NetIQ® Identity Manager

Administrator's Guide for Analyzer

February 2018

Legal Notice

For information about NetIQ legal notices, disclaimers, warranties, export and other use restrictions, U.S. Government restricted rights, patent policy, and FIPS compliance, see https://www.netiq.com/company/legal/.

Copyright (C) 2018 NetIQ Corporation. All rights reserved.

Contents

About this Book and the Library **5**

1 Overview **7**

Installing Analyzer . . . 7
Understanding Analyzer's User Interface . . . 7
Welcome View . . . 8
Project View . . . 8
Schema Map Editor . . . 10
Data Browser . . . 12
Analysis Profile Editor . . . 16
Analysis Results View . . . 17
Data Set Instances View . . . 18
Analysis Result Sets View . . . 19
Matching Profile Editor . . . 20
Uniqueness Results View . . . 22
Matching Results View . . . 23
IDS Trace View . . . 24
Additional Resources . . . 25

2 Configuring Analyzer **27**

Preferences . . . 27
Connections . . . 27
Data Browser . . . 27
Database Settings . . . 28
Matching. . . . 29
Reporting . . . 30
Connection Settings . . . 33
General Connection Info . . . 33
IDS Configuration . . . 33
Flat File Settings . . . 34
File Path . . . 34
Delimiter Settings . . . 34
Field Names . . . 35
Data Set Settings . . . 35
Data Browser Settings. . . . 36
IDS Trace Settings . . . 37

3 Using Analyzer **39**

Creating a Project . . . 39
Creating a New Identity Manager Project . . . 39
Importing an Existing Project . . . 39
Importing from Additional Sources . . . 40
Adding a Connection Profile. . . . 42
Creating a Connection Profile . . . 42
Importing a Connection Profile Information from Designer . . . 44

Importing a Connection Profile from a File . . . 45
Creating a Flat File Profile. . . . 45
Modifying a Schema Map. . . . 48
Creating a Data Set Definition . . . 50
Viewing Application Data . . . 52
Importing Data. . . . 52
Exporting Data . . . 52
Inspecting Data . . . 53
Filtering Data . . . 54
Inspecting the Current Data Set . . . 54
Modifying Data . . . 55
Analyzing Data. . . . 56
Creating and Customizing an Analysis Profile . . . 56
Analyzing a Data Set . . . 64
Cleaning Data . . . 65
Data Cleaning Variables. . . . 65
Data Cleaning Script Examples . . . 65
Managing Data Cleaning Scripts. . . . 66
Populating or Deleting Values. . . . 66
Comparing Data . . . 67
Evaluating a Data Set for Unique Values . . . 67
Matching Values between Data Sets . . . 69
Auditing and Reporting. . . . 71
Data Browser Audit . . . 72
Data Browser Reports . . . 72
Analysis Reports. . . . 72
Data Comparison Reports . . . 73

4 Data Security with Analyzer 75

Data in Analyzer Projects . . . 75
Data Transmission and Logging . . . 76

About this Book and the Library

The *Analyzer Administration Guide* provides conceptual information about NetIQ Analyzer for Identity Manager.

Intended Audience

This guide is intended for Identity Manager administrators and consultants.

Other Information in the Library

For more information about the library for Identity Manager, see the Identity Manager documentation website.

1 Overview

NetIQ Analyzer for Identity Manager is an Eclipse-based identity management toolset that helps you ensure that internal data quality policies are adhered to by providing data analysis, data cleansing, data reconciliation, and data monitoring/reporting. Analyzer lets you analyze, enhance, and control all data stores throughout the enterprise.

Before implementing an Identity Management solution, designers spend a significant amount of time analyzing and cleaning identity data, and modeling business rules to create identity data replication and synchronization policies that guarantee the data remains in a reliable state. Additionally, when an Identity solution is put into place, customers must verify and reconcile that these processes are performing as intended to maintain consistent and reliable data. The goal of Analyzer is to create a set of tools to resolve data quality issues and improve the Identity Manager deployment process.

Industry analysts note that Identity Management projects spend three to eight times more on design and implementation than they do on the software. Analyzer directly attacks these project-related costs by providing a powerful environment for cleaning and preparing identity data in order to streamline identity infrastructure implementations.

This section includes the following topics:

- "Installing Analyzer" on page 7
- "Understanding Analyzer's User Interface" on page 7
- "Additional Resources" on page 25

Installing Analyzer

For information about installing Analyzer, see Installing Analyzer in the *NetIQ Identity Manager Setup Guide for Linux* or Installing Analyzer in the *NetIQ Identity Manager Setup Guide for Windows*.

Understanding Analyzer's User Interface

Analyzer delivers its functionality through a new perspective for NetIQ Designer for Identity Manager. Much of the user interface remains the same, but individual UI elements provide different features in the Analyzer perspective than in the standard Designer perspective. The primary Analyzer UI elements include the following:

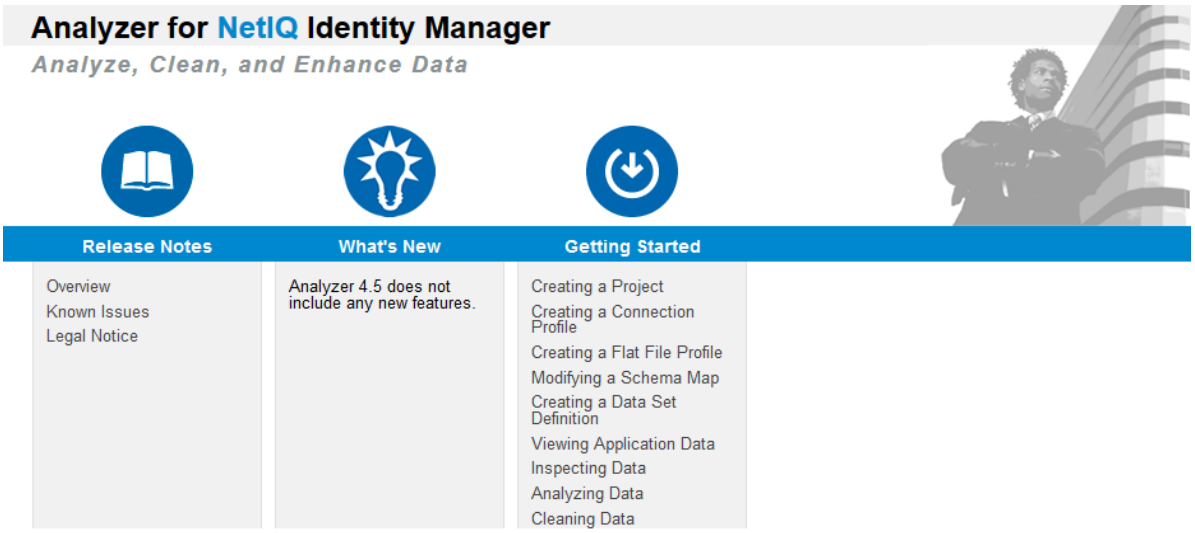
- "Welcome View" on page 8
- "Project View" on page 8
- "Schema Map Editor" on page 10
- "Data Browser" on page 12
- "Analysis Profile Editor" on page 16
- "Analysis Results View" on page 17

- “Data Set Instances View” on page 18
- “Analysis Result Sets View” on page 19
- “Matching Profile Editor” on page 20
- “Uniqueness Results View” on page 22
- “Matching Results View” on page 23
- “IDS Trace View” on page 24

Welcome View

The Welcome view is essentially Analyzer’s title page. It provides version information and links to the Analyzer Release Notes and help files.

Figure 1-1 *Analyzer Welcome Page*



Project View

The Project view displays the details of your current Analyzer projects. Analyzer actions create and manipulate the project components displayed in the Project view. The Project view includes a toolbar that you can use to work with your Analyzer project.

Figure 1-2 *Analyzer Project View*

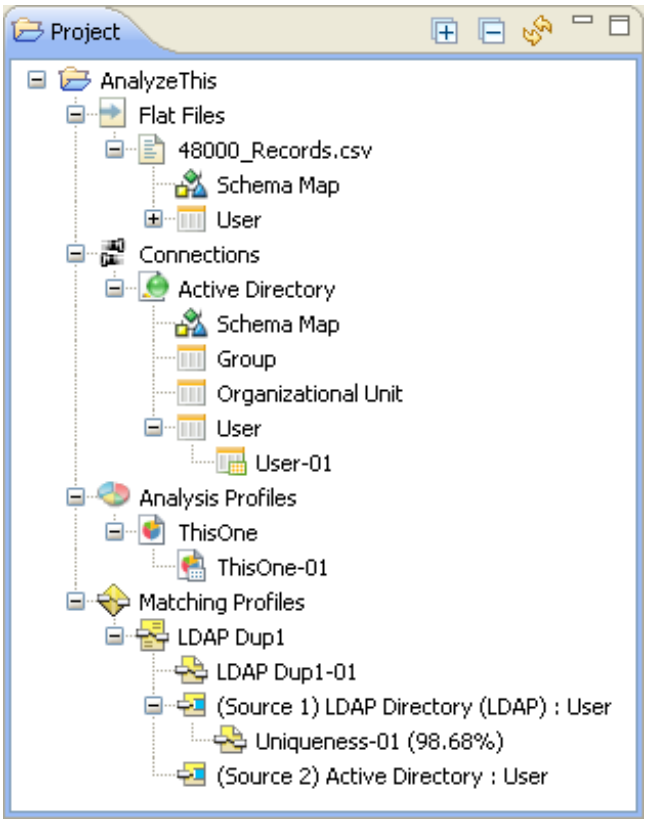


Table 1-1 *Icons in the Project View*

Icon	Description
	The **Expand All** icon expands the object hierarchy in the Project view.
	The **Collapse All** icon collapses the object hierarchy in the Project view.
	The **Refresh View** icon refreshes the Project view and reinitializes Analyzer's internal database.

Additionally, you can right-click any object in the Project View for a menu of actions you can perform on the selected object.

The Project View uses several icons to represent the various objects you can create in an Analyzer project:

Table 1-2 *Project View Icons*

Icon	Description
	Represents a specific project container. The project folder contains all objects related to a particular Analyzer project.

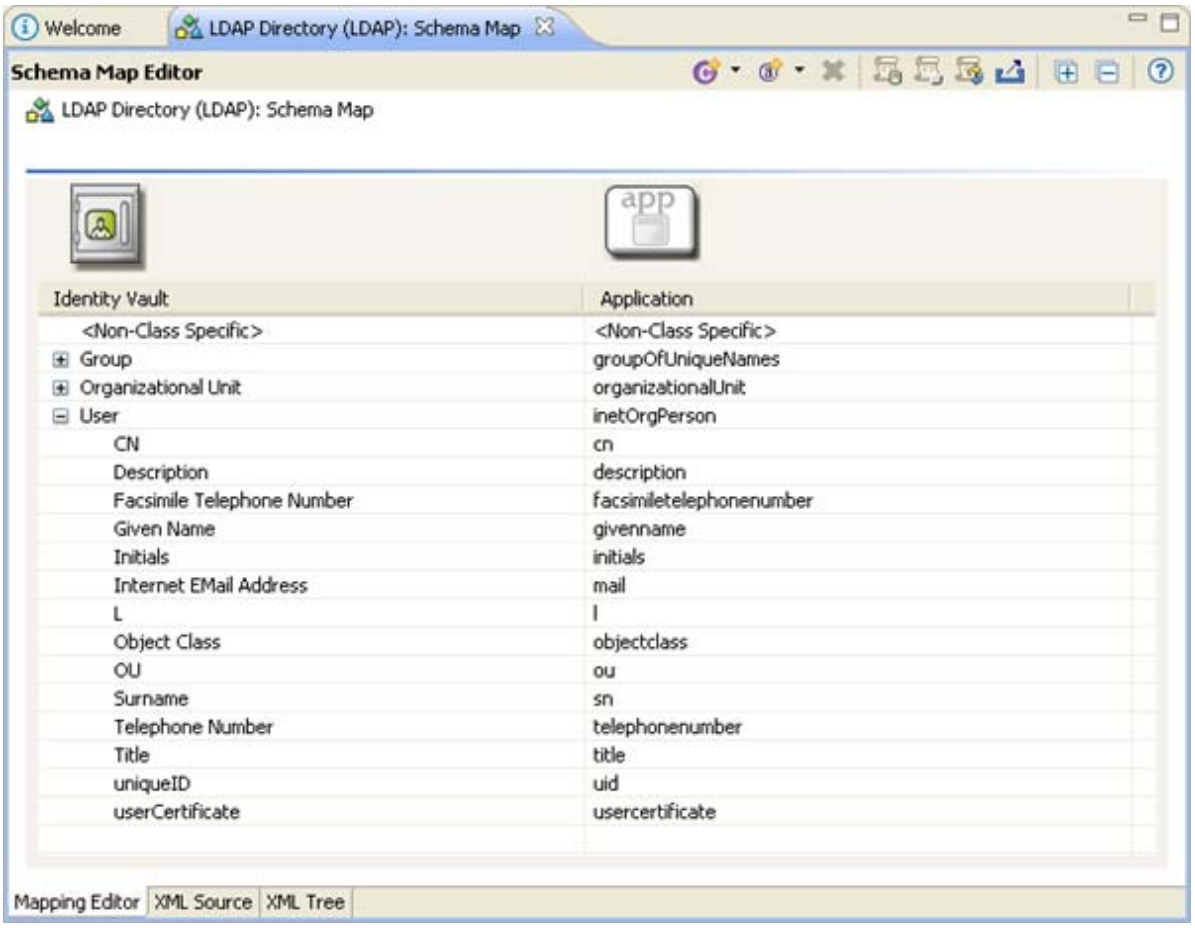
Icon	Description
	Represents the flat file connection container, where you can create and manage flat file profiles.
	Represents a specific flat file connection.
	Represents a specific schema map for the associated application or flat file connection.
	Represents a specific data set definition.
Unmodified Data Set	Represents a specific data set generated from the associated data set definition.
Modified Data Set	If you make changes to data set instance values after importing the data set, Analyzer indicates this by displaying the **Modified Data Set** icon. **NOTE:** Because managing modifications requires re-querying the database, the **Modified Data Set** icon persists even if you reset all values to their original state. To reset the Modified Data Set flag in this case, click **Update Changes to the Application** in the Data Browser toolbar.
	Represents the connection profile container, where you can create and manage connection profiles to different applications.
Driver Stopped Driver Running	Represents a specific application connection profile and identifies the associated driver as running or stopped.
	Represents the analysis profile container where you can create and manage analysis profiles.
	Represents a specific analysis profile.
	Represents specific analysis results, generated by applying the analysis profile to a specific data set instance.
	Represents the matching profile container, where you can create and manage matching profiles for data comparison.
	Represents a specific matching profile.
	Represents the results of a specific matching or uniqueness analysis, run against the data sets specified in the matching profile.
	Represents a specific data source configuration used in a matching profile.

Schema Map Editor

Analyzer's schema map associates an application's schema attributes to the corresponding schema attributes in Analyzer's base schema. This lets you ensure that your data analysis and cleaning operations properly associate similar values between the disparate systems.

To accomplish this, Analyzer leverages the schema mapping features in Designer. For detailed information about mapping schema in Identity Manager, see "Managing the Schema" in the *NetIQ Designer for Identity Manager Administration Guide*.

Figure 1-3 *Schema Map Editor*



The Schema Map editor includes a toolbar that you can use to work with schema:

Table 1-3 *Icons in the Schema Map Editor*

Icon	Description
	The **Add Class Mapping** icon lets you define a mapping between an Identity Vault class and an application class.
	The **Add Attribute Mapping** icon lets you define a mapping between an Identity Vault attribute and an application attribute.
	The **Delete** icon deletes the selected schema mapping entry in the Schema Map editor.
	The **Manage Identity Vault Schema** icon opens the Manage Schema dialog box, from which you can add, delete, and modify the object classes and attributes in the Identity Vault schema.
	The **Manage Application Schema** icon opens the Manage Schema dialog box, from which you can add, delete, and modify the object classes and attributes in the application schema.

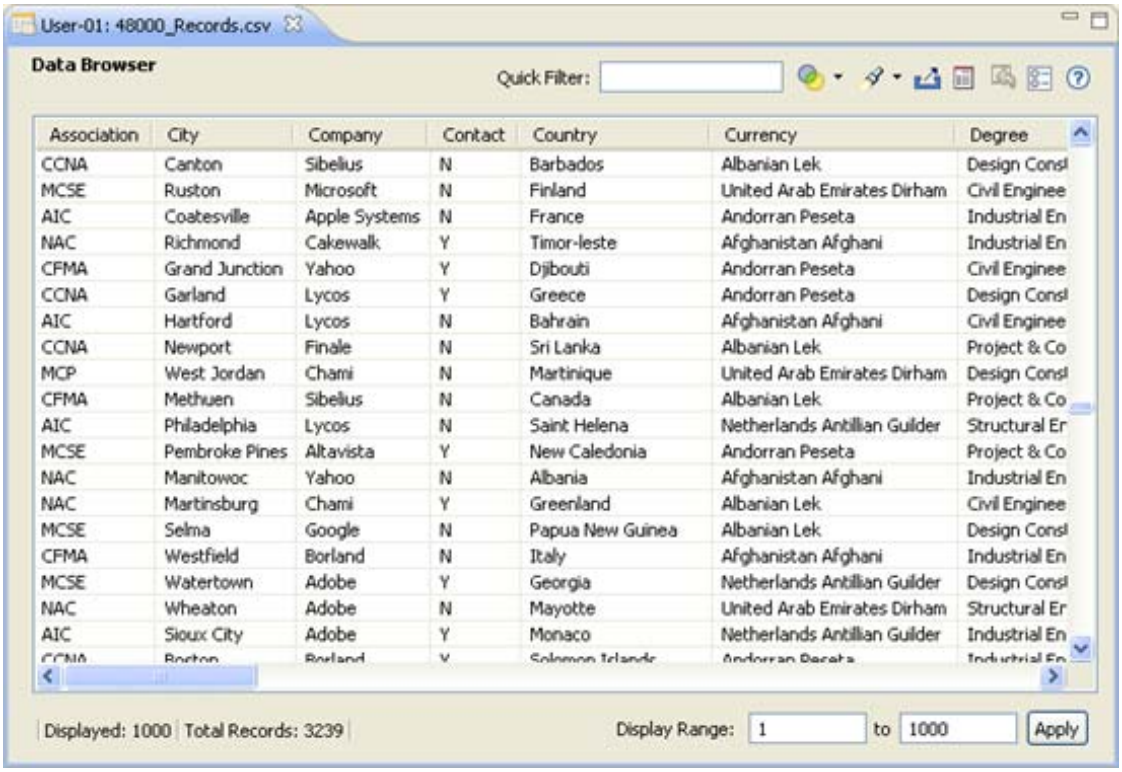
Icon	Description
	The **Refresh Application Schema** icon lets you query the application for an updated schema definition.
	The **Expand All** icon expands the schema hierarchy in the Schema Map editor.
	The **Collapse All** icon collapses the schema hierarchy in the Schema Map Editor.
	The **Export Schema Mapping Policy to XML File** icon lets you save the current schema mapping to an XML file.

Additionally, you can right-click in the Schema Map editor for a menu of actions you can perform.

Data Browser

The Data Browser displays the details of the selected data set instance. From this view you can examine and modify the data in the data set instance. Click in the Data Browser to select a cell. After selecting a cell, you can navigate the Data Browser with your keyboard arrow keys.

Figure 1-4 *Data Browser Editor*



Association	City	Company	Contact	Country	Currency	Degree
CCNA	Canton	Sibelius	N	Barbados	Albanian Lek	Design Consl
MCSE	Ruston	Microsoft	N	Finland	United Arab Emirates Dirham	Civil Enginee
AIC	Coatesville	Apple Systems	N	France	Andorran Peseta	Industrial En
NAC	Richmond	Cakewalk	Y	Timor-leste	Afghanistan Afghani	Industrial En
CFMA	Grand Junction	Yahoo	Y	Djibouti	Andorran Peseta	Civil Enginee
CCNA	Garland	Lycos	Y	Greece	Andorran Peseta	Design Consl
AIC	Hartford	Lycos	N	Bahrain	Afghanistan Afghani	Civil Enginee
CCNA	Newport	Finale	N	Sri Lanka	Albanian Lek	Project & Co
MCP	West Jordan	Chami	N	Martinique	United Arab Emirates Dirham	Design Consl
CFMA	Methuen	Sibelius	N	Canada	Albanian Lek	Project & Co
AIC	Philadelphia	Lycos	N	Saint Helena	Netherlands Antillian Guilder	Structural Er
MCSE	Pembroke Pines	Altavista	Y	New Caledonia	Andorran Peseta	Project & Co
NAC	Manitowoc	Yahoo	N	Albania	Afghanistan Afghani	Industrial En
NAC	Martinsburg	Chami	Y	Greenland	Albanian Lek	Civil Enginee
MCSE	Selma	Google	N	Papua New Guinea	Albanian Lek	Design Consl
CFMA	Westfield	Borland	N	Italy	Afghanistan Afghani	Industrial En
MCSE	Watertown	Adobe	Y	Georgia	Netherlands Antillian Guilder	Design Consl
NAC	Wheaton	Adobe	N	Mayotte	United Arab Emirates Dirham	Structural Er
AIC	Sioux City	Adobe	Y	Monaco	Netherlands Antillian Guilder	Industrial En

When using the Data Browser, you should be familiar with the following topics:

- "Data Browser Toolbar and Controls" on page 13
- "Data Browser Right-Click Menu" on page 14
- "Data Markers" on page 15

Data Browser Toolbar and Controls

The Data Browser editor includes a toolbar and other controls that provide access to common Data Browser actions.

Table 1-4 *Controls in the Data Browser Editor*

Icon	Description
Quick Filter:	The **Quick Filter** field lets you specify a keyword filter for the fields in the current data set instance. The Data Browser displays only those records that contain the keyword in one or more of its cells.
	The **Filter** icon lets you filter your view of the current data set instance. For more information, see "Filtering Data" on page 54.
	The **Inspector** icon opens the Inspect Data dialog box so you can test data in the Data Browser. For more information, see "Inspecting Data" on page 53.
	The **Export** icon lets you export the current data set instance or generate a report based on the current contents of the Data Browser. For more information, see "Exporting Data" on page 52 and "Auditing and Reporting" on page 71.
	The **Generate a Data Browser Report** icon lets you create a report of the data currently displayed in the Data Browser, including any changes you might have made since importing the data set instance. For more information, see "Data Browser Reports" on page 72. **NOTE:** The Data Browser reflects any query sort applied to the data set, but ignores any local sort settings in the Data Browser.
	The **Update changes to the application** icon opens the Update to Application dialog box, where you can select the changes you want to write back to the application. Click **Update** to complete the data update process.
	The **Display Settings** icon displays the Data Browser Display Settings page, where you can manage how the Data Browser editor displays data. For more information, see "Data Browser Settings" on page 36.
Query Sort	Click any column header to sort the entire data set instance by that column. Click the column header once to sort in ascending order. Click the column header again to sort in descending order. The Data Browser displays a sort order icon in the column header that indicates the current sort order (ascending or descending). The query sort is case sensitive, meaning that it sorts 0-9, then A-Z, then a-z. **IMPORTANT:** If you query sort an attribute with multiple values, the same record might appear on multiple pages in the Data Browser. For example, when sorting on a Title attribute with multiple values in each record, the John record might appear on the first page with Title values A - G, then on a later page with Title values H - P, then again with Title values Q - Z.

Icon	Description
	The Show icons let you select the types of records you want to display in the current page. Options include **Show all records on this page**, **Show only failed metrics in this page**, **Show only passed metrics in this page**, and **Show only modified records in this page**. **NOTE:** The Data Browser applies Show icons to each data page individually, rather than to the entire data set instance. This means that the number of pages in the data set instance does not change, but the number of records displayed on a page might change considerably.
	The Paging icons let you navigate between the pages of data in the current data set instance. You can jump to the first or last page, and browse to the previous or next page of data. You can set the actual page size (number of records per page) in **Window** > **Preferences**, then select **Analyzer** > **Data Browser** in the left navigation area.

Data Browser Right-Click Menu

You can right-click in the Data Browser to display a menu of actions you can perform.

Table 1-5 *Data Browser Right-Click Menu Options*

Menu Option	Description
Apply Metric to Column	Applies the metric you select to the specified column. Analyzer evaluates each value in the column against the metric to see if it matches the metric format.
Inspector	Opens the Inspector dialog box so you can define metrics for the different attributes in your data set. For more information, see "Inspecting Data" on page 53.
Clear	Allows you to clear currently applied filter or Inspector settings.
Manage Metrics	Opens the Metric Manager, where you can create, edit, and delete metrics, which allow you to test attributes for a specific data format. For more information, see "Analyzing a Data Set" on page 64.
Show Only	Specifies the types of records you want to display in the current page. Options include **Show all records on this page**, **Show only failed metrics in this page**, **Show only passed metrics in this page**, and **Show only modified records in this page**. **NOTE:** The Data Browser applies **Show Only** restrictions to each data page individually, rather than to the entire data set instance. This means that the number of pages in the data set instance does not change, but the number of records displayed on a page might change considerably.
Show all Multi-values	Enables displaying multiple values for those attributes that contain them. When this option is deselected, the Data Browser displays only the first attribute value. You can also modify this setting from the Display Settings dialog box.

Menu Option	Description
Sort Local Page	Sorts the currently displayed page of data from the data set instance. To use this feature, right-click in the column on which you want to sort, then select **Sort Local Page** to sort in ascending order. Repeat this process to sort in descending order. You can use the local sort and query sort together. The query sort orders the entire data set instance, while the local sort orders the current page of data. For example, if you query sort a data set instance by the Surname attribute, and local sort the current page by the Title attribute, then page through the resulting data, you might see Surnames starting with A-C in the first page (sorted by Title), surnames C-F in the second page (sorted by title), and so on.
Add Value	Adds a value to the selected cell. If a single-valued attribute already has a value, or **Show all Multi-values** is disabled and the attribute has a value, the **Add Values** option is disabled.
Edit Value	Edits the value in the selected cell.
Delete Value	Deletes the value in the selected cell.
Revert Value	Replaces a modified cell value with its original, unmodified value.
Edit Multi-Values	Opens the Edit Multi-Valued Cell dialog box so you can add, edit, or delete one or more of the attribute values. This option is only available when there are already two or more values in the selected attribute.
Export	Exports the data set definition to a flat file. For more information, see "Exporting Data" on page 52.
Update to Application	Saves Data Browser updates to the source application. For more information, see "Modifying Data" on page 55.

The Data Browser also provides keyboard-based access to the right-click menu. Keyboard usage varies slightly between Windows and Linux platforms:

Windows: Press Alt+m to open the right-click menu. Press the first letter of any menu option to execute that option. If multiple menu options start with the same letter, press the letter multiple times to cycle through the menu options that start with that letter, then press Enter.

Linux: Press Alt+m to open the right-click menu. Use use the up-arrow and down-arrow keys to select the option you want to execute, then press Enter.

Data Markers

When viewing and modifying data, the Data Browser uses different markers to indicate the state of the value in a given cell:

Figure 1-5 Data Browser Editor Data Markers

Association	City	Company	Contact	Country
AIC	Bandera	Chami	Y	USA
AIC	Gatlinburg	Google	N	USA
AIC	Anchorage	Google	Y	USA
NAC	Chandler	Altavista	N	USA
MCSE	Pullman	Novell	N	USA
CFMA	Eureka	Finale	N	USA
NAC	Rockville	Chami	N	USA
MCP	Cohoes	Finale	Y	India
CCNA	West Jordan	Apple Systems	N	China
MCP	East Lansing	Yahoo	Y	USA
AIC	Santa Barbara	Yahoo	Y	USA
CFMA	Meridian	Borland	Y	USA

Selected Record: Identified with a gray background.

Selected Cell: Identified with blue text and border.

Modified Value: Identified with a yellow background and pencil icon.

Added Value: Identified with a green background and plus sign icon.

Deleted Value: Identified with a pink background and X icon.

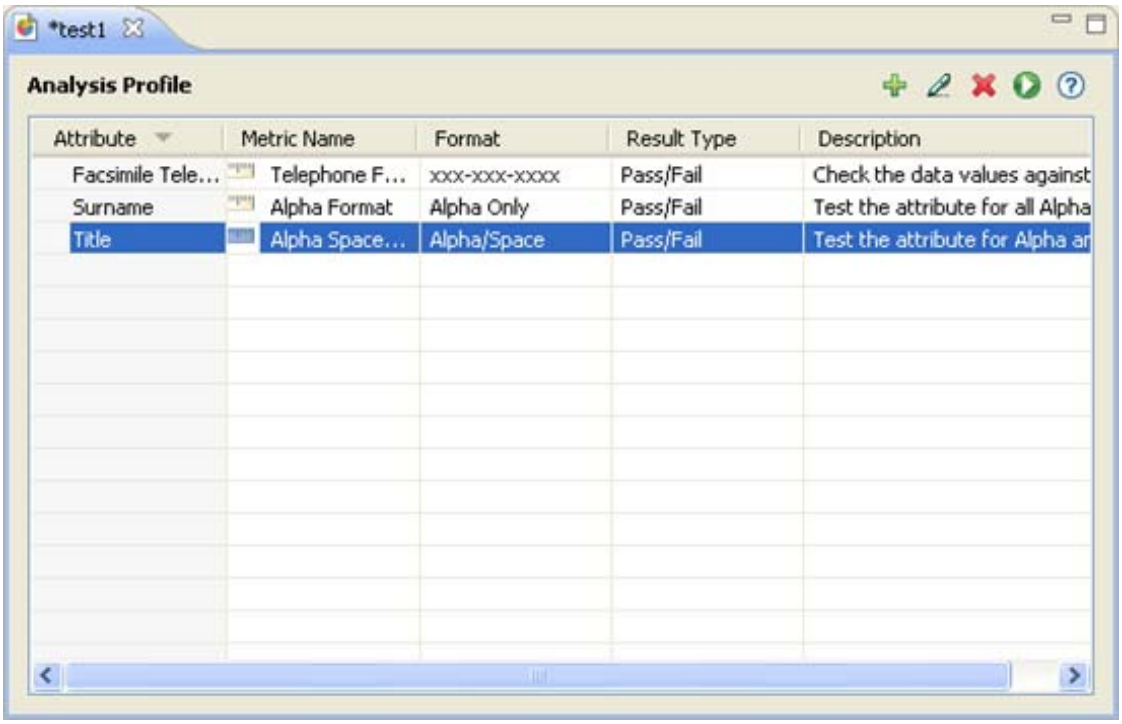
Metric Applied: Indicates that a metric is applied to the column. This filters the Data Browser content based on that column.

Failed Value: Identified with red text and a checkmark icon. This indicates that the value failed the metric applied to it.

Analysis Profile Editor

The Analysis Profile editor lets you configure a profile for analyzing one or more data set instances. Each analysis profile contains one or more metrics against which you can evaluate attribute values to see how the data conforms to your defined data format standards.

Figure 1-6 *Analysis Profile Editor*



The Analysis Profile editor includes a toolbar that you can use to configure the analysis profile. You can also right-click in the Analysis Profile editor to access the same tools from a drop-down menu.

Table 1-6 *Icons in the Analysis Profile Editor*

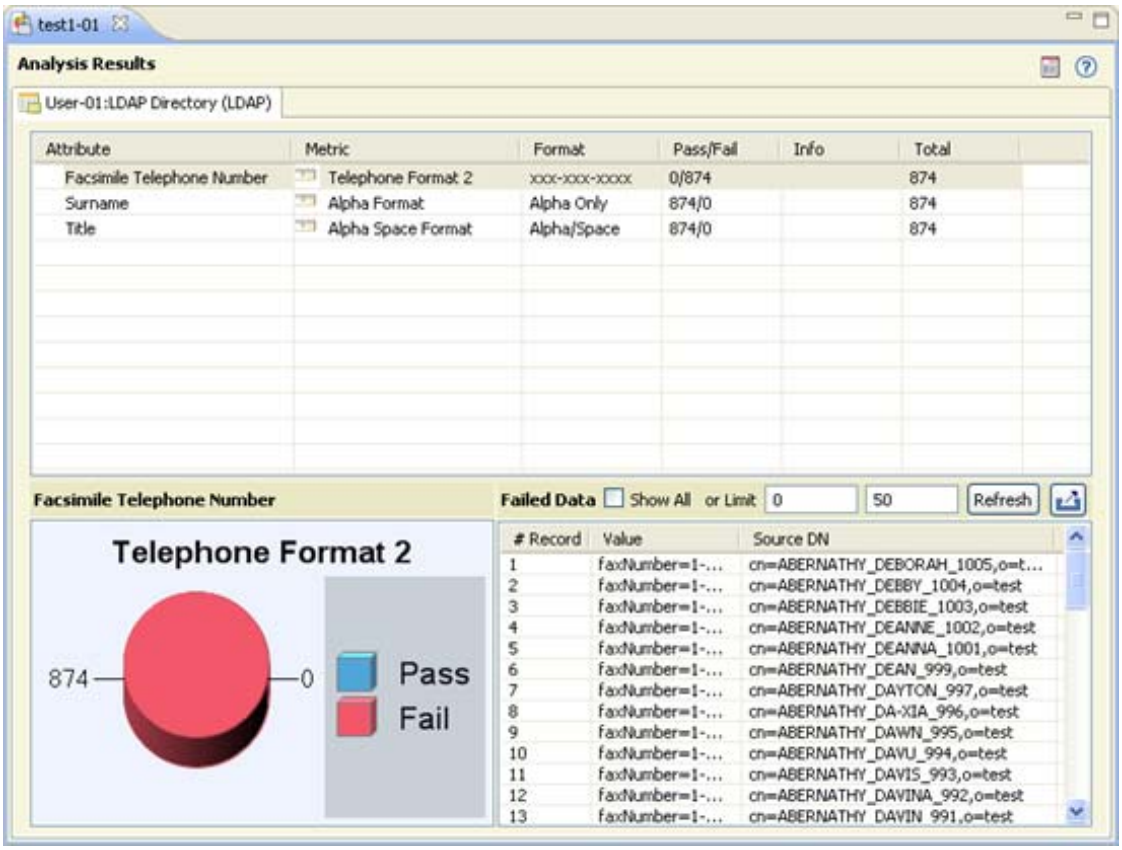
Icon	Description
	The **Add** icon opens the Analysis Metric Selector dialog box, where you can select an attribute/metric pair for your analysis. For more information, see "Creating and Customizing an Analysis Profile" on page 56.
	The Edit icon lets you edit the currently selected metric in the Analysis Profile Editor.
	The **Delete** icon removes the selected metric page from the current analysis profile.
	The **Analyze** icon performs an analysis (based on the analysis profile) against the selected data sets. For more information, see "Analyzing a Data Set" on page 64.

NOTE: To save changes to the Analysis Profile, use Ctrl+S or select Save from the Designer toolbar.

Analysis Results View

The Analysis Results view displays the results of a data analysis performed against a particular data set instance. It displays each metric used to test the data set instance. Select an attribute to display details about the results of the analysis for that attribute.

Figure 1-7 Analysis Results View



The Analysis Results view includes the following tools:

Table 1-7 Icons in the Analysis Results View

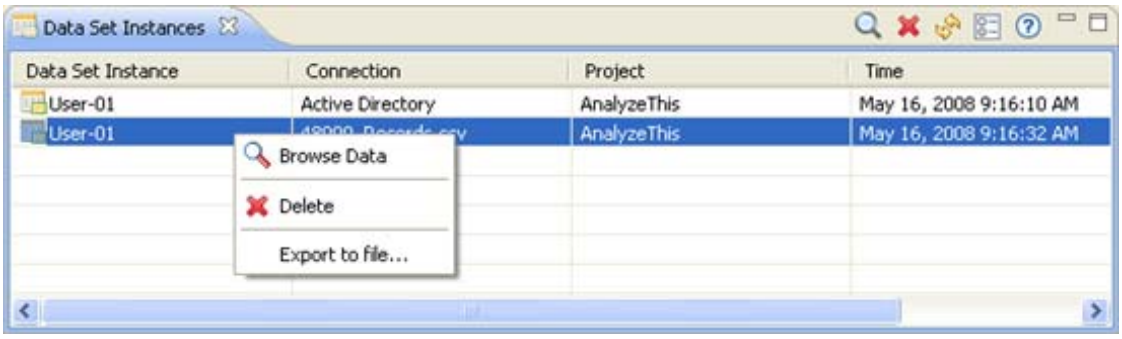
Icon	Description
	The **Generate Analysis Report** icon creates a report of the current analysis results. For more information, see "Analysis Reports" on page 72.
	The **Export to File** icon lets you export the failed data to a delimited file for further analysis. You can also right-click a metric and select **Export to File**. For information about the Export to File dialog box, see "Exporting Data" on page 52.

Data Set Instances View

The Data Set Instances view displays the current data set instances you have created in Analyzer. From this view, you can select a particular data set instance to browse, inspect, and analyze.

By default, Analyzer does not display the Data Set Instances view. To open it, select **Window** > **Show View** > **Data Set Instances**.

Figure 1-8 Data Set Instances View



The Data Set Instances view includes a toolbar that you can use to work with the data set instances you have created:

Table 1-8 Icons in the Data Set Instances View

Icon	Description
	The **Browse Data** icon opens the Data Browser for the selected data set. For more information, see "Viewing Application Data" on page 52.
	The **Delete** icon deletes the currently selected data set.
	The **Refresh the Data Set Instances view** icon updates the data set information displayed in the Data Set Instances view.
	The **Analyzer Preferences** icon opens the Preferences dialog box, from which you can configure Analyzer's database settings. For more information, see "Database Settings" on page 28.

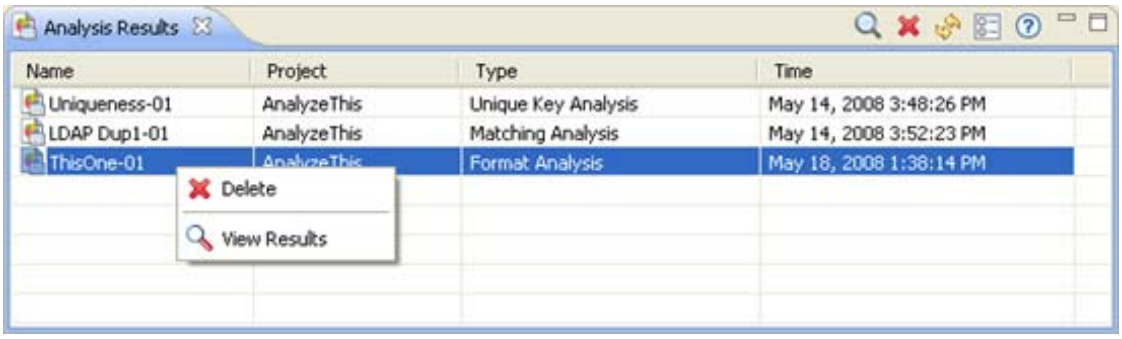
Additionally, you can right-click an entry in the data set instances view for a menu of actions you can perform. This menu includes **Browse Data** and **Delete** from Table 1-8, as well as an option to export the selected data set instance to a file.

Analysis Result Sets View

The Analysis Result Sets view displays the current analysis results sets you have generated in Analyzer. From this view, you can select a particular analysis results set to browse inspect, and analyze.

By default, Analyzer does not display the Analysis Result Sets view. To open it, select **Window** > **Show View** > **Analysis Results**.

Figure 1-9 Analysis Result Sets View



The Analysis Result Sets view includes a toolbar that you can use to work with the trace data:

Table 1-9 Icons in the Analysis Result Sets View

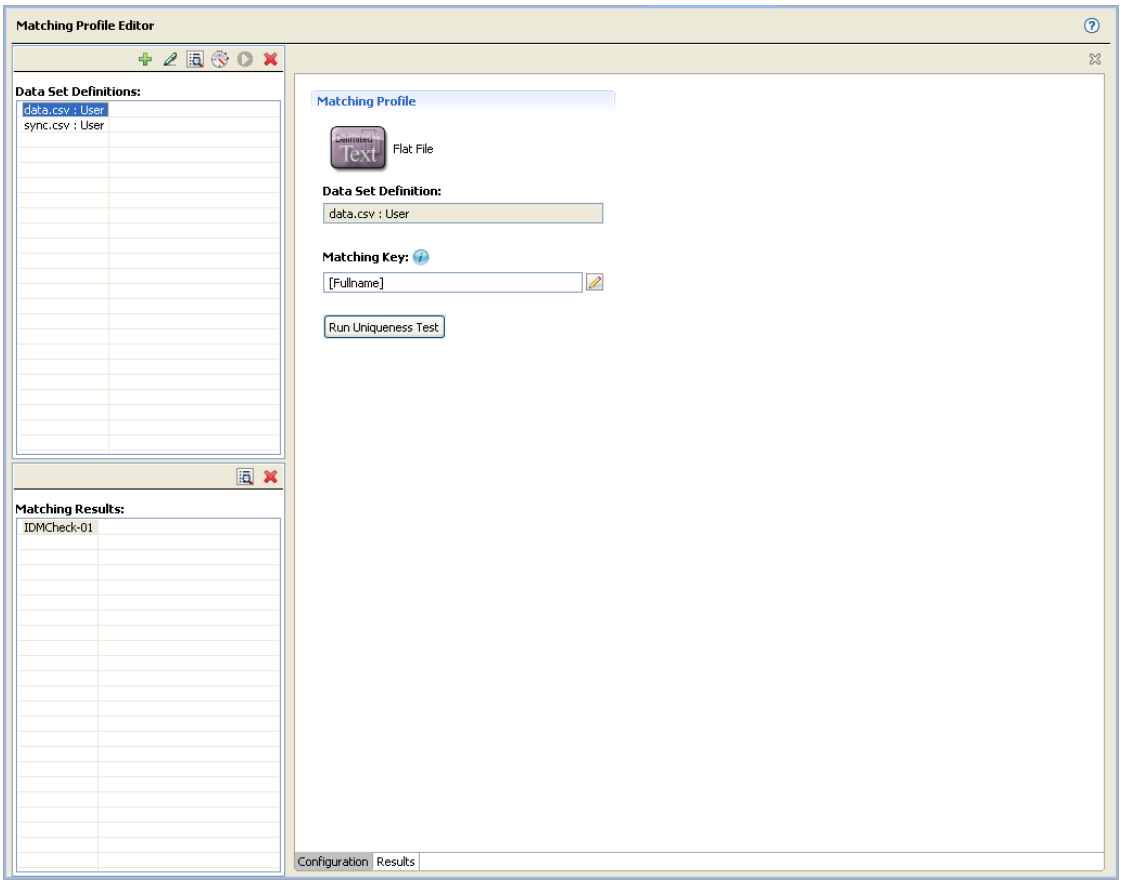
Icon	Description
	The **View Analysis Results** icon opens the Analysis Results view for the selected Analysis. For more information, see "Analyzing a Data Set" on page 64.
	The **Delete** icon deletes the currently selected analysis.
	The **Refresh the Analysis Results view** icon updates the analysis results information displayed in the Analysis Results view.
	The **Analyzer Preferences** icon opens the Preferences dialog box, where you can configure Analyzer's database. For more information, see "Database Settings" on page 28.

Additionally, you can right-click a result set for a menu of actions you can perform.

Matching Profile Editor

The Matching Profile editor lets you compare values in one or more data sets. You can check for duplicate values within a specified data set and check for matching values between two data sets.

Figure 1-10 *Matching Profile Editor*



The Matching Profile editor includes a toolbar with the following tools:

Table 1-10 Icons in the Matching Profile Editor

Icon	Description
	The Add icon adds new data set definitions for matching profile.
	The Edit icon lets you edit the selected data set definition.
	The View Uniqueness Results icon displays the uniqueness result of the selected data set definition. For more information about uniqueness results, see "Uniqueness Results View" on page 22. You can also click **Run Uniqueness Test** button to view the uniqueness result of a selected data set definition.
	The Uniqueness Test icon lets you test the matching key with the selected data set definition to make sure it is suitable for use in match testing. For more information about uniqueness testing, see "Evaluating a Data Set for Unique Values" on page 67.
	The **Analyze** icon starts the matching value analysis between two data sets. For more information about using the Matching Profile editor, see "Comparing Data" on page 67.
	The **Delete** icon removes the selected data set definition from the selected matching profile.

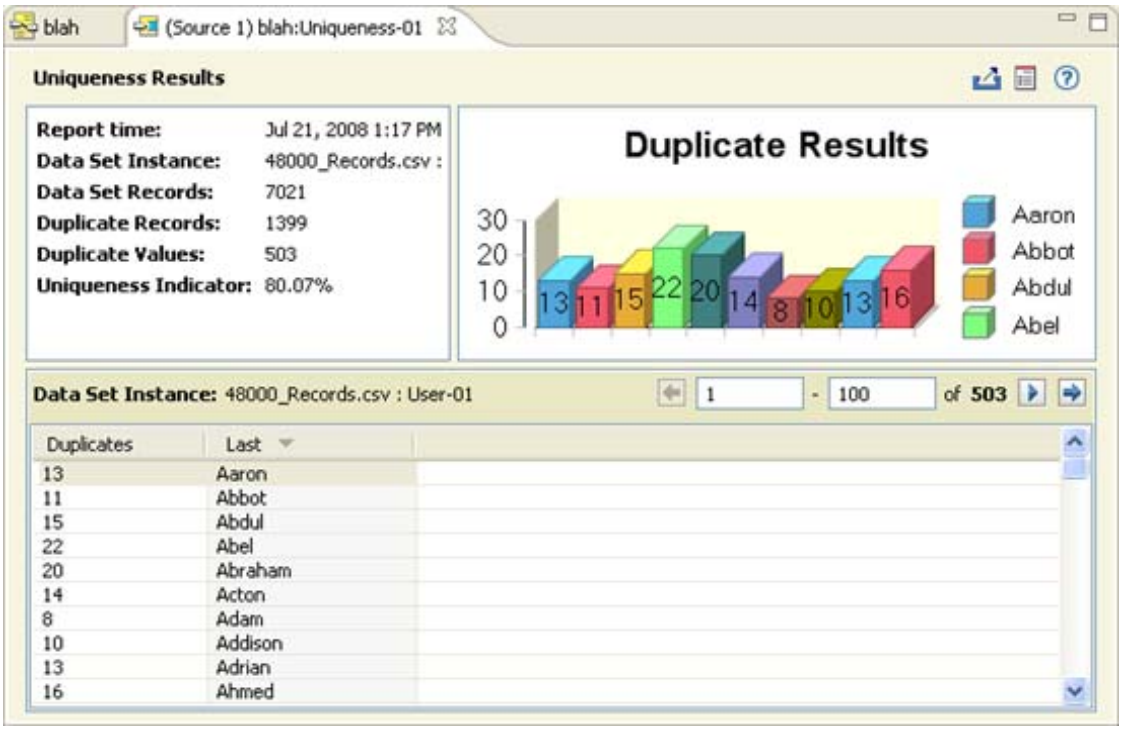
For more information on matching attributes between two data sets, see "Matching Values between Data Sets" on page 69.

Uniqueness Results View

The Uniqueness Results view displays the results of a uniqueness test. It provides details about the analysis configuration, a graphical view of the results, and a list of duplicate values encountered during the analysis.

For more information about the uniqueness test, see "Evaluating a Data Set for Unique Values" on page 67.

Figure 1-11 *Duplicates Results View*



The Duplicate Results view editor includes a toolbar with the following tools:

Table 1-11 *Icons in the Duplicate Results View*

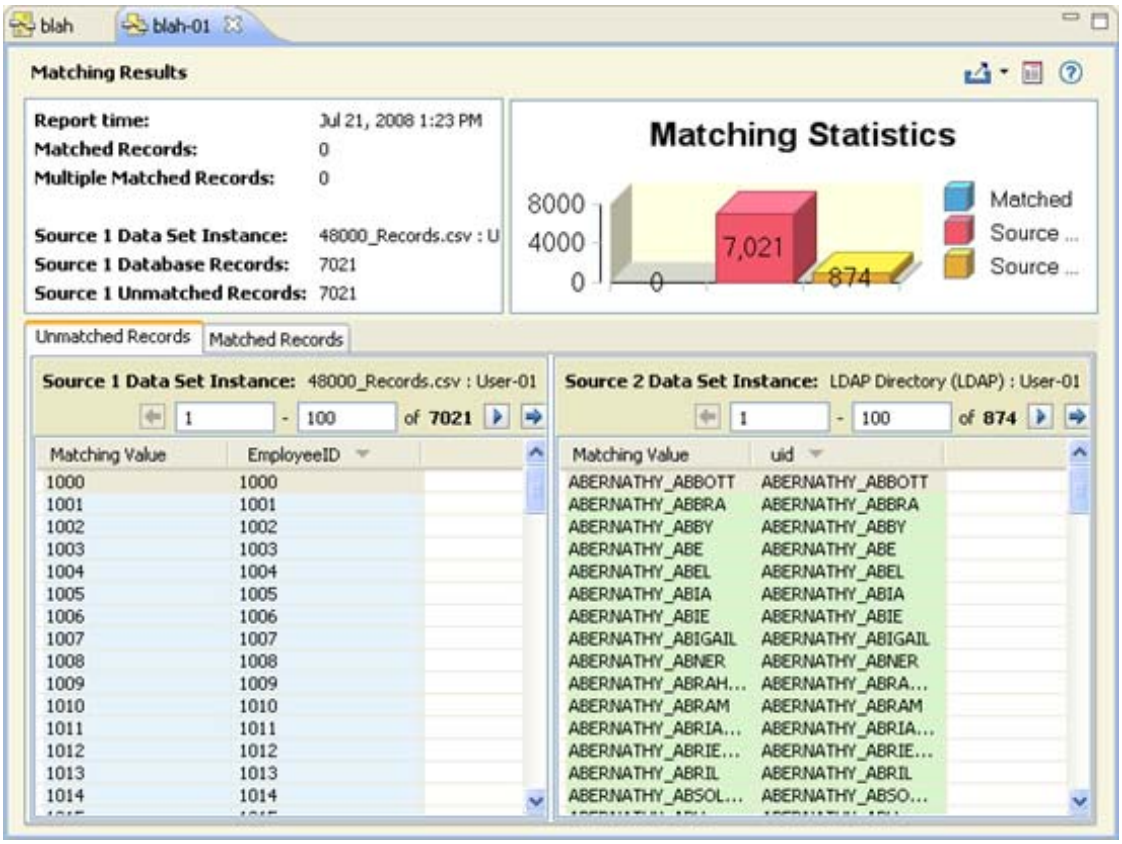
Icon	Description
	The **Export Duplicate** icon lets you export the records with duplicate values to a flat file for further analysis.
	The **Generate Uniqueness Report** icon lets you create a report of the analysis results.

Matching Results View

The Matching Results view displays the results of your matching values analysis. It provides details about the analysis configuration, a graphical view of the results, and a list of matching values encountered during the analysis.

For more information about testing for matching values, see "Matching Values between Data Sets" on page 69.

Figure 1-12 Matching Results View



The Matching Results view editor includes a toolbar with the following tools:

Table 1-12 Icons in the Matching Results View

Icon	Description
	The **Export to File** icon lets you export the unmatched records to a flat file for further analysis.
	The **Generate Matching Report** icon lets you create a report of the analysis results.

IDS Trace View

The Independent Driver Services (IDS) Trace view displays the communications between Analyzer and the selected application. This view is available primarily for troubleshooting connection problems when configuring and accessing data sets.

By default, Analyzer does not display the IDS Trace view. To open it, select **Window** > **Show View** > **IDS Trace**.

The IDS Trace view includes a toolbar that you can use to work with the trace data:

Table 1-13 Icons in the IDS Trace View

Icon	Description
	The **Export** icon lets you export the IDS trace data into an external text file so you can evaluate it offline.
	The **Clear** icon lets you clear the current IDS trace log.
	The **Preferences** icon opens the IDS trace settings page, where you can configure the IDS trace environment. For more information, see "IDS Trace Settings" on page 37.

Additional Resources

For more information on topics relevant to Analyzer, refer to the following Web sites:

- NetIQ Identity Manager documentation site (https://www.netiq.com/documentation/identity-manager-48/)
- Designer for Identity Manager documentation site
- Identity Manager drivers documentation site (https://www.netiq.com/documentation/identity-manager-48-drivers/)

2 Configuring Analyzer

This section describes the various NetIQ Analyzer for Identity Manager property pages. You can use the property pages to configure the operation of the various Analyzer components:

- "Preferences" on page 27
- "Connection Settings" on page 33
- "Flat File Settings" on page 34
- "Data Set Settings" on page 35
- "Data Browser Settings" on page 36
- "IDS Trace Settings" on page 37

Preferences

Analyzer preferences let you configure some general environment settings, and include the following:

- "Connections" on page 27
- "Data Browser" on page 27
- "Database Settings" on page 28
- "Matching" on page 29
- "Reporting" on page 30

Connections

The Connections preferences page lets you manage the Java Libraries that Analyzer needs to communicate with different applications. You can add and remove Java libraries (typically `.jar` files) from this page.

Data Browser

The Data Browser preferences page lets you manage how the Data Browser displays the data records in a data set instance. To access these settings, select **Window** > **Preferences**, then select **Analyzer** > **Data Browser** in the left navigation area.

Table 2-1 Analyzer Preferences - Data Browser

Setting	Description
Show Display settings for attributes when opening the Data Browser	Enables/disables displaying a warning dialog box about having more than 12 attributes with multiple values in a data set definition. Use this dialog box to limit the attributes displayed in the Data Browser to improve performance during data inspection and analysis.
Total number of rows per page	Specifies the number of records to display on a page in the Data Browser.

Database Settings

Analyzer database settings configure the database that Analyzer uses to organize and store its data. To access these settings, select Window > Preferences, then select Analyzer > Database Settings in the left navigation area.

You can also select the Preferences icon in the Data Set Instances view.

Table 2-2 Analyzer Preferences - Database Settings

Setting	Description
Select Alias	Specifies the database alias that Analyzer uses.
Add Alias	Opens the New Alias dialog box so you can configure a new database alias. The new alias requests the same information that is displayed in the Database Settings page.
Remove Alias	Deletes the currently selected database alias. However, you must have at least one database alias for Analyzer to run.
Driver	Specifies the driver that Analyzer uses to access the database. Analyzer supports either an internal HSQLDB, or an external MySQL database.
Use SSL	Enables SSL connections between Analyzer and an external MySQL database used as the Analyzer database. To use SSL, the MySQL JDBC library must be version 5.1.6 or higher. For example, `mysql-connector-java-5.1.6-bin.jar`. **NOTE:** MySQL must be configured to use SSL before enabling Use SSL. For more information, see "Using SSL for Secure Connections" (https://dev.mysql.com/doc/refman/5.0/en/secure-connections.html) in the *MySQL Reference Manual*.
JDBC URL	Specifies the JDBC URL where you can access the database. When you are using the internal HSQLDB, this field is not configurable.
Database	Specifies the database name.
Batch Size	Specifies the batch size of the records that Analyzer imports from the external database. By default, the batch size is set to 1000. You can configure batch size in Window > Preferences > Analyzer > Database Settings > Batch Size. This setting is only available if the external database is MySQL.

Setting	Description
Store Password	Saves the password in a local system. If it is a local database (HSQLDB), the password is by default stored locally. If it is an external database (MYSQL), user can select/deselect this option.
Username	Specifies a valid username with which Analyzer can access the database.
Password	Specifies the database access password for the specified user.
JDBC Paths	Specifies the path to the database's JDBC libraries. You must provide a classpath if you are using an external database with Analyzer.
Class Name	The class name for the driver.
Test Connection	Tests the database connection by using the specified database parameters.
Restore Defaults	Resets the database settings to Analyzer's default (internal database) configuration.

In case of Flat File import, Analyzer can read 1000 records at a time and write them to the Analyzer MySQL database server. The bigger the batch size, the faster is the import performance. The performance of importing data also depends on the packet size of the database server. The packet size is based on various factors including the number of fields or attributes, the size of the value of each field or attribute and the batch size. If you get `Error inserting data com.mysql.jdbc.PacketTooBigException: Packet for query is too large (9525623 > 1048576)` exception in the error log, you can change the value of the packet on the server by setting the **max_allowed_packet** variable. Analyzer's MySQL database server throws this exception because of the limited data size on the database server. You should reduce the Analyzer batch size and retry the import. Another workaround is to increase the value for the **max_allowed_packet** variable in the MySQL database server.

Analyzer allows you to change its internal database from the default HSQLDB to a MySQL database. You can configure database settings in **Window** > **Preferences** > **Analyzer** > **Database Settings**. When you use an external MySQL database, be aware of that the MySQL database uses the default character set from the operating system for encoding table fields. If an extended or double-byte character is not recognized by the default character set, Analyzer displays ??? in the Data Browser. To avoid this, set the operating system's default character set to UTF-8, or to a character set that includes all the extended or double-byte characters that Analyzer might import.

Ensure that you shift from the local database (HSQLDB) to the external database (MYSQL) if you are importing data more than 40,000 records.

Matching

Matching settings define the default configuration for Analyzer's matching analysis. To access these settings, select **Window** > **Preferences**, then select **Analyzer** > **Matching** in the left navigation area.

Table 2-3 Matching Preferences

Setting	Description
Uniqueness Test Percentage Required	Specifies the uniqueness threshold that a matching key/data set combination must achieve before Analyzer can run a matching analysis. Analyzer supports values between 95 and 100 percent.

Reporting

Reporting settings define the default configuration for the various reports you can generate from Analyzer. These are global settings that apply to all of your Analyzer projects. To access these settings, select **Window** > **Preferences**, then select **Analyzer** > **Reporting** in the left navigation area.

The Reporting settings page includes four tabs, one for each type of report you can generate. For more information about Analyzer reports, see "Auditing and Reporting" on page 71.

- "Data Browser" on page 30
- "Analysis" on page 31
- "Uniqueness" on page 31
- "Matching" on page 32

Data Browser

The Data Browser report settings lets you specify the following default settings.

Table 2-4 Data Browser Report Settings

Setting	Description
Report Title	Specifies a name for the Data Browser report you are generating. The default title is Data Browser Report.
Notes	Specifies optional details describing the report.
Display in Landscape Mode	Displays the report suitable for landscape printing (11" x 8.5" rather than 8.5" x 11").
Wrap text when values overflow	Allows cell data to wrap around within the report column. If you deselect this option, data that exceeds the width of the report column is truncated.
Display added/modified/deleted icons	Specifies that the report should include icon indicators for added, modified, and deleted data cells.
Display metric failures in red and with icons	Specifies that the report should indicate data cells that failed the analysis in red and with a failure icon.

Analysis

The Analysis report settings lets you specify the following default settings.

Table 2-5 *Analysis Report Settings*

Setting	Description
Report Title	Specifies a name for the Analysis report you are generating. The default title is Analysis Report.
Notes	Specifies optional details describing the report.
Display in Landscape Mode	Displays the report suitable for landscape printing (11” x 8.5” rather than 8.5” x 11”).
Wrap text when values overflow	Allows cell data to wrap around within the report column. If you deselect this option, data that exceeds the width of the report column is truncated.
Display a combined summary of all datasets	Specifies that the report should include a section that summarizes the analysis results across all analyzed data sets. This option is available only when the analysis results include data from multiple data sets.
Display graphic for each metric	Displays a graphical representation of the results of each metric in the analysis.
Display Failed Data/Details for each metric	For each metric in the analysis, displays a list of the failed records.
Max number of records to display in Failed Data or Patterns	Specifies a maximum number of records to display that failed the analysis. This setting is active only when **Display Failed Data/Details for each metric** is selected.

Uniqueness

The Uniqueness report settings lets you specify the following default settings.

Table 2-6 *Uniqueness Report Settings*

Setting	Description
Report Title	Specifies a name for the Uniqueness report you are generating. The default title is Uniqueness Report.
Notes	Specifies optional details describing the report.
Display in Landscape mode	Displays the report suitable for landscape printing (11” x 8.5” rather than 8.5” x 11”).
Wrap text when values overflow	Allows cell data to wrap around within the report column. If you deselect this option, data that exceeds the width of the report column is truncated.

Setting	Description
Display statistical information	Displays a statistical summary of the report data, including number of records, duplicate count, a uniqueness measure, and the total number of duplicate records.
Display graphic	Displays a graphical (bar graph) representation of the report data.
Display uniqueness data	Displays a list of the duplicate values along with the number of times the value occurs in the data set instance.
Max number of duplicate records to display	Specifies the maximum number of duplicate values to display. This setting is active only when **Display duplicate data** is selected.

Matching

The Matching report settings lets you specify the following default settings.

Table 2-7 *Matching Report Settings*

Setting	Description
Report Title	Specifies a name for the Matching Value report you are generating. The default title is Matching Report.
Notes	Specifies optional details describing the report.
Display in Landscape mode	Displays the report suitable for landscape printing (11” x 8.5” rather than 8.5” x 11”).
Wrap text when values overflow	Allows cell data to wrap around within the report column. If you deselect this option, data that exceeds the width of the report column is truncated.
Display statistical information	Displays a statistical summary of the report data, including number of records, duplicate count, a uniqueness measure, and the total number of duplicate records.
Display graphic	Displays a graphical (bar graph) representation of the report data.
Display unmatched data	Displays a list of unmatched values from the matching analysis.
Max number of unmatched records to display	Specifies the maximum number of unmatched values to display. This setting is active only when **Display unmatched data** is selected.
Display matched data	Displays a list of matched values from the matching analysis.
Max number of matched records to display	Specifies the maximum number of matched values to display. This setting is active only when **Display matched data** is selected.

Connection Settings

Analyzer connection settings govern the connections between Analyzer and your applications (data sources.) To access these settings from the Project view, right-click a connection profile, then select **Properties**.

Analyzer organizes its connection settings into two pages:

- "General Connection Info" on page 33
- "IDS Configuration" on page 33

General Connection Info

The General Info page displays the **Name** and **Notes** fields for the selected connection profile. You can modify the values in both of these fields. Click **Apply** or **OK** to save your changes.

IDS Configuration

The IDS Configuration page lets you configure the Independent Driver Services (IDS) settings for the selected connection profile. To access the IDS Configuration properties, right-click the appropriate connection profile in the Project View, then select **Properties** > **IDS Configuration**.

The IDS Configuration page has settings organized into the following tabs:

- "Configuration" on page 33
- "Authentication" on page 34
- "Parameters" on page 34

Configuration

The **Configuration** tab specifies whether the application driver used by the connection profile is locally or remotely hosted. When it is locally hosted, Analyzer displays the application driver associated with this connection profile. Click **Add Libraries** if the driver needs a third-party library to connect to the application.

When remotely hosted, Analyzer uses Identity Manager's Remote Loader to access a remotely hosted application driver. For more information about the Remote Loader, see Configuring the Remote Loader and Drivers in the *NetIQ Identity Manager Setup Guide for Linux* or Configuring the Remote Loader and Drivers in the *NetIQ Identity Manager Setup Guide for Windows*.

When connecting to a driver remotely, select **Enable SSL** to encrypt remote loader communications. You can select the self-signed default certificate, or specify an existing certificate. Either way, make sure the same certificate is installed on the Remote Loader server for the secure communications to work. For more information about secure connections with the Remote Loader, see Configuring the Remote Loader and Drivers in the *NetIQ Identity Manager Setup Guide for Linux* or Configuring the Remote Loader and Drivers in the *NetIQ Identity Manager Setup Guide for Windows*.

NOTE: When configuring SSL to use a default certificate, Analyzer gives you the opportunity to export the certificate to a file so you can install it on the Remote Loader server.

Authentication

The **Authentication** tab specifies the application-specific access credentials that Analyzer uses to connect to the application.

Parameters

The **Parameters** tab specifies the driver-specific settings available for this driver. For more information about a driver's settings, see the appropriate Identity Manager driver implementation guide (https://www.netiq.com/documentation/identity-manager-48-drivers/).

The LDAP driver includes the **Use SSL** setting that determines whether the driver uses SSL for communicating with the target application.When this setting is enabled, you must specify the path to the local keystore that contains the application certificate, or CA signing certificate. You can also enable Mutual Authentication, if desired.

NOTE: Prior to using SSL, you must manually export the application certificate and import it into your local keystore by using the Java keytool utility. More information about keytool is available in the Java keytool documentation (http://download.oracle.com/javase/1.5.0/docs/tooldocs/windows/keytool.html).

Flat File Settings

Analyzer flat file settings govern the connections between Analyzer and your data files. Analyzer organizes its connection settings into three pages:

- "File Path" on page 34
- "Delimiter Settings" on page 34
- "Field Names" on page 35

To access these settings from the Project view, right-click a flat file profile, then select **Properties**. You can modify the settings in any of these pages. Click **Apply** or **OK** to save your changes.

File Path

The File Path page contains the following settings:

Table 2-8 *File Path Settings*

Setting	Description
File Path	Specifies the path to the desired file.
File Path > Encoding	Specifies the character encoding used in the specified file. The default setting is UTF-8.

Delimiter Settings

The Delimiter Settings page contains the following settings:

Table 2-9 Delimiter Settings

Settings	Description
Name	Specifies a name for the flat file profile. This name is used for the profile object in the Project View. By default, this is the filename, but you can use any name you want.
Description	(Optional) Specifies a more detailed description for the flat file profile.
Object Class	Specifies the object class to which the attribute values in the selected file should be associated. For example, a file of user information should be associated with the User class.
Field Delimiter	Specifies the field delimiter used in the selected file. The default setting is a comma.
Text Delimiter	Specifies the character used to enclose values that contain the field delimiter as part of the value. You can choose between double quotes (“”) and single quotes (‘’).
First Row Contains Field Names	When this option is selected, Analyzer treats the first row of the data file as field names. Analyzer treats these field names as attribute names, so they should match the appropriate attribute names in the associated Object Class.

Field Names

The Field Names page lets you modify the field names used in the flat file profile.

By default, Analyzer displays generic field names if **First Row Contains Field Names** is not selected in the Delimited Text Settings dialog box. If **First Row Contains Field Names** is selected, Analyzer lists the values in the first data row as the field names. Regardless of the values initially displayed, you can adjust the field names after they are created.

Data Set Settings

Analyzer data set settings let you view and modify the configuration of existing data sets. To access these properties from the Project view, right-click a Data Set definition, then select **Properties**.

NOTE: You cannot modify an existing data set definition if you have a data set instance based on the current data set definition. To modify an existing data set definition, first delete any existing data set instances.

Analyzer organizes the data set properties into three pages:

General Info: The General Info page displays the **Name** and **Notes** fields for the selected connection profile. You can modify only the **Notes** field. Click **Apply** or **OK** to save your changes.

Classes Attributes Filter: The Classes Attributes Filter page includes tabs that let you view and modify the schema class, class attributes, and included in the data set. These settings are interdependent, so modifying class or attribute settings might require you to reconfigure other settings. Analyzer informs you if this is the case.

NOTE: By default, a data set definition includes all values in a multi-valued attribute when importing a data set. If you just want a single value imported for each multi-valued attribute, deselect **Enable multi-valued attribute support** on the **Attributes** tab.

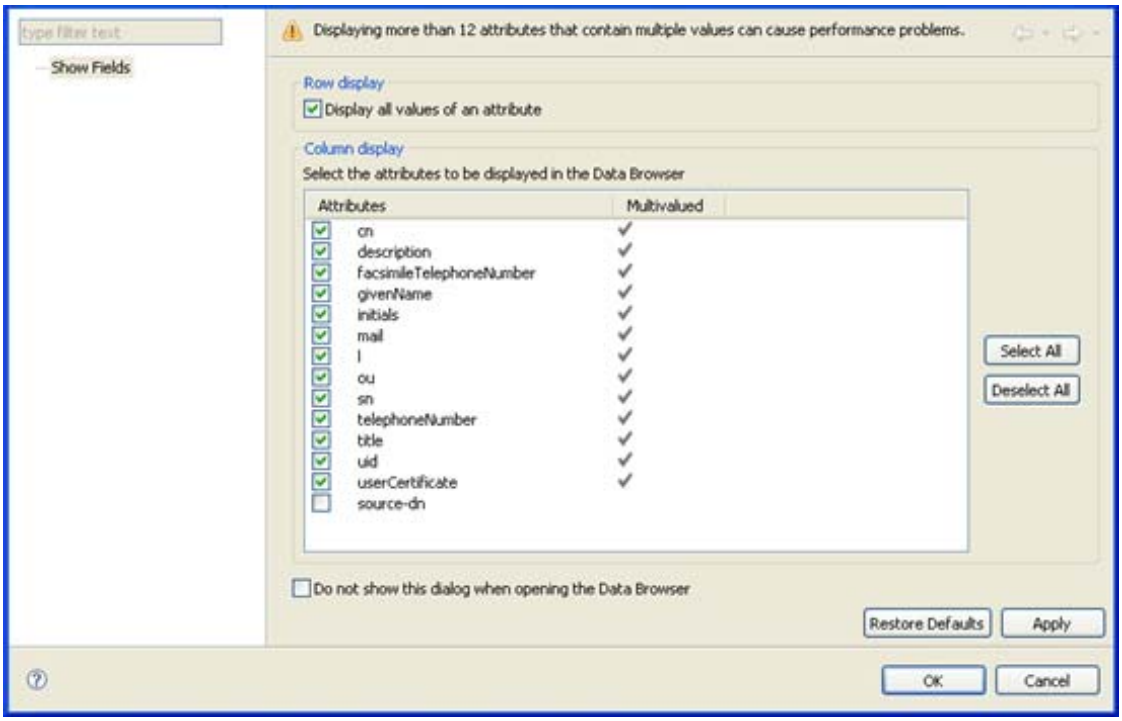
Search Scope: The Search Scope page displays the data set's scope. This setting is typically used with directory-based data sets.

Analyzer does not prevent users from modifying anything in a data set. If a user with appropriate rights to the source application modifies a value, for example a GUID or DN, Analyzer does not attempt to determine if the modification causes a problem when written out to the source application. To avoid causing unintended problems in the source application, users should be careful when modifying data and sending those modifications to the source application.

Analyzer allows a maximum of 41 columns in a data set. This is primarily due to the MySQL server limitation of 64K per column.

Data Browser Settings

The Data Browser settings govern how the Data Browser displays data imported from an application. To access these settings, select **Display Settings** in the Data Browser toolbar.



The Data Browser settings page lets you manage the following:

Table 2-10 Data Browser Settings

Setting	Description
Display all values of an attribute	Enables display of attributes with multiple values in the Data Browser. When disabled, the Data Browser displays only the first attribute value.
Select the attributes to be displayed in the Data Browser	Specifies the attributes (columns) in the current data set definition to display in the Data Browser. Limiting the number of displayed attributes that contain multiple values improves Data Browser performance.
Do not show this dialog when opening the Data Browser	Disables launching the Data Browser settings dialog box when Analyzer detects a large number (more than 12) of attributes with multiple values in the current data set definition. You can also enable or disable this feature from the Data Browser preferences (see "Data Browser" on page 27).

IDS Trace Settings

The IDS Trace settings let you configure Analyzer's trace environment. To access these settings from the IDS Trace view, click **Preferences** in the IDS Trace toolbar.

The IDS Trace settings let you specify three parameters:

Table 2-11 IDS Trace Settings

Setting	Description
Font	Specifies font characteristics for the IDS trace, including font, font size, and font style. Click **Change Font** to modify these settings.
Level	Specifies the level of detail to display in the IDS trace. The default trace level is 3, which is verbose and can cause performance problems. Only use it when troubleshooting the IDS connection. To improve performance, reduce the trace level to 1 for use during regular operations.
Maximum line to retain	Specifies the length of the IDS trace, in lines of text, before the trace starts dropping the oldest trace entries.

The IDS Trace view consumes significant resources. You should only open the IDS Trace view when you need the information. The IDS Trace level is set to 3 by default in order to track connection problems and errors. This trace level can cause performance issues with data browsing.

3 Using Analyzer

This section describes the primary tasks associated with using NetIQ Analyzer for Identity Manager. Tasks are presented in the order they would typically be used.

- "Creating a Project" on page 39
- "Adding a Connection Profile" on page 42
- "Creating a Flat File Profile" on page 45
- "Modifying a Schema Map" on page 48
- "Creating a Data Set Definition" on page 50
- "Viewing Application Data" on page 52
- "Inspecting Data" on page 53
- "Analyzing Data" on page 56
- "Cleaning Data" on page 65
- "Comparing Data" on page 67
- "Auditing and Reporting" on page 71

Creating a Project

After installing Analyzer, you can create a project for inspecting and analyzing application data. There are multiple ways to create a project.

- "Creating a New Identity Manager Project" on page 39
- "Importing an Existing Project" on page 39
- "Importing from Additional Sources" on page 40

Creating a New Identity Manager Project

1. Launch Analyzer.
2. Close the Welcome view to access the Project view.
3. Click **New Identity Manager Project**, or select **File > New Project** from the toolbar.
4. In the New Identity Manager Project dialog box, specify a name for the new project, then click **Finish**.

   A project object with the specified name appears in Analyzer's Project view.

Importing an Existing Project

1. Launch Analyzer.
2. Close the Welcome view to access the Project view.

3 Click **File System**.

4 Select whether the `.project` file is in an archive file or in a directory.

5 Click **Browse**, then browse to and select the archive file or the root directory.

6 Click **Finish** to import the `.project` file.

Importing from Additional Sources

When you import files into Analyzer to create a new project, there are different sources for the import. If you are only importing an existing project, follow the procedure in Importing an Existing Project. If the information you are importing is not an existing project, follow this procedure.

1 Launch Analyzer.

2 Close the Welcome view to access the Project view.

3 Select **File > Import** from the toolbar.

4 Select an import source, then click **Next**.

   Each import source displays a different configuration screen. For information on each import source, see Table 3-1

5 Click **Finish** to import the source.

There are two main categories for the import sources: general or Designer for Identity Manager.

Figure 3-1 *Import Source Categories*

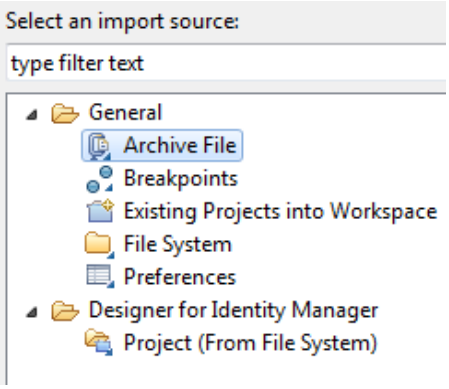


Table 3-1 *Import Sources Descriptions*

Import Sources	Description
Archive File	Imports files that are archived. ◆ **From archive file:** Browse to and select the archived file. ◆ **Filter Types:** Allows you to filter for specific types of files in the archive. ◆ **Select All:** Allows you to select all of the files in the archive to import. ◆ **Deselect All:** Allows you to deselect all of the files in the archive to import. ◆ **Into folder:** Browse to and select the destination location for the archived project. If you have multiple workspaces, you can import this file into multiple locations. ◆ **Overwrite existing resources without warning:** Select this option if you want your current project overwritten without warning.
Breakpoints	Imports predefined breakpoints for your project. ◆ **From file:** Browse to and select the breakpoint file. ◆ **Automatically update existing breakpoints:** Updates existing breakpoints. ◆ **Automatically create breakpoint working sets:** Creates the breakpoints in the working sets of data.
Existing Projects into Workspace	Imports an existing Analyzer project. ◆ **Select root directory:** Browse to and select the root directory of the existing Analyzer project. ◆ **Select archive file:** Browse to and select the Analyzer project that is in an archived file. ◆ **Projects:** If there are multiple projects in a folder or archived file, you can select one or more projects to import. ◆ **Copy projects into workspace:** Select this option to copy the projects into the existing workspace.
File System	Imports an existing project from the file system. ◆ **From directory:** Browse to and select the root directory of the Analyzer project. ◆ **Filter Types:** Allows you to filter for specific types of files in the directory. ◆ **Select All:** Allows you to select all of the files in the directory to import. ◆ **Deselect All:** Allows you to deselect all of the files in the directory to import. ◆ **Into folder:** Browse to and select the destination location for the Analyzer project. If you have multiple workspaces, you can import this file into multiple locations. ◆ **Overwrite existing resources without warning:** Select this option if you want your current project overwritten without warning. ◆ **Create complete folder structure:** Creates the complete folder structure for the project. ◆ **Create selected folders only:** Creates only the selected folders of the project.

Import Sources	Description
Preferences	Imports preferences from another instance of Analyzer. ◆ **From preference file:** Browse to and select the file that contains the preferences. ◆ **Import all:** Imports all preferences contained in the file. ◆ **Choose specific preferences to import:** Allows you to import specific preferences instead of all of the preferences defined in the file. ◆ **Preferences:** Select the desired preferences from the list. ◆ **Description:** A description of the selected preference. ◆ **Select All:** Selects all preferences. ◆ **Deselect All:** Deselects all preferences.
Project (From File System)	Imports a Designer project from the file system. ◆ **Select root directory:** Browse to and select the root directory of the Designer project. ◆ **Select archive file:** Browse to and select the Designer project that is in an archived file. ◆ **Projects:** If there are multiple projects in a folder or archived file, you can select one or more projects to import. ◆ **Copy projects into workspace:** Select this option to copy the projects into the existing workspace. ◆ **Open projects after import:** Select this option to automatically open the Designer project in Analyzer after the import process completes.

Adding a Connection Profile

After defining a project, you can create a connection profile, which provides the driver details necessary for Analyzer to communicate with the specified application data store. There are three ways to create connection profiles in your Analyzer project:

- "Creating a Connection Profile" on page 42
- "Importing a Connection Profile Information from Designer" on page 44
- "Importing a Connection Profile from a File" on page 45

NOTE: If you need to get data from a file, such as a spreadsheet, see "Creating a Flat File Profile" on page 45.

Creating a Connection Profile

To create a new connection profile from scratch:

1. In the Project view, expand the project with which you want to work.
2. Right-click the Connections object, then click **New Connection**.

3 In the Select an Application dialog box, select the type of application for which you want the connection profile, then click **Next**.

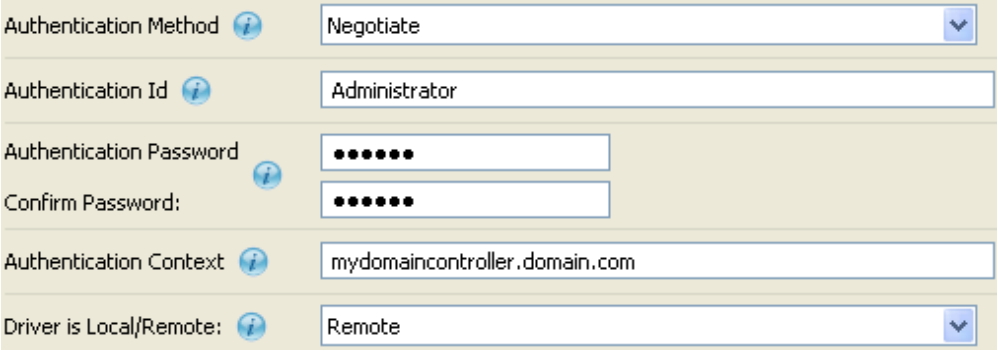
The category list organizes the many applications into similar groups. Selecting a category displays a subset of all available applications.

4 In the Select Driver Configuration dialog box, select the XML configuration file for the selected application, then click **Next**.

5 In the Name the Connection Profile dialog box, specify a name for the connection profile, then click **Next**.

Optionally, you can provide a detailed profile description in the **Notes** field.

6 In the Connection Settings dialog box, specify the necessary connection information to connect to the selected application, then click **Next**.



The particular settings vary by application, but common settings include the Hostname (or IP address), Username, Password, and whether the driver is Local (on the machine running Analyzer) or Remote (available through Remote Loader.)

NOTE: Do not use the following special characters in the password field: < &

Depending on the application, Analyzer might display more than one Connection Settings dialog box to collect the information it needs to configure the application connection.

7 In the Summary dialog box, review the settings for the connection profile.

- If necessary, click **Add Libraries** to add a third-party library (typically a `.jar` file) for your application. You can add these after the fact by selecting **Window** > **Preferences** and selecting **Analyzer** > **Connections**.
- Select **Create Data Set Definition** to have Analyzer automatically create default data set definitions based on information provided in the connection's preconfig settings. For more information about data set definitions, see "Creating a Data Set Definition" on page 50.

8 Click **Finish**

After creating the connection profile, you must start the driver in order to query the data store. You can do this when creating a data set that utilizes the driver, or you can right-click the connection profile in the Project View and select **Start Driver**.

NOTE: To use the SAP User driver, you must install the `sapjco.jar` library in Analyzer, and install the `librfc32.dll` and `sapjcorfc.dll` into the `Windows %systemroot%` folder (typically `C:\windows\system32`). Restart Analyzer after installing these files.

The Analyzer DB2 driver requires the `db2java.zip` and `db2jcc.jar` libraries to function properly. You can download these libraries from IBM (https://www.ibm.com/developerworks/).

Importing a Connection Profile Information from Designer

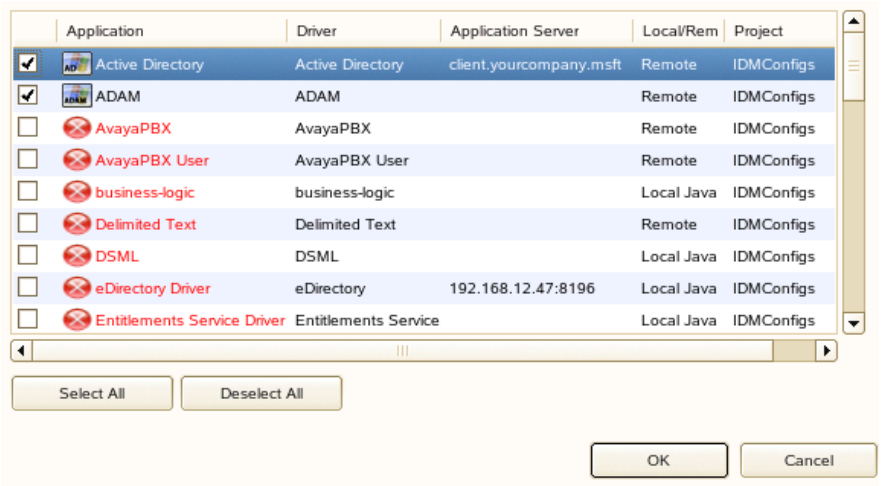
If your project is based on a Designer project, the information for each driver can be reused in Analyzer.

To import a connection profile based on an existing Designer project:

1 In the Project view, expand the project with which you want to work.

2 Right-click the Connections object, then click **Import from Designer**.

3 In the Application Found dialog box, select the applications you want to import as connection profiles, then click **OK**.



By default, Analyzer selects all applications for which Analyzer drivers exist. Applications for which Analyzer does not have drivers display in red.

IMPORTANT: Analyzer supports some Designer applications only through Identity Manager's Remote Loader. If you have such an application with a local configuration, Analyzer prevents you from importing it as a connection profile until you reconfigure the application to use Remote Loader. For more information about the Remote Loader, see the *Installing Remote Loader* in the *NetIQ Identity Manager Setup Guide for Linux* and *Installing Remote Loader* in the *NetIQ Identity Manager Setup Guide for Windows*.

If you have several Designer applications, click any of the column headings in the Applications Found dialog box to sort the list based on that column. Click once to sort in ascending order. Click again to sort in descending order.

4 In the New Connection Creation dialog box, click **OK**.

Analyzer lets you import the same Designer application into multiple connection profiles so you can create multiple variations for inspecting and analyzing data.

After importing a connection profile from a Designer project, you must start the driver in order to query the data store. You can do this when creating a data set that utilizes the driver, or you can right-click the connection profile in the Project View and select **Start Driver**.

NOTE: If the Analyzer driver won't start, you might need to add a third-party library (typically a `.jar` file) for the application. You can add third-party libraries by selecting **Window** > **Preferences > > Analyzer** > **Connections**, then click **Add** to select the `.jar` file.

Importing a Connection Profile from a File

To import a connection profile from a previously exported connection profile file:

1 In the Project view, expand the project with which you want to work.

2 Right-click the Connections object, then click **Import from file**.

3 Browse to and select the connection profile file you want to import, then click **Open**.

After importing a connection profile, you must start the driver in order to query the data store. You can do this when creating a data set that utilizes the driver, or you can right-click the connection profile in the Project View and select **Start Driver**.

NOTE: If an imported connection profile won't start, you might need to add a third-party library (typically a `.jar` file) for the application. You can add third-party libraries by selecting **Window** > **Preferences > > Analyzer** > **Connections**, then click **Add** to select the `.jar` file.

Creating a Flat File Profile

If you need to access data that is in a flat file, such as a comma-delimited file or spreadsheet, create a flat file profile. Flat file profiles provide the details necessary for Analyzer to access the specified data file.

NOTE: If you need to get data from an application data store, see "Adding a Connection Profile" on page 42.

1 In the Project view, expand the project with which you want to work.

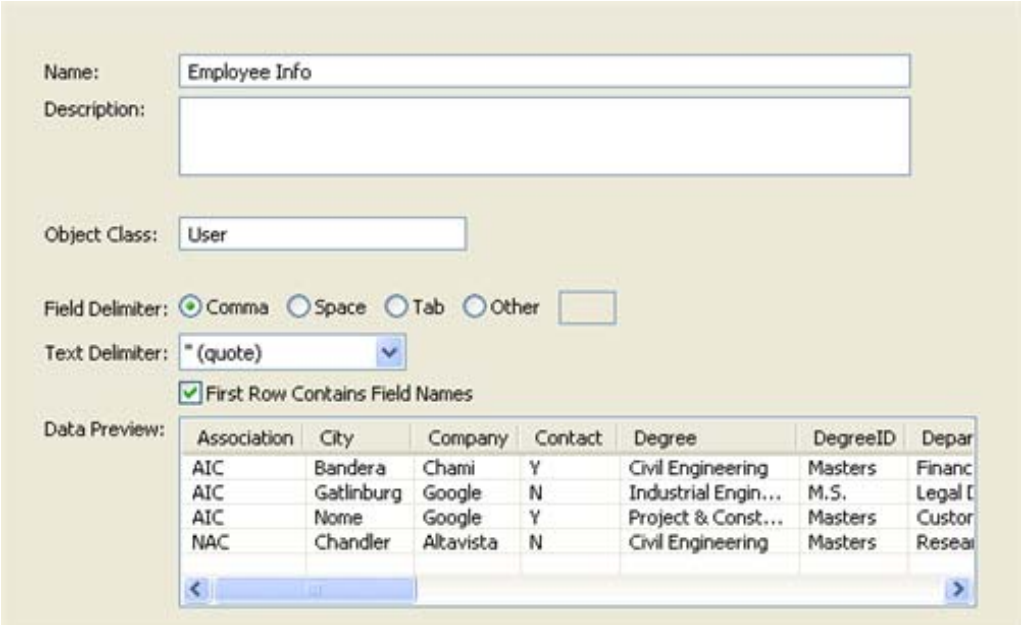
2 Right-click **Flat Files**, then select **New Flat File**.

3 In the Import Delimited Text File dialog box, provide the following information, then click **Next**.

File Path: Specify the path to the desired file.

Encoding: Specify the character encoding used in the specified file. The default setting is UTF-8.

4 In the Delimited Text Settings dialog box, provide the following information about the selected file, then click **Next**.



Name: Specify a name for the flat file profile. This name is used for the profile object in the Project View. By default, this is the filename, but you can use any name you want.

Description: (Optional) Specifies a more detailed description for the flat file profile.

Object Class: Specify the object class to which the attribute values in the selected file should be associated. For example, a file of user information should be associated with the User class.

Field Delimiter: Specify the field delimiter used in the selected file. The default setting is a comma.

Text Delimiter: Specify the character used to enclose values that contain the field delimiter as part of the value. You can choose between double quotes ("") and single quotes ('').

First Row Contains Field Names: When this option is selected, Analyzer treats the first row of the data file as field names. Analyzer treats these field names as attribute names, so they should match the appropriate attribute names in the associated Object Class.

5 (Conditional) In the Set Field Names dialog box, modify field names and data types as needed, then click **Finish**.

Field Name	Value	Type
Association	AIC	string
City	Bandera	string
Company	Chami	string
Contact	Y	string
Country	Christmas Island	string
Currency	Afghanistan Afghani	string
Degree	Civil Engineering	string
DegreeID	Masters	string
Department	Finances	string
Description		string
Email	magna@egestashendrerit.edu	string
EmployeeID	1000	string
Favorite_Color	violet	string
First	Jakeem	string

Field Name: To adjust a field name for a particular column (attribute) in the flat file data table, double-click the current Field Name entry and make your changes.

Value: Displays a sample value for the associated flat file data column. Values are for information only and cannot be modified.

Type: To adjust the data type for a particular column (attribute) in the flat file data table, double-click the Type entry and select the appropriate data type from the drop-down list. Analyzer's base schema provides the following Types:

Type	Description
string	Specifies that the attribute value is a character string.
teleNumber	Used by attributes whose values are telephone numbers. The length of telephone number strings must be between 1 and 32 characters. Two telephone numbers match when they are of the same length and their corresponding characters are identical, except that all spaces and hyphen characters are ignored during comparison.
int	Used by attributes represented as signed numeric values. Two Integer values match if they are identical. The comparison for ordering uses signed integer rules.
state	Used by attributes whose values are True (represented as 1) or False (represented as 0). The single-valued flag is set for this syntax type.
counter	Used by attributes whose values are incrementally modified numeric signed integers. Any attribute defined using Counter is a single-valued attribute. This syntax differs from Integer in that any value added to an attribute of this syntax is arithmetically added to the total, and any value deleted is arithmetically subtracted from the total.
dn	Used by attributes whose values are the names of objects in the eDirectory tree. Distinguished Names (DN) are not case sensitive, even if one of the naming attributes is case sensitive.

Type	Description
interval	Used by attributes whose values are signed numeric integers and represent intervals of time. The Interval syntax uses the same representation as the Integer syntax. The Interval value is the number of seconds in a time interval.
octet	Used by attributes whose values are strings of binary information not interpreted by eDirectory. These octet strings are non-Unicode strings. For two octet strings to match, they must be the same length, and the corresponding bit sequence (octet) must be identical.
time	Used by attributes whose values are unsigned integers and represent time expressed in seconds.

Modifying a Schema Map

Analyzer creates a schema map object in each connection profile and flat file profile. The schema map associates an application's schema attributes to the corresponding schema attributes in Analyzer's base schema. This lets you ensure that your data analysis and cleaning operations properly associate similar values between the disparate systems.

Using a schema map allows you to create analysis profiles independent from a specific connection profile or flat file profile. You can create a single analysis profile and apply it to multiple connection or flat file profiles because the analysis profile leverages the schema map to provide the schema information it needs to perform its analysis.

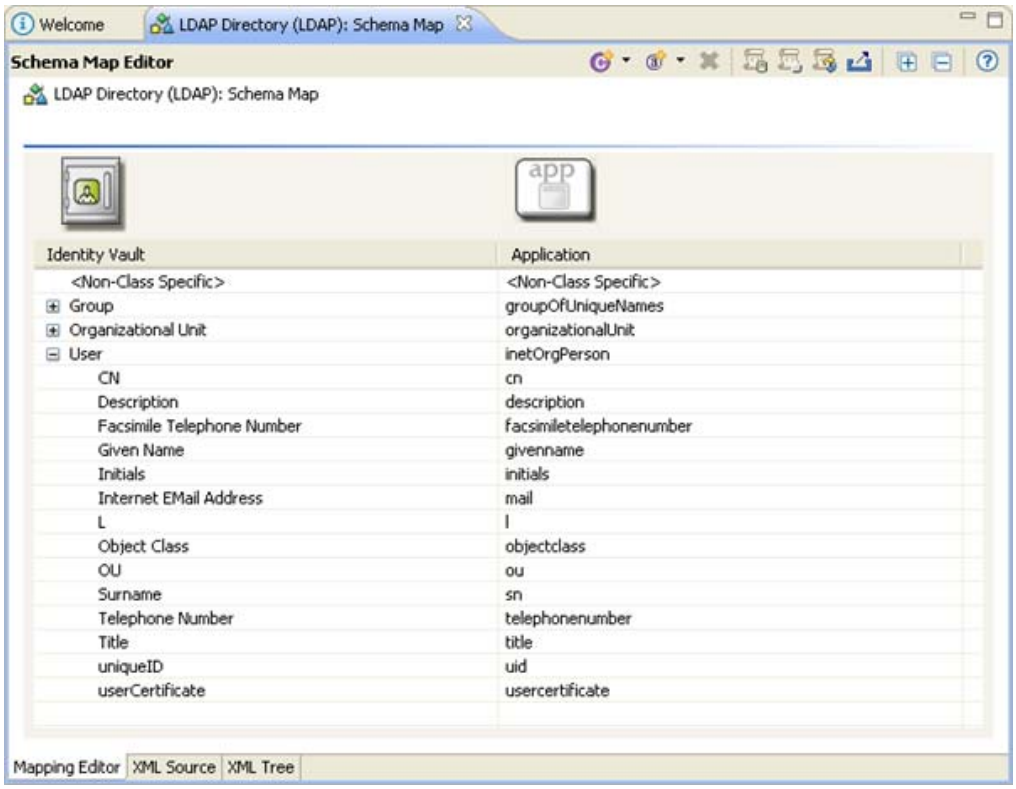
Use the Schema Map Editor to create or modify schema map settings for a given connection profile or flat file profile. The Schema Map Editor lets you do the following:

- Modify Analyzer's base schema classes and attributes
- Modify application schema classes and attributes
- Add class mappings
- Add attribute mappings

The Schema Map Editor includes several icons in a toolbar. For more information about these icons, see Schema Map Editor (page 10). The Schema Map Editor provides access to the Schema Manager from Designer for Identity Manager. For detailed information about Designer's Schema Manager, see "Managing the Schema" in the *NetIQ Designer for Identity Manager Administration Guide*.

To modify a schema mapping:

1. In the Project View, double-click the **Schema Map** icon to open the Schema Map editor.



Depending on the preconfigured settings for the selected application, the schema map might be prepopulated with basic mapping information.

2 In the Schema Map editor, specify a class mapping between the Identity Vault and the Application if none exists:

   2a Click **Add Class Mapping**.

   2b Under the Identity Vault, select the class you want to map from the drop-down list.

   2c Under the Application, select the class you want to map from the drop-down list.

3 In the Schema Map editor, specify the appropriate attribute mappings between the Identity Vault and the Application.

   3a Select the class mapping for which you want to specify attribute mappings.

   3b Click **Add Attribute Mapping**.

   3c Under the Identity Vault, select the attribute you want to map from the drop-down list.

   3d Under the Application, select the attribute you want to map from the drop-down list.

   3e Repeat steps 3a through 3d as needed to map the necessary attributes.

4 Click **OK** to save the changes.

Analyzer performs its data analysis solely based on the attribute name, and does not take the class name into account. Therefore, if you map attributes from different classes to the same application attribute, the analysis tests only the first mapped attribute it encounters. For example, in the following schema map, Analyzer tests only the name attribute mapped to the Group class, and ignores the mapping in the User class.

```
Class = Group
  |___ Attribute = gname ---> name
```

```
Class = User
  |___ Attribute = uname ---> name
```

This issue might also exist with the preconfigured schema maps that Analyzer includes with its drivers. The mappings might be correct to the attribute name, but not the class name.

Creating a Data Set Definition

A data set definition specifies the query used to generate a particular data set from the selected connection or flat file profile. You can create multiple data set definitions for the same profile in order to generate different data set instances.

1. In the Project View, browse to and select the profile for which you want to create a data set, then click **New Data Set**.
2. In the **Name** field of the Data Set Definition dialog box, specify a name for the new data set definition, then click **Next**.

   Optionally, provide a detailed description of the data set definition in the **Notes** field.
3. In the Select Classes dialog box, select the schema classes you want to use in the data set definition, then click **Next**.

   The Data Set Definition Wizard displays all available object classes for the selected profile. A flat file profile displays only the class you specified when creating the profile. For more information, see "Creating a Flat File Profile" on page 45.

   This dialog provides the following icons and options:

Icon or Option	Description
>	Adds the selected classes to the data set definition. Select multiple classes by using Ctrl+click and Shift+click. Alternatively, you can double-click a class to move it from one column to another.
<	Removes the selected classes from the data set definition. Select multiple classes by using Ctrl+click and Shift+click. Alternatively, you can double-click a class to move it from one column to another.
>>	Adds all available classes to the data set definition.
<<	Removes all classes from the data set definition.
Refresh Schema	Queries the application for an updated set of schema classes. Try using this option if the **Available Classes** list is empty.

4. In the Select Attributes dialog box, select the attributes that you want to include in the data set, then click **Next** or **Finish**.

   This is the last required step in the Data Set Definition Wizard. Click **Finish** to close the Wizard, or click **Next** to continue with some optional DSD configuration steps.

   This dialog provides the following icons and options:

Icon or Option	Description
Enable multi-valued attribute support	Select **Enable multi-valued attribute support** if you want to be able to see all attributes in a multi-valued attribute. This option is disabled if the application does not support multi-valued attributes. **IMPORTANT:** Multi-valued attribute support can add significant processing overhead to Analyzer operations. Because of this, Analyzer restricts the number of attributes in a data set to 60 when multi-valued attribute support is enabled.
>	Adds the selected attributes to the data set definition. Select multiple attributes by using Ctrl+click and Shift+click. Alternatively, you can double-click an attribute to move it from one column to another.
<	Removes the selected attributes from the data set definition. Select multiple attributes by using Ctrl+click and Shift+click. Alternatively, you can double-click an attribute to move it from one column to another.
>>	Adds all attributes to the data set definition.
<<	Removes all attributes from the data set definition.
⇧	Moves the selected attribute up in the attribute list, thereby changing the ordering of the attributes in the data set definition.
⇩	Moves the selected attribute down in the attribute list, thereby changing the ordering of the attributes in the data set definition.

5 (Optional) In the Narrow Data Import dialog box, specify a filter to restrict the data imported with the data set definition, then click **Next** or **Finish**.

To create a data filter:

5a In the **Filter** drop-down list, select the attribute where you want to apply the filter.

5b In the **Pattern** field, specify the text pattern against which you want to filter the data.

You can specify whole values, or use the asterisk (*) as a wildcard to specify a portion of a value. For example, smi*, *ith, *mit*.

6 (Optional) In the Select Search Base dialog box, specify the root container for gathering data for the data set definition, then click **Finish**.

This option is only applicable to hierarchical databases such as LDAP directories. Leave this field blank to search from the application root.

You can edit an existing data set definition by double-clicking the data set definition object in the Project view. The editor lets you see the same pages used to create the data set definition.

NOTE: You cannot modify an existing data set definition if you have a data set instance based on the current data set definition. To modify an existing data set definition, first delete any existing data set instances.

Viewing Application Data

After creating a data set definition, you can generate data set instances based on the data set definition. After generating a data set instance, you can also export it to a flat file.

- "Importing Data" on page 52
- "Exporting Data" on page 52

Regardless of the method you use to generate a data set instance, Analyzer lists the currently available data set instances in the Data Set Instances view. To view a particular data set instance, double-click its entry to open a Data Browser editor for that data set instance. This lets you easily move between data set instances when analyzing and inspecting data.

Importing Data

When you import data, Analyzer queries the data source (either application or file) based on the selected data set definition, and copies the resulting data set into the Analyzer database. This creates a snapshot of the data set that remains available between Analyzer sessions. However, if the source data changes over time, the imported data set can grow stale and outdated. To update the data, you must re-import the data set.

NOTE: For connection profiles, you must start the connection driver to import data. To do so, right-click the connection profile in the Project view, then select **Start Driver**.

To import a data set:

1. In the Project view, right-click the data set definition object that you want to browse, then click **Import Data**.
2. (Conditional) Select the attributes with multiple values that you want to display in the Data Browser, then click **OK**.

   Data set instances that contain many attributes with multiple values can cause significant performance issues when inspecting and analyzing data. The Data Browser settings page lets you manage attributes with multiple values. If the data set you are importing contains many these attributes, Analyzer opens the Data Browser settings page during the import process.

   For more information, see "Data Browser Settings" on page 36.

Analyzer retrieves data from the selected data set definition and displays it in the Data Browser editor.

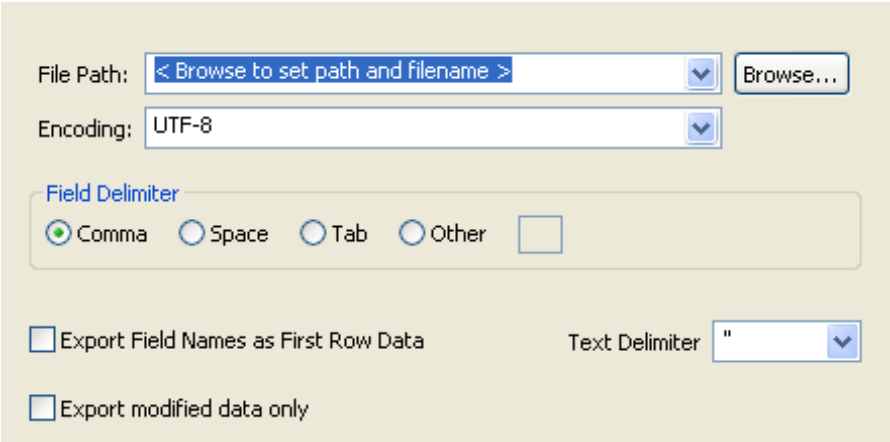
Exporting Data

Analyzer lets you export any data set instance to a flat file.

IMPORTANT: The Export To File Wizard takes into account any filters you have applied to the data set through the Data Browser editor. If you want to capture the entire data set, make sure you remove filters before exporting the data set to a file. For more information about filters, see "Filtering Data" on page 54.

To export a data set instance:

1 In the Data Set view, right-click the appropriate data set entry, then select **Export to File**.



2 In the Export to Flat File dialog box, specify the desired file characteristics, then click **OK**.

Analyzer creates the flat file based on your specifications.

File Path: Specify a name and location for the file.

Encoding: Specify the character encoding to use in the file. The default setting is UTF-8.

Field Delimiter: Specify the field delimiter to use in the file. The default setting is a comma.

Text Delimiter: Specify the character to use to enclose values that contain the field delimiter as part of the value. You can choose between double quotes (“”) and single quotes (’).

Export Field Names as First Row Data: When this option is selected, Analyzer includes the data set’s column headings as the first row of the file.

Export modified data only: Select this option to have the file contain only those values modified since Analyzer imported the data set.

3 (Optional) While Analyzer is creating the file, click **Details** to see more information about the file creation process. If the data set is large, click **Run in Background** to continue working while Analyzer creates the file.

Inspecting Data

After retrieving a data set, you use the Data Browser editor to examine its contents. To make this easier, the Data Browser lets you further refine the objects with which you are working by filtering the object list based on the criteria you select.

The Data Browser lets you perform three main tasks:

- “Filtering Data” on page 54
- “Inspecting the Current Data Set” on page 54
- “Modifying Data” on page 55

Filtering Data

The Data Browser's filter tool lets you restrict your view of the data set to those records that share a particular characteristic, such as all records that include the value "David".

There are two types of data set filters:

Quick Filter: The quick filter performs a global search across all attributes in the data set for those records that contain the filter string, and displays them in the Data Browser.

To specify a quick filter, simply type the desired value in the **Quick Filter** field, then press Enter.

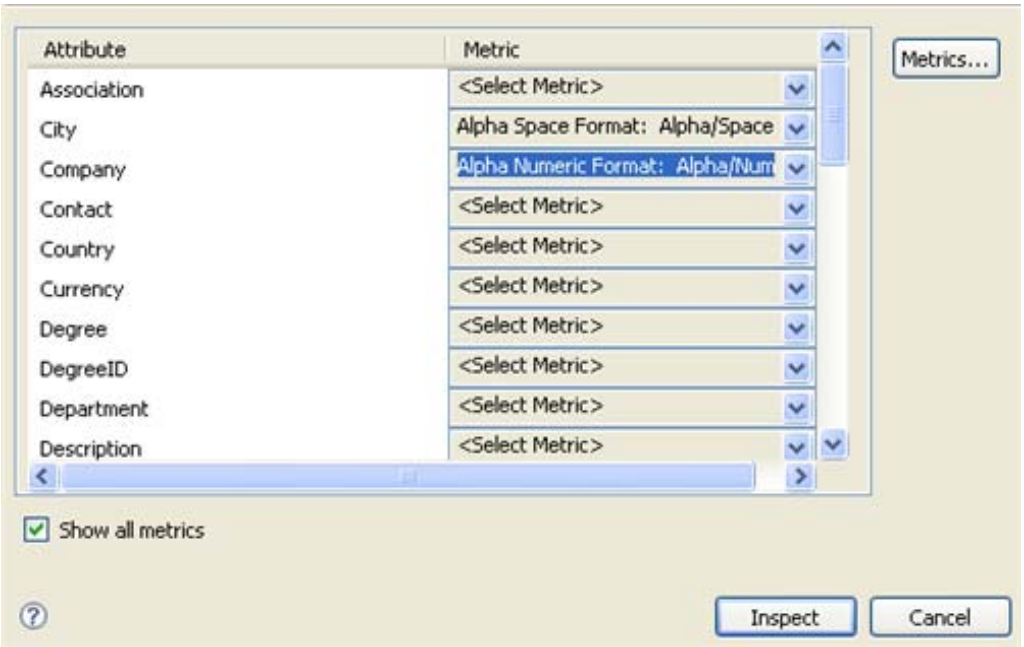
Detailed Filter: The detailed filter lets you specify additional filter information to create a more targeted data set filter. To specify a detailed filter:

1. In the Data Browser editor, click **Turn Filtering On or Off**
2. In the Filter dialog box, specify the filter attributes.
   - In the first field, select the attribute (column) on which you want to filter the data set.
   - In the second field, select the logical operator to use to filter the data set.
   - In the third field, specify the filter string.
3. Click **Apply**.

To remove an applied filter, click the **Filter** icon's drop-down arrow and select **Clear Filter**.

Inspecting the Current Data Set

1. In the Data Browser editor, click **Specify Metrics for Inspector**
2. In the Inspector dialog box, select the preferred data format for each attribute that you want to analyze, then click **Inspect**.



The drop-down menu for each attribute lists the appropriate metrics for the specified attribute. Select **Show all metrics** to display all the available metrics, even those that do not apply to the attribute's syntax.

If necessary, you can click **Metrics** to open the Metric Manager dialog box, where you can create or modify the existing metric list. For more information, see "Creating a New Test Metric" on page 57.

The Data Browser uses red highlights for the attribute values that do not conform to the specified data format, allowing you to modify the data as needed.

The Data Browser does not highlight an object's attribute if there is no defined value, or if the value already corresponds to the preferred data format.

To remove an applied filter, click the **Inspector** icon's dropdown arrow and select **Clear Inspector**.

Modifying Data

The Data Browser lets you update data in a data set. For example, you might update a phone number so that it conforms to the desired phone number format, or add a phone number that is not currently in the data set.

1. In the Data Browser editor, select the data cell you want to edit, then press Enter.

   If you are working with a large data set, you can filter the Data Browser editor down to a more relevant set of records. For more information, see "Filtering Data" on page 54.

2. Modify the cell data as desired.

   Single-value cell: Select a cell and press Enter, or double-click a cell to edit the cell value. Type your changes directly in the Data Browser editor. Press Delete to delete the value completely.

   Press Enter to capture your changes. When you select a different cell, the Data Browser marks the modified data cell so you can easily see the modified data.

   NOTE: To add a value to a single-value cell, press Insert to open the Multi-value Cell editor.

   Multi-value cell: Press Insert, Enter, or Delete to open the Multi-value Cell editor. Type your changes, then press **OK**.

   Click **New Value** to add a value to the cell. Click the **Delete** icon next to the value to delete it. Click the **Revert Value** icon next to the value to return to its original, unmodified value.

3. (Optional) Select **Export to File** to export the data set instance.

   This opens the Export to Flat File dialog box. For more information, see "Exporting Data" on page 52.

4. (Optional) Select **Generate a Data Browser Report** to generate a report about the data set instance.

5. (Optional) Select **Update changes to the application** to send data modifications back to the source application.

   This opens the Update to Application dialog box, where you can select the changes you want to write back to the application. Click **Update** to complete the data update process.

NOTE: Analyzer provides an audit trail for updates to application data. For more information, see "Data Browser Audit" on page 72.

Analyzing Data

After creating a data set instance, Analyzer lets you configure and perform analyses on those data set instances. Analyzer's data analysis refines and extends the data inspection capabilities available in the Data Browser.

Analyzing a data set involves two tasks:

- "Creating and Customizing an Analysis Profile" on page 56
- "Analyzing a Data Set" on page 64

Creating and Customizing an Analysis Profile

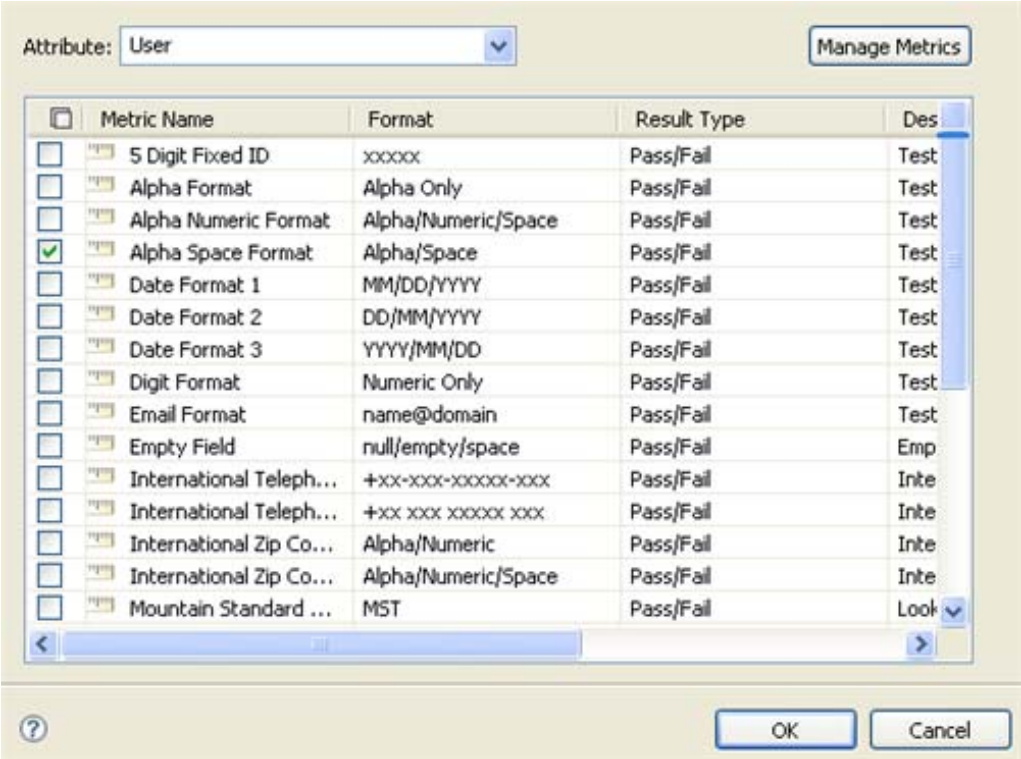
An *analysis profile* stores the details of the data analysis you want to perform. You can also create custom analysis metrics.

- "Creating an Analysis Profile" on page 56
- "Creating a New Test Metric" on page 57
- "Accessing Attributes within a Script" on page 61
- "Using Regex with Unicode Characters" on page 62

Creating an Analysis Profile

1. In the Project View, right-click the Analysis Profiles object, then select **New Analysis Profile**.
2. In the Name the Analysis Profile dialog box, specify a name for the analysis profile, then click **OK**.

   Analyzer creates a new Analysis Profile object in the selected project, and opens the Analysis Profile editor, from which you can finish configuring the analysis profile.
3. In the Analysis Profile editor, click **Add** to open the Analysis Metric Selector and configure an attribute-based analysis.



3a In the **Attribute** drop-down list, select the attribute to test.

3b Select the test metrics to use with this attribute, then click **OK**.

You can select from Analyzer's predefined metrics, or create a new metric. For more information, see "Creating a New Test Metric" on page 57.

4 (Optional) Repeat Step 3 above to configure additional attribute tests for the analysis profile.

5 Click **Save** in the toolbar, or click Ctrl+S to save the analysis profile.

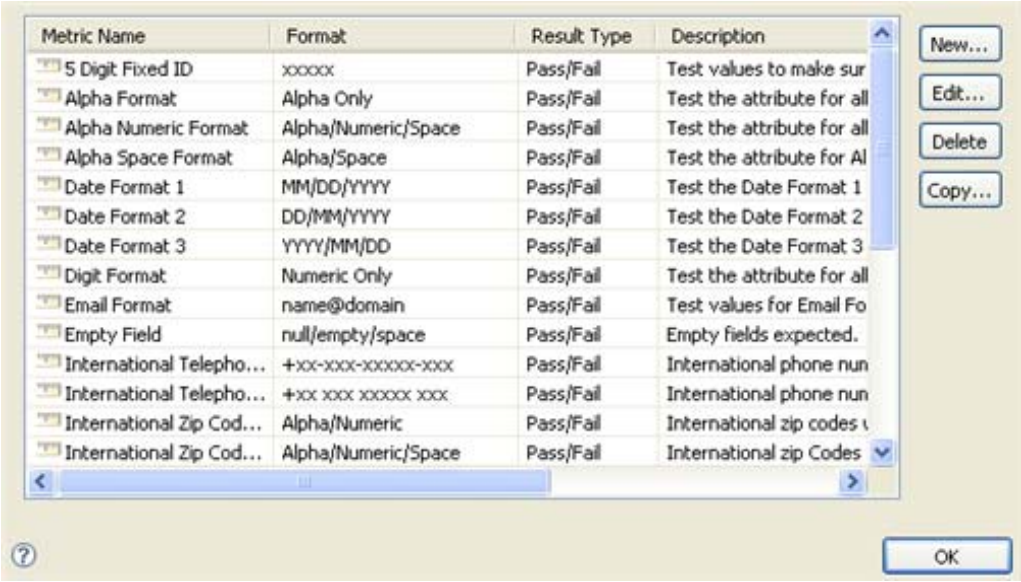
Creating a New Test Metric

Analyzer uses a set of predefined metrics, or you can create custom test metrics for use in data set analysis.

To create a new test metric:

1 In the Analysis Profile editor, click **Add**.

2 In the Analysis Metric Selector, click **Manage Metrics**.



3 In the Metric Manager, click **New**.

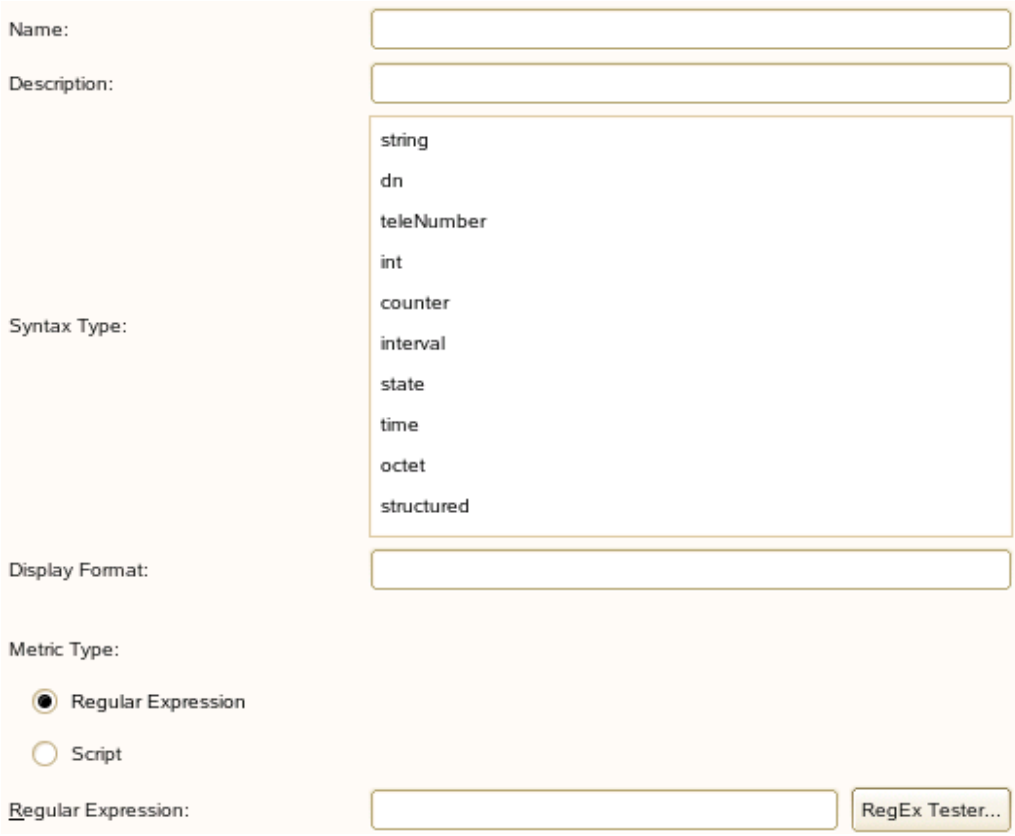
The Metric Manager provides three other icons for managing Analyzer's metric list:

Edit: Opens a dialog box from which you can edit the selected metric. You cannot edit predefined metrics.

Delete: Deletes the selected metric. You cannot delete predefined metrics.

Copy: Creates a custom metric based upon the currently selected metric. This is particularly useful for creating custom metrics based on predefined metrics.

4 In the New Metric dialog box, provide the requested information, then click **OK**.



Name: Specify a name for the test metric.

Description: Specify a longer text description for the test metric.

Syntax Type: Specify the data types (you can select more than one) for which Analyzer preselects this metric. For example, a String metric might automatically be associated with attributes of type String in the Analysis Profile editor.

Display Format: Describe the metric's regular expression in a human-readable format. This is informational only. For example, a telephone number metric might indicate its Display Format as (xxx) xxx-xxxx.

Metric Type: There are two different metric types: Regular Expressions and Script. Data tests can be performed with regular expressions or by using scripts. Select the desired type of metric and use the provided field to enter the regular expression or script.

- **Regular Expression:** Analyzer utilizes regular expression (regex) syntax to define metrics. You can view the regex syntax for a predefined metric by selecting it in Manage Metrics, then clicking **Edit**.

  NOTE: Regex is a complex language with unique syntax and rules. Make sure you are familiar with regex concepts and standards before trying to modify or create a metric. Analyzer metrics utilize regex as defined in the Java Platform Standard Edition 6 (http://download.oracle.com/javase/6/docs/api/java/util/regex/Pattern.html).

  For information about using regex with Unicode characters, see "Using Regex with Unicode Characters" on page 62.

- **Scripts:** Analyzer currently supports ECMAScript (the generic classification for JavaScript), Python, and Ruby. For more information, see:
  - ECMAScript Web page (http://www.ecmascript.org/).
  - Python Web page (https://www.python.org/)
  - Ruby Web page (http://www.ruby-lang.org/en/)

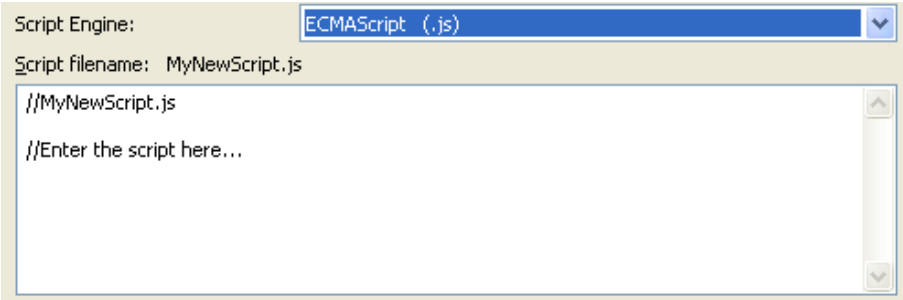
The following sections explain how implement the script metric.

- "Creating an Script Metric" on page 60
- "Copying and Modifying a Script" on page 60
- "Using the JavaScript Editor to Modify a Script" on page 60

Creating an Script Metric

The same script can be used to test and clean data. It can be applied as a test script or applied as a cleaning script. For more information about cleaning data, see "Cleaning Data" on page 65.

1. Select **Script** as the metric type.
2. Select the type of script:
   - **ECMAScript (.js)**
   - **python (.jy)**
   - **ruby (.rb)**
3. Create the script in the space provided, then click **OK**.



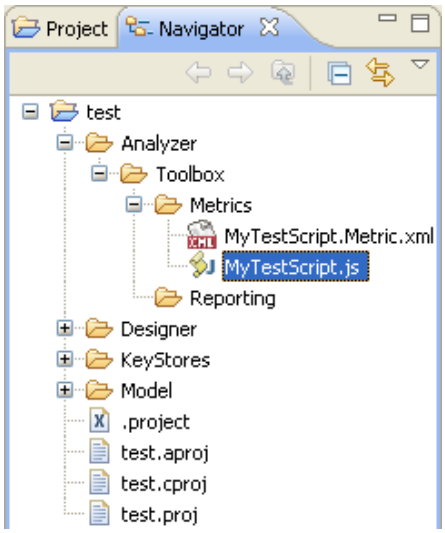
Copying and Modifying a Script

An easy way to create a new script is to copy an existing script, changing the **Name**, **Description**, **Syntax,** and **Format** fields as needed, then editing the script itself. A predefined script metric named Data Range can be used for this purpose. One advantage of this approach is that the predefined script contains comments describing how to access attributes within the script and how to return Pass/Fail status or cleaned values.

Using the JavaScript Editor to Modify a Script

An alternate way to modify a script is to use Eclipse's built-in JavaScript editor, which provides a full-screen color-coded view.

1. From the Analyzer toolbar select **Window** > **Show View** > **Navigator**.
2. In the Navigator view, select **project > Analyzer > Toolbox > Metrics > scripts**.



3 Right-click the `.js` file, then select **Open With > JavaScript Editor.**

4 Edit the script, then click **Save**.

5 Switch to the Data Browser view.

6 Right-click a cell in the desired column, then select **Apply Metric to column**.

7 Select the script file to run against the data.

Accessing Attributes within a Script

When you run a script, the value to be tested is sent into the script as the script variable `aimvalue`. The script should set the Boolean variable `aimReturnStatus` to `True` or `False` before returning. `True` indicates the value passed the test, and `False` indicates failure. For example, a very simple script to test that values don't exceed 64 characters might look like this:

```
aimReturnStatus = true;
  aimReturnValue = aimValue;

  if (aimValue != null && aimValue.length > 64)
     aimReturnStatus = false;
```

For an empty cell, for example an unpopulated value, `aimValue`, contains null, so the script needs to handle this.

If the script needs to access other attributes in the record in order to test the value in question, it can use script variables:

```
   aimValue              - the current value being tested
   aimArrayValue         - array of all values in the current cell
   aim_<attrName>        - the value of each attribute in the dataset
   aimArray_<attrName>   - array of values for each attribute

   aimReturnStatus       - boolean variable. Set to true if value passed
test, false if failed.
   aimReturnValue       - Used for data cleaning. Return with original value
to leave unmodified.
```

For example, if there is a Country attribute in the dataset, it can be accessed in the script as the variable `aim_Country`. If the attribute is empty, `aim_Country` contains null. If this attribute has multiple values, `aim_Country` contains the first value.

If the script needs to look at all values of a multi-valued attribute, it can use the `aimArrayValue` for an array of values in the current cell, or `aimArray_<attrName>` for an array of all the values of any attribute in the dataset. If a cell is empty the array variable is a 0-length array.

In this example, the metric returns a failure for each value of an attribute with more than one value.

```
aimReturnValue = aimValue;
aimReturnStatus = aimArrayValue.length > 1;
```

Using Regex with Unicode Characters

Click **RegEx Tester** to open a dialog box from which you can compile and test the regular expression before saving the new metric.

You can generalize any regex expression to work properly with Unicode characters. You can refer to Unicode characters in three different ways:

- "Character Code" on page 62
- "Character Category" on page 62
- "Character Blocks" on page 63

Character Code

Each Unicode character has a unique code, which takes the form `\u####`. For example:

- `\u0061` = a
- `\u0374-\u03F3` = Greek character set

Character Category

Each Unicode character is assigned to a character category, and potentially subcategory. This lets you easily refer to specific types of characters as a group. Character categories include Letters, Numbers, Punctuation, Symbols, Marks, Separators, and Other.

Some examples of using Unicode categories in regex include the following:

Regex	Description
`[A-Za-z]` or `\p{Alpha}`	ASCII characters from the Letters category.
`/p{L}`	A character from the Letters category, regardless of language.
`/p{Lu}`	A character from the Uppercase subcategory of the Letters category, regardless of language.

Regex	Description
`\p{Ll}`	A character from the Lowercase subcategory of the Letters category, regardless of language. **IMPORTANT:** Some letters are neither uppercase or lowercase, such as CJK Ideographs (Asian characters).
`[0-9]` or `\d`	ASCII digits 0 - 9.
`\p{Nd}`	A character from the Digits subcategory of the Numbers category, regardless of language. This includes Asian full-width digits.
`\p{punct}`	ASCII characters from the Punctuation category, which includes the following characters: !”#$%&'()*+,-./:;<=>?@[\]^_`{|}~.
`\p{P}`	Any character from the Punctuation category, regardless of language.
<*space*>	A space character.
`\s` or `\p{Space}`	Any whitespace character, which includes the following: <*space*> \t \n \x0B \f \r.
`\p{S}`	Any character from the Symbol category, regardless of language.
`\p{Sc}`	A character from the Currency subcategory of the Symbols category, regardless of language.
`[^A-Za-z]`	Any character that is not an ASCII letter.
`\P{L}`	Any character that is not in the Letters category, regardless of language.

NOTE: For more information on Regular Expression Categories and Subcategories, see Unicode Technical Standard #18 (http://www.unicode.org/reports/tr18/).

Character Blocks

Each Unicode character is part of a block, or character set. You can refer to entire blocks of Unicode characters by specifying the block name proceeded by “In.” For example:

Regex	Description
`\p{InGreek}`	Any character in the Greek block.
`^[\p{InBasic Latin}\p{InLatin-1 Supplement}]+$`	Only characters in the Basic or Latin-1 blocks.
`^[[\p{L}\d\s]&&[\p{InBasic Latin}\p{InLatin-1 Supplement}]]+$`	Only characters in the Basic or Latin-1 blocks of type Letter, Digit, or Whitespace.

A list of Unicode Block Names (http://www.unicode.org/Public/3.0-Update/Blocks-3.txt) is available on the Web.

Analyzing a Data Set

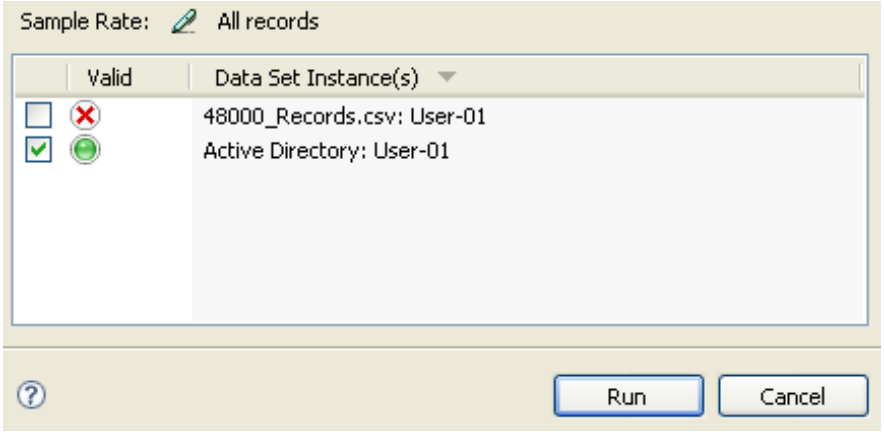
After creating and configuring an analysis profile, you can use the profile to analyze selected data set instances. You can use the analysis profile to analyze several different aspects of an application's data at once to determine how closely the data follows your defined rules for format and content.

IMPORTANT: You should make sure the connection's schema map is correct before attempting an analysis. If none of the tested attributes are mapped, the analysis does not run. If some of the attributes are mapped, the analysis runs but you get only partial results.

To analyze a data set:

1. In the Project View, browse to the analysis profile you want to run.
2. Right-click the Analysis Profile object, then select **Analyze**.

   Alternatively, you can click the **Analyze data** icon in the Analysis Profile editor for the appropriate analysis profile.
3. In the Analyze Data Sets dialog box, select the data set instances against which you want to run the analysis.



   When you select a data set instance, Analyzer reviews the analysis attributes to make sure they are properly mapped to the Application schema, and that they exist and are valid in the selected data set definition.

   If all checks return successfully, Analyzer displays a green icon next to that data set definition.

   If any check returns with one or more issues, Analyzer displays a yellow or red icon next to that data set definition. Click the icon for more information about the issues Analyzer has encountered.
4. (Optional) Click the **Sample Rate** icon to specify a sample rate for the analysis profile.

   This opens the Sample Rate Selector dialog box, from which you can select one of four sampling options:

   All Records: (Default) Analyzer analyzes all records in the data set.

   Number: Analyzer analyzes the specific number of records, starting at the beginning of the data set.

Sample every *n*th record: Analyzer analyzes every nth record in the data set. For example, specifying a value of 3 instructs Analyzer to analyze every third record.

NOTE: The following metrics bypass the sample rate setting, and always query all records: Populated Field, Pattern Frequency, Empty Field, and Unique Values.

---

5 In the Analysis Results view, select an individual attribute to see both graphical and detailed results of the analysis for that attribute.

Cleaning Data

Analyzer can clean bad data as well as test for the bad data. Data cleaning is based on the scripting capabilities of Script Metrics.

The same script is used for either testing or cleaning data. A script used for testing returns the Boolean variable `aimReturnStatus` as either `True` or `False` depending on whether the data passed the test. The same script returns a string variable `aimReturnValue` containing the cleaned data. When applied as a testing script, Analyzer looks at `aimReturnStatus`. When applied as a cleaning script, Analyzer looks at `aimReturnStatus` and, if different, modifies the data value. `aimReturnValue` is returned with the original value (`aimValue`) if no data modification is to take place. The modified value appears in the Data Browser highlighted in yellow with a pencil icon, just as if it had been modified individually.

- "Data Cleaning Variables" on page 65
- "Data Cleaning Script Examples" on page 65
- "Managing Data Cleaning Scripts" on page 66
- "Populating or Deleting Values" on page 66

Data Cleaning Variables

The input variables are the same as for Script Metrics, with the addition of `aimReturnValue`.

```
   aimValue              - the current value being tested
   aimArrayValue         - array of all values in the current cell
   aim_<attrName>        - the value of each attribute in the dataset
   aimArray_<attrName>   - array of values for each attribute

   aimReturnStatus       - boolean variable. Set to true if value passed
test, false if failed.
   aimReturnValue       - the new data value. Return with original value to
leave unmodified.
```

Data Cleaning Script Examples

The following script is both a testing and cleaning script to test for all lowercase letters, and to lowercase them if they are not already lowercase.

```
aimReturnStatus = true;
aimReturnValue = aimValue;

if (aimValue != null)
{
   lower = aimValue.toLowerCase();
   if ( !aimValue.equals(lower) )
   {
      aimReturnStatus = false;
      aimReturnValue = lower;
   }
}
```

When applied as a test script, the following script flags strings longer than 16 characters as failed. When applied as a cleaning script, it truncates the string and replaces the last 3 characters with "...".

```
limit = 16;

aimReturnStatus = true;
aimReturnValue = aimValue;

if (aimValue != null)
{
   len = aimValue.length;
   if ( len > limit)
   {
      aimReturnStatus = false;
      aimReturnValue = aimValue.substring(0,limit-3) + "...";
   }
}
```

Managing Data Cleaning Scripts

A data cleaning script is created, edited, copied, or deleted in the same manner as a testing script. The only difference that turns a testing script into a cleaning script is setting the `aimReturnValue` variable.

To see what values the cleaning script modifies, you can simply apply it with the **Apply Metric to column** option in the right-click menu. To clean the data, you can apply it as a cleaning script with the **Apply Cleaner Script to column** option.

Populating or Deleting Values

For an empty cell, the cleaning script is called with the value of `aimValue` as `null`. You can set a default value or construct one from other attributes set it in `aimReturnValue`.

```
aimReturnStatus = true;
aimReturnValue = aimValue;

if (aimValue == null)
{
   aimReturnValue = "Some Default Value";
}
else
{  //...
}
```

Similarly, a script can delete a value by setting `aimReturnValue` to null. Scripts should always initialize `aimReturnValue` to `aimValue` so it does not appear to Analyzer as a null.

Comparing Data

Analyzer lets you compare data records, both within a data set and between data sets. Data comparison support includes the following:

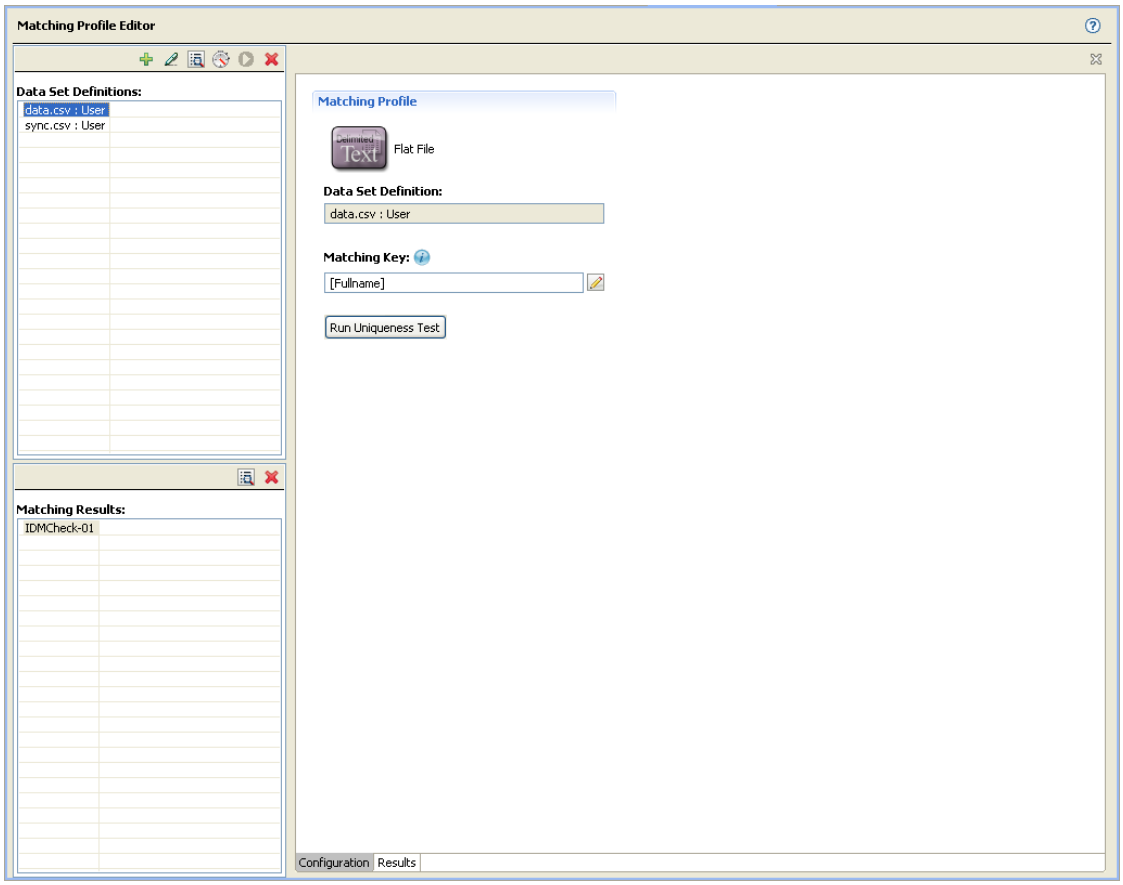
- "Evaluating a Data Set for Unique Values" on page 67
- "Matching Values between Data Sets" on page 69

Evaluating a Data Set for Unique Values

Analyzer lets you test a data set for a value's uniqueness within a particular attribute or set of attributes. This is useful because it allows you to look for a combination of attributes that provides a unique value, or identifier, for each record in the data set.

1. In the Project view, right-click the Matching Profile object, then select **New Matching Profile**.

   Analyzer creates a new Matching Profile object in the selected project, and opens the Matching Profile editor, from which you can finish the duplicate value search.



2 In the Matching Profile editor, specify the necessary information for one of the Matching Profiles.

The Matching Profile editor provides two source definitions for data matching purposes. However, for a uniqueness test, you need to provide information for only one source.

- **Data Set Definition:** Select a data set definition to search for duplicate values.
- **Matching Key:** Select one or more attributes to use as the matching key.
- **Display Attributes:** Select additional attributes that Analyzer displays in the data comparison results. Attributes used in the matching key are display attributes by default. Click to insert the additional display attributes associated with the data sets. The display attributes let you select additional attributes beyond the ones that are present in the Matching Key, to display in the matching test results. click **Subtract** to remove an existing attribute.

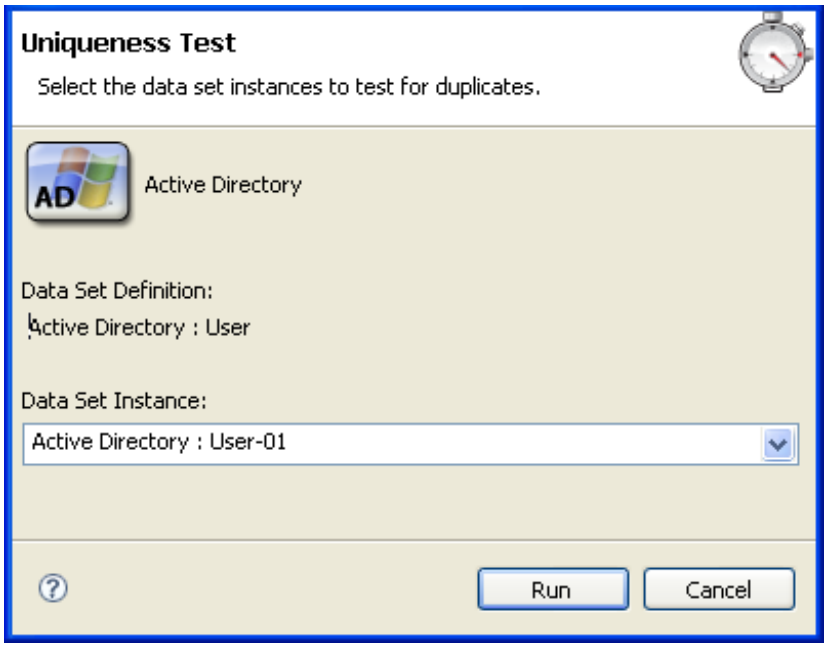
Click **Edit** to open the Matching Key editor.

Select at least one attribute from the drop-down list. Click **Add** to insert an additional attribute in the matching key, and click **Subtract** to remove an existing attribute.

The attributes you select become the matching key for the data comparison. Analyzer appends multiple attributes together to create the matching key. A match occurs between two records when the matching keys are identical.

3 In the **Uniqueness Test** drop-down list, select the data source you want to test.

4 In the Uniqueness Test dialog box, select a specific data set instance against which you want to run the uniqueness search, then click **Run**.



Analyzer performs the uniqueness test as defined and displays the results in the Uniqueness Results view. The Uniqueness Results view displays the details of the test, a graphical representation of the results, and a list of duplicate values.

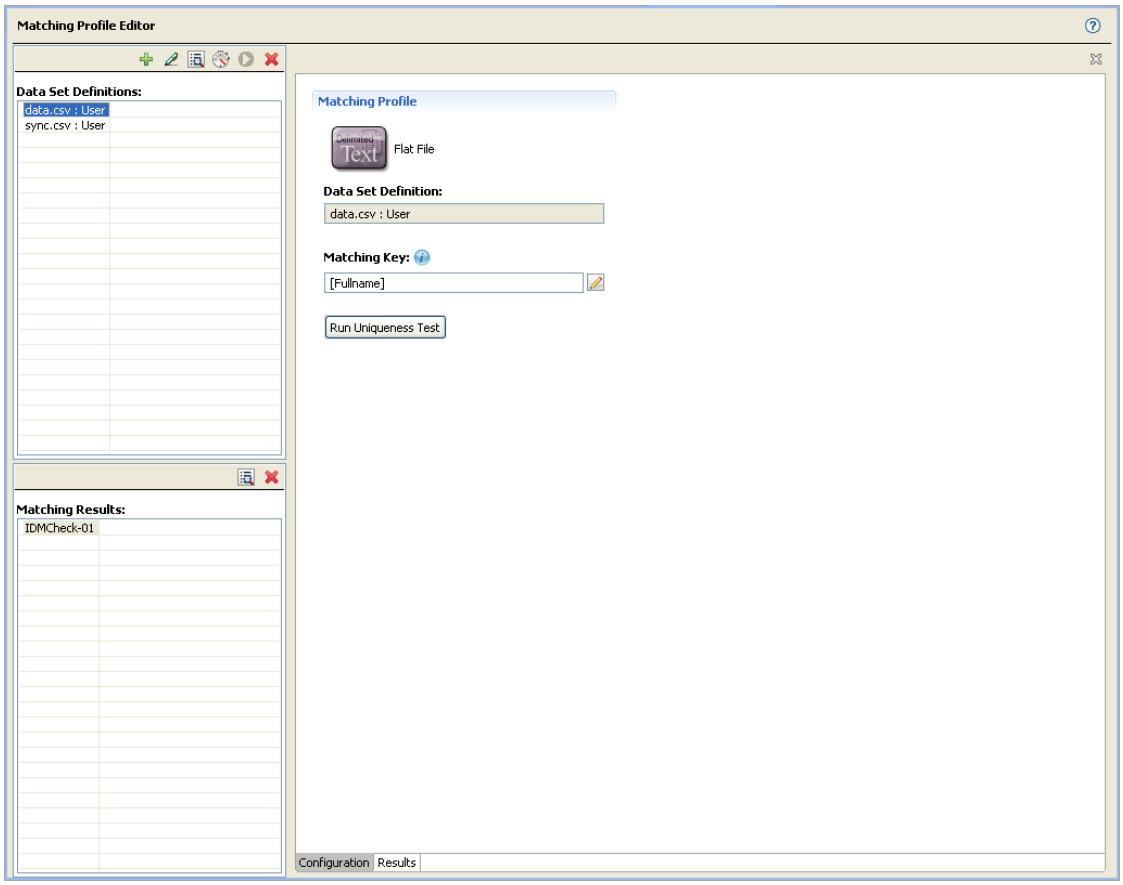
By default, the Uniqueness Results view displays 100 results on a page, but you can change that value as needed. You can also sort the results list on any of the displayed attributes by clicking that attribute's column heading.

Matching Values between Data Sets

Analyzer lets you compare records between two data sets to determine if attribute values are consistent between them. This lets you match attributes from one data set to another in order to plan for potential data mapping and synchronization with Identity Manager.

1 In the Project view, right-click the Matching Profiles object, then select **New Matching Profile**.

   Analyzer creates a new Matching Profile object in the selected project, and opens the Matching Profile editor, from which you can finish the value matching search.



2 In the Matching Profile editor, select the matching characteristics for both Source 1 and Source 2.

- **Matching Key:** Select one or more attributes to use in the matching value search.

  Click **Edit** to open the Matching Key editor.

  Select at least one attribute from the drop-down list. Click **Add** to insert an additional attribute in the matching key, and click **Subtract** to remove an existing attribute.

  The attributes you select become the matching key for the data comparison. Analyzer appends multiple attributes together to create the matching key. A match occurs between two records when the matching keys are identical.

  IMPORTANT: In order to run a Matching analysis, each Matching Key must pass a Uniqueness test at 95% or higher. For more information about Uniqueness tests, see "Evaluating a Data Set for Unique Values" on page 67.

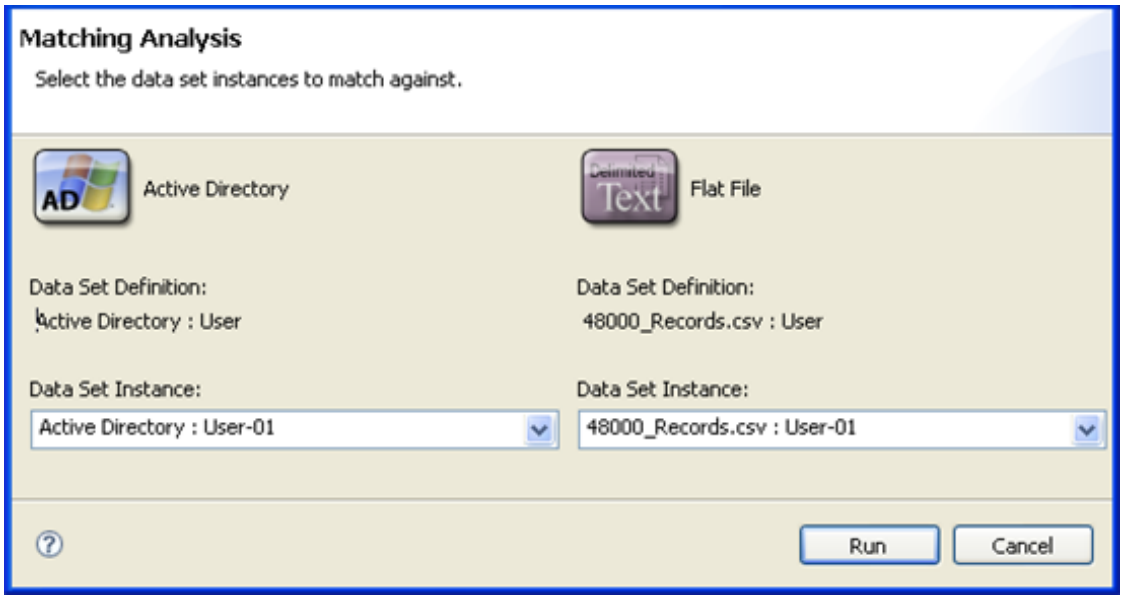
- **Display Attributes:** Select additional attributes that Analyzer displays in the data comparison results.

  Attributes used in the matching key are display attributes by default.

  Click **Add** to insert an additional attribute, and click **Subtract** to remove an existing attribute.

- **Data Set Definition:** Select any two data set definitions to use in the matching values search.

3 Click **Run Matching**.



4 In the Matching Analysis dialog box, select a specific data set instance for each data set against which you want to run the matching value search, then click **Run**.

Analyzer performs the matching record search as defined and displays the results in the Matching Results view. The Matching Results view displays the details of the search, a graphical representation of the results, and a list of all matched and unmatched values.

By default, the Matching Results view displays 100 results on a page, but you can change that value as needed. You can also sort the results list on any of the displayed attributes by clicking that attribute's column heading.

Analyzer handles multi-valued attributes and duplicate records by placing this information in the **Duplicate Match Records** tab. A record that contains attributes with multiple values has the potential to have several entries in a matching analysis. Analyzer only matches the first value, thereby eliminating multiple entries for the same record.

Auditing and Reporting

Analyzer provides both audit and report capabilities to help you capture and communicate the results of Analyzer's data inspection and analysis.

- "Data Browser Audit" on page 72
- "Data Browser Reports" on page 72
- "Analysis Reports" on page 72
- "Data Comparison Reports" on page 73

Analyzer leverages the open source JasperReports framework to generate its reports. For more information about JasperReports, see the JasperReports Project (http://www.jasperforge.org/).

JasperReports uses an XML template (`.jrxml`) to determine the report layout. Analyzer includes a basic template for generating reports suitable for printing on 8½" x 11" paper, portrait layout. However, you can modify the XML template to change the report configuration to include a company name or other customizations.

NOTE: Rather than modify the XML template manually, you can use a graphical design tool for the job. One example of these tools is iReport v2.0.5 (or later.) For more information, visit the iReport Project (https://community.jaspersoft.com/projects/ireport).

Data Browser Audit

Analyzer includes an XDAS library that automatically generates audit events from the Data Browser editor when you send data updates back to the application. For more information about using the Data Browser to update data in the source application, see "Modifying Data" on page 55.

To view these audit events, you must install an XDAS client that can receive the audit events from Analyzer. The Analyzer software package includes XDAS clients for both Linux and Windows.

Data Browser Reports

To create a Data Browser report:

1. In the Data Browser editor, click **Generate a Data Browser Report**.
2. In the Configure Data Browser Report dialog box, specify the report configuration, then click **OK**.

   You can configure default report settings in Analyzer preferences. For more information, see "Reporting" on page 30.
3. In the Report Viewer, Analyzer displays the Data Browser report based on your configuration.
   - Use the icons in the Report Viewer toolbar to navigate through the report and increase or decrease magnification.
   - To save the report, click **Save** and specify a name and format for the resulting report file.
   - To print the report, click **Print** to launch your system's default print dialog box.

   NOTE: The report displays each attribute in the data set instance in an equal-width column, so a display with many attributes might become difficult to read.

Analysis Reports

To create an Analysis report:

1. In the Analysis Results view, click **Generate an Analysis Report**.
2. In the Configure Analysis Report dialog box, specify the report configuration, then click **OK**.

   You can configure default report settings in Analyzer preferences. For more information, see "Reporting" on page 30.
3. In the Report Viewer, Analyzer displays the Analysis report based on your configuration.
   - Use the icons in the Report Viewer toolbar to navigate through the report and increase or decrease magnification.
   - To save the report, click **Save** and specify a name and format for the resulting report file.
   - To print the report, click **Print** to launch your system's default print dialog box.

Data Comparison Reports

Analyzer provides report capabilities for both uniqueness analysis and matching value analysis. For more information on these data comparison analyses, see "Comparing Data" on page 67.

- "Uniqueness Report" on page 73
- "Matching Value Report" on page 73

Uniqueness Report

1. In the Uniqueness Results view, click **Generate a Uniqueness Report** .
2. In the Configure Uniqueness Report dialog box, specify the report configuration, then click **OK**.

   You can configure default report settings in Analyzer preferences. For more information, see "Reporting" on page 30.
3. In the Report Viewer, Analyzer displays the Uniqueness report based on your configuration.
   - Use the icons in the Report Viewer toolbar to navigate through the report and increase or decrease magnification.
   - To save the report, click **Save** and specify a name and format for the resulting report file.
   - To print the report, click **Print** to launch your system's default print dialog box.

   NOTE: The report displays each attribute in the data set instance in an equal-width column, so a report with many attributes might become difficult to read.

Matching Value Report

To create a matching value report:

1. In the Matching Results view, click **Generate a Matching Report** .
2. In the Configure Matching Report dialog box, specify the report configuration, then click **OK**.

   You can configure default report settings in Analyzer preferences. For more information, see "Reporting" on page 30.
3. In the Report Viewer, Analyzer displays the Matching report based on your configuration.
   - Use the icons in the Report Viewer toolbar to navigate through the report and increase or decrease magnification.
   - To save the report, click **Save** and specify a name and format for the resulting report file.
   - To print the report, click **Print** to launch your system's default print dialog box.

   NOTE: The report displays each attribute in the data set instance in an equal-width column, so a report with many attributes might become difficult to read.

4 Data Security with Analyzer

Analyzer for NetIQ Identity Manager inspects and analyzes data from multiple source applications as part of an Identity Manager project. This data might include sensitive company or personal information that needs to be protected. Because of this, it is important to use Analyzer within the context of a broader data security model that helps prevent the loss or exposure, either intentional or otherwise, of sensitive information.

Furthermore, because Analyzer is regularly used as part of a consulting engagement for Identity Manager, consultants might have ready access to sensitive organizational and employee information. Organizations should take steps to account for this within their standard data security measures.

A company that is using Analyzer to analyze and clean its data prior to implementing Identity Manager should create an environment that prevents the exposure of sensitive company or personal information. When using Analyzer, the opportunity for data exposure occurs at two primary points. You can help mitigate the risk of data exposure by implementing a few common-sense practices.

- "Data in Analyzer Projects" on page 75
- "Data Transmission and Logging" on page 76

Data in Analyzer Projects

By their very nature, Analyzer projects might contain sensitive information. To mitigate the risk of data exposure, use the following techniques:

- **Use an external Analyzer Database:** Analyzer's internal HSQLDB stores all project information locally where you run Analyzer. Using an external MySQL database to store Analyzer project information gives you better control over where this information is stored and how accessible it is. For information about configuring the Analyzer database, see "Database Settings" on page 28.
- **Have consultants use local hardware:** Because projects, data set instances, and reports can all contain sensitive data, consultants should use local computers and systems when working with Analyzer. This helps prevent sensitive data from leaving the site. If a consultant brings a computer on-site, all Analyzer projects should be deleted before leaving the site to help protect sensitive information.

  Even when using an external Analyzer database, projects contain sensitive connection information, including hostnames, usernames, and passwords that should be protected.

  NOTE: If consultants install their own licenses (activation keys) for Analyzer while using a customer's local hardware, they should delete the license files from the local machine when their consulting engagement ends. Analyzer stores the license file in the user's home directory.

- **Limit project exports:** To prevent the copying and distribution of sensitive information, limit the export of Analyzer projects and connection profiles.

- **Limit reports:** To prevent the copying and distribution of sensitive information, limit report generation that contains sensitive data. For more information about reports, see "Auditing and Reporting" on page 71.

Data Transmission and Logging

As with any applications that access data in a networked environment, sensitive data is transmitted as applications interact. Although much of this risk should be addressed by standard network security techniques, there are some Analyzer-specific issues you should consider:

- **Use flat files for creating data sets:** Using flat files with Analyzer eliminates connections to live applications and data, and the risk associated with sharing application access information. It also provides more control over the scope of data available to Analyzer. For more information about using flat files, see "Creating a Flat File Profile" on page 45.
- **Audit application interactions:** When using application connections with Analyzer, you should use Analyzer's auditing capabilities to track events between Analyzer and the application. For more information about auditing, see "Auditing and Reporting" on page 71.
- **Local driver connections are not encrypted:** When you use local drivers instead of the Remote Loader, be aware that the connections used by the local driver are not encrypted. This might render usernames and passwords visible on the wire. Because of this, you should use temporary user accounts and passwords, and restrict the rights of these temporary accounts as much as possible to limit the risk of this data being intercepted.
- **IDS Trace displays passwords in clear text:** Currently, the IDS Trace feature in Analyzer does not obfuscate passwords in its trace files. Because of this, you should use temporary user accounts and passwords, and restrict the rights of these temporary accounts as much as possible to limit the risk of this data being intercepted.