Oracle Enterprise Data Quality

The Director User Interface and Its Key Objects

Release 12.2.1
Table of Contents

Labs for the Director User Interface and its Key Objects ................................................................. 4
  Lab 1: Import Data ............................................................................................................................ 5
  Lab 2: Create and Run a Process ..................................................................................................... 10
  Lab 3: Examine the Data ................................................................................................................ 13
Labs for the Director User Interface and its Key Objects
Lab 1: Import Data
The instructions in this lab assume that you are using the EDQ-12.2.1-Trn virtual machine. However, if you have instead installed Oracle Enterprise Data Quality on a Windows machine, then alternative instructions, marked with asterisks, are provided where necessary.

Create a Project
1. In the Project Browser, right-click Projects and select New Project...

2. Give the new project a name of Training Project and optionally add a description. Click Next.
3. Ensure that All Groups is selected in Project Permissions. This will enable any user to view and use the project. Click Finish to create the new project.
Add a Data Store

1. Expand the newly created Training Project by clicking the + symbol beside it, right-click Data Stores and select New Data Store...

2. In the Data is accessed from drop down leave Server selected and, in the Category drop down, select Database* (also note what other options are available in these two drop downs).

   *If you are not using the EDQ-12.2.1-Trn virtual machine, and have instead installed EDQ on a Windows Machine, select Other instead of Database in the Category drop down.

What other options are available in the Data is accessed from and Category drop downs?

Data is accessed from

Category

Ensure that you still have Server and Database selected in the two drop-downs before you move to the next step.

3. In the Type list select Oracle* and click Next>:

   *If you are not using the EDQ-12.2.1-Trn virtual machine, and have instead installed EDQ on a Windows Machine, select Server-based DB files instead of Oracle in the Type list.
4. In the Data Store Configuration screen, enter the following values*:
   a. **Database host**: localhost
   b. **Port**: 1521
   c. **Database name (SID)**: orcl
   d. **User name**: servicemanagement
   e. **Password**: servman

   *If you are not using the EDQ-12.2.1-Trn virtual machine, and have instead installed EDQ on a Windows Machine, then, instead of the values above, enter servicemanagement.jmp in the File in server work area field.

5. Click **Test…** to ensure that you can successfully connect to your new Data Store. A dialog box should report success. If it does, just click **OK** to dismiss it. If it reports failure, check the values that you have entered and try again*. Once you have successfully tested the connection, click **Next >**.

   *If you are not using the EDQ-12.2.1-Trn virtual machine, and have instead installed EDQ on a Windows Machine, then also ensure that you have copied the servicemanagement.jmp file from edq_training_assets_12.2.1\Data Files\fundamentals to your EDQ instance’s landing area, as described in Lab 6 of the Overview module (note that on Windows 7 machines, the default location of the EDQ landing area is C:\ProgramData\Oracle\Enterprise Data Quality\oedq_local_home\landingarea).
6. Give the data store a name of **Connection to Service Management Database** and, optionally, add a description. Click **Finish** to save the Data Store.

**Add a Snapshot**

1. Right-click **Staged Data** and select **New Snapshot...**
2. Select your newly created **Connection to Service Management Database** data store. This is where the data for the snapshot will come from. Click **Next >**.
3. In the Table Selection screen, ensure that the CUSTOMERS table is selected, and click **Next >**.
4. In the Column Selection screen, leave all the columns selected and just click **Next >**.
5. In Filter Options screen, leave the default option selected and just click **Next >** to continue.
6. In Sampling Options screen, leave the default **All** option selected and just click **Next >** to continue.
7. In the No Data Handling screen leave the ‘No Data’ Reference Data field blank, and then click **Next >**.
8. In the Snapshot Name screen, change the name of your new snapshot to **Customers Table** and click **Finish** to create the snapshot.

   Notice that after a short delay whilst the CUSTOMERS Table is read, the **Results Browser** is populated with the data from the Service Management database. Taking the Snapshot causes Enterprise Data Quality to stage the data from the database’s Customers table – meaning that a copy of this table is placed in Enterprise Data Quality repository. From now on you will be working with this copy. This approach has the advantage of avoiding unnecessary calls on the database itself, which could, for example, be in use for production purposes.
Lab 2: Create and Run a Process

Create a process

1. Right-click Processes and select New Process...
2. Processes start with a Reader and this needs a source for its data. Select the Customer Table snapshot that you created in the previous exercise and click Next >.

3. In Analysis just click Next to continue (we will come back to what this means later).
4. Give your process a Name of Customer Data Quick Stats and click Finish to create the process.
   Notice that the Process Canvas becomes active, the process is created and a Reader processor appears. The green circular arrow icon on any processor means it has yet to be executed.

Add a processor

1. In the Tool Palette, enter the word 'quick' into the Search box. This will allow you to quickly find the Quickstats Profiler:
The **Quickstats Profiler** belongs to the **Profiling** family of processors. These are stored together under the **Profiling** icon:

1. Drag the **Quickstats Profiler** onto the Process Canvas, dropping it to the right of the **Reader**.
2. Hover over the output triangle of the **Reader** processor. The pointer icon turns to a hand. Click and drag from the output triangle of the **Reader** processor so that the connector line reaches the input triangle of the **Quickstats Profiler**.

The **Quickstats Profiler** dialog will appear:
Note the message in red informing you that this processor needs at least one attribute to be set as an input.

4. Click the Select All icon, shown in the screenshot above. This will select all of the input attributes. Click OK to save the processor. Note that the new processor also shows the 'not yet run' green circular arrow.

Once a process is run its results will be stored in the Enterprise Data Quality repository and the green 'not yet run' icons will disappear. As subsequent processors are added then only these will need to be run saving time, provided upstream processors are left unchanged.

Run the process

1. The process now has a Reader and a Quickstats Profiler, both with their green 'not yet run' icons displayed. Click the Run icon ( ) in the toolbar to run the process. Whilst the process is running, the Process Canvas will turn light blue. When the process has finished the 'not yet run' icons will disappear to show that the processors have results associated with them.
Lab 3: Examine the Data

View the data in the Results Browser

1. Click the Reader processor to see the input data stored as a snapshot. This will be displayed in the Results Browser.
2. Next click the Quickstats Profiler to see the output of the processor. The Quickstats Profiler produces a summary of the data. For each attribute it displays the number of records without data, the number of distinct values, the number of singletons and the number of duplicate values (a singleton is a value that only appears once in a data set). Blue text is hyperlinked. You can click blue text to drill-down to see the data underneath.
3. Click the hyperlink for the 3 duplicate rows for the CU_NO attribute.
4. A second summary appears showing details for those 3 duplicate rows. Click again on the hyperlink marked ‘3’ to see the three underlying rows of data:

The three rows of data are displayed.

Note that these are test rows and ought to be removed from the database. The Quickstats Profiler has quickly enabled us to find these.
Create and export a Results Book

1. Click the **Add page in Results Book** icon ( ) in the Toolbar of the Results Browser.

2. Click **New Results Book**...

3. Name the new results book **Duplicated Customer Numbers** and click **Finish**.

4. Click **Next >** to add a page to the results book.

5. Click **Finish**, using the default mapping as it is displayed.
   
   Note that the new Results Book appears in the Project Browser. Any number of pages can be added to this book.

6. Finally, export the Results Book to MS Excel by clicking **the Export to Excel** icon in the Results Browser toolbar ( ).

7. Find the exported file and open it in MS Excel or a similar Spreadsheet application.