 1. La deuxième fois que vous exécutez un job, qu'il s'agisse du même job ou d'un autre, son ID est 2, etc.

Stage Nom du stage ayant contribué aux données du rapport, tel que spécifié dans le stage Report d'Enterprise Designer.

Date	Jour, mois et année de création du rapport.
Écraser les rapports existants	Remplace les rapports précédents portant le même nom de fichier par le nouveau rapport. Si vous ne sélectionnez pas cette option et qu'il existe un rapport portant le même nom que le nouveau rapport, le job se termine correctement, mais le nouveau rapport n'est pas enregistré. Un commentaire apparaît dans l'historique d'exécution, indiquant que le rapport n'a pas été enregistré.

7. Cliquez sur **OK**.

Lorsque vous exécutez un job, l'Historique de l'exécution possède une colonne vous indiquant si des rapports sont associés à celui-ci. Une icône vide signifie qu'il n'y a pas de rapport, une icône document indique qu'il existe un rapport, et plusieurs icônes documents vous informent qu'il y a plusieurs rapports. Vous pouvez utiliser Détails du job pour afficher, enregistrer ou imprimer le rapport.

Remarque : Pour supprimer un rapport, cliquez avec le bouton droit sur l'icône rapport sur le canevas puis sélectionnez **Suppression**.

Affichage des rapports

Pour afficher des rapports, exécuter d'abord le job, puis réaliser l'une des actions suivantes :

- Dans Enterprise Designer, la fenêtre **Détails d'exécution** s'ouvre quand vous exécutez votre job. Sélectionnez le rapport que vous voulez consulter.
- Dans Management Console, sous le nœud Exécution, cliquez sur **Historique**, puis sélectionnez le job dont vous voulez voir les rapports, et cliquez sur **Détails**.

Utilisation de rapports personnalisés

Les modules Spectrum™ Technology Platform fournissent des rapports utiles au reporting de base. Cependant, si vos besoins de reporting ne sont pas couverts par les rapports standard, vous pouvez créer vos propres rapports personnalisés et les inclure dans votre flux de données.

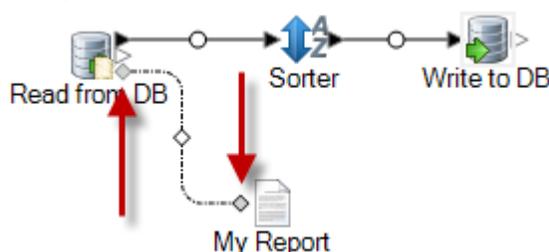
1. Créez le modèle de rapport à l'aide de l'outil de conception de rapport de votre choix. Votre outil de conception doit pouvoir exporter le rapport au format JasperReports (.jrxml).
2. Copiez votre fichier .jrxml dans le dossier `server\app\import` sur le serveur Spectrum™ Technology Platform.

En quelques secondes, le modèle de rapport sera importé dans le système et rendu disponible dans Enterprise Designer.

3. Dans Enterprise Designer, ouvrez le job auquel vous souhaitez ajouter votre rapport personnalisé.
4. Sur le côté gauche de la fenêtre, sous Palette, cliquez sur **Rapports**.
5. Faites glisser votre rapport personnalisé sur le canevas.
6. Indiquez la source de données du rapport en procédant de l'une des manières suivantes :

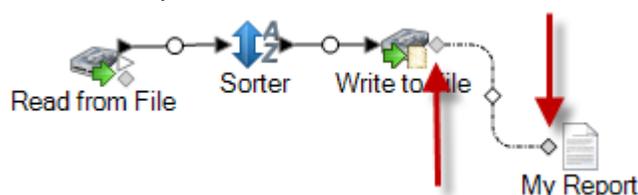
Option	Description
--------	-------------

Pour créer un rapport sur l'entrée du flux de données	Connectez le rapport au stage source sur lequel vous souhaitez créer un rapport à l'aide du port de rapport en forme de losange gris, comme indiqué ici :
--	---



le rapport sera basé sur les données d'entrée du flux de données et ne reflètera pas le traitement qui a lieu dans le flux de données.

Pour créer un rapport sur la sortie du flux de données	Connectez le rapport au stage récepteur sur lequel vous souhaitez créer un rapport à l'aide du port de rapport en forme de losange gris, comme indiqué ici :
---	--

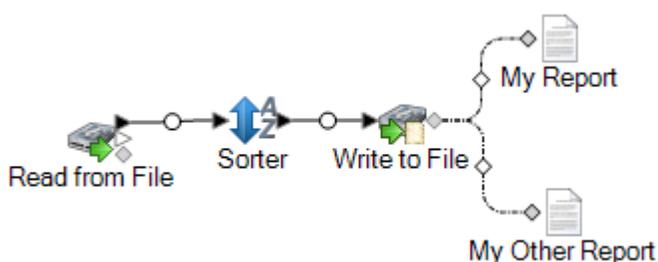


le rapport sera basé sur les données de sortie du flux de données et reflètera l'effet du flux de données sur les données.

Pour utiliser une requête imbriquée dans le modèle de rapport	Si le modèle de rapport contient une requête SQL imbriquée dans l'élément <code><queryString></code> du fichier JRXML, double-cliquez sur l'icône du rapport et cochez la case Utiliser une requête imbriquée , puis sélectionnez la connexion de base de données à utiliser pour la requête.
--	--

Remarque : Si vous devez définir une connexion de base de données, ouvrez Management Console et accédez à **Ressources**, puis **Connexions**.

Vous pouvez connecter plusieurs rapports à une source ou à un récepteur, comme indiqué ici :



7. Si le rapport contient des paramètres définis par l'utilisateur :
 - a) Double-cliquez sur l'icône du rapport sur le canevas.
 - b) Dans l'onglet **Paramètres**, indiquez les valeurs à utiliser pour les paramètres définis par l'utilisateur du rapport.
8. Facultatif : Si nécessaire, cliquez avec le bouton droit de la souris sur le canal et mettez en correspondance les champs de la source ou du récepteur avec les champs dans le rapport.

Considérations en matière de performances

Consignes de conception en vue d'une performance optimale

Une conception attentive de vos flux de données dans le but d'accroître la performance est le meilleur moyen d'atteindre une bonne performance avec Spectrum™ Technology Platform. Les consignes suivantes décrivent les techniques dont vous pouvez vous servir pour optimiser les performances de vos flux de données.

Minimiser le nombre de stages

Spectrum™ Technology Platform obtient de hautes performances à travers le processus parallèle. Chaque stage d'un flux est exécuté de manière asynchrone dans son propre thread. Cependant, il est possible d'entraîner l'« overhead » des processeurs lors de l'exécution de certains types de flux de données. Dans ce cas, le système passe autant de temps, voire plus, à gérer les threads qu'à réellement « travailler ». Nous avons vu des flux de données qui comportaient jusqu'à 130 stages individuels et qui livrent une très faible performance sur des serveurs de plus petites tailles avec un ou deux processeurs.

Donc, la première chose à considérer pour concevoir des flux de données qui livrent une bonne performance est d'utiliser autant de stages que nécessaire, mais pas plus. Quelques exemples où plus de stages que nécessaires sont utilisés :

- Utiliser plusieurs routeurs conditionnels quand un seul suffirait
- Définir plusieurs stages Transformer au lieu de combiner les transformations dans un seul stage

Heureusement, il est généralement possible de reconcevoir ces flux de données afin d'ôter les stages redondants ou inutiles et ainsi améliorer la performance.

Pour les flux complexes, songez à utiliser des flux intégrés ou des sous-flux afin de réduire l'encombrement sur le canevas et de faciliter l'affichage et le parcours du flux. L'utilisation de flux intégrés ne présente pas d'avantages en termes de performances lors de l'exécution, mais cela facilite l'utilisation de flux dans Enterprise Designer. L'utilisation de sous-flux pour simplifier les flux complexes peut améliorer les performances d'Enterprise Designer lorsque vous modifiez des flux.

Réduire la longueur d'enregistrement

Puisque les données sont transmises entre les stages s'exécutant de manière simultanée, il faut aussi prendre en considération la longueur des enregistrements entrants. Généralement, une entrée possédant une longueur d'enregistrement longue sera plus longue à traiter qu'une entrée possédant une longueur d'enregistrement plus courte, tout simplement en raison du nombre plus élevé de données à lire, écrire et trier. Les flux de données intégrant plusieurs opérations de tri bénéficieront particulièrement d'une réduction de la longueur d'enregistrement. Dans le cas où la longueur d'enregistrement est très longue, il est parfois plus rapide de supprimer les champs inutiles de l'entrée avant d'exécuter le job Spectrum™ Technology Platform, puis de réintroduire les champs supprimés dans le fichier de sortie qui en résulte.

Utiliser les tris de manière appropriée

Une autre méthode à prendre en considération est la minimisation des opérations de tri. Le tri prend souvent plus de temps que d'autres opérations, ce qui peut devenir problématique au fur et à mesure qu'augmentent le nombre et la taille des enregistrements d'entrée. Toutefois, de nombreux stages Spectrum™ Technology Platform nécessitent, ou privilégient, des données d'entrée triées. Les modules Universal Addressing et Enterprise Geocoding, par exemple, donnent un performance optimale quand l'entrée est triée par pays et par code postal. Les stages comme Intraflow Match et Interflow Match nécessitent que l'entrée soit triée par le champ « Grouper par ». Dans certains cas, vous pouvez utiliser une application de tri externe afin de pré-trier les données d'entrée, ce qui peut s'avérer plus rapide que de trier à l'intérieur du flux de données Spectrum™ Technology Platform.

Options de performances d'exécution des stages

Les options de performances d'exécution contrôlent la manière dont les stages individuels d'un flux de données sont exécutés et fournissent des paramètres que vous pouvez utiliser pour améliorer les performances de votre flux de données. Les paramètres à votre disposition dépendent de la manière dont votre environnement Spectrum™ Technology Platform a été configuré.

- L'option **Local** est le paramètre par défaut dans lequel les stages s'exécutent sur le serveur Spectrum™ Technology Platform local et utilisent une instance d'exécution. Le paramètre des instances d'exécution peut être augmenté, utilisant ainsi le traitement parallèle et l'amélioration des performances.

- L'option **Distribué** est généralement utilisée dans un environnement en cluster, qui implique l'installation d'un équilibreur de charge et de plusieurs serveurs Spectrum™ Technology Platform.
- Vous pouvez utiliser l'option **Distant** si votre environnement se compose de plusieurs serveurs Spectrum™ Technology Platform mais n'est pas configuré pour le traitement distribué. Cette option permet à un autre serveur de procéder au traitement d'un stage.

Taille de pool de base de données et instances d'exécution

Dans la plupart des environnements Spectrum™ Technology Platform, il existe plusieurs flux en cours d'exécution en même temps, qu'ils s'agisse de traitements par lots ou de services répondant à un service Web ou à des requêtes d'API. Pour optimiser le traitement simultané, vous pouvez utiliser le paramètre de taille de pool de base de données, qui limite le nombre de requêtes simultanées gérées par une base de données Spectrum, et les instances d'exécution, qui contrôlent le nombre d'instances d'un stage de flux exécutées de manière simultanée. Ces deux paramètres doivent être réglés ensemble pour obtenir des performances optimales.

Taille de pool de base de données

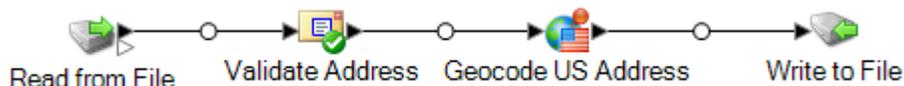
Les ressources de base de données contiennent des données de référence utilisées par certains stages, comme les données postales utilisées pour valider des adresses ou les données de géocodage utilisées pour géocoder des adresses. Ces ressources de base de données peuvent être configurées pour accepter plusieurs demandes simultanées provenant des stages de dataflows ou des services qui les utilisent, améliorant ainsi les performances des dataflows ou des demandes de service. La taille de pool de base de données détermine le nombre maximal de requêtes simultanées qu'une base de données Spectrum peut traiter. Par défaut, les bases de données Spectrum disposent d'une taille de pool de 4, ce qui signifie que la base de données peut traiter quatre demandes simultanément.

La taille de pool optimale varie en fonction du module. En général, vous verrez les meilleurs résultats en définissant la taille de pool entre la moitié à deux fois le nombre d'unités centrales sur le serveur, avec une taille de pool optimale de la plupart des modules identique au nombre d'unités centrales. Par exemple, si votre serveur dispose de quatre unités centrales que vous souhaitez expérimenter avec une taille de pool comprise entre 2 (la moitié du nombre d'unités centrales) et 8 (deux fois le nombre d'unités centrales) avec la taille optimale étant probablement 4 (le nombre d'unités centrales).

Lors de la modification de la taille de pool, vous devez également considérer le nombre d'instances d'exécution indiquées dans le dataflow pour les stages accédant à la base de données. Considérez par exemple un flux de données disposant d'un stage Geocode US Address configuré pour utiliser une instance d'exécution. Si vous définissez la taille de pool de la base de données de géocodage des États-Unis, vous ne verrez pas d'amélioration des performances, car il n'y aurait qu'une instance d'exécution et donc il n'y aurait qu'une demande à la fois dans la base de données. Cependant, si vous souhaitez augmenter le nombre d'instances d'exécution de Geocode US Address à quatre, vous pouvez voir une amélioration de performances dans la mesure où il y aurait quatre instances de Geocode US Address accédant à la ressource de base de données simultanément, utilisant ainsi le pool entier.

Instances d'exécution (Runtime)

Chaque stage dans un flux de données fonctionne de manière asynchrone dans son propre thread et est indépendant de tout autre stage. Ceci permet le traitement parallèle des stages dans un flux de données, ce qui vous permet d'utiliser plus d'une instance d'exécution pour un stage. Ceci s'avère utile dans les flux de données dans lesquels certains stages traitent les données plus vite que d'autres. Cela peut mener à une distribution non équilibrée du travail parmi les threads. Par exemple, prenons un flux de données composé des stages suivants :



Selon la configuration de ces stages, il est possible que le stage Validate Address traite les enregistrements plus vite que le stage Geocode US Address. Si tel est le cas, à un certain moment de l'exécution du flux de données, tous les enregistrements auront été traités par Validate Address, mais Geocode US Address aura encore des enregistrements à traiter. Afin d'améliorer la performance de ce flux de données, il est nécessaire d'améliorer la performance du stage qui est le plus lent - dans notre cas, Geocode US Address. Une manière d'accomplir cela est de spécifier plusieurs instances d'exécution du stage. Régler le nombre d'instances d'exécution sur deux, par exemple, signifie qu'il y aura deux instances de ce stage, chacune s'exécutant dans son propre thread, disponible pour le traitement de vos enregistrements.

En règle générale, le nombre d'instances d'exécution devrait être au minimum égal au nombre d'instances du composant distant. Pour des informations sur les composants distants, reportez-vous à *Guide d'administration*. Bien que la spécification de plusieurs instances d'exécution puisse vous aider à améliorer la performance, définir une valeur trop élevée peut être néfaste pour vos ressources système, ce qui aura pour résultat une diminution de la performance.

Remarque : L'utilisation de plusieurs instances d'exécution n'améliore la performance que lors de l'exécution de jobs ou de requêtes de services avec plus d'un enregistrement.

Procédure de réglage fin

Pour trouver les bons réglages pour la taille de pool de base de données et le nombre d'instances d'exécution, il convient d'expérimenter différents réglages pour trouver ceux qui maximisent les ressources du serveur sans les surcharger ni réduire les performances.

Remarque : Avant d'affiner la taille de pool de base de données, nous vous conseillons d'optimiser la taille de pool de flux de données. Pour plus d'informations sur l'optimisation de la taille de pool de flux de données, reportez-vous à la section [Taille de pool de flux de données](#).

1. Commencez par rechercher des exemples de données à utiliser pour tester différents réglages. L'exemple de jeu de données doit être suffisamment important pour que le temps d'exécution soit mesurable et qu'il puisse être validé à des fins de cohérence. Les exemples de données doivent aussi être représentatifs des données réelles que vous souhaitez traiter. Par exemple, si vous effectuez un test des performances de géocodage, assurez-vous que vos données de

test disposent d'un nombre équivalent d'enregistrements pour tous les pays que vous avez l'intention de géocoder.

2. Si vous testez un service ou un flux de données qui requiert l'utilisation d'une ressource de base de données, comme des bases de données postales ou des bases de données de géocodage, assurez-vous que vous disposez de la dernière version de la base de données installée.
3. Une fois les exemples de données prêts et les toutes dernières ressources de base de données installées, créez un flux de données simple qui lit les données d'un fichier, les traite à l'aide du stage que vous souhaitez optimiser et les écrit dans un fichier. Par exemple, si vous souhaitez tester les paramètres de performances de Validate Address, créez un flux de données composé de Read From File, Validate Address et Write To File.
4. Définissez la taille de pool de ressource de base de données sur 1 :
 - a. Ouvrez Management Console.
 - b. Accédez à **Ressources > Bases de données Spectrum**.
 - c. Sélectionnez la ressource de base de données à optimiser et cliquez sur le bouton Modifier .
 - d. Dans le champ **Taille de pool**, saisissez 1.
 - e. Cliquez sur **OK**.
5. Définissez les instances d'exécution du stage sur 1 :
 - a. Ouvrez le flux de données dans Enterprise Designer.
 - b. Double-cliquez sur le stage que vous souhaitez définir de sorte qu'il utilise plusieurs instances d'exécution.
 - c. Cliquez sur **Exécution**.

Remarque : Tous les stages ne sont pas capables d'utiliser plusieurs instances d'exécution. Si aucun bouton **Exécution** n'apparaît au bas de la fenêtre du stage, cela signifie que le stage n'est pas capable d'utiliser plusieurs instances d'exécution.

- d. Sélectionnez **Local** et indiquez 1.
 - e. Cliquez sur **OK** pour fermer la fenêtre **Performances d'exécution**, puis cliquez sur **OK** pour fermer le stage.
6. Calculez les performances de base en exécutant le flux de données plusieurs fois et en enregistrant les valeurs moyennes pour chacun des éléments suivants :
 - Temps écoulé
 - Utilisation de l'UC
 - Utilisation de la mémoire

Conseil : Vous pouvez utiliser la console JMX pour analyser les performances. Pour plus d'informations, reportez-vous à la section [Analyse des performances à l'aide de la console JMX](#).
7. Exécutez plusieurs instances simultanées du job, s'il s'agit d'un scénario qui doit être pris en charge. Enregistrez le temps écoulé, l'utilisation de l'UC et l'utilisation de la mémoire pour chaque scénario.

Conseil : Vous pouvez utiliser un analyseur de fichiers pour exécuter plusieurs instances d'un job à la fois. Pour plus d'informations, reportez-vous à la section [Déclenchement d'un flux avec un fichier de contrôle](#) à la page 167.

8. Augmentez la taille de pool des ressources de base de données et le paramètre d'instances d'exécution des stages.
9. Redémarrez le serveur.
10. Exécutez de nouveau le flux de données, en enregistrant le temps écoulé, l'utilisation de l'UC et l'utilisation de la mémoire.
11. Continuez à augmenter la taille de pool des ressources de base de données et les instances d'exécution des stages jusqu'à ce que vous commenciez à voir les performances diminuer.
12. Si vous testez les performances de géocodage, répétez cette procédure en utilisant un seul pays et plusieurs pays.

Traitement distribué

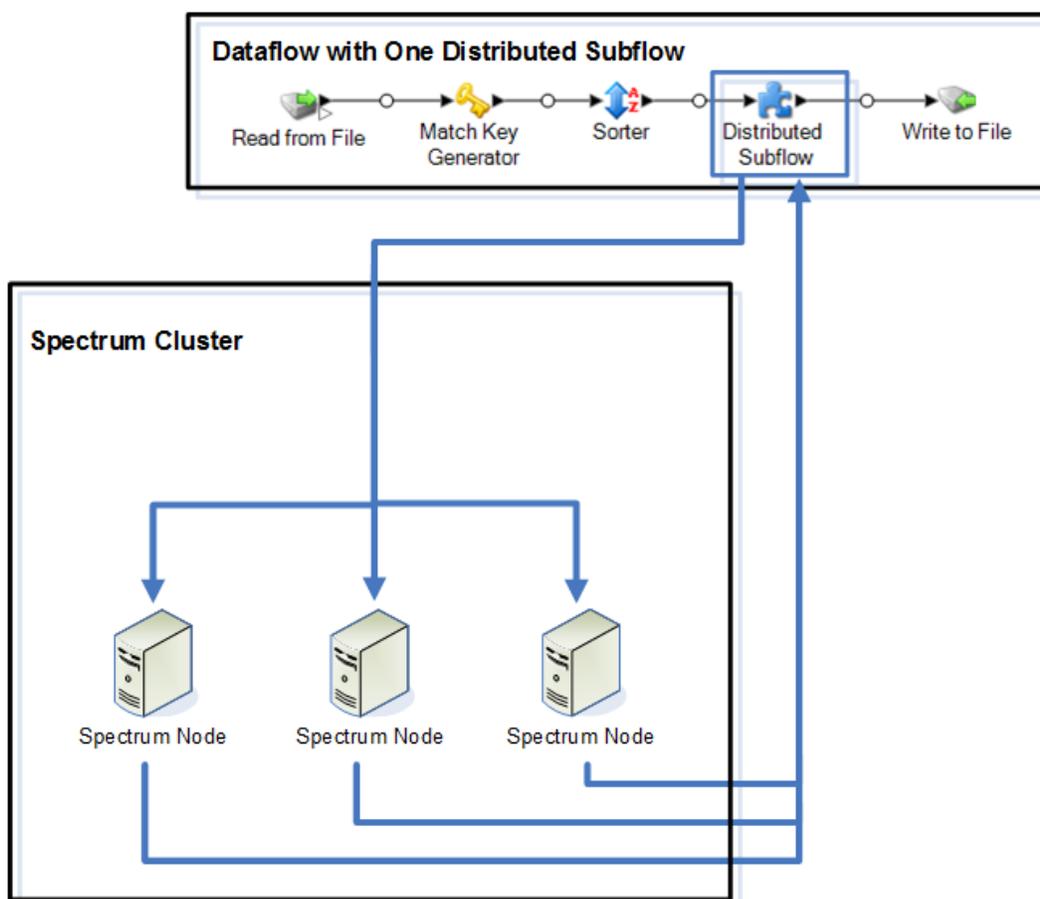
Si vous avez une tâche très complexe, ou si vous êtes en train de traiter des données très volumineuses, contenant par exemple des millions d'enregistrements, vous pouvez peut-être améliorer les performances de flux de données en distribuant le traitement du flux de données auprès de plusieurs instances du serveur Spectrum™ Technology Platform sur un ou plusieurs serveurs physiques.

La solution la plus évolutive pour le traitement distribué est d'installer Spectrum™ Technology Platform dans un cluster. Pour obtenir des instructions d'installation et de configuration d'un cluster, reportez-vous au *Guide d'installation*.

Remarque : Même s'il est également possible d'utiliser le traitement distribué sur un seul serveur Spectrum™ Technology Platform, les informations suivantes décrivent l'utilisation du traitement distribué dans un cluster. Si vous utilisez un seul serveur, le traitement distribué des sous-flux est divisé en microlots et traité par ce serveur et non par le cluster.

Une fois votre environnement en cluster configuré, vous pouvez créer un traitement distribué dans un flux de données en créant des sous-flux pour les parties du flux de données que vous souhaitez distribuer auprès de plusieurs serveurs. Spectrum™ Technology Platform gère automatiquement la distribution du traitement, une fois que vous avez défini seulement quelques options de configuration pour le sous-flux.

Le schéma suivant illustre le traitement distribué :



À mesure que les enregistrements sont lus dans le sous-flux, les données sont regroupées en lots. Ces lots sont ensuite écrits dans le cluster et automatiquement distribués au nœud du cluster qui traite le lot. Ce traitement est appelé un microflux. Un sous-flux peut être configuré pour permettre le traitement simultané de plusieurs microflux, améliorant potentiellement les performances du flux de données. Lorsque l'instance distribuée a terminé le traitement d'un microflux, elle renvoie la sortie au flux de données parent.

Plus vous disposez de nœuds Spectrum™ Technology Platform, plus vous pouvez traiter de microflux simultanément, ce qui vous permet de mettre votre environnement à l'échelle selon les besoins, afin d'obtenir les performances requises.

Une fois configuré, un environnement en cluster est facile à maintenir, dans la mesure où tous les nœuds du cluster synchronisent automatiquement leur configuration, ce qui signifie que les paramètres que vous appliquez dans Management Console et les flux de données que vous concevez dans Enterprise Designer sont automatiquement disponibles pour toutes les instances.

Conception d'un flux de données pour le traitement distribué

Le traitement distribué prend des parties de votre flux de données et distribue le traitement de ces parties auprès d'un cluster de serveurs Spectrum™ Technology Platform. Par exemple, votre flux de données procède peut-être au géocodage et vous souhaitez distribuer le traitement de géocodage

entre plusieurs nœuds Spectrum™ Technology Platform d'un cluster afin d'améliorer les performances.

1. Décidez des stages de votre flux de données que vous souhaitez distribuer, puis créez un sous-flux contenant les stages que vous souhaitez distribuer.

N'utilisez pas les stages suivants dans un sous-flux à utiliser pour le traitement distribué :

- Sorter
- Unique ID Generator
- Record Joiner
- Interflow Match

Les ensembles de stages suivants doivent être utilisés ensemble dans un sous-flux pour le traitement distribué :

- les stages de correspondance (Intraflow Match et Transactional Match) et les stages de consolidation (Filter, Best of Breed et Duplicate Synchronization).
- Aggregator et Splitter

N'incluez pas d'autres sous-flux dans le sous-flux (sous-flux imbriqués).

Prenez en compte les éléments suivants, si vous allez exécuter des opérations de correspondance dans un sous-flux utilisé pour le traitement distribué :

- Le tri doit être effectué dans le travail et non dans le sous-flux. Vous devez désactiver le tri dans le stage et définir le tri au niveau du travail.
- L'analyse de correspondance n'est pas prise en charge dans un sous-flux distribué
- Les numéros de collection seront réutilisés dans un groupe de lots de microflux

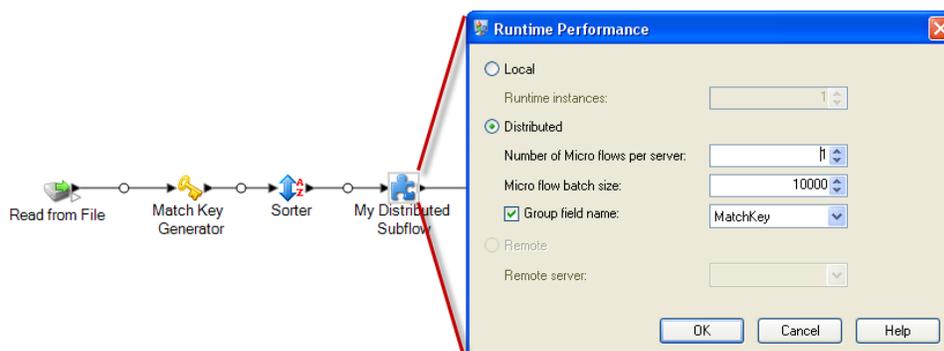
Utiliser un stage Write Exception dans un sous-flux peut produire des résultats inattendus. Ajoutez plutôt ce stage à votre flux de données au niveau du job.

2. Une fois que vous avez créé votre sous-flux pour la partie du flux de données à distribuer, ajoutez le sous-flux au flux de données parent et reliez-le à un stage ascendant et descendant. Les sous-flux utilisés pour le traitement distribué ne disposent peut-être que d'un port d'entrée.
3. Faites un clic droit sur le sous-flux, puis sélectionnez **Options**.
4. Sélectionnez **Distributed**.
5. Saisissez le nombre de micro-flux à envoyer à chaque serveur.
6. Saisissez le nombre d'enregistrement que doit contenir chaque lot de micro-flux.
7. Facultatif : (Facultatif) Vérifiez **Nom de champ de groupe** et sélectionnez le nom du champ suivant lequel les lots de micro-flux doivent être regroupés.

Si vous fournissez un champ de groupe, la taille de vos lots peut être supérieure au nombre que vous avez indiqué dans le champ **Micro flow batch size**, car un groupe n'est pas divisé en plusieurs lots. Par exemple, si vous indiquez une taille de lot de 100, mais que vous disposez de 108 enregistrements dans le même groupe, ce lot contiendra 108 enregistrements. De même, si vous indiquez une taille de lot de 100 et qu'un nouveau groupe de 28 enregistrements portant

le même ID commence à l'enregistrement 80, vous disposerez de 108 enregistrements dans ce lot.

L'exemple suivant présente un flux de données, dans lequel un sous-flux nommé My Distributed Subflow a été configuré pour s'exécuter en mode distribué :



Exécution d'un stage sur un serveur distant

Si votre administrateur système a activé les serveurs distants dans Management Console, des stages dans votre flux de données peuvent exécuter leur traitement sur un serveur distant. L'utilisation de serveurs distants permet d'améliorer les performances en répartissant le traitement du flux de données sur plusieurs serveurs Spectrum™ Technology Platform.

Votre administrateur système peut avoir déjà désigné certains stages à exécuter sur un serveur distant. Si un stage est déjà routé vers un serveur distant, vous verrez apparaître une étoile rouge dans le coin supérieur gauche de l'icône du stage sur le canevas dans Enterprise Designer.

Cette procédure explique comment configurer le traitement distant pour un stage dans un flux de données.

1. Ouvrez le flux de données dans Enterprise Designer.
2. Double-cliquez sur le stage que vous voulez router vers un serveur distant.
3. Cliquez sur **Exécution**.

La fenêtre **Performances d'exécution** apparaît.

4. Cliquez sur **Distant** et sélectionnez le serveur distant vers lequel vous voulez router le processus de ce stage.
5. Cliquez sur **OK**.

Dépannage des erreurs de serveur distant

Cette section expose les erreurs éventuelles que vous pouvez rencontrer lors de l'utilisation des serveurs distants.

Module non licencié

Le serveur distant doit disposer de la licence à la fois pour le module et pour le mode de l'exécution vous tentez d'exécuter, soit par lot, soit en temps réel. La licence du serveur distant peut être différente de la licence du serveur local. Connectez-vous au serveur distant à l'aide de Management Console et vérifiez que la licence correcte est installée. Vous devez vous connecter avec un compte disposant de privilèges administratifs afin d'afficher les informations de licence.

Serveur distant non disponible

Si le serveur distant n'est pas exécuté ou n'est pas accessible pour quelque raison que ce soit, les services distants deviendront disponibles dans Enterprise Designer et Management Console. Vous verrez une icône de danger de couleur jaune dans la barre d'état située dans la partie inférieure de l'écran :



Cliquez sur cette icône pour voir apparaître un message d'erreur détaillant quels serveurs distants ne sont pas disponibles.

De plus, dans Enterprise Designer tout stage utilisant un stage distant sera remplacé par une icône vous indiquant que le stage n'est plus disponible :



Le routage a été modifié

Si vous supprimez ou annulez le déploiement d'un service installé à la fois localement et à distance et qui a été routé via un serveur distant, puis si vous cliquez sur ce service dans Management Console, vous verrez un indicateur de changement de routage (un point d'exclamation clignotant) à côté du bouton de routage, dans l'onglet Options de ce service. Cet indicateur signifie que le routage a été modifié pour ce service.



Optimisation des stages

Optimisation du rapprochement

Le rapprochement est généralement l'une des opérations les plus longues de toute implémentation de qualité des données, ce qui rend indispensable de s'assurer que le rapprochement s'opère de la manière la plus performante possible. Il existe toujours un équilibre entre les résultats de correspondance et la performance. Si chaque enregistrement contenu dans un fichier est comparé à tous les autres enregistrements, vous pouvez être tout à fait sûr que toutes les correspondances seront identifiées. Cependant, cette approche n'est plus viable quand le volume des données grandit.

Par exemple, prenez un fichier d'entrée comportant 1 million d'enregistrements : le rapprochement de chaque enregistrement avec tous les autres nécessiterait environ 1 trillion de comparaisons afin d'évaluer chaque règle de rapprochement.

Étant donné que la plupart des enregistrements contenus dans un fichier ne correspondent pas, l'approche générale afin de résoudre ce problème consiste à définir une match key et à ne comparer que les enregistrements qui possèdent la même match key. Bien définir la match key est la variable la plus cruciale qui affecte la performance d'un moteur de rapprochement. Pour bien définir votre match key, vous devez comprendre comment le moteur de rapprochement traite les enregistrements ainsi que les options qui sont à votre disposition.

La méthode de rapprochement par défaut réalise une comparaison exhaustive de l'enregistrement dans une file de correspondance afin d'identifier le nombre maximal d'enregistrements. En raison de cela, cette méthode est souvent la plus longue des méthodes de rapprochement. Avec la méthode de rapprochement par défaut, le premier enregistrement de la file devient l'enregistrement suspect. L'enregistrement suivant y est comparé, et s'il correspond, celui-ci est écrit comme doublon. Si ce n'est pas le cas, il est ajouté comme suspect, puis le prochain enregistrement est comparé aux deux suspects actifs. Supposez la file de correspondance suivante :

ID unique	Match key
1	123A
2	123A
3	123A
4	123A
5	123A
6	123A
7	123A
8	123A
9	123A
10	123A

D'abord, l'enregistrement 2 serait comparé à l'enregistrement 1. À supposer qu'il ne corresponde pas, l'enregistrement 2 serait ajouté comme suspect. Puis, l'enregistrement 3 serait comparé aux enregistrements 1 et 2, et ainsi de suite. Si aucun enregistrement ne correspond, le nombre total

de comparaisons serait 45. Si certains enregistrement correspondent, le nombre de comparaison sera moindre. Pour une file de correspondance d'une taille donnée N, le nombre maximum de comparaisons sera $N \times (N-1) \div 2$. Quand la taille de la file est petite, cela ne se remarque pas, mais dès que la file s'agrandit, l'impact devient significatif. Par exemple, une taille de file de 100 engendrerait jusqu'à 4 450 comparaisons, et une taille de file de 500 engendrerait jusqu'à 124 750 comparaisons.

Définition d'une match key appropriée

Afin de définir une match key qui soit appropriée, prenez en considérations les éléments suivants :

- La chose la plus importante à ne pas oublier est que la plupart des enregistrements ne correspondent pas. Donc, vous ne voulez comparer que les enregistrements susceptibles de correspondre.
- Seuls les enregistrements possédant la même match key seront comparés.
- La performance constitue la considération clé :
 - La match key détermine la taille de la file de correspondance.
 - Pour un nombre d'enregistrements donné, quand la taille de la file de correspondance double, le temps d'exécution double.
 - Une match key « serrée » livre une meilleure performance. Une match key « serrée » est une clé qui est spécifique, qui contient plus de caractères issus de peut-être plus de champs.
 - Une match key « lâche » peut donner plus de correspondances. Une match key « lâche » est une clé qui sera moins spécifique, qui contiendra moins de caractères issus de peut-être moins de champs.

Équilibrage entre performance et résultats de correspondance

Pour trouver un bon équilibre entre performance et résultats, prenez en compte la règle de rapprochement et la densité des données.

- Prenez en compte les règles de rapprochement :
 - Les champs nécessitant une correspondance parfaite pourraient être compris dans la match key.
 - Générez un clé appropriée pour la règle de rapprochement. Par exemple, pour une règle de rapprochement phonétique, une match key phonétique est probablement appropriée.
 - Une match key est souvent composée de parties de tous les champs rapprochés.
 - Ayez connaissance des effets résultant du manque de données.
- Examinez la densité des données :
 - Par exemple, dans la correspondance d'adresses, une match key aurait de plus fortes chances d'être plus serrée si tous les enregistrements concernent une seule ville au lieu d'un ensemble de données national.
 - Songez à la file de correspondance la plus grande, et non à une file de taille moyenne. Passez en revue le relevé de Match Summary (synthèse des rapprochements) afin de trouver la plus longue file de correspondance

- Si vous employez une correspondance transactionnelle, les mêmes considérations s'appliquent à l'instruction SELECT dans Candidate Finder.

Match Key Express

Dans un fichier typique, la plupart des enregistrements doublons correspondent soit parfaitement, soit quasi-parfaitement. La définition d'une match key de type express permet au moteur de rapprochement de réaliser une comparaison initiale des match keys expresses pour déterminer que deux enregistrements sont des doublons. Ceci peut améliorer la performance de manière significative en évitant d'avoir à évaluer la totalité des règles de rapprochement de niveau de champ.

Méthodes Intraflow Match

La méthode de correspondance Intraflow Match par défaut compare l'ensemble des enregistrements portant la même match key. Pour une taille de file d'attente de correspondance N , la méthode par défaut effectue des comparaisons $N-1$ à $N \times (N-1)$. Si tous les enregistrements correspondent, le nombre de comparaisons est $N-1$. Si aucun enregistrement ne correspond, le nombre de comparaisons est $N \times (N-1)$. En règle générale, le nombre de comparaisons figure quelque part dans la partie supérieure de cette plage.

Si la performance est votre priorité, songez à utiliser la méthode de correspondance de fenêtre coulissante au lieu de la méthode par défaut. La méthode de correspondance de fenêtre coulissante compare chaque enregistrement aux W enregistrements suivants (où W est la taille de la fenêtre). Pour une taille de fichier N donnée, la méthode de fenêtre coulissante ne réalise pas plus de $N \times W$ comparaisons. Ceci peut résulter en une meilleure performance, mais certaines correspondances peuvent être manquées.

Optimisation de Candidate Finder

Candidate Finder sélectionne les enregistrements candidats dans une base de donnée en vue de leur comparaison par Transactional Match (correspondance transactionnelle). Puisque la correspondance transactionnelle compare l'enregistrement suspect avec la totalité des enregistrements candidats renvoyés par Candidate Finder, la performance de Transactional Match est proportionnelle au nombre de comparaisons.

Cependant, certaines actions peuvent être réalisées afin d'améliorer la performance de Candidate Finder. Pour maximiser la performance de Candidate Finder, un administrateur de bases de données, ou un développeur disposant d'une connaissance profonde des index et des schémas de bases de données, doit ajuster l'instruction SQL SELECT dans Candidate Finder. L'un des problèmes les plus courants concernant la performance est qu'une requête contenant une option JOIN nécessite de balayer la table entière. Dans ce cas, songez à ajouter un index ou à utiliser UNION à la place de JOIN. En règle générale, les requêtes SQL devraient être examinées et optimisées par des personnes qualifiées.

Optimisation des Transforms

Le stage Transformer fournit un ensemble d'opérations prédéfinies qui peuvent être réalisées sur les données d'entrée. Généralement, ces transforms prédéfinies s'exécutent plus vite que les

transforms personnalisées, étant déjà compilées. Toutefois, quand on définit un grand nombre de transforms, l'exécution d'une transform personnalisée sera souvent plus rapide. Par exemple, pour raccourcir un certain nombre de champs, la transform personnalisée suivante s'exécutera en général plus vite que neuf transforms raccourcir séparés.

```
data['AddressLine1'] = (data['AddressLine1'] != null) ?
data['AddressLine1'].trim() : null;
data['AddressLine2'] = (data['AddressLine2'] != null) ?
data['AddressLine2'].trim() : null;
data['AddressLine3'] = (data['AddressLine3'] != null) ?
data['AddressLine3'].trim() : null;
data['AddressLine4'] = (data['AddressLine4'] != null) ?
data['AddressLine4'].trim() : null;
data['City'] = (data['City'] != null) ? data['City'].trim() : null;
data['StateProvince'] = (data['StateProvince'] != null) ?
data['StateProvince'].trim() : null;
data['PostalCode'] = (data['PostalCode'] != null) ?
data['PostalCode'].trim() : null;
data['LastName'] = (data['LastName'] != null) ? data['LastName'].trim()
: null;
data['FirstName'] = (data['FirstName'] != null) ?
data['FirstName'].trim() : null;
```

Optimisation de l'écriture vers la base de données

Par défaut, le stage Write to DB valide à chaque fois qu'une rangée est insérée dans la table. Néanmoins, pour améliorer les performances, activez l'option **Validation de lot**. Lorsque cette option est activée, une validation est effectuée après le nombre d'enregistrements indiqué. En fonction de la base de données, ceci peut améliorer de manière considérable la vitesse d'écriture.

Lorsque vous sélectionnez une taille de lot, prenez en considération les éléments suivants :

- **Taux d'arrivée des données au stage Write To DB** : si les données arrivent à un taux plus lent que ce que la base de données peut traiter, la modification de la taille de lot ne permettra pas d'améliorer les performances générales du dataflow. Par exemple, les dataflow avec la validation d'adresse ou le géocodage ne peuvent pas bénéficier d'un accroissement de taille.
- **Trafic réseau** : pour les réseaux lents, l'augmentation de la taille des lots à une taille de lot moyenne (1 000 à 10 000) se traduit par de meilleures performances.
- **Vitesse de chargement et/ou de traitement de la base de données** : pour les bases de données dotées d'une importante puissance de traitement, l'augmentation de la taille de lot permettra d'améliorer les performances.
- **Instances d'exécution multiples** : si vous utilisez plusieurs instances d'exécution du stage Write to DB, une taille de lot volumineuse va consommer beaucoup de mémoire ; il est donc préférable d'utiliser une taille de lot petite ou moyenne (100 à 10 000).
- **Restauration de base de données** : chaque fois qu'une instruction échoue, l'ensemble du lot est restauré. Plus la taille du lot est volumineuse, plus l'opération de restauration est longue.

Optimisation de la validation d'adresse

Validate Address livre une performance optimale quand les enregistrements d'entrée sont triés par code postal. Ceci est dû à la manière dont les données de référence sont chargées en mémoire. Une entrée qui est triée livrera une performance parfois plusieurs fois plus rapide qu'une entrée qui n'a pas été triée. Puisque certains enregistrements ne contiendront aucune donnée dans le champ code postal, l'ordre de tri suivant est recommandé :

1. Country (uniquement si les enregistrement traités appartiennent à plusieurs pays)
2. PostalCode
3. StateProvince
4. City

Optimisation du géocodage

Les stages de géocodage délivrent des performances optimales quand les enregistrements d'entrée sont triés par code postal. Ceci est dû à la manière dont les données de référence sont chargées en mémoire. Une entrée qui est triée livrera une performance parfois plusieurs fois plus rapide qu'une entrée qui n'a pas été triée. Puisque certains enregistrements ne contiendront aucune donnée dans le champ code postal, l'ordre de tri suivant est recommandé :

1. PostalCode
2. StateProvince
3. City

Vous pouvez également optimiser les stages de géocodage en testant différents modes de correspondance. Le mode de correspondance contrôle la manière dont le stage de géocodage détermine si un résultat de géocodage est une correspondance proche. Vous pouvez définir le mode de correspondance sur **Approximatif** et vérifier si les résultats répondent à vos besoins. Le mode **Approximatif** donne généralement de meilleurs résultats que les autres modes de correspondance.

Optimisation de Geocode US Address

Le stage Geocode US Address comporte plusieurs options qui affectent les performances. Ces options se trouvent dans le fichier suivant :

```
SpectrumLocation\server\modules\geostan\java.properties
```

egm.us.multimatch.max.records	Spécifie le nombre maximal de correspondances à renvoyer. Un nombre faible entraîne une meilleure performance, mais aux dépens des correspondances.
egm.us.multimatch.max.processing	Spécifie le nombre de recherches à réaliser. Un nombre faible entraîne une meilleure performance, mais aux dépens des correspondances.
FileMemoryLimit	Contrôle la quantité de données de référence initialement chargées en mémoire.

Versions de flux de données

La fonctionnalité Versions dans Enterprise Designer vous permet de conserver un historique de révision de vos flux de données. Vous pouvez afficher les versions précédentes d'un flux de données, exposer des versions plus anciennes pour leur exécution et conserver un historique de vos modifications au cas où vous auriez besoin de revenir à une version précédente d'un flux de données.

Enregistrement d'une version de flux de données

Il existe deux façons d'enregistrer une version de votre flux de données dans Enterprise Designer :

- Exposez votre flux de données. Chaque fois que vous exposez un flux de données, soit en sélectionnant **Fichier > Exposer/Ne plus exposer et Enregistrer**, soit en cliquant sur l'ampoule dans la barre d'outils, Enterprise Designer enregistre automatiquement une version du flux de données.
- Enregistrez manuellement une version dans le volet **Versions** dans Enterprise Designer.

Remarque : Une version de flux de données n'est pas créée lorsque vous enregistrez simplement un flux de données.

La procédure suivante explique comment enregistrer une version manuellement dans le volet **Versions** d'Enterprise Designer.

1. Dans Enterprise Designer, ouvrez le flux de données.
2. Si le volet **Versions** n'est pas visible, sélectionnez **Vue > Versions**
3. Assurez-vous que la dernière version enregistrée est sélectionnée dans la liste **Versions**. Il s'agit de la version en haut de la liste.
4. Cliquez sur l'icône plus verte dans le volet **Versions**.

Une nouvelle version du flux de données est enregistrée et ajoutée au volet **Versions**.

Affichage d'une version de flux de données

Vous pouvez afficher une version antérieure d'un flux de données. Cela vous permet de voir comment un flux de données a été désigné auparavant, avant que des modifications récentes n'aient été apportées. Vous ne pouvez qu'afficher les versions antérieures et non les modifier. Pour modifier une version antérieure, celle-ci doit d'abord être promue sur la dernière version enregistrée.

1. Dans Enterprise Designer, ouvrez le flux de données.
2. Si le volet **Versions** n'est pas visible, sélectionnez **Vue > Versions**

- Sélectionnez la version à afficher.

La version sélectionnée est affichée sur le canevas du flux de données.

Modification d'une version de flux de données

Vous pouvez modifier une version précédente d'un flux de données en la promouvant à la dernière version enregistrée. La promotion d'une version de flux de données la fait passer à la dernière version enregistrée, la rendant disponible pour la modification.

Remarque : Avant d'effectuer cette procédure, la dernière version enregistrée existante sera écrasée par la version que vous promouvez et modifiez. Pour préserver une copie de la dernière version enregistrée existante, enregistrez-la sous une version avant de promouvoir la version antérieure.

- Dans Enterprise Designer, ouvrez le flux de données.
- Si le volet **Versions** n'est pas visible, sélectionnez **Vue > Versions**
- Sélectionnez la version à modifier.
- Cliquez sur l'icône de promotion.



La version sélectionnée est promue à la dernière version enregistrée. Vous pouvez désormais modifier le flux de données.

Modification des propriétés de version

Lorsque vous enregistrez une version de flux de données, il obtient un numéro de version par défaut. Vous pouvez modifier le numéro de version et ajouter des commentaires pour documenter les modifications ou l'objectif de la version.

- Dans Enterprise Designer, ouvrez le flux de données.
- Si le volet **Versions** n'est pas visible, sélectionnez **Vue > Versions**
- Sélectionnez la version à modifier.
- Cliquez sur l'icône des propriétés :



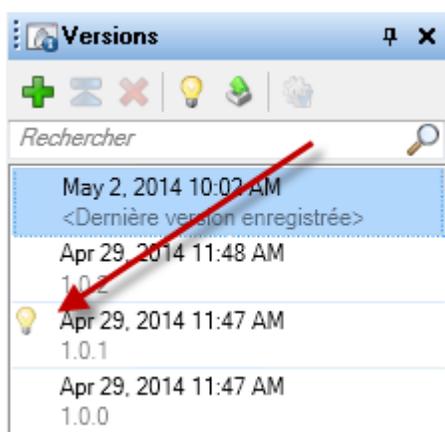
5. Dans le champ **Name**, entrez le nom de la version. Vous pouvez utiliser des numéros de version ou tout nom significatif. Ce nom est entièrement de votre choix.
6. Dans le champ **Comment**, vous pouvez entrer un commentaire plus long qui décrit, plus en détails, l'objectif de la version des modifications que vous apportez. L'ajout d'un commentaire est facultatif.
7. Cliquez sur **OK**.

Exposition d'une version

Si vous avez enregistré plusieurs versions d'un flux de données, vous pouvez choisir la version à exposer pour l'exécution.

1. Dans Enterprise Designer, ouvrez le flux de données.
2. Si le volet **Versions** n'est pas visible, sélectionnez **Vue > Versions**
3. Dans le volet **Versions**, sélectionnez la version du flux de données à exposer.
4. Sélectionnez **Fichier > Exposer/Ne plus exposer et Enregistrer**

La version sélectionnée est désormais exposée et disponible pour l'exécution. La version avec l'ampoule en regard est la version exposée, comme indiqué ici :



Lorsqu'un flux de données est exposé, le bouton de l'ampoule dans la barre d'outils Enterprise Designer indique que le flux de données est exposé comme indiqué ici :



L'ampoule indique que le flux de données est exposé même si vous visionnez une version différente que la version exposée. Si vous cliquez sur l'ampoule lors du visionnage d'une version non exposée, cela basculera la version exposée avec la version que vous visionnez actuellement. Si vous cliquez sur l'ampoule lors du visionnage de la version exposée, cela annulera l'exposition du flux de données.

3 - Inspection et tests

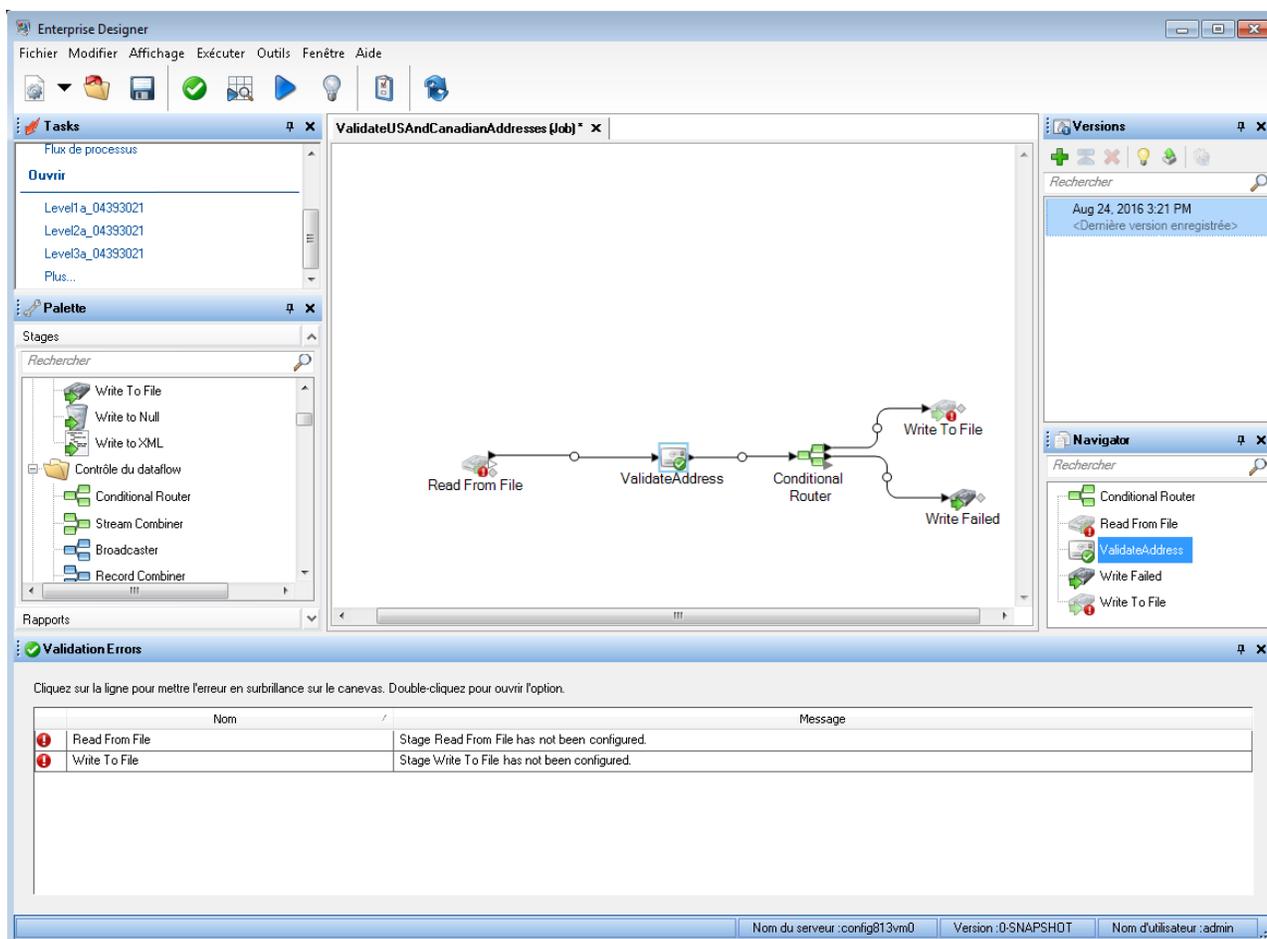
In this section

Recherche d'erreurs dans un flux	145
Inspection d'un flux de données	146
Test d'un service avec Management Console	151

Recherche d'erreurs dans un flux

Enterprise Designer recherche automatiquement les erreurs dans un flux lorsque vous exécutez un flux, lorsque vous exécutez une inspection, lorsque vous exposez un flux ou lorsque vous enregistrez un flux exposé. Vous pouvez également rechercher les erreurs en cliquant sur le bouton de validation .

Lorsqu'une erreur est détectée, le volet **Validation** apparaît au bas de la fenêtre Enterprise Designer. Cliquez sur une erreur pour mettre en surbrillance l'erreur sur le canevas. Double-cliquez sur une erreur pour ouvrir la fenêtre d'options de l'élément contenant l'erreur.



The screenshot shows the Enterprise Designer interface with a dataflow diagram. The diagram includes stages: Read From File, ValidateAddress, Conditional Router, Write To File, and Write Failed. The 'Validation Errors' pane at the bottom is active, displaying the following table:

Validation Errors	
Cliquez sur la ligne pour mettre l'erreur en surbrillance sur le canevas. Double-cliquez pour ouvrir l'option.	
Nom	Message
Read From File	Stage Read From File has not been configured.
Write To File	Stage Write To File has not been configured.

At the bottom of the window, the status bar shows: Nom du serveur : config813vm0 | Version : 0-SNAPSHOT | Nom d'utilisateur : admin

Inspection d'un flux de données

Pour voir l'effet de votre flux de données sur les données d'entrée à différents points du flux de données, utilisez l'outil d'inspection d'Enterprise Designer. L'inspection vous permet de vérifier que le flux de données a l'effet souhaité sur vos données, d'isoler les problèmes ou d'identifier les enregistrements contenant des défauts.

1. Indiquez les données à utiliser pour l'inspection.

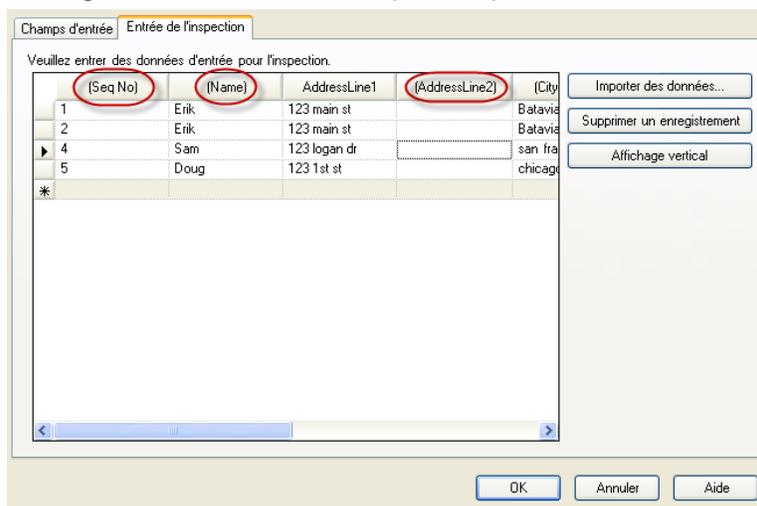
Ces données doivent être représentatives des données réelles, ou, si vous dépannez un problème spécifique, les données doivent être celles qui provoquent le problème que vous cherchez à résoudre. Il existe deux manières d'indiquer les données à utiliser pour l'inspection, suivant que vous inspectez un service ou un job.

Scénario	Description
Pour indiquer les données d'inspection d'un job	Lors de l'inspection d'un job, les données utilisées pour l'inspection sont les données indiquées dans le stage source. L'outil d'inspection des données peut traiter un maximum de 50 enregistrements, qui sont par défaut les 50 premiers enregistrements du fichier ou de la base de données d'entrée. Si vous souhaitez utiliser des données commençant ailleurs qu'au premier enregistrement, double-cliquez sur le stage Read From File et renseignez le champ Enregistrement de départ dans l'onglet Exécution.
Pour indiquer les données d'inspection d'un service	<p>Les flux de données de service utilisent un stage Input pour définir leur entrée. Parce qu'un stage Input n'a pas accès aux données lorsque vous modifiez le flux de données, vous devez définir des données d'inspection dans le stage Input sous l'onglet Données d'inspection. Vous pouvez indiquer jusqu'à 50 enregistrements.</p> <p>Il existe plusieurs façons de saisir des données d'inspection dans un stage Input.</p> <ul style="list-style-type: none"> • Si vous ne souhaitez utiliser que quelques enregistrements pour l'inspection, vous pouvez saisir manuellement les données en question. <p>Conseil : Si vous souhaitez enregistrer les données d'inspection que vous saisissez pour les réutiliser dans un autre stage, vous pouvez exporter les données d'inspection dans un fichier texte en cliquant sur Exporter les données.</p> <ul style="list-style-type: none"> • Si vos données sont contenues dans un fichier CSV ou TXT, vous pouvez importer ces données en cliquant sur Importer des données. Les données doivent utiliser l'un de ces délimiteurs :

Scénario Description

- \t
 - |
 - ,
 - ;
- Vous pouvez copier des données délimitées provenant d'une autre application et les coller dans l'éditeur de données d'inspection.

L'onglet **Entrée à inspecter** indique les données devant passer directement en englobant le nom du champ entre parenthèses, comme ci-après :



Remarque : Certains types de champ présentent des restrictions quand ils sont utilisés pour l'inspection :

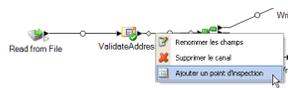
- Les champs double et float doivent contenir uniquement des données numériques. Le champ peut-être comporter jusqu'à 16 chiffres et 6 décimales. La notation exponentielle n'est pas prise en charge dans l'inspection.
- Les champs integer et long doivent contenir uniquement des données numériques.

2. Ils indiquent les points du flux de données où vous souhaitez afficher les données.

Scénario Description

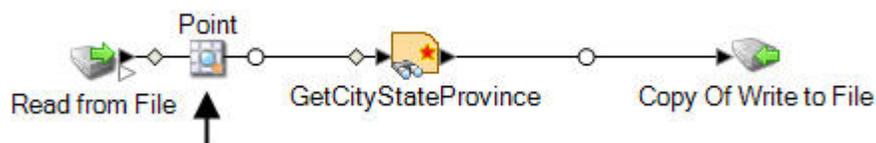
Pour ajouter un point d'inspection à un canal

Cliquez avec le bouton droit de la souris sur le nœud Renommer d'un canal et sélectionnez **Ajouter un point d'inspection**.



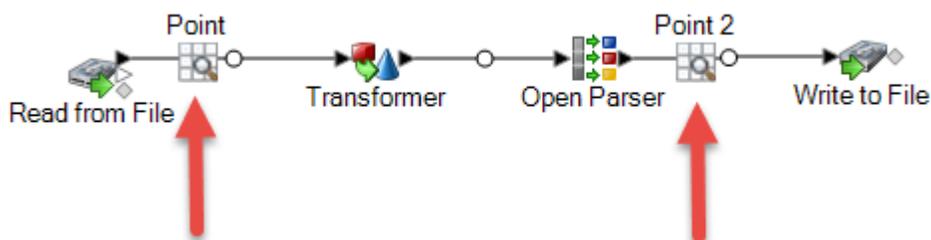
Scénario **Description**

Un point est ajouté au job :



Pour comparer des enregistrements à deux points d'un flux de données

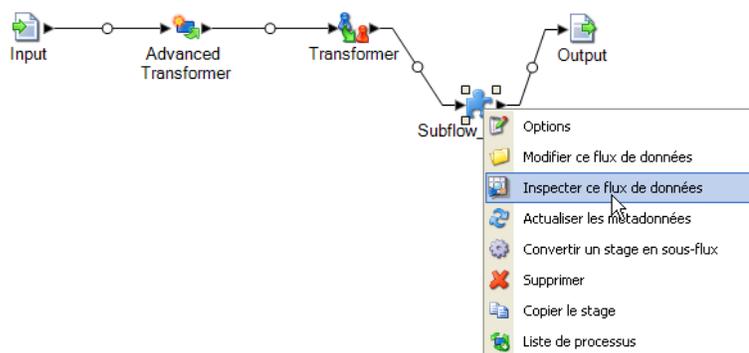
Ajoutez deux points d'inspection au niveau des points du flux de données que vous souhaitez comparer :



Conseil : Si vous utilisez l'inspection pour identifier un problème, inspectez en premier les points d'approche du flux de données, puis approfondissez jusqu'à l'endroit du problème éventuel.

Pour inspecter un sous-flux imbriqué dans un job ou dans un service

Cliquez avec le bouton droit de la souris sur le stage du sous-flux et sélectionnez **Inspecter ce flux de données**:



Les données d'entrée (dans un job) ou les données d'inspection (dans un service) sont transmises automatiquement au sous-flux, de sorte qu'il n'est

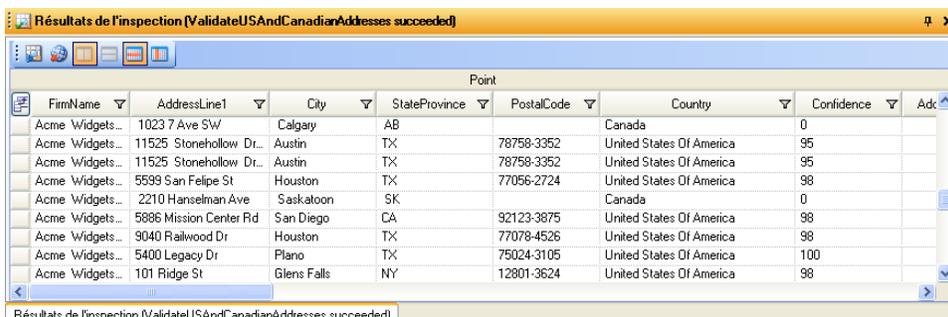
Scénario Description

pas nécessaire de saisir les données d'inspection dans le stage d'entrée du sous-flux.

Remarque : Lorsque vous inspectez un sous-flux, la version exposée du sous-flux apparaît. Si vous modifiez le sous-flux et souhaitez réexécuter l'inspection, vous devez exposer la nouvelle version.

3. Sélectionnez **Exécuter > Inspecter le flux actif** ou cliquez sur le bouton **Inspecter le flux actif** de la barre d'outils.

Si vous avez indiqué un point d'inspection, le volet **Résultats de l'inspection** affiche les données inspectées dans une vue horizontale. Vous pouvez modifier la mise en page de la vue en utilisant les icônes de la barre d'outils au-dessus du tableau. Si votre inspection de données est de type hiérarchique, celle-ci ne pourra pas être affichée de manière verticale.



Résultats de l'inspection (ValidateUSAndCanadianAddresses succeeded)

FirmName	AddressLine1	City	StateProvince	PostalCode	Country	Confidence	Adt
Acme Widgets...	1023 7 Ave SW	Calgary	AB		Canada	0	
Acme Widgets...	11525 Stonehollow Dr.	Austin	TX	78758-3352	United States Of America	95	
Acme Widgets...	11525 Stonehollow Dr.	Austin	TX	78758-3352	United States Of America	95	
Acme Widgets...	5599 San Felipe St	Houston	TX	77056-2724	United States Of America	98	
Acme Widgets...	2210 Hanselman Ave	Saskatoon	SK		Canada	0	
Acme Widgets...	5886 Mission Center Rd	San Diego	CA	92123-3875	United States Of America	98	
Acme Widgets...	9040 Railwood Dr	Houston	TX	77078-4526	United States Of America	98	
Acme Widgets...	5400 Legacy Dr	Plano	TX	75024-3105	United States Of America	100	
Acme Widgets...	101 Ridge St	Glens Falls	NY	12801-3624	United States Of America	98	

Résultats de l'inspection (ValidateUSAndCanadianAddresses succeeded)

Remarque : Les données de date et d'heure sont affichées au format spécifié dans les options de conversion de type.

Conseil : Vous pouvez déplacer un point d'inspection en le faisant glisser jusqu'à un autre canal. Les données de l'inspection se mettent à jour automatiquement.

Si vous avez indiqué deux points d'inspection, le volet **Résultats de l'inspection** affiche les enregistrements tels qu'ils existent au niveau des deux points. Le volet gauche affiche le point d'inspection le plus à gauche dans le flux de données et le volet droit affiche le point d'inspection le plus à droite dans le flux de données. Cliquez sur un enregistrement dans le volet droit pour mettre en surbrillance l'enregistrement correspondant dans le volet gauche, afin de voir comment l'enregistrement a été modifié entre les deux points d'inspection.

Point 2				Point			
FirmName	AddressLine1	City	StateProvince	FirmName	AddressLine1	City	StateProvince
Acme Widgets...	11525 Stonehollow Dr...	Austin	TX	Acme Widgets...	11525 Stonehollow Dr...	Austin	TX
Acme Widgets...	11525 Stonehollow Dr...	Austin	TX	Acme Widgets...	11525 Stonehollow Dr...	Austin	TX
Acme Widgets...	5599 San Felipe St	Houston	TX	Acme Widgets...	5599 San Felipe St	Houston	TX
Acme Widgets...	2210 Hanselman Ave	Saskatoon	SK	Acme Widgets...	2210 Hanselman Ave	Saskatoon	SK
Acme Widgets...	5886 Mission Center Rd	San Diego	CA	Acme Widgets...	5886 Mission Center Rd	San Diego	CA
Acme Widgets...	9040 Railwood Dr	Houston	TX	Acme Widgets...	9040 Railwood Dr	Houston	TX
Acme Widgets...	5400 Legacy Dr	Plano	TX	Acme Widgets...	5400 Legacy Dr	Plano	TX
Acme Widgets...	101 Ridge St	Glens Falls	NY	Acme Widgets...	101 Ridge St	Glens Falls	NY
Acme Widgets...	4321 N Ballard Rd	Appleton	WI	Acme Widgets...	4321 N Ballard Rd	Appleton	WI

Chaque colonne représente un champ du flux de données. Les colonnes sont organisées dans l'ordre alphabétique. Les nouveaux champs ajoutés entre les points d'inspection sont affichés dans le volet droit après les colonnes d'origine. Pour réorganiser les colonnes, cliquez dessus et glissez-les dans l'ordre de votre choix.

Ces situations influencent la manière dont les résultats d'inspection sont affichés pour deux points d'inspection :

- S'il existe un stage Sorter entre les deux points d'inspection, les enregistrements des résultats d'inspection sont triés tels qu'ils l'étaient avant le stage Sorter. Le tri est ignoré sur le deuxième point d'inspection, afin que vous puissiez comparer les enregistrements correspondants de chaque point d'inspection côte à côte.
 - S'il existe des stages, entre les deux points d'inspection, qui créent de nouveaux enregistrements, comme un stage Aggregator, les enregistrements affichés sur le second point d'inspection n'auront pas d'enregistrement correspondant sur le premier point d'inspection.
 - Les enregistrements qui existent au deuxième point d'inspection, mais pas au premier, sont affichés au bas de la liste d'enregistrements du deuxième point d'inspection.
4. Lorsque vous mettez à jour ou que vous modifiez le dataflow, cliquez sur **Exécuter > Inspecter le flux actif** pour actualiser les résultats d'inspection.
 5. Quand vous fermez le volet Résultats de l'inspection, les données d'inspection sont perdues. De même, quand vous fermez un job, les points d'inspection et les données d'inspection sont perdues. Pour enregistrer les résultats d'inspection dans un fichier :
 - a) Dans la grille de résultats d'inspection, sélectionnez les lignes que vous souhaitez enregistrer. Vous pouvez sélectionner toutes les données en cliquant avec le bouton droit de la souris dans n'importe quel volet et en cliquant sur **Sélectionner tout**.
 - b) Sélectionnez **Copier** dans le menu contextuel.
 - c) Ouvrez l'application dans laquelle vous voulez enregistrer les données (par exemple, Microsoft Excel ou Bloc-notes).
 - d) Dans l'application, collez les données.
 - e) Enregistrez le fichier.

Test d'un service avec Management Console

Management Console propose une fonction d'aperçu qui vous permet d'envoyer des données de test à un service et d'afficher les résultats.

1. Dans un navigateur Web, accédez à l'URL suivante :

`http://server.port/managementconsole`

Où : *server* est l'adresse IP ou le nom de serveur de votre serveur Spectrum™ Technology Platform et *port* le port HTTP utilisé par Spectrum™ Technology Platform. Par défaut, le port HTTP est 8080.

2. Accédez au menu **Services** et sélectionnez le module contenant le service pour lequel vous souhaitez tester.
3. Cliquez sur le service à tester.
4. Cliquez sur **Aperçu**.
5. Entrez les données d'entrée à utiliser pour votre test. Pour importer des données à partir d'un fichier, cliquez sur le bouton Importer .
6. Cliquez sur **Exécuter l'aperçu**.

4 - Exécution d'un flux

In this section

Exécution d'un job ou d'un process flow	153
Exposition d'un service	176
Options de Runtime	178
Configuration des notifications par courrier électronique pour un flux de données	182

Exécution d'un job ou d'un process flow

Exécution d'un flux de données dans Enterprise Designer

La procédure suivante explique comment exécuter manuellement un job ou un processflow.

1. Dans Enterprise Designer, sélectionnez **Fichier > Ouvrir** et ouvrez le flux de données à exécuter.
2. Validez un flux de données avant de l'exécuter afin de vous assurer qu'il ne contient aucune erreur. Pour valider un flux de données, sélectionnez **Exécuter > Valider**.
3. Sélectionnez **Exécuter > Exécuter le flux actuel**.

Exécution d'un job à partir de la ligne de commande

Avant de pouvoir exécuter un job à partir de la ligne de commande, il doit être exposé. Pour exposer un job, ouvrez-le dans Enterprise Designer et sélectionnez **Fichier > Exposer/Ne plus exposer et Enregistrer**.

Pour exécuter un job à partir de la ligne de commande, vous devez installer l'utilitaire Job Executor sur le système dans lequel exécuter le job. L'Exécuteur de job est disponible depuis la page d'accueil de Spectrum™ Technology Platform sur le serveur Spectrum™ Technology Platform (par exemple, <http://monserver:8080>).

Utilisation

```
java -jar jobexecutor.jar -u UserID -p Password -j Job [Optional Arguments]
```

Obligatoire	Argument	Description
Non	-?	Imprime les informations d'utilisation.
Non	-d <i>delimiter</i>	Définit un délimiteur d'instance/état. Ceci apparaît uniquement dans les sorties synchrones.
Non	-e	Utilisez une connexion HTTPS sécurisée pour communiquer avec le serveur Spectrum™ Technology Platform.
Non	-f <i>property file</i>	Indique un chemin vers un fichier de propriétés d'un job. Un fichier de propriété de job contient des arguments de Job Executor. Pour plus d'informations sur les fichiers de propriétés

Obligatoire	Argument	Description
		d'un job, voir Utilisation d'un fichier de propriétés de job à la page 161.
Non	<code>-h host name</code>	Indique le nom ou l'adresse IP du serveur Spectrum™ Technology Platform.
Non	<code>-i poll interval</code>	Indique la fréquence de vérification des jobs achevés, en secondes. Ceci ne s'applique qu'en mode synchrone.
Oui	<code>-j job name</code>	Une liste de jobs séparés par des virgules à exécuter. Les noms de jobs sont sensibles à la casse. Les jobs sont lancés selon l'ordre répertorié.
Non	<code>-n email list</code>	Indique une liste d'adresses email supplémentaires séparées par des virgules pour les notifications de job configurées.
Non	<code>-o property file</code>	<p>Indique un chemin vers un fichier de propriétés des options de dataflow. Utilisez un fichier de propriétés d'options de dataflow pour définir des options pour les stages du dataflow. Pour définir des options de dataflow à l'aide d'un fichier de propriétés, vous devez configurer le dataflow pour qu'il expose les options de stage lors de l'exécution. Pour plus d'informations, reportez-vous à la section Ajout d'options d'exécution de flux de données à la page 178.</p> <p>Par exemple, un fichier de propriétés d'options de dataflow pour un dataflow contenant un stage Assign GeoTAX Info peut ressembler à ceci :</p> <pre>OutputCasing=U UseStreetLevelMatching=N TaxKey=T Database.GTX=gs1</pre>
Oui	<code>-p password</code>	Mot de passe de l'utilisateur.
Non	<code>-r</code>	Spécifiez cet argument pour renvoyer un rapport détaillé sur le job. Cette option ne fonctionne que si vous spécifiez également <code>-w</code> . Le rapport contient les informations suivantes :

Obligatoire	Argument	Description
		<ul style="list-style-type: none"> • Position 1—Nom du job • Position 2—ID du processus du job • Position 3—État • Position 4—Date/Heure de début (JJ/MM/AAAAA HH:MM:SS) • Position 5—Date/Heure de fin (JJ/MM/AAAAA HH:MM:SS) • Position 6—Nombre d'enregistrements ayant réussi • Position 7—Nombre d'enregistrements ayant échoué • Position 8—Nombre d'enregistrements non conformes • Position 9—Actuellement non employé <p>Par exemple :</p> <pre>MySimpleJob 4 succeeded 04/09/2010 14:50:47 04/09/2010 14:50:47 100 0 0 </pre> <p>Les informations sont délimitées à l'aide du délimiteur spécifié dans l'argument <code>-d</code>.</p>
Non	<code>-s port</code>	Le socket (port) sur lequel le serveur Spectrum™ Technology Platform s'exécute. La valeur par défaut est 8080.
Non	<code>-t timeout</code>	Configure le délai d'attente (en secondes) pour le mode synchrone. La valeur par défaut est 3 600. La valeur maximum est 2147483. Il s'agit d'un délai d'attente global agrégé qui représente la durée d'attente maximale que tous les jobs générés se terminent.
Oui	<code>-u user name</code>	Le nom de connexion de l'utilisateur.
Non	<code>-v</code>	Renvoie une sortie détaillée.
Non	<code>-w</code>	Exécute Job Executor en mode synchrone. Cela signifie que l'exécution de Job Executor se poursuit jusqu'à ce que le job soit terminé. Si vous ne spécifiez pas <code>-w</code> , Job Executor s'arrête après le démarrage du job, sauf si le job lit ou écrit dans des fichiers sur le serveur. Dans ce cas, Job Executor s'exécutera jusqu'à

Obligatoire	Argument	Description
		ce que tous les fichiers locaux soient traités, puis il s'arrêtera.
Non	<i>StageName=Protocol:FileName</i>	Remplace le fichier d'entrée ou de sortie indiqué dans Read from File ou Write to File. Pour plus d'informations, reportez-vous à la section Remplacement des emplacement des fichiers Job à la page 156.
Non	<i>StageName:schema=Protocol:SchemaFile</i>	Remplace la définition de disposition de fichier spécifiée dans Read from File ou Write to File par une définition définie dans un fichier de schéma. Pour plus d'informations, reportez-vous à la section Remplacement du format de fichier dans la ligne de commande à la page 158.

Exemple d'utilisation de Job Executor

L'exemple suivant affiche la ligne de commande et sa sortie :

```
D:\spectrum\job-executor>java -jar jobexecutor.jar -u user123
-p "mypassword" -j validateAddressJob1 -h spectrum.example.com
-s 8888 -w -d "%" -i 1 -t 9999validateAddressJob1%105%succeeded
```

Dans cet exemple, la sortie indique que le job nommé « validateAddressJob1 » a été exécuté (avec l'identifiant 105) sans erreur. Le résultat aurait aussi pu être « failed » (échec) ou « running » (en cours d'exécution).

Remplacement des emplacement des fichiers Job

Lorsque vous exécutez un job dans la ligne de commande à l'aide de Job Executor ou de l'utilitaire Administration, vous pouvez remplacer le fichier d'entrée indiqué dans le stage source du dataflow (tel que Read from File), ainsi que le fichier de sortie spécifié dans le stage récepteur du dataflow (tel que Write to File).

Pour ce faire, dans Job Executor, indiquez les éléments suivants à la fin de la commande de la ligne de commande de Job Executor :

```
StageName=protocole:nom de fichier
```

Dans l'utilitaire Administration, utilisez l'argument `--l` dans la commande `job execute` :

```
--l StageName=Protocol:FileName
```

Où :

StageName

L'intitulé du stage affiché sous l'icône du stage dans le dataflow dans Enterprise Designer. Par exemple, si le stage est intitulé « Read from File », vous devez indiquer `Read from File` comme nom de stage.

Pour indiquer un stage au sein d'un sous-flux ou dataflow intégré, préfacez le nom de stage par le nom du sous-flux ou dataflow intégré, suivi d'un point, puis du nom de stage :

NomDeSous-FluxOuFluxDeDonnéesIntégré.NomDeStage

Par exemple, pour indiquer un stage nommé Write to File dans un sous-flux nommé Subflow1, vous devez procéder comme suit :

`Subflow1.Write to File`

Pour indiquer un stage dans un dataflow intégré se trouvant dans un autre dataflow intégré, ajoutez le dataflow parent, en séparant chacun par un point. Par exemple, si Embedded Dataflow 2 se trouve à l'intérieur d'Embedded Dataflow 1 et que vous souhaitez indiquer le stage Write to File dans Embedded Dataflow 2, vous devez procéder comme suit :

`Embedded Dataflow 1.Embedded Dataflow 2.Write to File`

Protocol

Protocole de communication. L'un des éléments suivants :

file Utilisez le protocole file si le fichier figure sur le même ordinateur que le serveur Spectrum™ Technology Platform. Par exemple, sur Windows, spécifiez :

`"file:C:/myfile.txt"`

Sur Unix ou Linux, spécifiez :

`"file:/testfiles/myfile.txt"`

esclient Utilisez le protocole esclient si le fichier se trouve sur l'ordinateur sur lequel vous exécutez le job s'il s'agit d'un autre ordinateur que celui qui exécute le serveur Spectrum™ Technology Platform. Utilisez le format suivant :

`esclient:ComputerName/path to file`

Par exemple,

`esclient:mycomputer/testfiles/myfile.txt`

Remarque : Si vous exécutez le job sur le serveur lui-même, vous pouvez utiliser le protocole file ou esclient, mais vous obtiendrez probablement de meilleures performances si vous utilisez le protocole file.

Si le nom d'hôte du serveur Spectrum™ Technology Platform ne peut pas être résolu, il se peut que vous receviez un message d'erreur « Une erreur s'est produite lors de l'accès au fichier ». Pour résoudre ce problème, ouvrez ce fichier sur le serveur :

`SpectrumLocation/server/app/conf/spectrum-container.properties`. Définissez la propriété `spectrum.runtime.hostname` sur l'adresse IP du serveur.

esfile Utilisez le protocole esfile si le fichier se trouve sur un serveur de fichiers. Le serveur de fichiers doit être défini dans Management Console comme une ressource. Utilisez le format suivant :

```
esfile://serveur de fichiers/chemin d'accès au fichier
```

Par exemple,

```
esfile://myserver/testfiles/myfile.txt
```

Où myserver est une ressource de serveur de fichiers FTP définie dans Management Console.

webhdfs Utilisez le protocole webhdfs si le fichier se trouve sur un serveur de fichiers distribué Hadoop. Le serveur HDFS doit être défini dans Management Console comme une ressource. Utilisez le format suivant :

```
webhdfs://serveur de fichiers/chemin d'accès au fichier
```

Par exemple,

```
webhdfs://myserver/testfiles/myfile.txt
```

Où myserver est une ressource de serveur de fichiers HDFS définie dans Management Console.

FileName

Chemin complet vers le fichier à utiliser comme entrée ou sortie.

Remarque : Vous devez utiliser des barres obliques (/) pour les chemins d'accès, et non des barres obliques inversées.

Pour spécifier plusieurs substitutions, séparez chacune par une virgule.

Exemple de remplacement de fichier

La commande suivante de Job Executor utilise le fichier C:/myfile_input.txt comme fichier d'entrée pour le stage Read from File et le fichier C:/myfile_output.txt comme fichier de sortie pour le stage Write to File.

```
java -jar jobexecutor.jar -j Job1 -u Bob1234 -p "" "Read from File"="file:C:/myfile_input.txt" "Write to File"="file:C:/myfile_output.txt"
```

Remplacement du format de fichier dans la ligne de commande

Lorsque vous exécutez un job à l'aide de Job Executor ou de l'utilitaire Administration, vous pouvez remplacer la mise en forme de fichier (ou schéma) du fichier spécifié dans les stages Read from File et Write to File du flux de données.

Pour ce faire, dans Job Executor, indiquez les éléments suivants à la fin de la commande de la ligne de commande de Job Executor :

StageName: schéma =protocole:SchemaFile

Dans l'utilitaire Administration, utilisez l'argument `--l` dans la commande `job execute` :

--lStageName:schema=Protocol:SchemaFile

Où :

StageName

L'intitulé du stage affiché sous l'icône du stage dans le dataflow dans Enterprise Designer. Par exemple, si le stage est intitulé « Read from File », vous devez indiquer `Read from File` comme nom de stage.

Pour indiquer un stage au sein d'un sous-flux ou dataflow intégré, préfacez le nom de stage par le nom du sous-flux ou dataflow intégré, suivi d'un point, puis du nom de stage :

NomDeSous-FluxOuFluxDeDonnéesIntégré.NomDeStage

Par exemple, pour indiquer un stage nommé Write to File dans un sous-flux nommé Subflow1, vous devez procéder comme suit :

Subflow1.Write to File

Pour indiquer un stage dans un dataflow intégré se trouvant dans un autre dataflow intégré, ajoutez le dataflow parent, en séparant chacun par un point. Par exemple, si Embedded Dataflow 2 se trouve à l'intérieur d'Embedded Dataflow 1 et que vous souhaitez indiquer le stage Write to File dans Embedded Dataflow 2, vous devez procéder comme suit :

Embedded Dataflow 1.Embedded Dataflow 2.Write to File

Protocol

Protocole de communication. L'un des éléments suivants :

file Utilisez le protocole file si le fichier figure sur le même ordinateur que le serveur Spectrum™ Technology Platform. Par exemple, sur Windows, spécifiez :

"file:C:/myfile.txt"

Sur Unix ou Linux, spécifiez :

"file:/testfiles/myfile.txt"

esclient Utilisez le protocole esclient si le fichier se trouve sur l'ordinateur sur lequel vous exécutez le job s'il s'agit d'un autre ordinateur que celui qui exécute le serveur Spectrum™ Technology Platform. Utilisez le format suivant :

esclient:ComputerName/path to file

Par exemple,

esclient:mycomputer/testfiles/myfile.txt

Remarque : Si vous exécutez le job sur le serveur lui-même, vous pouvez utiliser le protocole file ou esclient, mais vous obtiendrez probablement de meilleures performances si vous utilisez le protocole file.

Si le nom d'hôte du serveur Spectrum™ Technology Platform ne peut pas être résolu, il se peut que vous receviez un message d'erreur « Une erreur s'est produite lors de l'accès au fichier ». Pour résoudre ce problème, ouvrez ce fichier sur le serveur : `SpectrumLocation/server/app/conf/spectrum-container.properties`. Définissez la propriété `spectrum.runtime.hostname` sur l'adresse IP du serveur.

esfile Utilisez le protocole esfile si le fichier se trouve sur un serveur de fichiers. Le serveur de fichiers doit être défini dans Management Console comme une ressource. Utilisez le format suivant :

```
esfile://serveur de fichiers/chemin d'accès au fichier
```

Par exemple,

```
esfile://myserver/testfiles/myfile.txt
```

Où myserver est une ressource de serveur de fichiers FTP définie dans Management Console.

webhdfs Utilisez le protocole webhdfs si le fichier se trouve sur un serveur de fichiers distribué Hadoop. Le serveur HDFS doit être défini dans Management Console comme une ressource. Utilisez le format suivant :

```
webhdfs://serveur de fichiers/chemin d'accès au fichier
```

Par exemple,

```
webhdfs://myserver/testfiles/myfile.txt
```

Où myserver est une ressource de serveur de fichiers HDFS définie dans Management Console.

SchemaFile

Chemin d'accès complet au fichier qui définit la disposition à utiliser.

Remarque : Vous devez utiliser des barres obliques (/) pour les chemins d'accès, et non des barres obliques inversées.

Pour créer un fichier de schéma, définissez la disposition que vous souhaitez dans Read from File ou Write to File, puis cliquez sur le bouton **Exporter** pour créer un fichier XML définissant la disposition.

Remarque : Vous ne pouvez pas remplacer le type de données d'un champ dans un fichier de schéma lorsque vous utilisez Job Executor. La valeur dans l'élément <Type>, qui est un enfant de l'élément <FieldSchema>, doit correspondre au type du champ indiqué dans le stage Read from File ou Write to File du flux de données.

Exemple de remplacement de format de fichier

La commande suivante de Job Executor utilise le fichier `C:/myschema.xml` comme définition de disposition pour le fichier lu par le stage Read from File.

```
java -jar jobexecutor.jar -j Job1 -u Bob1234 -p "" "Read from File":schema="file:C:/myschema.xml"
```

Utilisation d'un fichier de propriétés de job

Un fichier de propriétés de job contient des arguments qui contrôlent l'exécution des jobs lorsque vous utilisez Job Executor ou l'utilitaire Administration pour exécuter un job. Utilisez un fichier de propriétés de job si vous souhaitez réutiliser des arguments en indiquant un seul argument dans la ligne de commande (`-f`), plutôt que d'indiquer chaque argument individuellement dans la ligne de commande.

Pour créer un fichier de propriété, créez un fichier texte avec un argument sur chaque ligne. Par exemple :

```
d %
h spectrum.mydomain.com
i 30
j validateAddressJob1
u user
p password
s 8888
t 9999
w true
```

Le fichier de propriété de job peut contenir ces arguments :

Obligatoire	Argument	Description
Non	?	Imprime les informations d'utilisation.
Non	d <i>delimiter</i>	Définit un délimiteur d'instance/état. Ceci apparaît uniquement dans les sorties synchrones.
Non	e	Utilisez une connexion HTTPS sécurisée pour communiquer avec le serveur Spectrum™ Technology Platform.
Non	h <i>hostname</i>	Indique le nom ou l'adresse IP du serveur Spectrum™ Technology Platform.
Non	i <i>pollinterval</i>	Indique la fréquence de vérification des jobs achevés, en secondes. Ceci ne s'applique qu'en mode synchrone.
Oui	j <i>jobname</i>	Une liste de jobs séparés par des virgules à exécuter. Les noms de jobs sont sensibles à la casse. Les jobs sont lancés selon l'ordre répertorié.

Obligatoire	Argument	Description
Non	<code>n</code> <i>emallist</i>	Indique une liste d'adresses email supplémentaires séparées par des virgules pour les notifications de job configurées.
Oui	<code>p</code> <i>password</i>	Mot de passe de l'utilisateur.
Non	<code>r</code>	<p>Renvoie une liste délimitée contenant les informations suivantes à propos du job écrit dans une sortie standard :</p> <ul style="list-style-type: none"> • Position 1—Nom du job • Position 2—ID du processus du job • Position 3—État • Position 4—Date/Heure de début (JJ/MM/AAAAA HH:MM:SS) • Position 5—Date/Heure de fin (JJ/MM/AAAAA HH:MM:SS) • Position 6—Nombre d'enregistrements ayant réussi • Position 7—Nombre d'enregistrements ayant échoué • Position 8—Nombre d'enregistrements non conformes • Position 9—Actuellement non employé <p>Les informations sont délimitées à l'aide du délimiteur spécifié dans l'argument <code>-d</code>. Par exemple :</p> <pre>MySimpleJob 4 succeeded 04/09/2010 14:50:47 04/09/2010 14:50:47 100 0 0 </pre>
Non	<code>s</code> <i>port</i>	Le socket (port) sur lequel le serveur Spectrum™ Technology Platform s'exécute. La valeur par défaut est 8080.
Non	<code>t</code> <i>timeout</i>	Configure le délai d'attente (en secondes) pour le mode synchrone. La valeur par défaut est 3 600. La valeur maximum est 2147483. Il s'agit d'un délai d'attente global agrégé qui représente la durée d'attente maximale que tous les jobs générés se terminent.
Oui	<code>u</code> <i>username</i>	Le nom de connexion de l'utilisateur.
Non	<code>v</code>	Renvoie une sortie détaillée.
Non	<code>w</code>	Spécifie d'attendre que les jobs se terminent en mode synchrone.

Utilisation d'arguments de ligne de commande et d'un fichier de propriétés

Une combinaison mêlant entrée de ligne de commande et entrée de fichier de propriétés est également valide. Par exemple :

```
java -jar jobexecutor.jar -f /dcb/job.properties -j job1
```

Dans ce cas, les arguments de la ligne de commande ont priorité sur les arguments indiqués dans le fichier de propriétés. Dans l'exemple ci-dessus, le job `job1` aurait priorité sur un job indiqué dans le fichier de propriétés.

Exécution d'un processflow à partir de la ligne de commande

Pour exécuter un processflow depuis la ligne de commande, utilisez Process Flow Executor. Installez Process Flow Executor à partir de la page d'accueil de Spectrum™ Technology Platform (par exemple, <http://monserveur:8080>).

Remarque : Vous pouvez également utiliser l'utilitaire Administration pour exécuter des processflow depuis la ligne de commande.

Utilisation

```
java -jar pflowexecutor.jar -r ProcessFlowName -u UserID -p Password [Arguments facultatifs]
```

Obligatoire	Argument	Description
Non	-?	Imprime les informations d'utilisation.
Non	-d <i>DelimiterCharacter</i>	Définit un délimiteur à utiliser pour séparer les informations de statut affichées lorsque vous exécutez la commande. L'emplacement par défaut est « ». Par exemple, à l'aide du caractère par défaut, les informations suivantes sont affichées à la ligne de commande lorsque vous exécutez un processflow nommé « MyProcessflow » : MyProcessflow 1 Succeeded
Non	-e	Utilisez une connexion HTTPS pour communiquer avec le serveur Spectrum™ Technology Platform. Remarque : Si vous indiquez les substitutions de fichier cet argument ne doit pas être le dernier argument spécifié.
Non	-f <i>PropertyFile</i>	Indique un chemin vers un fichier de propriétés. Pour plus d'informations sur les fichiers de propriétés, voir Utilisation d'un fichier de propriétés d'un processflow à la page 165.
Non	-h <i>HostName</i>	Indique le nom ou l'adresse IP du serveur Spectrum™ Technology Platform.
Non	-i <i>PollInterval</i>	Indique la fréquence de vérification des jobs achevés, en secondes. La valeur par défaut est « 5 ».
Oui	-p <i>Password</i>	Mot de passe de l'utilisateur. Obligatoire.
Oui	-r <i>ProcessFlowNames</i>	Une liste de flux /de processus séparés par des virgules à exécuter. Obligatoire. Remarque : Si vous indiquez les substitutions de fichier cet argument ne doit pas être le dernier argument spécifié.

Obligatoire	Argument	Description				
Non	<code>-s Port</code>	Le socket (port) sur lequel le serveur Spectrum™ Technology Platform s'exécute. La valeur par défaut est 8080.				
Non	<code>-t Timeout</code>	Cette option est désapprouvée et est ignorée.				
Oui	<code>-u UserName</code>	Le nom de connexion de l'utilisateur. Obligatoire.				
Non	<code>-v Verbose</code>	Renvoie une sortie détaillée où <i>Verbose</i> prend l'une des valeurs suivantes : <table data-bbox="552 567 1250 661"> <tr> <td>true</td> <td>Renvoie une sortie détaillée.</td> </tr> <tr> <td>false</td> <td>Ne renvoie pas de sortie détaillée.</td> </tr> </table>	true	Renvoie une sortie détaillée.	false	Ne renvoie pas de sortie détaillée.
true	Renvoie une sortie détaillée.					
false	Ne renvoie pas de sortie détaillée.					
Remarque : Si vous indiquez les substitutions de fichier cet argument ne doit pas être le dernier argument spécifié.						
Non	<code>-w WaitToComplete</code>	Cette option est désapprouvée et est ignorée.				
Non	<code>StageName=FileName</code>	Remplace le fichier d'entrée ou de sortie indiqué dans le job. Pour plus d'informations, reportez-vous à la section Remplacement des emplacements de fichiers Process Flow .				

Exemples

Il s'agit d'une entrée de ligne de commande basique, comprenant nom de flux et ID utilisateur ainsi que mot de passe :

```
java -jar pflowexecutor.jar -r MyFlow1 -u Bob1234 -p "mypassword1"
```

Cet exemple illustre les mêmes informations que ci-dessus mais avec des arguments supplémentaires :

```
java -jar pflowexecutor.jar -r Flow1 -u Bob1234 -p "mypassword1" -h spectrum.example.com -s 8080 -w -d "%" -i 1
```

L'exemple suivant illustre la ligne de commande et le résultat :

```
D:\spectrum\pflow-executor>java -jar pflowexecutor.jar -u Bob1234 -p "mypassword1" -r validateAddressFlow1 -h spectrum.example.com -s 8080 -w -d "%" -i 1 -t 9999
validateAddressJob1%111% succeeded
```

Dans cet exemple, le processflow nommé validateAddressFlow1 a été exécuté (avec l'identifiant 111). Aucune erreur n'est survenue. Le résultat aurait aussi pu être « failed » (échec) ou « running » (en cours d'exécution).

Utilisation d'un fichier de propriétés d'un processflow

Un fichier de propriétés contient des arguments que vous pouvez réutiliser en spécifiant le chemin d'accès au fichier de propriétés avec l'argument `-f` dans l'exécuteur de processflow. Le fichier de propriétés doit contenir, au minimum, le processflow (`r`), l'ID d'utilisateur (`u`) et le mot de passe (`p`).

1. Ouvrez un éditeur de texte.
2. Indiquez un argument pour chaque ligne comme indiqué dans l'exemple suivant. Pour une liste des arguments, voir [Exécution d'un processflow à partir de la ligne de commande](#) à la page 163.

Remarque : Vous ne pouvez pas utiliser de fichier de propriétés pour remplacer des fichiers d'entrée ou de sortie. Le remplacement des fichiers d'entrée et de sortie peut uniquement être effectué via des arguments de ligne de commande.

```
d=%
h=myserver.mydomain.com
i=30
u=user
p=password
r=MyFlow1
s=8888
```

3. Enregistrez le fichier avec une extension de fichier `.properties` (par exemple, « exemple.properties »).
4. Lorsque vous exécutez l'Exécuteur de processflow, indiquez le chemin d'accès au fichier de propriétés en utilisant l'argument `-f`. Une combinaison mêlant entrée de ligne de commande et entrée de fichier de propriétés est également valide. Les arguments de la ligne de commande ont priorité sur les arguments indiqués dans le fichier de propriétés.

```
java -jar pflowexecutor.jar -f /dca/flow.properties -r MyFlow2
```

Dans l'exemple ci-dessus, le processflow MyFlow2 a priorité sur un job indiqué dans le fichier de propriétés.

Planification d'un dataflow

Un calendrier de dataflow exécute automatiquement un job ou un process flow automatiquement à une heure donnée. Vous pouvez programmer un dataflow à exécuter une fois, ou configurer son exécution périodique.

1. Si ce n'est pas déjà fait, exposez le job ou le processflow.
Vous pouvez exposer des jobs et des process flows en ouvrant le job ou le process flow dans Enterprise Designer et en sélectionnant **Fichier > Exposer/Ne plus exposer et Enregistrer**.
2. Ouvrez Management Console.
3. Naviguez jusqu'à **Exécution** puis cliquez sur **Planification**.
4. Cliquez sur **Ajouter** pour créer un nouveau calendrier ou, si vous voulez modifier un calendrier existant, choisissez le calendrier puis cliquez sur **Modifier**.
5. Dans la fenêtre **Ajouter Tâche** ou **Modifier Tâche**, choisissez les paramètres pour cette tâche.
 - **Nom de la tâche** : Le nom que vous voulez donner à cette tâche planifiée. Il s'agit du nom qui sera affiché dans la liste des tâches.
 - **Type du flux** : choisissez le type de processus que vous planifiez, soit un job ou un processflow.
 - **Nom du flux** : Sélectionnez le job ou processflow que vous voulez planifier. Seuls les jobs et processflow que vous avez enregistrés et exposés sont disponibles ici. Si le job ou processflow que vous voulez ne figure pas, ouvrez le job ou processflow dans Enterprise Designer et sélectionnez **Fichier > Exposer/Ne plus exposer et Enregistrer**.
 - **Initialiser la tâche** : cochez cette case pour exécuter le job ou processflow au moment spécifié. Décochez cette case pour suspendre le calendrier.
 - **Calendrier** : spécifiez la date et l'heure auxquelles vous voulez exécuter le job ou processflow.
6. Si le flux utilise des fichiers pour l'entrée ou la sortie, ces fichiers doivent résider sur le serveur Spectrum™ Technology Platform ou sur un serveur de fichiers défini en tant que ressource externe dans Management Console. Ceci s'applique tant aux jobs qu'aux activités de job dans un processflow. Si un stage source ou récepteur fait référence à un fichier sur un ordinateur client uniquement, exécutez l'une des procédures suivantes :

Option	Description
Option 1 : modifier le dataflow	<p>Déplacez le fichier dans le serveur Spectrum™ Technology Platform ou le serveur de fichiers, puis modifiez le dataflow :</p> <ol style="list-style-type: none"> 1. Ouvrez le flux dans Enterprise Designer. 2. Double-cliquez sur le stage source ou récepteur. 3. Dans le champ File name, cliquez sur le bouton naviguer. 4. Cliquez sur Ordinateur distant puis sélectionnez le fichier que vous voulez. <p>Remarque : Si vous exécutez Enterprise Designer sur le même ordinateur que le serveur Spectrum™ Technology Platform, un clic sur l'Ordinateur distant aura apparemment le même effet qu'un clic sur Mon ordinateur. Cependant, vous devez sélectionner le fichier à l'aide de l'icône Machine Distante afin que le système reconnaisse que le fichier appartient bien au serveur Spectrum™ Technology Platform.</p>

Option	Description
Option 2 : remplacez l'emplacement du fichier de dataflow quand le calendrier s'exécute.	<p>Dans le fichier de propriétés d'exécution d'un job, vous pouvez remplacer le fichier d'entrée indiqué dans le stage source du dataflow (tel que Read from File), ainsi que le fichier de sortie spécifié dans le stage récepteur du dataflow (tel que Write to File).</p> <ol style="list-style-type: none"> 1. Cliquez sur Options. 2. Sous Emplacement des fichiers du stage sélectionnez le stage faisant référence à un fichier local. 3. Cliquez sur Modifier et sélectionnez le fichier sur le serveur Spectrum™ Technology Platform.

7. Si vous voulez que le job ou processflow s'exécute dans un calendrier répété, cochez la case **Récurrence de Tâche** puis cliquez sur le bouton **Récurrence** et complétez les champs.
8. Si le dataflow est configuré pour la notification par courrier électronique, vous pouvez spécifier les destinataires supplémentaires pour les notifications qui seront envoyées lorsque le dataflow s'exécute.
 - a) Cliquez sur **Options**.
 - b) Sous Notification, cliquez sur **Ajouter**.
 - c) Saisissez l'adresse email à laquelle vous voulez que la notification soit envoyée. Par exemple, moi@monentreprise.com.
 - d) Cliquez sur **OK**.

Remarque : La notification doit être configurée dans Management Console pour que les notifications par courrier électronique fonctionnent. En outre, vérifiez que le dataflow a été configuré pour prendre en charge la notification. Pour ce faire, ouvrez le dataflow dans Enterprise Designer, puis sélectionnez **Modifier > Notifications**.

9. Cliquez sur **OK**.

Déclenchement d'un flux avec un fichier de contrôle

Un flux peut être exécuté automatiquement lorsqu'un fichier de contrôle est détecté dans un répertoire surveillé. Cette fonctionnalité s'avère utile dans les situations où le flux doit attendre la fin d'un autre processus avant d'être exécuté. Par exemple, vous pouvez disposer d'un flux qui a besoin d'un fichier d'entrée généré par un autre processus d'entreprise. Vous pouvez configurer l'autre processus pour qu'il place un fichier de contrôle dans un dossier, et configurer Spectrum™ Technology Platform pour qu'il exécute un flux lorsque ce fichier de contrôle apparaît.

Remarque : Assurez-vous que le fichier de contrôle est placé dans le dossier surveillé uniquement après que tous les fichiers requis par le flux sont en place et prêts pour le traitement.

1. Si ce n'est pas déjà fait, exposez le flux.

Pour exposer un flux, ouvrez-le dans Enterprise Designer et sélectionnez **Fichier > Exposer/Ne plus exposer et Enregistrer**.

2. Ouvrez Management Console.
3. Accédez à **Flux > Calendriers**.
4. Cliquez sur le bouton Ajouter .
5. Dans le champ **Nom**, saisissez le nom à attribuer à ce calendrier. Il s'agit du nom qui sera affiché dans la liste des calendriers.
6. Dans le champ **Flux**, saisissez le job ou le process flow que vous souhaitez exécuter. Seuls les jobs et processflow que vous avez enregistrés et exposés sont disponibles ici.
7. Une fois que vous avez spécifié un flux, des champs supplémentaires apparaissent sous le champ **Flux**, un champ pour chacun des stages source (tels que Read from File) et collecteurs du flux (tel que Write to File). Ces champs affichent les fichiers qui seront utilisés lors de l'exécution du flux par ce calendrier. Par défaut, le flux utilise les fichiers spécifiés dans ses sources et dans ses récepteurs. Vous pouvez spécifier des fichiers différents à utiliser lorsque ce calendrier s'exécute en remplaçant le chemin d'accès au fichier par le chemin d'accès vers un autre fichier. Par exemple, si votre flux comporte un stage Read from File qui lit les données de `C:\FlowInput\Customers.csv`, mais que vous souhaitez utiliser les données de `C:\FlowInput\UpdatedCustomers.csv` lorsque ce calendrier est exécuté, vous devez indiquer `C:\FlowInput\UpdatedCustomers.csv` dans le champ Read from File.

Remarque : Pour modifier les fichiers utilisés dans les stages source et collecteur, vous devez disposer d'autorisations de lecture pour le type d'entité sécurisée **Ressources - Serveurs de fichiers**.

Notez que lorsqu'un flux est déclenché par un calendrier, les fichiers utilisés par un flux doivent se trouver sur le serveur Spectrum™ Technology Platform ou sur un serveur de fichiers défini comme une ressource externe dans Management Console. Ceci s'applique tant aux jobs qu'aux activités de job dans un processflow. Si un stage source ou récepteur fait référence à un fichier sur un ordinateur client uniquement, exécutez l'une des procédures suivantes :

Option	Description
Option 1 : modifier le dataflow	<p>Déplacez le fichier dans le serveur Spectrum™ Technology Platform ou le serveur de fichiers, puis modifiez le dataflow :</p> <ol style="list-style-type: none"> 1. Ouvrez le dataflow dans Enterprise Designer. 2. Double-cliquez sur le stage source ou récepteur. 3. Dans le champ File name, cliquez sur le bouton naviguer. 4. Cliquez sur Ordinateur distant puis sélectionnez le fichier que vous voulez.

Remarque : Si vous exécutez Enterprise Designer sur le même ordinateur que le serveur Spectrum™ Technology

Option	Description
	Platform, un clic sur l'Ordinateur distant aura apparemment le même effet qu'un clic sur Mon ordinateur. Cependant, vous devez sélectionner le fichier à l'aide de l'icône Machine Distante afin que le système reconnaisse que le fichier appartient bien au serveur Spectrum™ Technology Platform.
Option 2 : remplacez l'emplacement du fichier de dataflow quand le calendrier s'exécute.	Vous pouvez remplacer les références de fichiers contenues dans le flux lorsque ce calendrier s'exécute. Pour ce faire, remplacez le fichier par défaut indiqué dans chaque champ source et collecteur par un chemin d'accès à un fichier sur le serveur Spectrum™ Technology Platform ou une ressource de serveur de fichiers défini dans Management Console.

8. Dans le champ **Déclencheur**, choisissez **Fichier de contrôle**.
9. Dans le champ **Fichier de contrôle**, indiquez le chemin complet et le nom du fichier de contrôle qui déclenchera le flux. Vous pouvez spécifier un nom de fichier exact ou vous pouvez utiliser l'astérisque (*) en tant que caractère générique. Par exemple, *.trg déclenche le flux lorsqu'un fichier avec une extension .trg apparaît dans le dossier.

La présence d'un fichier de contrôle indique que tous les fichiers requis pour le flux sont en place et prêts à être utilisés dans le flux.

Le fichier de contrôle peut être un fichier vide. Pour les jobs, le fichier de contrôle peut indiquer des substitutions aux chemins de fichier configurés dans les stages Write to File ou Read from File. Pour utiliser un fichier de contrôle pour remplacer les chemins de fichier, spécifiez les noms de stage Read from File ou Write from File avec le fichier d'entrée ou de sortie comme derniers arguments, comme suit :

```
stagename=filename
```

Par exemple :

```
Read\ from\ File=file:C:/myfile_input.txt
Write\ to\ File=file:C:/myfile_output.txt
```

Le nom du stage indiqué dans le fichier de contrôle doit correspondre au libellé du stage montré sous l'icône de stage dans le flux de données. Par exemple, si le stage d'entrée est étiqueté « Read From File » vous devez spécifier :

```
Read\ From\ File=file:C:/inputfile.txt
```

Si le stage d'entrée est étiqueté « Clients de l'Illinois », vous devez spécifier :

```
Illinois\ Customers=file:C:/inputfile.txt
```

Lorsque vous remplacez un emplacement Read from File ou Write to File, veillez à suivre les directives suivantes :

- Démarrez le chemin par le protocole « file: ». Par exemple, sous Windows, spécifiez "file:C:/myfile.txt", et, sous Unix ou Linux, spécifiez "file:/testfiles/myfile.txt".
- Le contenu du fichier doit utiliser un codage de caractères compatible avec la norme ISO-8559-1 (Latin-1) ASCII.
- Vous devez utiliser des barres obliques (/) pour les chemins d'accès, et non des barres obliques inversées.
- Les espaces dans les noms de stage doivent être échappés avec une barre oblique inverse.
- Les noms de stage sont sensibles à la casse.

Remarque : S'il existe plusieurs calendriers qui utilisent un déclencheur de fichier de contrôle, il est important qu'ils surveillent chacun des fichiers de contrôle différents. Sinon, le même fichier de contrôle peut déclencher plusieurs jobs ou processflow et générer ainsi un comportement inattendu. À des fins organisationnelles, nous recommandons de placer tous les fichiers requis et le fichier de contrôle dans un répertoire dédié.

10. Dans le champ **Intervalle de sondage** indiquez la fréquence de vérification de la présence du fichier de contrôle. Par exemple, si vous indiquez 10, l'analyseur recherche toutes les 10 secondes pour voir si le fichier de contrôle est dans le dossier surveillé.

La valeur par défaut est de 60 secondes.

11. Dans la **dossier de job** champ, indiquez un dossier dans lequel se trouve le fichier de contrôle temporairement lors de l'exécution du flux. Spectrum™ Technology Platform copie le fichier dans le dossier surveillé dans le dossier de job avant d'exécuter le flux. Cela supprime le dossier surveillé, ce qui empêche au flux d'être à nouveau renvoyé par le même fichier de contrôle.
12. Dans le champ **Options du dossier de job**, précisez ce qu'il faut faire avec les fichiers dans le dossier de job lorsque le flux termine l'exécution.

Conserver Laisse les fichiers à leur emplacement actuel avec leur nom actuel. Si vous sélectionnez cette option, les fichiers du dossier de job peuvent être écrasés à chaque exécution de ce calendrier.

Déplacer vers Déplace les fichiers du dossier de job vers un dossier de votre choix. Cela vous permet de conserver les fichiers qui se trouvaient dans le dossier de job en les déplaçant vers un autre emplacement afin qu'ils ne sont pas remplacés la prochaine fois que le moniteur exécute. Vous pouvez également utiliser cette option pour déplacer les fichiers dans un autre dossier surveillé pour déclencher un traitement en aval, comme un autre dataflow ou un autre processus.

Renommer avec l'horodatage Ajoute un horodatage au nom du fichier dans le dossier de job. Cela vous permet de conserver une copie des fichiers dans le dossier de job dans la mesure où le fichier renommé aura un nom unique et ne sera donc pas être écrasé la prochaine fois que le moniteur exécutera un flux de données.

Supprimer Supprime les fichiers du dossier de job après la fin de l'exécution du flux .

13. Si le dataflow est configuré pour la notification par courrier électronique, vous pouvez spécifier les destinataires supplémentaires pour les notifications qui seront envoyées lorsque le dataflow s'exécute. Les destinataires que vous indiquez ici recevront des notifications en plus des destinataires spécifiés dans les paramètres de notification du flux. Pour configurer un flux pour qu'il envoie des notifications, ouvrez le flux dans Enterprise Designer et accédez à **Modifier > Notifications**.
14. Cliquez sur **Enregistrer**.

Exemple : Surveillé dossier et le dossier de job

Imaginons que vous avez un atelier de carrosserie. Chaque jour que vous souhaitez envoyer aux clients du jour précédent un bon pour une remise sur vos service futurs. Pour ce faire, vous disposez d'un dataflow qui prend la liste de clients pour la journée, vérifie que les noms ont la casse correcte et valide les adresses. La liste de clients du jour est générée par un autre système chaque soir. Cet autre système génère un fichier contenant la liste des clients, et vous souhaitez utiliser ce fichier comme entrée pour le dataflow.

Le système qui génère la liste des clients les dépose dans un dossier nommé `DailyCustomerReport`. Il place également un fichier déclencheur vide dans le dossier lorsqu'il est terminé. Vous pouvez donc configurer Spectrum™ Technology Platform pour gérer ce dossier, en indiquant les éléments suivants pour le fichier déclencheur :

```
C:\DailyCustomerReport\*.trg
```

Cela indique à Spectrum™ Technology Platform d'exécuter le dataflow chaque fois qu'un fichier avec une extension `.trg` apparaît dans ce dossier. Vous pouvez également préciser un nom de fichier spécifique, mais dans cet exemple, nous utilisons un caractère générique.

Lorsqu'un fichier `.trg` est détecté dans le dossier `DailyCustomerReport`, Spectrum™ Technology Platform doit le déplacer vers un autre dossier avant d'exécuter le dataflow. Le fichier doit être déplacé car sinon, il serait détecté à nouveau lors du prochain intervalle de sondage, cela se traduit par la nouvelle exécution du dataflow. Par conséquent, le fichier est déplacé dans un « dossier de job » dans lequel il se trouve au cours de l'exécution du flux de données. Vous choisissez le dossier de job `C:\SpectrumWorkingFolder`.

Une fois que le dataflow a terminé le traitement de la liste des clients, vous souhaitez que le fichier déclencheur soit déplacé vers un autre emplacement dans lequel il

déclenchera un autre processus de facturation. Par conséquent, vous sélectionnez l'option **Atteindre** et choisissez un dossier nommé `C:\DailyBilling`.

Donc dans cet exemple, le fichier déclencheur commence dans `C:\DailyCustomerReport` et est ensuite déplacé vers le dossier de travail `C:\SpectrumWorkingFolder`. Une fois que le dataflow est terminé, le fichier déclencheur est déplacé vers `C:\DailyBilling` pour lancer le processus de facturation.

Affichage de l'état et de l'historique du flux

Vous pouvez afficher un historique du job, du flux de processus et de l'exécution du service dans Management Console et Enterprise Designer.

Dans Management Console

Pour afficher l'état et l'historique du flux dans Management Console, sélectionnez **Flux > Historique**. L'onglet **flux** affiche l'historique des flux de processus et des jobs et l'onglet **Transactions** affiche l'historique des services.

Remarque : Pour l'historique de flux, le nombre d'enregistrements affiché lorsque vous survolez la colonne **Résultats** reflète le nombre total d'enregistrements écrits en sortie par tous les collecteurs de flux de données. Ce nombre peut être différent du nombre d'enregistrements en entrée si le flux de données combine des enregistrements, divise des enregistrements ou crée de nouveaux enregistrements.

Par défaut, l'historique des transactions est désactivé, car l'activation de l'historique des transactions peut avoir un impact négatif sur les performances. Si vous souhaitez consulter l'historique des transactions, vous devez activer la journalisation de l'historique des transactions en cliquant sur l'interrupteur **Journalisation des transactions**. Pour afficher l'activité de l'utilisateur, utilisez le journal d'audit auquel vous pouvez accéder sous **Système > Journaux**.

La liste d'historique de flux est automatiquement mise à jour toutes les 30 secondes. Si vous souhaitez la mettre à jour plus tôt, cliquez sur le bouton Actualiser .

Dans Enterprise Designer

Pour afficher l'état et l'historique du flux dans Enterprise Designer, sélectionnez **Vue > Historique d'exécution**.

La liste d'historique de flux est automatiquement mise à jour toutes les 30 secondes. Si l'affichage de l'historique d'exécution est lent, décochez la case **Actualiser automatiquement**.

L'onglet **Jobs** sert à suivre le statut d'un job et à mettre en pause, reprendre ou annuler des jobs en cours ainsi qu'à supprimer des jobs terminés.

Remarque : Le nombre d'enregistrements affiché dans l'onglet **Jobs** reflète le nombre total d'enregistrements écrits en sortie par tous les récepteurs de flux de données. Ce nombre peut être différent du nombre d'enregistrements en entrée si le flux de données combine des enregistrements, divise des enregistrements ou crée de nouveaux enregistrements.

- La colonne **Succès** indique le nombre total d'enregistrements écrits en sortie par tous les récepteurs de flux de données présentant une valeur vierge dans le champ Statut.
- La colonne **Échec** indique le nombre total d'enregistrements écrits en sortie par les récepteurs du flux de données présentant une valeur F dans le champ Statut.
- La colonne **Non conformes** indique le nombre total d'enregistrements sortant de tous les ports d'erreur de stage source.

L'onglet **Processflow** sert à suivre l'état des processflow et à annuler les processflow en cours ainsi qu'à supprimer les processflow terminés. Si vous cliquez sur le signe plus à côté de tout processflow, vous aurez accès aux informations du Statut de l'activité du processflow. Les informations suivantes sont comprises dans cette zone :

ActivityName	Comprend les noms de toutes les activités, y compris les activités ayant réussi, qui composent le processflow.
State	État de l'activité (échec, réussi, en cours, annulé).
ReturnCode	Code indiquant le résultat du processflow. L'un des éléments suivants :
1	Le processflow a échoué.
0	Le processflow s'est correctement terminé.
-1	Le processflow a été annulé.
Autres chiffres	Si le processflow contient une activité Run Program, le programme externe peut renvoyer des codes qui lui sont propres. Toutes les valeurs de la colonne ReturnCode différentes de -1, 0 et 1 proviennent du programme externe. Pour obtenir des explications sur les codes de renvoi du programme externe, consultez sa documentation.
Started	Date et heure de début de l'activité.
Finished	Date et heure de fin de l'activité.
Commentaires	Tout commentaire associé à l'activité.

Téléchargement de l'historique de flux

Vous pouvez télécharger les informations affichées sur la page Historique de Management Console vers un fichier Excel.

1. Ouvrez Management Console.
2. Accédez à **Flux > Historique**.
3. Pour télécharger les informations d'historique pour les services, cliquez sur **Historique des transactions**. Pour télécharger l'historique des jobs et flux de processus, laissez l'onglet **Flux** sélectionné.

4. Cliquez sur le bouton Télécharger .

Conseil : Si vous souhaitez ne télécharger que certaines entrées dans la liste d'historique, modifiez les paramètres de filtre pour afficher uniquement l'historique que vous souhaitez télécharger.

Définition des paramètres par défaut des enregistrements non conformes

Un enregistrement non conforme est un enregistrement d'entrée que Spectrum™ Technology Platform ne parvient pas à analyser. Par défaut, si les données d'entrée d'un job contiennent un enregistrement non conforme, le job se termine. Vous pouvez modifier ce paramètre afin d'autoriser davantage d'enregistrements d'entrée non conformes, voire même autoriser un nombre illimité d'enregistrements non conformes. Cette procédure explique comment définir le seuil par défaut des enregistrements non conformes des jobs sur votre système.

Remarque : Vous pouvez écraser le seuil par défaut des enregistrements non conformes d'un job en ouvrant le job dans Enterprise Designer, puis en accédant à **Modifier > Options de job**.

1. Ouvrez Management Console.
2. Accédez à **Flux > Paramètres par défaut**.
3. Cliquez sur **Enregistrements non conformes**.
4. Sélectionnez l'une des options suivantes :

Terminer les jobs contenant ce nombre d'enregistrements non conformes

Sélectionnez cette option pour terminer les jobs si les données d'entrée contiennent un ou plusieurs enregistrements non conformes. Saisissez le nombre d'enregistrements non conformes devant déclencher la fin du job. La valeur par défaut est 1.

Ne pas terminer les flux ayant des enregistrements non conformes

Sélectionnez cette option pour autoriser un nombre illimité d'enregistrements non conformes dans les données d'entrée.

Configuration des paramètres de rapport par défaut

Les rapports sont générés par les jobs contenant un stage Report. Les rapports peuvent inclure des synthèses de traitement, comme le nombre d'enregistrements traités par le job ou des formats postaux comme USPS CASS 3553. Certains modules fournissent des rapports prédéfinis. Vous pouvez également créer des rapports personnalisés. La configuration des paramètres par défaut des rapports établit les paramètres par défaut utilisés pour enregistrer les rapports. Les paramètres

par défaut peuvent être modifiés pour un job ou pour un rapport donné au sein d'un job, via Enterprise Designer.

La procédure suivante explique comment définir les options de reporting par défaut de votre système.

1. Ouvrez Management Console.
2. Accédez à **Flux > Paramètres par défaut**.
3. Cliquez sur **Rapports**.
4. Sélectionnez le format à utiliser pour enregistrer les rapports. Les rapports peuvent être enregistrés au format HTML, PDF ou texte.
5. Sélectionnez l'emplacement d'enregistrement des rapports.

Enregistrer les rapports dans l'historique de tâches Enregistre les rapports sur le serveur dans le cadre de l'historique des jobs. Cela permet aux utilisateurs de Management Console et d'Enterprise Designer d'afficher facilement des rapports, dans la mesure où ces derniers sont disponibles dans l'historique d'exécution.

Enregistrer les rapports dans un fichier Enregistre les rapports dans un fichier à l'emplacement de votre choix. Ceci s'avère utile si vous souhaitez partager des rapports avec des personnes qui ne sont pas des utilisateurs Spectrum™ Technology Platform. Cette option s'avère également utile si vous souhaitez créer une archive de rapports à un emplacement différent. Pour afficher les rapports ainsi enregistrés, vous pouvez utiliser tout outil capable d'ouvrir le format de rapport, comme une visionneuse PDF pour les rapports PDF ou un navigateur Web pour les rapports HTML.

6. Si vous avez sélectionné **Enregistrer les rapports dans un fichier**, renseignez ces champs.

Emplacement des rapports Dossier dans lequel vous souhaitez enregistrer les rapports.

Ajouter au nom du rapport Spécifie les informations de variable à inclure dans le nom du fichier. Vous pouvez sélectionner une ou plusieurs des options suivantes :

ID de job ID unique attribué à l'exécution d'un job. La première fois que vous exécutez un job sur votre système, son ID est 1. La deuxième fois que vous exécutez un job, qu'il s'agisse du même job ou d'un autre, son ID est 2, etc.

Stage Nom du stage ayant contribué aux données du rapport, tel que spécifié dans le stage Report d'Enterprise Designer.

Date Jour, mois et année de création du rapport.

Écraser les rapports existants Remplace les rapports précédents portant le même nom de fichier par le nouveau rapport. Si vous ne sélectionnez pas cette option et qu'il existe un rapport portant le même nom que le nouveau rapport, le job se termine correctement, mais le nouveau rapport n'est pas enregistré.

Un commentaire apparaît dans l'historique d'exécution, indiquant que le rapport n'a pas été enregistré.

Exposition d'un service

Exposition d'un service sous forme de service Web

Les services Spectrum™ Technology Platform peuvent être mis à disposition sous forme de services Web REST et/ou SOAP. Pour mettre un service à disposition sous forme de service Web sur votre serveur, procédez comme suit :

1. Ouvrez Enterprise Designer.
2. Ouvrez le service que vous souhaitez exposer sous forme de service Web.
3. Accédez à **Modifier > Options de services Web**.
4. Pour rendre le service disponible sous forme de service Web SOAP, cochez la case **Exposer sous forme de service Web SOAP**.
5. Pour rendre le service disponible sous forme de service Web REST, cochez la case **Exposer sous forme de service Web REST** et procédez comme suit.
 - a) Si vous souhaitez remplacer la destination par défaut, indiquez la destination que vous souhaitez utiliser dans le champ **Chemin d'accès**.

Le chemin d'accès est facultatif. Par défaut, la destination d'un service Web REST est la suivante :

```
http://server:port/rest/service_name/results.qualifier
```

Si vous souhaitez utiliser une autre destination, le chemin d'accès que vous indiquez est ajouté après le nom de service. Par exemple, si vous spécifiez *Amériques/Expédition* dans le champ **Chemin d'accès**, votre destination JSON sera semblable à ce qui suit :

```
http://myserver:8080/rest/MyService/Americas/Shipping/results.json
```

Vous pouvez utiliser des champs et des options du dataflow comme noms de variable dans le chemin d'accès en cliquant sur le menu déroulant **Insérer variable** et en sélectionnant le champ ou l'option que vous souhaitez utiliser. La variable est représentée dans le chemin d'accès à l'aide de la notation `${Option.Name}` pour les options de dataflow et de la notation `${Data.Name}` pour les champs de dataflow.

- b) Par défaut, les services Web REST prennent en charge la méthode GET et renvoient des données aux formats XML et JSON. Vous pouvez définir des méthodes HTTP et des formats

de sortie supplémentaires en cliquant sur **Ajouter** pour ajouter une ressource au service Web.

Quand vous ajoutez une ressource, vous pouvez sélectionner la méthode HTTP (**GET** ou **POST**). Les formats de données pris en charge sont répertoriés ci-dessous. Il se peut que vous ne disposiez pas de tous ces formats, car certains formats ne sont disponibles que si certains modules sont installés sur votre serveur Spectrum™ Technology Platform.

XML Format XML par défaut. Utilisez ce format si vous souhaitez utiliser XML comme format dans les requêtes et les réponses et qu'il n'existe aucun format XML spécialisé pour le type de données que vous souhaitez traiter.

JSON Format JSON par défaut. Utilisez ce format si vous souhaitez utiliser JSON comme format dans les requêtes et les réponses et qu'il n'existe aucun format JSON spécialisé pour le type de données que vous souhaitez traiter.

GeoJSON Format JSON spécialisé approprié aux services qui gèrent des données géographiques. La prise en charge est garantie uniquement pour Géométrie et pour les types de plate-forme natifs suivants :

- booléen
- double
- flottant
- integer
- bigdecimal
- long
- string
- date
- time
- datetime
- timespan

Si vous tentez d'exposer un flux avec tout autre type, vous ne pouvez pas spécifier GeoJSON (une erreur s'affiche lors de la conception). En outre, GeoJSON accepte une seule géométrie. Si la sortie contient plusieurs champs de géométrie, le système recherche un champ appelé « geometry » suivi d'un champ appelé « obj ». Si ces champs n'existent pas, le premier champ de géométrie est sélectionné.

c) Cliquez sur **OK**.

La nouvelle ressource est ajoutée au service Web.

6. Cliquez sur **OK** lorsque vous avez terminé de configurer les options de service Web.
7. Cliquez sur l'ampoule grise de la barre d'outils pour exposer le service.

Lorsqu'un dataflow est exposé, le bouton de l'ampoule dans la barre d'outils Enterprise Designer indique que le dataflow est exposé comme indiqué ici :



Pour vérifier que le service est maintenant exposé sous forme de service Web, accédez à l'une des URL suivantes :

- Pour REST : `http://server:port/rest`
- Pour SOAP : `http://server:port/soap`

Où *serveur* est le nom de l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* le port utilisé pour la communication HTTP.

Exposition d'un service pour l'accès à l'API

Pour qu'un service soit accessible via l'API Spectrum™ Technology Platform, vous devez l'exposer.

1. Ouvrez le service dans Enterprise Designer.
2. Cliquez sur l'ampoule grise de la barre d'outils pour exposer le service.

Lorsqu'un dataflow est exposé, le bouton de l'ampoule dans la barre d'outils Enterprise Designer indique que le dataflow est exposé comme indiqué ici :



Options de Runtime

Ajout d'options d'exécution de flux de données

Les options d'exécution de flux de données vous permettent de contrôler le comportement des stages lorsque vous exécutez le flux de données. Ceci s'avère utile lorsque vous souhaitez avoir la possibilité de modifier le comportement du flux de données lors de son exécution. Par exemple, vous pouvez souhaiter indiquer une base de données source pour un stage Read from DB lorsque vous exécutez le flux de données, au lieu d'utiliser la base de données indiquée dans le stage Read from DB du flux de données.

Cette procédure explique comment exposer des options qui peuvent être définies lors de l'exécution. Après avoir effectué cette procédure, vous pourrez définir des options de flux de données lors de l'exécution à l'aide de ces techniques :

- Pour les jobs, vous pourrez indiquer des options d'exécution à l'aide d'un fichier de propriétés d'options de flux de données et de l'argument `-o` de Job Executor.
- Pour les services, vous pourrez indiquer des options d'exécution sous forme d'options de l'API.
- Pour les services exposés sous forme de service Web, vous pourrez indiquer des options d'exécution sous forme de paramètres dans la requête.
- Pour les sous-flux, les options d'exécution seront héritées du flux de données parent et exposées via l'une des méthodes ci-dessus, selon le type de flux de données parent (job, service ou service exposé sous forme de service Web).

Pour ajouter des options d'exécution à un flux de données :

1. Ouvrez le flux de données dans Enterprise Designer.
2. Pour configurer les options d'exécution d'un stage dans un flux de données intégré, ouvrez ce dernier.
3. Cliquez sur l'icône Options de flux de données de la barre d'outils ou sur **Édition > Options de flux de données**. La boîte de dialogue **Options de flux de données** apparaît.
4. Cliquez sur **Ajouter**. La boîte de dialogue **Définir l'option de flux des données** apparaît.
5. Dans le champ **Option name**, indiquez le nom à utiliser pour cette option. Il s'agit du nom de l'option qui devra être spécifié lors de l'exécution afin de définir cette option.
6. Dans le champ **Label**, vous pouvez spécifier un autre libellé ou conserver le même en tant que nom d'option.
7. Saisissez une description de l'option dans le champ **Description**.
8. Dans le champ **Target**, déterminez si cette option doit être appliquée à tous les stages dans le flux de données ou à certains stages uniquement.

Stage(s) sélectionné(s)

Sélectionnez cette option pour qu'elle ne soit appliquée qu'aux stages que vous spécifiez.

Tous les stages

Sélectionnez cette option pour qu'elle soit appliquée à tous les stages dans le flux de données.

Inclut les transformations

Sélectionnez cette option pour que l'option d'exécution soit mise à disposition pour les transformations personnalisées dans les stages Transformer dans le flux de données. Si vous choisissez cette option, vous pouvez accéder à la valeur indiquée lors de l'exécution dans le script Groovy d'une transformation personnalisée à l'aide de la syntaxe suivante :

```
options.get("optionName")
```

Par exemple, pour accéder à une option nommée `casing`, vous devez inclure cela dans votre script de transformation personnalisée :

```
options.get("casing")
```

- Si vous choisissez **Selected stage(s)** dans le champ **Target**, la table **Map dataflow options to stages** affiche la liste des stages contenus dans le flux de données. Sélectionnez l'option à exposer en tant qu'option de flux de données. Vous verrez les champs **Valeur par défaut** et **Valeurs conformes** se remplir de données lorsque vous sélectionnez votre premier élément.

Remarque : Vous pouvez sélectionner plusieurs options de telle sorte que l'option de flux de données contrôle plusieurs options de stages. Si vous faites cela, chaque option que vous sélectionnez doit partager des valeurs conformes. Par exemple, une option présente des valeurs de Y et N, chacune des options supplémentaires doit disposer de Y ou de N dans son ensemble de valeurs, et vous ne pouvez autoriser la valeur en commun à être disponible qu'au moment de l'exécution. Ainsi, si vous sélectionnez une option avec les valeurs Y et N, vous ne pouvez pas sélectionner une option dotée des valeurs E, T, M et L, mais vous pouvez sélectionner une option avec les valeurs P, S et N car les deux options partagent la valeur « N ». Cependant, seule cette valeur « N » est disponible pour cette option, pas « Y », « P » ou « S ».

- Si vous voulez limiter les valeurs qu'un utilisateur peut sélectionner, modifiez les options disponibles dans le champ **Legal values** en cliquant sur l'icône située directement à droite de ce champ.
- Pour modifier la valeur par défaut, indiquez une valeur différente dans le champ **Valeur par défaut**.

Remarque : Pour un service, vous ne pouvez modifier que les valeurs par défaut avant d'exposer le service pour la première fois. Une fois que vous exposez le service, vous ne pouvez plus modifier les valeurs par défaut à l'aide d'Enterprise Designer. Au lieu de cela, vous devez utiliser Management Console. Pour plus d'informations, reportez-vous à la section [Spécification des Options de services par défaut](#) à la page 181.

- Cliquez sur **OK**.
- Continuez d'ajouter les options souhaitées.
- Cliquez sur **OK** dans la boîte de dialogue des **Options du flux de données** lorsque vous avez terminé d'ajouter vos options.
- Si vous ajoutez une option d'exécution à un flux de données intégré, vous devez définir le flux de données parent de l'option d'exécution, ainsi que tous les flux de données ancêtres, pour que les options soient disponibles lors de l'exécution. Pour ce faire, ouvrez le flux de données contenant le flux de données intégré et exposez l'option que vous venez de créer. Si nécessaire, ouvrez le parent de ce flux de données et définissez-y l'option, etc., jusqu'à ce que l'option de flux de données soit définie pour tous les ancêtres.

Par exemple, imaginons que vous disposiez d'un flux de données nommé « A » contenant un flux de données intégré nommé « B », contenant à son tour un flux de données intégré nommé

« C », de sorte que votre hiérarchie de flux de données intégrés ressemble à : A > B > C. Si vous souhaitez exposer une option nommée Casing dans un stage dans le flux de données intégré « C », vous devez ouvrir le flux de données intégré « C » et y définir l'option. Ensuite, vous devez ouvrir le flux de données intégré « B » et y définir l'option. Pour finir, vous devez ouvrir le flux de données « A » et y définir l'option, pour qu'elle soit disponible lors de l'exécution.

Le flux de données est désormais configuré pour permettre la spécification des options lors de l'exécution.

Spécification des Options de services par défaut

Les options de services par défaut contrôlent le comportement par défaut de chaque service sur votre système. Vous pouvez spécifier une valeur par défaut pour chaque option dans un service. Les paramètres des options par défaut prennent effet quand un appel d'API ou une requête de service Web ne définit pas de manière explicite de valeur pour une option donnée. Ces options de service par défaut sont aussi utilisées par défaut lorsque vous créez un dataflow dans Enterprise Designer à l'aide de ce service.

Remarque : Pour un service, vous ne pouvez modifier que les valeurs par défaut avant d'exposer le service pour la première fois. Une fois que vous exposez le service, vous ne pouvez plus modifier les valeurs par défaut à l'aide d'Enterprise Designer. Au lieu de cela, vous devez utiliser Management Console.

1. Ouvrez Management Console.
2. Cliquez sur **Services**.
3. Cochez la case en regard du service de votre choix, puis cliquez sur le bouton Modifier .
4. Définissez les options du service. Pour plus d'informations sur les options du service, reportez-vous au guide des solutions du module du service.
5. Cliquez sur **Enregistrer**.

Suppression d'options d'exécution de flux de données

Les options d'exécution de flux de données vous permettent de définir des options de stage lors de l'exécution. Les options des stages peuvent être définies lors de l'appel du job via un processflow ou via l'outil de ligne de commande Job Executor. Cette procédure explique comment supprimer une option d'exécution de flux de données de sorte que l'option ne puisse plus être définie lors de l'exécution.

1. Ouvrez le job, service ou sous-flux.
2. Cliquez sur l'icône **Options de flux de données** ou cliquez sur **Édition > Options de flux de données**. La boîte de dialogue **Options de flux de données** apparaît.

3. Surlignez l'option que vous voulez supprimer et cliquez sur **Supprimer**.

Configuration des notifications par courrier électronique pour un flux de données

Un flux de données peut être configuré pour envoyer une notification par courrier électronique contenant des informations sur le statut du job. Par exemple, vous pouvez souhaiter envoyer une notification par courrier électronique en cas d'échec d'un flux de données. La notification par courrier électronique peut contenir des informations telles que le nom du flux, l'adresse de début et heure de fin, le nombre d'enregistrements traités et plus.

Remarque : Un serveur de courrier doit être configuré dans Management Console avant de pouvoir configurer la notification d'un flux de données. Pour plus d'informations, voir le *Guide d'administration Spectrum™ Technology Platform*.

1. Avec un flux de données ou un processflow ouvert dans Enterprise Designer, sélectionnez **Édition > Notifications**.
2. Cliquez sur **Ajouter**.
3. Dans le champ **Send Notification To**, saisissez l'adresse électronique à laquelle les courriers électroniques doivent être envoyés.
4. Sélectionnez les événements à propos desquels vous voulez recevoir une notification.
5. Dans le champ **Subject**, entrez le texte que vous voulez voir apparaître dans la ligne objet du courrier électronique.
6. Dans le champ **Message**, saisissez le texte que vous voulez voir apparaître dans le corps du courrier électronique.

Vous pouvez choisir d'inclure des informations sur le job dans le courrier électronique en cliquant sur **Cliquer ici pour insérer un champ dans l'objet ou le message**. Voici quelques exemples d'informations de job : heure de début, heure de fin et nombre d'enregistrements ayant échoué.

7. Cliquez sur **Aperçu** si vous souhaitez visualiser la notification.
8. Cliquez sur **OK**. La boîte de dialogue **Notifications** réapparaît avec la nouvelle notification.
9. Cliquez sur **OK**.

5 - Combinaison de flux en un process flow

In this section

Introduction aux processflow	184
Conception des processflow	184

Introduction aux processflow

Un processflow exécute une série d'activités, comme des jobs et des applications externes. Chaque activité du flux de processus s'exécute à la fin de l'activité précédente. Les processflow s'avèrent utiles si vous souhaitez exécuter plusieurs flux de données en séquence ou un programme externe. Par exemple, un processflow peut exécuter un job de normalisation des noms, valider des adresses, puis utiliser une application externe pour trier les enregistrements dans le bon ordre pour l'obtention de réductions sur les tarifs postaux. Un tel processflow ressemble à ceci :



Dans cet exemple, les jobs Standardize Names et Validate Addresses sont des jobs exposés sur le serveur Spectrum™ Technology Platform. Exécuter le programme appelle une application externe, et une activité Réussi indique la fin du processflow.

Conception des processflow

Création d'un processflow

Pour créer un processflow, utilisez Enterprise Designer pour créer une séquence d'activités exécutant des jobs et/ou des applications externes.

1. Ouvrez Enterprise Designer.
2. Sélectionnez **Fichier > Nouveau > Processflow**.
3. Ajoutez la première action que vous souhaitez que le processflow réalise. Vous pouvez effectuer l'une des opérations suivantes :
 - Pour exécuter un job, faites glisser l'icône du job du dossier **Activités** de la palette jusqu'au canevas.
 - Pour exécuter un programme externe, faites glisser l'icône Exécuter le programme du dossier **Activités** de la palette jusqu'au canevas.
4. Ajoutez la deuxième action que vous souhaitez que le processflow réalise.

Vous pouvez ajouter un job en faisant glisser l'icône du job jusqu'au canevas, ou ajouter un programme externe en faisant glisser l'icône Exécuter le programme jusqu'au canevas.

- Reliez les deux activités en cliquant sur le triangle gris à droite de la première icône, puis en la faisant glisser pour la relier à l'icône grise à gauche de la deuxième icône.

Par exemple, si votre processflow commence par exécuter un job nommé GeocodeAddress, puis un programme externe, votre processflow prend la forme suivante :



- Ajoutez des activités supplémentaires selon les besoins.
- Une fois que vous avez ajouté toutes les activités que vous souhaitez exécuter dans le processflow, faites glisser une activité Réussi jusqu'au canevas et reliez-la à la dernière activité du processflow.

Par exemple, ce processflow contient deux jobs (« Standardize Names » et « Validate Addresses ») et une activité Exécuter le programme. À la fin de ce processflow figure l'activité Réussi :



- Double-cliquez sur les activités que vous avez placées sur le canevas pour configurer les options d'exécution. Vous pouvez également double-cliquer sur la connexion entre les activités pour configurer les options de transition.

Utilisation d'une variable pour référencer un fichier

Dans un processflow, les variables s'avèrent utiles si vous souhaitez que plusieurs activités du processflow référencent le même fichier. À l'aide d'une variable, vous pouvez définir le fichier à un emplacement, puis référencer la variable dans toutes les activités en aval devant référencer le fichier. Si le fichier change, vous pouvez modifier la définition de variable sans avoir à modifier toutes les activités en aval.

Lorsque vous ajoutez une activité de job à un processflow, l'activité crée automatiquement les variables de chaque source et de chaque collecteur de données dans le flux de données. Si, dans le flux de données, vous souhaitez utiliser des fichiers non définis dans la source ou le collecteur de données d'un job, vous pouvez créer des variables.

Lorsque vous ajoutez une activité Exécuter le programme, aucune variable n'est créée par défaut. Si vous souhaitez utiliser des variables avec une activité Exécuter le programme, vous devez les créer.

Cette procédure explique comment créer une variable dans une activité de job ou dans une activité Exécuter le programme.

- Ouvrez le processflow dans Enterprise Designer.
- Double-cliquez sur l'activité de job ou l'activité Exécuter le programme dans laquelle vous souhaitez définir la variable.

Remarque : Les variables peuvent uniquement être référencées par les activités suivant l'activité dans laquelle vous avez défini la variable. Par conséquent, veillez à définir la variable dans une activité qui précède les activités dans lesquelles vous souhaitez l'utiliser.

3. Cliquez sur l'onglet **Variables**.
4. Créez la variable.

Option	Description
Pour créer une nouvelle variable pour un fichier d'entrée...	En regard de Entrées , cliquez sur Ajouter . Dans le champ Nom , saisissez le nom de la variable. Il s'agit du nom que les activités en aval vont référencer. Dans le champ Emplacement , sélectionnez l'une des options suivantes :
	Utiliser le fichier indiqué dans la tâche Sélectionnez cette option pour utiliser le fichier défini dans le stage source dans le job. Cette option est disponible uniquement si vous définissez une variable pour une activité de job.
	Accéder au fichier sur le serveur Sélectionnez cette option si vous souhaitez sélectionner un fichier à attribuer à cette variable.
Pour créer une nouvelle variable pour un fichier de sortie...	Référencer un fichier d'activité en amont Sélectionnez cette option si vous souhaitez utiliser un fichier attribué à une variable existante d'un stage en amont.
	En regard de Entrées , cliquez sur Ajouter . Dans le champ Nom , saisissez le nom de la variable. Il s'agit du nom que les activités en aval vont référencer. Dans le champ Emplacement , sélectionnez l'une des options suivantes :
	Accéder au fichier sur le serveur Sélectionnez cette option si vous souhaitez sélectionner un fichier à attribuer à cette variable.
	Fichier temporaire géré par le serveur Sélectionnez cette option si vous souhaitez que cette variable fasse référence à un fichier temporaire qui sera automatiquement créé et supprimé au besoin. Cette option s'avère utile dans les cas où un fichier n'est employé que comme étape intermédiaire dans un processflow et n'est plus nécessaire une fois ce processflow est terminé.

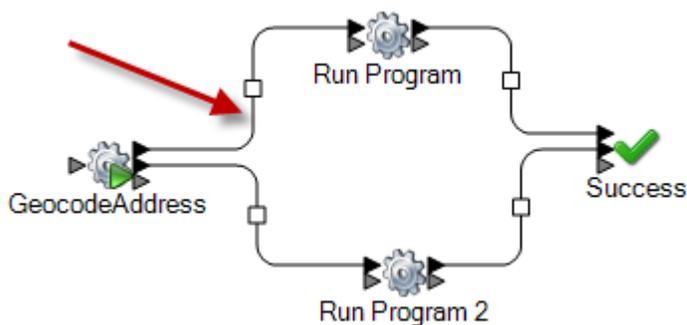
5. Cliquez sur **OK** pour fermer la fenêtre **Ajouter une variable**.
6. Cliquez sur **OK** pour fermer la fenêtre d'options de l'activité.
7. Pour référencer la variable dans une activité en aval :
 - a) Double-cliquez sur l'activité qui doit référencer la variable.
 - b) Sélectionnez le stage d'entrée que vous souhaitez voir référencer à la variable, puis cliquez sur **Modifier**.
 - c) Dans le champ **Location**, sélectionnez Reference an upstream activity's file....

Ajout d'une logique conditionnelle à un processflow

Vous pouvez ajouter une logique conditionnelle à un processflow de sorte que des activités différentes soient exécutées en fonction du code de renvoi d'une activité précédente. Par exemple, vous pouvez exécuter une activité si un job renvoie un code de renvoi 1 et une autre activité si un job renvoie un code de renvoi 0. Vous pouvez ainsi construire un embranchement conditionnel dans votre processflow.

1. Ouvrez le processflow dans Enterprise Designer.
2. Double-cliquez sur la transition entre deux activités du flux.

Une *transition* est la ligne reliant deux activités. Par exemple, la ligne entre l'activité GeocodeAddress et l'activité Exécuter le programme illustrée ici est une transition :



La fenêtre **Options de transition** apparaît.

3. Sélectionnez le type de transition que vous souhaitez ajouter.

Simple	Sélectionnez cette option si vous souhaitez toujours exécuter ce chemin d'accès dans le processflow.
Conditionnel	Sélectionnez cette option si vous souhaitez uniquement exécuter ce chemin d'accès dans le processflow si l'activité en amont renvoie un ou des codes de renvoi spécifiques ou une plage de codes de renvoi.

Sinon Sélectionnez cette option si vous souhaitez exécuter ce chemin d'accès dans le processflow uniquement si les conditions des autres transitions découlant de l'activité ne sont pas remplies.

Remarque : Une seule transition **Autre** peut exister dans les transitions découlant d'une activité.

4. Cliquez sur **OK**.
5. Pour configurer les transitions déclenchant une activité, cliquez avec le bouton droit de la souris sur l'activité, sélectionnez **Modes d'entrée**, puis l'une des options suivantes :

Première entrée La première entrée de cette activité, qu'elle se produise via une transition **Simple**, **Conditionnel** ou **Sinon**, déclenche l'exécution de l'activité. Les autres transitions sont ignorées.

Toutes les entrées L'activité n'est pas exécutée, sauf si toutes les transitions de cette activité sont effectuées.

6. Pour configurer les transitions effectuées découlant d'une activité, cliquez avec le bouton droit de la souris sur l'activité, sélectionnez **Modes de sortie**, puis l'une des options suivantes :

Première sortie La première transition évaluée comme true est effectuée. Les autres transitions sont ignorées, même si leurs conditions sont évaluées sur true.

Toutes les sorties Toutes les transitions évaluées comme true sont effectuées.

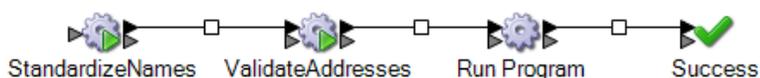
Suppression d'un processflow

1. Allez sur **Fichier > Gérer**. La boîte de dialogue **Gérer** apparaît.
2. Cliquez avec le bouton droit sur le stage que vous voulez supprimer et sélectionnez **Supprimer**.
3. Cliquez sur **OK**.

Activités

Job

L'activité d'un job consiste à exécuter un job dans le cadre d'un processflow. L'exemple suivant illustre un processflow exécutant deux activités de job : Standardize Names et Validate Addresses.



Pour ajouter une activité de job à un processflow, faites glisser l'activité du job du dossier Activités de la palette jusqu'au canevas.

Remarque : Pour qu'un job puisse être utilisé dans un processflow, il doit être exposé. Si le job de votre choix n'est pas exposé, ouvrez le job dans Enterprise Designer et sélectionnez **Fichier > Exposer**.

Double-cliquez sur l'activité du job pour configurer l'onglet **Options** et l'onglet **Variables**.

Onglet Options

L'onglet Options affiche les options d'exécution disponibles pour le job. Vous pouvez modifier les options d'exécution de sorte que, lorsque le job est exécuté dans ce processflow, les options que vous définissez ici sont utilisées pour le job. Par exemple, si l'une des options du flux de données du job permet de contrôler les unités à utiliser pour la distance et que cette option est définie par défaut sur miles, vous pouvez l'écraser ici afin de renvoyer la distance en kilomètres lorsque le job est exécuté via ce processflow.

Onglet Variables

L'onglet Variables affiche les stages sources et de collecteurs de données dans le flux de données. Vous pouvez écraser les fichiers d'entrée et de sortie indiqués dans le flux de données, de sorte que, lors de l'exécution du job dans le processflow, des fichiers d'entrée et/ou de sortie différents soient utilisés.

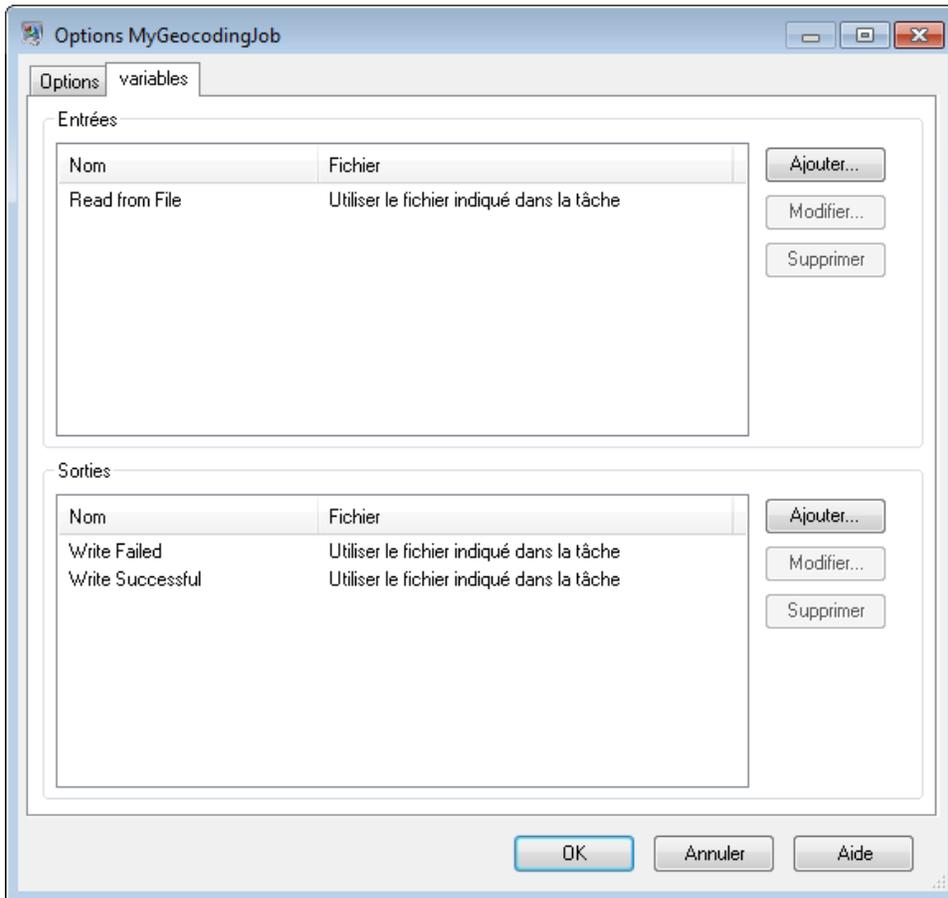
Écrasement des fichiers d'entrée et de sortie

Par défaut, un job exécuté dans un processflow utilise les fichiers d'entrée et de sortie définis dans les stages sources et de collecteurs de données du job. Cependant, vous pouvez écraser les fichiers d'entrée et de sortie définis dans le job de sorte que, lors de l'exécution du processflow, le job utilise les fichiers d'entrée et de sortie que vous indiquez dans l'activité du job dans le processflow, au lieu des fichiers indiqués dans les stages sources et de collecteurs de données du job. Vous pouvez écraser les fichiers d'entrée et de sortie en indiquant un fichier spécifique ou en utilisant une variable pour faire référence à un fichier défini dans une activité en amont.

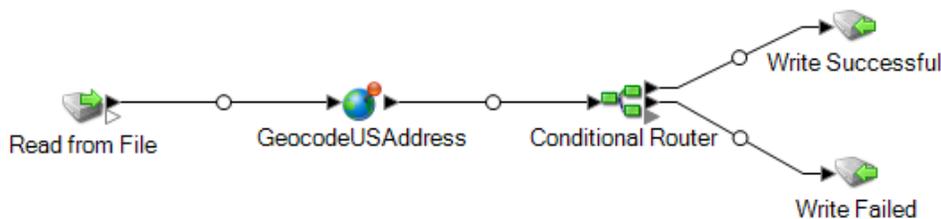
1. Ouvrez le processflow dans Enterprise Designer.
2. Double-cliquez sur l'activité du job dont vous souhaitez écraser un fichier d'entrée ou de sortie.
3. Cliquez sur l'onglet **Variables**.

Cliquez sur l'onglet **Variables**. Les variables répertoriées sous **Entrées** correspondent aux stages sources du job. Les variables répertoriées sous **Sorties** correspondent aux stages de collecteurs de données du job.

Par exemple, imaginons que vous disposiez d'un processflow contenant une activité de job d'un job nommé MyGeocodingJob. Dans l'onglet **Variables** des options de l'activité figurent les éléments suivants :



Chaque variable répertoriée correspond au nom d'un stage source ou de collecteur de données du flux de données MyGeocodingJob. Dans cet exemple, l'onglet **Variables** affiche une source (Read from File) et deux collecteurs de données (Write Failed et Write Successful). Si vous ouvriez le flux de données MyGeocodingJob dans Enterprise Designer, vous verriez apparaître les éléments suivants :



4. Dans l'onglet **Variables**, sélectionnez le stage source ou le stage de sortie que vous souhaitez écraser, puis cliquez sur **Modifier**.
5. Réalisez l'une des actions suivantes :

Option	Description
Pour écraser un fichier d'entrée...	Dans le champ Emplacement , sélectionnez l'une des options suivantes :
	Utiliser le fichier indiqué dans la tâche Sélectionnez cette option pour utiliser le fichier défini dans le stage source dans le job.
	Accéder au fichier sur le serveur Sélectionnez cette option pour écraser le fichier défini dans le job et utiliser un autre fichier de votre choix.
Référencer un fichier d'activité en amont	Sélectionnez cette option si vous souhaitez remplacer le fichier défini dans le job par un fichier dont le nom et l'emplacement sont définis dans le stage Read from File ou Write to File d'une activité en amont ou dans la variable d'une activité en amont. Utilisez cette option si le fichier de sortie d'une activité précédente correspond à l'entrée de cette activité. L'avantage de cette option réside dans le fait que si le stage Write to File d'une activité en amont est modifié pour pointer vers un autre fichier, cette activité continue à pointer vers le fichier correct. Un autre avantage réside dans le fait que vous n'avez pas besoin de connaître le chemin d'accès au fichier ni le nom du fichier d'entrée des activités en amont pour pointer vers lui.
Pour écraser un fichier de sortie...	Dans le champ Emplacement , sélectionnez l'une des options suivantes :
	Accéder au fichier sur le serveur Sélectionnez cette option pour écraser le fichier défini dans le job et utiliser un autre fichier de votre choix.
	Fichier temporaire géré par le serveur Sélectionnez cette option si vous souhaitez que cette variable fasse référence à un fichier temporaire qui sera automatiquement créé et supprimé au besoin. Cette option s'avère utile dans les cas où un fichier n'est employé que comme étape intermédiaire dans un processflow et n'est plus nécessaire une fois ce processflow est terminé.

6. Cliquez sur **OK** pour fermer la fenêtre **Modifier la variable**.
7. Cliquez sur **OK** pour fermer la fenêtre d'options de l'activité.

Lors de l'exécution du processflow, le job utilise les fichiers que vous avez indiqués dans l'activité du processflow au lieu des fichiers indiqués dans le job lui-même.

Clear Cache

Une activité Clear Cache efface les données du cache global dans le cadre d'un processflow. Elle ne supprime pas le cache ; elle efface uniquement les données du cache.

1. Faites glisser l'activité **Clear cache** jusqu'au canevas.
2. Double-cliquez sur l'activité Effacer le cache.
3. Sélectionnez le cache. Vous pouvez également sélectionner plusieurs caches dont vous souhaitez effacer les données.
4. Exécutez le processflow.

Execute SQL

Une activité Execute SQL vous permet d'exécuter des instructions SQL avant et après l'exécution d'un flux de données ou d'un programme externe.

Load to Hive

Apache Hive est une infrastructure d'entrepôt de données construite par-dessus Hadoop pour fournir des fonctionnalités de synthèse, de requête et d'analyse. Pour utiliser Hive pour interroger la source de données sous-jacente, utilisez son propre langage de requête, HiveQL.

Hive prend en charge les formats de fichier Hadoop ci-dessous :

- TEXTFILE
- SEQUENCE FILE
- ORC
- RCFILE
- PARQUET
- AVRO

Remarque : Le format de fichier AVRO est pris en charge dans la version Hive 0.14 et les versions supérieures.

L'activité **Load to Hive** vous permet de charger des données dans une table Hive à l'aide d'une connexion JDBC. Grâce à cette connexion, les données sont lues depuis un fichier Hadoop spécifié et chargées soit dans une table existante de la connexion sélectionnée, soit dans une nouvelle table créée dans la connexion sélectionnée.

Pour charger les données dans une nouvelle table, il faut que le schéma de la table soit défini.

Remarque : Spectrum ne prend pas en charge les données hiérarchiques, même si Hive les prend en charge.

Création d'une connexion Hive

1. Ouvrez l'activité **Load to Hive**.

2. Dans le champ **Nom de fichier**, saisissez le nom du fichier à lire. Cliquez sur Parcourir [...] pour sélectionner le fichier à lire.
3. Dans le champ **Type de fichier**, saisissez le format du fichier à lire. Le format de fichier par défaut sélectionné est `Delimited`.

Si le **Type de fichier** sélectionné est `Delimited` ou `Sequence`, les champs **Séparateur de champ** et **Séparateur d'enregistrement** sont affichés. Sinon, ils ne sont pas affichés.

4. Dans le champ **Séparateur de champ**, sélectionnez le caractère de séparation de chaque champ consécutif d'un enregistrement.
5. Sélectionnez la connexion de la base de données Hive que vous souhaitez utiliser dans le champ **Connexion**.
 - a) Pour ajouter, modifier et supprimer des connexions, cliquez sur **Gérer**.
La fenêtre **Gestionnaire de connexion à la base de données** s'ouvre.
 - b) Cliquez sur **Ajouter** pour créer une connexion ou sur **Modifier** pour modifier une connexion existante.
La fenêtre **Propriétés de la connexion** s'ouvre.
 - c) Saisissez le **Nom de la connexion**.
 - d) Dans le champ **Pilote de base de données**, sélectionnez un pilote de base de données Hive pour la connexion.
 - e) Indiquez tous les détails de la connexion, à savoir **Utilisateur**, **Mot de passe**, **Hôte**, **Port** et **Instance**.
 - f) Pour tester les détails de la connexion, cliquez sur **Tester**.
 - g) Si le test de connexion réussit, cliquez sur **OK**.
La fenêtre **Propriétés de la connexion** se ferme.
 - h) Cliquez sur **OK**.
La fenêtre **Gestionnaire de connexion à la base de données** se ferme.
6. Dans le champ **Table/Vue**, sélectionnez la table dans laquelle vous souhaitez écrire, ou saisissez le nom d'une nouvelle table à créer.
Si vous créez une nouvelle table dans le champ **Table/Vue**, la case **Externe** est cochée. Sinon, si vous sélectionnez une table existante, la case **Externe** reste décochée.
7. Pour créer la nouvelle table externe à la base de données Hive, cochez la case **Externe**.

Important : En cas de tables externes :

1. Les enregistrements existants ne peuvent pas être écrasés, et les nouveaux enregistrements ne peuvent pas être ajoutés. Vous êtes uniquement autorisé à créer de nouvelles tables externes et à les renseigner à l'aide des enregistrements.
2. Si vous sélectionnez un fichier placé dans un dossier donné, tous les fichiers placés dans ce dossier sont automatiquement sélectionnés. Par conséquent, assurez-vous que tous les fichiers placés dans le dossier donné présentent le même format.

Pour en savoir plus sur les tables EXTERNES, cliquez [ici](#).

8. Pour remplacer tous les enregistrements existants de la table, cochez la case **Écraser**. Cela supprime les enregistrements existants de la table sélectionnée et ajoute que les enregistrements lus du fichier à la table.
9. La grille affiche les noms et les types de données des colonnes de la table sélectionnée. Si vous avez spécifié une nouvelle table dans le champ **Table/Vue**, utilisez les boutons **Ajouter**, **Modifier** et **Supprimer** en regard de la grille pour ajouter des colonnes pour définir la table et spécifier leurs types de données respectifs. Utilisez les boutons **Déplacer vers le haut** et **Déplacer vers le bas** pour spécifier la séquence des colonnes de la table.

Remarque : Les boutons **Ajouter**, **Modifier**, **Supprimer**, **Déplacer vers le haut** et **Déplacer vers le bas** restent désactivés si vous sélectionnez une table existante dans le champ **Table/Vue**.

Important :

1. Vérifiez que les types de données de tous les champs du fichier correspondent aux types de données des colonnes de table respectives, sauf si tous les types de données sont de type Chaîne. Sinon, le chargement de données peut entraîner des données incohérentes.
 2. Assurez-vous que le nombre de champs du fichier correspond au nombre de colonnes de la table. Sinon, les données des champs supplémentaires du fichier sont ignorées.
 3. Hive accepte les noms de table et de colonne en minuscules uniquement. Si vous saisissez les noms en lettres majuscules, Hive les convertit en minuscules. Le schéma qui en résulte affiche tous les noms en minuscules.
10. Cliquez sur **OK**.

Remarque : Si vous décidez de créer une table et de définir ses colonnes, cette table est créée lors de l'exécution. L'activité **Load to Hive** est uniquement prévue pour concevoir la structure de la table. Lors de l'exécution, la table conçue est créée et les données lues du fichier y sont écrites.

Run Hadoop MapReduce Job

L'activité **Run Hadoop MapReduce Job** vous permet d'exécuter tout job MapReduce sur un cluster Hadoop, en mettant en correspondance le fichier JAR approprié. Vous pouvez utiliser cette activité pour exécuter un job MapReduce du SDK Spectrum™ Data Quality for Big Data ou tout job MapReduce externe.

Remarque : Si le job MapReduce échoue, un message d'erreur s'affiche avec l'état d'exécution du job.

Pour obtenir une liste des configurations de job prises en charge sur les plates-formes Windows et Linux, consultez le tableau **Configurations de job prises en charge** à la page 200.

Champ	Description
Serveur Hadoop	Liste des serveurs Hadoop configurés. Pour des informations sur le mappage des serveurs de fichiers HDFS via Management Console, reportez-vous au <i>Guide d'administration</i> .
Chemin d'accès Jar	Chemin d'accès au fichier JAR approprié pour l'exécution du job MapReduce Hadoop. Remarque : Le fichier JAR doit être présent à l'emplacement client externe ou sur le serveur Spectrum. Il ne doit pas être placé sur le cluster Hadoop.
Classe de pilote	Sélectionnez l'un des éléments suivants : Valeur par défaut Pour exécuter un job externe en saisissant simplement le <i>nom de classe</i> et les <i>arguments</i> du job, sélectionnez <i>Par défaut</i> . Si <i>Par défaut</i> est sélectionné, les champs Nom de classe et Arguments sont affichés. Configurer Pour saisir des propriétés de job supplémentaires pour tout job externe ou pour exécuter l'un des jobs du SDK Qualité des Big Data Spectrum, sélectionnez <i>Configurer</i> . Lors de la sélection de <i>Configurer</i> , le champ Type de job s'affiche.
Type de job	Sélectionnez l'un des éléments suivants : Spectrum Pour exécuter l'un des jobs du SDK Qualité des Big Data Spectrum, sélectionnez <i>Spectrum</i> . Lors de la sélection de <i>Spectrum</i> , le champ Jobs Spectrum s'affiche. Générique Pour spécifier des propriétés supplémentaires pour tout job externe, sélectionnez <i>Générique</i> .

Champ	Description
Jobs Spectrum	<p>Sélectionnez le job requis dans la liste des jobs du SDK Qualité des Big Data Spectrum. La liste comprend ces jobs :</p> <ul style="list-style-type: none"> • Validation d'adresse • Advanced Transformer • Best of Breed • Duplicate Synchronization • Filtre • Groovy • Intraflow Match • Interflow Match • Joiner • Match Key Generator • Open Name Parser • Open Parser • Table Lookup • Transactional Match • Validate Address • Validate Address Global <p>Lors de la sélection du job Spectrum de votre choix :</p> <ol style="list-style-type: none"> 1. Les champs Nom du job, Nom de classe et Arguments sont automatiquement renseignés. <p>Tous les champs automatiquement renseignés peuvent être modifiés, si nécessaire, sauf le champ Nom de classe.</p> <p>Important : Pour le job Spectrum sélectionné, le champ Nom de classe automatiquement renseigné ne doit pas être modifié, sinon le job ne peut pas être exécuté.</p> 2. La grille Propriétés est automatiquement renseignée à l'aide des propriétés de configuration requises du job Spectrum sélectionné, avec leurs valeurs par défaut. <p>Vous pouvez ajouter ou importer davantage de propriétés ainsi que modifier les propriétés automatiquement renseignées, si nécessaire.</p>
Nom de classe	Nom complet de la classe de pilote du job.

Champ	Description
Arguments	<p>Liste des arguments séparés par des espaces. Ils sont transmis à la classe du pilote lors de l'exécution du job.</p> <p>Par exemple,</p> <pre>23Dec2016 /home/Hadoop/EYInc.txt</pre> <ol style="list-style-type: none"> <p>Ces variables peuvent être transmises comme arguments dans la liste d'arguments, qui sont définis de sorte à accepter des valeurs d'exécution dans le stage source ou dans ce stage en cours du process flow.</p> <p>Par exemple, si, dans la sortie du stage précédent du process flow, la variable <code>SalesStartRange</code> est définie, vous pouvez inclure cette variable dans cette liste d'arguments comme <code>\${SalesStartRange}</code> ainsi que d'autres arguments obligatoires, comme suit :</p> <pre>23Dec2016 /home/Hadoop/EYInc.txt \${SalesStartRange}</pre> <p>Dans le cas où un argument donné contient un espace, encadrez-le à l'aide de guillemets doubles.</p> <p>Il peut s'agir par exemple de <code>"/home/Hadoop/Sales Records"</code>.</p> <p>Jobs du SDK Qualité des Big Data Spectrum - Arguments :</p> <p>Pour exécuter les jobs <i>MapReduce</i> du SDK Qualité des Big Data Spectrum, transmettez les différents fichiers de configuration sous forme de liste d'arguments. Chaque clé d'argument accepte le chemin d'accès à un seul fichier de propriétés de configuration, où chaque fichier contient plusieurs propriétés de configuration.</p> <p>La syntaxe de la liste d'arguments des propriétés de configuration est la suivante :</p> <pre>[-config <Path to configuration file>] [-debug] [-input <Path to input configuration file>] [-conf <Path to MapReduce configuration file>] [-output <Path of output directory>]</pre> <p>Par exemple, pour un job <i>MapReduce MatchKeyGenerator</i> :</p> <pre>-config /home/hadoop/matchkey/mkgConfig.xml -input /home/hadoop/matchkey/inputFileConfig.xml -conf /home/hadoop/matchkey/mapReduceConfig.xml -output /home/hadoop/matchkey/outputFileConfig.xml</pre> <p>Remarque : Si la même clé de propriété de configuration est spécifiée dans le champ Arguments et dans la grille Propriétés, mais que chacune pointe vers des fichiers de configuration différents, le fichier indiqué dans la grille Propriétés est celui retenu.</p> <p>Les exemples de propriétés de configuration sont livrés avec le SDK Data and Address Quality for Big Data et placés sous <code><Big Data Quality bundle>\samples\configuration</code>.</p>

Onglet Général

Champ	Description	Exigence
Nom du job	Nom du job MapReduce Hadoop.	Obligatoire
Chemin d'accès d'entrée	Chemin d'accès au fichier d'entrée du job.	Obligatoire
Chemin d'accès de sortie	Chemin d'accès au fichier de sortie du job.	Obligatoire
Remplacer la sortie	Indique si le chemin d'accès de sortie spécifié doit être remplacé au cas où il existe déjà. Remarque : Si cette case est laissée décochée et si le système constate que le chemin d'accès de sortie configuré existe lors de l'exécution, Hadoop lève une exception et le process flow est abandonné.	Facultatif
Classe Mapper	Nom complet de la classe qui gère les fonctionnalités Mapper pour le job.	Obligatoire
Classe Reducer	Nom complet de la classe qui gère les fonctionnalités Reducer pour le job.	Facultatif
Classe Combiner	Nom complet de la classe qui gère les fonctionnalités Combiner pour le job.	Facultatif
Classe Partitioner	Nom complet de la classe qui gère les fonctionnalités Partitioner pour le job.	Facultatif
Nombre de réducteurs	Nombre de réducteurs utilisés pour exécuter le job MapReduce.	Facultatif
Format d'entrée	Format des données d'entrée.	Obligatoire
Format de sortie	Format des données de sortie.	Obligatoire
Classe de clé de sortie	Type de données des clés des paires clé-valeur de sortie.	Obligatoire
Classe de valeur de sortie	Type de données des valeurs des paires clé-valeur de sortie.	Obligatoire

Onglet Propriétés

Pour spécifier des propriétés supplémentaires pour l'exécution du job requis, utilisez cet onglet pour définir autant de paires propriété-valeur que nécessaire. Vous pouvez ajouter les propriétés requises directement dans la grille une par une.

Sinon, pour importer les propriétés d'un fichier, cliquez sur **Importer**. Accédez à l'emplacement du fichier de propriétés respectif et sélectionnez le fichier au format XML. Les propriétés contenues dans le fichier importé sont copiées dans la grille. Le fichier de propriétés doit être au format XML et doit suivre la syntaxe :

```
<configuration>
  <property>
    <name>key</name>
    <value>some_value</value>
    <description>A brief description of the
      purpose of the property key.</description>
  </property>
</configuration>
```

Vous pouvez importer directement le fichier de propriétés Hadoop *mapred.xml*, ou créer vos propres fichiers à ce format XML.

Remarque :

1. Si la même propriété est définie ici et dans Management Console, les valeurs définies ici remplacent celles définies dans Management Console.
2. Si la même propriété existe à la fois dans la grille et dans le fichier de propriétés importé, la valeur importée du fichier remplace la valeur existante dans la grille pour la même propriété.
3. Vous pouvez importer plusieurs fichiers de propriétés l'un après l'autre, si nécessaire. Les propriétés contenues dans chaque fichier importé sont ajoutées à la grille.
4. Assurez-vous que le fichier de propriétés est présent sur le serveur Spectrum™ Technology Platform.
5. La balise `<description>` est facultative pour chaque clé de propriété d'un fichier de propriétés de configuration.
6. Les données de référence doivent être placées localement sur les nœuds de données pour que les jobs pertinents puissent être exécutés. Cette propriété n'est disponible que pour les jobs qui utilisent des données de référence, tels que *Advanced Transformer*, *Validate Address Global* et *Validate Address*. La propriété est : `pb.bdq.reference.data.location`.

Onglet Dépendances

Dans cet onglet, ajoutez la liste des fichiers d'entrée et des fichiers Jar nécessaires pour exécuter le job.

Une fois que le job est exécuté, les fichiers de référence et les fichiers de référence Jar ajoutés ici sont disponibles dans le cache distribué du job.

Fichiers de référence

Pour ajouter les différents fichiers requis comme entrée pour exécuter le job, cliquez sur **Ajouter**, accédez à l'emplacement respectif sur votre système ou cluster local et sélectionnez le fichier en question.

Pour supprimer tout fichier ajouté à la liste, sélectionnez le fichier en question et cliquez sur **Supprimer**.

Fichiers JAR de référence

Pour ajouter les fichiers Jar requis pour exécuter le job, cliquez sur **Ajouter**, accédez à l'emplacement respectif sur votre système ou cluster local et sélectionnez le fichier Jar en question.

Pour supprimer tout fichier ajouté à la liste, sélectionnez le fichier en question et cliquez sur **Supprimer**.

Configurations de job prises en charge

Ce tableau répertorie les job pris en charge en modes par défaut et configuré sur les plates-formes Windows et Linux :

stages	Windows		Linux	
	Mode par défaut	Mode configuré	Mode par défaut	Mode configuré
MatchKeyGenerator	Pris en charge	Pris en charge	Pris en charge	Pris en charge
Intraflow Match	Pris en charge	Pris en charge	Pris en charge	Pris en charge
Interflow Match	Pris en charge	Pris en charge	Pris en charge	Pris en charge
Transactional Match	Pris en charge	Pris en charge	Pris en charge	Pris en charge
Table Lookup	Pris en charge	Pris en charge	Pris en charge	Pris en charge
Advanced Transformer	Pris en charge	Pris en charge	Pris en charge	Pris en charge
Open Name Parser	Pris en charge	Pris en charge	Pris en charge	Pris en charge
Duplicate Synchronization	Pris en charge	Pris en charge	Pris en charge	Pris en charge
Filtre	Pris en charge	Pris en charge	Pris en charge	Pris en charge
Best of Breed	Pris en charge	Pris en charge	Pris en charge	Pris en charge

stages	Windows		Linux	
	Mode par défaut	Mode configuré	Mode par défaut	Mode configuré
Open Parser	Pris en charge	Pris en charge	Pris en charge	Pris en charge
Address Validation (GAV)				
Groovy	Pris en charge	Pris en charge	Pris en charge	Pris en charge
Joiner	Pris en charge	Pris en charge	Pris en charge	Pris en charge
Validate Address Global (AD)	Pris en charge	Pris en charge	Pris en charge	Pris en charge
Validate Address (C1P)			Pris en charge	Pris en charge

Remarque :

- Les stages GAV et C1P ne sont pas pris en charge sur la plate-forme Windows sur laquelle le répertoire d'installation de Spectrum contient des espaces.
- Le comportement des stages GAV et C1P doit être vérifié de nouveau sur la plate-forme Windows sur laquelle le répertoire d'installation de Spectrum ne contient aucun espace.
- Les données de référence sur la stratégie HDFS ne sont pas prises en charge.

Variables

Entrées

Dans cette grille, parmi les champs reçus d'une activité source, sélectionnez les champs à utiliser dans cette activité.

Sorties

Dans cette grille, sélectionnez les champs à inclure dans la sortie de l'activité de destination dans ce flux de processus.

Run Hadoop Pig

Run Hadoop Pig exécute un script Apache Pig. Apache Pig est un langage de haut niveau qui permet d'exprimer des programmes d'analyse de données. Il dispose de l'infrastructure nécessaire à l'évaluation de ces programmes. Tous les programmes Pig peuvent être mis en parallèle, ce qui leur permet de gérer de très grands jeux de données.

Run Hadoop Pig vous permet de sélectionner des opérations Pig, de saisir tous les paramètres requis et de faire générer automatiquement votre script Pig par le système. Vous pouvez exécuter le script Pig sur n'importe quel serveur Hadoop.

Run Hadoop Pig fonctionne uniquement sur les serveurs de fichiers Hadoop. Apache Hadoop 1.x et 2.x sont tous les deux pris en charge.

Pour définir les options de **Run Hadoop Pig** :

1. Glissez et déposez l'activité **Run Hadoop Pig** sur le canevas.
2. Faites un clic droit sur l'activité **Run Hadoop Pig** et sélectionnez **Options**.
3. Le champ du nom de serveur indique le serveur Hadoop sur lequel se trouve le fichier à traiter.
4. Cliquez sur le bouton de navigation (**[...]**) pour accéder au fichier à traiter.
5. Sélectionnez le type de fichier. Run Hadoop Pig prend en charge les fichiers au format Delimited et au format Delimited Sequence.
6. Sélectionnez le délimiteur et le caractère de protection appropriés.
7. Cliquez sur **Ajouter** dans la section Champs et ajoutez les champs présents dans le fichier à traiter. Pour les fichiers Sequence, le premier champ est considéré comme la clé et les autres champs font partie des valeurs délimitées.
8. Sélectionnez l'opération **Raccourcir**, le cas échéant. L'opération Raccourcir coupe les espaces blancs du champ d'entrée, avant de le traiter.
9. Accédez à l'onglet Opérations. Cliquez sur **Ajouter** pour commencer à ajouter les opérations Pig à effectuer sur le fichier. Cette opération ouvre l'éditeur d'opérations.
10. Sélectionnez une opération à effectuer. Les différentes opérations sont les suivantes :
 - Tri : trie les données dans l'ordre alphabétique.
 - Filtre : filtre les données selon vos besoins.
 - Agrégat : effectue des opérations statistiques, telles que Somme et Compte, sur les données.
 - Distinct : sélectionne tous les enregistrements uniques du champ spécifié.
 - Limite : limite le nombre d'enregistrements traités à un nombre spécifié.
11. Utilisez les boutons **Déplacer vers le haut** et **Déplacer vers le bas** pour modifier l'ordre des opérations.
12. Une fois que vous avez sélectionné les opérations et saisi l'entrée requise pour le traitement des opérations, cliquez sur **Ajouter** pour enregistrer votre sélection et revenir à l'éditeur d'options Pig.
13. Le script Pig est automatiquement généré en fonction des opérations sélectionnées.
 - L'éditeur vous permet de remplacer le script Pig généré par votre propre script, selon les besoins. Cliquez sur l'option **Modifier un script** et saisissez votre propre script dans la zone de texte Script Pig. Dans ce cas, le bouton **Régénérer** est activé. Si vous souhaitez revenir au script généré par le système, cliquez sur **Régénérer** dans la section Script Pig pour générer le script Pig.
 - Pour valider la syntaxe du script, cliquez sur le bouton **Valider**.

14. Vous pouvez spécifier le fichier de sortie sous l'onglet **Variables**. Le fichier de sortie peut être utilisé par les activités ultérieures.
15. Cliquez sur **OK** pour enregistrer le script Pig. Par défaut, le type de fichier de sortie est le même que le type de fichier d'entrée. Vous pouvez le modifier à l'aide du script Pig généré.

Opérations Hadoop Pig

Les différentes opérations Pig sont les suivantes :

1. Tri : trie les données dans l'ordre alphabétique. L'opération de tri est décrite en détail à la section **Tri des enregistrements d'entrée**.
2. Filtre : vous permet de filtrer les données selon vos besoins. L'opération de filtre est décrite en détail à la section **Filtrage des enregistrements d'entrée**.
3. Agrégat : vous permet d'effectuer des opérations statistiques comme Somme, Compte et autres, sur les données.

Sélectionnez les opérations d'agrégation pour chaque champ de votre choix.

- Somme : calcule la somme des valeurs dans le champ.
- Moyenne : calcule la moyenne de toutes les valeurs dans le champ.
- Maxi. : calcule la valeur maximale des valeurs dans le champ.
- Mini. : calcule la valeur minimale des valeurs dans le champ.
- Compte : calcule le nombre total de valeurs dans le champ.

Remarque : Si vous sélectionnez l'opération Distinct, seules les valeurs uniques sont comptées.

4. Distinct : si vous sélectionnez cette option, l'opération Compte Agrégat compte uniquement les valeurs uniques dans le champ.
5. Limite : saisissez une valeur supérieure à zéro pour limiter le nombre d'enregistrements traités à cette valeur.

Exécuter le programme

L'activité Exécuter le programme exécute une application externe dans le cadre d'un processflow.

Tableau 9 : Options d'Exécuter le programme

Nom de l'option	Description
Nom du programme	Le chemin vers l'exécutable que vous voulez lancer.

Nom de l'option	Description
Arguments	Indique les arguments de ligne de commande à transmettre au programme indiqué dans le champ Program name . Séparez les arguments par des espaces. Vous pouvez avoir recours à des variables définies dans l'onglet Variables comme arguments en cliquant sur Insérer variable . Pour plus d'informations sur les variables, voir Utilisation d'une variable pour référencer un fichier à la page 185.
Délai d'attente (en secondes)	Indique le délai de réponse du programme spécifié dans le champ Program name . Si le programme ne répond pas dans le nombre de secondes indiqué, le processflow échoue.
Variables d'environnement	<p>Indique les valeurs des variables d'environnement à utiliser lors de l'exécution d'un programme. Si vous indiquez des valeurs à cet endroit, le programme utilisera ces variables d'environnement au lieu de celles spécifiées sur votre système. Sinon, les variables d'environnement indiquées sur votre système seront utilisées. Remarquez que si le programme appelé emploie plusieurs variables d'environnement, vous devez définir des valeurs pour toutes ou pour aucune de ces variables. Si vous indiquez des valeurs ici, les définitions de variable d'environnement sur votre système ne seront pas modifiées.</p> <p>Cliquez sur Add et saisissez le nom de la variable dans le champ Variable Name. Par exemple, « JAVA_HOME ». Entrez la valeur de la variable dans le champ Variable Value. Par exemple : « C:\Program Files\Java\jdk1.6.0_17. » Au lieu de saisir une valeur, vous pouvez cliquer sur Insérer variable pour la définir sur la valeur d'une variable définie dans l'onglet Variables.</p>

Indication de fichiers d'entrée et de sortie

Vous pouvez appeler une application externe depuis un processflow à l'aide de l'activité Exécuter le programme. Vous pouvez indiquer le fichier contenant les données que vous souhaitez envoyer à l'application externe ainsi que le fichier de sortie dans lequel vous souhaitez que l'application externe écrive.

1. Dans un processflow, double-cliquez sur une activité Exécuter le programme.
2. Cliquez sur l'onglet **Variables**.
3. Pour indiquer un fichier d'entrée...
 - a) Cliquez sur le bouton **Ajouter** dans la section **Entrées**.

- b) Dans le champ **Nom**, saisissez un nom significatif pour ce fichier. Ce nom est entièrement de votre choix.
 - c) Dans le champ **Emplacement**, sélectionnez l'une des options suivantes :

Accéder au fichier sur le serveur	Sélectionnez cette option si vous souhaitez naviguer jusqu'au fichier d'entrée de votre choix pour le sélectionner.
Référencer un fichier d'activité en amont	Sélectionnez cette option si vous souhaitez utiliser un fichier attribué à une variable existante d'un stage en amont.
 - d) Cliquez sur **OK**.
4. Pour indiquer un fichier de sortie...
- a) Cliquez sur le bouton **Ajouter** dans la section **Sorties**.
 - b) Dans le champ **Nom**, saisissez un nom significatif pour ce fichier. Ce nom est entièrement de votre choix.
 - c) Dans le champ **Emplacement**, sélectionnez l'une des options suivantes :

Accéder au fichier sur le serveur	Sélectionnez cette option si vous souhaitez naviguer jusqu'au fichier de sortie de votre choix pour le sélectionner.
Fichier temporaire géré par le serveur	Sélectionnez cette option si vous souhaitez que la sortie du programme soit écrite dans un fichier temporaire qui sera automatiquement créé et supprimé, au besoin. Cette option s'avère utile dans les cas où un fichier n'est employé que comme étape intermédiaire dans un processflow et n'est plus nécessaire une fois ce processflow est terminé.
 - d) Cliquez sur **OK**.
5. Cliquez sur **OK** pour fermer la fenêtre **Options d'Exécuter le programme**.

Utilisation d'un fichier de contrôle avec un programme externe

Vous pouvez appeler une application externe depuis un processflow à l'aide de l'activité Exécuter le programme. Dans ce cas, vous pouvez utiliser un fichier de contrôle contenant les paramètres de configuration de l'application externe. Par exemple, vous pouvez appeler VeriMove™ à l'aide d'une activité Exécuter le programme et indiquer un fichier de contrôle à utiliser lors de l'exécution.

1. Dans un processflow, double-cliquez sur une activité Exécuter le programme.
2. Cliquez sur l'onglet **Variables**.
3. Dans la section **Fichiers de contrôle**, cliquez sur **Ajouter**.
4. Dans le champ **Name**, nommez le fichier de contrôle. Ce nom est entièrement de votre choix.
5. Dans le champ **Contents**, indiquez le contenu du fichier de contrôle.

Pour référencer un fichier à l'aide d'une variable définie dans **Entrées** ou **Sorties**, cliquez sur **Insérer une variable**.

6. Cliquez sur **OK** pour fermer la fenêtre **Ajouter un fichier de contrôle**.
7. Cliquez sur l'onglet **Options**.
8. Cliquez sur **Insérer une variable**.
9. Sélectionnez le nom du fichier de contrôle que vous avez créé, puis cliquez sur **OK**.
Une variable pointant vers votre fichier de contrôle est ajoutée au champ **Arguments**.
10. Si nécessaire, modifiez le champ **Arguments** pour qu'il utilise les arguments de ligne de commande nécessaires pour indiquer un fichier de contrôle. Pour obtenir davantage d'informations, reportez-vous à la documentation sur l'application externe.
11. Cliquez sur **OK** pour fermer la fenêtre **Options d'Exécuter le programme**.

Spark Sorter

L'activité **Spark Sorter** vous permet de trier une énorme quantité d'enregistrements. Cette activité utilise des bibliothèques Apache Spark pour donner de la puissance à la fonction et est exécutée sur votre serveur Spectrum™ Technology Platform.

Actuellement, les fichiers *délimités*, présents sur le serveur Spectrum™ Technology Platform, sont acceptés pour lire les enregistrements d'entrée.

Remarque : Les fichiers présents sur des serveurs distants ne sont pas pris en charge.

Champ	Description
Server name	Indique l'emplacement du fichier que vous sélectionnez comme entrée. Étant donné que l'activité Spark Sorter accepte uniquement les fichiers situés sur Spectrum™ Technology Platform, ce champ affiche Spectrum™ Technology Platform.
Nom de fichier	Indique le chemin du fichier. Cliquez sur le bouton d'ellipse (...) pour accéder au fichier souhaité. Vous pouvez lire plusieurs fichiers en utilisant un caractère générique pour lire les données de plusieurs fichiers du répertoire. Les caractères génériques * et ? sont pris en charge. Par exemple, vous pouvez utiliser *.CSV pour lire tous les fichiers portant une extension .CSV placés dans le répertoire. Pour que plusieurs fichiers puissent être lus, tous doivent présenter la même mise en page (les mêmes champs aux mêmes emplacements). Tout enregistrement ne correspondant pas à la mise en page indiquée dans l'onglet Champs est traité comme un enregistrement non conforme. Avertissement : Si le Spectrum™ Technology Platform est en cours d'exécution sur les systèmes Unix ou Linux, rappelez-vous que les noms de fichiers et les chemins sur ces plates-formes sont sensibles à la casse.

Champ	Description
Type d'enregistrement	<p>Le format des enregistrements dans le fichier. Actuellement, les formats de fichier <i>délimités</i> sont acceptés comme entrée.</p> <p>Délimité Un fichier texte dans lequel les enregistrements sont séparés par un caractère de fin de ligne (EOL) comme un retour chariot ou un retour à la ligne (CR ou LF) et chaque champ est séparé par un caractère désigné comme une virgule.</p>
Chiffrement de caractères	<p>Chiffrement des caractères du fichier d'entrée.</p> <p>Le chiffrement <code>UTF-8</code> est pris en charge. Pour plus d'informations sur UTF, reportez-vous à unicode.org/faq/utf_bom.html.</p>
Séparateur de champs	<p>Indique le caractère utilisé pour séparer des champs dans un fichier délimité.</p> <p>Par exemple, cet enregistrement utilise une barre verticale () comme séparateur de champ :</p> <pre>7200 13TH ST MIAMI FL 33144</pre> <p>Ces caractères disponibles à définir en tant que séparateurs de champs sont :</p> <ul style="list-style-type: none"> • Espace • Tabulation • Virgule • Point (.) • Point-virgule • Conduite <p>Si le fichier utilise un caractère différent comme séparateur de champ, cliquez sur le bouton de sélection pour sélectionner un autre caractère délimiteur.</p>
Qualificateur de texte	<p>Le caractère utilisé pour entourer les valeurs de texte dans un fichier délimité.</p> <p>Par exemple, cet enregistrement utilise des guillemets doubles (") en tant que qualificateur de texte.</p> <pre>"7200 13TH ST" "MIAMI" "FL" "33144"</pre> <p>Ces caractères disponibles à définir en tant que qualificateurs de texte sont :</p> <ul style="list-style-type: none"> • Guillemets simples (') • Guillemets doubles (") <p>Si le fichier utilise un délimiteur de texte, cliquez sur le bouton de sélection pour sélectionner un autre caractère délimiteur.</p>

Champ	Description
Séparateur d'enregistrements	<p>Indique le caractère utilisé pour séparer des enregistrements dans un fichier séquentiel ligne par ligne ou délimité. Ce champ n'est pas disponible si vous activez la case Utiliser EOL par défaut.</p> <p>Les paramètres de séparateur d'enregistrements disponibles sont :</p> <p>Unix (U+000A) Un caractère de saut de ligne sépare les enregistrements. Il s'agit du séparateur d'enregistrement standard pour les systèmes Unix.</p> <p>Macintosh (U+000D) Un caractère de retour chariot sépare les enregistrements. Il s'agit du séparateur d'enregistrement standard pour les systèmes Macintosh.</p> <p>Windows (U+000D U+000A) Un retour chariot suivi d'un saut de ligne sépare les enregistrements. Il s'agit du séparateur d'enregistrement standard pour les systèmes Windows.</p> <p>Si votre fichier utilise un séparateur d'enregistrement, cliquez sur le bouton de sélection pour sélectionner un autre caractère séparateur.</p>
Utiliser EOL par défaut	<p>Indique que le séparateur d'enregistrements du fichier est le caractère de fin de ligne (EOL) par défaut utilisé sur le système d'exploitation sur lequel le serveur Spectrum™ Technology Platform est exécuté.</p> <p>Ne sélectionnez pas cette option si le fichier utilise un caractère EOL qui est différent du caractère EOL par défaut utilisé sur le système d'exploitation du serveur. Par exemple, si le fichier utilise un EOL Windows, mais que le serveur fonctionne sous Linux, ne cochez pas cette option. Au lieu de cela, sélectionnez l'option Windows dans le champ Record separator.</p>
La première ligne est l'enregistrement d'en-tête.	<p>Indique si le premier enregistrement dans un fichier délimité contient des informations d'en-tête, mais aucune donnée.</p> <p>Par exemple, cet extrait de fichier illustre une ligne d'en-tête dans le premier enregistrement.</p> <pre>"AddressLine1" "City" "StateProvince" "PostalCode" "7200 13TH ST" "MIAMI" "FL" "33144" "One Global View" "Troy" "NY" 12180</pre>
Sortie	<p>Indique le chemin d'accès au fichier de sortie du serveur Spectrum™ Technology Platform. Cliquez sur le bouton d'ellipse (...) pour accéder au répertoire de sortie et au nom de fichier de votre choix.</p> <p>Avertissement : Si le Spectrum™ Technology Platform est en cours d'exécution sur les systèmes Unix ou Linux, rappelez-vous que les noms de fichiers et les chemins sur ces plates-formes sont sensibles à la casse.</p>
Écraser	<p>Indique que le fichier de sortie doit effectuer un écrasement s'il existe un fichier portant le même nom tel que spécifié dans le champ Sortie.</p>

Champ	Description
Concaténer	Indique que tous les fichiers de partie Spark doivent être concaténés en un seul fichier de sortie à l'emplacement Sortie spécifié.
Aperçu	<p>Une fois que le fichier d'entrée est sélectionné dans le champ Nom du fichier, la grille Aperçu affiche les 100 premiers enregistrements du fichier de sortie existant.</p> <p>Pour afficher correctement toutes les valeurs de colonne distinctes, cliquez sur Régénérer dans l'onglet Champs.</p>

Onglet Champs

L'onglet **Champs** définit les noms, types et positions des champs du fichier. Pour plus d'informations, voir :

[Définition de champs dans un fichier d'entrée délimité](#) à la page 210

Onglet Tri

L'onglet **Tri** définit les champs permettant de trier les enregistrements d'entrée avant qu'ils ne soient envoyés dans le flux de données. Pour plus d'informations, reportez-vous à la section [Tri des enregistrements](#) à la page 212.

Onglet Configuration

Pour spécifier des propriétés supplémentaires pour l'exécution du job requis, utilisez cet onglet pour définir autant de paires propriété-valeur que nécessaire. Vous pouvez ajouter les propriétés requises directement dans la grille une par une.

Sinon, pour importer les propriétés d'un fichier, cliquez sur **Importer**. Accédez à l'emplacement du fichier de propriétés respectif et sélectionnez le fichier au format XML. Les propriétés contenues dans le fichier importé sont copiées dans la grille. Le fichier de propriétés doit être au format XML et doit suivre la syntaxe :

```
<configuration>
  <property>
    <name>key</name>
    <value>some_value</value>
    <description>A brief description of the
      purpose of the property key.</description>
  </property>
</configuration>
```

Remarque :

1. Si la même propriété est définie ici et dans Management Console, les valeurs définies ici remplacent celles définies dans Management Console.

2. Si la même propriété existe à la fois dans la grille et dans le fichier de propriétés importé, la valeur importée du fichier remplace la valeur existante dans la grille pour la même propriété.
3. Vous pouvez importer plusieurs fichiers de propriétés l'un après l'autre, si nécessaire. Les propriétés contenues dans chaque fichier importé sont ajoutées à la grille.
4. Assurez-vous que le fichier de propriétés est présent sur le serveur Spectrum™ Technology Platform.
5. La balise <description> est facultative pour chaque clé de propriété d'un fichier de propriétés de configuration.
6. Les données de référence doivent être placées localement sur les nœuds de données pour que les jobs pertinents puissent être exécutés. Cette propriété n'est disponible que pour les jobs qui utilisent des données de référence, tels que *Advanced Transformer*, *Validate Address Global* et *Validate Address*. La propriété est : `pb.bdq.reference.data.location`.

L'onglet Exécution

Nom du champ	Description
Nom de fichier	Affiche le nom de fichier sélectionné dans le premier onglet.
Enregistrement de départ	Si vous voulez ignorer les enregistrements au début du fichier lors de la lecture des enregistrements dans le dataflow, spécifiez le premier enregistrement que vous souhaitez lire. Par exemple, si vous voulez sauter les 50 premiers enregistrements dans un fichier, spécifiez 51. Le 51ème enregistrement sera le premier enregistrement lu dans le dataflow.
Tous les enregistrements	Sélectionnez cette option si vous souhaitez lire tous les enregistrements à partir de l'enregistrement spécifié dans le champ Starting record jusqu'à la fin du fichier.
Nombre max d'enregistrements	Sélectionnez cette option si vous souhaitez uniquement lire un certain nombre d'enregistrements à partir de l'enregistrement spécifié dans le champ Starting record . Par exemple, si vous voulez lire les 100 premiers enregistrements, sélectionnez cette option et entrez 100.

Définition de champs dans un fichier d'entrée délimité

L'onglet **Champs** définit les noms, types et positions des champs du fichier. Après avoir défini un fichier d'entrée sur l'onglet **Propriétés de fichier**, vous pouvez définir les champs.

Si le fichier d'entrée ne contient pas d'enregistrement d'en-tête, ou si vous souhaitez définir manuellement les champs, suivez ces étapes dans l'onglet **Champs** :

1. Pour définir les champs déjà présents dans le fichier d'entrée, cliquez sur **Régénérer**. Ensuite, cliquez sur **Détecter le type**. Cela définira automatiquement le type de données pour chaque champ en fonction des 50 premiers enregistrements dans le fichier.
2. Pour ajouter des champs supplémentaires dans la sortie, cliquez sur **Ajouter**.
3. Dans le champ **Name**, choisissez le champ à ajouter ou saisissez le nom du champ.
4. Dans le champ **Type**, vous pouvez laisser le type de données défini sur `string` si vous ne prévoyez pas d'effectuer des opérations mathématiques sur les données. Cependant, si vous avez l'intention d'effectuer de telles opérations, sélectionnez un type de données approprié. Cela convertira les données de chaîne à partir du fichier en un type de données qui activera la manipulation appropriée des données dans le flux de données.

L'activité prend en charge ces types de données :

bigdecimal	Type de données numérique prenant en charge 38 points de précision décimaux. Utilisez ce type de données pour les données utilisées dans les calculs mathématiques requérant un haut niveau de précision, en particulier ceux impliquant des données financières. Le type de données <code>bigdecimal</code> prend en charge des calculs plus précis que le type de données <code>double</code> .
double	Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre 2^{-1074} et $(2-2^{-52}) \times 2^{1023}$. En notation E, la plage de valeurs est comprise entre <code>-1.79769313486232E+308</code> et <code>1.79769313486232E+308</code> .
flottant	Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre 2^{-149} et $(2-2^{-23}) \times 2^{127}$. En notation E, la plage de valeurs est comprise entre <code>-3.402823E+38</code> et <code>3.402823E+38</code> .
integer	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre -2^{31} (<code>-2,147,483,648</code>) et $2^{31}-1$ (<code>2,147,483,647</code>).
long	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre -2^{63} (<code>-9,223,372,036,854,775,808</code>) et $2^{63}-1$ (<code>9,223,372,036,854,775,807</code>).
chaîne	Séquence de caractères.

5. Dans le champ **Position**, entrez la position de ce champ dans l'enregistrement.

Par exemple, dans ce fichier d'entrée, `AddressLine1` est en position 1, `City` est en position 2, `StateProvince` est en position 3, et `PostalCode` est en position 4.

```
"AddressLine1"|"City"|"StateProvince"|"PostalCode"
"7200 13TH ST"|"MIAMI"|"FL"|"33144"
"One Global View"|"Troy"|"NY"|"12180"
```

6. Si vous souhaitez que tout caractère d'espacement superflu soit supprimé au début et à la fin d'une chaîne de valeur dans un champ, cochez la case **Trim**.

Tri des enregistrements

L'onglet **Tri** vous permet de définir les critères en fonction desquels les différents champs doivent être triés dans les enregistrements de sortie avant qu'ils ne soient envoyés dans le flux de données.

1. Dans l'onglet **Tri**, cliquez sur **Ajouter**.
2. Cliquez sur la flèche de déroulement dans la colonne **Nom du champ** et sélectionnez le champ que vous voulez trier. Les champs disponibles à la sélection dépendent des champs définis dans ce fichier d'entrée.
3. Dans la colonne **Ordre**, indiquez si vous souhaitez que le champ soit trié dans l'ordre Ascending ou Descending.
4. Dans le champ **Type**, vous pouvez laisser le type de données défini sur `string` si vous ne prévoyez pas d'effectuer d'opérations mathématiques ou d'horodatage sur les données. Cependant, si vous avez l'intention d'effectuer ce genre d'opérations, sélectionnez un type de données approprié. Cela convertira les données `string` du fichier en un type de données qui permettra la manipulation appropriée des données dans le flux de données.
5. Si vous souhaitez que tout caractère d'espacement superflu soit supprimé au début et à la fin d'une chaîne de valeur dans un champ, cochez la case **Trim**.
6. Dans le champ **Traiter nul comme**, indiquez si vous souhaitez traiter les valeurs `null` du champ donné d'un enregistrement comme valeur `la plus grande` ou `la plus petite` parmi tous les enregistrements.
7. Répétez cette opération jusqu'à l'ajout de tous les champs d'entrée que vous souhaitez utiliser pour le tri. Changez l'ordre du tri en surlignant la ligne pour le champ que vous voulez déplacer et cliquez sur **Vers le haut** ou **Vers le bas**.

Variables

Entrées

Dans cette grille, parmi les champs reçus d'une activité source, sélectionnez les champs à utiliser dans cette activité.

Sorties

Dans cette grille, sélectionnez les champs à inclure dans la sortie de l'activité de destination dans ce flux de processus.

Submit Spark Job

L'activité **Submit Spark Job** vous permet d'exécuter tout job Spark, soit sur un cluster Hadoop, soit sur un cluster Spark. Grâce à cette activité, vous pouvez exécuter un job Spark du SDK Qualité des Big Data Spectrum™ ou tout job Spark externe.

Actuellement, vous pouvez soumettre un job Spark à l'un des deux types de cluster :

- YARN
- Spark

Modes de déploiement

Pour un job Spark, vous pouvez utiliser le mode de déploiement *cluster* ou *client*. Ces modes de déploiement déterminent si la classe de pilote du job Spark s'exécute sur le *cluster* ou sur le *client* Spectrum™ Technology Platform.

Pour simplifier, vous pouvez exécuter un job Spark dans l'un des modes de déploiement :

1. Mode YARN Cluster
2. Mode YARN Client
3. Mode Spark Client

Remarque : Les modes *client* YARN ou Spark sont recommandés lorsque le serveur Spectrum est installé et exécuté depuis l'environnement de cluster.

Pour obtenir une liste complète des configurations de job prises en charge sur les plates-formes Windows et Linux, consultez le tableau [Configurations de job prises en charge](#) à la page 221.

Champ	Description				
Nom du job	Nom du job.				
Serveur Hadoop	Liste des serveurs Hadoop configurés. Pour des informations sur le mappage des serveurs de fichiers HDFS via Management Console, reportez-vous au <i>Guide d'administration</i> .				
Chemin d'accès Jar	Chemin d'accès au fichier JAR approprié pour l'exécution du job Spark. Remarque : Le chemin d'accès Jar doit pointer vers un répertoire figurant sur l'ordinateur du serveur Spectrum.				
Type de job	Sélectionnez l'un des éléments suivants : <table border="0"> <tr> <td style="vertical-align: top;">Spectrum</td> <td>Pour exécuter l'un des jobs du SDK Qualité des Big Data Spectrum, sélectionnez <i>Spectrum</i>. Lors de la sélection de <i>Spectrum</i> , le champ Jobs Spectrum s'affiche.</td> </tr> <tr> <td style="vertical-align: top;">Générique</td> <td>Pour spécifier des propriétés supplémentaires pour tout job externe, sélectionnez <i>Générique</i>.</td> </tr> </table>	Spectrum	Pour exécuter l'un des jobs du SDK Qualité des Big Data Spectrum, sélectionnez <i>Spectrum</i> . Lors de la sélection de <i>Spectrum</i> , le champ Jobs Spectrum s'affiche.	Générique	Pour spécifier des propriétés supplémentaires pour tout job externe, sélectionnez <i>Générique</i> .
Spectrum	Pour exécuter l'un des jobs du SDK Qualité des Big Data Spectrum, sélectionnez <i>Spectrum</i> . Lors de la sélection de <i>Spectrum</i> , le champ Jobs Spectrum s'affiche.				
Générique	Pour spécifier des propriétés supplémentaires pour tout job externe, sélectionnez <i>Générique</i> .				

Champ	Description
Jobs Spectrum	<p>Sélectionnez le job requis dans la liste des jobs du SDK Qualité des Big Data Spectrum. La liste comprend ces jobs :</p> <ul style="list-style-type: none"> • Validation d'adresse • Advanced Transformer • Best of Breed • Duplicate Synchronization • Filtre • Groovy • Intraflow Match • Interflow Match • Joiner • Match Key Generator • Open Name Parser • Open Parser • Table Lookup • Transactional Match • Validate Address • Validate Address Global <p>Lors de la sélection du job Spectrum de votre choix :</p> <ol style="list-style-type: none"> 1. Les champs Nom du job, Nom de classe et Arguments sont automatiquement renseignés. <p>Tous les champs automatiquement renseignés peuvent être modifiés, si nécessaire, sauf le champ Nom de classe.</p> <p>Important : Pour le job Spectrum sélectionné, le champ Nom de classe automatiquement renseigné ne doit pas être modifié, sinon le job ne peut pas être exécuté.</p> 2. La grille Propriétés est automatiquement renseignée à l'aide des propriétés de configuration requises du job Spectrum sélectionné, avec leurs valeurs par défaut. <p>Vous pouvez ajouter ou importer davantage de propriétés ainsi que modifier les propriétés automatiquement renseignées, si nécessaire.</p>
Nom de classe	Nom complet de la classe de pilote du job.

Champ	Description
Arguments	<p>Liste des arguments séparés par des espaces. Ils sont transmis à la classe du pilote lors de l'exécution du job.</p> <p>Par exemple,</p> <pre>23Dec2016 /home/Hadoop/EYInc.txt</pre> <ol style="list-style-type: none"> 1. Ces variables peuvent être transmises comme arguments, qui sont définis de sorte à accepter des valeurs d'exécution dans le stage source ou dans ce stage en cours du process flow. Par exemple, si, dans la sortie du stage précédent du process flow, la variable <code>SalesStartRange</code> est définie, vous pouvez inclure cette variable dans cette liste d'arguments comme <code>\${SalesStartRange}</code> ainsi que d'autres arguments obligatoires, comme suit : <pre>23Dec2016 /home/Hadoop/EYInc.txt \${SalesStartRange}</pre> 2. Dans le cas où un argument donné contient un espace, encadrez-le à l'aide de guillemets doubles. Il peut s'agir par exemple de <code>"/home/Hadoop/Sales Records"</code>. <p>Jobs du SDK Qualité des Big Data Spectrum - Arguments :</p> <p>Pour exécuter les jobs <i>Spark</i> du SDK Qualité des Big Data Spectrum, transmettez les différents fichiers de configuration sous forme de liste d'arguments. Chaque clé d'argument accepte le chemin d'accès à un seul fichier de propriétés de configuration, où chaque fichier contient plusieurs propriétés de configuration.</p> <p>La syntaxe de la liste d'arguments des propriétés de configuration est la suivante :</p> <pre>[-config <Path to configuration file>] [-debug] [-input <Path to input configuration file>] [-conf <Path to Spark configuration file>] [-output <Path of output directory>]</pre> <p>Par exemple, pour un job Spark MatchKeyGenerator :</p> <pre>-config /home/hadoop/spark/matchkey/matchKeyGeneratorConfig.xml -input /home/hadoop/spark/matchkey/inputFileConfig.xml -output /home/hadoop/spark/matchkey/outputFileConfig.xml</pre> <p>Remarque : Si la même clé de propriété de configuration est spécifiée dans le champ Arguments et dans la grille Propriétés, mais que chacune pointe vers des fichiers de configuration différents, le fichier indiqué dans la grille Propriétés est celui retenu.</p> <p>Les exemples de propriétés de configuration sont livrés avec le SDK Data and Address Quality for Big Data et placés sous <code><Big Data Quality bundle>\samples\configuration</code>.</p>

Propriétés générales

Champ	Description
Maître	Sélectionnez l'une des options permettant d'exécuter le job Spark :
	YARN Pour lancer et gérer le job Spark à l'aide de YARN.
	Spark Pour lancer et gérer le job Spark à l'aide de l'application Spark.
URL Spark	URL permettant d'accéder au cluster Spark au format <code><hostname of Spark cluster>:<port of Spark cluster></code> . Ce champ n'est visible que si vous sélectionnez <code>Spark</code> dans le champ Maître .
Mode de déploiement	Sélectionnez l'une des options :
	Client Pour exécuter le pilote du job Spark sur le client Spectrum™ Technology Platform.
	Cluster Pour exécuter le pilote du job Spark sur un cluster.

Champ	Description
-------	-------------

Propriétés

Champ Description

Dans la grille, sous la colonne **Propriétés**, saisissez les noms des propriétés, et, sous la colonne **Valeur**, saisissez les valeurs des propriétés correspondantes.

Il existe certaines propriétés obligatoires selon le type de **Maître** et de **Mode de déploiement**.

Propriétés obligatoires YARN

<code>yarn.resourcemanager.hostname</code>	Adresse IP du ResourceManager YARN.
<code>yarn.resourcemanager.address</code>	Adresse, y compris l'adresse IP et le port du ResourceManager YARN au format <code><hostname>:<port></code> .

Propriétés du mode de déploiement Client

<code>spark.driver.host</code>	Adresse IP du poste sur lequel doit être exécuté le pilote Spark.	Obligatoire
<code>spark.client.mode.temp.location</code>	<p>Chemin d'accès au dossier <code>temp</code> sur le serveur Spectrum à utiliser pour les jobs Universal Addressing :</p> <ul style="list-style-type: none"> • Validate Address • Validate Address Global <p>Remarque : Nous vous recommandons vivement d'utiliser cette propriété pour les jobs Universal Addressing pour garantir l'utilisation du dossier <code>temp</code> spécifié pour obtenir des résultats</p>	Facultatif

Champ	Description
Propriétés du mode de déploiement Client	
	intermédiaires.

Ainsi :

1. Pour le mode YARN Cluster, les deux premières propriétés sont obligatoires.
2. Pour le mode YARN Client, les trois premières propriétés sont obligatoires.
3. Pour le mode SPARK Client, la troisième propriété est obligatoire.

Remarque : Vous pouvez définir les propriétés obligatoires ci-dessus soit en créant la connexion dans Management Console, soit dans cette activité Spark. Si les mêmes propriétés sont définies dans Management Console et dans l'activité du job Spark, les valeurs attribuées dans l'activité Spark sont applicables.

En plus de ces propriétés obligatoires, vous pouvez saisir ou importer autant d'autres propriétés que nécessaire pour exécuter le job.

Champ	Description
Importer	<p>Pour importer des propriétés d'un fichier, cliquez sur Importer. Accédez à l'emplacement du fichier de propriétés respectif et sélectionnez le fichier au format XML. Les propriétés contenues dans le fichier importé sont copiées dans la grille Propriétés.</p> <p>Remarque :</p> <ol style="list-style-type: none"> 1. Si la même propriété est définie ici et dans Management Console, les valeurs définies ici remplacent celles définies dans Management Console. 2. Le fichier de propriétés doit être au format XML et doit suivre la syntaxe : <pre><configuration> <property> <name>key</name> <value>some_value</value> <description>A brief description of the purpose of the property key.</description> </property> </configuration></pre> <p>Créez vos propres fichiers de propriétés en utilisant le format XML ci-dessus.</p> <ol style="list-style-type: none"> 3. Si la même propriété existe à la fois dans la grille et dans le fichier de propriétés importé, la valeur importée du fichier remplace la valeur existante dans la grille pour la même propriété. 4. Vous pouvez importer plusieurs fichiers de propriétés l'un après l'autre, si nécessaire. Les propriétés contenues dans chaque fichier importé sont ajoutées à la grille. 5. Assurez-vous que le fichier de propriétés est présent sur le serveur Spectrum™ Technology Platform. 6. La balise <code><description></code> est facultative pour chaque clé de propriété d'un fichier de propriétés de configuration. 7. Les données de référence doivent être placées localement sur les nœuds de données pour que les jobs pertinents puissent être exécutés. Cette propriété n'est disponible que pour les jobs qui utilisent des données de référence, tels que <i>Advanced Transformer</i>, <i>Validate Address Global</i> et <i>Validate Address</i>. La propriété est : <code>pb.bdq.reference.data.location</code>.

Dépendances

Dans cet onglet, ajoutez la liste des fichiers d'entrée et des fichiers Jar nécessaires pour exécuter le job.

Une fois que le job est exécuté, les fichiers de référence et les fichiers de référence Jar ajoutés ici sont disponibles dans le cache distribué du job.

Fichiers de référence

Pour ajouter les différents fichiers requis comme entrée pour exécuter le job, cliquez sur **Ajouter**, accédez à l'emplacement respectif sur votre système ou cluster local et sélectionnez le fichier en question.

Pour supprimer tout fichier ajouté à la liste, sélectionnez le fichier en question et cliquez sur **Supprimer**.

Fichiers JAR de référence

Pour ajouter les fichiers Jar requis pour exécuter le job, cliquez sur **Ajouter**, accédez à l'emplacement respectif sur votre système ou cluster local et sélectionnez le fichier Jar en question.

Pour supprimer tout fichier ajouté à la liste, sélectionnez le fichier en question et cliquez sur **Supprimer**.

Remarque : Le chemin d'accès Jar doit pointer vers un répertoire figurant sur l'ordinateur du serveur Spectrum.

Configurations de job prises en charge

Ce tableau répertorie les configurations de job Spark prises en charge sur les plates-formes Windows et Linux :

stages	Windows			Linux		
	Yarn+Client	Yarn+Cluster	Spark+Client	Yarn+Client	Yarn+Cluster	Spark+Client
MatchKeyGenerator	Pris en charge					
Intraflow Match	Pris en charge					
Interflow Match	Pris en charge					
Transactional Match	Pris en charge					
Table Lookup	Pris en charge					
Advanced Transformer	Pris en charge					
Open Name Parser	Pris en charge					
Duplicate Synchronization	Pris en charge					

	Windows			Linux		
stages	Yarn+Client	Yarn+Cluster	Spark+Client	Yarn+Client	Yarn+Cluster	Spark+Client
Filtre	Pris en charge					
Best of Breed	Pris en charge					
Open Parser	Pris en charge					
Address Validation (GAV)						
Groovy	Pris en charge					
Joiner	Pris en charge					
Validate Address Global (AD)	Pris en charge					
Validate Address (C1P)				Pris en charge	Pris en charge	Pris en charge

Remarque :

- Les stages GAV et C1P ne sont pas pris en charge sur la plate-forme Windows sur laquelle le répertoire d'installation de Spectrum contient des espaces.
- Le comportement des stages GAV et C1P doit être vérifié de nouveau sur la plate-forme Windows sur laquelle le répertoire d'installation de Spectrum ne contient aucun espace.
- Les données de référence sur la stratégie HDFS ne sont pas prises en charge.
- Le job Addressing avec la fonction *Grouper par* n'est pas pris en charge en mode Yarn+Cluster.

Variables

Entrées

Dans cette grille, parmi les champs reçus d'une activité source, sélectionnez les champs à utiliser dans cette activité.

Sorties

Dans cette grille, sélectionnez les champs à inclure dans la sortie de l'activité de destination dans ce flux de processus.

Réussie

Une activité réussie indique la fin d'un processflow. Un processflow doit avoir au moins une activité réussie.

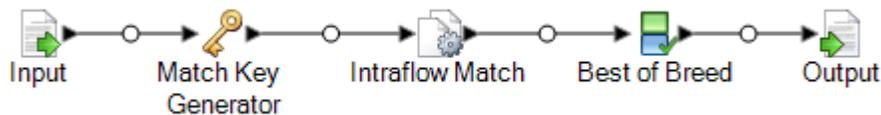
6 - Création de composants de flux réutilisables

In this section

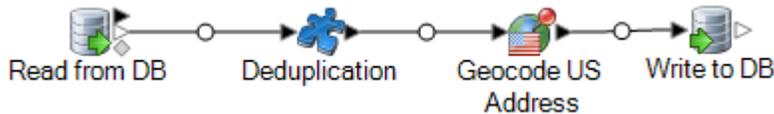
Introduction aux sous-flux	225
Utilisation d'un sous-flux comme source	225
Utilisation d'un sous-flux au milieu d'un flux de données	227
Utilisation d'un sous-flux comme collecteur de données	228
Modification d'un sous-flux	229
Suppression d'un sous-flux	229
Exposition et annulation de l'exposition d'un sous-flux	230
Conversion d'un stage en un sous-flux	230

Introduction aux sous-flux

Un sous-flux est un dataflow qui peut être réutilisé par d'autres dataflows. Les sous-flux s'avèrent utiles quand vous souhaitez créer un processus réutilisable, facile à intégrer à des dataflows. Par exemple, vous pouvez souhaiter créer un sous-flux qui effectue une déduplication en utilisant certains paramètres de chaque stage, pour pouvoir utiliser le même processus de déduplication dans différents dataflows. Pour ce faire, vous pouvez créer un sous-flux comme suit :



Vous pouvez ensuite utiliser ce sous-flux dans un dataflow. Par exemple, vous pouvez utiliser le sous-flux de déduplication dans un dataflow qui effectue un géocodage de sorte que les données soient dédupliquées avant l'opération de géocodage :



Dans cet exemple, les données d'une base de données sont lues, puis transmises au sous-flux de déduplication, où elles sont traitées par Match Key Generator, Intraflow Match, puis Best of Breed. Ensuite, pour finir, elles quittent le sous-flux et sont transmises au stage suivant du dataflow parent, dans ce cas, Geocode US Address. Les sous-flux sont représentés par une icône morceau de puzzle, comme illustré ci-dessus.

Les sous-flux enregistrés et exposés s'affichent dans le dossier **Stages définis par l'utilisateur** d'Enterprise Designer.

Utilisation d'un sous-flux comme source

Vous pouvez utiliser un sous-flux comme premier stage d'un flux de données pour lire les données d'une source, voire même pour réaliser un traitement des données avant de les transmettre au flux de données parent. Vous pouvez créer un sous-flux aussi simple qu'un stage source unique, configuré de sorte que vous puissiez le réutiliser dans d'autres flux de données ; ou vous pouvez créer un sous-flux plus complexe qui lit les données et les traite d'une certaine manière avant de les transmettre au flux de données parent.

1. Dans Enterprise Designer, cliquez sur **Fichier > Nouveau > Flux de données > Sous-flux**.
2. Faites glisser la source de données appropriée de la palette jusqu'au canevas et configurez-la.

Par exemple, si vous souhaitez que le sous-flux lise les données d'un fichier séparé par des virgules, vous pouvez faire glisser un stage Read from File jusqu'au canevas.

3. Si vous souhaitez que le sous-flux traite les données d'une certaine manière avant de les envoyer au flux de données parent, ajouter d'autres stages, si besoin, pour réaliser le pré-traitement de votre choix.
4. À la fin du flux de données, ajoutez un stage Output et configurez-le.

Cette opération permet d'envoyer les données du sous-flux au flux de données parent.

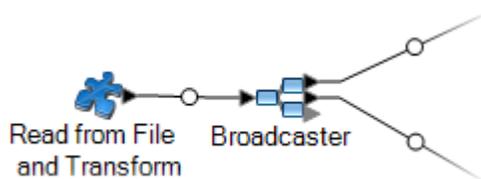
Par exemple, si vous créez un sous-flux qui lit les données d'un fichier, puis utilise un stage Transformer pour couper les espaces blancs et normaliser la casse d'un champ, votre sous-flux peut prendre la forme suivante :



5. Double-cliquez sur le stage Output et sélectionnez les champs que vous souhaitez transmettre au flux de données parent.
6. Sélectionnez **Fichier > Enregistrer** et enregistrez le sous-flux.
7. Sélectionnez **Fichier > Exposer** pour que le sous-flux puisse être inclus dans des flux de données.
8. Dans le flux de données auquel inclure le sous-flux, faites glisser le sous-flux de la palette jusqu'au canevas.
9. Reliez le sous-flux au stage de flux de données de votre choix.

Remarque : Étant donné que le sous-flux contient un stage source et non un stage Input, seule l'icône du sous-flux comporte un port de sortie. Le sous-flux ne peut être utilisé que comme source dans le flux de données.

Le flux de données parent utilise maintenant le sous-flux que vous avez créé comme entrée. Par exemple, si vous avez créé un sous-flux nommé « Read from File and Transform », que vous ajoutez le sous-flux et que vous le reliez à un stage Broadcaster, votre flux de données prend la forme suivante :



Utilisation d'un sous-flux au milieu d'un flux de données

Vous pouvez utiliser un sous-flux au milieu d'un flux de données pour que le traitement de votre choix puisse être réutilisable dans d'autres flux de données. En effet, le sous-flux devient un stage personnalisé dans votre flux de données.

1. Dans Enterprise Designer, cliquez sur **Fichier > Nouveau > Flux de données > Sous-flux**.

2. Faites glisser un stage Input de la palette jusqu'au canevas.

Cette opération permet d'envoyer les données du flux de données parent au sous-flux.

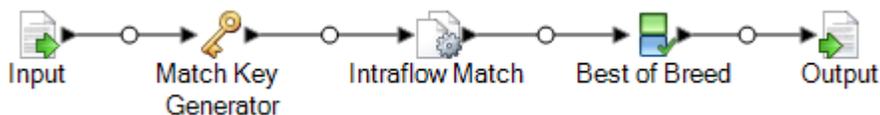
3. Double-cliquez sur le stage Input et ajoutez les champs que le sous-flux va recevoir du flux de données dans lequel il est utilisé.

4. Après avoir configuré le stage Input, ajoutez d'autres stages, si besoin, pour réaliser le traitement de votre choix.

5. À la fin du flux de données, ajoutez un stage Output.

Cette opération permet d'envoyer les données du sous-flux au flux de données parent.

Par exemple, vous pouvez souhaiter créer un sous-flux qui effectue une déduplication en utilisant certains paramètres de chaque stage, pour pouvoir utiliser le même processus de déduplication dans différents flux de données. Pour ce faire, vous pouvez créer un sous-flux comme suit :



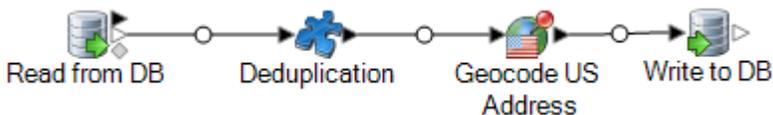
6. Sélectionnez **Fichier > Enregistrer** et enregistrez le sous-flux.

7. Sélectionnez **Fichier > Exposer** pour que le sous-flux puisse être inclus dans des flux de données.

8. Dans le flux de données auquel inclure le sous-flux, faites glisser le sous-flux de la palette jusqu'au canevas.

9. Reliez le sous-flux au stage de flux de données de votre choix.

Par exemple, vous pouvez utiliser le sous-flux de déduplication dans un flux de données qui effectue un géocodage de sorte que les données soient dédupliquées avant l'opération de géocodage :



Utilisation d'un sous-flux comme collecteur de données

Vous pouvez utiliser un sous-flux comme dernier stage d'un flux de données pour écrire les données dans un fichier ou une base de données, voire même pour réaliser un traitement des données avant de les écrire vers la destination de sortie. Vous pouvez créer un sous-flux aussi simple qu'un stage de collecteur de données unique, configuré de sorte que vous puissiez le réutiliser dans d'autres flux de données ; ou vous pouvez créer un sous-flux plus complexe qui traite les données d'une certaine manière avant de les écrire vers la destination de sortie.

1. Dans Enterprise Designer, cliquez sur **Fichier > Nouveau > Flux de données > Sous-flux**.
2. Faites glisser un stage Input de la palette jusqu'au canevas.
3. Double-cliquez sur le stage Input et ajoutez les champs que le sous-flux va recevoir du flux de données dans lequel il est utilisé.
4. Après avoir configuré le stage Input, ajoutez d'autres stages, si besoin, pour réaliser le post-traitement de votre choix.
5. À la fin du flux de données, ajoutez le collecteur de données approprié.

Par exemple, si vous avez créé un sous-flux qui utilise un stage Transformer pour couper les espaces blancs et standardiser la casse d'un champ, l'écrire dans une base de données, votre sous-flux prend la forme suivante :



6. Sélectionnez **Fichier > Enregistrer** et enregistrez le sous-flux.
7. Sélectionnez **Fichier > Exposer** pour que le sous-flux puisse être inclus dans des flux de données.
8. Dans le flux de données auquel inclure le sous-flux, faites glisser le sous-flux de la palette jusqu'au canevas et reliez-le au dernier stage du flux de données.

Remarque : Étant donné que le sous-flux contient un stage de collecteur de données et non un stage Output, seule l'icône du sous-flux comporte un port d'entrée. Le sous-flux ne peut être utilisé que comme collecteur de données dans le flux de données.

Le flux de données parent utilise maintenant le sous-flux que vous avez créé comme un collecteur de données. Par exemple, si vous avez créé un sous-flux nommé « Transform and Write to DB », que vous ajoutez le sous-flux et que vous le reliez à un stage Geocode US Address, votre flux de données prend la forme suivante :



Modification d'un sous-flux

1. Ouvrez le sous-flux dans Enterprise Designer.
2. Avant de modifier un sous-flux, prenez en considération l'impact de la modification sur les flux de données exploitant le sous-flux. Pour savoir quels flux de données utilisent le sous-flux, sélectionnez **Outils > Utilisé par**.
3. Modifiez le sous-flux selon les besoins.

Remarques :

- Lorsque vous supprimez un stage Input ou Output ou que vous ajoutez un stage Input ou Output, Enterprise Designer affiche un message d'avertissement vous rappelant que d'autres flux de données utilisent le sous-flux et vous propose de voir quels flux de données exploitent le sous-flux. Si vous continuez d'enregistrer le stage réutilisable, Enterprise Designer fait en sorte que tous les flux de données utilisés par le sous-flux ne sont pas exposés.
 - Si vous changez un sous-flux d'une autre manière, par exemple en modifiant le nom d'un fichier ou les configurations du stage, Enterprise Designer affiche un message d'avertissement vous rappelant que d'autres flux de données utilisent le sous-flux et vous propose de visualiser quels flux de données exploitent le sous-flux. Vous pouvez continuer sans que ces flux de données n'aient été rendus exposés.
4. Une fois les modifications apportées, sélectionnez **Fichier > Enregistrer**.
 5. Sélectionnez **Afficher > Actualiser** pour que les modifications soient reflétées dans le flux de données parent.

Remarque : Si vous disposez de plusieurs versions du sous-flux, n'oubliez pas que la version utilisée dans le dataflow parent est la version exposée. Lorsque vous modifiez un sous-flux, veillez à exposer sa version la plus récente, afin que vos modifications soient appliquées aux dataflows utilisant le sous-flux.

Suppression d'un sous-flux

Si vous essayez de supprimer un sous-flux exposé, Enterprise Designer affiche un message d'avertissement vous rappelant que d'autres flux de données utilisent le sous-flux que vous vous apprêtez à supprimer. Si vous continuez la suppression du sous-flux, Enterprise Designer rend tous les flux de données non exposés.

Exposition et annulation de l'exposition d'un sous-flux

Pour qu'un sous-flux puisse être utilisé dans un dataflow, il doit être exposé. Pour exposer un sous-flux, ouvrez-le dans Enterprise Designer et accédez à **Fichier > Exposer/Ne plus exposer et Enregistrer**. Cela rendra le sous-flux disponible et utilisable dans d'autres dataflows.

Remarque : Si vous disposez de plusieurs versions du sous-flux, n'oubliez pas que la version utilisée dans le dataflow parent est la version exposée. Lorsque vous modifiez un sous-flux, veillez à exposer sa version la plus récente, afin que vos modifications soient appliquées aux dataflows utilisant le sous-flux.

Pour annuler l'exposition d'un sous-flux, ouvrez-le dans Enterprise Designer et sélectionnez **Fichier > Exposer/Ne plus exposer et Enregistrer**. Lorsqu'un sous-flux est non exposé, Enterprise Designer affiche un message d'avertissement vous rappelant que d'autres dataflows utilisent le sous-flux que vous vous apprêtez à modifier. Si vous continuez à ne pas exposer le sous-flux, Enterprise Designer convertit tous les dataflows qui utilisent ce dataflow en non exposés.

Conversion d'un stage en un sous-flux

1. Créez un nouveau job, service ou sous-flux.
2. Ajoutez les stages que vous souhaitez inclure dans le job, service ou sous-flux.
3. Si vous souhaitez configurer le stage à cet étape, faites un clic-droit sur le stage et cliquez sur **Options**. Configurez ensuite les options du stage comme souhaité puis cliquez sur **OK**.
4. Cliquez avec le bouton droit sur le stage que vous voulez convertir et sélectionnez **Convertir un stage en sous-flux**. La boîte de dialogue **Enregistrer sous** apparaît.
5. Entrez le nom que vous souhaitez donner aux sous-flux et cliquez sur **OK**, puis enregistrez le service. Ce nom doit être unique sur le système. Trois événements se produisent :
 - Le système crée un nouveau sous-flux qui inclut les éléments suivants :
 - le stage que vous avez sélectionné
 - une entrée de flux de données pour chaque port d'entrée du stage
 - une sortie de flux de données pour chaque port de sortie du stage
 - des connexions entre le stage et ses entrées et sorties
 - Le système remplace le stage que vous avez sélectionné par le nouveau sous-flux.
 - Le système expose le nouveau sous-flux. Vous le verrez dans l'Explorateur du serveur et dans la section Stages définis par l'utilisateur de la boîte à outils.

Une fois que vous avez créé un sous-flux et que vous l'avez utilisé dans d'autres flux de données, vous pouvez voir les autres flux de données qui sont en train d'employer ce sous-flux. Ouvrez le sous-flux et accédez à **Outils > Utilisé par**. (Vous pouvez également cliquer avec le bouton droit de la souris sur le sous-flux dans Explorateur du serveur et sélectionner **Utilisé par**.) Il s'affiche une liste de flux de données employant le sous-flux existant, ce qui vous permet de voir les flux de données affectés si vous modifiez le sous-flux existant.

7 - Exemples de flux

In this section

Introduction	233
Intégration entre SugarCRM OnPremises et Microsoft Dynamics 365 Online	234
Intégration entre Salesforce et Oracle Eloqua	237

Introduction

L'Exemple de modèle est un ensemble de modèles pré-configurés à l'aide duquel vous pouvez démontrer les capacités de migration de Spectrum. Il supprime le besoin de créer des dataflows à partir de rien. Pour démontrer les capacités de migration, vous devez créer plusieurs connexions et importer ces flux dans votre système. Cette section décrit le processus complet de migration des données à l'aide d'exemples de flux.

Localisation de l'exemple de modèle

L'exemple de modèle est livré avec le fichier zip du programme d'installation de Spectrum Technology Platform. Suivant votre répertoire d'installation, vous pouvez trouver l'exemple de modèle sur le chemin d'accès suivant :

```
Program Files\Pitney  
Bowes\Spectrum\server\modules\metadata-insights\connectors\samples
```

Le dossier contient tous les fichiers requis pour configurer le dataflow.

Création de connexions

Vous devez créer deux connexions pour importer et déployer les fichiers requis pour configurer le dataflow. Pour créer les connexions, procédez comme suit :

1. Installez le dernier serveur Spectrum Platform Server.
2. Ouvrez la page d'accueil de Spectrum Server.
3. Cliquez sur **Outils client de la plateforme** , puis sur **Web**
4. Ouvrez la vue **Management Console**.
5. Cliquez sur le menu **Ressources**, puis sur **Sources de données**, et créez deux connexions.

Importation et déploiement des fichiers sur le serveur

Une fois la connexion en place, vous avez maintenant besoin d'un utilitaire à ligne de commande pour importer et déployer les fichiers sur le serveur.

1. Ouvrez la page d'accueil de Spectrum Platform, cliquez sur **Outils client de la plateforme**, puis sur **Ligne de commande** et téléchargez les fichiers suivants :
 - a. Job Executor, qui est un fichier .jar individuel
 - b. Utilitaire Administration, qui est un fichier zippé (spectrum-cli)
2. Décompressez le fichier spectrum-cli.
3. Copiez tous les fichiers du dossier Exemple de modèle dans le dossier spectrum-cli. Cela vous évite d'avoir à saisir le chemin d'accès complet à chaque fichier avec chaque commande.
4. Exécutez l'utilitaire cli.cmd qui se trouve dans le dossier spectrum-cli-12.1. L'interface de ligne de commande de Spectrum Platform est lancée.

5. Connectez-vous au serveur Spectrum Server à l'aide des commandes suivantes : `connect <nom de serveur>:<port> --u <nom d'utilisateur> --p <mot de passe>`. La connexion réussie est confirmée par le message : `Connecté au serveur<nom de serveur> :< port>`.
6. Saisissez et exécutez les commandes suivantes dans l'ordre indiqué pour importer et déployer exemples de dataflows :
 - a. Importez modelstore à l'aide de la commande : `spectrum> modelstore bulkImport --importDependency true`. Cette commande importe modelstore avec toutes les dépendances.
 - b. Déployez modelstore à l'aide de la commande : `modelstore deploy --n <name>`.
 - c. Importez dbconnection à l'aide de la commande : `dbconnection import --f <name.json>`.
 - d. Importez des dataflows à l'aide de la commande : `dataflow import --f <name.df>`.

Remarque :

Supprimez les espaces supplémentaires autour des commandes des noms de fichier.

Remplacez l'espace réservé <nom> par le nom exact du fichier.

Affichage des fichiers importés

Une fois l'importation terminée, vous pouvez utiliser ces dataflows pour démontrer les fonctionnalités de Spectrum. Les fichiers sont stockés aux emplacements suivants :

1. **Model Store** : dans les onglets Physical Model et Model Store de Metadata Insights
2. **Dataflow** : dans l'Explorateur Serveur de l'application Enterprise Designer
3. **JSON** : dans le Model Store de Management Console

Les flux importés peuvent être utilisés pour migrer des données d'un système vers un autre à l'aide d'Enterprise Designer. Des étapes détaillées pour chacun des flux sont mentionnées dans leurs sections respectives suivantes.

Remarque : Ces exemples de modèles font partie du fichier du programme d'installation Spectrum et ne fonctionnent pas de manière autonome. Plus aucune assistance n'est disponible pour ces exemples.

Intégration entre SugarCRM OnPremises et Microsoft Dynamics 365 Online

Cette section illustre la migration de Accounts et Contacts de SugarCRM OnPremises vers Microsoft Dynamics 365 Online. S'il existe une association de Contacts et de Accounts dans SugarCRM, la migration fait en sorte de la conserver également dans le système Microsoft Dynamics 365 Online.

L'Exemple de modèle contient les fichiers suivants :

Tableau 10 : Fichiers du dossier Exemple de modèle

Type de fichier	Nom de fichier
Dataflow	<ul style="list-style-type: none"> • SugarCRMAccount_Sync_MSAccount.df • SugarCRMContact_Sync_MDSContact.df
Connexion DB	<ul style="list-style-type: none"> • MSDynamics_MS.json • SugarCRM_MS.json
ModelStore	<ul style="list-style-type: none"> • mi_modelStore_MS Dynamics_MS.smims • mi_modelStore_SugarCRM_MS.smims

Procédez comme suit pour migrer des données de SugarCRM OnPremise vers MS Dynamics 365 Online.

1. Créez les connexions suivantes :
 - a. **SugarCRM_TestConnection** de type SugarCRM. Cliquez sur Test et un message de réussite s'affiche : *La connexion SugarCRM_OnPremises à la source de données a bien été établie.*
 - b. **MSDynamics_TestConnection** de type Microsoft Dynamics 365. Cliquez sur Test et un message de réussite s'affiche : *La connexion Microsoft Dynamics 365 Online à la source de données a bien été établie.*

Pour plus de détails sur la façon de créer des connexions, reportez-vous aux sections [Connexion à SugarCRM](#) et [Connexion à Microsoft Dynamics 365 Online](#).

2. Importez et déployez des fichiers à l'aide de l'utilitaire de ligne de commande Spectrum. Utilisez les commandes suivantes dans l'ordre indiqué :
 - Importez modelstore à l'aide de la commande :


```
modelstore bulkimport --importDependency true
```
 - Déployez modelstore à l'aide de la commande :


```
modelstore deploy --modelStoreName MSDynamics_MSmodelstore deploy --modelStoreName SugarCRM_MS
```
 - Importez dbconnection à l'aide de la commande :


```
dbconnection import --f MSDynamics_MS.jsondbconnection import --f SugarCRM_MS.json
```
 - Importez des dataflows à l'aide de la commande :

```
dataflow import --f SugarCRMAccount_Sync_MSDAccount.dfd
dataflow import --f SugarCRMContact_Sync_MDSContact.df
```

3. Lancez l'application Enterprise Designer. Vous pouvez télécharger le programme d'installation exécutable de cette application depuis la section **Bureau de Outils client de la plateforme** sur la page d'accueil de Spectrum Platform.
4. Connectez-vous à l'aide de vos informations d'identification Spectrum.
5. Cliquez sur **Afficher** dans le menu, puis sur **Explorateur Serveur**
6. Double-cliquez d'abord sur le job de dataflow **SugarCRMAccount_Sync_MSDAccount**.
7. Double-cliquez sur le stage **Read from DB_SugarCRM**.
 - a. Modifiez la valeur du champ `date_entered`, si nécessaire, puis cliquez sur **OK**.

```
Select "SugarCRM_PM"."Accounts"."email1",
       "SugarCRM_PM"."Accounts"."name",
       "SugarCRM_PM"."Accounts"."phone_office",
       "SugarCRM_PM"."Accounts"."date_entered" From
       "SugarCRM_PM"."Accounts" Where
       "SugarCRM_PM"."Accounts"."date_entered" Like '2017-08-28%'
```

- b. Cliquez sur **OK** pour continuer.
8. Exécutez le flux en cliquant sur le bouton **Exécuter** de la barre d'outils d'Enterprise Designer.
9. Accédez à la page d'accueil MS Dynamics et cliquez sur **Accounts** sous l'onglet **Customers** sur la page **Sales**. Notez que **Accounts** a été migré.
10. Double-cliquez sur le job **SugarCRMContact_Sync_MDSContacts**.
11. Double-cliquez sur le stage **Read from_MSDynamics_DB**.
 - a. Modifiez la valeur du champ `Créé le`, si nécessaire, puis cliquez sur **OK**.

```
Select "MSDynamics_PM"."account"."name",
       "MSDynamics_PM"."account"."telephone1",
       "MSDynamics_PM"."account"."emailaddress1",
       "MSDynamics_PM"."account"."createdon",
       "MSDynamics_PM"."account"."accountid" From
       "MSDynamics_PM"."account"
       Where "MSDynamics_PM"."account"."createdon" Like '2017-09-11%'
```

- b. Cliquez sur **OK** pour continuer.
 12. Exécutez le flux en cliquant sur le bouton **Exécuter** de la barre d'outils d'Enterprise Designer.
 13. Accédez à la page d'accueil MS Dynamics et cliquez sur **Contacts** sous l'onglet **Customers** de la page **Sales**. Notez que les **Contacts** s'affichent sur la page.
- L'affichage des comptes et des contacts sur la page MS Dynamics termine le processus de migration.

Intégration entre Salesforce et Oracle Eloqua

Cette section illustre la migration de Accounts et de Contacts de Salesforce vers les systèmes Oracle Eloqua. S'il existe une association de Contacts et de Accounts dans Salesforce, la migration fait en sorte de la conserver également dans le système Oracle Eloqua.

L'exemple de modèle contient les fichiers suivants :

Tableau 11 : Fichiers du dossier Exemple de modèle

Type de fichier	Nom de fichier
Dataflow	<ul style="list-style-type: none"> • <i>SalesforceAccount_Sync_OracleEloquaAccount.df</i> • <i>SalesforceContact_Sync_OracleEloquaContact.df</i>
Connexion DB	<ul style="list-style-type: none"> • <i>Eloqua_MS.json</i> • <i>Salesforce_MS.json</i>
ModelStore	<ul style="list-style-type: none"> • <i>mi_modelStore_Eloqua_MS.smims</i> • <i>mi_modelStore_Salesforce_MS.smims</i>

Pour migrer des données de Salesforce vers Oracle Eloqua, procédez comme suit :

1. Créez les connexions suivantes :
 - a. **Salesforce_TestConnection** avec le type Salesforce. Cliquez sur Test et un message de réussite s'affiche : *La connexion Salesforce à la source de données a bien été établie.*
 - b. **Eloqua_TestConnection** avec le type Oracle Eloqua. Cliquez sur Test et un message de réussite s'affiche : *La connexion Eloqua à la source de données a bien été établie.*

Pour plus de détails sur la façon de créer des connexions, reportez-vous aux sections [Connexion à Salesforce](#) et [Connexion à Oracle Eloqua](#).

2. Importez et déployez des fichiers à l'aide de l'utilitaire de ligne de commande Spectrum. Utilisez les commandes suivantes dans l'ordre indiqué :
 - Importez modelstore à l'aide de la commande :


```
modelstore bulkimport --importDependency true
```
 - Déployez modelstore à l'aide de la commande :

```
modelstore deploy --modelStoreName Eloqua_MSmodelstore deploy
--modelStoreName Salesforce_MS
```

- Importez dbconnection à l'aide de la commande :

```
dbconnection import --f Eloqua_MS.jsondbconnection import --f
Salesforce_MS.json
```

- Importez des dataflows à l'aide de la commande : #

```
dataflow import --f
SalesforceAccount_Sync_OracleEloquaAccount.dfdataflow import --f
SalesforceContact_Sync_OracleEloquaContact.df
```

3. Lancez l'application Enterprise Designer. Vous pouvez télécharger le programme d'installation exécutable de cette application depuis la section **Bureau de Outils client de la plateforme** sur la page d'accueil de Spectrum Platform.
4. Connectez-vous à l'aide de vos informations d'identification Spectrum.
5. Cliquez sur **Afficher** dans le menu, puis sur **Explorateur Serveur**
6. Double-cliquez d'abord sur le job de dataflow **SalesforceAccount_Sync_OracleEloquaAccount**.
7. Double-cliquez sur le stage **Read from DB**.
 - a. Modifiez la valeur du champ CreatedDate, si nécessaire, puis cliquez sur **OK**.

```
Select "SalesforcePM"."Account".* From "SalesforcePM"."Account"
where CreatedDate > '2017-01-01'
```

- b. Cliquez sur **OK** pour continuer.

8. Exécutez le flux en cliquant sur le bouton **Exécuter** de la barre d'outils d'Enterprise Designer.
9. Accédez à la page d'accueil d'Eloqua et cliquez sur **Accounts**. Notez que les comptes ont migré vers Eloqua.
10. Revenez à Enterprise Designer.
11. Double-cliquez sur le job **SalesforceContact_Sync_OracleEloquaContact**.
12. Double-cliquez sur le stage **Read from DB**.
 - a. Modifiez la valeur du champ CreatedDate, si nécessaire, puis cliquez sur **OK**.

```
Select "SalesforcePM"."Contact".* From "SalesforcePM"."Contact"
where CreatedDate > '2017-01-01'
```

- b. Cliquez sur **OK** pour continuer.

13. Exécutez le flux en cliquant sur le bouton **Exécuter** de la barre d'outils d'Enterprise Designer.
14. Accédez à la page d'accueil d'Eloqua et cliquez sur **Contacts**. Notez que les contacts ont migré vers Eloqua.

L'affichage des comptes et des contacts dans Oracle Eloqua termine le processus de migration.

8 - A propos de Spectrum™ Technology Platform

In this section

Qu'est-ce que Spectrum™ Technology Platform ?	240
Architecture de gestion des données d'entreprise	241
Architecture Spectrum™ Technology Platform	245
Modules et composants	249

Qu'est-ce que Spectrum™ Technology Platform ?

Spectrum™ Technology Platform est un système permettant d'améliorer l'exhaustivité, la validité, la cohérence, la précision et la transmission en temps opportun de vos données par la standardisation, la vérification et l'amélioration. En recevant l'assurance que ses données sont précises, complètes, et à jour, votre entreprise est à même de mieux comprendre et mieux communiquer avec ses clients.

Spectrum™ Technology Platform facilite la conception et l'implémentation de votre politique métier concernant la qualité des données en remplissant les fonctions suivantes.

Parsing, normalisation des noms, et validation des noms.

Pour obtenir la plus grande précision de standardisation possible, vous devez décomposer les chaînes de données en plusieurs champs. Spectrum™ Technology Platform fournit des fonctionnalités d'analyse avancées vous permettant d'analyser les noms de personnes, les noms d'entreprises et de nombreux autres termes et abréviations. De plus, vous pouvez créer votre propre liste de termes personnalisés à utiliser comme base pour vos opérations d'analyse/extraction. Le module Universal Name fournit ces fonctionnalités.

Déduplication et consolidation

L'identification d'entité unique vous permet de consolider des enregistrements, d'éliminer des doublons et de développer des enregistrements de type Best of Breed (Best of Breed). Un enregistrement de type "Best of Breed" est un enregistrement composite élaboré à l'aide de données provenant d'autres enregistrements. Le module Advanced Matching et le module Data Normalization fournissent ces fonctionnalités.

Validation d'adresse

La validation d'adresse applique des règles en provenance des autorités postales appropriées afin d'écrire une adresse sous sa forme normalisée et même la valider comme adresse livrable. La validation d'adresse peut vous aider à être éligible à certaines réductions sur les tarifs postaux et améliorer la délivrabilité de vos courriers. Le module Universal Addressing et le module Address Now fournissent ces fonctionnalités.

Géocodage

Le géocodage est la procédure qui consiste à déterminer des coordonnées géographiques (latitude et longitude) à partir d'une adresse complète ou partielle. Le géocodage peut être utilisé pour la création de cartes, mais ce n'est pas sa seule application. Les données de localisation sous-jacentes peuvent vous servir dans vos décisions professionnelles. En renversant le processus, vous pouvez saisir un géocode (un point représenté par ses coordonnées de longitude et de latitude) et recevoir des informations d'adresse se rapportant à ce géocode. Le module Enterprise Geocoding fournit ces fonctionnalités.

Location Intelligence

La Location Intelligence (localisation avancée) génère de nouvelles informations sur vos données en évaluant, analysant et modélisant des liens entre territoires. En utilisant le traitement de Location Intelligence, vous pouvez vérifier des lieux et transformer des informations en renseignements commerciaux précieux. Le module Location Intelligence fournit ces fonctionnalités.

Master Data Management

Master Data Management vous permet de créer des vues de données de référence centrées sur les relations de vos actifs de données critiques. Le module Data Hub vous permet d'identifier les influenceurs et les relations non évidentes, de détecter les fraudes et d'améliorer la qualité, l'intégration et l'accessibilité de vos informations.

Assignment de juridiction fiscale

L'assignation de juridiction fiscale prend une adresse et en détermine les juridictions fiscales qui s'appliquent à l'emplacement correspondant à l'adresse en question. Une assignation la plus précise possible des juridictions fiscales peut réduire vos risques financiers et vos responsabilités en matière de réglementation.

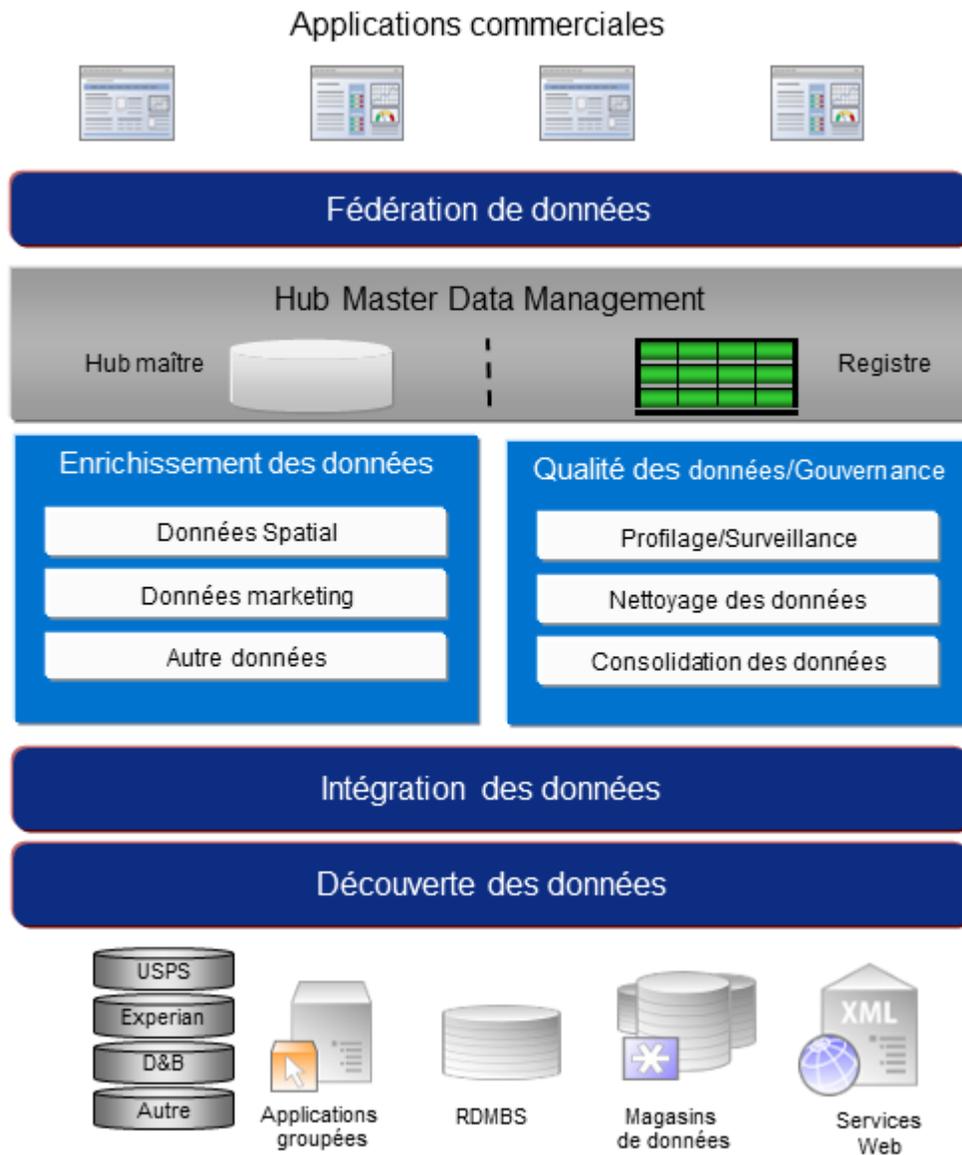
Le logiciel Spectrum™ Technology Platform de Pitney Bowes intègre les toutes dernières circonscriptions juridiques avec l'adresse postale exacte des enregistrements de votre client, vous permettant d'ajouter à vos enregistrements les informations sur l'État, le pays, le canton, la municipalité et des informations particulières à la juridiction fiscale. Quelques exemples d'utilisation de l'assignation de juridiction fiscale :

- Taxe sur les ventes et la consommation
- Impôt mobilier
- Taxe sur les primes d'assurance

Le module Enterprise Tax Module fournit ces fonctionnalités.

Architecture de gestion des données d'entreprise

Avec Spectrum™ Technology Platform, vous pouvez générer une procédure complète de gestion des données d'entreprise, ou vous pouvez l'utiliser comme une solution plus ciblée. Le schéma suivant illustre une solution complète qui extrait les données d'une source, les fiabilise et les enrichit, et qui alimente un hub de gestion de données de référence qui offre une vue unique des données pour différentes applications métier.



Data Discovery

Data Discovery est le processus de balayage de vos ressources de données visant à obtenir un inventaire complet de votre paysage de données. Spectrum™ Technology Platform peut balayer les données structurées, non structurées et semi-structurées à l'aide d'une grande variété de techniques de profilage de données. Les résultats du balayage sont utilisés pour générer automatiquement une bibliothèque de documentation décrivant les actifs de données de votre entreprise et pour créer un référentiel de métadonnées. Cette documentation et le référentiel de métadonnées associé vous fournissent les informations dont vous avez besoin avant de commencer des projets d'intégration de données, de gestion de la qualité des données, de gouvernance de données ou de gestion de données de référence.

Pour plus d'informations sur le module Data Discovery Spectrum™ Technology Platform, contactez votre responsable de compte.

Data Integration

Une fois que vous disposez d'un inventaire de votre paysage de données, vous devez envisager le mode d'accès aux données que vous devez gérer. Spectrum™ Technology Platform peut se connecter à vos données via plusieurs sources, soit directement, soit par l'intermédiaire d'une intégration à vos technologies d'accès aux données existantes. Il prend en charge des fonctionnalités d'intégration de données par lots et en temps réel pour une grande variété de besoins d'entreprise, notamment l'entreposage de données, la gestion de la qualité des données, l'intégration système et la migration. Spectrum™ Technology Platform peut accéder aux données de bases de données RDBMS, de magasins de données, de fichiers XML, de fichiers plats, etc. Spectrum™ Technology Platform prend en charge les requêtes SQL avec des agrégations et des jointures complexes et constitue un outil de développement de requêtes visuel. De plus, Spectrum™ Technology Platform peut accéder aux données via les services Web REST et SOAP.

Spectrum™ Technology Platform peut déclencher un traitement par lots en fonction de l'apparence d'un ou de plusieurs fichiers sources dans un dossier spécifié. Ce déclenchement de « dossier à chaud » permet de surveiller les chargements FTP et de les traiter à mesure qu'ils se produisent.

Certaines de ces fonctions d'intégration de données nécessitent une licence pour le module Enterprise Data Integration. Pour plus d'informations, contactez votre responsable de compte.

Enfin, Spectrum™ Technology Platform peut s'intégrer à des applications groupées, telles que SAP.

Qualité des données/gouvernance

Les procédures de gestion de la qualité et de gouvernance des données recherchent les enregistrements en double, les informations incohérentes et les informations inexacts dans vos données.

Le rapprochement de doublons identifie les enregistrements doublons potentiels ou les relations entre les enregistrements, si les données sont de type nom et adresse, ou appartiennent à un autre type d'informations clients. Spectrum™ Technology Platform vous permet d'indiquer un ensemble cohérent de règles de correspondance pour votre entreprise en utilisant des méthodes de correspondance booléennes, des méthodes de calcul du score, des seuils, des algorithmes et des pondérations afin de déterminer si un groupe d'enregistrements contient des doublons. Spectrum™ Technology Platform prend en charge une personnalisation étendue pour vous permettre d'ajuster ces règles aux besoins uniques de votre entreprise.

Une fois les enregistrements doublons identifiés, vous pouvez souhaiter consolider les enregistrements. Spectrum™ Technology Platform vous permet d'indiquer la manière de lier ou de fusionner les enregistrements doublons pour pouvoir créer l'enregistrement le plus complet et précis à partir de toute collecte d'informations client. Par exemple, un seul enregistrement de type best-of-breed (premier de groupe) peut être généré à partir de tous les enregistrements d'un même foyer. Le module Advanced Matching permet d'identifier les doublons et de les éliminer.

Les procédures de gestion de qualité des données normalisent également vos données. La normalisation est une procédure critique car les éléments de données normalisées sont requis pour

atteindre les meilleurs résultats possibles pour mettre en correspondance et identifier les relations entre les enregistrements. Tandis que plusieurs modules réalisent une normalisation d'un type ou d'un autre, c'est le module Data Normalization de la Spectrum™ Technology Platform qui fournit l'ensemble des fonctions de normalisation le plus complet. En outre, le module Universal Name fournit des fonctions de fiabilisation des données de noms personnel et de noms d'entreprises.

Des données normalisées ne sont pas forcément des données précises. Spectrum™ Technology Platform peut comparer vos données aux données de référence connues et à jour à des fins de correction. Les sources utilisées pour cette procédure peuvent inclure des autorités réglementaires, telles que le service postal des Etats-Unis, Postal Service, des fournisseurs de données tiers tels Experian ou D&B, ou encore les sources de référence internes de votre entreprise, comme des données comptables. Spectrum™ Technology Platform est particulièrement efficace pour la validation de données d'adresse. Il peut valider ou normaliser les adresses de 250 pays et territoires dans le monde. Il existe deux modules de validation des adresses : le module Address Now et le module Universal Addressing.

Pour déterminer celui qui vous convient, parlez de vos besoins avec votre responsable de compte.

Alors que Spectrum™ Technology Platform peut gérer automatiquement un large éventail de problèmes de qualité des données, dans certains cas, une révision manuelle par un data steward est souhaitable. Pour cela, le module Business Steward permet d'indiquer les règles qui déclencheront une révision manuelle et fournit un outil Web pour la révision des enregistrements d'exception. Cela inclut l'accès intégré aux outils tiers, tels que les cartes Bing et les données Experian qui aident les data stewards dans la procédure de révision et de résolution.

Enrichissement des données

Les procédures d'enrichissement des données améliorent vos données avec des informations supplémentaires. L'enrichissement peut être basé sur des données spatiales, des données marketing ou des données d'autres sources à utiliser pour ajouter des détails supplémentaires à vos données. Par exemple, si vous disposez d'une base de données d'adresses de clients, vous pouvez géocoder l'adresse pour déterminer les coordonnées de latitude/longitude de l'adresse et stocker ces coordonnées dans l'enregistrement. Vos données client peuvent ensuite être utilisées pour réaliser divers calculs spatiaux, par exemple pour rechercher l'agence bancaire la plus proche du client. Spectrum™ Technology Platform vous permet d'enrichir vos données de différentes informations, comme le géocodage (à l'aide du module Enterprise Geocoding), l'affectation de juridictions fiscales (à l'aide du module Enterprise Tax), les calculs géospatiaux (à l'aide du module Location Intelligence) et les directions d'itinéraire à pied et à bord d'un véhicule entre des points (à l'aide du module Enterprise Routing).

Master Data Management Hub

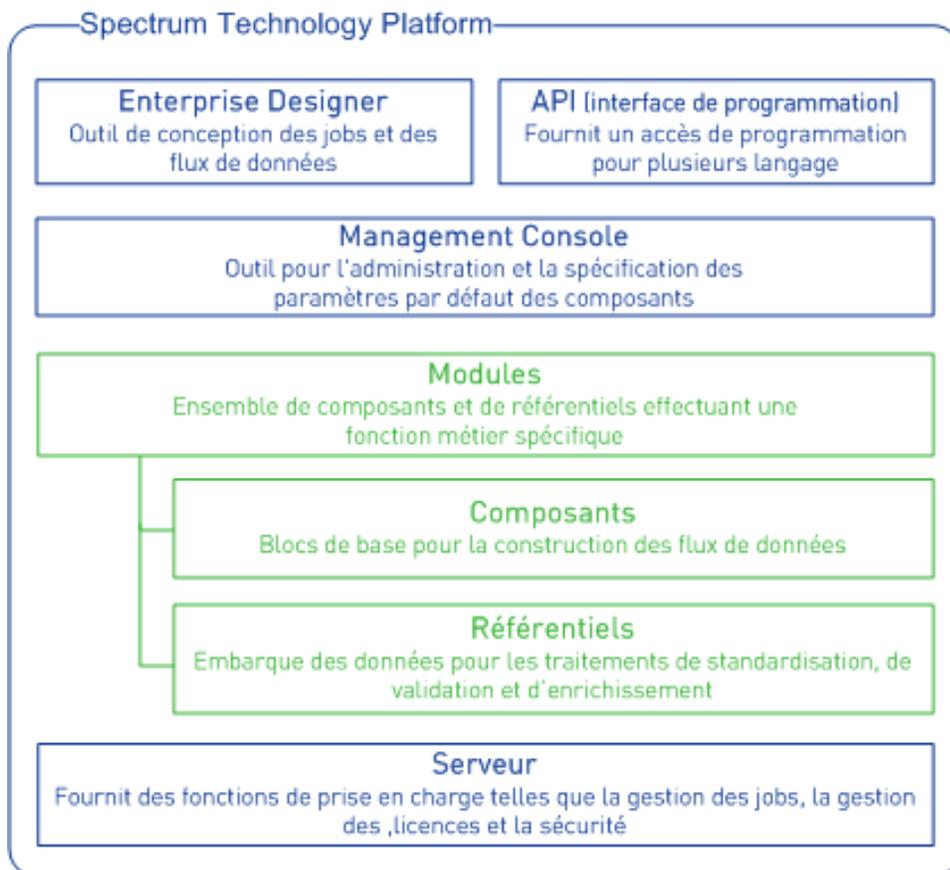
Le hub de gestion de données maître (MDM) permet une modélisation rapide des entités et de leurs relations complexes entre les rôles, les procédures et les interactions. Il fournit des fonctions d'analyse de réseau social intégrées qui vous permettent de comprendre les influenceurs, de prévoir les désabonnements, de détecter les relations non évidentes et les modèles frauduleux, et de fournir des recommandations.

Spectrum™ Technology Platform prend en charge deux approches pour le hub MDM. Dans l'approche de hub maître, les données sont gérées dans une base de données MDM unique et les applications accèdent aux données à partir de la base de données MDM. Dans l'approche registre, les données sont gérées dans chaque application métier et le registre du hub MDM contient les clés permettant de trouver les enregistrements associés. Par exemple, un enregistrement de client peut exister dans une base de données de saisie de commande et une base de données de support clientèle. Le registre MDM contient une clé unique qui permettrait d'accéder aux données client dans les deux emplacements.

Le module Data Hub fournit des fonctions MDM.

Architecture Spectrum™ Technology Platform

Spectrum™ Technology Platform de Pitney Bowes comprend un serveur qui exécute un certain nombre de modules. Ces modules proposent différentes fonctions, comme la validation d'adresses, le géocodage et le parsing avancé, entre autres. Le schéma suivant illustre l'architecture de Spectrum™ Technology Platform.



Serveur

Spectrum™ Technology Platform s'appuie principalement sur le serveur. Le serveur gère le traitement des données, synchronise les données du référentiel et gère la communication. Il met en œuvre les fonctionnalités de gestion des travaux et de sécurité.

Modules

Les modules regroupent des ensembles de fonctionnalités qui exécutent une fonction spécifique. Par exemple, le module Universal Addressing normalise les adresses conformément aux normes postales. Le module Enterprise Tax détermine les juridictions fiscales qui s'appliquent à une adresse donnée. Les modules sont regroupés ensemble pour résoudre des problèmes d'entreprise courants et ils sont distribués sous licence ensemble sous forme de lots.

Composants

Les modules sont constitués de composants qui remplissent une fonction spécifique dans un flux ou sous forme de service. Par exemple, le composant Geocode US Address du module Enterprise Geocoding prend une adresse et renvoie les coordonnées de latitude et longitude correspondantes ; le composant Get City State Province du module Universal Addressing prend un code postal et renvoie la ville et l'état/la province où ce code postal se situe.

Les composants disponibles sur votre système dépendent du lot Spectrum™ Technology Platform pour lequel vous possédez une licence.

Bases de données

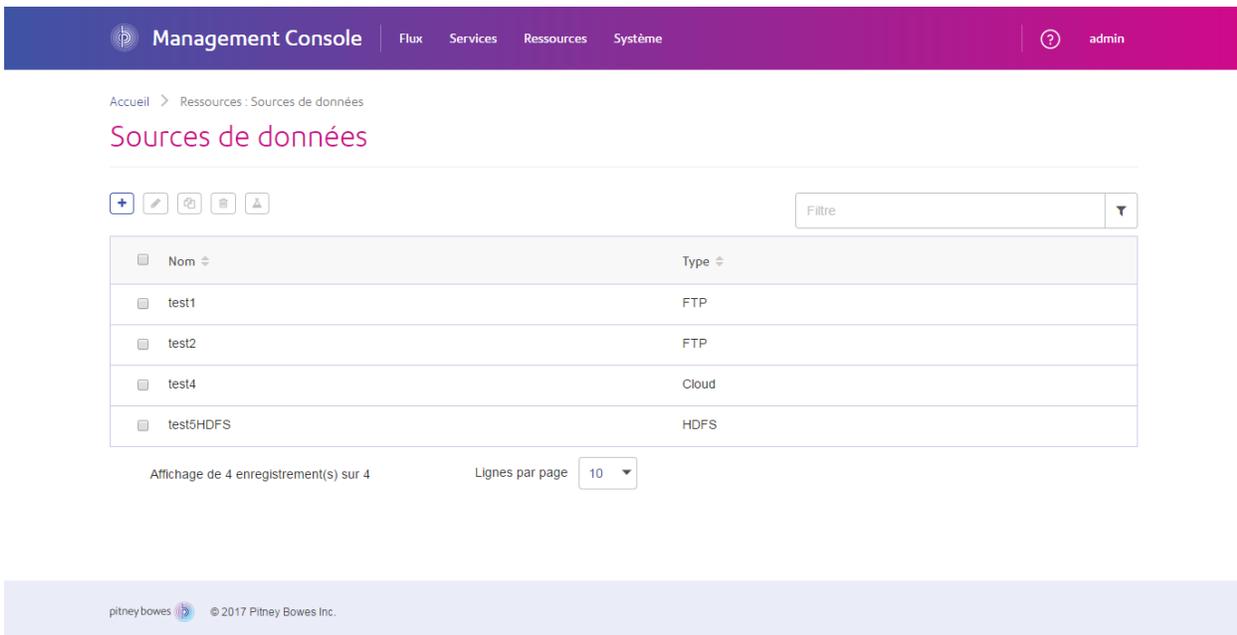
Certains modules dépendent des bases de données contenant des données de référence. Par exemple, le module Universal Addressing a besoin d'avoir accès aux données de U.S. Postal Service afin de vérifier et de normaliser les adresses aux États-Unis. Les bases de données sont installées séparément, et certaines sont mises à jour régulièrement pour vous fournir les données les plus récentes.

Les modules possèdent des bases de données requises et facultatives. Les bases de données facultatives fournissent les données nécessaires pour certaines fonctionnalités qui peuvent améliorer le processus de Spectrum™ Technology Platform.

Management Console

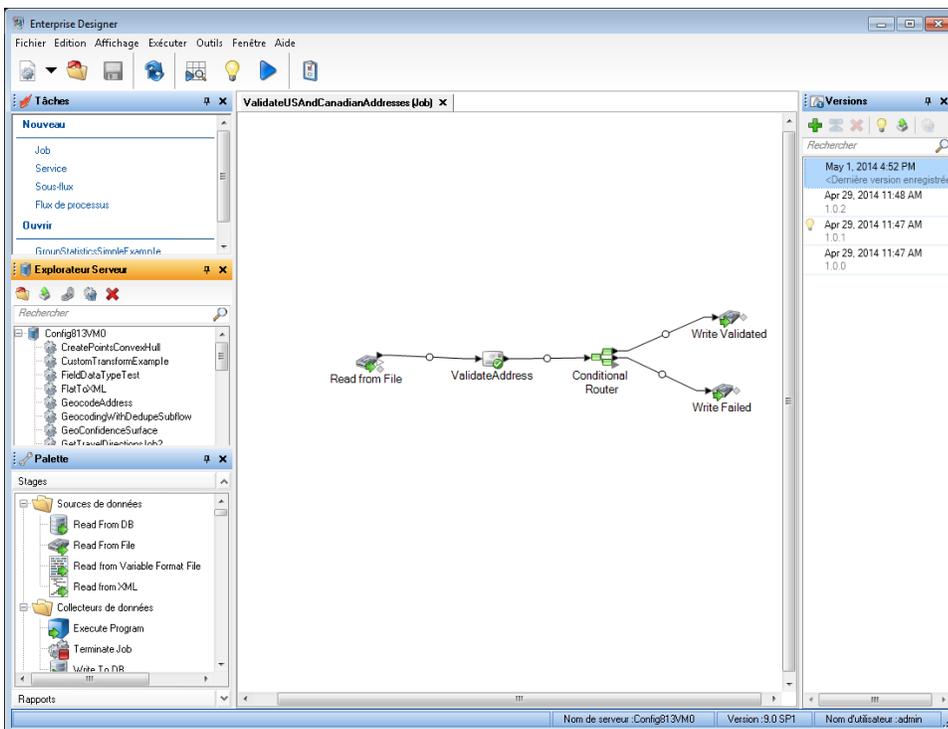
Management Console est un outil d'administration de Spectrum™ Technology Platform. Vous pouvez utiliser Management Console aux fins suivantes :

- Définir les connexions entre Spectrum™ Technology Platform et vos données
- Indiquer les paramètres par défaut pour les services et les flux
- Gérer les comptes utilisateurs, y compris autorisations et mots de passe.
- Afficher des journaux
- Afficher des licences, notamment des informations d'expiration de licence



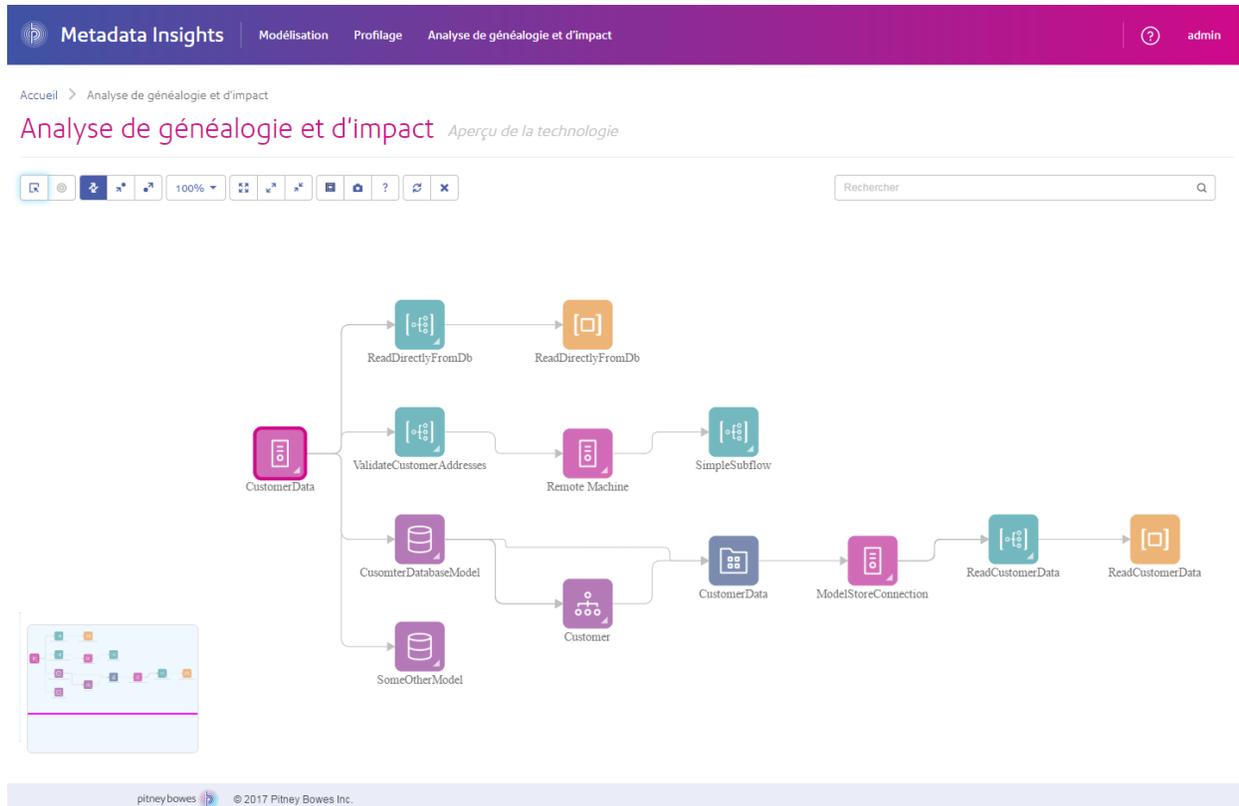
Enterprise Designer

Enterprise Designer est un outil permettant de créer des jobs, des services, des sous-flux et des flux de processus Spectrum™ Technology Platform. Il utilise une interface glisser-déposer facile à utiliser qui vous permet de créer des flux de données complexes au format graphique.



Metadata Insights

Metadata Insights vous donne le contrôle dont vous avez besoin pour fournir à votre entreprise des informations tirées de données précises et dans les temps impartis. Utilisez Metadata Insights pour développer des modèles de données, afficher le dataflow de la source vers l'application commerciale et évaluer la qualité de vos données grâce au profilage. Grâce à ces informations, vous pouvez identifier les ressources de données à utiliser pour répondre à des questions professionnelles spécifiques, adapter et optimiser les processus afin d'améliorer l'utilité et la cohérence des données au sein de l'entreprise toute entière et résoudre les problèmes de données.



Services Web et API

Vous pouvez intégrer les fonctionnalités Spectrum™ Technology Platform dans vos applications à l'aide de services Web et d'API de programmation. Ces interfaces fournissent une intégration simple, un traitement rationalisé des enregistrements et une prise en charge de la rétrocompatibilité des versions futures.

L'API Spectrum™ Technology Platform est disponible pour ces langages :

- C
- C++
- COM
- Java
- .NET

Les services Web sont disponibles via SOAP et REST.

Modules et composants

Tableau 12 : Modules et composants

Module	Description	Composants
Module Advanced Matching	Rapproche les enregistrements dans des fichiers d'entrée et/ou entre ceux-ci.	Best of Breed Candidate Finder Duplicate Synchronization Filtre Interflow Match Intraflow Match Match Key Generator Transactional Match
Module Business Steward	Identifie les enregistrements d'exception et fournit un outil basé sur un navigateur pour vérifier manuellement les enregistrements d'exception.	Exception Monitor Read Exceptions Write Exceptions
Country Identifier	Prend un nom de pays ou une combinaison de code postal et d'État/province puis renvoie le code pays ISO à deux chiffres, le code de l'UPU (Union postale universelle) à trois caractères, et le nom de pays en français.	Country Identifier

Module	Description	Composants
Metadata Insights	Vous donne le contrôle dont vous avez besoin pour fournir à votre entreprise des informations tirées de données précises et dans les temps impartis. Développe des modèles de données, vous offre une vue du dataflow de la source vers une application d'entreprise et évalue la qualité de vos données grâce au profiling. Vous permet d'identifier les ressources de données que vous devez utiliser pour répondre aux questions commerciales données et optimiser les processus afin d'améliorer l'utilité et la cohérence des données de votre entreprise.	Logical Models et Physical Models Model Store Profil Lineage et Analyse d'impact
Module Data Hub	Lie et analyse les données en identifiant leurs relations et tendances.	Write to Hub Read From Hub Query Hub Graph Visualization
Module Data Integration	Fournit des fonctionnalités utiles pour l'entreposage de données, la gestion de la qualité des données, l'intégration système et la migration.	Field Selector Generate Time Dimension Query Cache Write To Cache
Module Data Normalization	Élimine les incohérences dans les données.	Advanced Transformer Open Parser Table Lookup Transliterator
Enterprise Data Integration	Se connecte aux données de différentes sources pour une variété de besoins d'entreprise, notamment l'entreposage de données, la gestion de la qualité des données, l'intégration système et la migration.	Call Stored Procedure Field Selector Generate Time Dimension Query Cache Write To Cache

Module	Description	Composants
Module Enterprise Geocoding	Détermine les coordonnées géographiques d'une adresse. Détermine également l'adresse d'une latitude et longitude données.	<p>Geocode Address AUS</p> <p>Geocode Address GBR - déprécié. Utilisez le stage Geocoding du module Global Geocoding.</p> <p>Geocode Address Global</p> <p>Geocode Address World</p> <p>Geocode US Address</p> <p>GNAF PID Location Search</p> <p>Reverse APN Lookup</p> <p>Reverse Geocode Address Global</p> <p>Reverse Geocode US Location</p>
Module Enterprise Routing	Obtient les directions de conduite ou de marche, calcule le temps et la distance de conduite, et identifie les emplacements dans un certain temporel et géographique à partir d'un point de départ.	<p>Obtenir les données de l'itinéraire</p> <p>Get Travel Boundary</p> <p>Obtenir la matrice de coût de transport</p> <p>Obtenir les directions de trajet</p> <p>Mise à jour permanente</p>
Module Enterprise Tax	Détermine les juridictions fiscales qui s'appliquent à un lieu donné.	<p>Assign GeoTAX Info</p> <p>Calculate Distance</p>
Module FCC Screening	Aide les banques et les institutions financières à détecter efficacement les infractions financières, à réduire les faux positifs et à maintenir une capacité de détection solide, comme l'exigent les organismes de réglementation.	<p>Groupes de partis</p> <p>Listes</p> <p>Screen</p> <p>Alertes</p>
Module GeoConfidence	Détermine la probabilité qu'une adresse ou une intersection de rues se situe à l'intérieur d'une zone donnée.	<p>Geo Confidence Surface</p> <p>CreatePointsConvexHull</p>

Module	Description	Composants
Module Global Addressing	Fournit une normalisation et une validation d'adresse améliorées. Suggère également automatiquement des adresses à mesure de la saisie et renvoie immédiatement des candidats en fonction de votre saisie. Divise les chaînes d'adresse postale en éléments d'adresse individuels à l'aide de techniques Machine Learning.	Global Address Parser Global Address Validation Global Type Ahead
Module Global Geocoding	Détermine les coordonnées géographiques d'une adresse. Détermine également l'adresse d'une latitude et longitude données. Le géocodage interactif est une fonction à saisie anticipée de GGM. Key Lookup utilise une clé pour géocoder des adresses.	Global Geocode Global Reverse Geocode Global Interactive Geocoding Global Key Lookup
Global Sentry	Tentatives de rapprochement des transactions avec les listes de surveillance fournies par les gouvernements qui contiennent des données de différents pays.	Global Sentry Global Sentry Address Check Global Sentry ID Number Check Global Sentry Name Check Global Sentry Other Data Check
Module Location Intelligence	Exécute l'analyse Point In Polygon (Point In Polygon) et l'analyse radiale par rapport à une variété de bases de données géospatiales.	Closest Site Find Nearest Point In Polygon Query Spatial Data Read Spatial Data Spatial Calculator Spatial Union Write Spatial Data

Module	Description	Composants
Module SAP	Permet à Spectrum™ Technology Platform d'être en interface avec les applications du Module SAP Customer Relationship Management.	<ul style="list-style-type: none"> SAP Generate Match Key SAP Generate Match Score SAP Generate Search Key SAP Generate Search Key Constant SAP Generate Search Key Metaphone SAP Generate Search Key Substring SAP Validate Address With Candidates
Module Universal Addressing	Normalise et valide des adresses selon les normes des autorités postales.	<ul style="list-style-type: none"> Get Candidate Addresses Get City State Province Get Postal Codes Validate Address Validate Address AUS Validate Address Global
Module Universal Name	Analyse des noms personnels, des noms d'entreprises, des adresses, et plusieurs autres termes et abréviations.	<ul style="list-style-type: none"> Name Parser (déprécié) Name Variant Finder Open Name Parser

Notices

© 2018 Pitney Bowes. Tous droits réservés. MapInfo et Group 1 Software sont des marques commerciales de Pitney Bowes Software Inc. Toutes les autres marques et marques commerciales sont la propriété de leurs détenteurs respectifs.

Avis USPS®

Pitney Bowes Inc. détient une licence non exclusive pour la publication et la vente de bases de données ZIP + 4® sur des supports optiques et magnétiques. Les marques de commerce suivantes appartiennent à United States Postal Service : CASS, CASS Certified, DPV, eLOT, FASTforward, First-Class Mail, Intelligent Mail, LACS^{Link}, NCOA^{Link}, PAVE, PLANET Code, Postal Service, POSTNET, Post Office, RDI, Suite^{Link}, United States Postal Service, Standard Mail, United States Post Office, USPS, ZIP Code et ZIP + 4. Cette liste de marques de commerce appartenant à U.S. Postal Service n'est pas exhaustive.

Pitney Bowes Inc. détient une licence non exclusive de USPS® pour le traitement NCOA^{Link®}.

Les prix des produits, des options et des services de Pitney Bowes Software ne sont pas établis, contrôlés ni approuvés par USPS® ni par le gouvernement des États-Unis. Lors de l'utilisation de données RDI™ pour déterminer les frais d'expédition de colis, le choix commercial de l'entreprise de distribution de colis à utiliser n'est pas fait par USPS® ni par le gouvernement des États-Unis.

Fournisseur de données et avis associés

Les produits de données contenus sur ce support et utilisés au sein des applications Pitney Bowes Software sont protégés par différentes marques de commerce et par un ou plusieurs des copyrights suivants :

© Copyright United States Postal Service. Tous droits réservés.

© 2014 TomTom. Tous droits réservés. TomTom et le logo TomTom logo sont des marques déposées de TomTom N.V.

© 2016 HERE

Source : INEGI (Instituto Nacional de Estadística y Geografía)

Basées sur les données électroniques © National Land Survey Sweden.

© Copyright United States Census Bureau

© Copyright Nova Marketing Group, Inc.

Des portions de ce programme sont sous © Copyright 1993-2007 de Nova Marketing Group Inc. Tous droits réservés.

© Copyright Second Decimal, LLC

© Copyright Canada Post Corporation

Ce CD-ROM contient des données provenant d'une compilation dont Canada Post Corporation possède le copyright.

© 2007 Claritas, Inc.

Le jeu de données Geocode Address World contient des données distribuées sous licence de GeoNames Project (geonames.org) fournies sous la licence Creative Commons Attribution License (« Attribution License ») à l'adresse : <http://creativecommons.org/licenses/by/3.0/legalcode>. Votre utilisation des données GeoNames (décrites dans le Manuel de l'utilisateur Spectrum™ Technology Platform) est régie par les conditions de la licence Attribution License et tout conflit entre votre accord avec Pitney Bowes Software, Inc. et la licence Attribution License sera résolu en faveur de la licence Attribution License uniquement s'il concerne votre utilisation des données GeoNames.



3001 Summer Street
Stamford CT 06926-0700
USA

www.pitneybowes.com