Pitney Bowes Software

DB Crawler

User Guide

Spectrum™ Technology Platform

Data Discovery Module

Version 8

April 2013

Global IDs Product Suite® DB Crawler Manual

Copyright © 2001 – 2013 Global IDs Inc.

All rights reserved. Printed in USA.

This software and documentation contain proprietary information of Global IDs Inc., they are provided under a license agreement containing restrictions on use and disclosure and is also protected by copyright law. Reverse engineering of the software is prohibited. No part of this document may be reproduced or transmitted in any form, by any means (electronic, photocopy, recording, or otherwise) without prior consent of Global IDs Inc.

Use, duplication, or disclosure of the Software by the U.S. Government is subject to the restrictions set forth in the applicable software license agreement as provided in DFARS 227.7202-1 (a) and 227.7702-3 (a) (1995), DFARS 252.227-7013(c)(1)(ii) (OCT 1988), FAR 12.212(a) (1995), FAR 52.227-19, or FAR 52.227-14(ALT III), as applicable.

This information in this document is subject to change without notice. If you find any problems in the documentation, please report them to us in writing. Global IDs Inc. does not warrant that this documentation is error free.

Global IDs Product Suite is the registered trademark of the Global IDs Inc. in the United States and in jurisdictions throughout the world. All other company and product names may be trade names or trademarks of their respective owners.

Global IDs products contain GUI Components from JIDE™ Software, yFiles from yWorks™, JGo: GoDiagram™ from Northwoods Software Corporation, StelsEngine from J-Stels™ Software, EasyCharts™ from ObjectPlanet, Install4j from ej-technologies GmbH, Nmap™ - Free Security Scanner For Network Exploration & Security Audits(originally written by Gordon Lyon). Java™ is a registered trademark of Sun Microsystems and Apache Derby™ from Apache Software Foundation. All rights reserved.

DISCLAIMER: Global IDs Inc. provides this documentation ‘as is’ without warranty of any kind, either express or implied, including, but not limited to, the implied warranties of non-infringement , merchantability, or use for a particular purpose. The information provided in this documentation may include many technical inaccuracies or typographical errors. Global IDs could make improvements and/or changes in the products described in this documentation at many time without notice.

Contents

List of Figures 4
About DB Crawler Application 6
Discovering Tables 6
Viewing discovered tables 7
Discovering Metadata 8
Viewing Metadata 9
Discovering Relationships 10
Viewing Relationships 10
Running Data Discovery and Profiling Workflow 12
Generating Reports 13
Index 15

List of Figures

Figure 1: Discover Tables - Step 1 ..... *6*
Figure 2: Viewing discovered tables - hierarchy pane *7*
Figure 3: Viewing discovered tables - Tables tab *8*
Figure 4: Discover Metadata - Step 1 ..... *8*
Figure 5: Viewing metadata of tables *9*
Figure 6: Viewing record count and column count of tables *9*
Figure 7: Viewing documented relationships *11*
Figure 8: Viewing documented relationships of a table *11*
Figure 9: Data Discovery and Profiling Workflow *12*
Figure 10: Landscape Metadata Report *13*
Figure 11: Metadata Report *13*

DB Crawler

What it includes:

- About DB Crawler Application
- Discovering Tables
- Discovering Metadata
- Discovering Relationships
- Running Data Discovery and Profiling Workflow
- Generating Reports

About DB Crawler Application

DB Crawler helps you to scan structured/relational data sources. It discovers tables, metadata and documented or explicitly specified Primary Key-Foreign Key relationships in your database. This information would be used by other modules to analyze, clean, monitor and integrate your data. It populates a metadata repository with the information scanned, that can be searched by business and technical users.

Discovering Tables

Actions in grey need to be completed sequentially before the current action is performed. The current action is marked in red.

Pre-requisites

- Add Connection
- Discover Tables

This topic teaches you to discover database tables and views from a schema.
To discover tables:

- Right click a schema from the left hierarchy pane and select **Discover Tables** option. **Discover Tables** wizard appears.
- On Step 1 of the wizard:
  - Enter **Username**
  - Enter **Password**
  - Select the appropriate checkbox to
    - discover views
    - discover synonymous tables
    - find row count from system tables

Figure 1: Discover Tables - Step 1

Click **Next**.

- On Step 2 of the wizard:
  - Select the appropriate radio button to run the task now or schedule it for a later time
  - Assign a priority level by choosing between five priority options **(Background, Low, Normal, High, Critical)**
  - Select the **Run tasks in parallel** check box to enable running more than one task at the same time and specify **Maximum tasks to run in parallel**
  - You can also add time restrictions by clicking on the **Add Time Restrictions** link

Click **Finish**.

Viewing discovered tables

Once tables in a schema are discovered you can expand the schema from the hierarchy pane to view all tables.

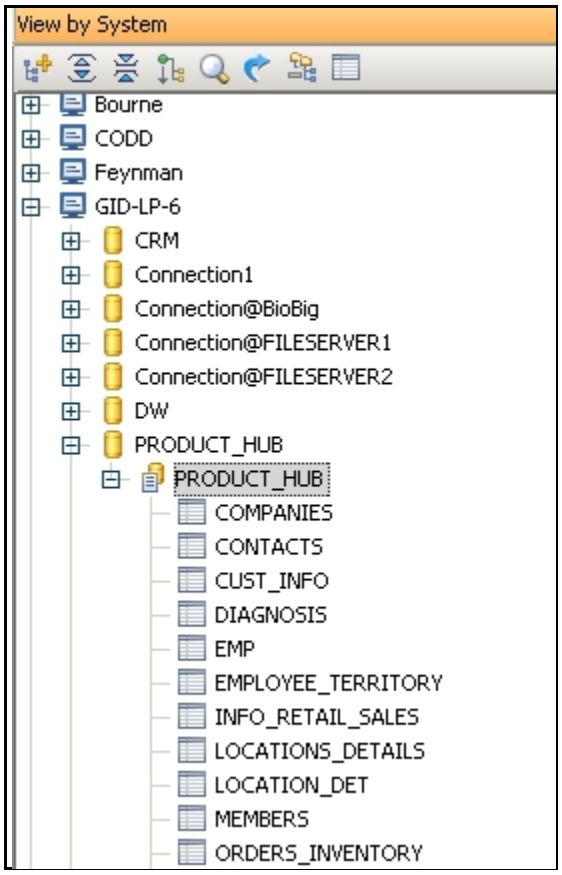

Figure 2: Viewing discovered tables - hierarchy pane

You can view table names, record count, column count and table descriptions (if any) from the **Tables** tab in the display pane on the right when a schema is selected from the left hierarchy pane.

Tables | SQL Editor | Relationships | Table Descriptions | Schema Visualization

Table Name	Record Count	Column Count	Description	Metadata Discovered	Timestamp
COMPANIES	1575	14		✓	2010 02 17, 08:32:47
CONTACTS	2957	8		✓	2010 02 17, 08:32:47
CUST_INFO	107	12	Master Table	✓	2010 02 17, 08:32:47
DIAGNOSIS	58	12		✓	2011 06 23, 10:26:45
EMP	9	17	Employee Table	✓	2010 02 17, 08:32:47
EMPLOYEE_TERRITORY	49	2		✓	2010 02 17, 08:32:47
INFO_RETAIL_SALES	4248	18		✓	2010 02 17, 08:32:47
LOCATIONS_DETAILS	4	2		✓	2010 02 17, 08:32:47
LOCATION_DET	53	3		✓	2010 02 17, 08:32:47
MEMBERS	52	12	Member Master Table	✓	2011 06 23, 10:26:42
ORDERS_INVENTORY	7090	19		✓	2010 02 17, 08:32:47

Figure 3: Viewing discovered tables - Tables tab

Discovering Metadata

Actions in grey need to be completed sequentially before the current action is performed. The current action is marked in red.

Pre-requisites

- Add Connection
- Discover Table
- Discover Metadata

This topic teaches you to discover metadata of database tables and views.

To discover metadata:

- Right click the desired schema from the left hierarchy pane and select **Discover Metadata** option. **Discover Metadata** wizard appears
- On Step 1 of the wizard:
  - Select tables to discover their metadata

Figure 4: Discover Metadata - Step 1

Click **Next**.

- On Step 2 of the wizard:
  - select the appropriate radio button to run the task now or schedule it for a later time
  - Assign a priority level by choosing between five priority options **(Background, Low, Normal, High, Critical)**
  - Select the **Run tasks in parallel** check box to enable running more than one task at the same time and specify **Maximum tasks to run in parallel**
  - You can also add time restrictions by clicking on the **Add Time Restrictions** link
  - Click **Finish**.

Viewing Metadata

To view discovered metadata of a table, select the table from hierarchy pane and click on corresponding **Metadata** tab. All columns along with metadata information are displayed in a list view.

Metadata | Data | Relationships | Indexes | Query | Descriptions

Name	Data Type	Size	Decimal digits	Primary Key	Allows Nulls	Default Value	Order
EMPLOYEE_ID	NUMBER	6	0	☑	☐		1
FIRST_NAME	VARCHAR2	20	N/A	☐	☑		2
LAST_NAME	VARCHAR2	25	N/A	☐	☐		3
EMAIL	VARCHAR2	25	N/A	☐	☐		4
PHONE_NUMBER	VARCHAR2	20	N/A	☐	☑		5
HIRE_DATE	DATE	7	N/A	☐	☐		6
JOB_ID	VARCHAR2	10	N/A	☐	☐		7
SALARY	NUMBER	8	2	☐	☑		8
COMMISSION_PCT	NUMBER	2	2	☐	☑		9
MANAGER_ID	NUMBER	6	0	☐	☑		10
DEPARTMENT_ID	NUMBER	4	0	☐	☑		11
TEMP	NUMBER	5	0	☐	☑		12

Figure 5: Viewing metadata of tables

To view the column count and documented descriptions of all tables in a schema, select a schema from the hierarchy pane and click on the corresponding **Tables** tab from the display pane.

Tables | SQL Editor | Relationships | Table Descriptions | Schema Visualization

Table Name	Record Count	Column Count	Description	Metadata Discovered	Timestamp
COMPANIES	1575	14		☑	2010 02 17, 08:32:47
CONTACTS	2957	8		☑	2010 02 17, 08:32:47
CUST_INFO	107	12	Master Table	☑	2010 02 17, 08:32:47
DIAGNOSIS	58	12		☑	2011 06 23, 10:26:45
EMP	9	17	Employee Table	☑	2010 02 17, 08:32:47
EMPLOYEE_TERRITORY	49	2		☑	2010 02 17, 08:32:47
INFO_RETAIL_SALES	4248	18		☑	2010 02 17, 08:32:47
LOCATIONS_DETAILS	4	2		☑	2010 02 17, 08:32:47
LOCATION_DET	53	3		☑	2010 02 17, 08:32:47
MEMBERS	52	12	Member Master Table	☑	2011 06 23, 10:26:42
ORDERS_INVENTORY	7090	19		☑	2010 02 17, 08:32:47

Figure 6: Viewing record count and column count of tables

Discovering Relationships

Actions in grey need to be completed sequentially before the current action is performed. The current action is marked in red.

Pre-requisites

- Add Connection
- Discover Tables
- Discover Metadata
- Discover Relationships

This topic teaches you to discover documented relationships among database tables and views.

To discover relationships:

- Right click the desired schema from the left hierarchy pane and select **Discover Relationships** option. **Discover Relationships** wizard appears.
- In the wizard:
  - select the appropriate radio button to run the task now or schedule it for a later time
  - Assign a priority level by choosing between five priority options **(Background, Low, Normal, High, Critical)**
  - Select the **Run tasks in parallel** check box to enable running more than one task at the same time and specify **Maximum tasks to run in parallel**
  - You can also add time restrictions by clicking on the **Add Time Restrictions** link

Click **Finish**.

Viewing Relationships

After the documented relationships are discovered you can select the schema from the hierarchy pane and corresponding **Relationships** tab from the display pane to view how tables in a schema are related.

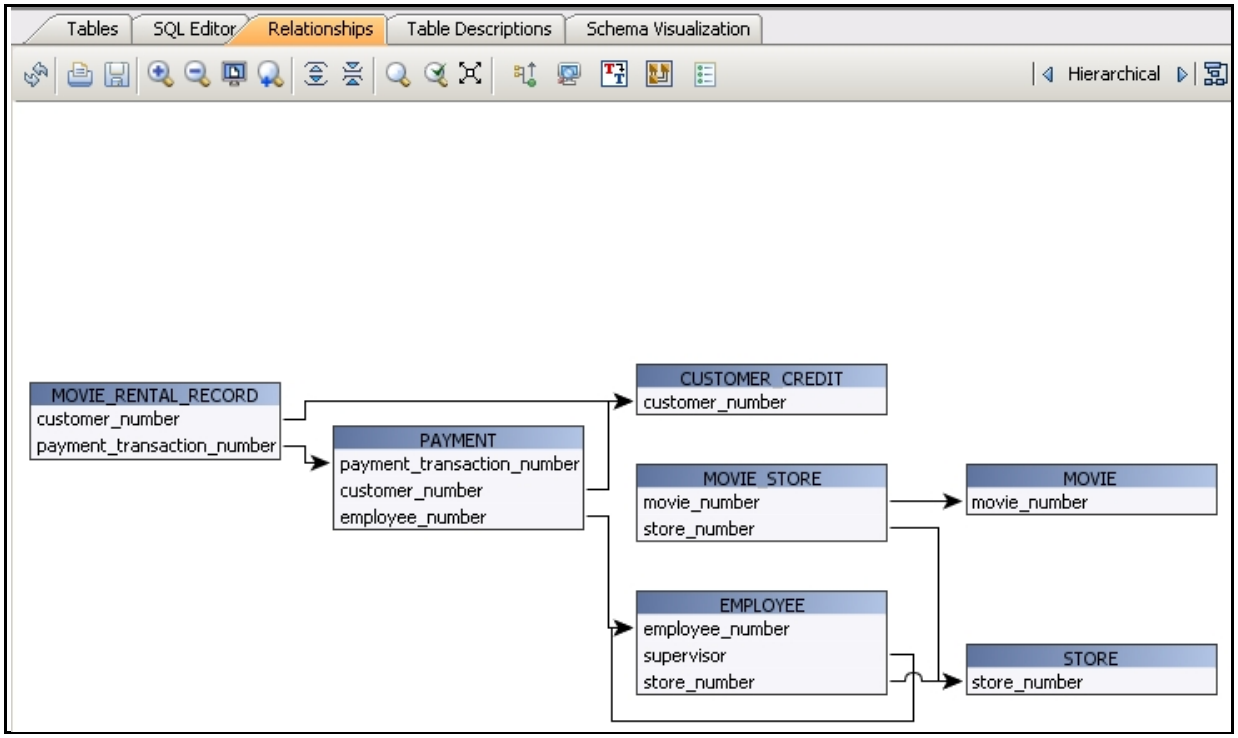

Figure 7: Viewing documented relationships

You can also select a table from the hierarchy pane to view how other tables are related to it.

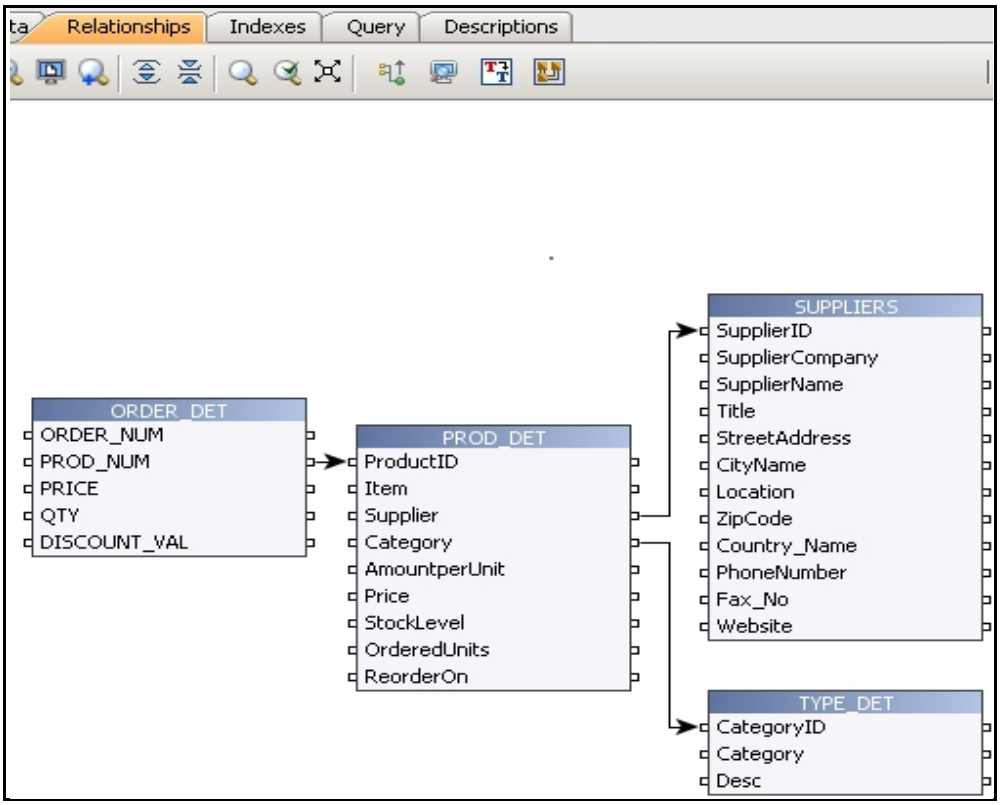

Figure 8: Viewing documented relationships of a table

Running Data Discovery and Profiling Workflow

In order to reduce manual intervention, this workflow has been introduced in DB Crawler to automate the process of data discovery and profiling. Instead of running every data discovery and profiling activity manually, all you have to do is to run this workflow by right-clicking on a desired schema and selecting **Run Data Discovery and Profiling Workflow**. The software will do everything for you automatically. The process completes on the basis of the parameters provided by the user before the workflow begins its execution.

The flowchart displays all the actions the workflow performs:

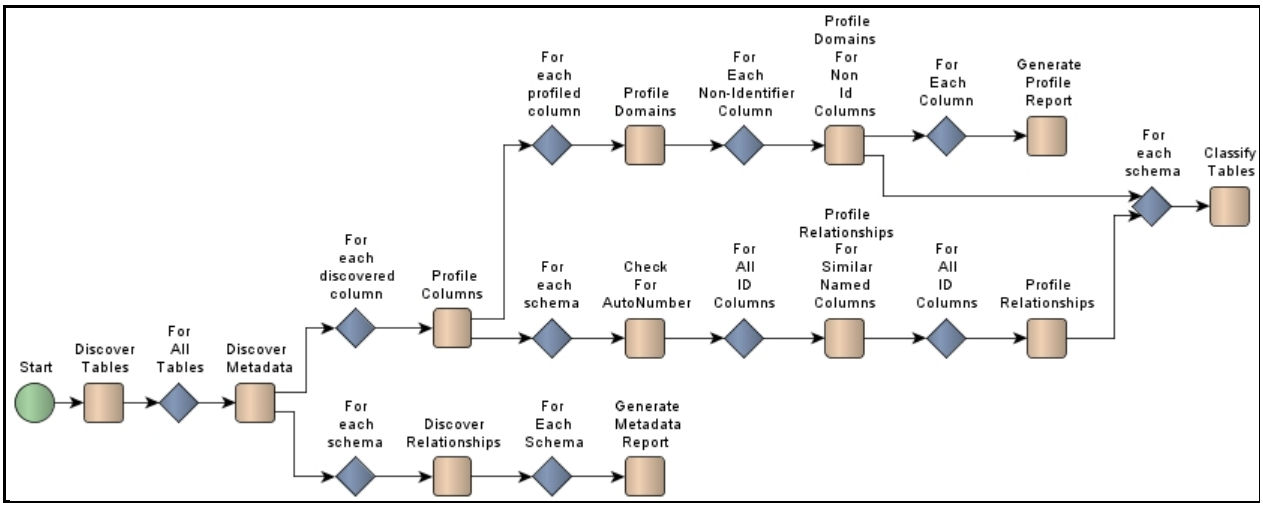

Figure 9: Data Discovery and Profiling Workflow

Deliverables

- Metadata Report
- Profile Report

Note: You can also run this workflow with data and metadata monitoring by right-clicking on a desired schema and selecting the **Run Data Discovery and Profiling Workflow (with Monitoring)** option from DB Crawler.

Generating Reports

Reports can be generated in DB Crawler from the **Reports** tab which forms a part of the left vertical tabs. The following reports can be created in HTML or PDF from DB Crawler.

- **Landscape Metadata Report** - Select the root node of the left hierarchy to create this report.

Landscape Metadata Report

Description : Landscape Metadata Report
Last Modified : 16 Aug 2012, 05:19:42
Report Created by : debashmita.roy@globalids.com
Created on : 16 Aug 2012, 05:12:26

Number of Systems : 7
Number of Databases : 15
Number of Schemas : 358
Number of Tables : 2334
Number of Columns : 13896
Number of Records : 50,896,247

Database Summary

System	Instance Name	Database Type	#Schemas	#Scanned Schemas	#Tables	#Columns	#Records
Bourne(192.168.33.101)	DEMO1	Derby	14	0	0	0	0
herschel(192.168.33.112)	globalid	Derby	12	1	309	3,455	60,892
harvey(192.168.33.113)	globalid	Derby	12	1	309	3,455	60,785
TEST-LINUX(192.168.33.14)	SAMPLES	Cache	27	24	142	267	76,985
GID-LP-6(192.168.33.191)	CRM	Derby	15	3	49	239	733,833
GID-LP-6(192.168.33.191)	SALESDB	Derby	14	3	49	503	737,530
GID-LP-6(192.168.33.191)	PRODUCT_HUB	Derby	13	1	18	180	22,659
GID-LP-6(192.168.33.191)	DW	Derby	14	1	9	68	0
GID-LP-6(192.168.33.191)	FILESERVER1	Derby	12	1	4	46	7,223
GID-LP-6(192.168.33.191)	FILESERVER2	Derby	12	1	4	46	7,223
GID-LP-6(192.168.33.191)	BioBig	Derby	13	2	69	272	3,810,527
GID-LP-6(192.168.33.191)	SALESDB	Derby	12	1	20	197	29,207
Feynman(192.168.33.8)	globalid	Oracle	89	12	634	2,837	9,661,249
CODD(192.168.33.9)	coddgid	Oracle	78	5	629	2,298	35,345,329
CODD(192.168.33.9)	KDRM	MSSQLServer	21	1	89	33	342,805

Figure 10: Landscape Metadata Report

- **Metadata Report** - Select the schema from the left hierarchy to create the schema metadata report.

DEMO4 Metadata Report

Metadata Retrieved On :21 Jul 2010, 13:14:00
Time of Report Creation :16 Aug 2012, 05:26:50
Description : Report on documented table and column metadata of DEMO4
Created by :debashmita.roy@globalids.com

System IP :192.168.33.8
System Name :Feynman
Database Type :Oracle
Database Instance Name :globalid
Schema Name :DEMO4
Total Number of Columns :146

Database Product Version Information :

Oracle9i Enterprise Edition Release 9.2.0.1.0 - Production With the Partitioning, OLAP and Oracle Data Mining options JServer Release 9.2.0.1.0 - Production

Number of table(s):44

Table Name	No. of Records	Total Columns	Table Description
AGENT	0	2	
ASSOCIATION	0	4	
BILL_TO_CUSTOMER	0	2	
COMPETITOR	0	2	
CONTACT	0	2	
CONTRACTOR	0	2	
CUSTOMER	0	2	
CUSTOMER_RELATIONSHIP	0	5	
CUST_MASTER	2956	8	
DEPARTMENT	0	2	
DISTRIBUTION_CHANNEL	0	2	
DISTRIBUTION_CHANNEL_RELATIONS	0	5	
DISTRIBUTOR	0	2	
DIVISION	0	2	
EMPLOYEE	0	2	

Figure 11: Metadata Report

NOTE: DB Crawler is one of the main applications in the Networked Databases layer. The software has the ability to generate separate DB Crawler applications for each database type you have connected to using the software. This means, if you have connected to 10 different types of databases using the software, a total of 10 DB Crawler applications will be created dynamically, one for each database type. For example, if you have connected to *Oracle* database, the software will dynamically generate an application named *Oracle DB Crawler* which will show all the Oracle database instances in your landscape only. These dynamically created applications are visible from the Networked Databases layer in Application Manager.

Index

A

Add Time Restrictions 7, 9, 10

D

Discovering Metadata
- discover metadata of database tables and views 8

Discovering Relationships
- discover documented relationships 10

Discovering Tables
- discover synonymous tables 6
- discover views 6
- find row count from system tables 6

G

Generating Reports
- Landscape Metadata Report 13
- Metadata Report 13
- Reports 13

V

Viewing discovered tables
- Tables tab 7

Viewing Metadata
- Metadata tab 9
- Tables tab 9

Viewing Relationships
- Relationships tab 10

W

Workflow
- Flowchart 12
- Running Data Discovery and Profiling Workflow 12