AdventureWorks Sample for DQS

This sample is based on the AdventureWorks sample database in SQL Server. To demonstrate Data Quality Services (DQS) capabilities, some data in the database was replaced with incorrect data and some new columns were added.

# Sample Scenarios

The sample covers the following scenarios in DQS:

* [Scenario 1: Create knowledge base from an existing knowledge base (.dqs file)](#_Scenario_1:_Create)
* [Scenario 2: Perform cleansing activity](#_Scenario_2:_Perform)
* [Scenario 3: Perform matching activity](#_Scenario_3:_Perform)

**NOTE**: This sample is not suitable for Knowledge Discovery demonstration.

## Scenario 1: Create knowledge base from an existing knowledge base (.dqs file)

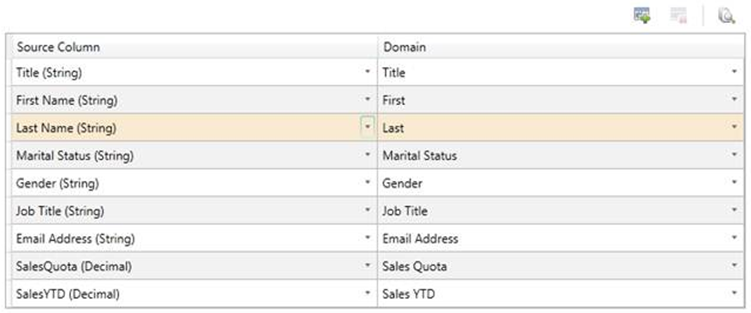
In this scenario, you will use the AdventureWorks.dqs file to create a knowledge base. Later, the new knowledge base created by you will be used to cleanse and match records in the AdventureWorksSampleData.xls file.

1. Create a knowledge base from the AdventureWorks.dqs file. For step-by-step instructions to do so, see [Import a Knowledge Base from a .dqs File](http://msdn.microsoft.com/en-us/library/hh213048(SQL.110).aspx).
2. Select **Domain Management**, and then click **Next**. All the domains from the .dqs file will be imported to the new knowledge base.
3. Observe the following domains:
   1. **Email Address**: The Email validation rule in the **Domain Rules** tab validates an email address using simple and regular expressions to ensure that the email address is in a valid format, and ends with “adventure-works.com”.
   2. **Job Title:**  Domain values and Term-Based Relations.
   3. **Sales Validation:** Cross-domain validation rules in the **CD Rules** tab: checks for cases when total sales are less than sales quota, and flag it as invalid.
   4. **Title Enrichment:** Cross-domain correction rules in the **CD Rules** tab: Checks for the gender and accordingly corrects the salutation.
   5. **First:** A linked domain to the **First Name** domain. The First domain is part of the **Sales Validation** composite domain, whereas the **First Name** domain is part of the **Employee Name** composite domain.
4. Click **Finish** to publish the knowledge base.

**NOTE**: There are two composite domains in the knowledge base: **Title Enrichment** and **Employee Name**. The **Title Enrichment** composite domain will be used in the cleansing scenario and the **Employee Name** composite domain will be used in the matching scenario.

## Scenario 2: Perform cleansing activity

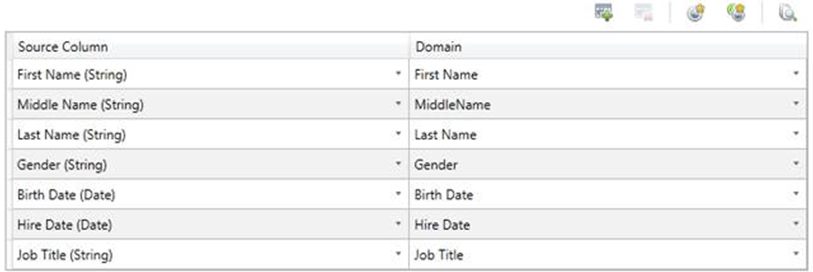
1. Create a new data quality project, select the knowledge base created in scenario 1, select the **Cleansing** activity, then click **Next**.
2. In the Mapping stage, select **Excel File** as the data source, and then select the AdventureWorksSampleData.xls file. Map the columns in the Excel file with the domains as shown below:



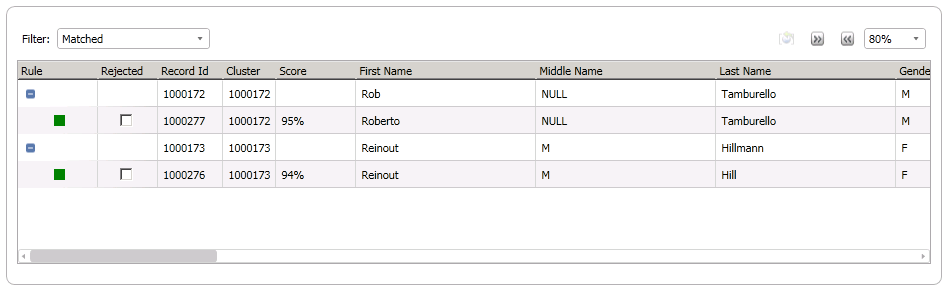
1. Click **Next** to go to the Cleanse stage. Click **Start** to run the computer-assisted cleansing.
2. After, the cleansing is done, click **Next** to go to the interactive cleansing stage.
3. In the interactive cleansing stage, view and update domain values as required in the 5 tabs on the right pane.
4. Click **Next**, and then export the cleansed data to a SQL Server database, csv, or an Excel file.

## Scenario 3: Perform matching activity

1. Create a new data quality project, select the knowledge base created in scenario 1, select the **Matching** activity, then click **Next**.
2. In the Mapping stage, select **Excel File** as the data source, and then select the AdventureWorksSampleData.xls file. Map the columns in the Excel file with the domains as shown below:



1. Click **Next** to go to the Match stage. Click **Start** to run the matching. After the matching is done, two clusters of Matched records are shown.



1. Eliminate the duplicate records, and then proceed to export the processed data.