

---

---

*Data Quality Management*  
*using*  
*DecisionSpace® Data Quality*  
*Release 5000.10.3.0*  
*Volume II*

---

---

© 2015 Halliburton

**HALLIBURTON** | Landmark

## Copyright and Trademarks

© 2015 Halliburton  
All Rights Reserved

This publication has been provided pursuant to an agreement containing restrictions on its use. The publication is also protected by Federal copyright law. No part of this publication may be copied or distributed, transmitted, transcribed, stored in a retrieval system, or translated into any human or computer language, in any form or by any means, electronic, magnetic, manual, or otherwise, or disclosed to third parties without the express written permission of:

**Halliburton | Landmark**  
10200 Bellaire Blvd., Houston, Texas 77072-5206, USA  
P.O. Box 42810, Houston, Texas 77242-2810, USA  
Phone: 281.575.3000, Fax: 713.839.2015  
Internet: <https://www.landmarksoftware.com>

### Trademarks

3D Drill View, 3D Drill View KM, 3D Surveillance, 3DFS, 3DView, Active Field Surveillance, Adaptive Mesh Refining, ADC, Advanced Data Transfer, Analysis Model Layering, ARIES, ARIES DecisionSuite, Asset Data Mining, Asset Decision Solutions, Asset Development Center, Asset Development Centre, Asset Journal, Asset Performance, AssetConnect, AssetConnect Enterprise, AssetConnect Enterprise Express, AssetConnect Expert, AssetDirector, AssetJournal, AssetLink, AssetLink Advisor, AssetLink Director, AssetLink Observer, AssetObserver, AssetObserver Advisor, AssetOptimizer, AssetPlanner, AssetPredictor, AssetSolver, AssetSolver Online, AssetView, AssetView 2D, AssetView 3D, Barrier Assurance Monitoring, BLITZPAK, Buckle, CartoSnap, CasingLife, CasingSeat, CDS Connect, CGMage Builder, Channel Trim, COMPASS, Contract Generation, Corporate Data Archiver, Corporate Data Store, Data Analyzer, DataManager, DataServer, DataStar, DataVera, DBPlot, Decision Management System, DecisionSpace, DecisionSpace 3D Drill View, DecisionSpace 3D Drill View KM, DecisionSpace AssetLink, DecisionSpace AssetPlanner, DecisionSpace AssetSolver, DecisionSpace Atomic Meshing, DecisionSpace Base Module, DecisionSpace Data Quality, DecisionSpace Desktop, DecisionSpace Dropsite, DecisionSpace Geoscience, DecisionSpace GIS Module, DecisionSpace GRC Module, DecisionSpace Nexus, DecisionSpace Reservoir, DecisionSuite, Deeper Knowledge, Broader Understanding, Depth Team, Depth Team Explorer, Depth Team Express, Depth Team Extreme, Depth Team Interpreter, DepthTeam, DepthTeam Explorer, DepthTeam Express, DepthTeam Extreme, DepthTeam Interpreter, Desktop Navigator, DESKTOP-PVT, DESKTOP-VIP, DEX, DIMS, Discovery, Discovery 3D, Discovery Asset, Discovery Framebuilder, Discovery PowerStation, Discovery Suite, DMS, Drillability Suite, Drilling Desktop, DrillModel, DrillNET, Drill-to-the-Earth-Model, Drillworks, Drillworks ConnectML, Drillworks Predict, DSS, Dynamic Frameworks to Fill, Dynamic Reservoir Management, Dynamic Surveillance System, EDM, EDM AutoSync, EDT, eLandmark, Engineer's Data Model, Engineer's Desktop, Engineer's Link, ENGINEERING NOTES, eNotes, ESP, Event Similarity Prediction, ezFault, ezModel, ezSurface, ezTracker, ezTracker2D, ezValidator, FastTrack, Field Scenario Planner, FieldPlan, FieldPlan Express, For Production, FrameBuilder, Frameworks to Fill, FZAP!, GeoAtlas, GeoDataLoad, GeoGraphix, GeoGraphix Exploration System, Geologic Interpretation Component, Geometric Kernel, GeoProbe, GeoProbe GF DataBase, GeoSmith, GES, GES97, GesFull, GESXplorer, GMAplus, GMI Imager, Grid3D, GRIDGENR, H. Clean, Handheld Field Operator, HHFO, High Science Simplified, Horizon Generation, I<sup>2</sup> Enterprise, iDIMS, iEnergy, Infrastructure, iNotes, Iso Core, IsoMap, iWellFile, KnowledgeSource, Landmark (*as service*), Landmark (*as software*), Landmark Decision Center, LandNetX, Landscape, Large Model, Lattix, LeaseMap, Limits, LithoTect, LogEdit, LogM, LogPrep, MagicDesk, Make Great Decisions, MathPack, MDS Connect, MicroTopology, MIMIC, MIMIC+, Model Builder, NETool, Nexus (*as service*), Nexus (*as software*), Nexus View, Object MP, OneCall, OpenBooks, OpenJournal, OpenLink, OpenSGM, OpenVision, OpenWells, OpenWire, OpenWire Client, OpenWire Server, OpenWorks, OpenWorks Development Kit, OpenWorks Production, OpenWorks Well File, Operations Management Suite, PAL, Parallel-VIP, Parametric Modeling, Permedia, Petris WINDS Enterprise, PetrisWINDS, PetroBank, PetroBank Explorer, PetroBank Master Data Store, PetroWorks, PetroWorks Asset, PetroWorks Pro, PetroWorks ULTRA, PLOT EXPRESS, PlotView, Point Gridding Plus, Pointing Dispatcher, PostStack, PostStack ESP, PostStack Family, Power Interpretation, PowerCalculator, PowerExplorer, PowerExplorer Connect, PowerGrid, PowerHub, PowerModel, PowerView, PrecisionTarget, Presgraf, PressWorks, PRIZM, Production, Production Asset Manager, PROFILE, Project Administrator, ProMAGIC Connect, ProMAGIC Server, ProMAX, ProMAX 2D, ProMax 3D, ProMAX 3DPDSM, ProMAX 4D, ProMAX Family, ProMAX MVA, ProMAX VSP, pSTAx, Query Builder, Quick, Quick+, QUICKDIF, Quickwell, Quickwell+, Quiklog, QUIKRAY, QUIKSHOT, QUIKVSP, RAVE, RAYMAP, RAYMAP+, Real Freedom, Real Time Asset Management Center, Real Time Decision Center, Real Time Operations Center, Real Time Production Surveillance, Real Time Surveillance, Real-time View, Recall, Reference Data Manager, Reservoir, Reservoir Framework Builder, RESev, ResMap, Resolve, RTOC, SCAN, SeisCube, SeisMap, SeisMapX, Seismic Data Check, SeisModel, SeisSpace, SeisVision, SeisWell, SeisWorks, SeisWorks 2D, SeisWorks 3D, SeisWorks PowerCalculator, SeisWorks PowerJournal, SeisWorks PowerSection, SeisWorks PowerView, SeisXchange, Semblance Computation and Analysis, Sierra Family, SigmaView, SimConnect, SimConvert, SimDataStudio, SimResults, SimResults+, SimResults+3D, SIVA+, SLAM, Smart Change, Smart Deploy, Smart Flow, Smart Skills, Smart Start, Smart Sustain, Smart Transform, Smart Vision, SmartFlow, smartSECTION, smartSTRAT, Spatializer, SpecDecomp, StrataMap, StrataModel, StratAmp, StrataSim, StratWorks, StratWorks 3D, StreamCalc, StressCheck, STRUCT, Structure Cube, Surf & Connect, SurfNet, SynTool, System Start for Servers, SystemStart, SystemStart for Clients, SystemStart for Servers, SystemStart for Storage, Tanks & Tubes, TDQ, Team Workspace, TERAS, T-Grid, The Engineer's DeskTop, Total Drilling Performance, TOW/cs, TOW/cs Revenue Interface, TracPlanner, TracPlanner Xpress, Trend Form Gridding, Trimmed Grid, Tubular Basic, Turbo Synthetics, Unconventional Essentials, VESPA, VESPA+, VIP, VIP-COMP, VIP-CORE, VIPDataStudio, VIP-DUAL, VIP-ENCORE, VIP-EXECUTIVE, VIP-Local Grid Refinement, VIP-THERM, vSpace, vSpace Blueprint, vSpace Onsite, WavX, Web Editor, Well H. Clean, Well Seismic Fusion, Wellbase, Wellbore Planner, Wellbore Planner Connect, WELLCAT, WELLPLAN, WellSolver, WellXchange, WOW, Xsection, You're in Control. Experience the difference, ZAP!, ZEH, ZEH Plot, ZetaAnalytics, Z-MAP, Z-MAP Plus, and ZPS are trademarks, registered trademarks, or service marks of Halliburton.

All other trademarks, service marks and product or service names are the trademarks or names of their respective owners.

### Note

The information contained in this document is subject to change without notice and should not be construed as a commitment by Halliburton. Halliburton assumes no responsibility for any error that may appear in this manual. Some states or jurisdictions do not allow disclaimer of expressed or implied warranties in certain transactions; therefore, this statement may not apply to you.

### **Third Party Licenses and Attributions**

Halliburton acknowledges that certain third party code has been bundled with, or embedded in, its software. The licensors of this third party code, and the terms and conditions of their respective licenses, may be found at the following location:

*PathNameInInstallationDir/Third\_Party.pdf*

### **Disclaimer**

The programs and documentation may provide links to external web sites and access to content, products, and services from third parties. Halliburton is not responsible for the availability of, or any content provided on, third party web sites. You bear all risks associated with the use of such content. If you choose to purchase any products or services from a third party, the relationship is directly between you and the third party. Halliburton is not responsible for: (a) the quality of third party products or services; or (b) fulfilling any of the terms of the agreement with the third party, including delivery of products or services and warranty obligations related to purchased products or services. Halliburton is not responsible for any loss or damage of any sort that you may incur from dealing with any third party.



# *Data Quality Management using DecisionSpace® Data Quality Release 5000.10.3.0*

## **Volume II**

<b>Chapter 7: Data Consolidation in DecisionSpace Data Quality .....</b>	<b>7-1</b>
Chapter Overview .....	7-2
The Merge Process.....	7-3
Exercise: Adding a Merge Phase .....	7-3
Merge Setup.....	7-6
Exercise: Adding a Merge Group .....	7-6
Configure Merge .....	7-9
Exercise: Using the Configure Merge Tool .....	7-9
Merge Flags for Target Tables .....	7-15
Merge Flags for Target Columns .....	7-15
Using Merge .....	7-18
Exercise: Running the Merge Tool.....	7-18
Exercise: Modifying Data while Merging .....	7-22

<b>Chapter 8: Requirements Administrator in DecisionSpace Data Quality .....</b>	<b>8-1</b>
Chapter Overview .....	8-2
Repository .....	8-3
Exercise: Populating the Repository.....	8-3
Service Levels.....	8-7
Exercise: Creating a Service Level.....	8-7
Exercise: Adding a Service Level .....	8-12
Exercise: Editing a Service Level.....	8-15
Exercise: Deleting a Service Level.....	8-18
Exercise: Exporting a Service Level .....	8-21
Exercise: Importing a Service Level .....	8-24
Managing the Repository.....	8-27
Components of Repository.....	8-27
Exercise: Adding a Component .....	8-28
Exercise: Editing a Component .....	8-34
Exercise: Deleting a Component .....	8-37
Exercise: Associating an Element with a Standards Reference Table .....	8-42
Requirements .....	8-46
Exercise: Adding a Requirement .....	8-46
Exercise: Editing a Requirement .....	8-54
Exercise: Deleting a Requirement .....	8-57
Exercise: Validating a Requirement.....	8-61
Rules Editor.....	8-63
Exercise: Managing Advanced Requirements .....	8-63
<b>Chapter 9: Configuring the Data Quality Web Dashboard .....</b>	<b>9-1</b>
Chapter Overview .....	9-2
Submodel Analysis .....	9-3
Exercise: Publishing Submodels .....	9-3

Grouping Master Data Information.....	9-6
Exercise: Configuring Columns Display Groups .....	9-6
Exercise: Specifying a Column Dashboard Name .....	9-8
Updating Statistics on the Web Dashboard .....	9-12
 <b>Chapter 10: Project Administration.....</b>	 10-1
<b>Chapter Overview .....</b>	<b>10-2</b>
Roles & Permissions.....	10-2
<b>Managing Users .....</b>	<b>10-5</b>
Exercise: Creating a User .....	10-5
Exercise: Editing a User .....	10-6
Exercise: Removing a User.....	10-7
<b>Advanced Scheduling .....</b>	<b>10-8</b>
Exercise: Adding a Scheduled Event .....	10-8
Exercise: Adding Jobs to Events .....	10-10
Exercise: Scheduling Events.....	10-12
Exercise: Running a Job Group on the Server .....	10-13
Exercise: Run Job Group Locally.....	10-14
<b>Viewing Job Group Charts.....</b>	<b>10-16</b>
Exercise: Viewing Job Group Performance .....	10-16
Exercise: Viewing Job Group Sequencing .....	10-17
<b>Managing Jobs .....</b>	<b>10-18</b>
Exercise: Opening Job Administrator.....	10-19
Exercise: Stopping Current Running Jobs .....	10-20
Exercise: Viewing Server Logs .....	10-21
<b>Managing Connections.....</b>	<b>10-22</b>
Exercise: Editing a Data Owner Connection .....	10-23
Exercise: Editing a Workspace Connection .....	10-26
Exercise: Removing a Data Owner/Workspace Connection .....	10-28

<b>Workshop A: Connecting DSDQ to an OpenWorks Data Source .....</b>	A-1
Overview .....	A-2
Creating a New Project.....	A-4
Evaluating Data using the HealthCheck Phase .....	A-15
Rapid HealthCheck Activity.....	A-15
Exercise: Running Table Analysis on All Tables Task .....	A-15
Exercise: Running Column Analysis on All Columns Task .....	A-18
Exercise: Running Table Analysis on Modeled Tables Task.....	A-21
Exercise: Running Column Analysis on Modeled Tables Task.....	A-24
Detailed HealthCheck Activity.....	A-28
Exercise: Configuring the Detailed HealthCheck Tool .....	A-28
Exercise: Running the Detailed HealthCheck Task .....	A-38
Resolving Data Quality Issues using the Clean Phase .....	A-43
Adding a Clean Phase.....	A-43
Rapid Clean Activity .....	A-45
Exercise: Configuring the Rapid Clean Tool .....	A-45
Exercise: Running the Test Rapid Clean Task.....	A-48
Exercise: Running the Rapid Clean Task.....	A-51
Detailed Clean Activity .....	A-55
Exercise: Configuring the Detailed Clean Tool .....	A-55
Exercise: Running the Detailed Clean Task .....	A-61
Validating Corrected Data in the OpenWorks Data Source .....	A-66
Viewing Results on the DSDQ Web Dashboard .....	A-69

<b>Workshop B: Connecting DSDQ to an EDM Data Source</b>	... B-1
Overview	..... B-2
Creating a New Project	..... B-4
Evaluating Data using the HealthCheck Phase	..... B-15
Rapid HealthCheck Activity	..... B-15
Exercise: Running Table Analysis on All Tables Task	..... B-15
Exercise: Running Column Analysis on All Columns Task	..... B-18
Exercise: Running Table Analysis on Modeled Tables Task	..... B-21
Exercise: Running Column Analysis on Modeled Tables Task	..... B-24
Detailed HealthCheck Activity	..... B-28
Exercise: Configuring the Detailed HealthCheck Tool	..... B-28
Exercise: Running the Detailed HealthCheck Task	..... B-38
Resolving Data Quality Issues using the Clean Phase	..... B-43
Adding a Clean Phase	..... B-43
Rapid Clean Activity	..... B-45
Exercise: Configuring the Rapid Clean Tool	..... B-45
Exercise: Running the Test Rapid Clean Task	..... B-48
Exercise: Running the Rapid Clean Task	..... B-52
Detailed Clean Activity	..... B-56
Exercise: Configuring the Detailed Clean Tool	..... B-56
Exercise: Running the Detailed Clean Task	..... B-62
Validating Corrected Data in the EDM Data Source	..... B-67
Viewing Results on the DSDQ Web Dashboard	..... B-71



## *Chapter* 7

# *Data Consolidation in DecisionSpace*

## *Data Quality*

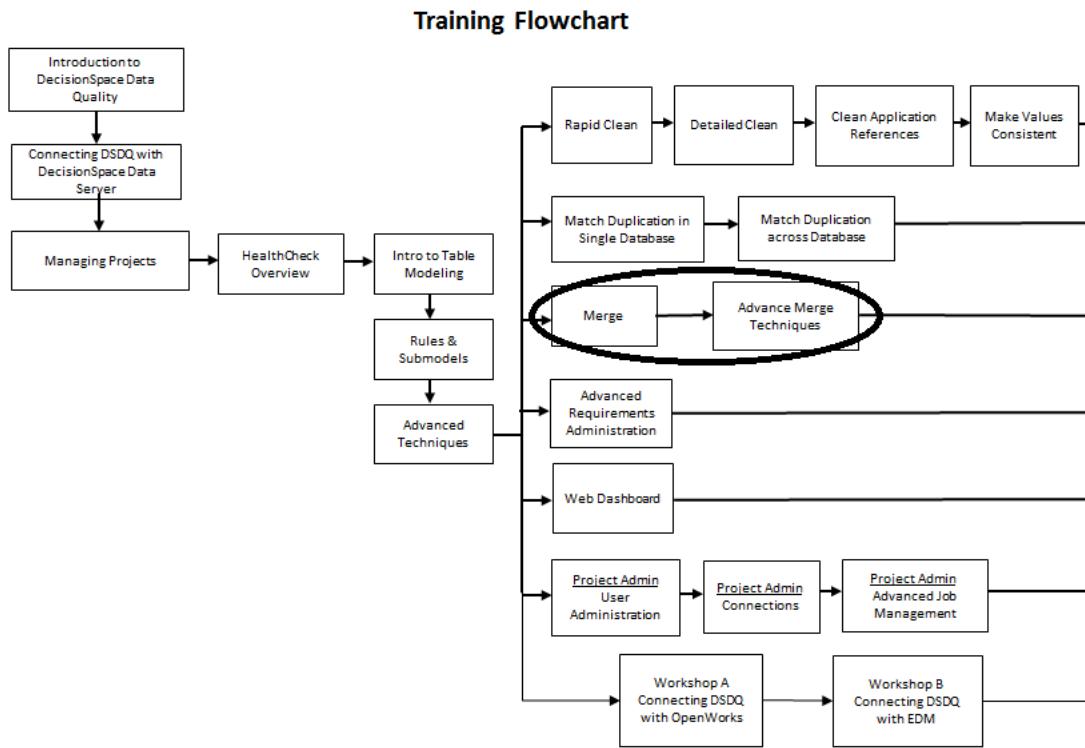
The Merge Phase enables you to consolidate data that is fit to be included in your dataset. Merge can apply individual policies to choose correct business values or transform data, and output the data to different databases or data formats.

# Chapter Overview

In this chapter, you will learn about:

- The Merge Process
- Setting up Merge Groups
- Using Configure Merge
- Using Run Merge

Topics covered in each chapter are outlined in the following illustration. Those specific to the current chapter will be circled in black for your reference:



---

## The Merge Process

---

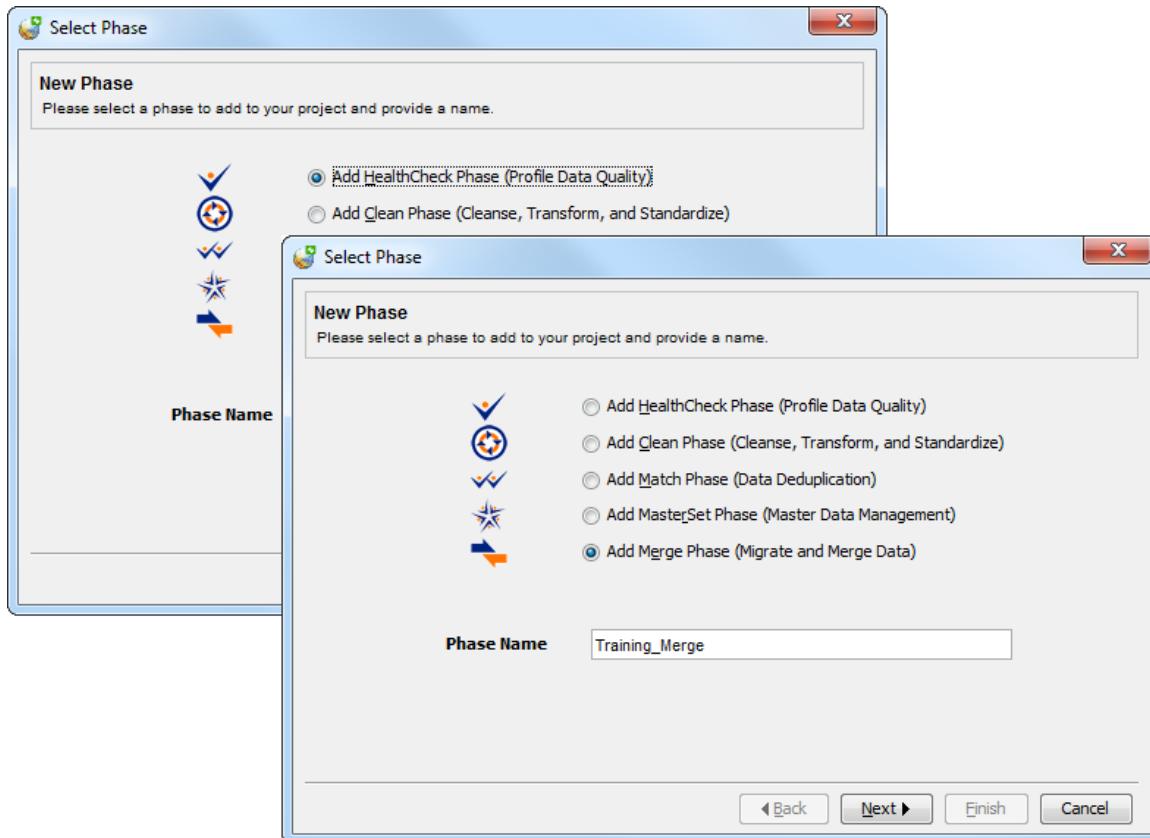
Once your data has passed through the various Phases of DSDQ and cleaned as per applicable policies, it needs to be consolidated in one database. The **Merge** module enables you to migrate/merge data in one data set. This also ensures that bad data is not loaded during the merge Phase.

### ***Exercise: Adding a Merge Phase***

The Merge phase is added to the project for the processing of your data. This phase consists of the Merge setup & Merge activities which allow you to setup Merge groups and run the **Merge Task**. In this exercise, you will be adding a new Merge Phase **Training\_Merge** using **DSDQ\_Training** and **DSDQ\_Training\_Source** as the source and target connection respectively, to an already created **OpenWorks** database.

To Add a Merge Phase:

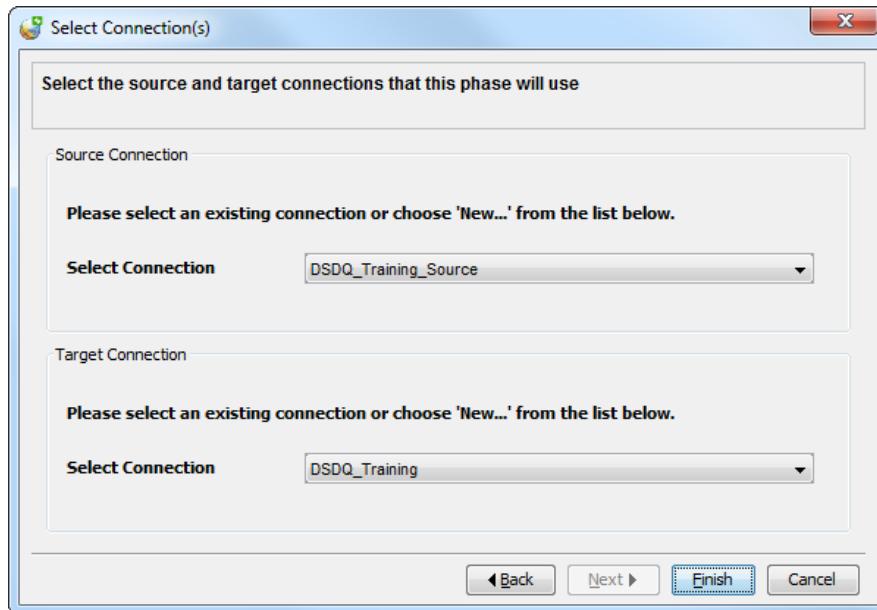
1. Click the **Add New Phase**  button on the project toolbar.  
The **Select Phase** window appears with the **Add HealthCheck Phase (Profile Data Quality)** option selected by default.



2. Select the **Add Merge Phase (Migrate and Merge Data)** option.
3. Enter **Training\_Merge** in the **Phase Name** field.

4. Click **Next** to continue.

The **Select Connection(s)** window appears with selection options for Source and Target Connections.



5. Select **DSDQ\_Training\_Source** from the **Select Connection** drop-down list of the Source Connection section.
6. Select **DSDQ\_Training** from the **Select Connection** drop-down list of the Target Connection section.
7. Click **Finish**.

The **Merge** Phase is created and displays in the DecisionSpace Data Quality Project Window.

**Note**

Select the source database from the **Select Connection** drop-down list. If the desired connection is not listed, add a new one [refer to **Chapter 2: Creating Connections** for instructions].

## Merge Setup

---

The Merge Setup Activity helps you in managing Merge groups. This activity consists of 2 tools:

- Setup Data Configurations
- Configure Merge

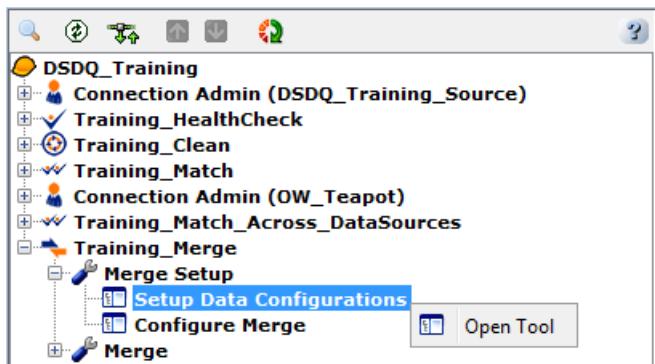
You can create, edit and delete Merge groups with the **Setup Data Configurations** Tool. The **Configure Merge** Tool allows you to map tables and columns.

### **Exercise: Adding a Merge Group**

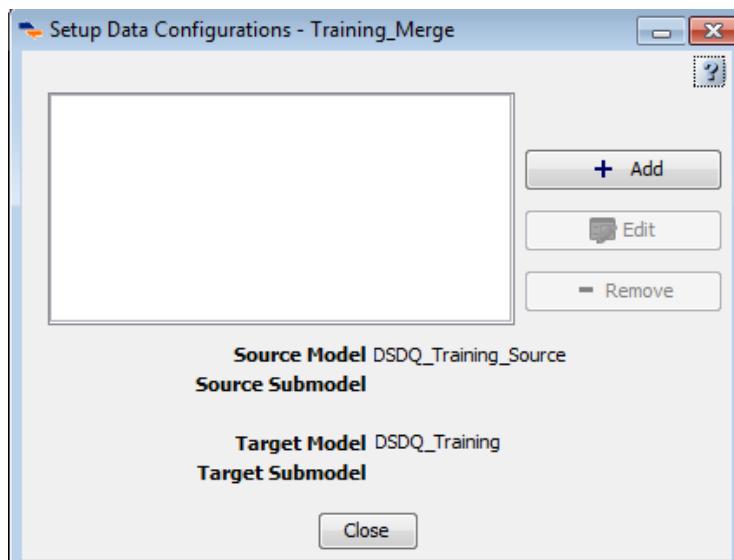
Merge groups allow you to effectively manage different merge configurations. Any settings that you want to apply to the source and target data are contained in Merge groups. These groups can be added from the **Setup Data Configurations** Tool. You will need to enter a name for the group, select a Source submodel and a Target Submodel.

To add a Merge Group:

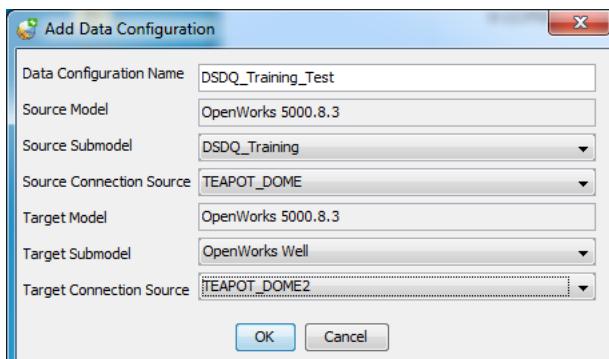
1. Click  on the DecisionSpace Data Quality Tree to expand the **Training\_Merge** Phase.
2. Click  to expand the **Merge Setup** Activity.
3. Double-click the **Setup Data Configurations** Tool or right-click the **Setup Data Configurations** Tool and select **Open Tool** from the pop-up menu



The **Setup Data Configurations** window appears.



4. Click the **+ Add** button to add a Merge Group. The **Add Data Configuration** dialog box appears.



5. Enter **DSDQ\_Training\_Test** in the **Data Configuration Name** field.

The **Source Model** field is pre-populated with the source connection selected during the process of adding a Merge Phase.

6. Select **DSDQ\_Training** from the **Source Submodel** drop-down list.

7. Select **TEAPOT\_DOME** from the **Source Connection Source** drop-down list.

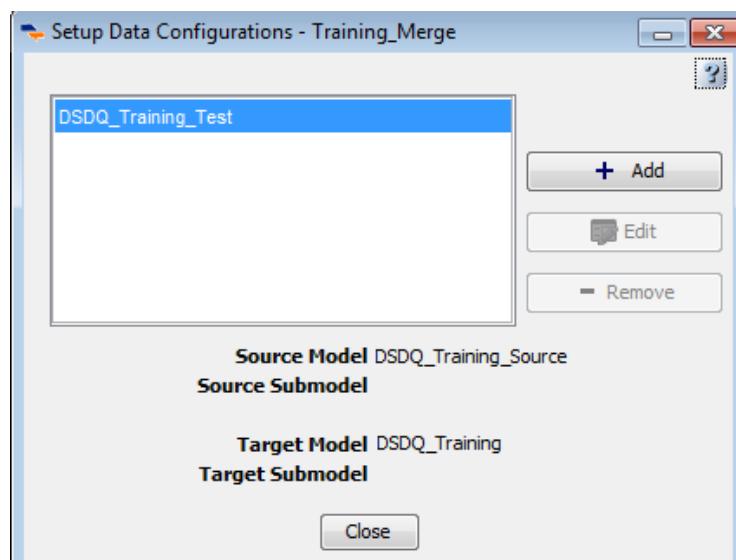
The **Target Model** field is pre-populated with the target connection selected during the process of adding a Merge Phase.

8. Select **OpenWorks Well** from the **Target Submodel** drop-down list.

9. Select **TEAPOT\_DOME2** from the **Target Connection Source** drop-down list.

10. Click **OK**.

The **DSDQ\_Training\_Test** is added and displays in the **Setup Data Configurations** window.



11. Click **Close** to exit the **Setup Data Configurations** window.

#### Note

You can also edit or remove a merge group. Click the **Edit** Button to rename the merge group or click the **Remove** button to delete the merge group.

---

## Configure Merge

---

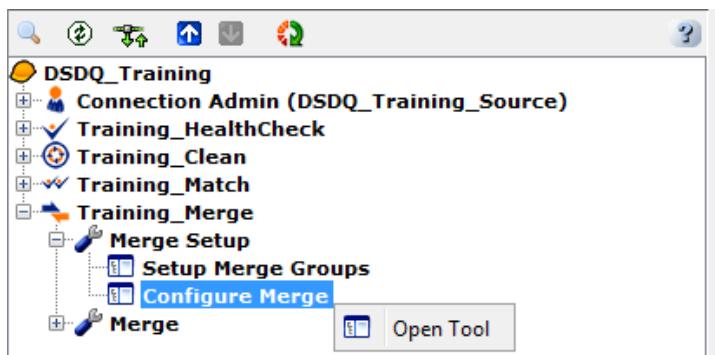
The **Configure Merge** Tool maps tables and columns. While using this tool, you will select a subset of data for testing. If the merge results are correct, you will run the **Merge** Phase to apply the changes to the dataset. The **Configure Merge** Tool is also used to apply rules that can alter or modify the data before it is migrated to the database.

### ***Exercise: Using the Configure Merge Tool***

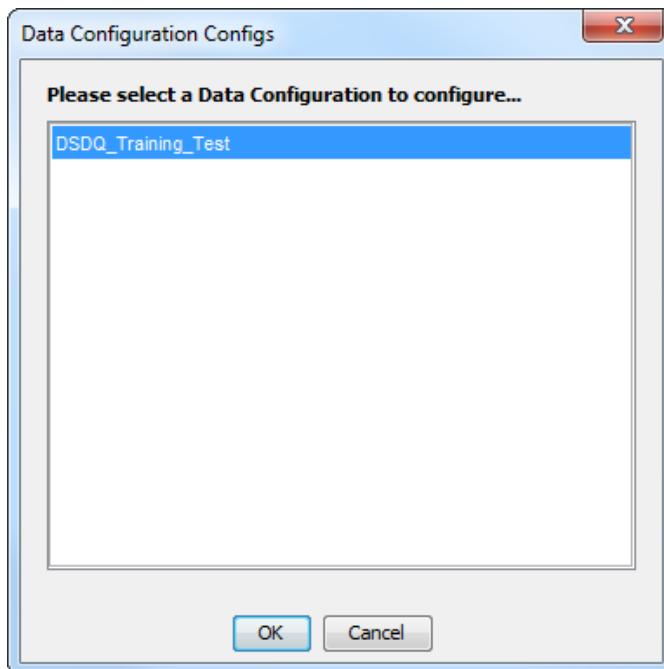
To map tables and columns, you will use the **Configure Merge** Tool. You will need to select the desired element, table or column from the **Source Model** and **Target Model** Tree and map them before they can be tested. Once testing is complete and it is confirmed that correct mapping has been done, you can run the **Merge** Tool.

To use the Configure Merge Tool:

1. Double-click the **Configure Merge** Tool or right-click the **Configure Merge** Tool and select **Open Tool** from the pop-up menu.



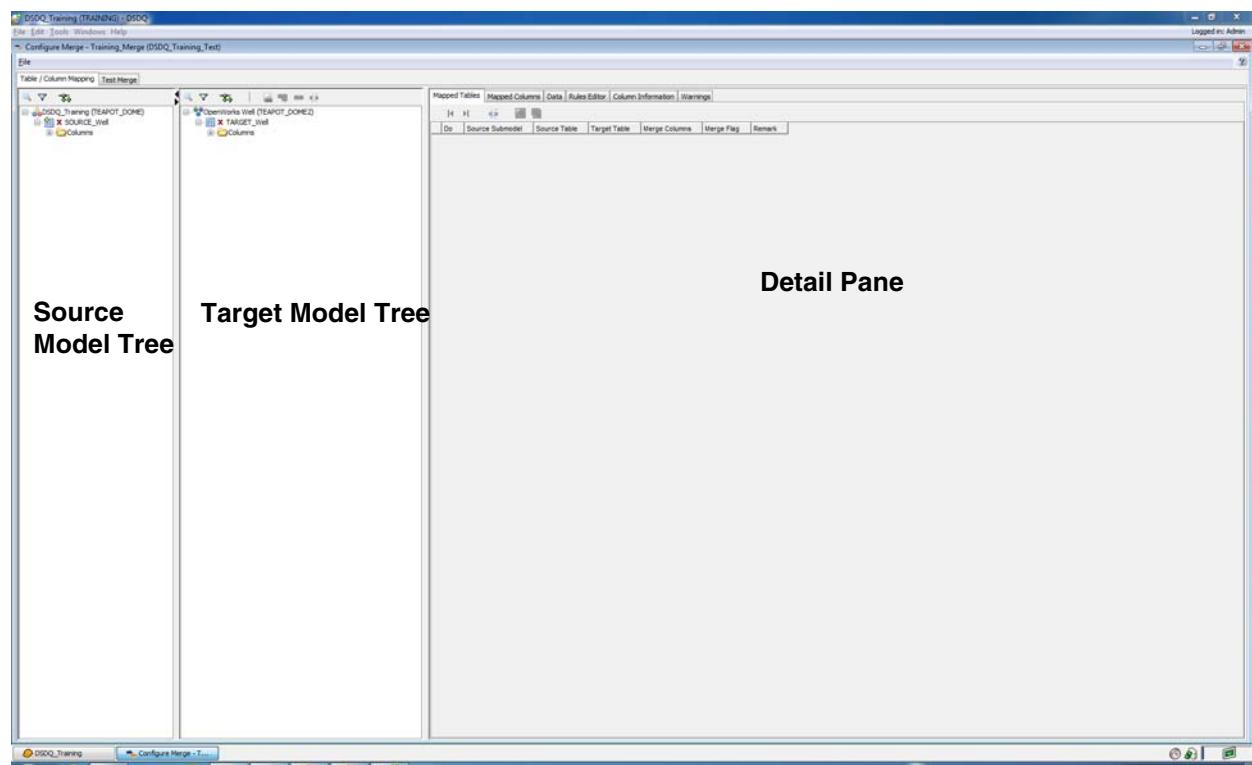
The **Data Configuration Configs** window appears.



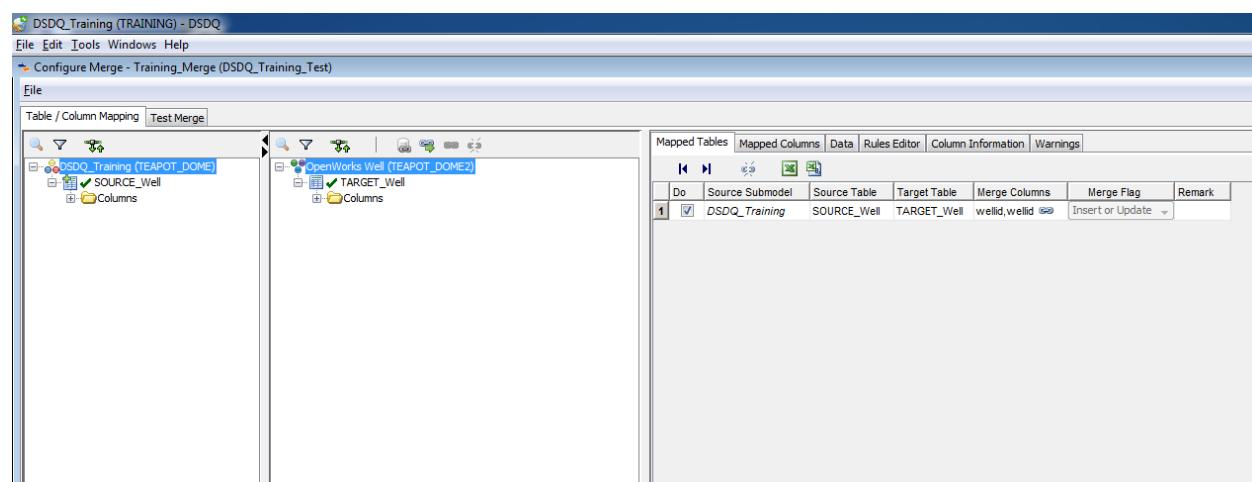
2. Select the **DSDQ\_Training\_Test** option.
3. Click **OK**.

The **Configure Merge** window appears, displaying tables and columns for the Source Model Tree, Target Model Tree and the Detail pane. The Detail Panel has six tabs: **Mapped Tables**, **Mapped Columns**, **Data**, **Rules Editor**, **Column Information**, and **Warnings**:

<b>Mapped Tables</b>	This area displays information about the tables that have been mapped in the source and target trees. Clicking the <b>Link</b>  button under the "Merge Column" header causes the <b>Record Association</b> dialog to be launched and allows configuration of merge columns.
<b>Mapped Columns</b>	This area displays the mapping between the target column and all its mapped sources.
<b>Data</b>	<p>This area displays the source and target tables' data. Selecting a column in either the Source or Target tree highlights the corresponding column in the data view if available. If the selected column has already been mapped, the mapped column data is highlighted in each of the corresponding trees.</p> <p>Data in the Data Detail Pane can be sorted by clicking on a column header. Column information on any column can be viewed by right-clicking on a column header and selecting column Info from the pop-up menu. Other columns information can be viewed by right-clicking on a column header and selecting columns filter from the pop-menu.</p>
<b>Rules Editor</b>	Used to apply rules to specific columns. Simply drag the rule to the target column that the rule has to be applied to. The Methods tab is automatically populated with the relevant information. Make changes to the fields as needed.
<b>Column Information</b>	Displays basic information about the selected column: "Data Type", "Column Size", etc. The tab is divided into two vertical panes: the left one holds the source column information, and the right one holds target column information.
<b>Warnings</b>	Displays any inconsistencies between the mapped columns, e.g. source column length are greater than the target column length. Initially this tab is blank. When the first warning is logged, the tab name turns red and a warning icon appears next to its name.

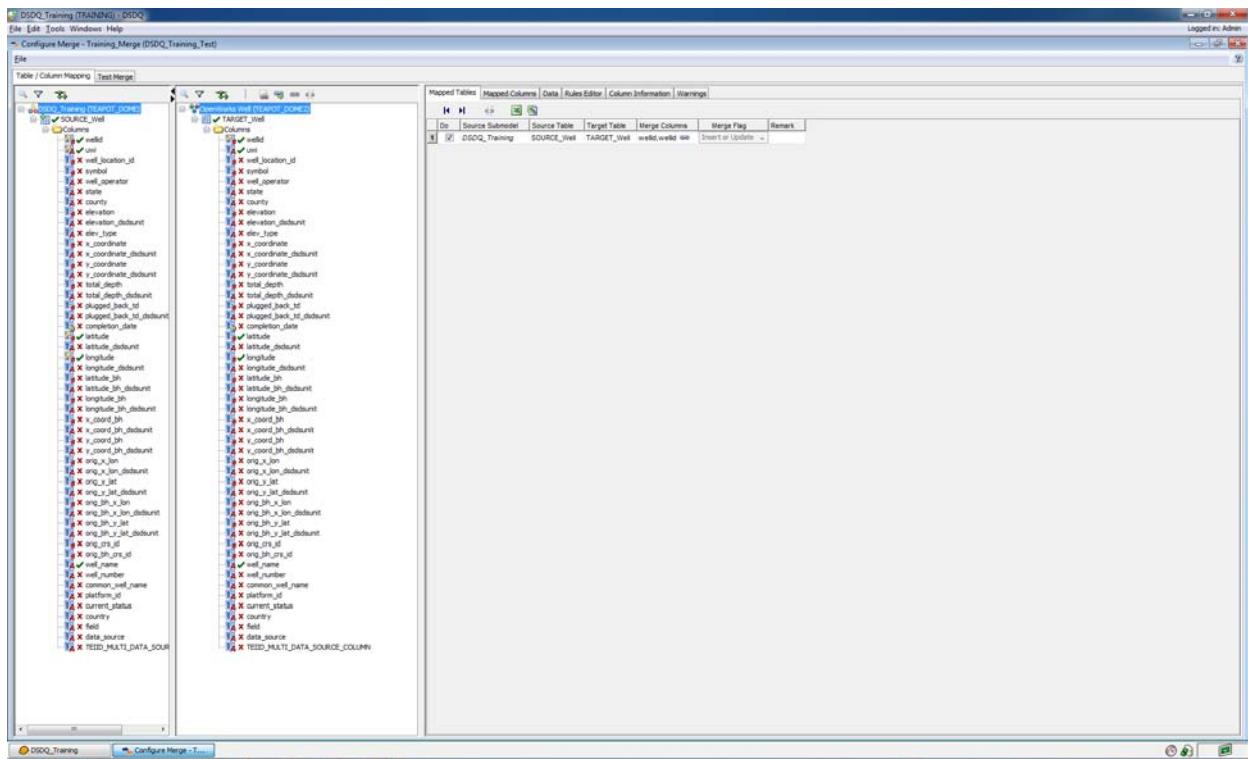


4. Select the **Source\_Well** table from the Source Model Tree.
5. Select the **Target\_Well** table from the Target Model Tree.
6. Click the **Auto Map Selected Tables/Columns**  button on the Target Model Tree toolbar  
A green check mark appears adjacent to the selected tables.



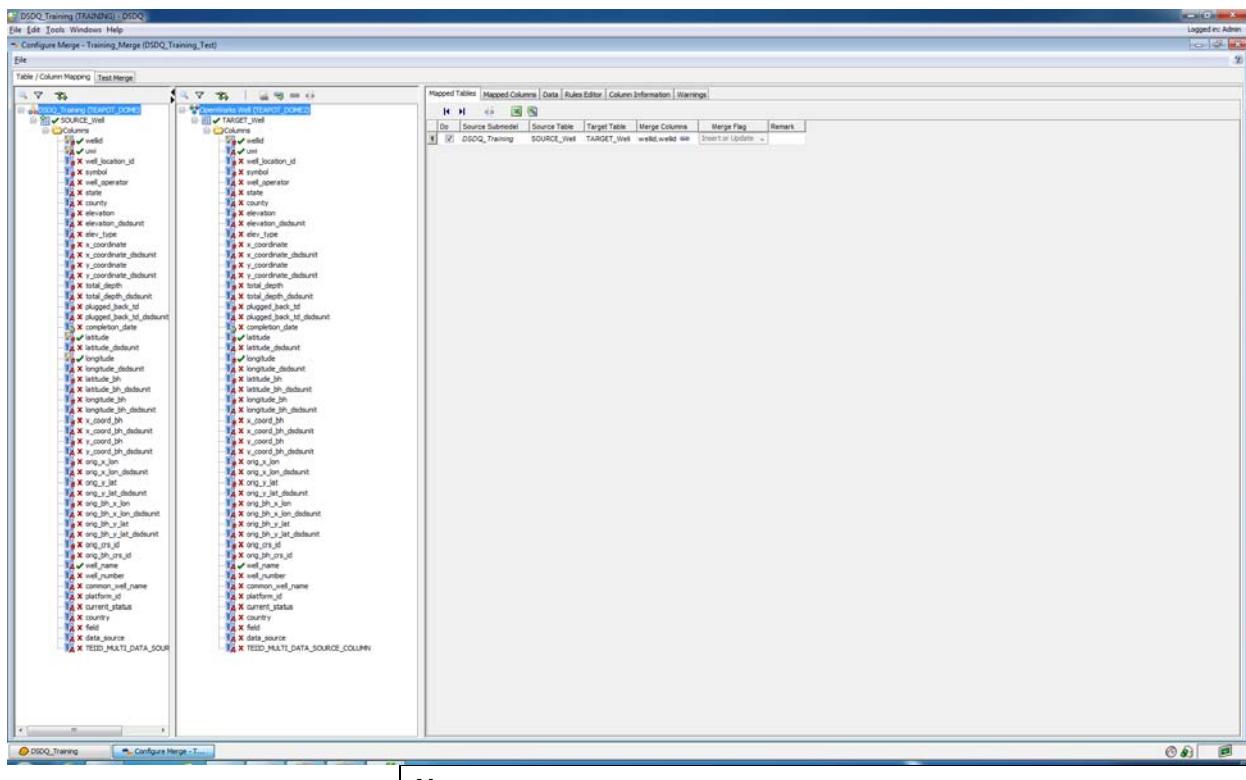
7. Click  to expand the columns on the **Source Model Tree**.

8. Click to expand the columns on the **Target Model Tree**.



9. Select the **Country** element from the **Source\_Well** and **Target\_Well** table.
10. Click the **Map Selected Tables/Columns** button on the **Target Model Tree** toolbar.

A green check mark appears adjacent to the **Country** column in both the Source Model Tree & Target Model Tree.



#### Note

The source and target base tables need to be mapped before their children can be mapped. The number of key columns set in the source table must match the number of key columns set in the target table. If the numbers are not equal, a rule such as 'Generate Sequence Number' may need to be applied to the source or target tables to match the keys.

To unmap the merged columns, select the mapped pair in the lower half of the Record Association window, and click the **Unmap Merge Columns** button.

11. Optionally, repeat steps 9 and 10 to map desired elements for merging.

## Merge Flags for Target Tables

Several merge flags options are available to merge the source table to the target table. Merge flags allow configuring merge differently for every table. The Insert and Update merge flag is set as the default. The available options for the merge flags follows:

Merge Flags for Target Tables Options	Action
Insert Only	Only inserts records that exist in the source table but not in the target table.
Insert or Update	Inserts records that exist in the source table but not in the target table, or carries out updates of records.
Update Only	Only updates records that exist in the source table.
Replace	Deletes all records that exist in the target table and writes all records from source to target.

## Merge Flags for Target Columns

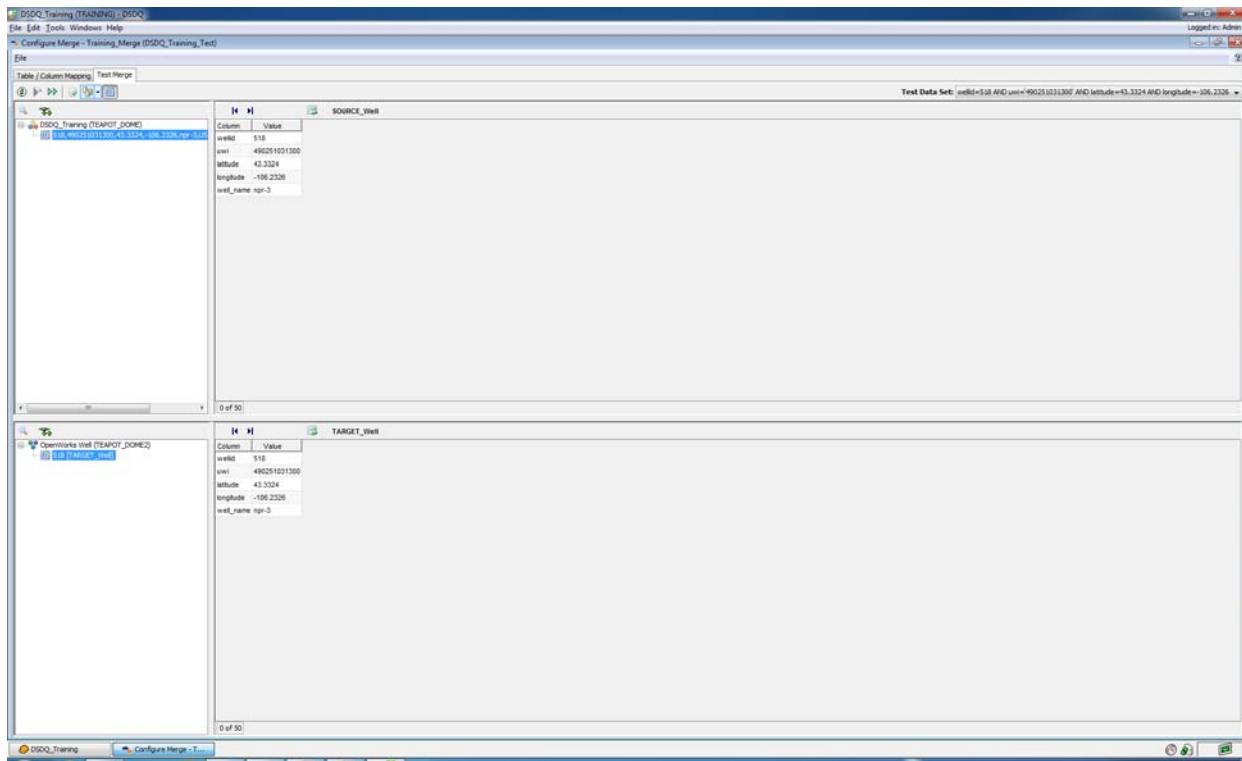
Merge flags allow configuring merge differently for every column. The **Update Always** merge flag is set as the default.

The available options for the merge flags are:

Merge Flags for Target Columns Options	Action
Update Always	Overwrites the target data.
Update Null Values	Inserts the source value if the target is blank.
Append	Appends the source value to the target value.
Append with a Semi-Colon Delimiter	Appends the source value to the target value and places a semi-colon in front of the source value.
Append with Column Name	Places a semi-colon, column name, equal sign, and source value after the target value.

12. Select the **Test Merge** tab.

The test is automatically executed for the first record, the source and target data are displayed one record at a time.



13. Click the **Next Data Set** button to test the next record.

14. Repeat step **10** to check all records.

15. To re-test the current record, click the **Run Migrate|Merge** button again.

**Note**

If there is an “APPEND” flag set for a column, or an append rule is applied to the column, the values in this column will be appended the number of times that the **Run Migrate|Merge** button is executed. For example, if SOURCE\_COLUMN\_A is mapped to TARGET\_COLUMN\_A with the APPEND flag, and the source value is HELLO, the first time the merge is tested, the target will have the value HELLO. If the same record is tested a second time, the target value will be HELLOHELLO. To avoid multiple appending, click the **Refresh All Data** button in the **Test Merge** toolbar, and then click the **Run Migrate|Merge** button.

16. To view just the affected mapped columns in the **Data** table view, click on the **View Affected Columns**  button in the **Test Merge** toolbar.
17. To **Test Merge** against database constraints, click the **Test Against Database**  button in the **Test Merge** toolbar.
18. If the table only has a single row, click the **Switch View** button to view the table in a vertical orientation. Clicking the button again will return to the horizontal view.
19. Select **File > Exit** from the menu bar on the **Configure Merge** window.

---

## Using Merge

---

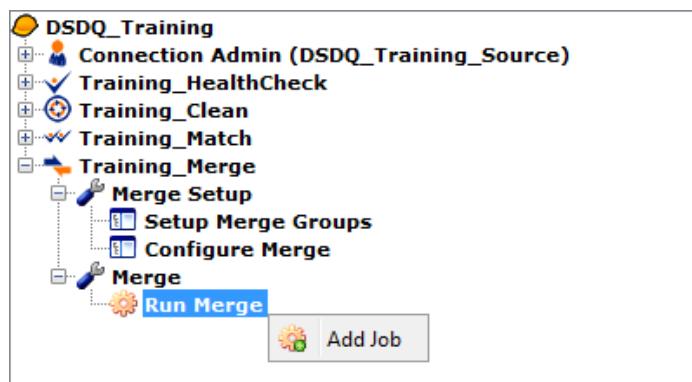
The Merge Activity is used after merge configurations have been tested on your data subset. This activity moves the data to the target database and generates a merge activity report.

### **Exercise: Running the Merge Tool**

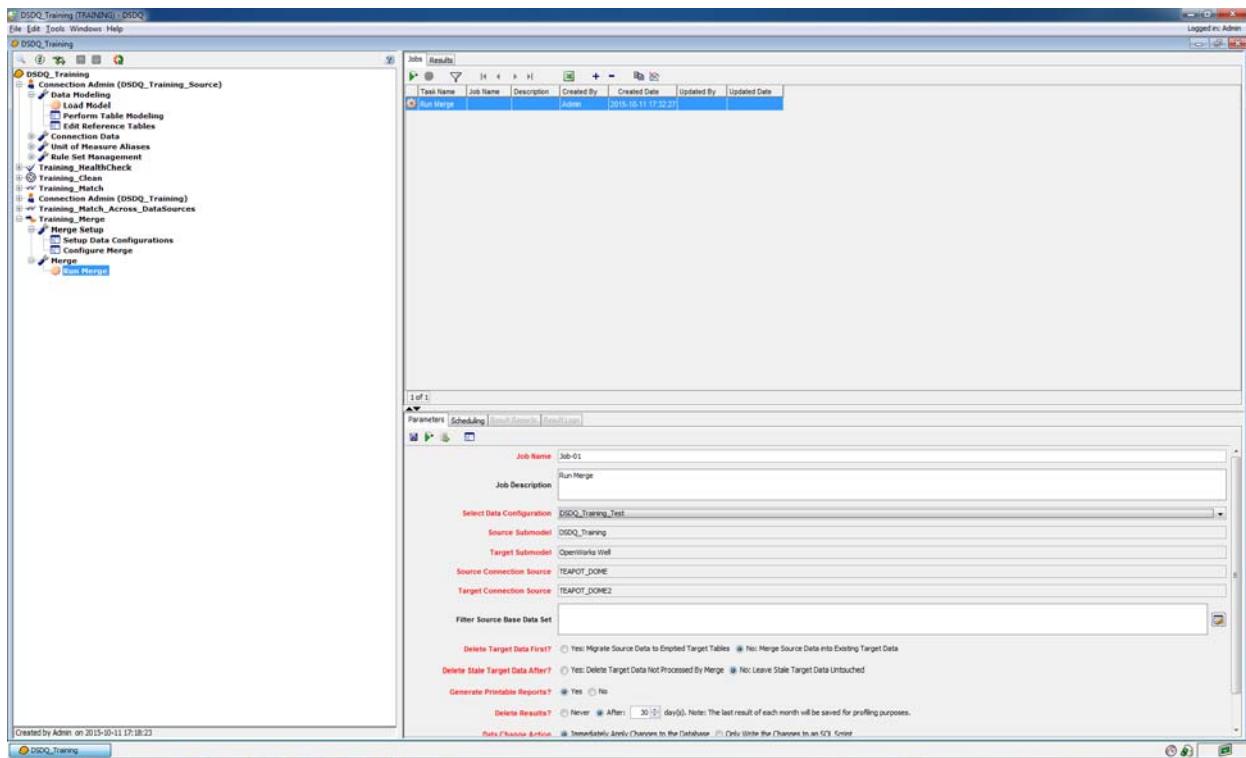
The **Run Merge** Tool is used to move the merge data to the target database. The parameters you would enter for the **Merge** Tool include: Job Name and Merge Group. After the Merge operation is complete, you can check the results on the **Result Reports** tab of the **Jobs and Results Information Pane**.

To run the Merge Tool:

1. Click to expand the **Merge** Activity in the DecisionSpace Data Quality Tree.
2. Double click the **Run Merge** Task or right-click the **Run Merge** Task and select **Add Job** from the pop-up menu.



A new job is initiated and displays on the **Jobs and Results Listing Pane**.



3. Enter **Job-01** in the **Job Name** field.
4. Enter **Run Merge** in the **Job Description** field.
5. Select **DSDQ\_Training\_Test** from the **Select Data Configuration** drop-down list.  
The **Source Submodel** and **Target Submodel** fields populate automatically.
6. Optionally, set a filter on the data set.
7. Select the **No: Merge Source Data into Existing Target Data** option for **Delete Target Data First**.

#### Note

**Delete Target Data First** allows the merge job to be run as a migration or as a merge. Selecting **Yes: Migrate Source Data to Emptied Target Tables** truncates the target table and then merges the source, while selecting **No: Merge Source Data into Existing Target Data** allows the target data to persist and try to merge the Source data into the target.

**Warning**

Selecting **Yes: Migrate Source Data to Emptied Target Tables** will completely remove all information in the target tables.

**Warning**

Selecting **Yes: Delete Target Data Not Processed By Merge** will permanently delete all orphan rows in the target submodel and tables in your target submodel that are not mapped to a table in your source submodel.

8. Select the **No: Leave Stale Target Data Untouched** option for the **Delete Stale Target Data After**.
9. Select the **Yes** option for **Generate Printable Reports?**
10. Select the **After** option for **Delete Results?** Set the number of days as **30**.
11. Select the **Immediately Apply Changes to the Database** option for **Data Change Action**.

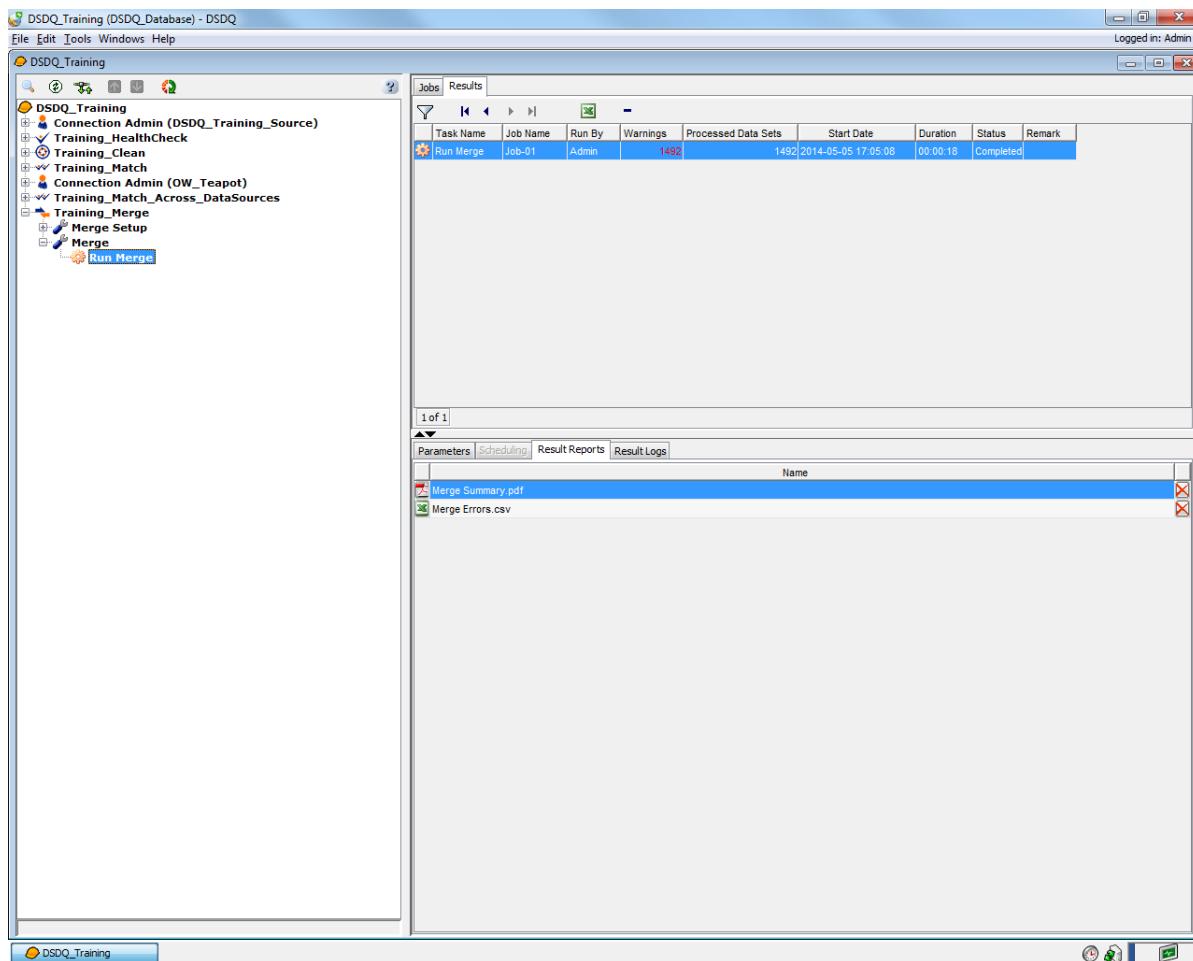
**Note**

Selecting **Immediately Apply Changes to the Database** will apply the job changes directly to the configured data source. Or if you want to write a Sql script of the changes, select **Only Write the Changes to an SQL Script**. This creates a “Database Changes.sql” file to be produced and can be viewed upon job completion in the **Results Reports** tab.

12. Select the **No** option for **Enable Audit Logging**.
13. Click  to save changes in the **Parameter** tab.
14. Click .

The **Run Merge Task** is executed and displays results in the **Result Reports** tab of the **Job and Results Information Pane**.

15. Select the **Results** tab on the **Jobs and Results Listing Pane** to view the values in the **Result Reports** tab on the **Job and Results Information Pane**.



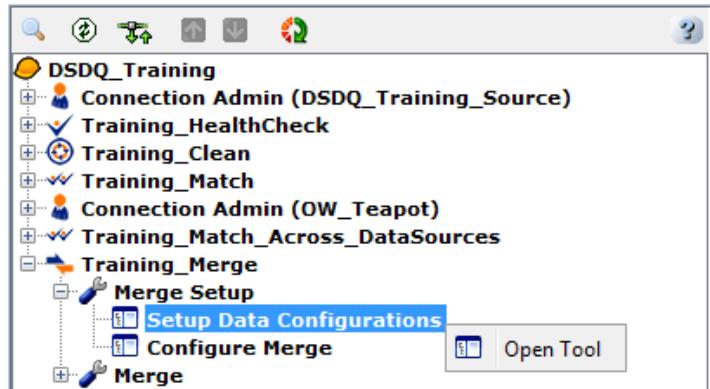
16. Click on the **Result Reports** tab to display Merge Summary in PDF format.

Merge Summary						<b>HALLIBURTON</b> Landmark Software & Services
Project:	DSDQ_Training	Phase:	Training_Merge	Task:	Merge	
Job:	Job-01	Merge Group:	DSDQ_Training_Test	Result Date:	2014-01-15 15:00:41	
Source Table	Target Table	Total Rows	Rows Processed	Rows Succeeded	Rows Failed	Target Table Remarks
Well	Well	1388	1388	1388	0	

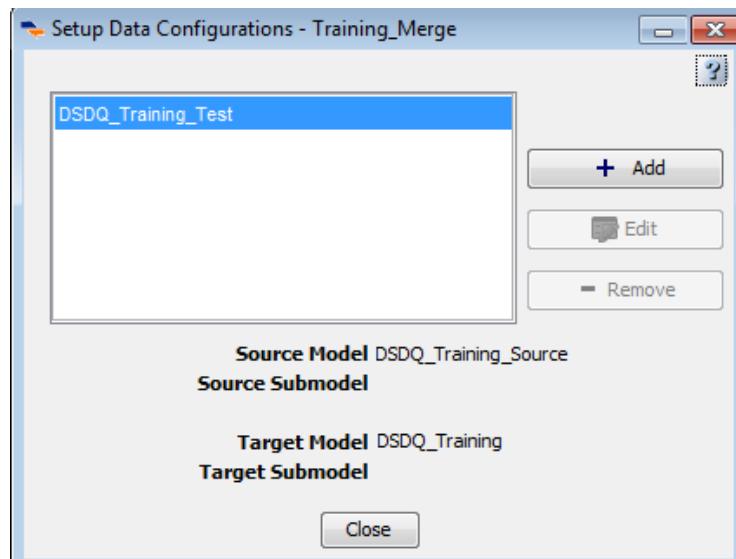
## Exercise: Modifying Data while Merging

To modify data while merging:

1. Double-click the **Setup Data Configurations** Tool or right-click the **Setup Data Configuration** Tool and select **Open Tool** from the pop-up menu.



The **Setup Data Configuration** window appears.

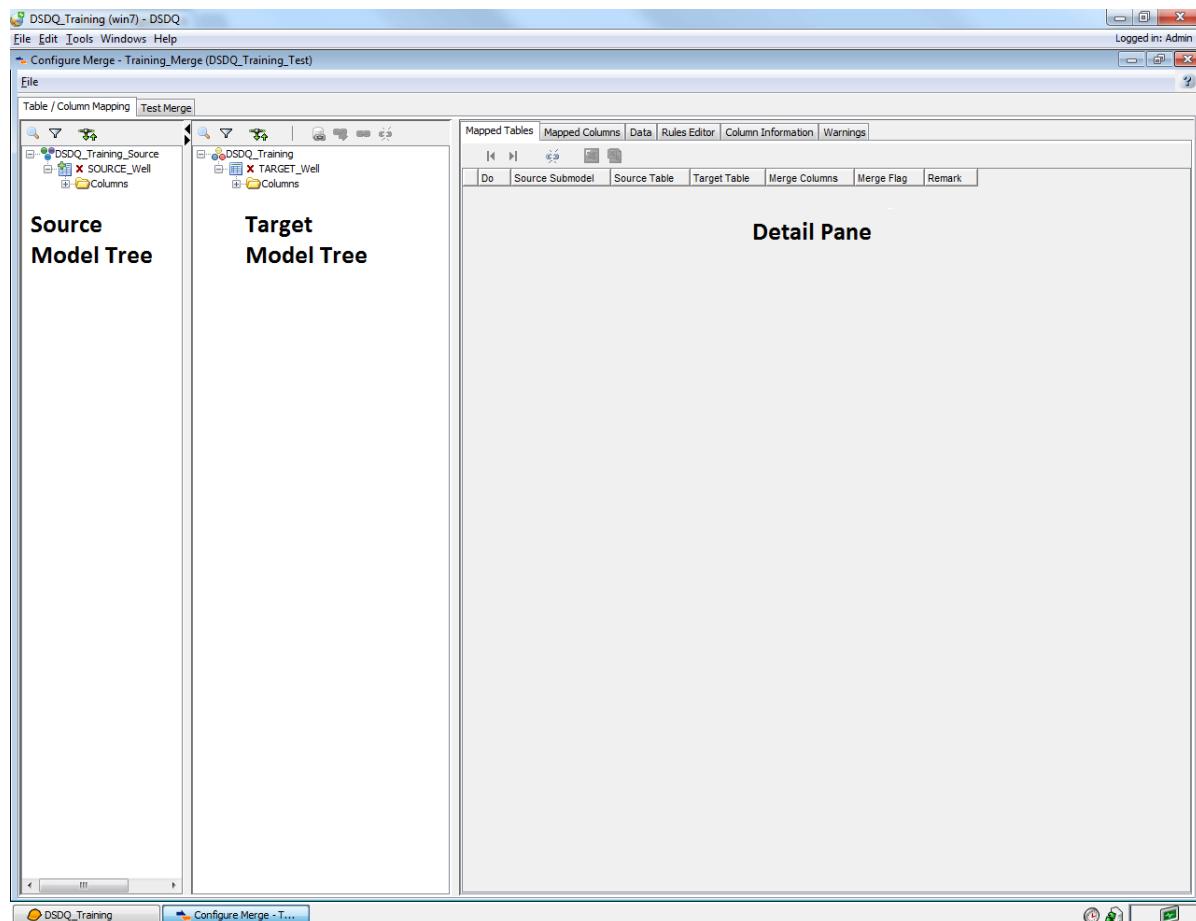


2. Select the **DSDQ\_Training\_Test** option.

3. Click **OK**.

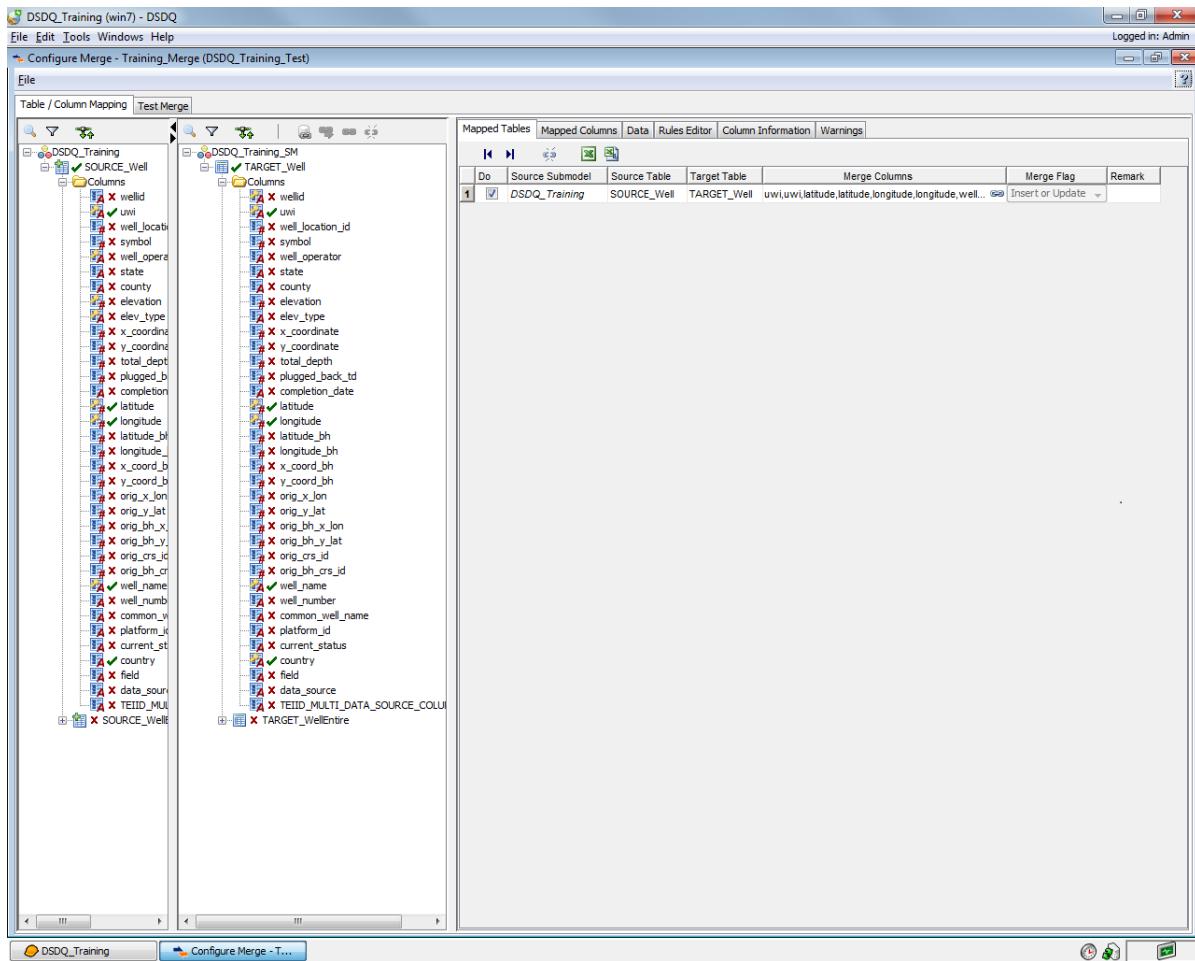
The **Configure Merge** window appears, displaying tables and

columns for the Source Model Tree, Target Model Tree and the Detail pane.



4. Click to expand the **Columns** folder on the **Source Model Tree**.

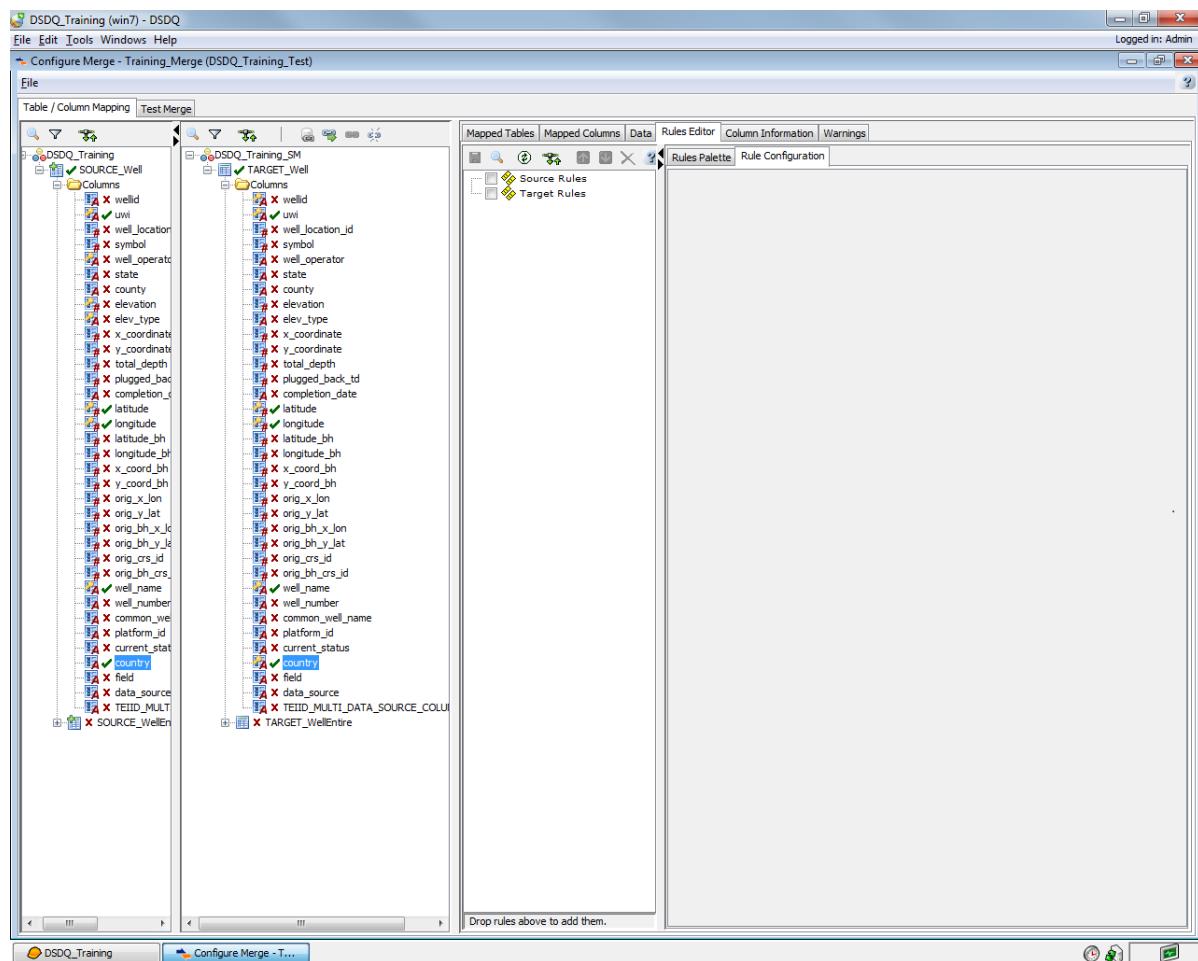
- Click **+** to expand the **Columns** folder on the **Target Model Tree**.



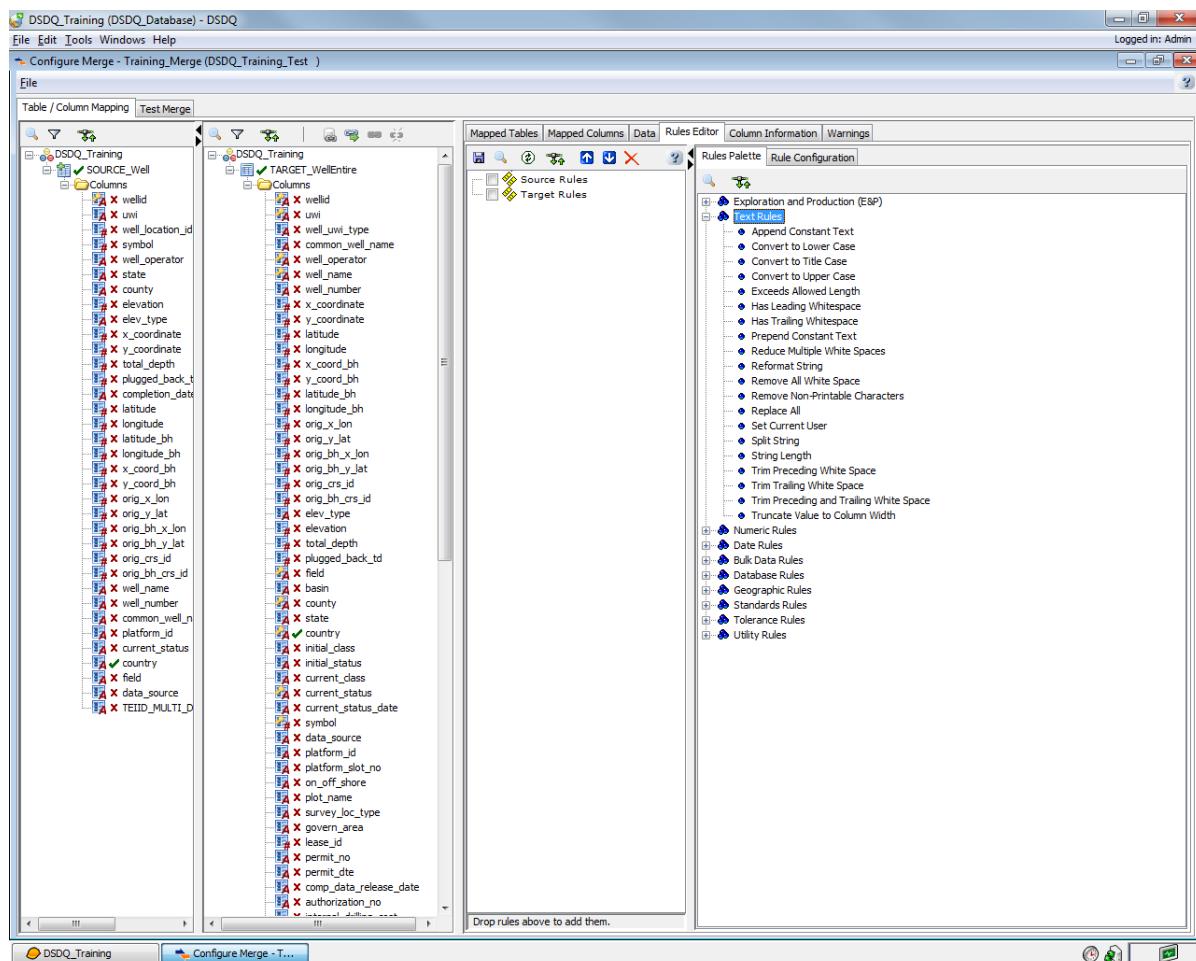
- Select the **Country** element from the Source\_Well and Target\_Well tables.

7. Select the **Rule Editor** tab from the **Detail Pane**.

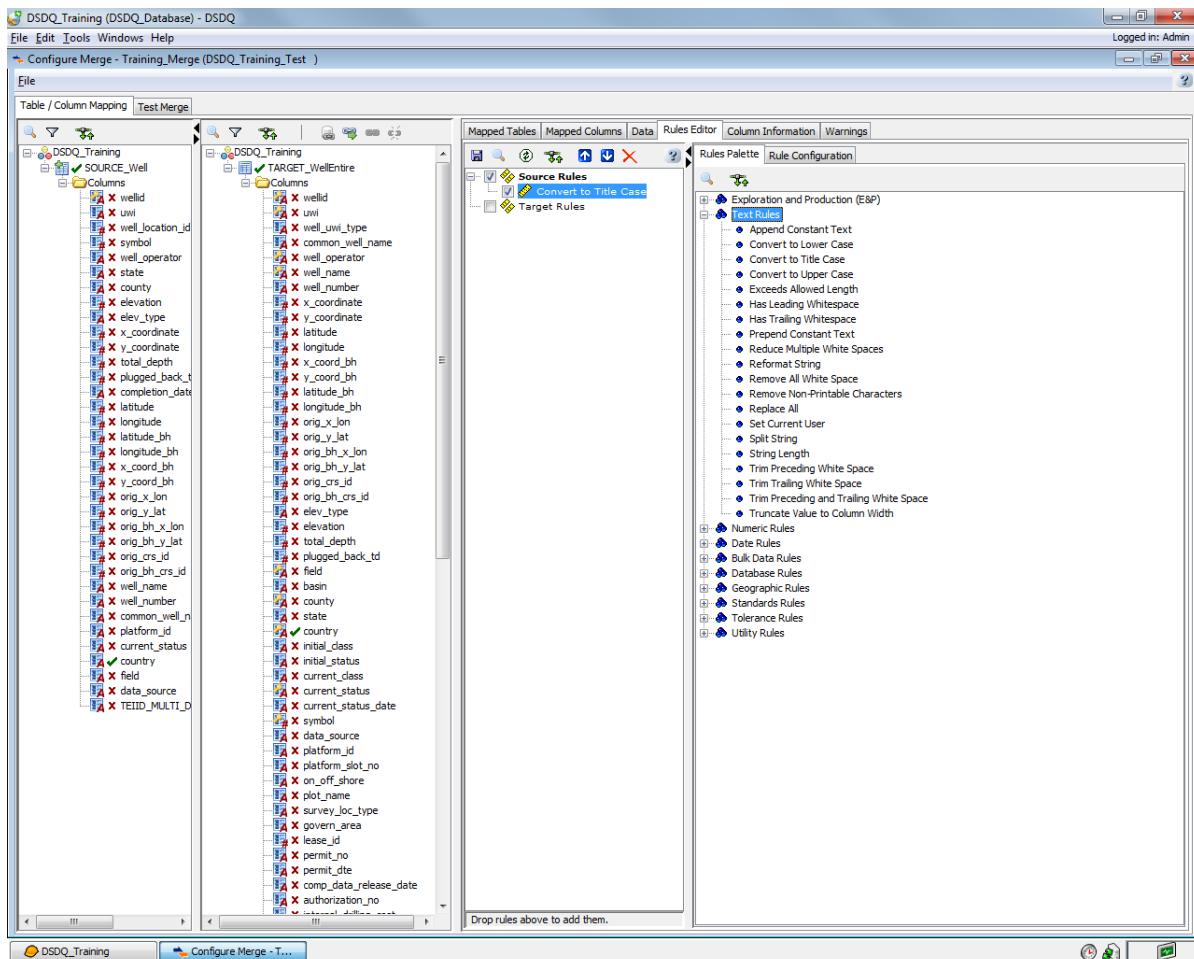
The **Rule Configuration** tab displays in the **Detail Pane** of the **Configure Merge** window.



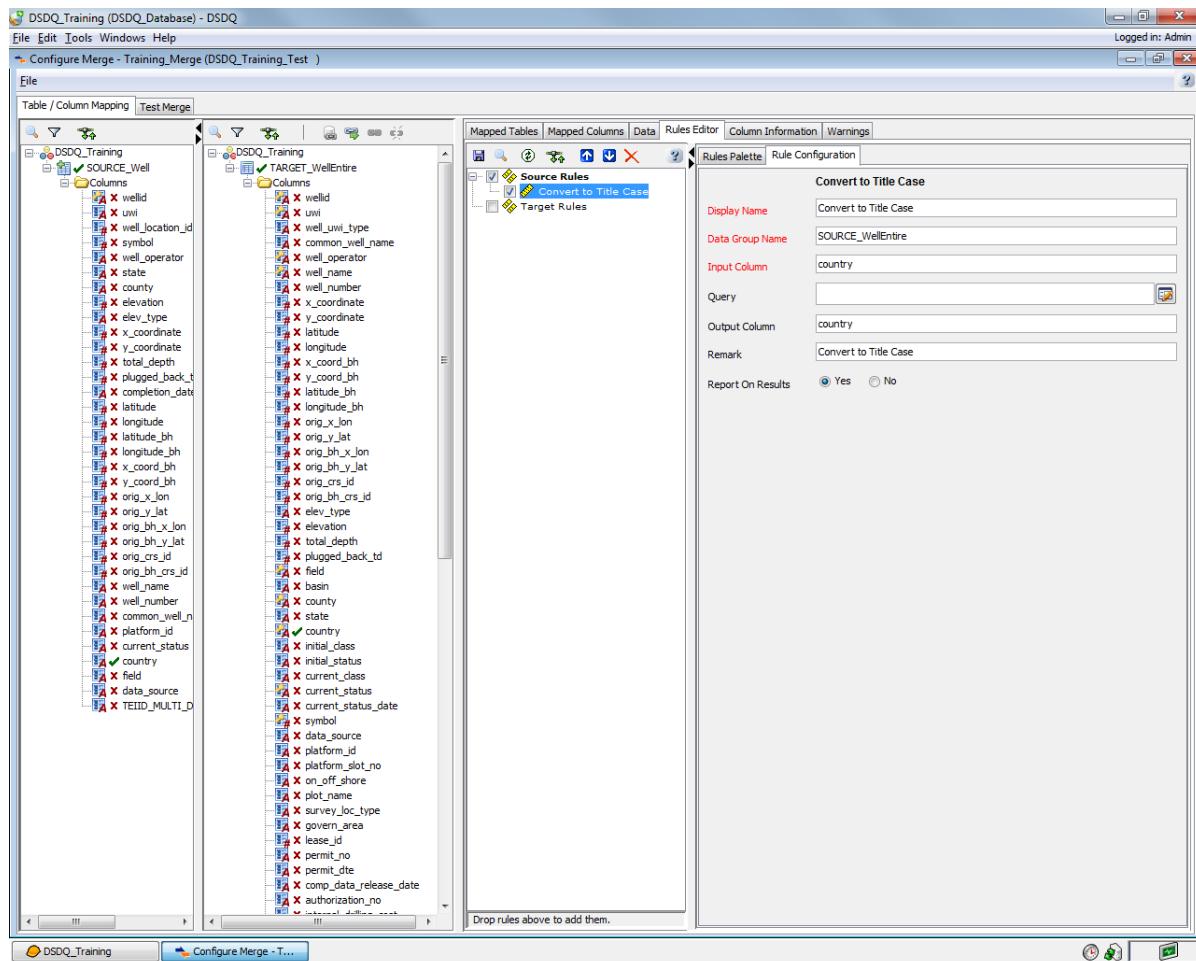
8. Select the **Rules Palette** tab and click to expand the **Text Rules**.



9. Drag and drop the **Convert to Title Case** rule from the **Text Rules** to the **Source Rules**.



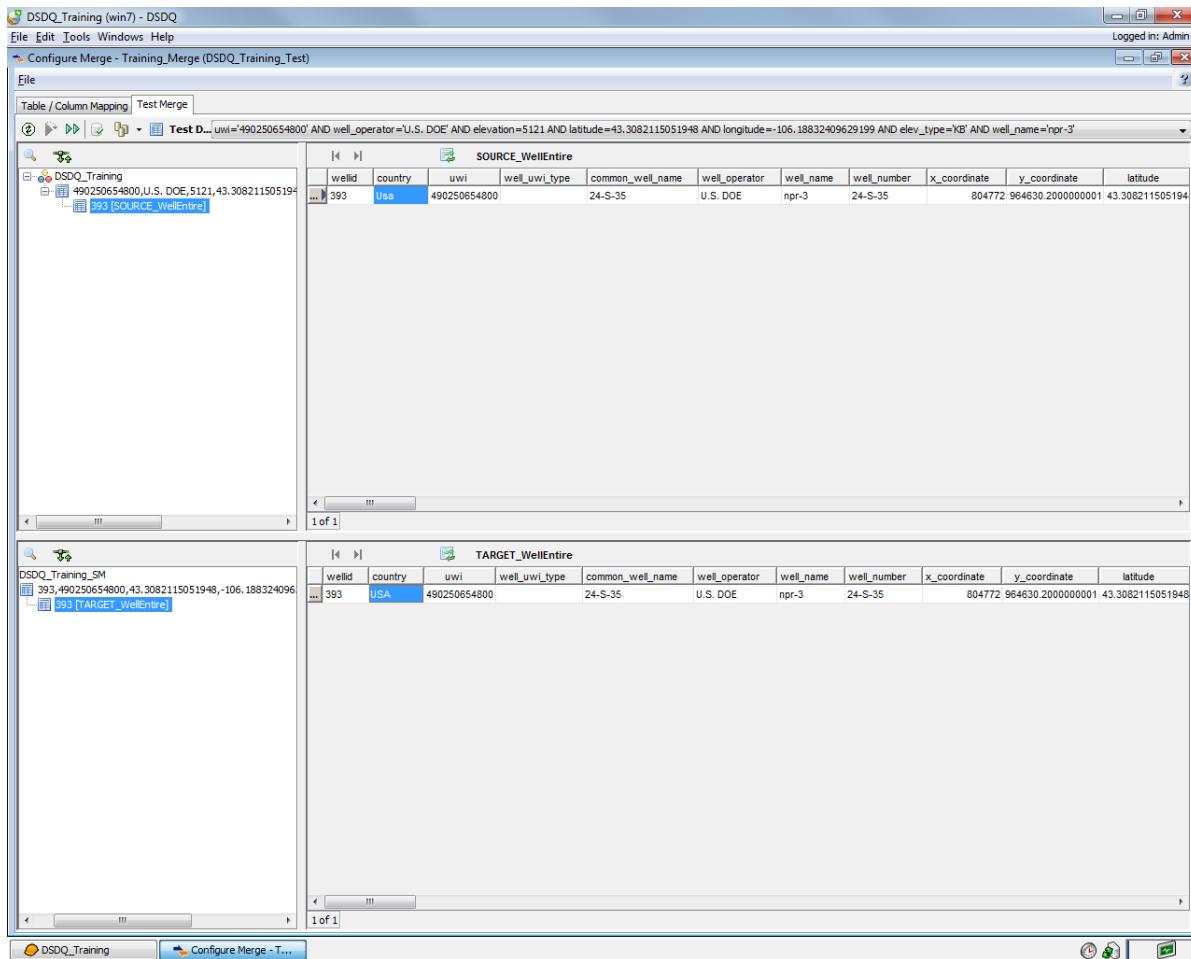
The **Rule Configuration** tab displays by default.



10. Enter **Convert to Title Case** in the **Display Name** field.
11. Enter **SOURCE\_WellEntire** in the **Data Group Name** field.
12. Enter **country** in the **Input Column** field.
13. Optionally, select a **Query** from the **Query** field.
14. Enter **country** in the **Output Column** field.
15. Enter **Convert to Title Case** in the **Remark** field.
16. Select the **Yes** option for the **Report On Results**.
17. Click to save changes in the **Rule Configuration** tab.

**18. Select the **Test Merge** tab.**

The test is automatically executed for the first record.



**19. Click the **Next Data Set** ➤ button to test the next record.**

20. Repeat step **19** to check all records.

21. Select **File > Exit** from the menu bar on the **Configure Merge** window.



## *Chapter* 8

# ***Requirements Administrator in DecisionSpace Data Quality***

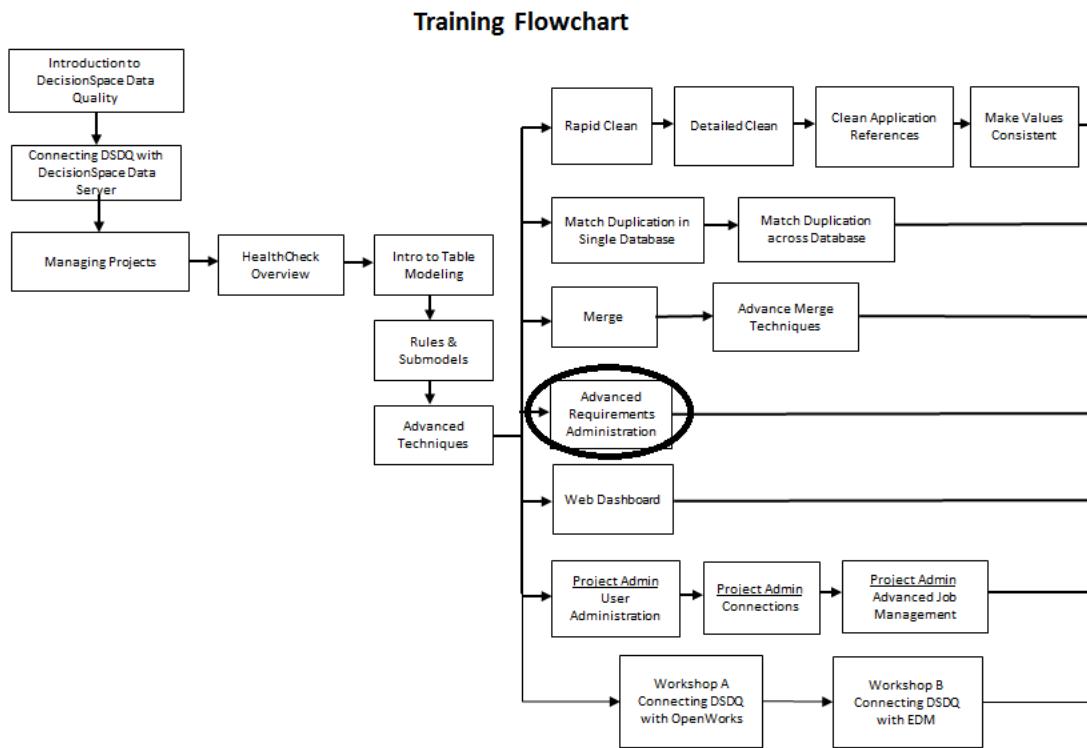
The Requirements Administrator tool manages the Data Quality Rules Repository. The tool is divided into two sections with respective folder trees: Repository and Service Level. The Repository Tree contains the original values in the repository and should only be updated if a new global requirement needs to be added or modified. Service Levels are copies and subsets from the Repository that can be modified without altering the original (master) version.

# Chapter Overview

In this chapter, you will learn about:

- Managing the Repository
- Managing the Service Levels
- Managing the Requirements
- Rules Editor.

Topics covered in each chapter are outlined in the following illustration. Those specific to the current chapter will be circled in black for your reference:



## Repository

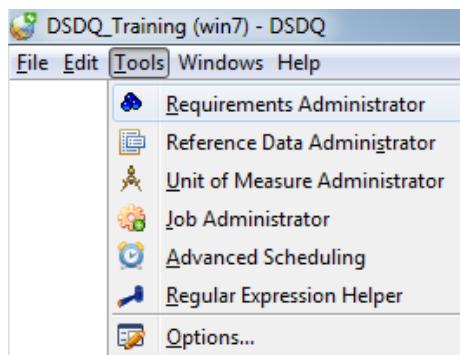
---

The Repository Tree Pane contains the original values in the repository and should only be updated if a new global requirement needs to be added or modified. All local changes are saved in the Service Level Tree Pane.

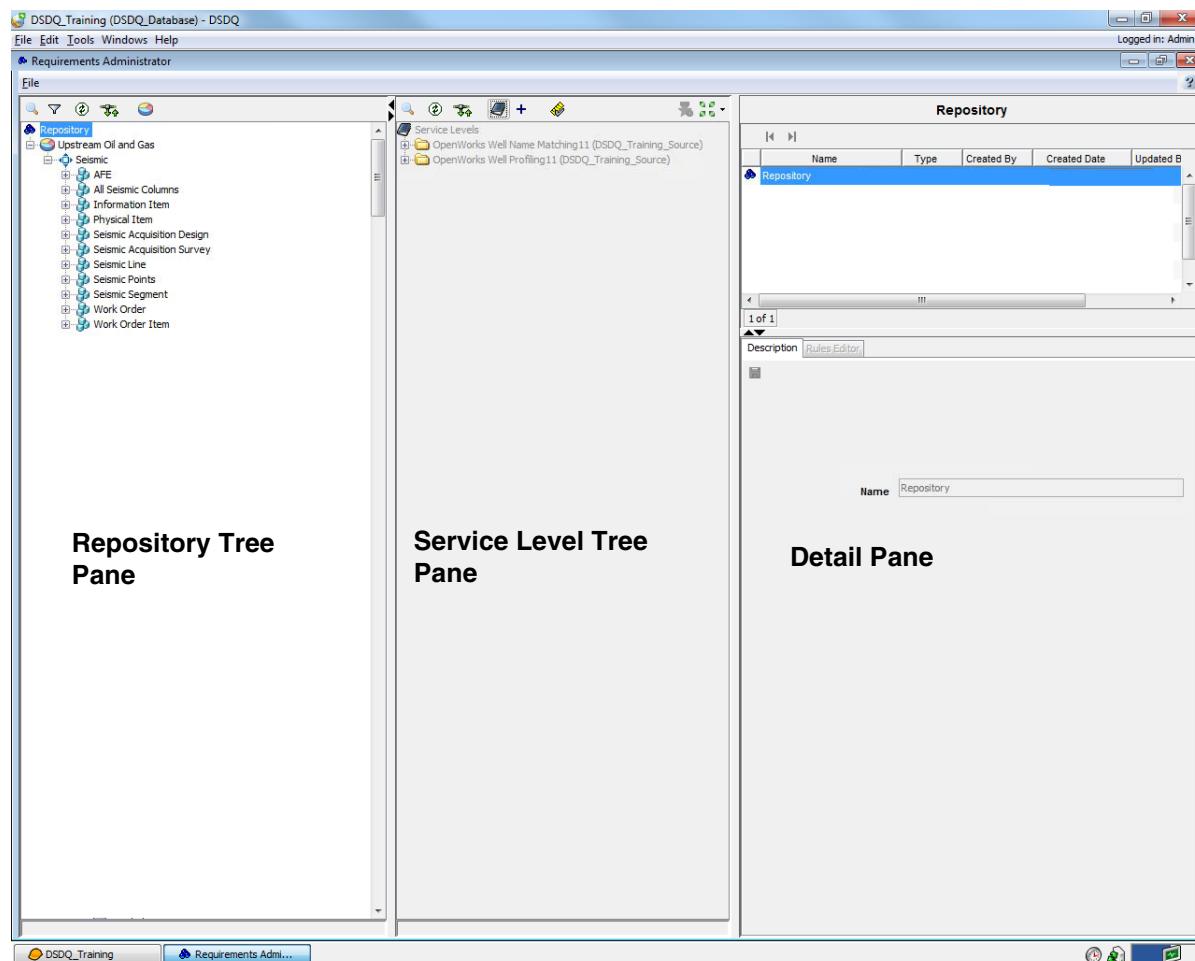
### ***Exercise: Populating the Repository***

To populate the Repository:

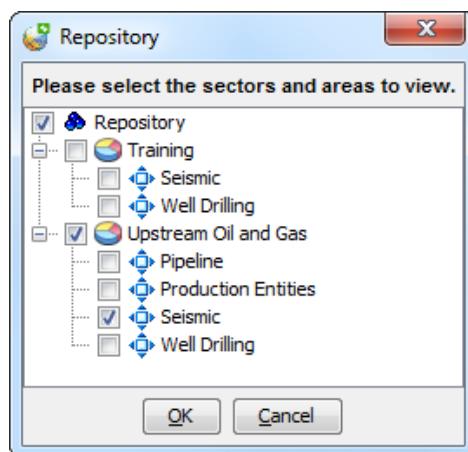
1. Select **Tools > Requirements Administrator** from the menu bar on the DSDQ Project window.



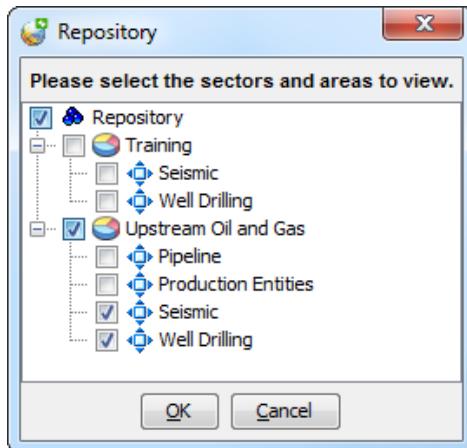
The Requirements Administrator window appears.



2. Click on the **Repository Tree** pane toolbar.  
The **Repository** dialog box appears.

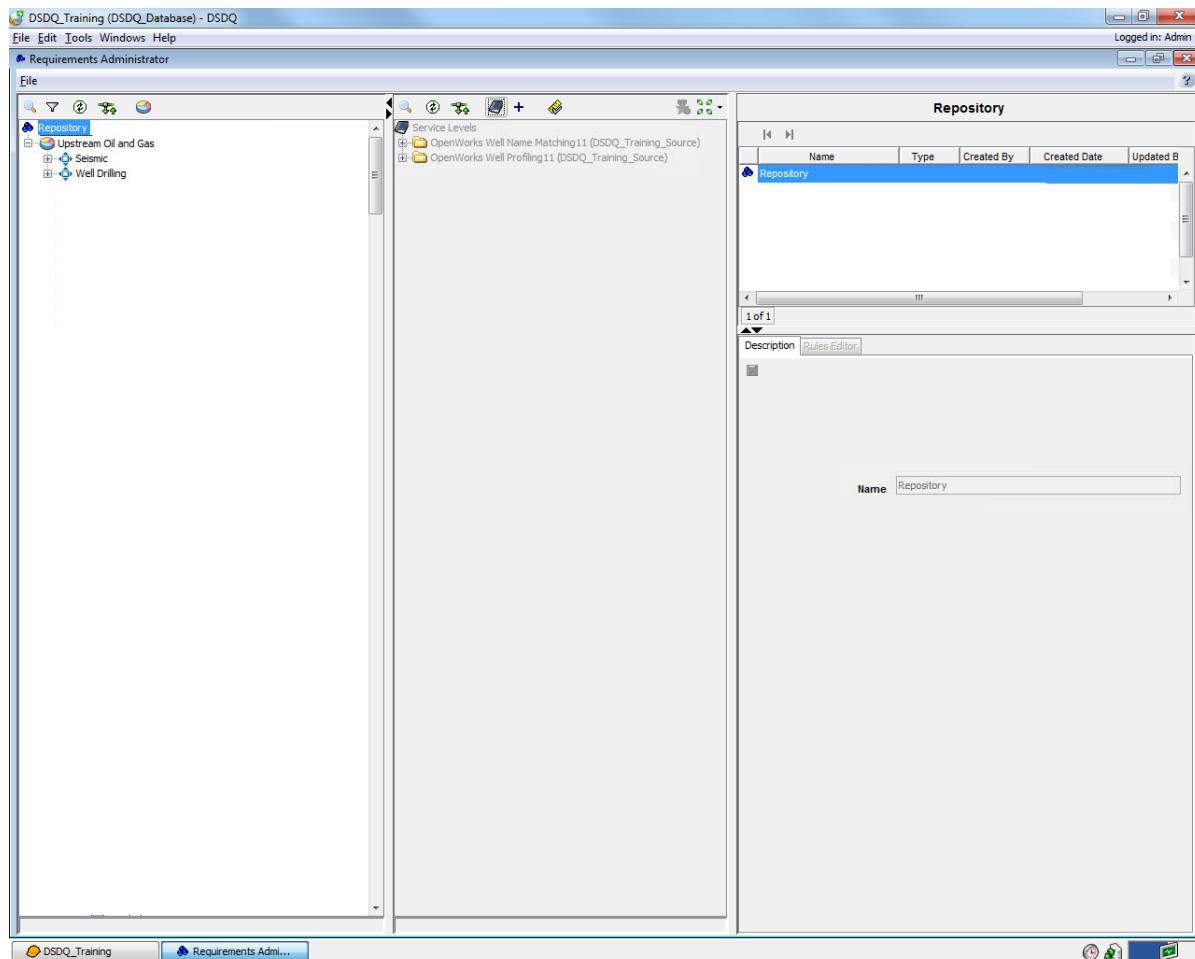


3. Select the **Well Drilling** area from the **Upstream Oil and Gas** sector.



4. Click **OK**.

The selected repository area appears in the **Repository Tree Pane**.



**Note**

You can select and add multiple areas from the **Repository** dialog box.

5. Select **File > Exit** from the menu bar on the **Requirements Administrator** window.

**Note**

You can select an Area from the **Repository Tree** in the **Requirements Administrator** window to view its details that are populated in the **Details** Pane.

## Service Levels

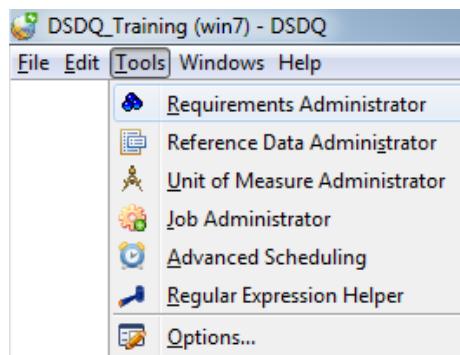
---

Service levels create a subset of requirements to be run against a dataset. One or many service levels can be created for a single dataset. Service levels are copies and subsets from the repository that can be modified without altering the original (master) version.

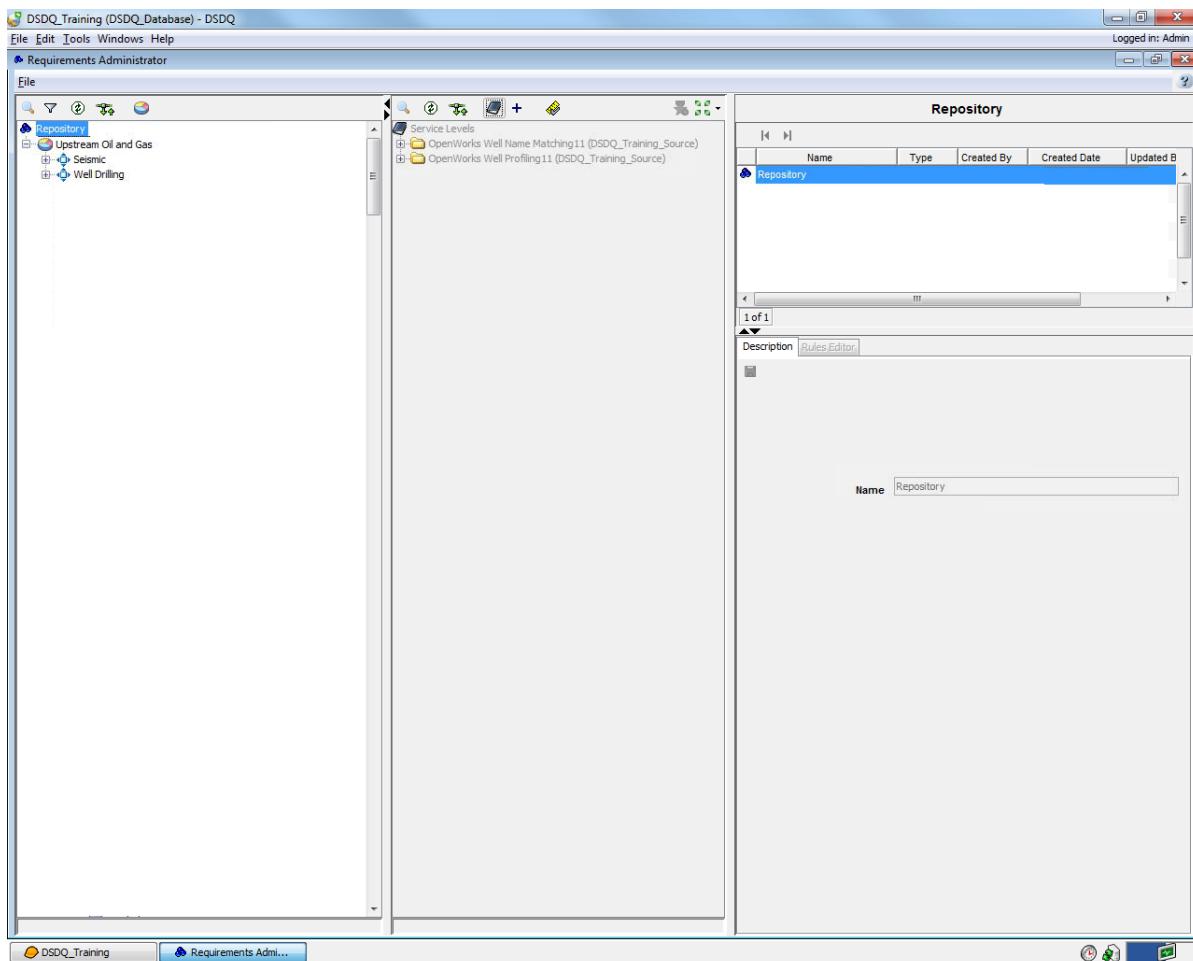
### **Exercise: Creating a Service Level**

To create a service level:

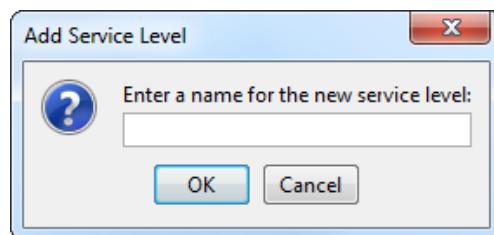
1. Select **Tools > Requirements Administrator** from the menu bar on the **DSDQ Project** window.



The Requirements Administrator window appears.



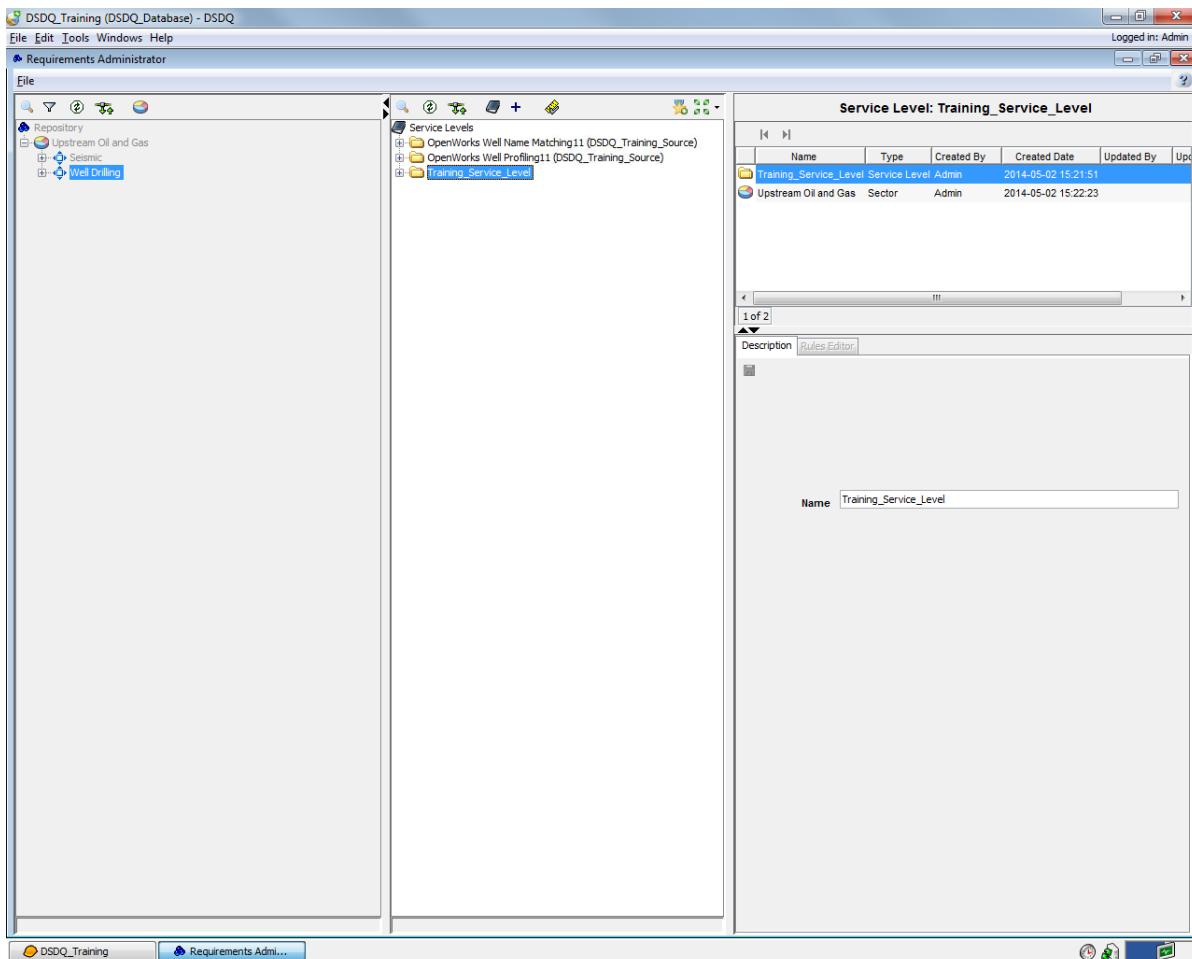
2. Click on the Service Level Tree Pane toolbar.  
The Add Service Level dialog box appears.



3. Enter **Training\_Service\_Level** in the **Enter a name for the new service level** field.

4. Click **OK**.

The service level is added in the Service Level Tree.



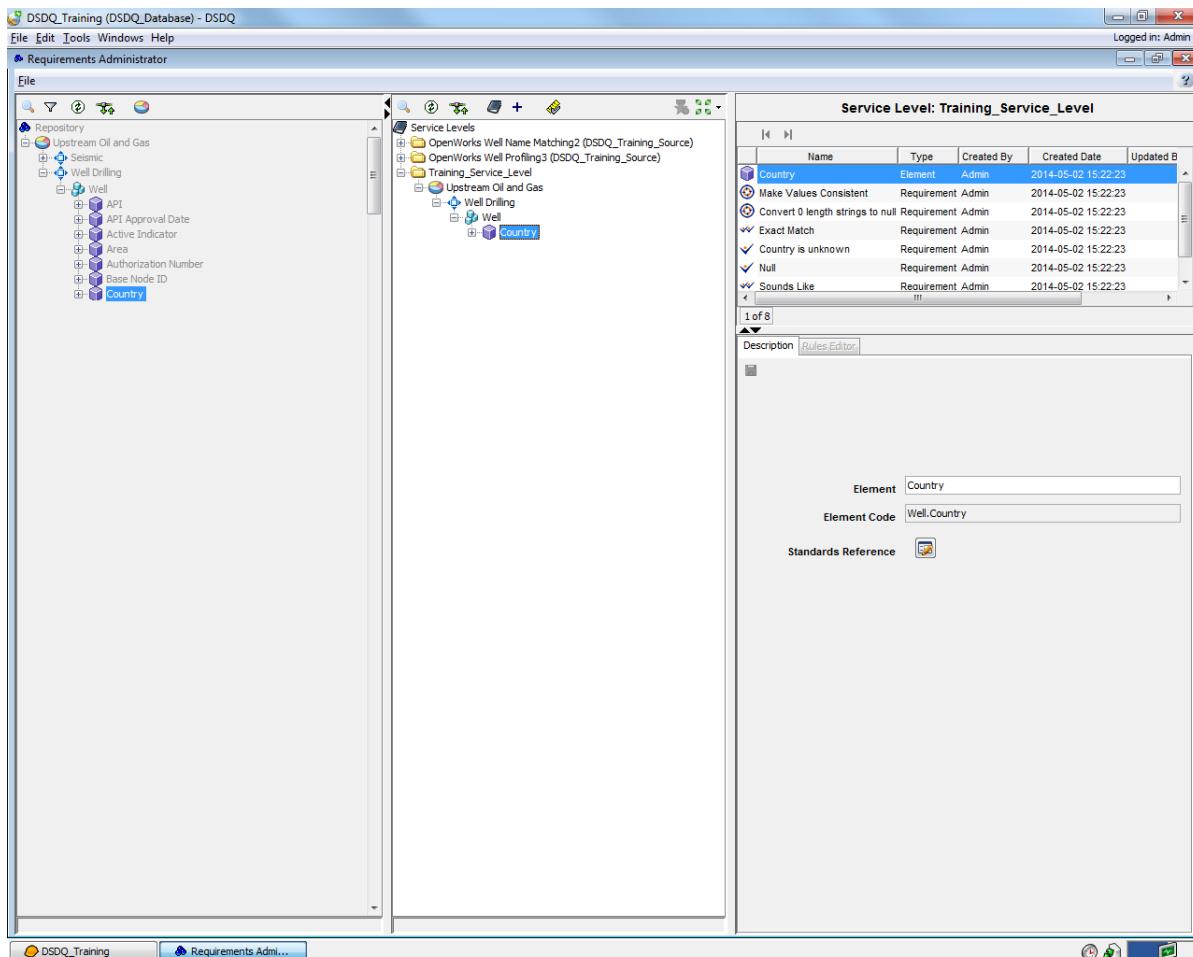
5. Optionally, add more service levels by repeating steps **2** and **3**.

6. Click **+** to expand the **Upstream Oil and Gas** sector in the **Repository Tree**.

7. Click **+** to expand the **Well Drilling** area.

8. Click **+** to expand the **Well** element group.

9. Drag and drop the **Country** element onto the **Training\_Service\_Level** service level.  
 The element is added in the Service Level Tree.



#### Note

If an Element Group is dragged onto a Service Level, all of its associated Elements and Requirements are copied as well. If an Element is dragged onto a Service Level without the Element Group, the application adds the Element Group that the Element belongs to automatically if it does not already exist in the Service Level (similar behavior occurs for Requirements as well). The drag-and-drop operation between the trees is saved automatically.

10. Optionally, add more elements to the service level by repeating steps 5 to 9.

11. Optionally, select a service level from the Service Level Tree to view its details that are populated in the **Details** Pane.

**Note**

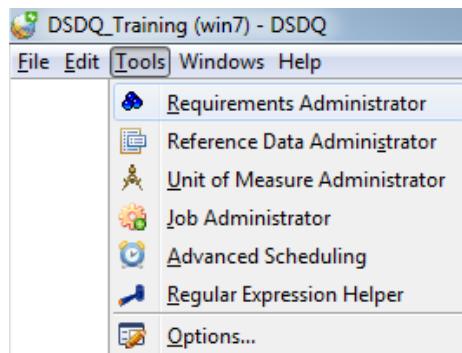
You can publish a service level to the repository. Select the desired Service Level from the Service Level Tree and click the **Publish Service Level to Repository** button from the Service Level Tree Pane toolbar.

12. Select **File > Exit** from the menu bar on the **Requirements Administrator** window.

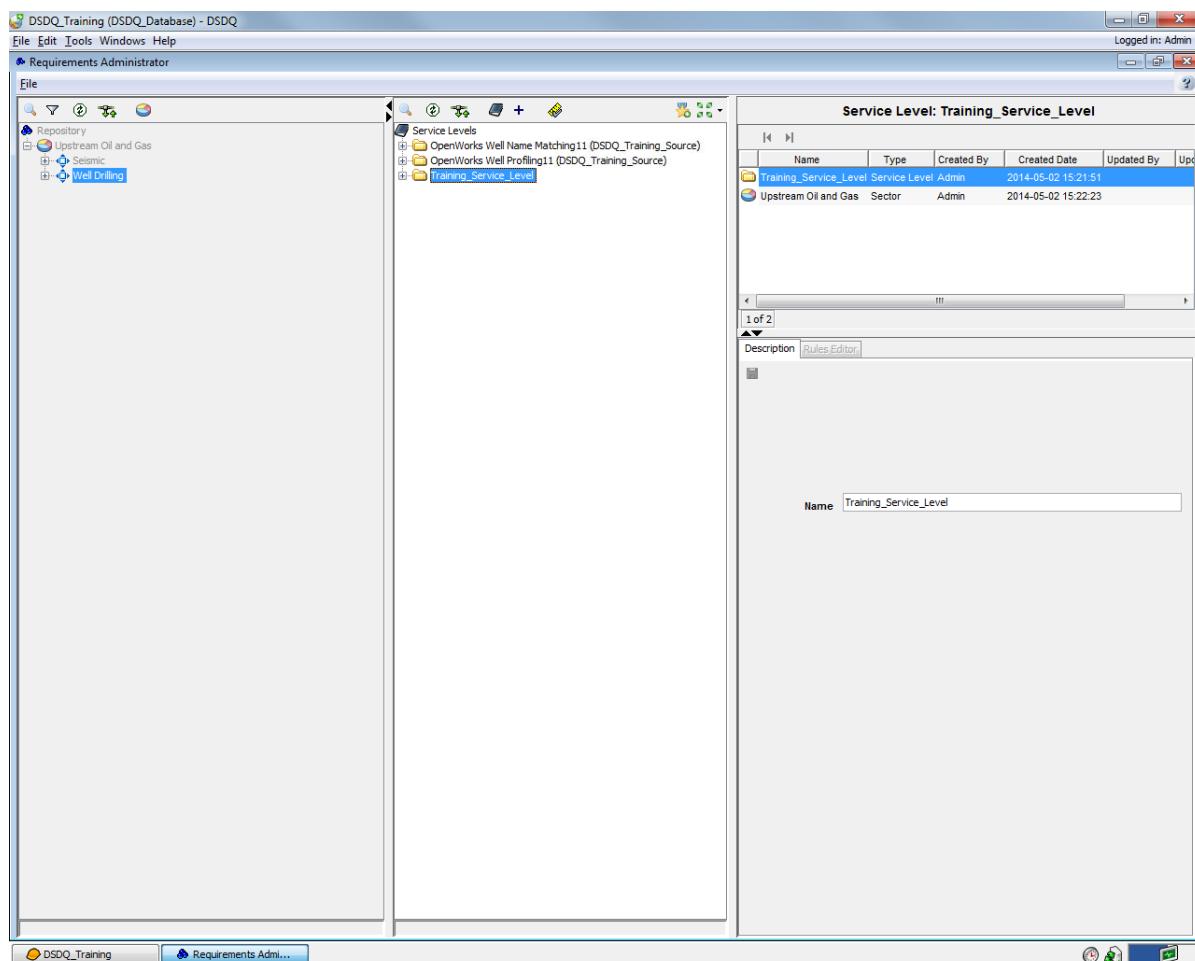
## Exercise: Adding a Service Level

To add a service level:

1. Select **Tools > Requirements Administrator** from the menu bar on the **DSDQ Project** window.

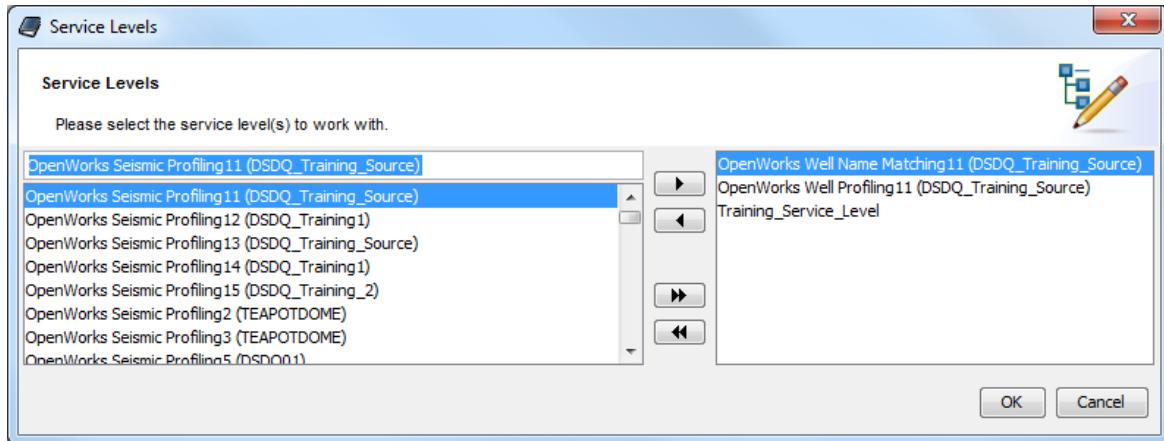


The **Requirements Administrator** window appears.



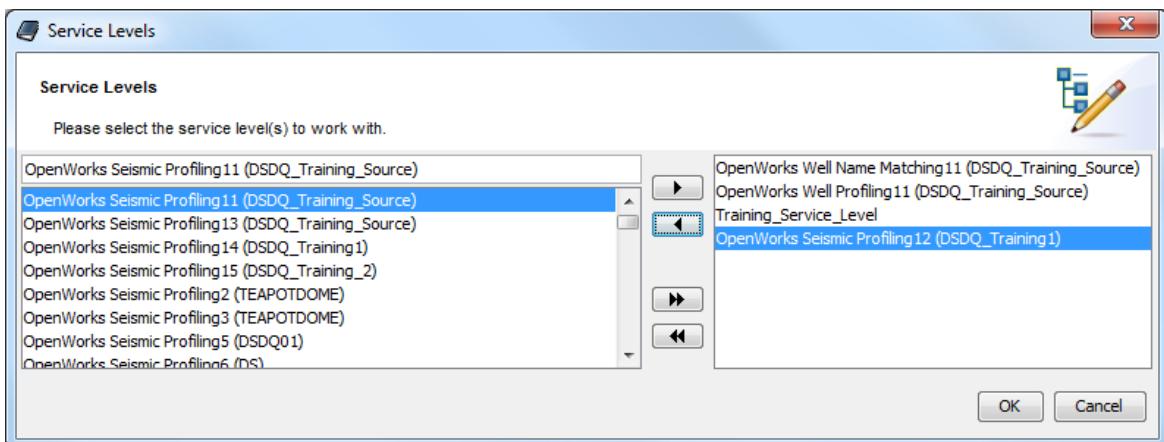
2. Click  on the Service Level Tree toolbar.

The Service Levels dialog box appears.



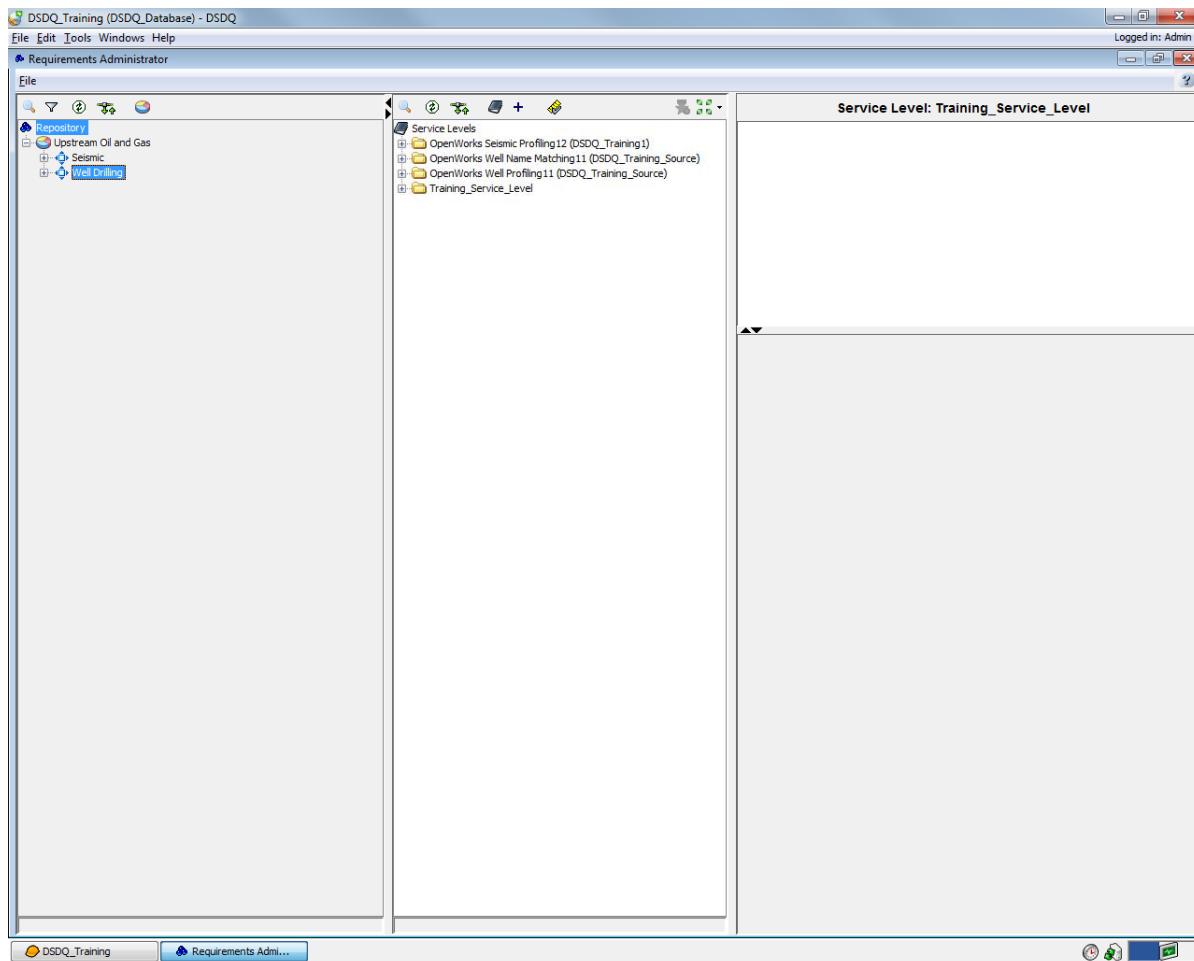
3. Select **OpenWorks Seismic Profiling 12(DSDQ\_Training1)** from the left pane and click .

The selected service level is moved to the right pane of the Service Levels dialog box.



4. Click **OK**.

The selected service level appears in the **Service Level Tree Pane**.

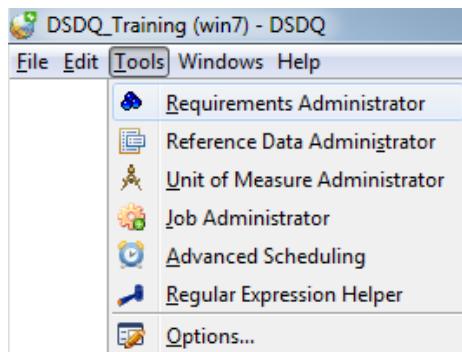


5. Optionally, you can add more than one service level by repeating steps **2** to **4**.
6. Select **File > Exit** from the menu bar on the **Requirements Administrator** window.

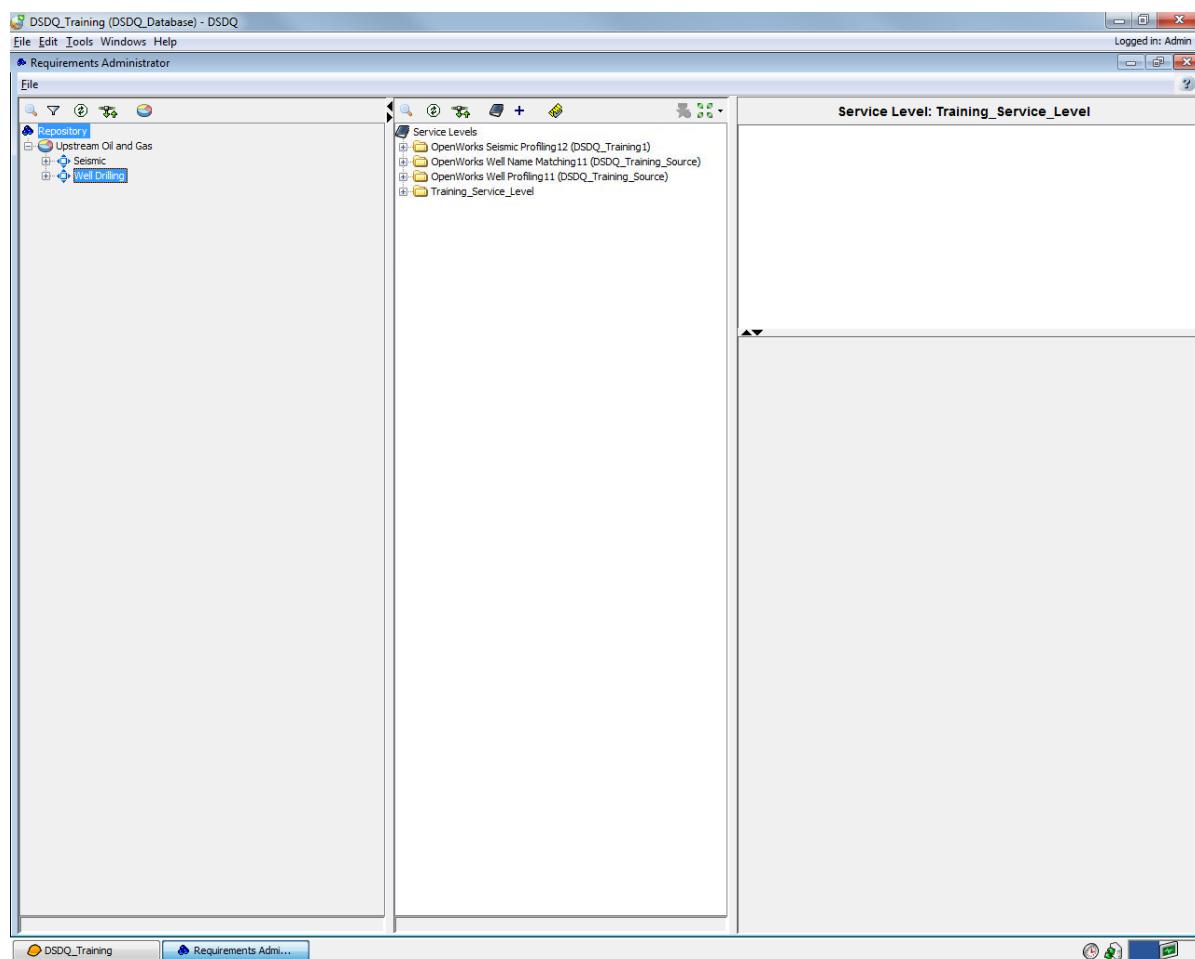
## Exercise: Editing a Service Level

To edit a service level:

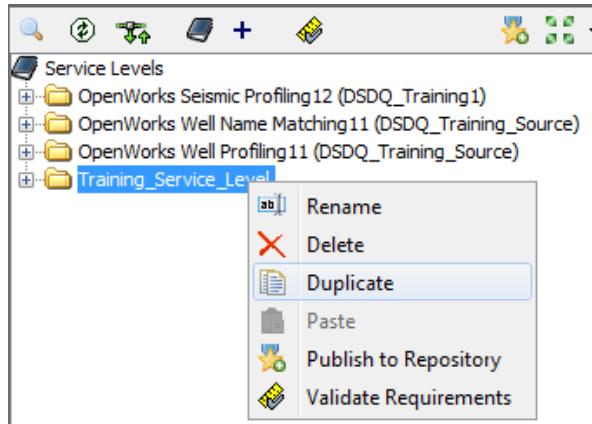
1. Select **Tools > Requirements Administrator** from the menu bar on the **DSDQ Project** window.



The **Requirements Administrator** window appears.



2. Right-click **Training\_Service\_Level** on the Service Level Tree Pane and select **Duplicate** from the pop-up menu.

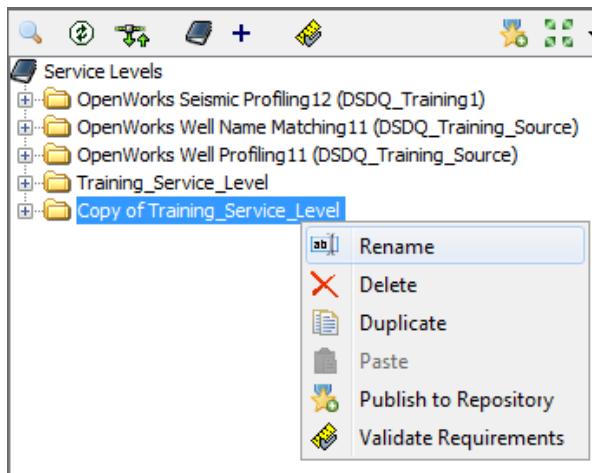


A duplicate service level is added in the Service Level Tree.

**Note**

The words “Copy of” are prefixed to the name of the duplicate service level.

3. Right-click **Copy of Training\_Service\_Level** and select **Rename** from the pop-up menu.

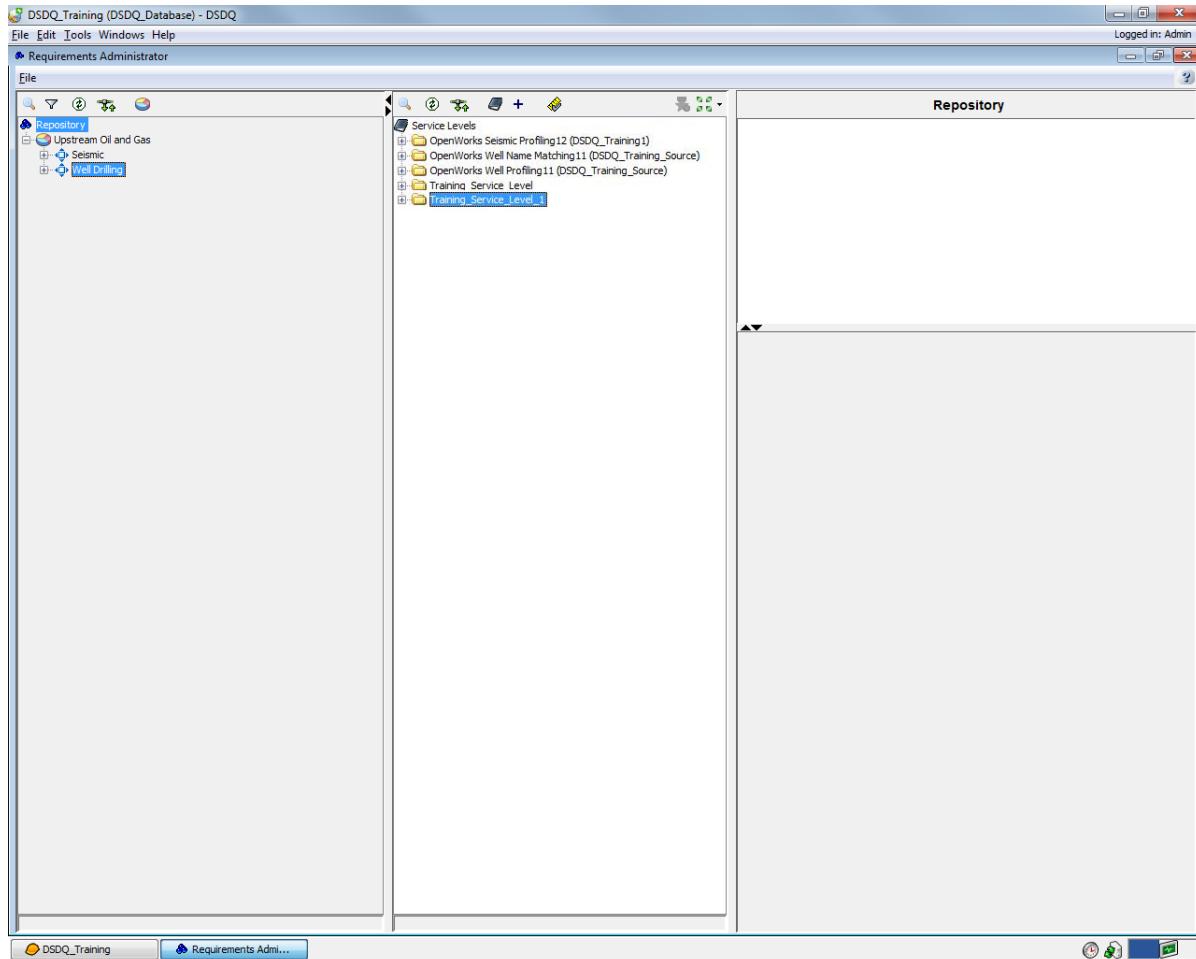


The **Rename Service Level Item** dialog box appears.



4. Enter **Training\_Service\_Level\_1** in the **Enter a new name for the service level item** field.
5. Click **OK**.

The edited service level appears in the **Service Level Tree Pane**.

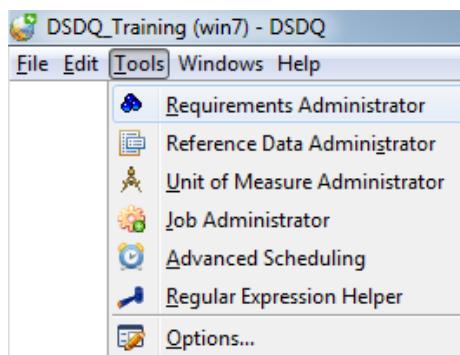


6. Select **File > Exit** from the menu bar on the **Requirements Administrator** window.

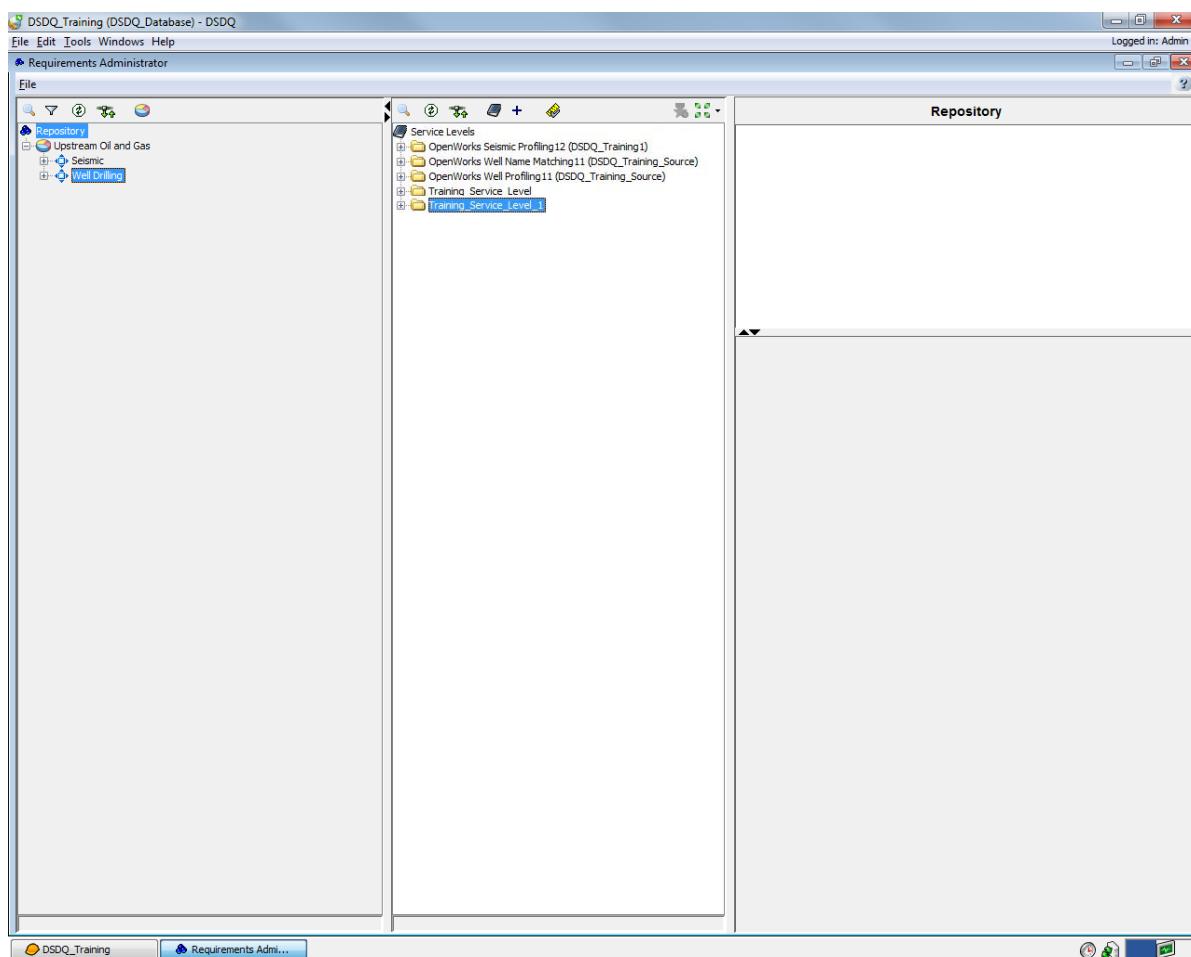
## Exercise: Deleting a Service Level

To delete a service level:

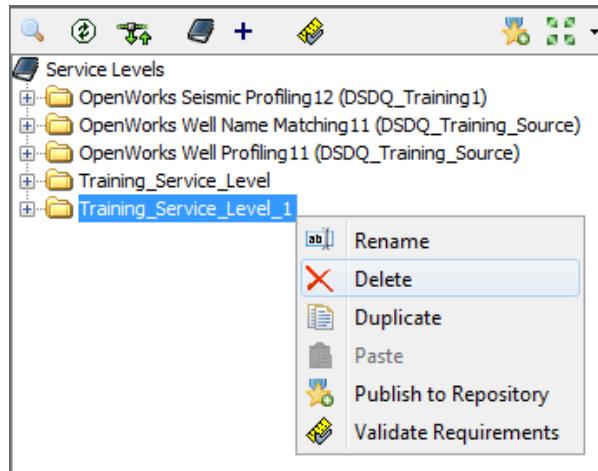
1. Select **Tools > Requirements Administrator** from the menu bar on the DSDQ Project window.



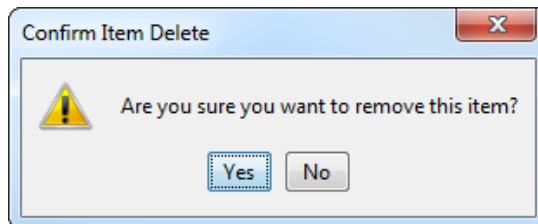
The **Requirements Administrator** window appears.



2. Right-click **Training\_Service\_Level\_1** and select **Delete** from the pop-up menu.

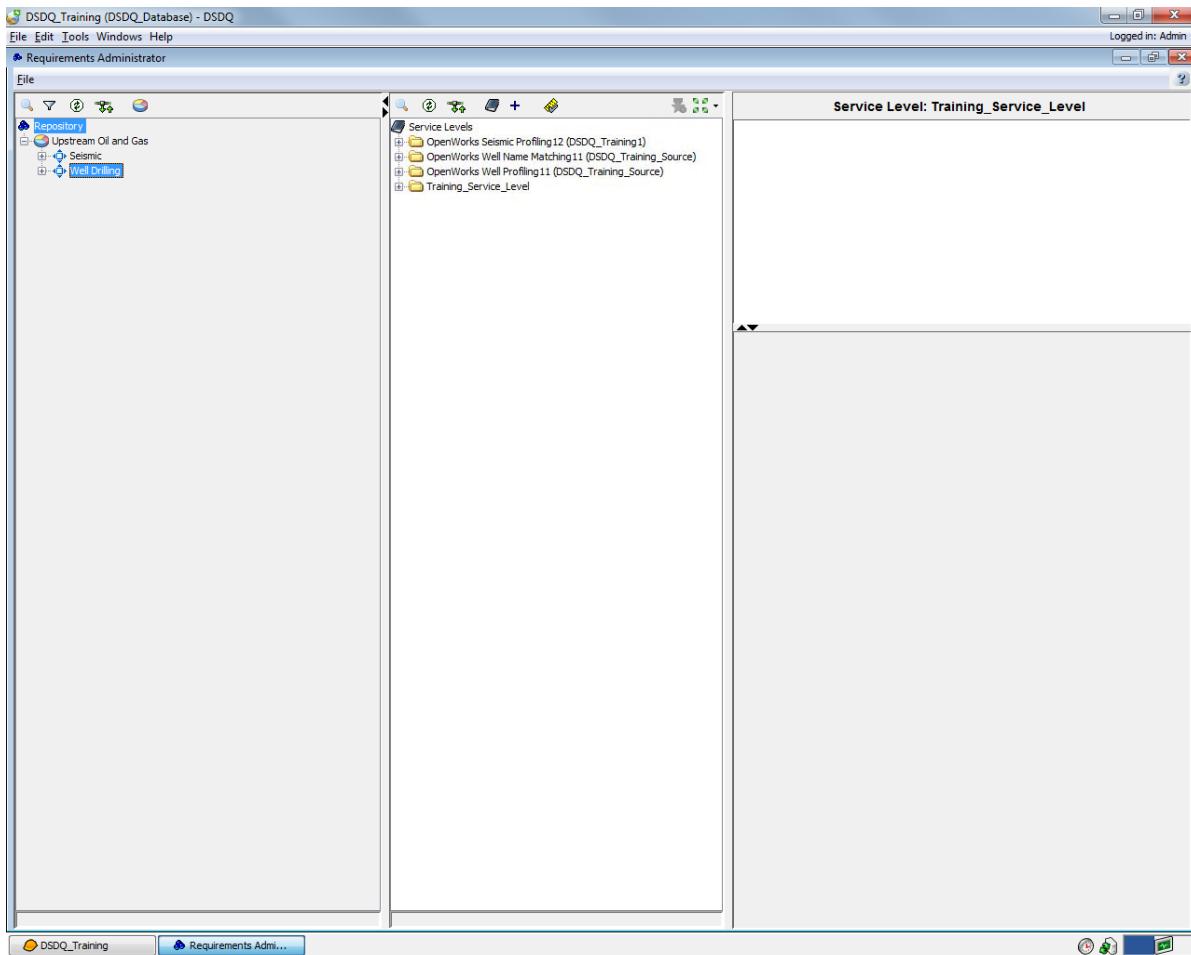


The **Confirm Item Delete** dialog box appears.



3. Click Yes.

The selected service level is deleted from the **Service Level Tree**.

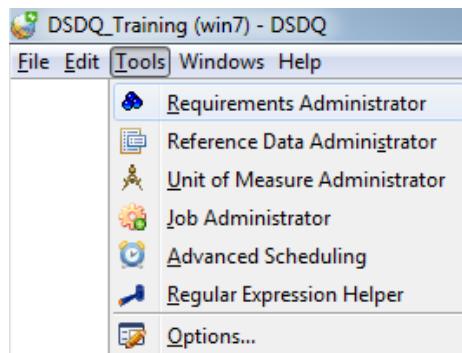


4. Select **File > Exit** from the menu bar on the **Requirements Administrator** window.

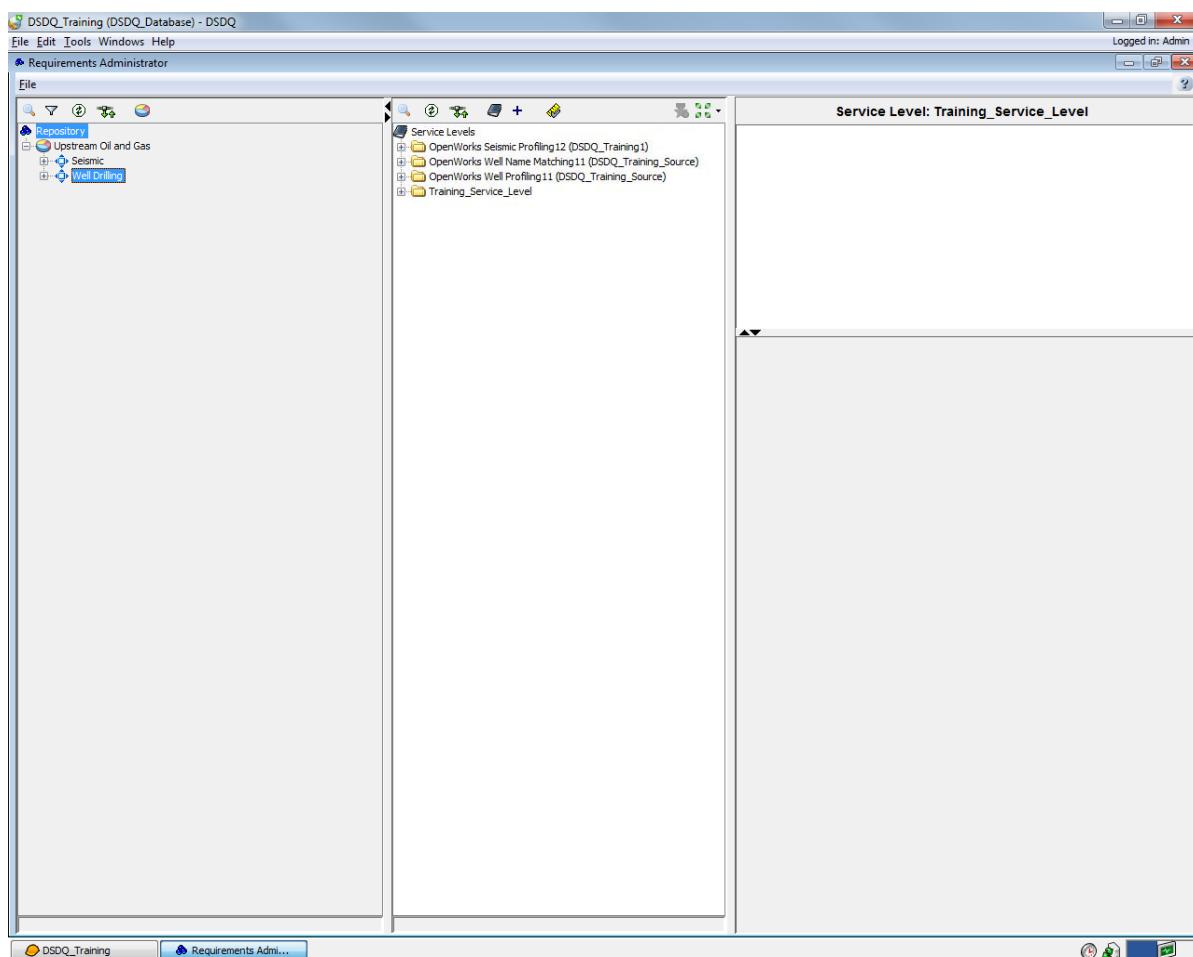
## Exercise: Exporting a Service Level

To export a service level:

1. Select **Tools > Requirements Administrator** from the menu bar on the **DSDQ Project** window.

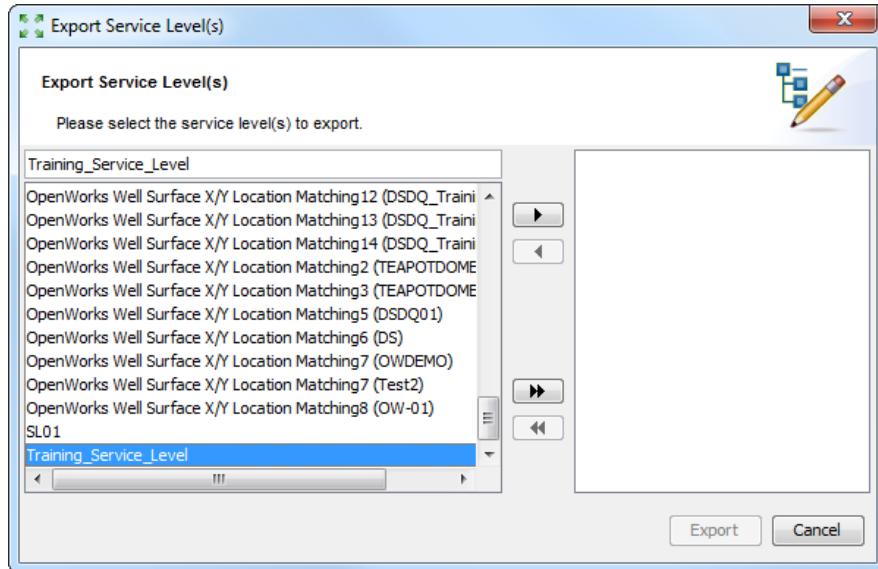


The **Requirements Administrator** window appears.

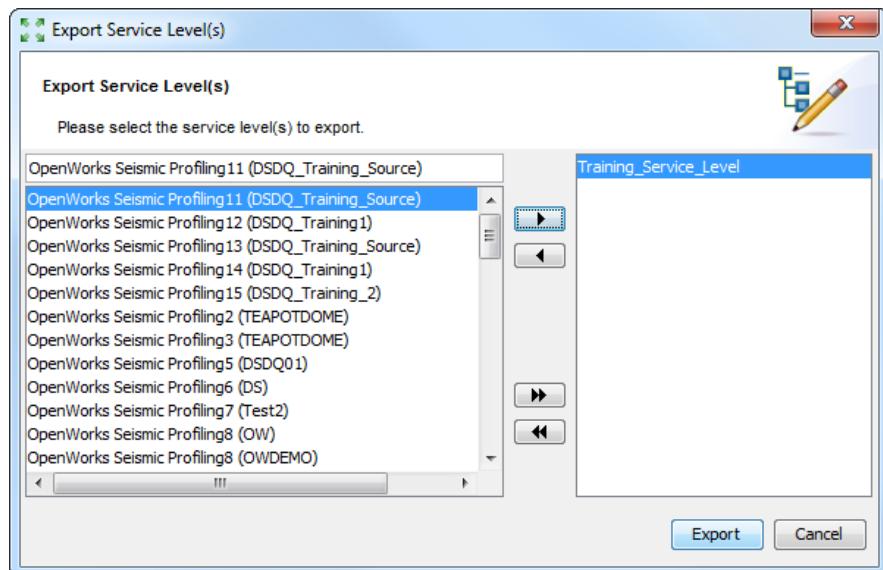


2. Click the arrow ▾ on the **Import/Export Service Levels** icon  and select the **Export Service Levels** option from the drop-down menu.

The **Export Service Level(s)** dialog box appears.

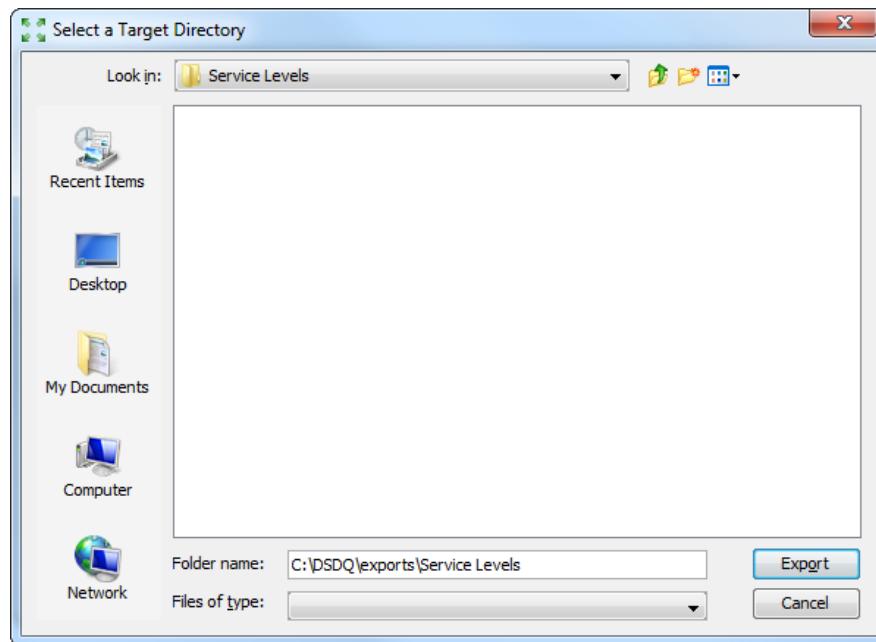


3. Select **Training\_Service\_Level**.
4. Click  to move the selected service level to the right pane of the **Export Service Level(s)** dialog box.



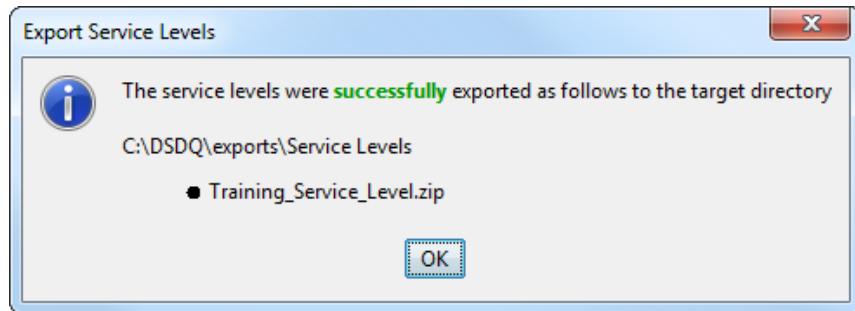
5. Click **Export**.

The **Select a Target Directory** dialog box appears.



6. Select an export path and click **Export**.

The **Export Service Levels** confirmation window appears.



7. Click **OK**.

The selected service level is exported into their own zip files at the desired location.

8. Select **File > Exit** from the menu bar on the **Requirements Administrator** window.

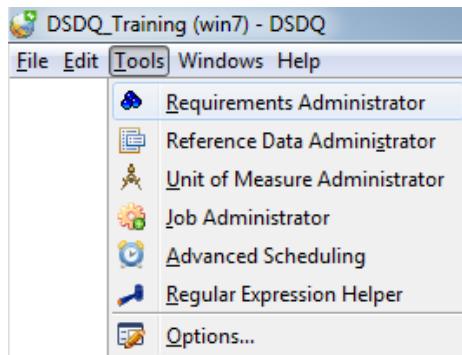
**Note**

This functionality is only accessible to users with Administrator rights.

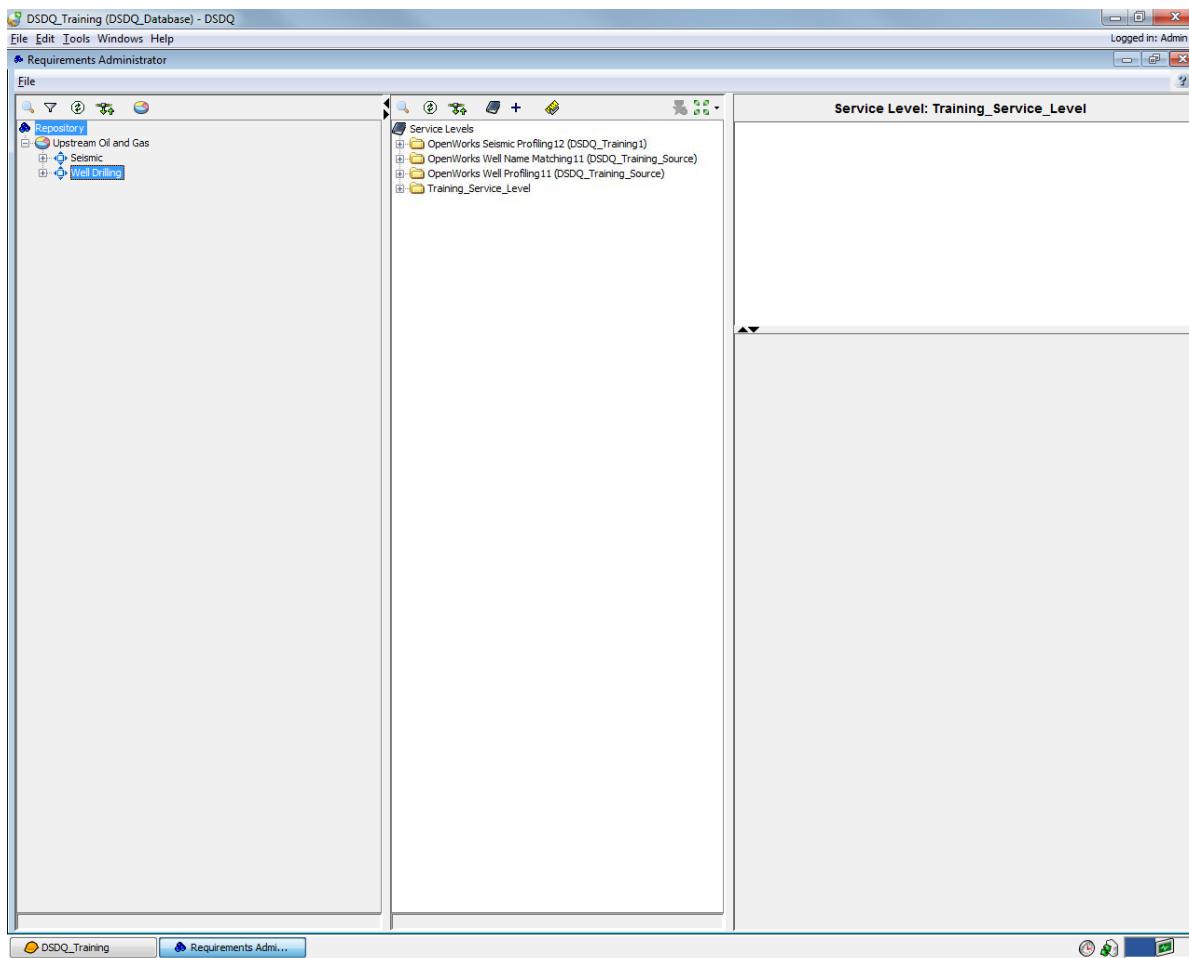
## Exercise: Importing a Service Level

To import a service level:

1. Select **Tools > Requirements Administrator** from the menu bar on the **DSDQ Project** window.

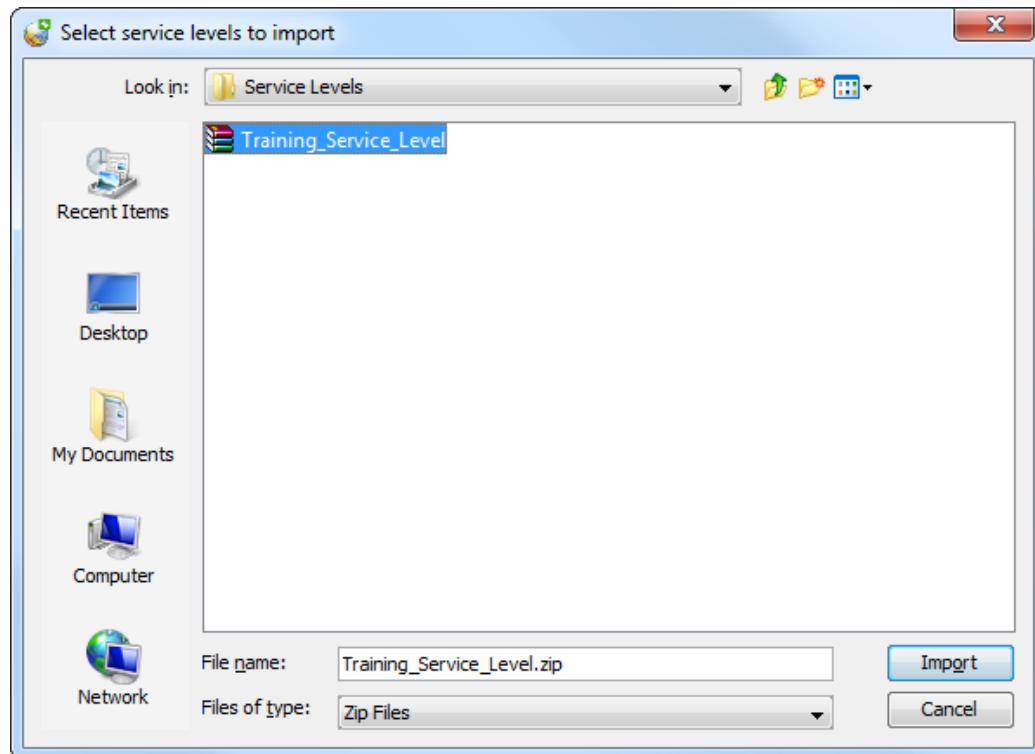


The **Requirements Administrator** window appears.

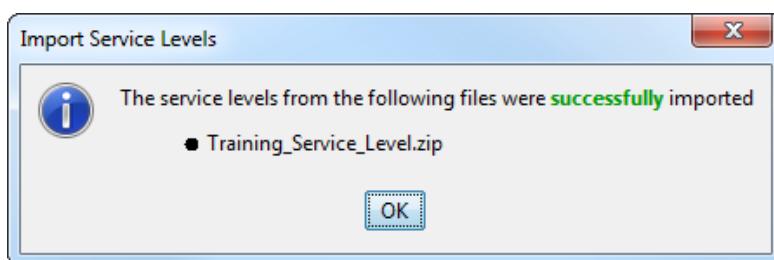


2. Click the arrow ▾ on the **Import/Export Service Levels** icon  and select the **Import Service Levels** option from the drop-down menu.

The **Select service levels to import** dialog box appears.



3. Select the **Training\_Service\_Level.zip** file and click **Import**.  
The **Import Service Levels** confirmation dialog box appears.



4. Click **OK**.

The selected service level is imported in the Service Level Tree pane.

**Note**

If the selected service level import file matches a service level that already exists in the current Data Quality install, a confirmation dialog box appears informing you that a backup of the matching service level will be created if the import operation is continued. Click **Yes** to continue or **No** to abort the Import process.

5. Select **File > Exit** from the menu bar on the **Requirements Administrator** window.

**Note**

This functionality is only accessible to users with Administrator rights.

## Managing the Repository

---

When the Requirements Administrator initially opens, no requirements or service levels are displayed in the Repository or Service Level Trees. To select the Master Repository, click on the **Repository Tree** Pane of the Requirements Administrator window. The Repository Tree is activated (background whitened) and the **Service Level Tree** Pane background is grayed out.

### **Components of Repository**

The components that make up the Repository Tree are:

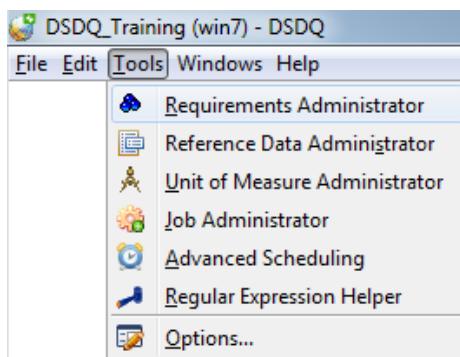
- **Sector**  – e.g.: Upstream Oil and Gas  
A sector can be identified as an industry. For example, Finance or Utilities.
- **Area**  – e.g.: Well Drilling  
An area is a smaller division within the sector. For example, a Financial Sector can have Financial Accounting and Production Accounting areas, as they have different requirements.
- **Element Group**  – e.g.: Well  
An element group is another sub-division of requirements. For example, a Financial Accounting area can have Sales and Purchases element groups.
- **Element**  – e.g.: Country  
An element is associated to a particular database column. For example, the Purchase Date element has all the requirements to validate the purchase date in the database.
- **HealthCheck Requirement**  - e.g. Greater than 2000
- **Clean Requirement**  - e.g. Convert to Lower Case
- **Match Requirement**  - e.g. Is Exactly
- **Master Registry HealthCheck Requirement**  - e.g. Equivalency Check

A Requirement carries out a data check. For example, a Null Check requirement contained within the Purchase Date element would check if all values are populated in the purchase date field of the database.

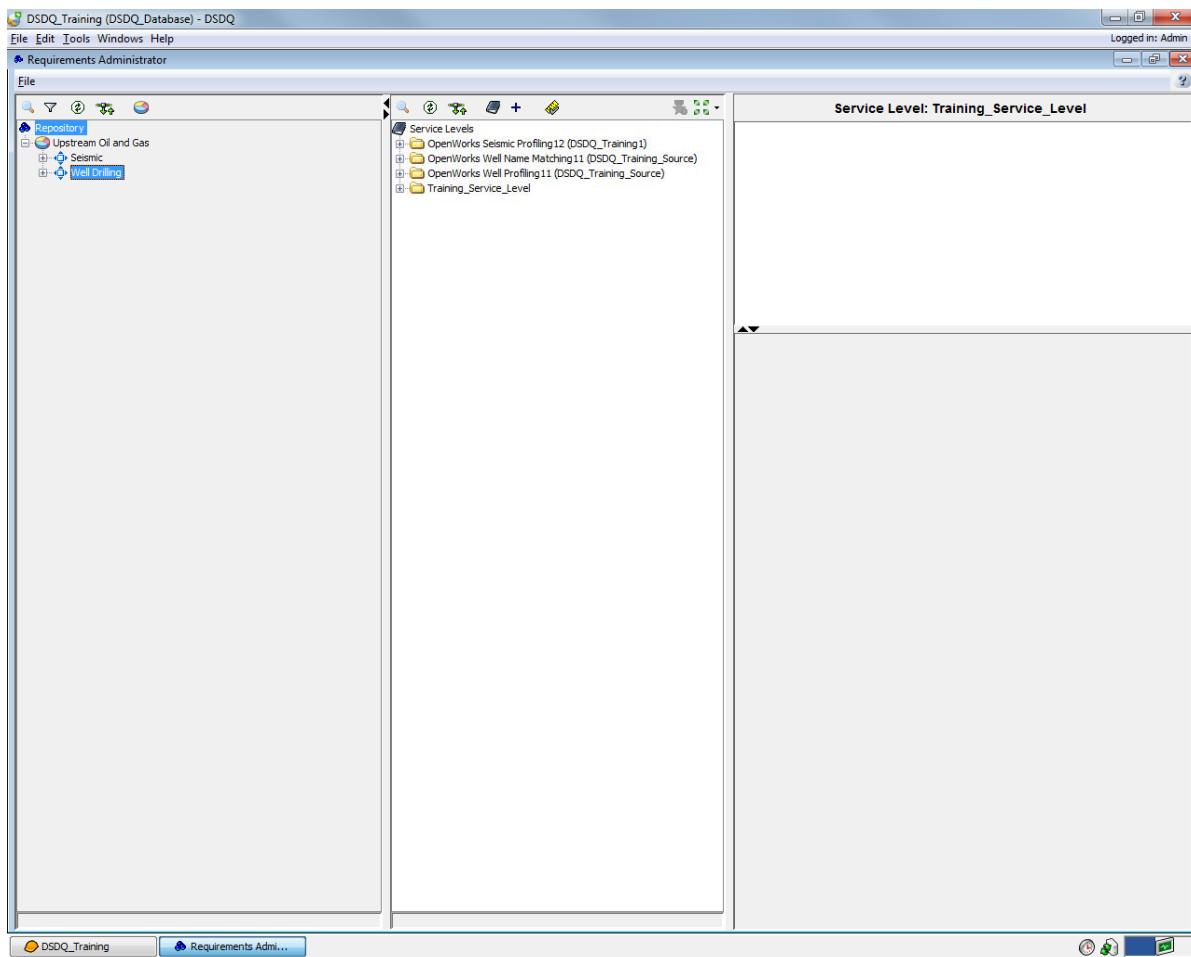
### ***Exercise: Adding a Component***

To add a component:

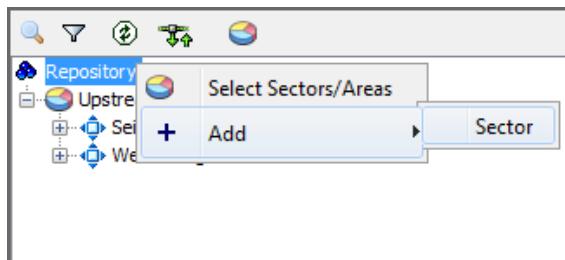
1. Select **Tools > Requirements Administrator** from the menu bar on the **DSDQ Project** window.



The Requirements Administrator window appears.



2. Right-click **Repository** on the **Repository Tree** Pane and select **Add > Sector** from the pop-up menu.

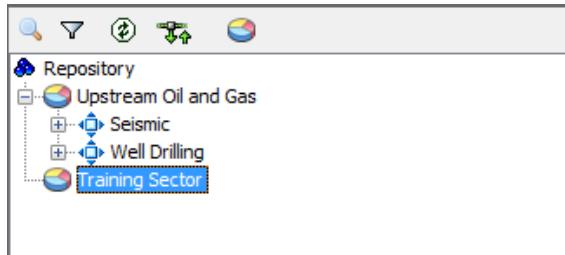


The **Enter Name** dialog box appears.

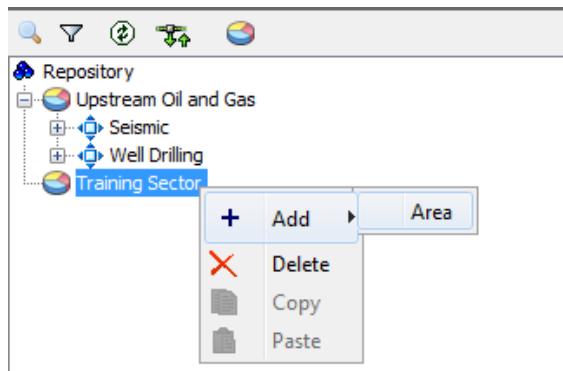


3. Enter **Training Sector** in the **Please enter a name for the new item** field.
4. Click **OK**.

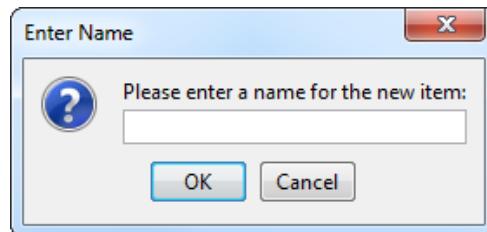
The sector is added and displays in the Repository Tree.



5. Right-click **Training Sector** on the **Repository Tree** Pane and select **Add > Area** from the pop-up menu.

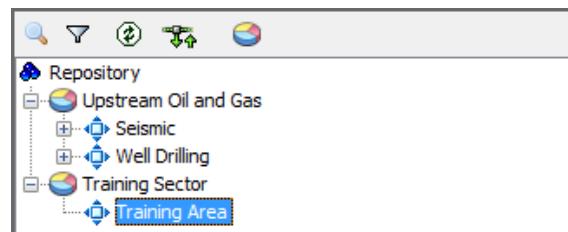


The **Enter Name** dialog box appears.

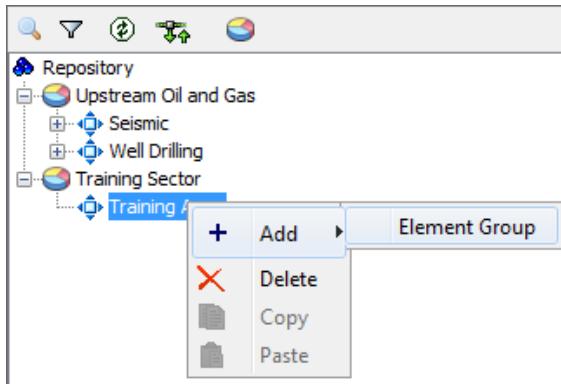


6. Enter **Training Area** in the **Please enter a name for the new item** field.
7. Click **OK**.

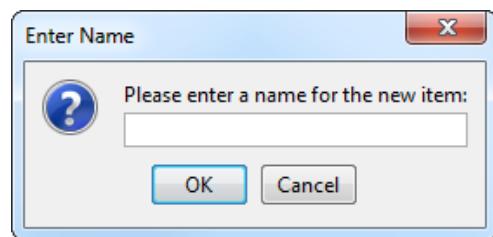
The area is added and displays in the Repository Tree.



8. Right-click **Training Area** on the Repository Tree Pane and select **Add > Element Group** from the pop-up menu.

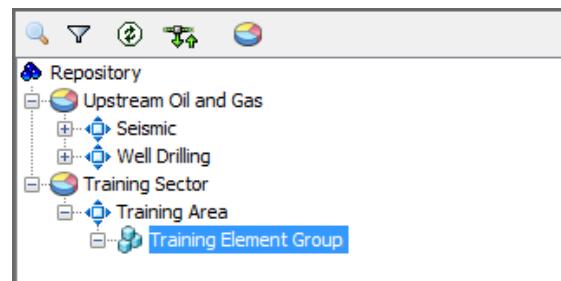


The **Enter Name** dialog box appears.

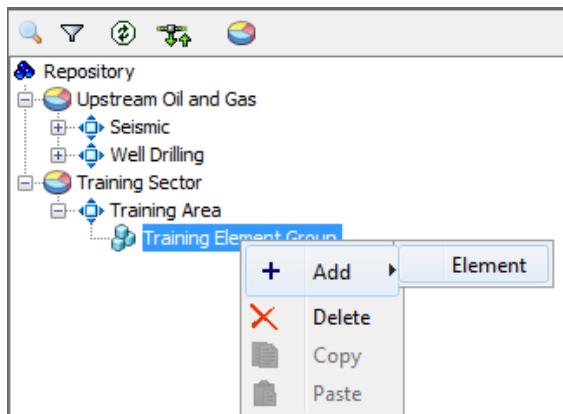


9. Enter **Training Element Group** in the **Please enter a name for the new item** field.
10. Click **OK**.

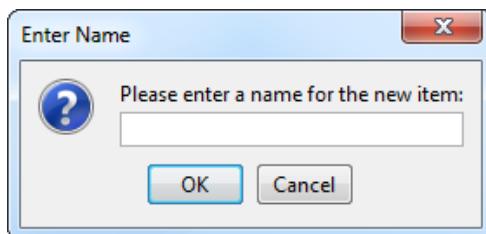
The element group is added and displays in the Repository Tree.



11. Right-click **Training Element Group** on the **Repository Tree** Pane and select **Add > Element** from the pop-up menu.



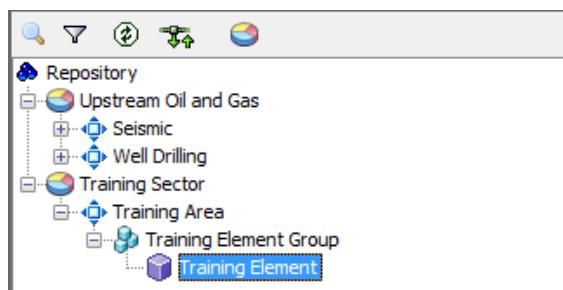
The **Enter Name** dialog box appears.



12. Enter **Training Element** in the **Please enter a name for the new item** field.

13. Click **OK**.

The element is added and displays in the Repository Tree.

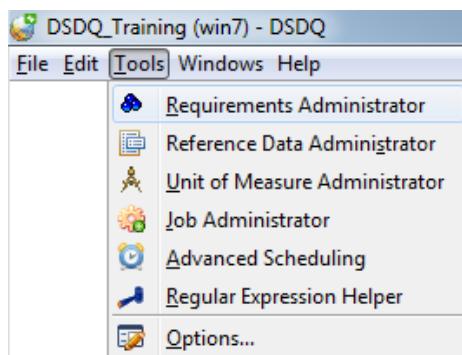


14. Select **File > Exit** from the menu bar on the **Requirements Administrator** window.

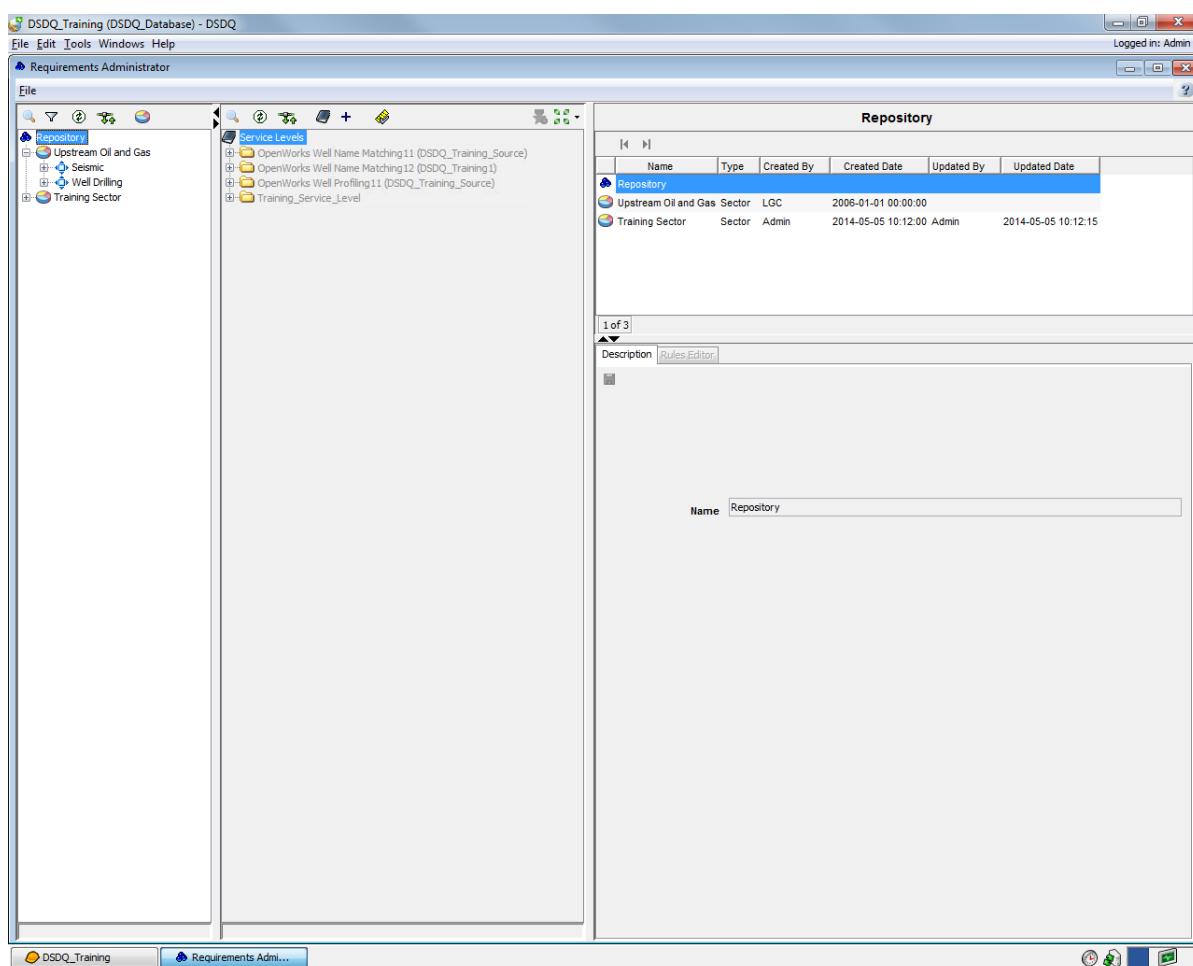
## Exercise: Editing a Component

To edit a component:

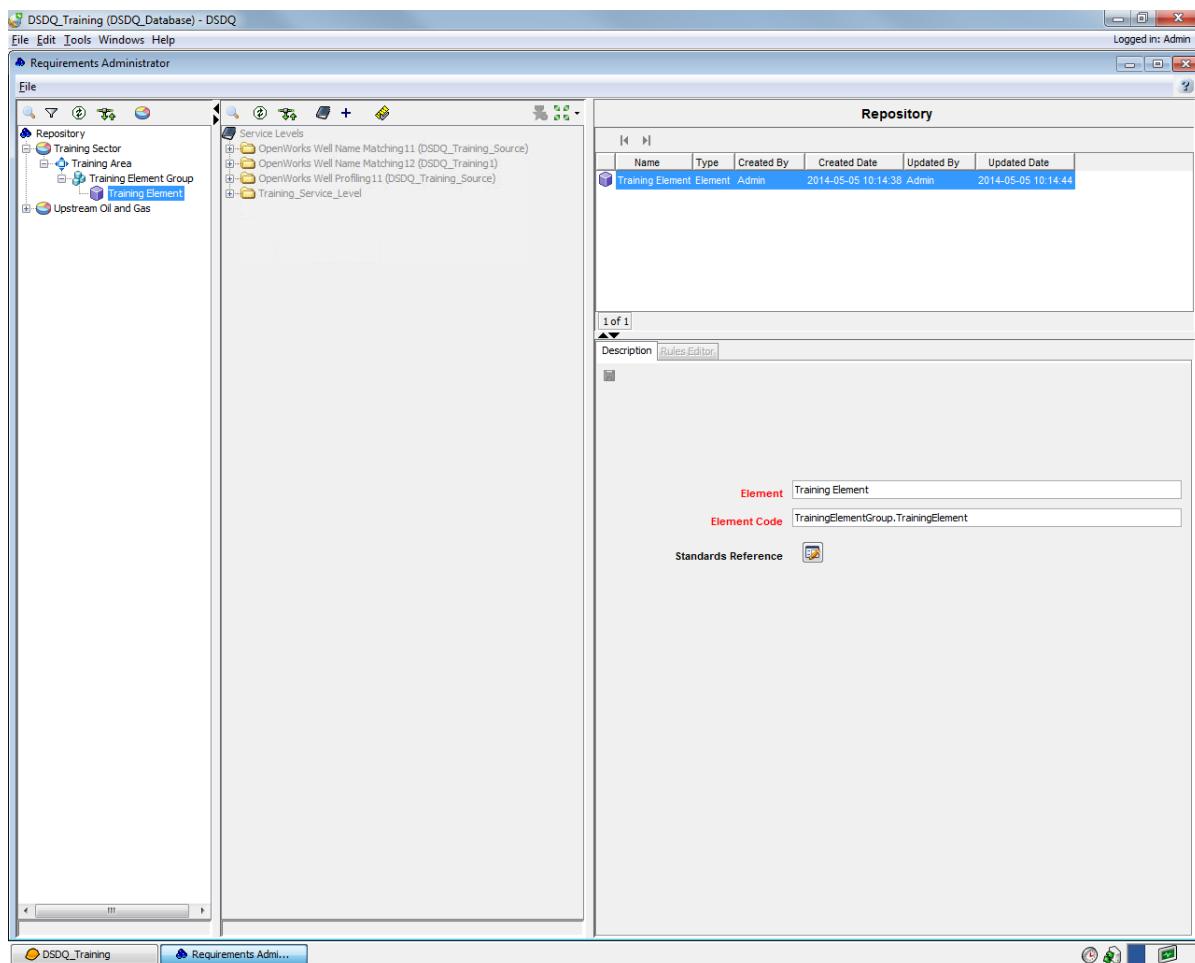
1. Select **Tools > Requirements Administrator** from the menu bar on the **DSDQ Project** window.



The **Requirements Administrator** window appears.



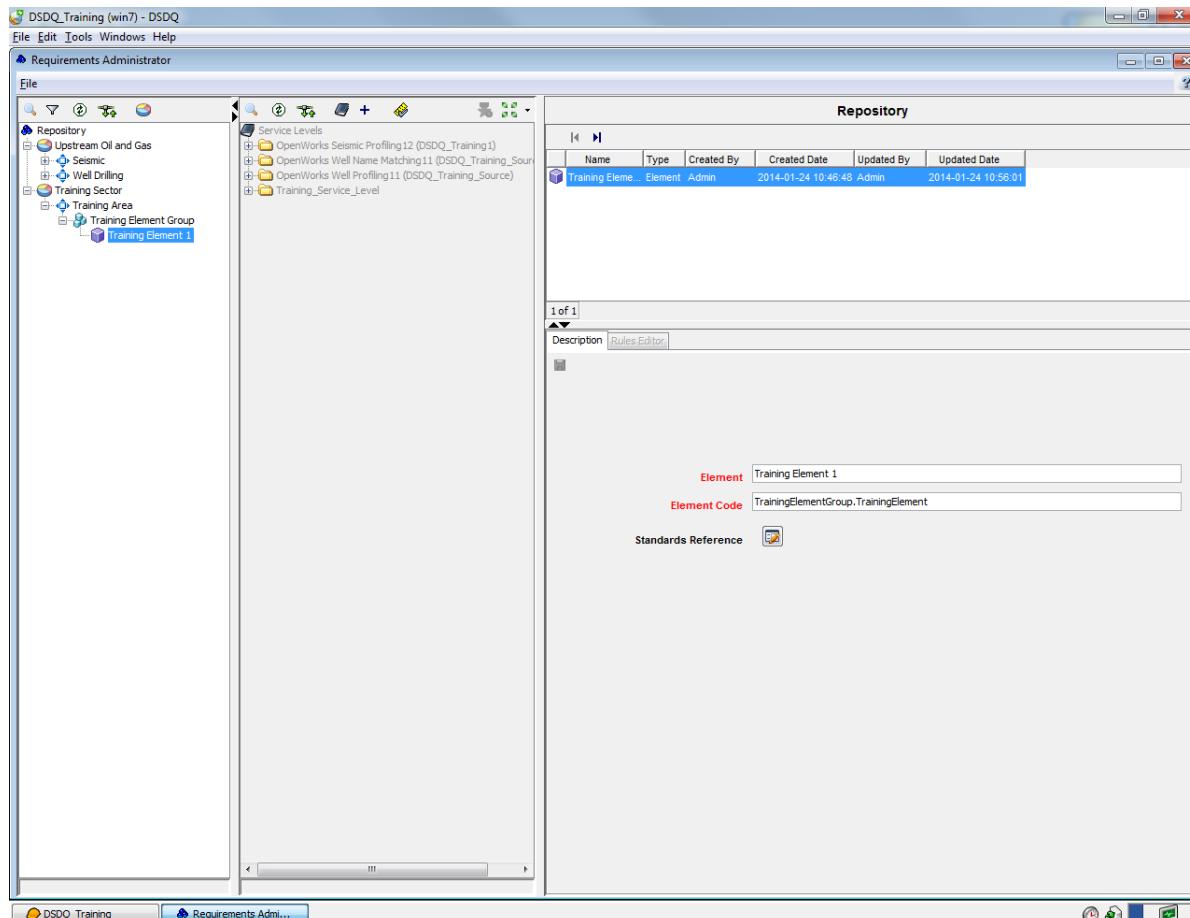
2. Click to expand **Training Sector** in the **Repository Tree**.
3. Click to expand **Training Area**.
4. Click to expand **Training Element Group**.
5. Select **Training Element** in the **Repository Tree**.  
Information about **Training Element** appears in the **Description** tab of the Details Pane.



6. Enter **Training Element 1** in the **Element** field of the **Description** tab.
7. Do not change the value of the **Element Code** field.

8. Click .

The updated element displays in the Repository Tree Pane.



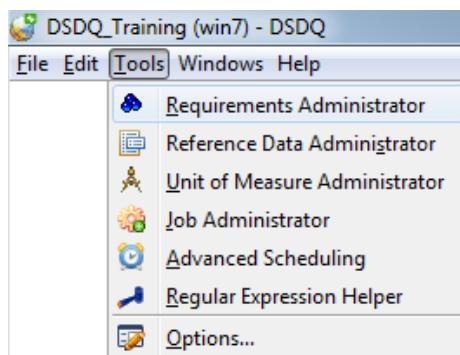
**Note**

To duplicate an element, right-click on the desired element and select **Copy** from the pop-up menu. Right-click on the desired area and select **Paste** from the pop-up menu to **copy** the element to the desired area.

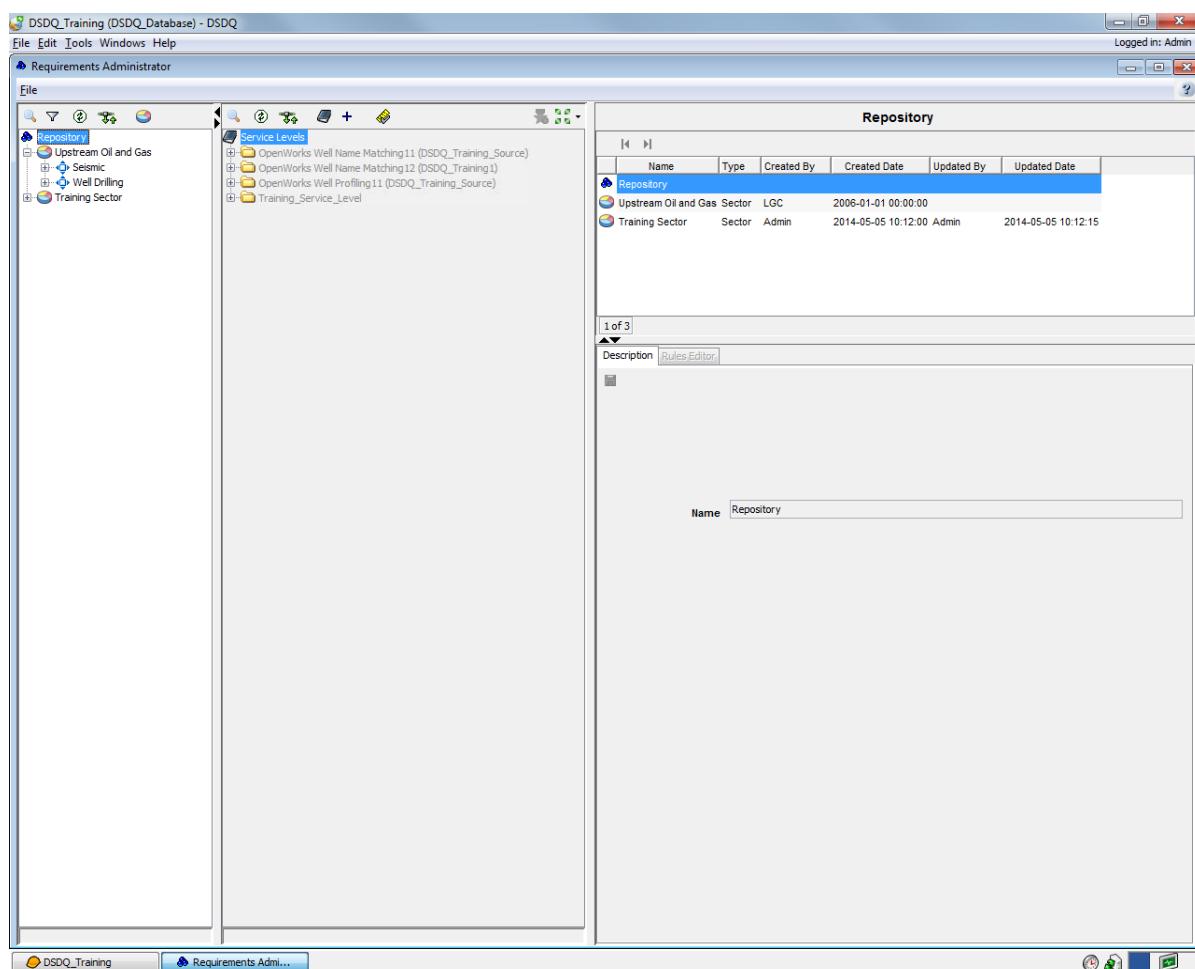
## Exercise: Deleting a Component

To delete a component:

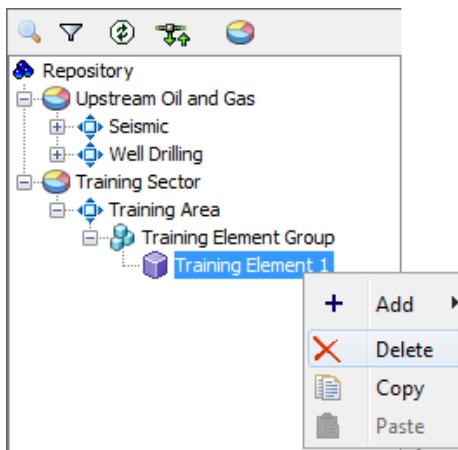
1. Select **Tools > Requirements Administrator** from the menu bar on the **DSDQ Project** window.



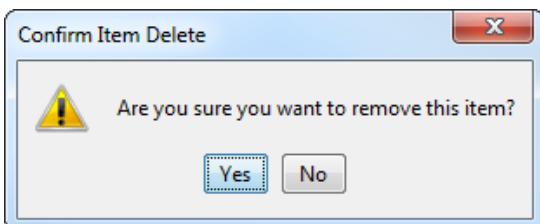
The **Requirements Administrator** window appears.



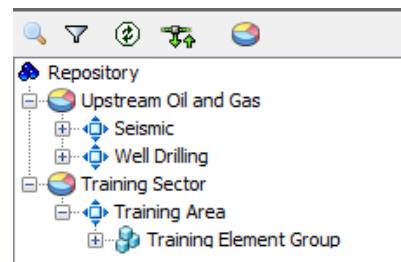
2. Click to expand **Training Sector** in the **Repository Tree**.
3. Click to expand **Training Area**.
4. Click to expand **Training Element Group**.
5. Right-click **Training Element** and select **Delete** from the pop-up menu.



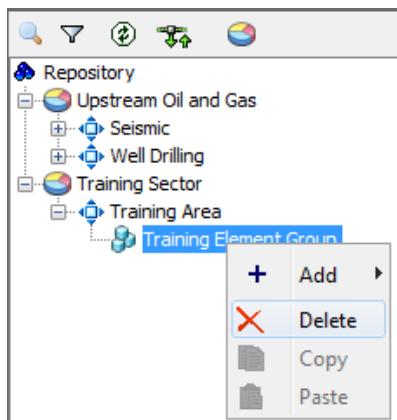
The **Confirm Item Delete** dialog box appears.



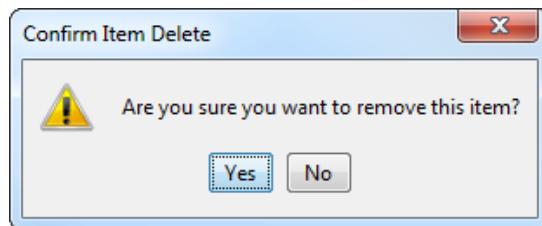
6. Click **Yes**.  
The selected element is deleted.



7. Right-click **Training Element Group** and select **Delete** from the pop-up menu.

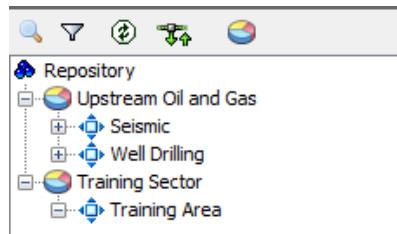


The **Confirm Item Delete** dialog box appears.

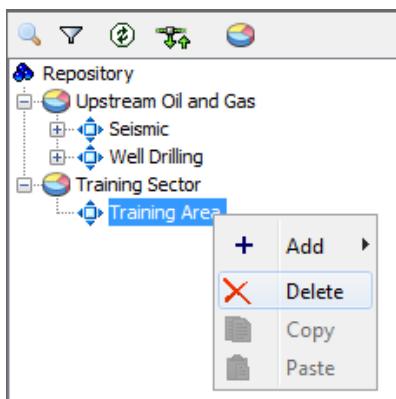


8. Click **Yes**.

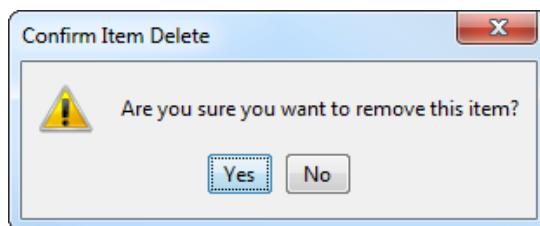
The selected element group is deleted.



9. Right-click **Training Area** and select **Delete** from the pop-up menu.

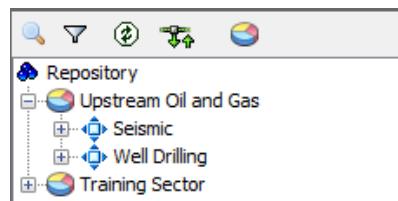


The **Confirm Item Delete** dialog box appears.

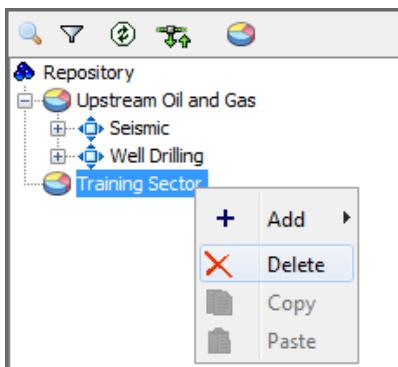


10. Click **Yes**.

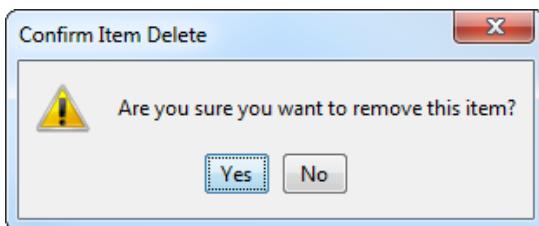
The selected area is deleted.



11. Right-click **Training Sector** and select **Delete** from the pop-up menu.

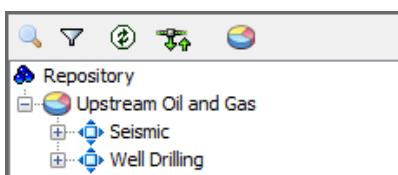


The **Confirm Item Delete** dialog box appears.



12. Click **Yes**.

The selected sector is deleted.

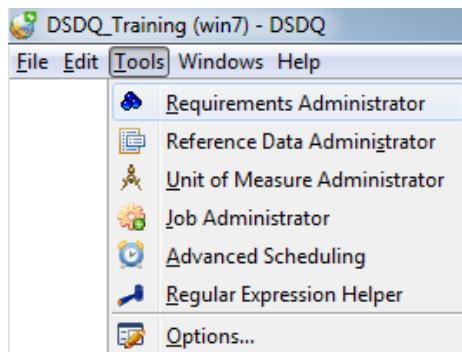


13. Select **File > Exit** from the menu bar on the **Requirements Administrator** window.

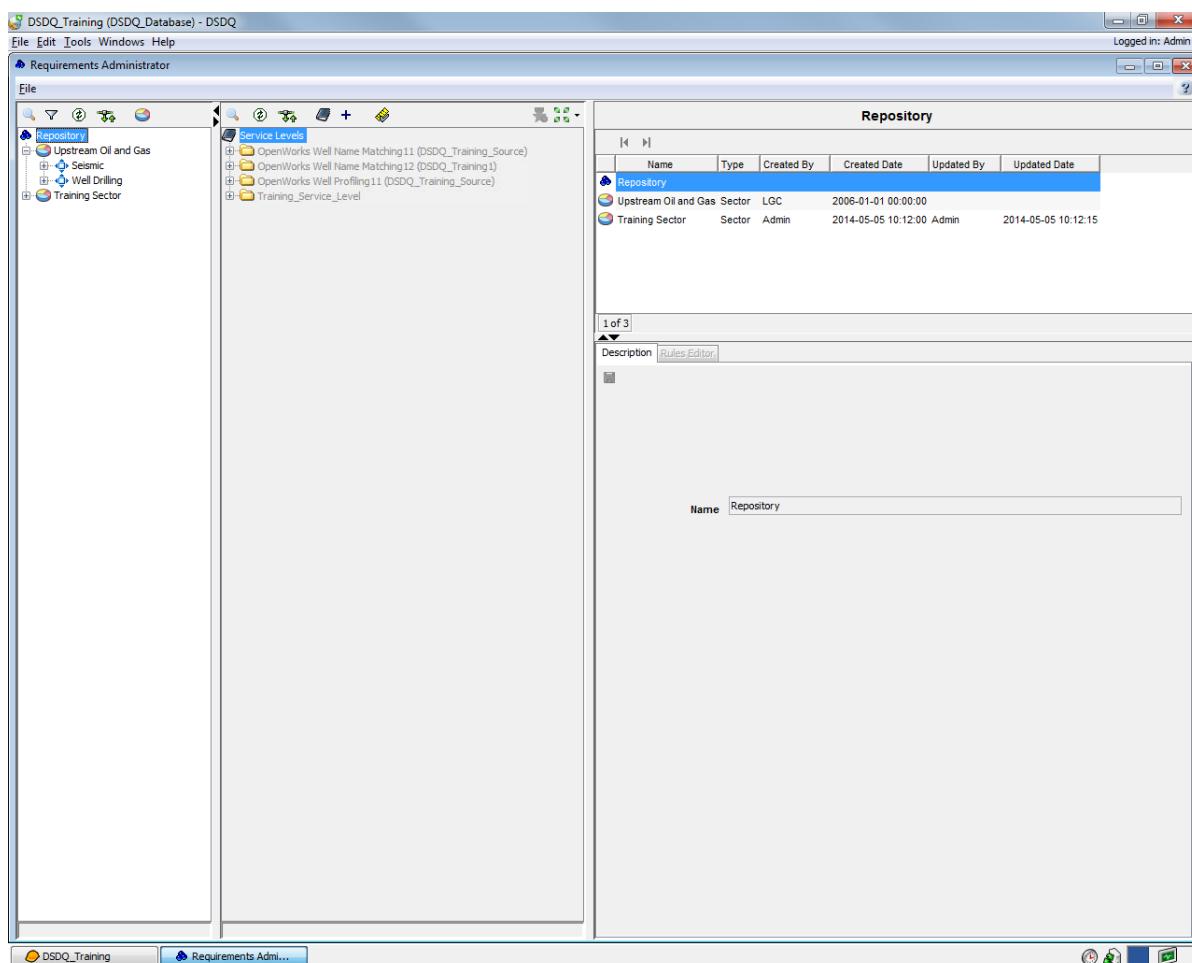
## **Exercise: Associating an Element with a Standards Reference Table**

To associate an element with a Standards Reference Table:

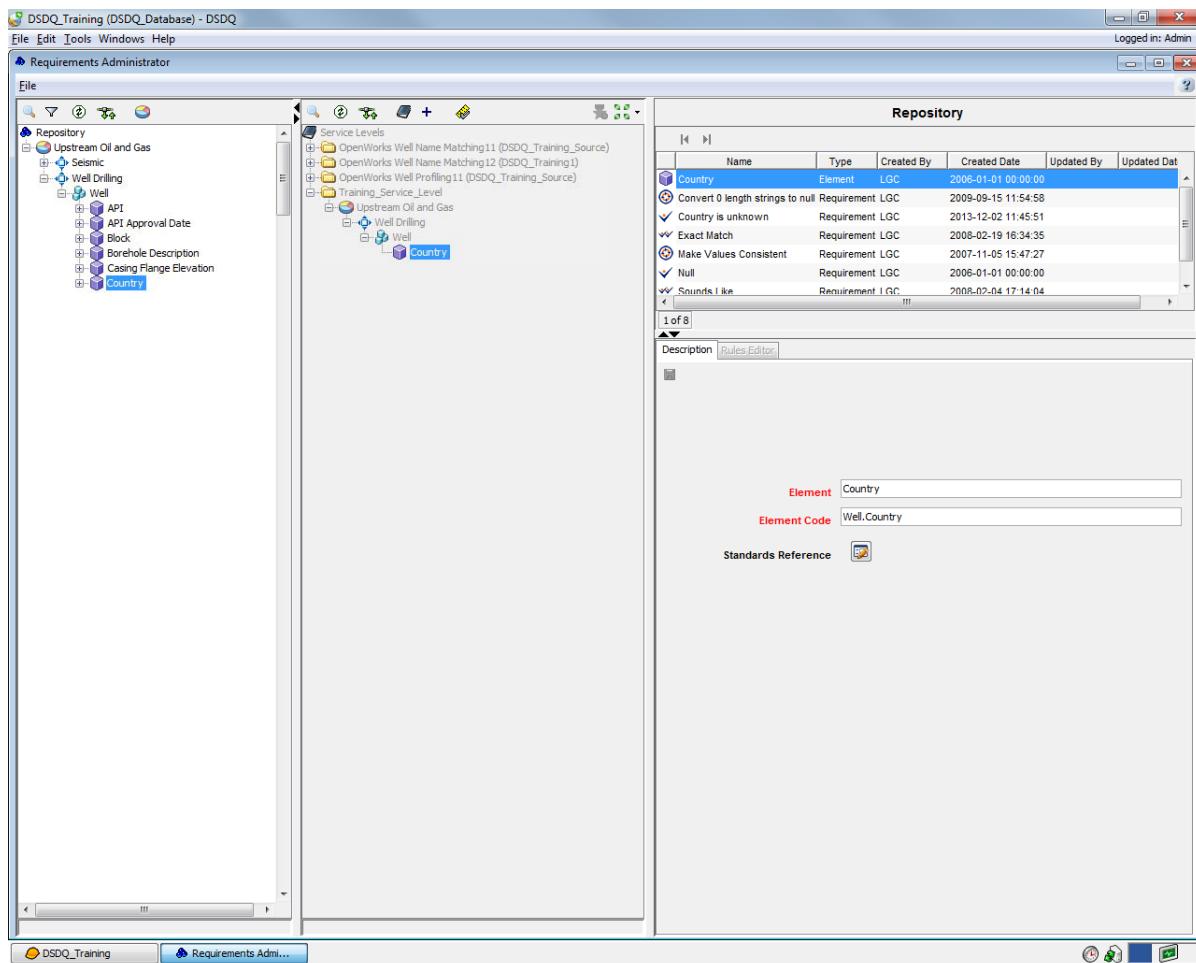
1. Select **Tools > Requirements Administrator** from the menu bar on the **DSDQ Project** window.



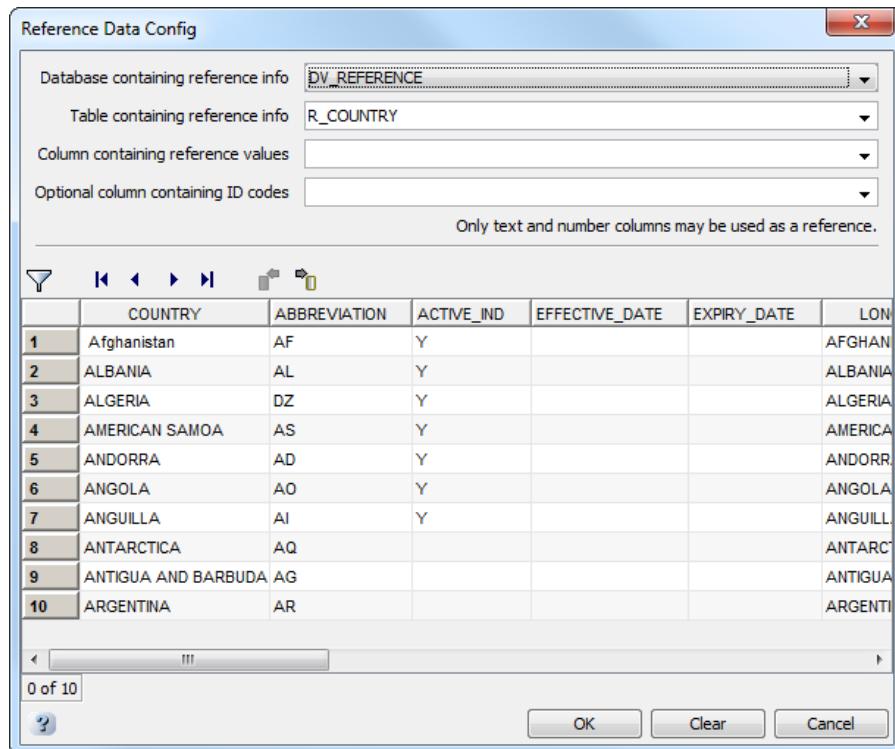
The **Requirements Administrator** window appears.



2. Click to expand the **Well Drilling** area in the **Repository Tree**.
3. Click to expand the **Well Element** group.
4. Select the **Country** element in the **Repository Tree**.  
Information about the **Country** element appears in the **Description** tab of the Details Pane.



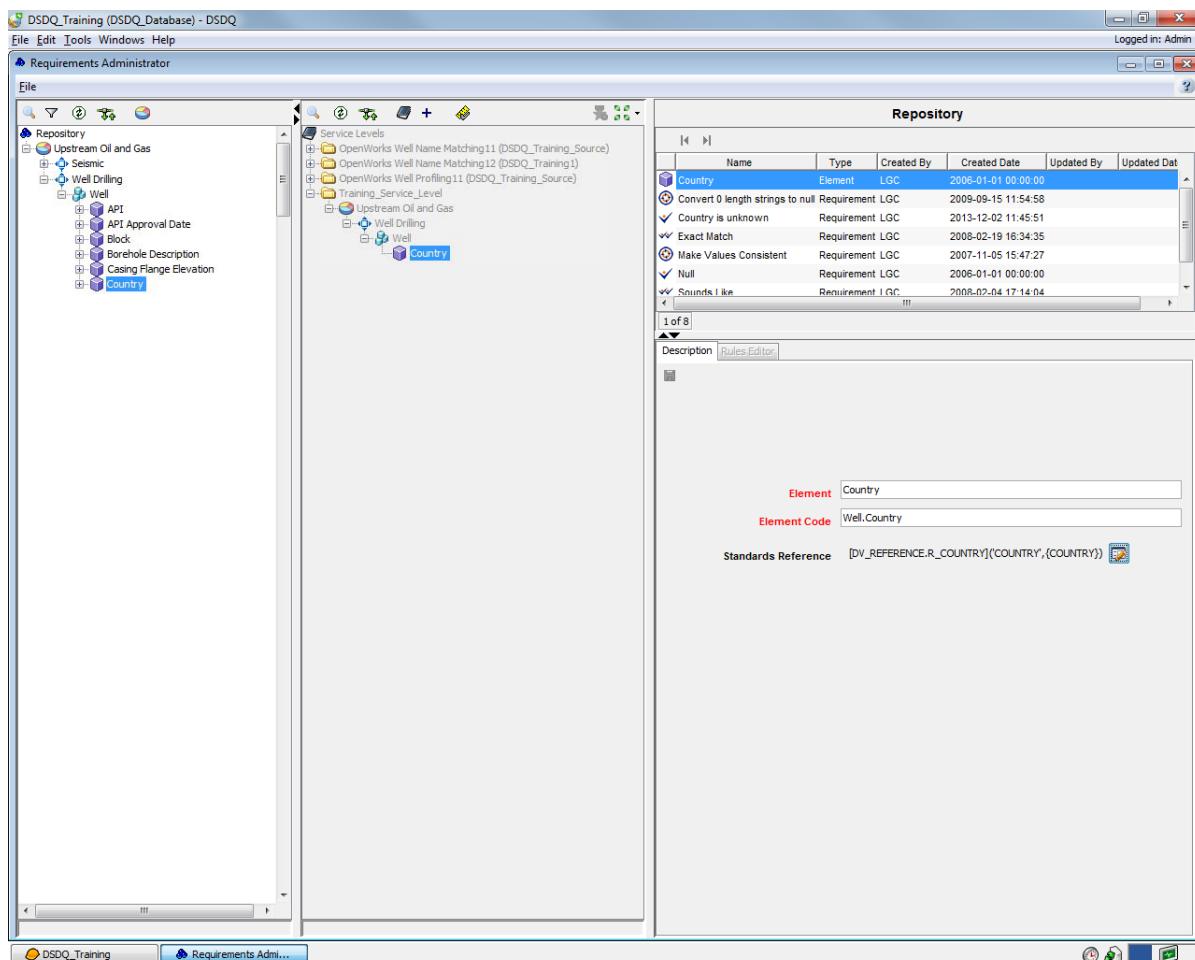
5. Click the **Standard Reference**  button in the **Description** tab.  
The **Reference Data Config** window appears.



6. Select **DV\_Reference** from the **Database containing reference info** drop-down list.
7. Select **R\_COUNTRY** from the **Table containing reference info** drop-down list.
8. Select **COUNTRY** from the **Column containing reference values** drop-down list.
9. Enter **Country** in the **Option column containing ID codes** field.

**10. Click **OK**.**

The selected element is associated with a **Selected Reference Table** and displays in the **Description** tab of the **Detail Pane**.



**11. Click **OK**.**

**12. Select **File > Exit** from the menu bar on the Requirements Administrator window.**

## Requirements

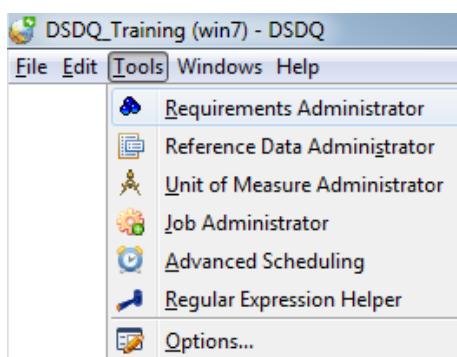
---

Requirements are defined to perform various actions on the data. Once the requirements are finalized, they can be run against a dataset. You can also add advanced requirements for **HealthCheck**, **Clean**, and **Match**.

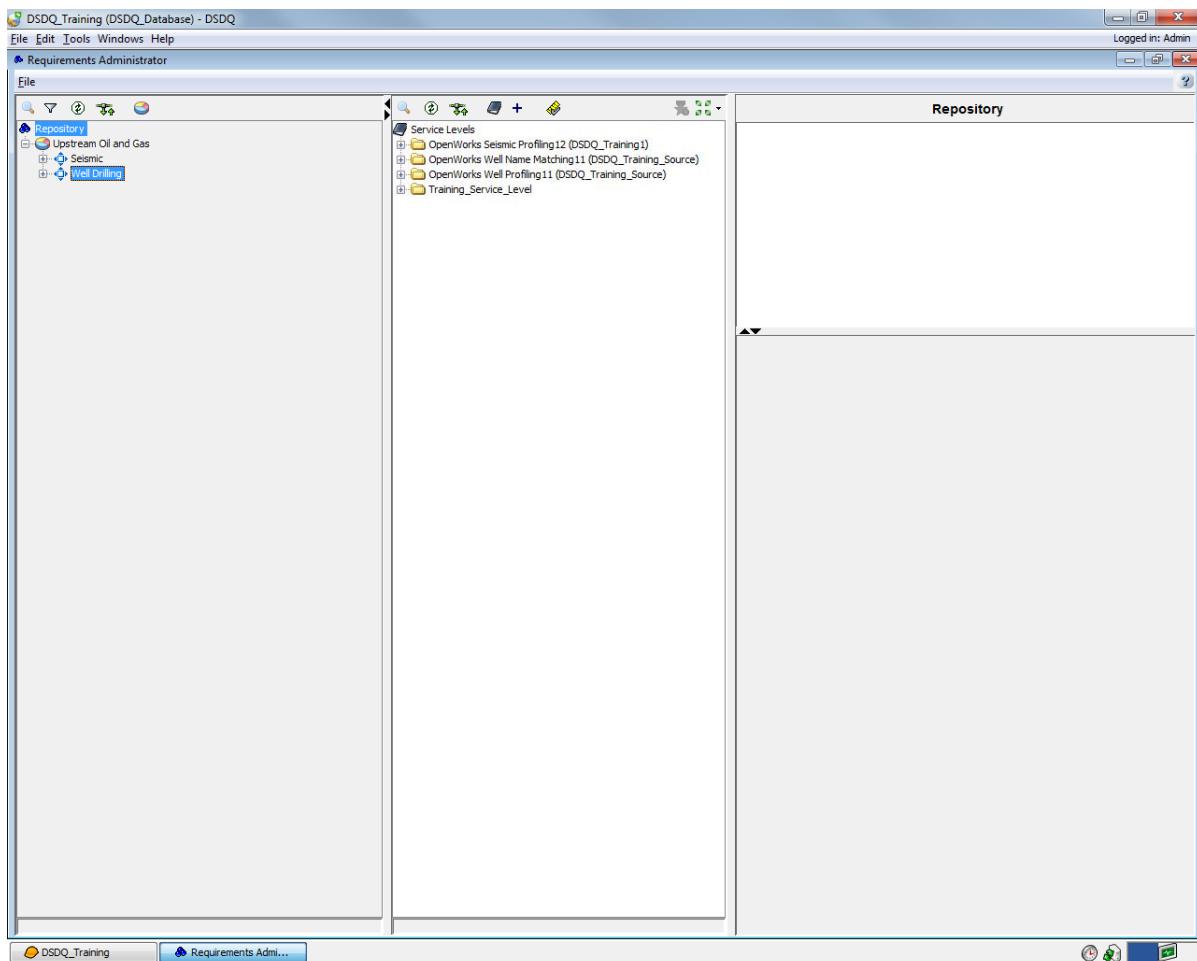
### **Exercise: Adding a Requirement**

To add a requirement:

1. Select **Tools > Requirements Administrator** from the menu bar on the **DSDQ Project** window.

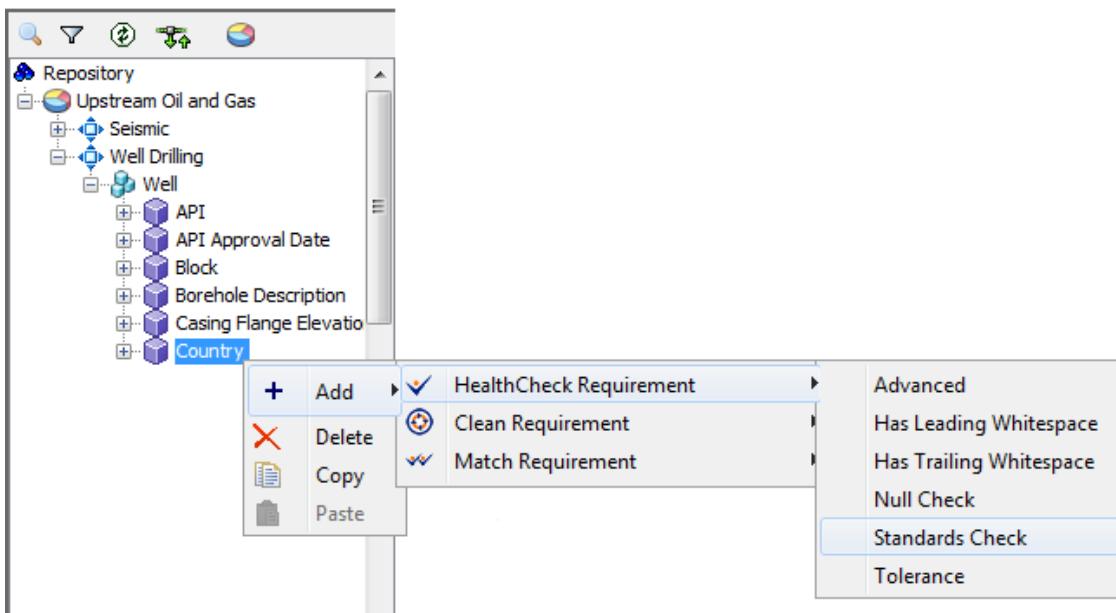


The Requirements Administrator window appears.

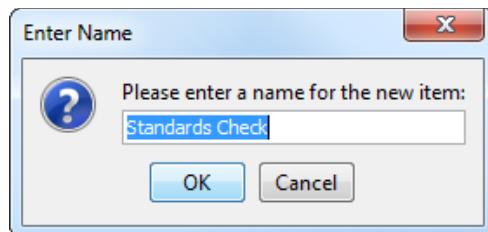


2. Click to expand the **Well Drilling** area in the **Repository Tree**.
3. Click to expand the **Well** element group.
4. Right-click the **Country** element in the **Repository Tree** and select **Add > HealthCheck Requirement > Standards Check** from the

pop-up menu.



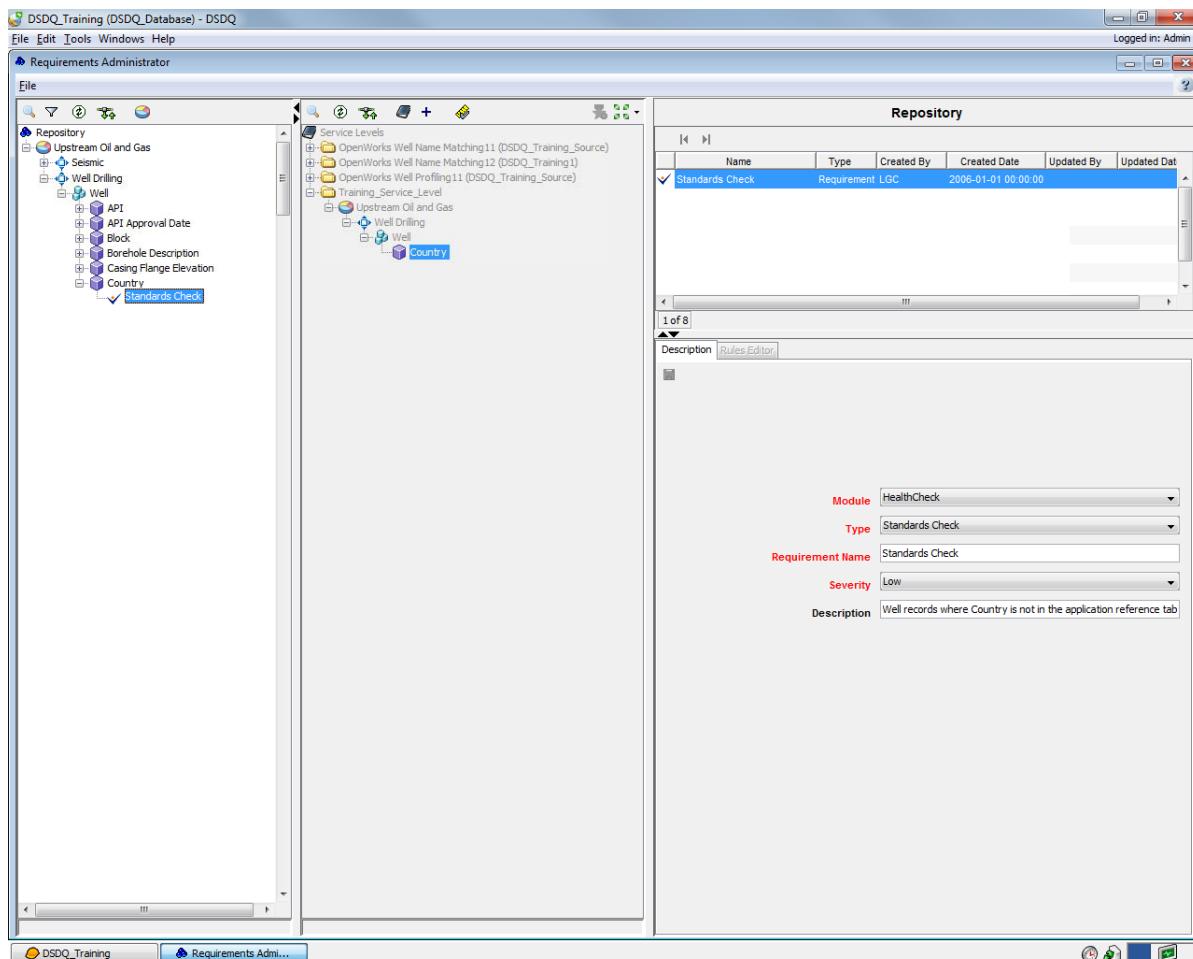
The Enter Name dialog box appears.



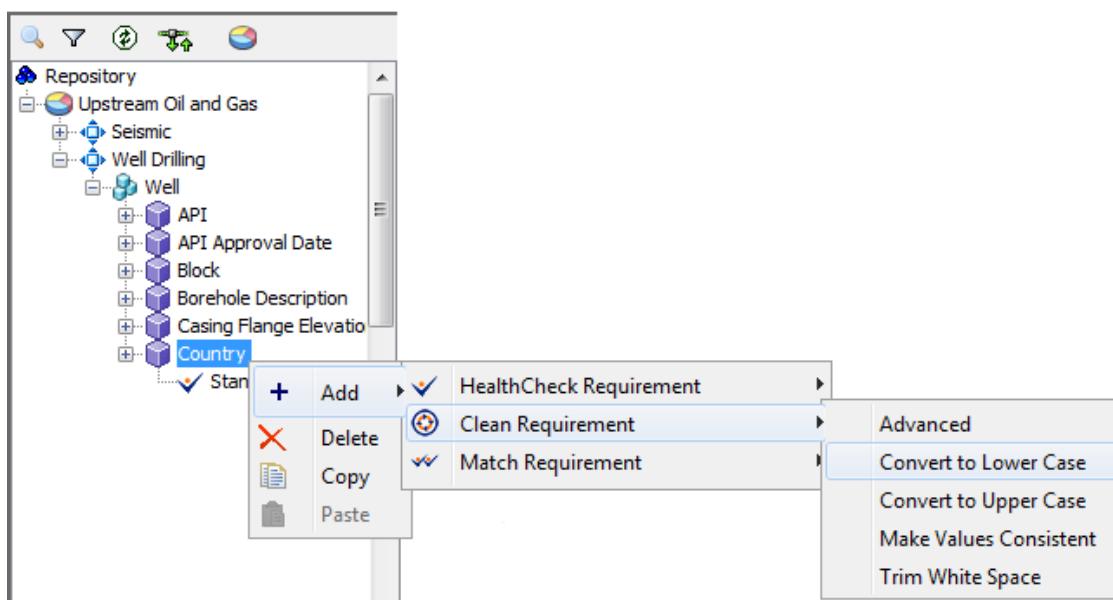
5. Optionally, specify a user-defined name for the requirement.

6. Click **OK**.

The selected requirement is added to the **Country** element and displays in the **Repository Tree**.



7. Right-click the **Country** element in the **Repository Tree** and select **Add > Clean Requirement > Convert to Lower Case** from the pop-up menu.



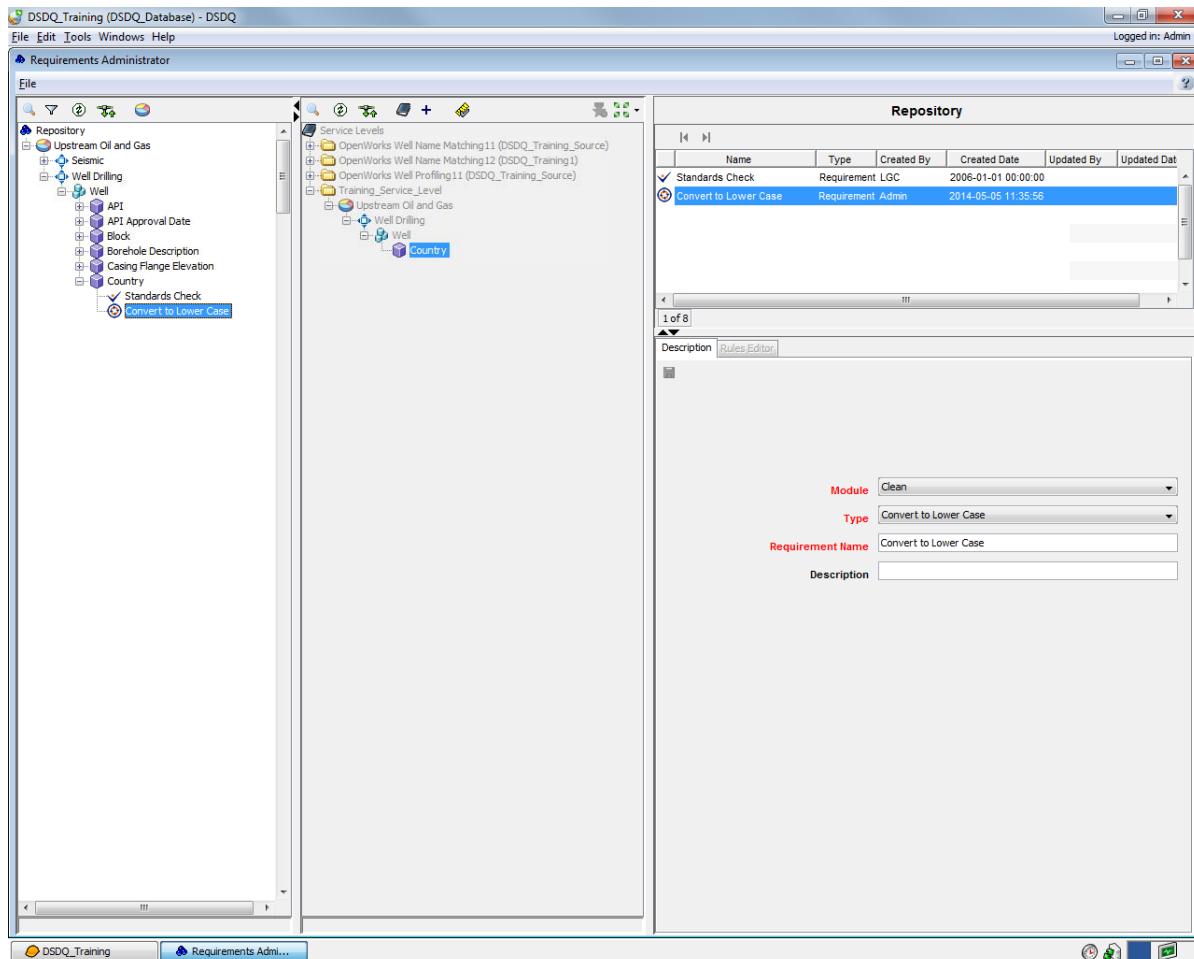
The **Enter Name** dialog box appears.



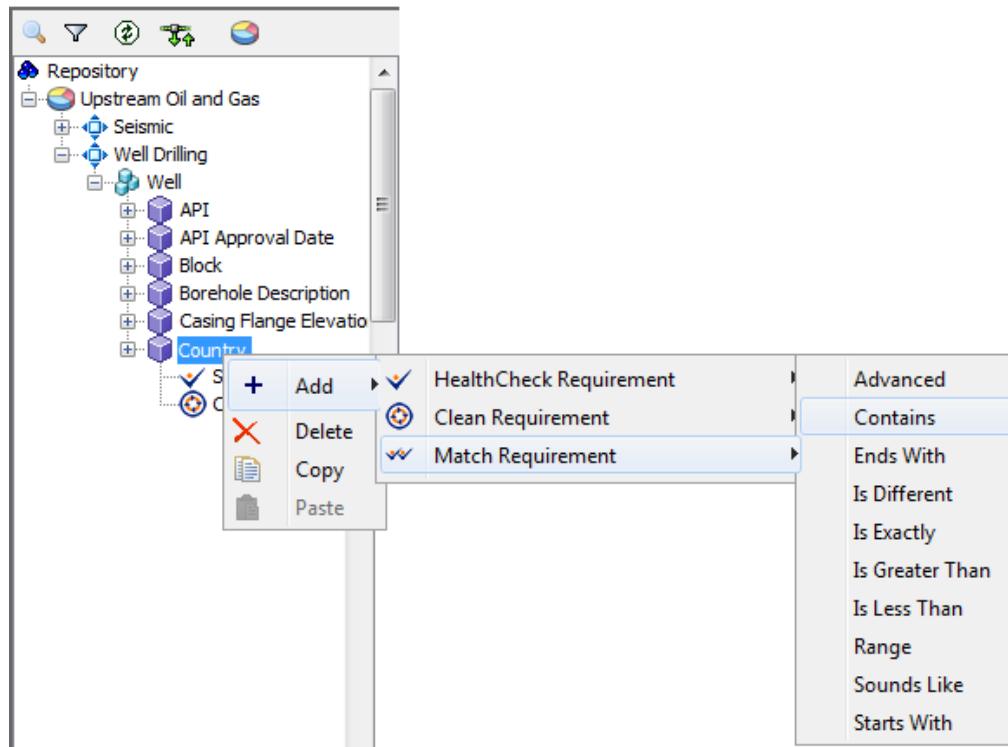
8. Optionally, specify a user-defined name for the requirement.

9. Click **OK**.

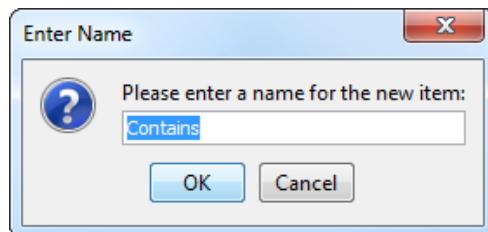
The selected requirement is added to the **Country** element and displays in the **Repository Tree**.



10. Right-click the **Country** element in the **Repository Tree** and select **Add > Match Requirement > Contains** from the pop-up menu.



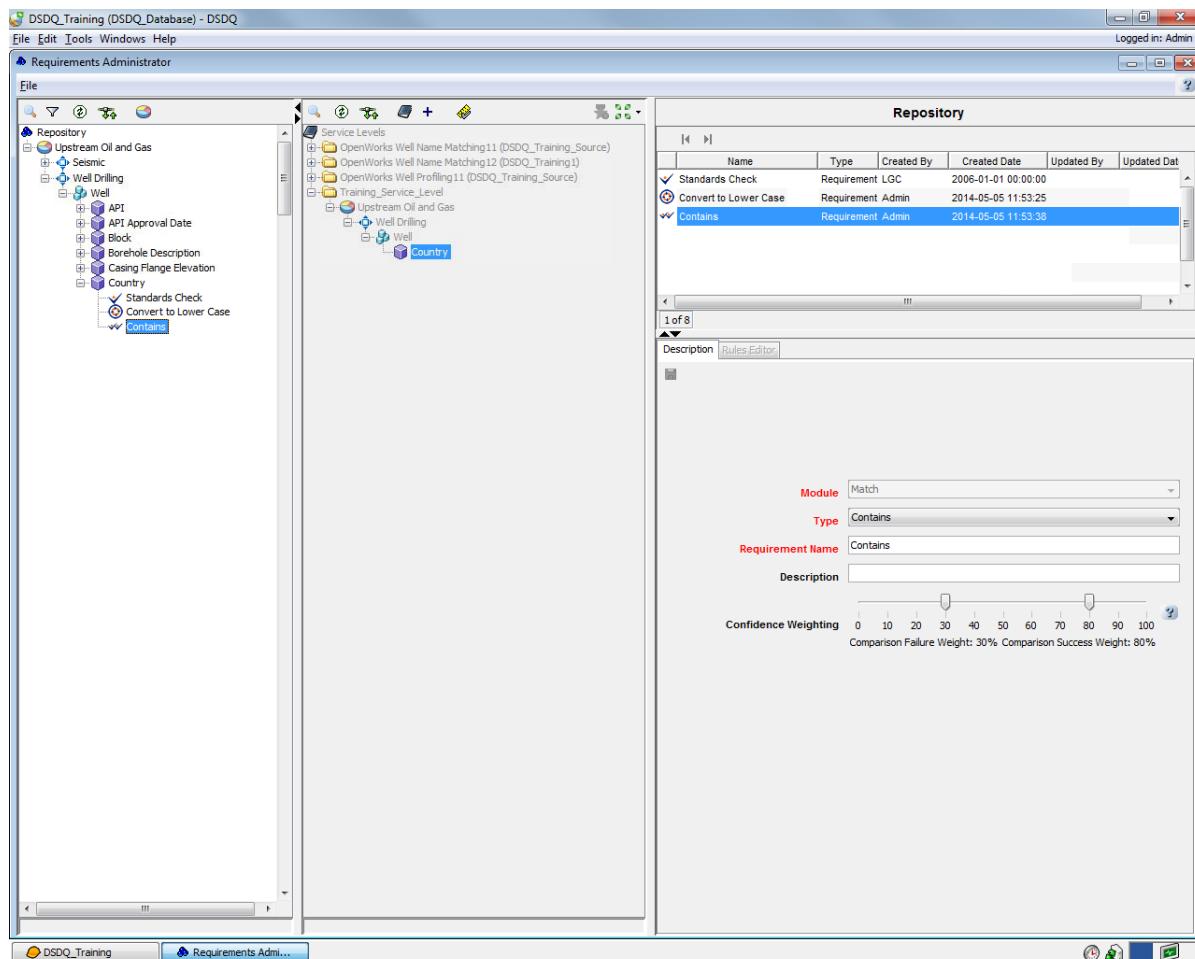
The **Enter Name** dialog box appears.



11. Optionally, specify a user-defined name for the requirement.

**12. Click OK.**

The selected requirement is added to the **Country** element and displays in the **Repository Tree**.



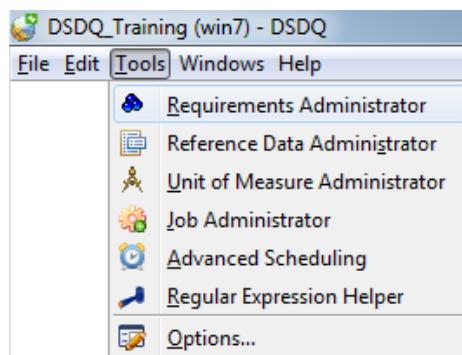
**13. Optionally repeat steps 4 to 6 to add more requirements.**

**14. Select File > Exit from the menu bar on the Requirements Administrator window.**

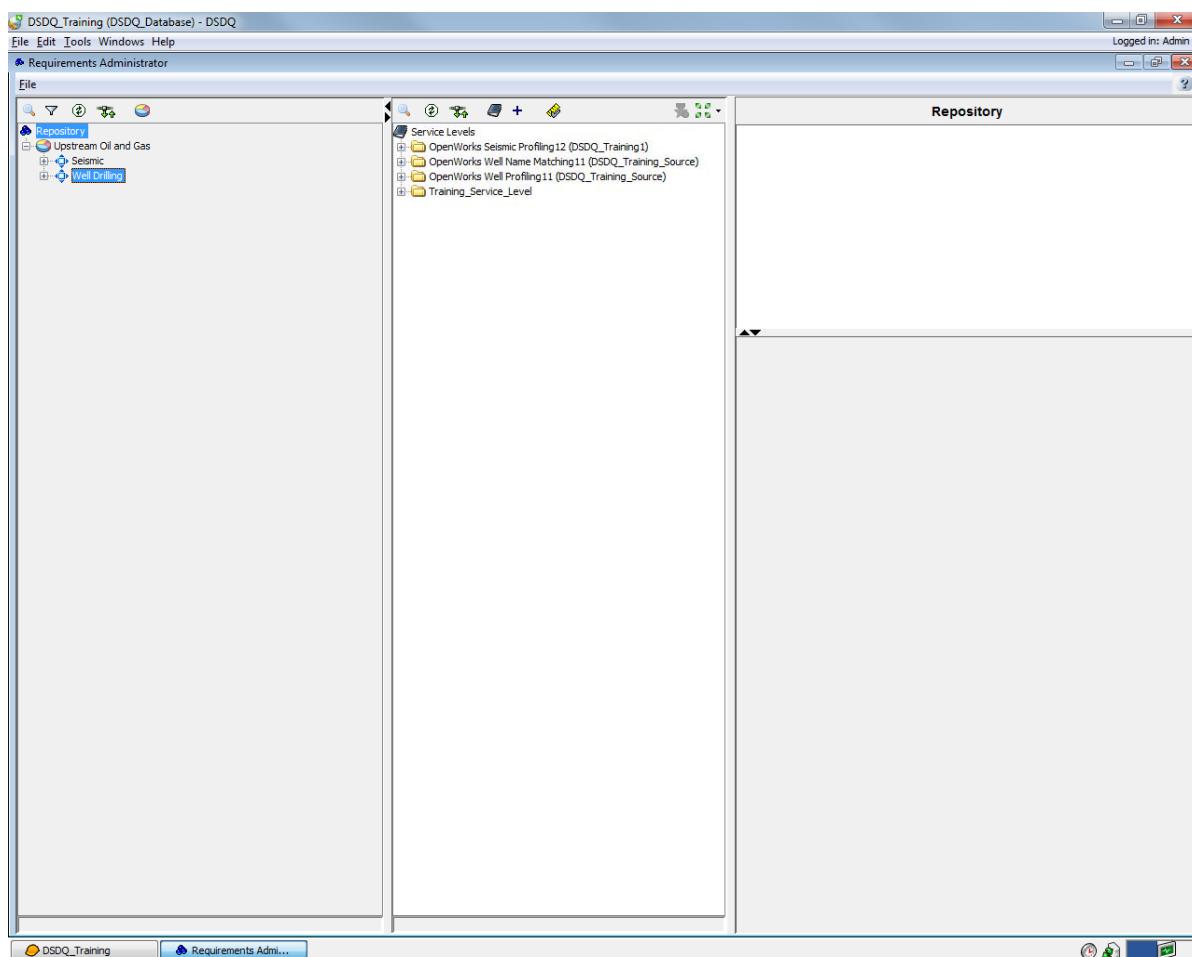
## Exercise: Editing a Requirement

To edit a Requirement:

1. Select **Tools > Requirements Administrator** from the menu bar on the **DSDQ Project** window.

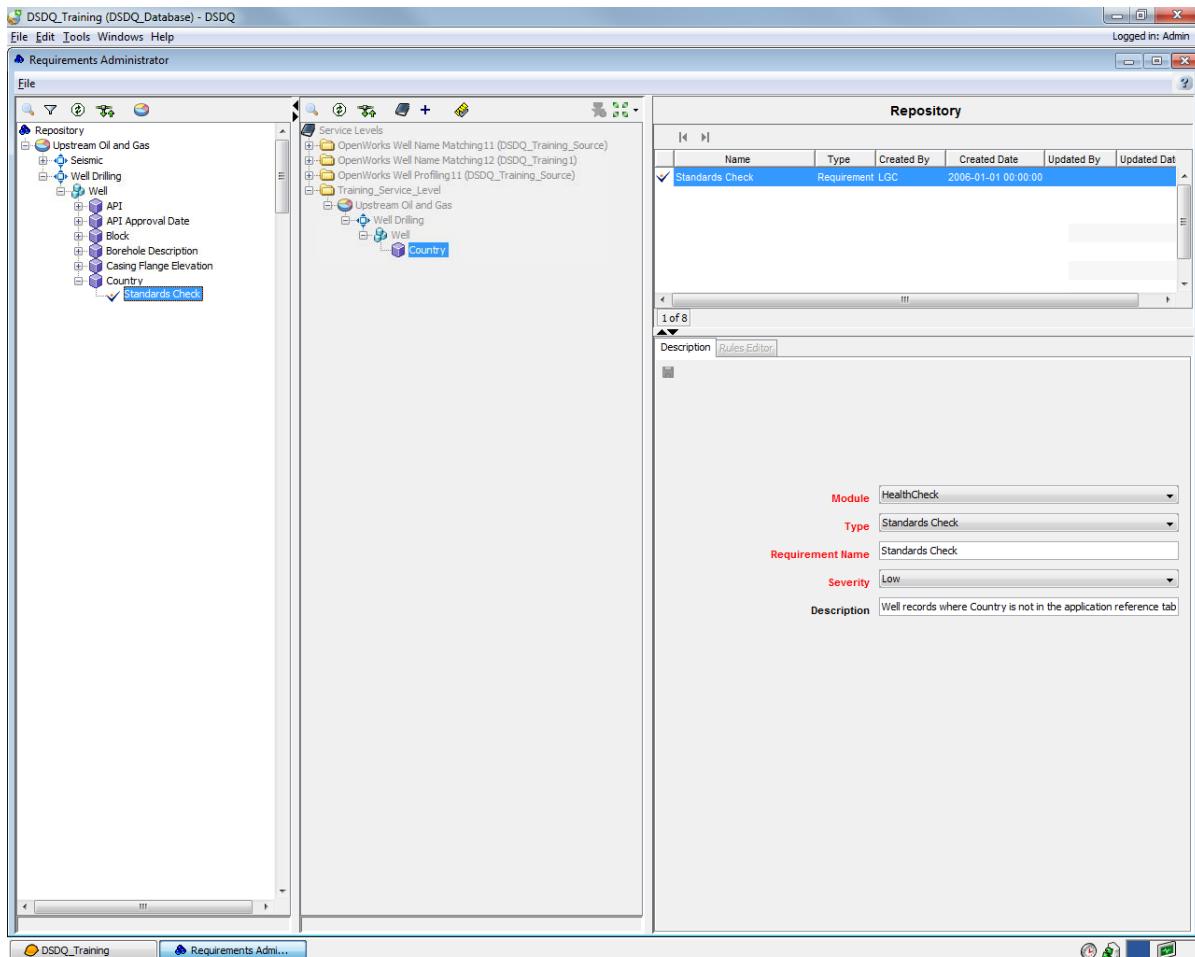


The **Requirements Administrator** window appears.



2. Click to expand the **Well Drilling** area in the **Repository** Tree.

3. Click to expand the **Well** element group.
4. Click to expand the **Country** element.
5. Select the **Standards Check Requirement** in the **Repository Tree**. Information about the **Standards Check** requirement appears in the **Description** tab of the **Details Pane**.



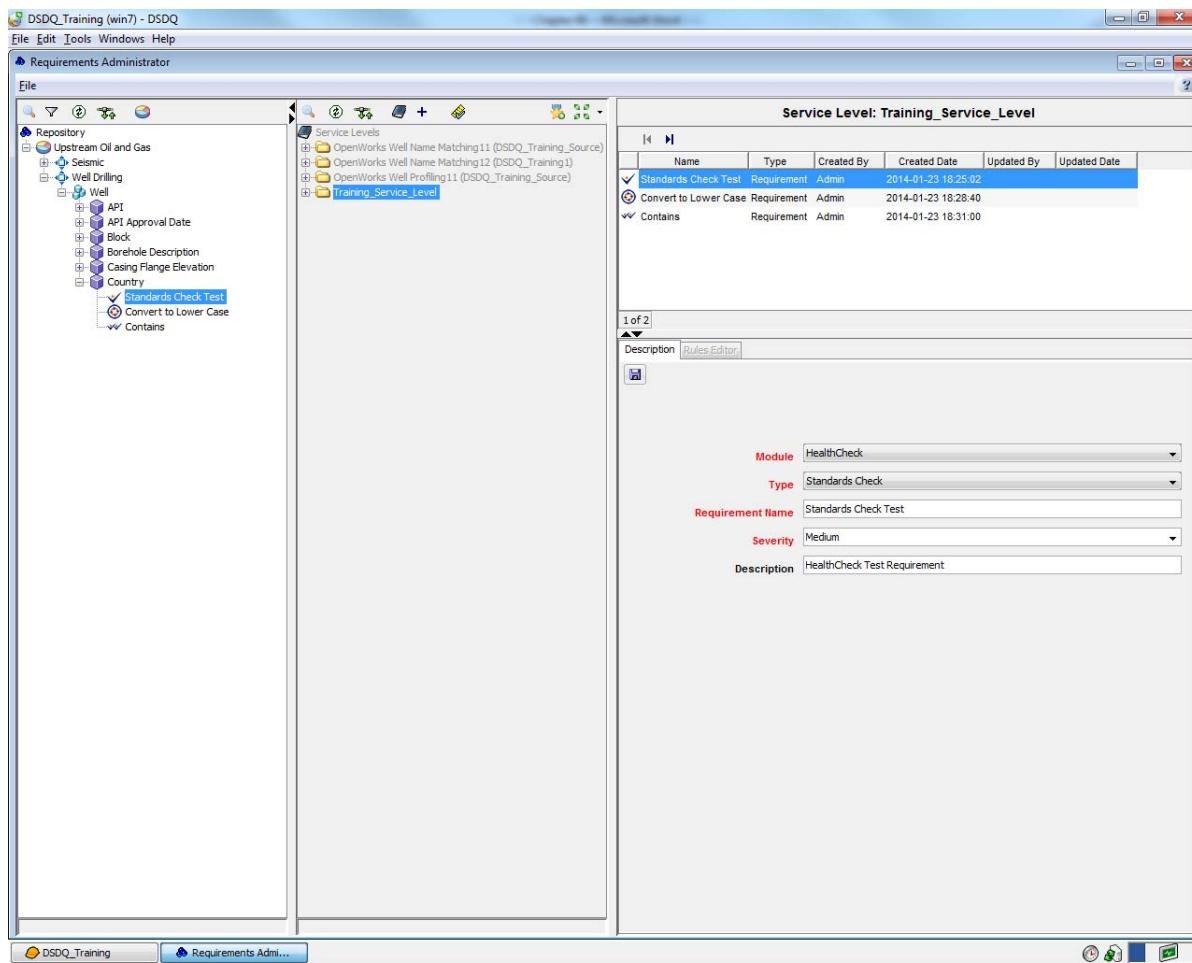
6. Enter **Standards Check** in the **Requirement Name** field of the **Description** tab.

**7. Enter Well records where Country is not in the application reference tab in the Description field.**

**Note**

The Module parameter of a requirement can only be changed between **HealthCheck** and **Clean** phases for Advanced requirement type. The **Severity** parameter is only available for **HealthCheck** requirements. Changing the parameter to High will cause the application to produce high level reporting when running the **Run Detailed HealthCheck Task**.

- 8. Click  to save changes in the **Description** tab.**  
The updated requirement displays in the Repository Tree Pane.



**Note**

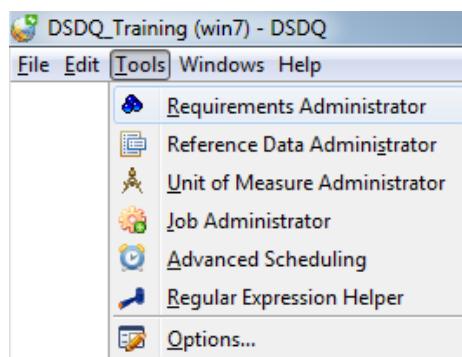
To duplicate a requirement, right-click on the desired requirement and select **Copy** from the pop-up menu. Right-click on the Element and select **Paste** from the pop-up menu to copy the requirement to the desired Element..

9. Select **File > Exit** from the menu bar on the **Requirements Administrator** window.

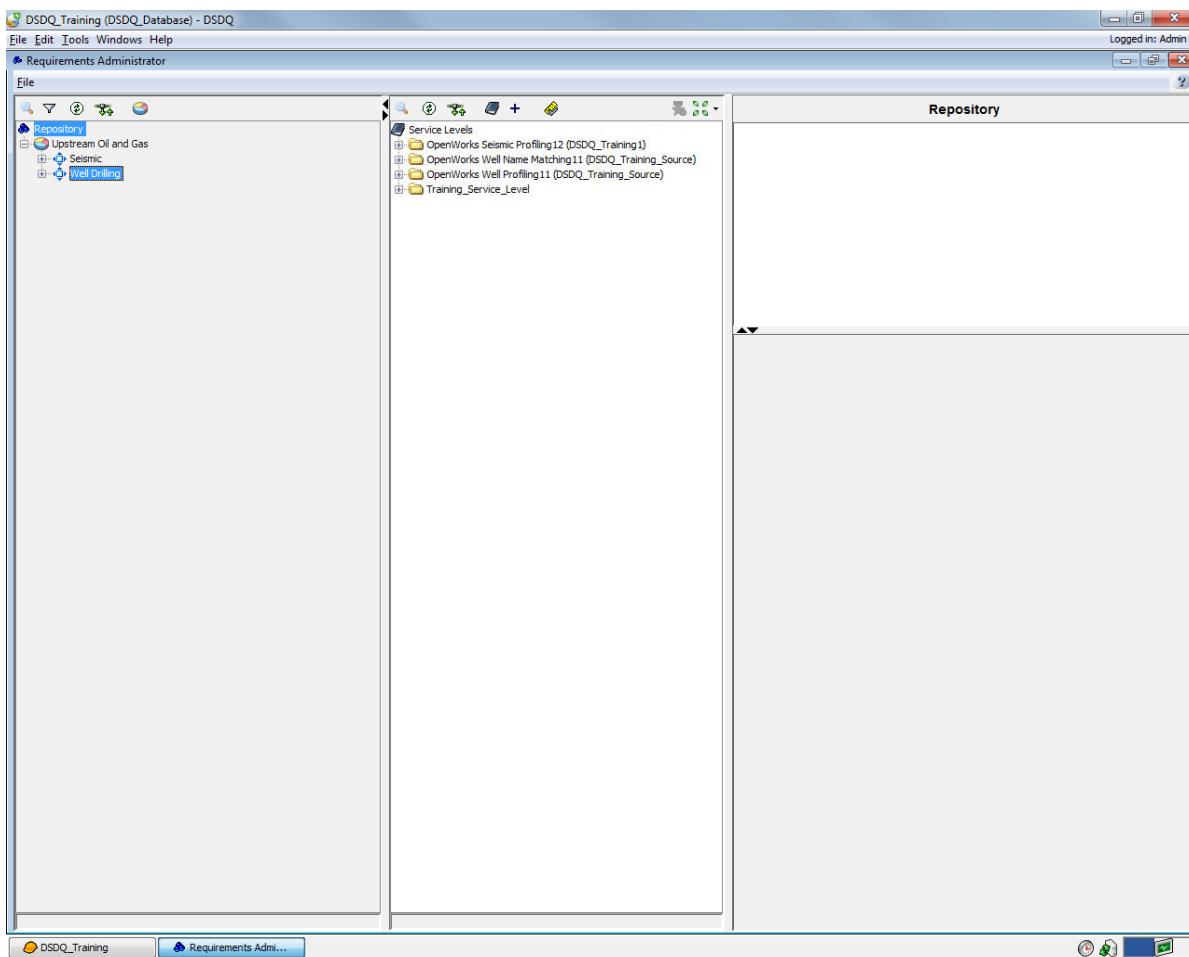
### **Exercise: Deleting a Requirement**

To delete a Requirement:

1. Select Tools > Requirements Administrator from the menu bar on the DSDQ Project window.

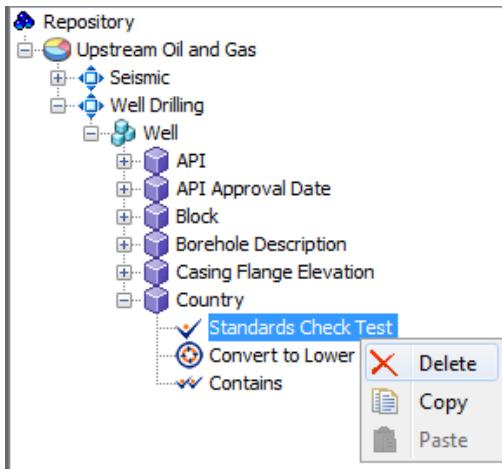


The Requirements Administrator window appears.

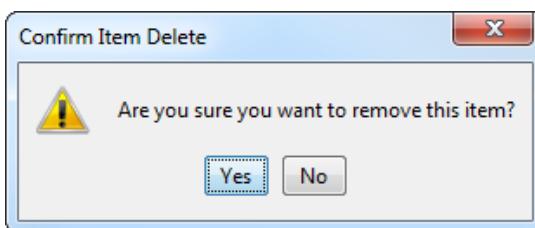


2. Click to expand the **Well Drilling** Area in the Repository Tree.
3. Click to expand the **Well** Element Group.
4. Click to expand the **Country** Element.

5. Right-click the **Standards Check Test** Requirement and select **Delete** from the pop-up menu.

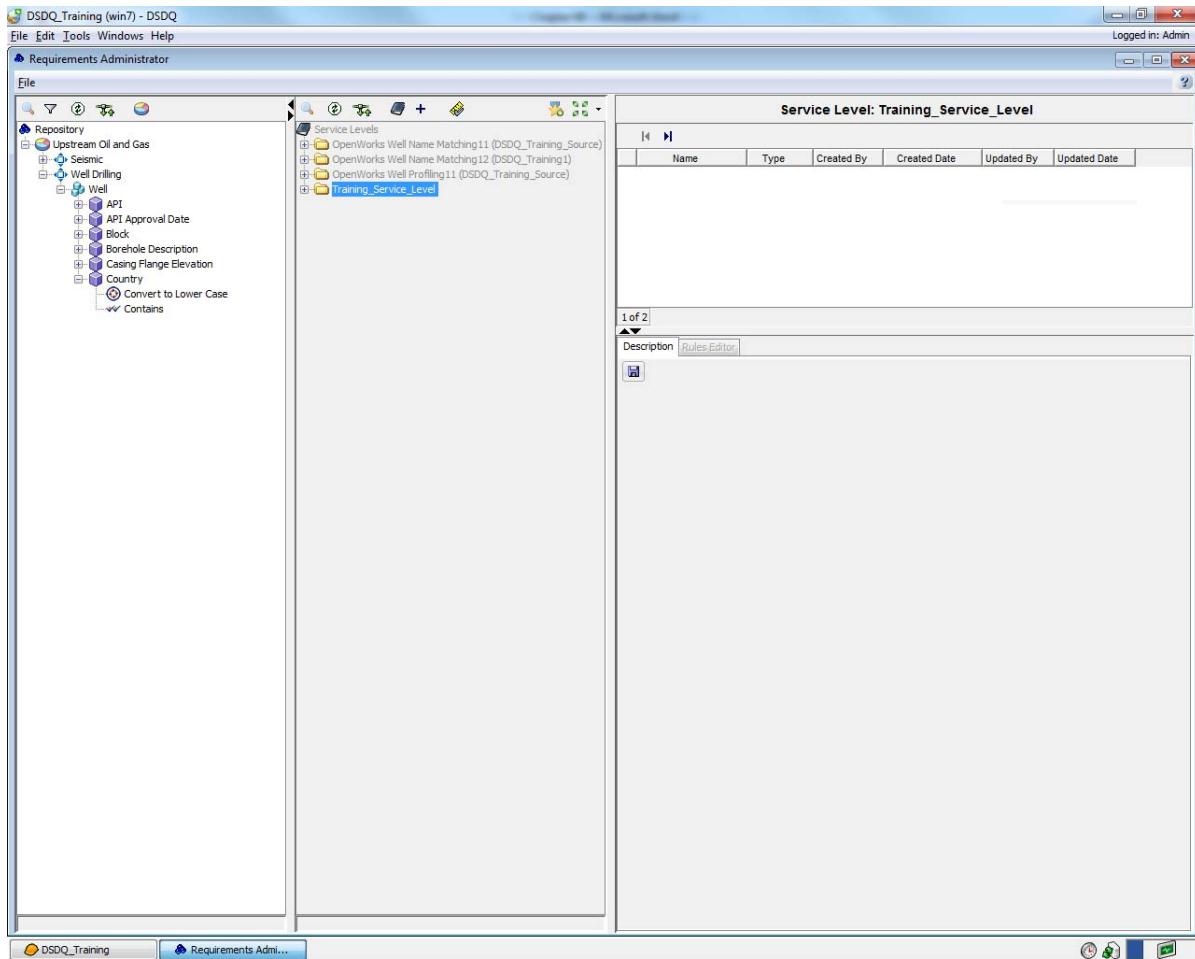


The **Confirm Item Delete** dialog box appears.



6. Click Yes.

The selected requirement is deleted.

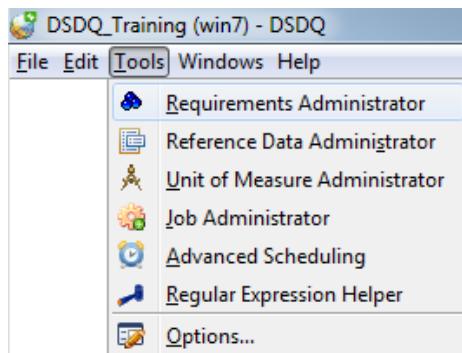


7. Select **File > Exit** from the menu bar on the **Requirements Administrator** window.

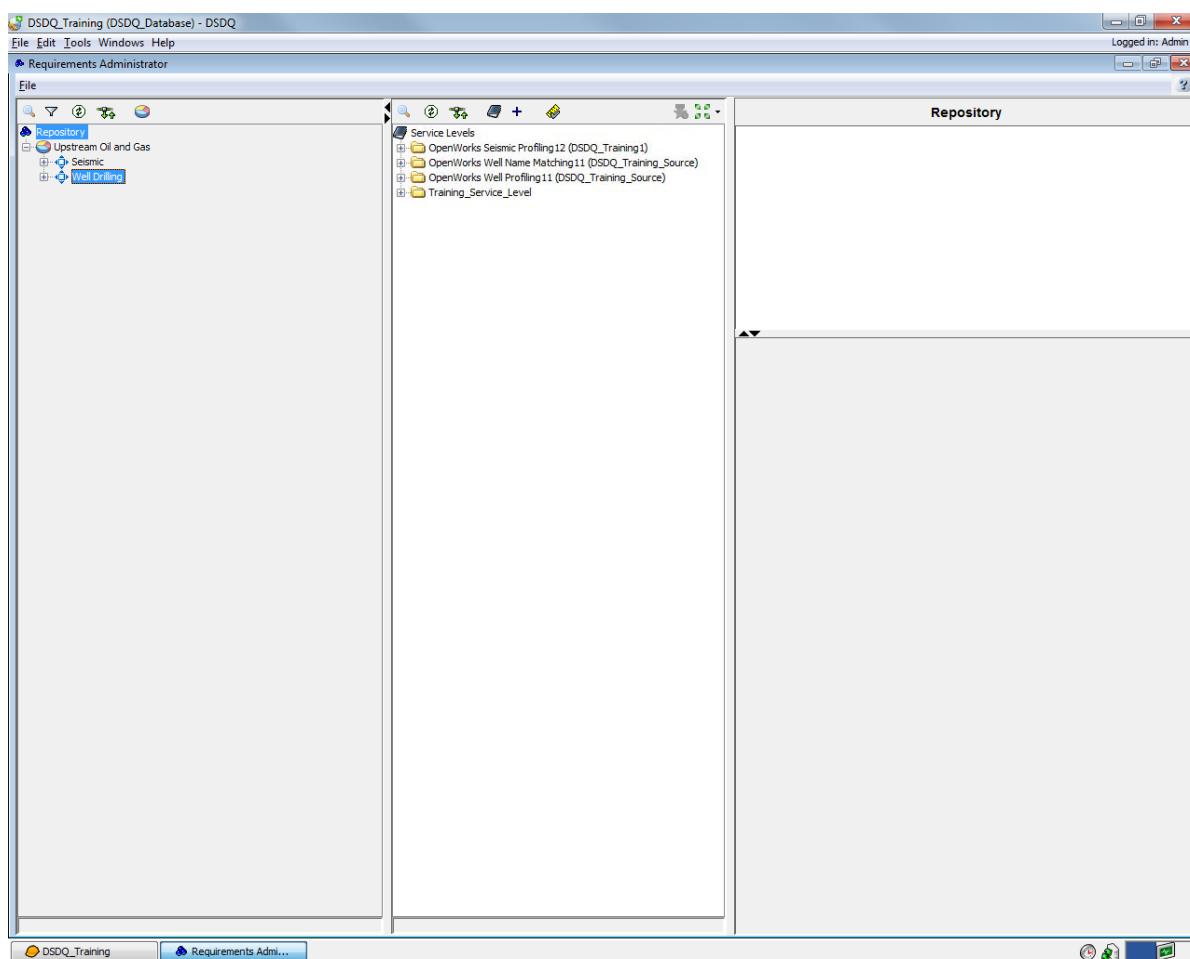
## Exercise: Validating a Requirement

To validate a Requirement:

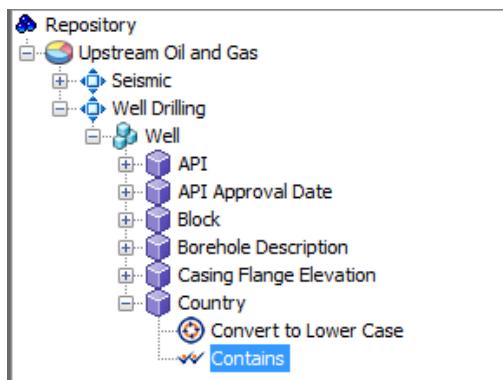
1. Select **Tools > Requirements Administrator** from the menu bar on the **DSDQ Project** window.



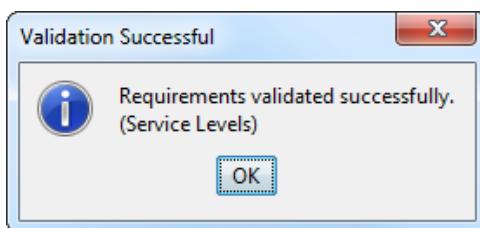
The **Requirements Administrator** window appears.



2. Click to expand the **Well Drilling** area in the Repository Tree.
3. Click to expand the **Well** element group.
4. Click to expand the **Country** element.
5. Select the **Contains** requirement in the Repository Tree.



6. Click on the Service Level Tree Pane toolbar.  
A Validation Successful dialog box appears.



7. Click **OK**.

## Rules Editor

---

The Rules Editor creates custom requirements. These requirements can only be created for the Advanced Requirement. An Advanced Requirement allows the user to create a Rule Set, which is a combination of multiple requirements. An Advanced Requirement enables the user to perform complex data manipulation. Rules can be made using a combination of HealthCheck, Clean, and Match Requirements. They are grouped by the type of work they perform in the Rules Editor.

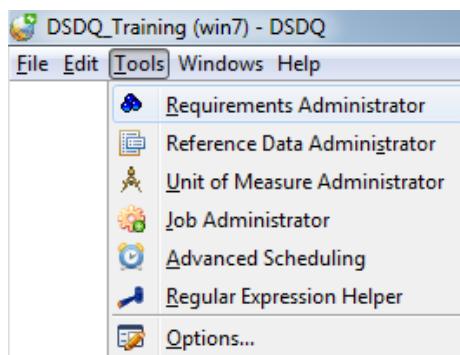
**Note**

This section is for advanced users who understand basic programming and database querying.

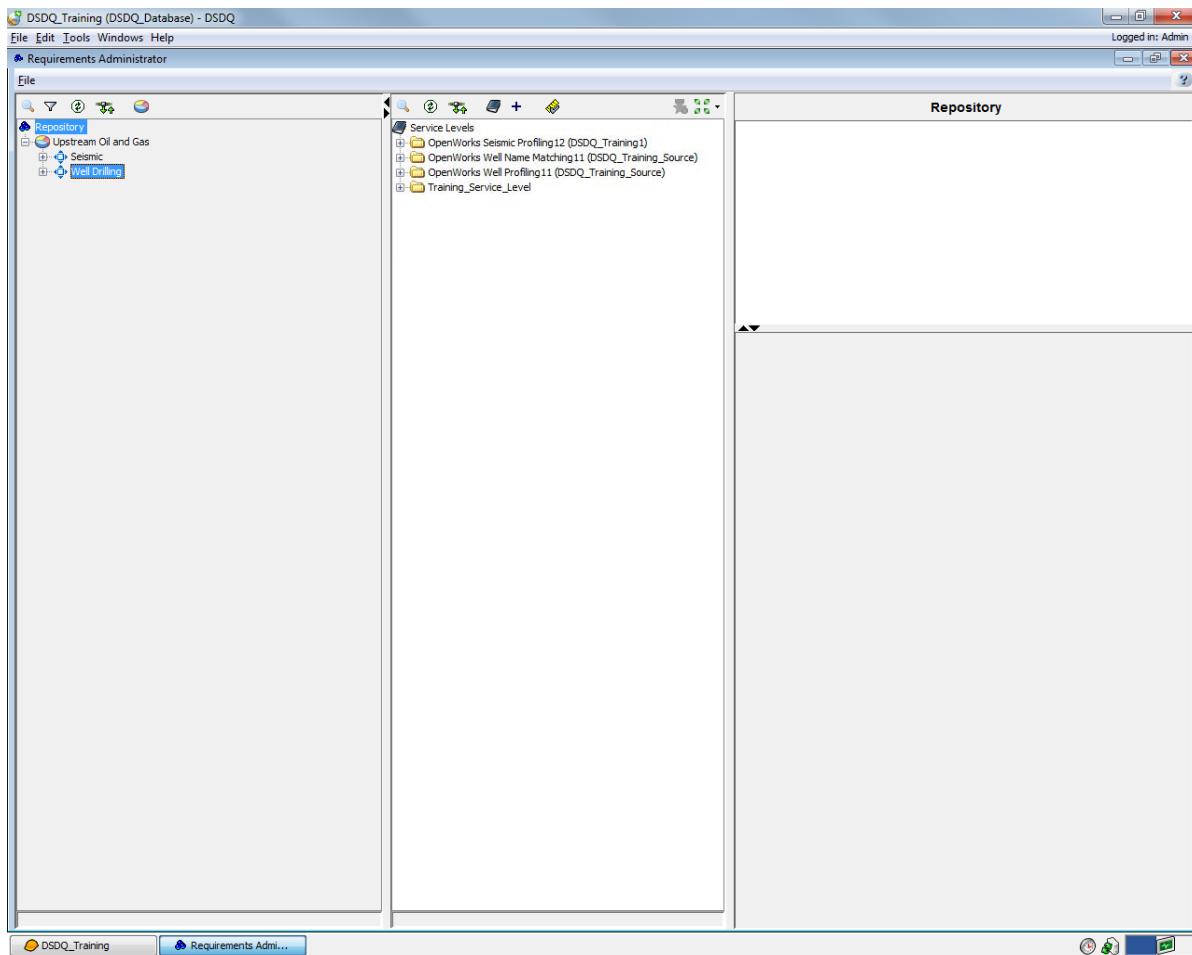
### **Exercise: Managing Advanced Requirements**

To manage advanced requirements:

1. Select **Tools > Requirements Administrator** from the menu bar on the **DSDQ Project** window.

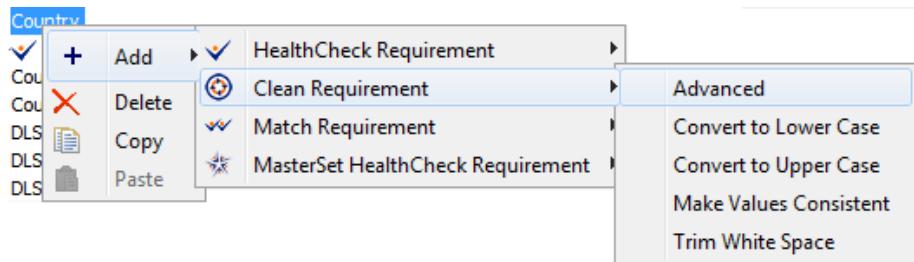


The Requirements Administrator window appears.

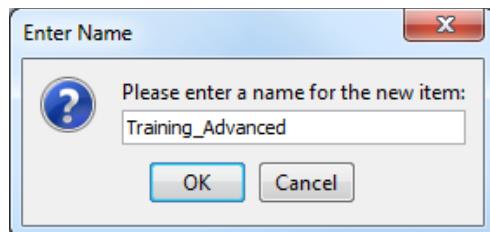


2. Click to expand the **Upstream Oil and Gas** sector in the Repository Pane.
3. Click to expand the **Well Drilling** area in the Repository Tree Pane.
4. Click to expand the **Well** element group in the Repository Tree Pane.

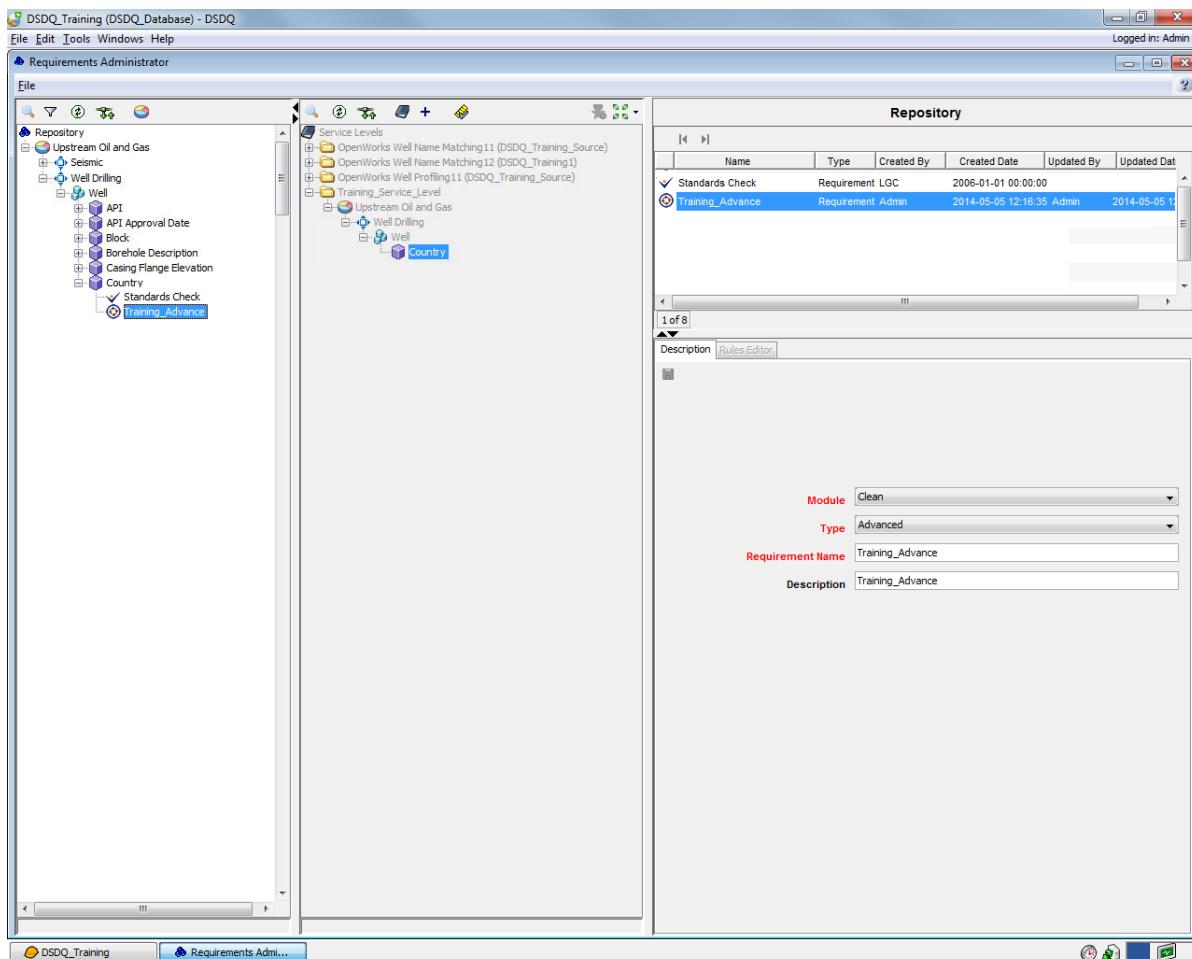
5. Right-click on the **Country** Element and select **Add > Clean Requirements > Advanced** from the pop-up menu.



The **Enter Name** dialog box appears.

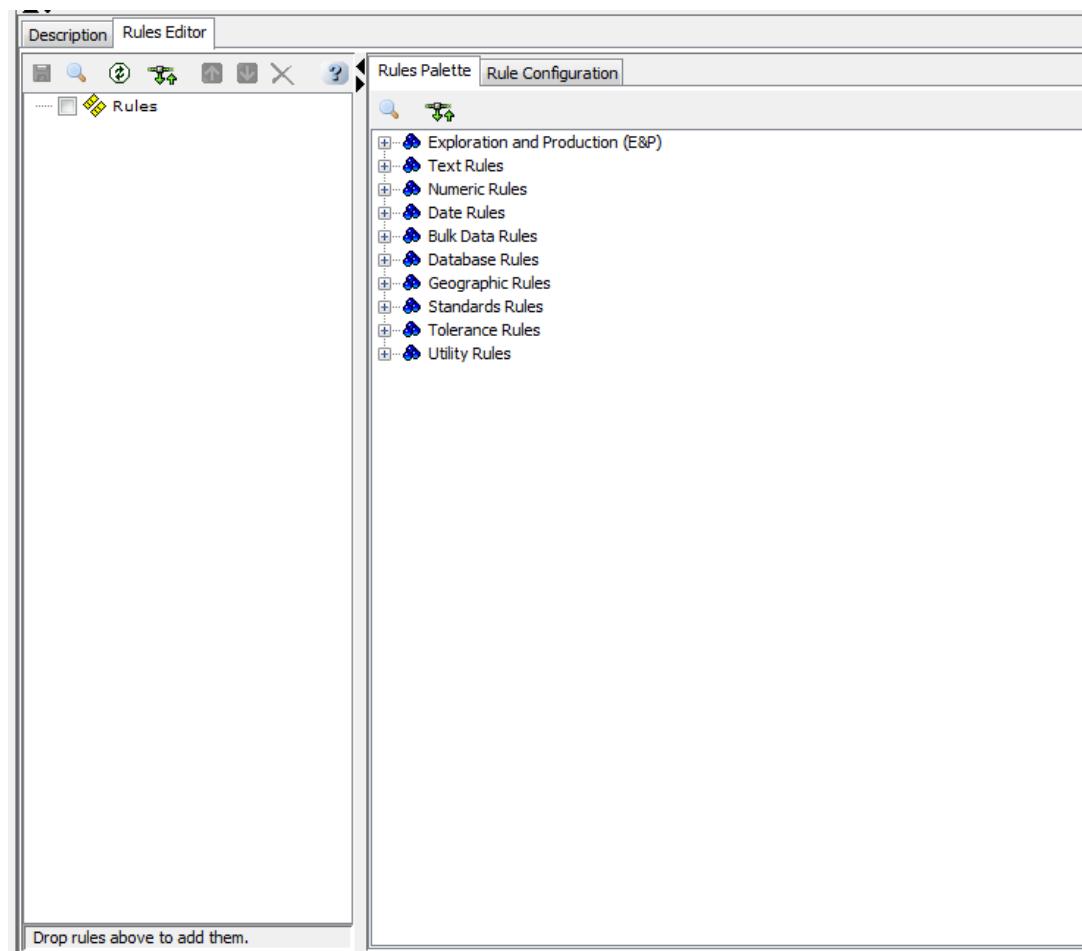


6. Optionally, enter a name and click **OK**.

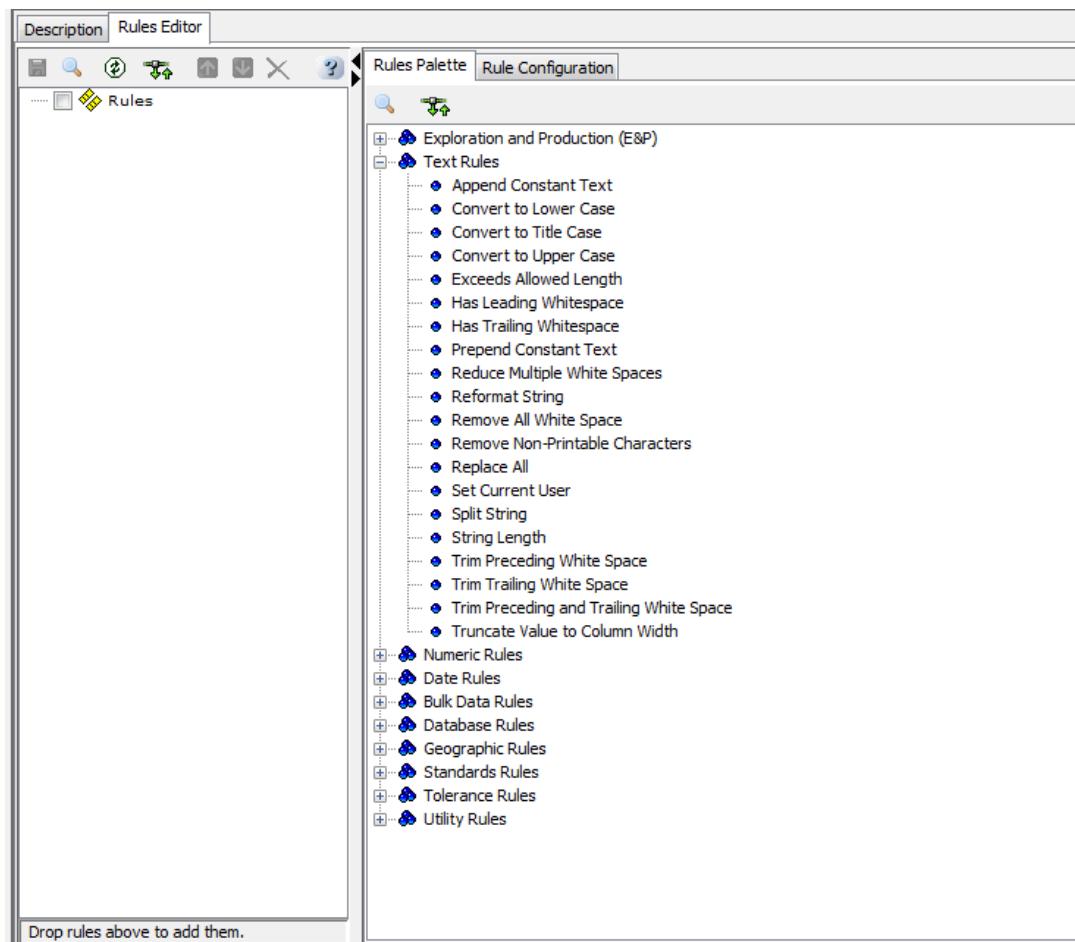


7. Select **Clean** from the **Module** drop-down list.
8. Select **Advanced** from the **Type** drop-down list.
9. Enter **Training\_Advanced** in the **Requirement Name** field.
10. Enter **Training\_Advanced** in the **Description** field.

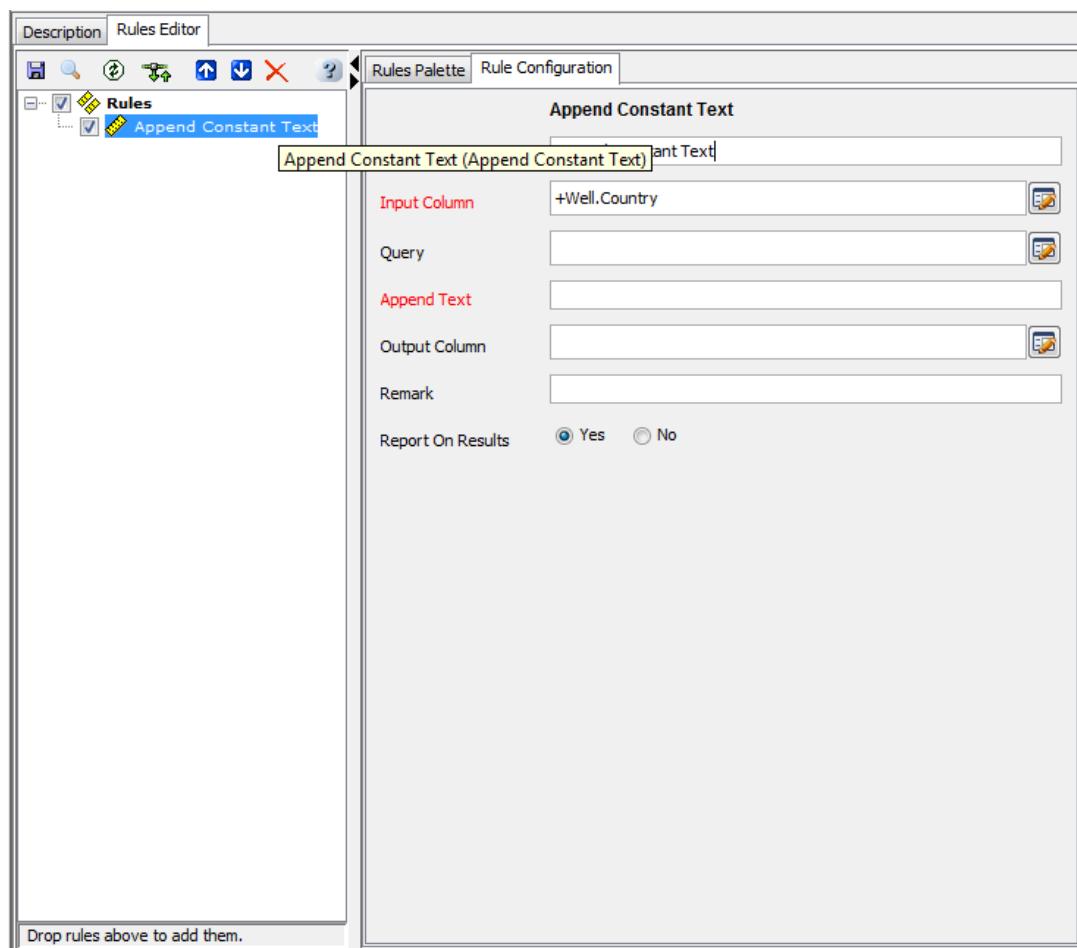
11. Select the **Rules Editor** tab.



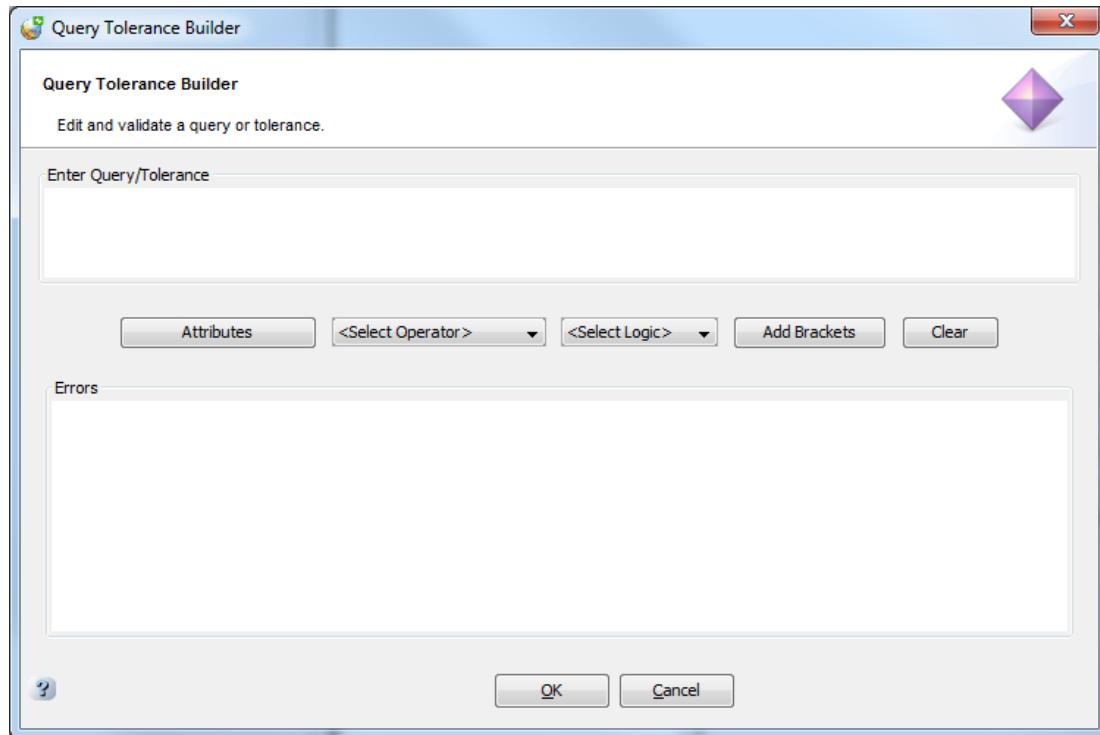
12. Click to expand **Text Rules** in the **Rules Palette** window.



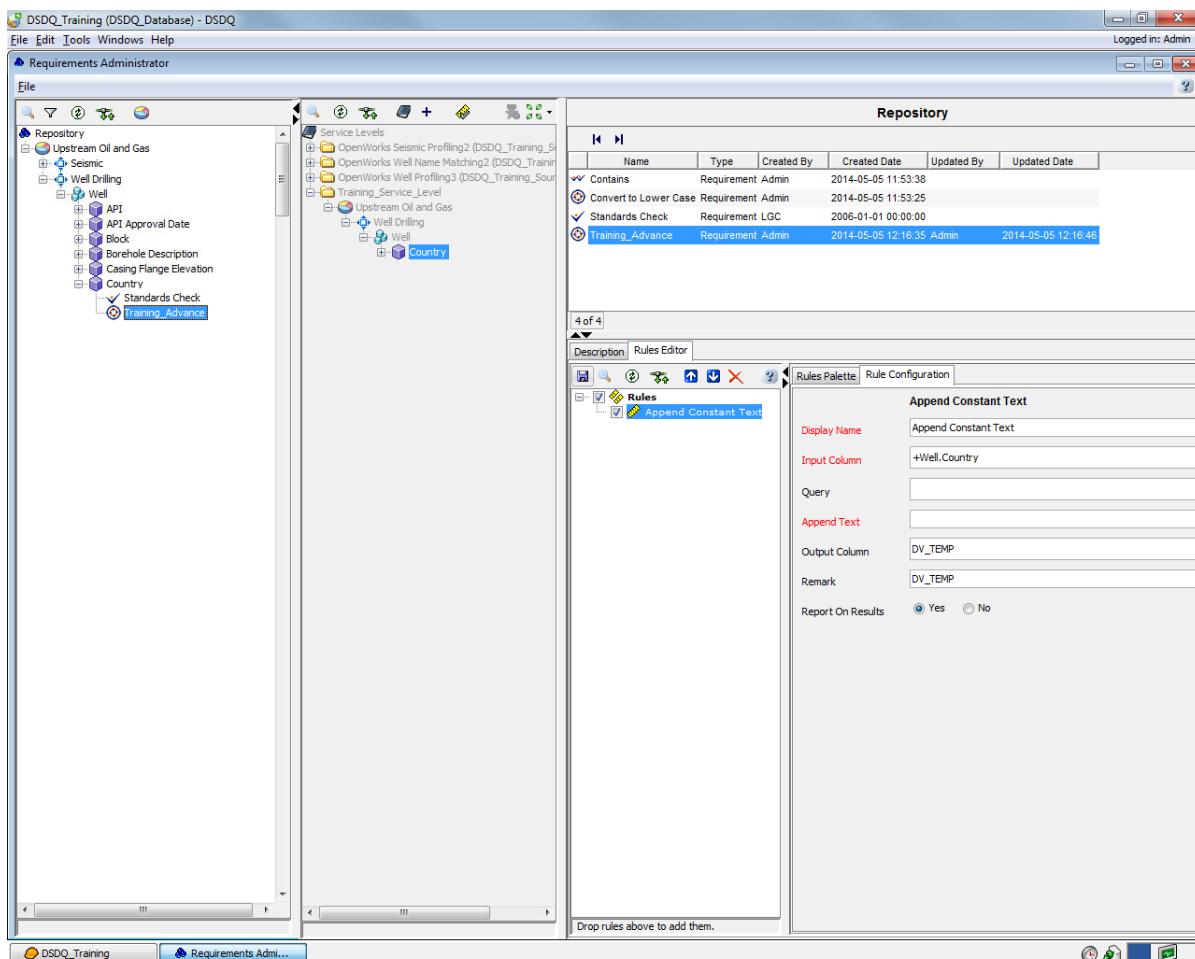
13. Drag **Convert to Lower Case** Rule from the **Rules Palette** to the **Rules Pane**.  
The **Rule Configuration** tab opens automatically.



14. Click  adjacent to the **Query** field to enter/select the query.  
The **Query Tolerance Builder** window appears.



15. Enter the required query and click **OK**.



16. Select **DV\_Temp** from the **Output Column** field.

17. Enter **DV\_TEMP** in the **Remark** field.

18. Select the **Yes** option for **Report on Results**.

19. Click to save your rule.

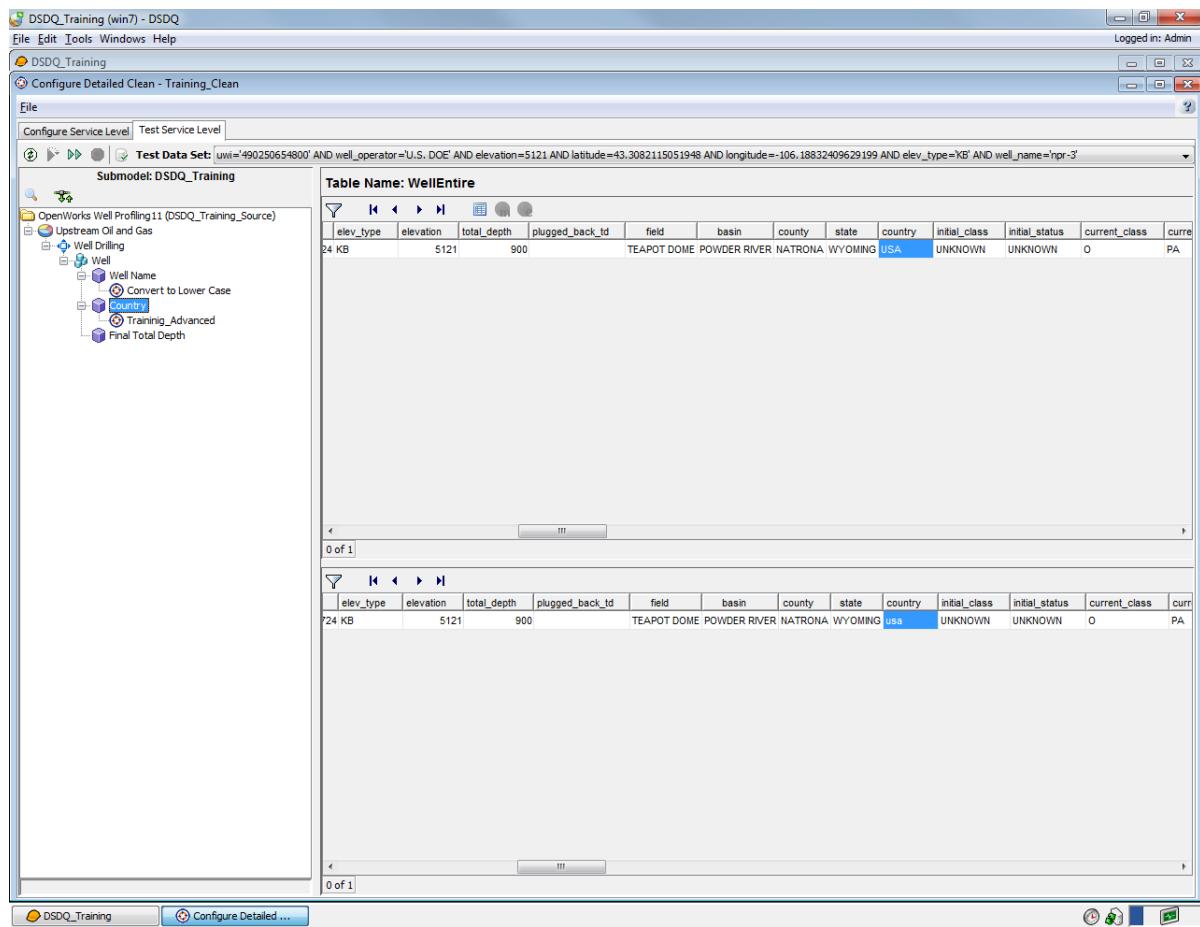
The advanced rule is created.

20. Click to expand the **Training\_Clean** Phase.

21. Click to expand the **Detail Clean** Activity.

22. Double-click the **Configure Detail Clean** Tool or right-click the **Configure Detail Clean** Tool and select **Open Tool** from the pop-up menu.

23. Select the **Test Service Level** tab to view advance rule results.



24. Select **File > Exit** from the menu bar on the **Requirements Administrator** window.

# **Chapter 9**

# **Configuring the Data Quality Web Dashboard**

The Decision Space Data Quality Web Dashboard allows you to view Data Quality results and search Master Data stores. You can access the Web Dashboard through your office Intranet using a web browser on your workstation, tablet or mobile device. The default link to the Data Quality Web Dashboard is ***http://SERVER NAME:8091***. Note that **SERVER\_NAME** is the name of the server machine where you have installed the Data Quality application.



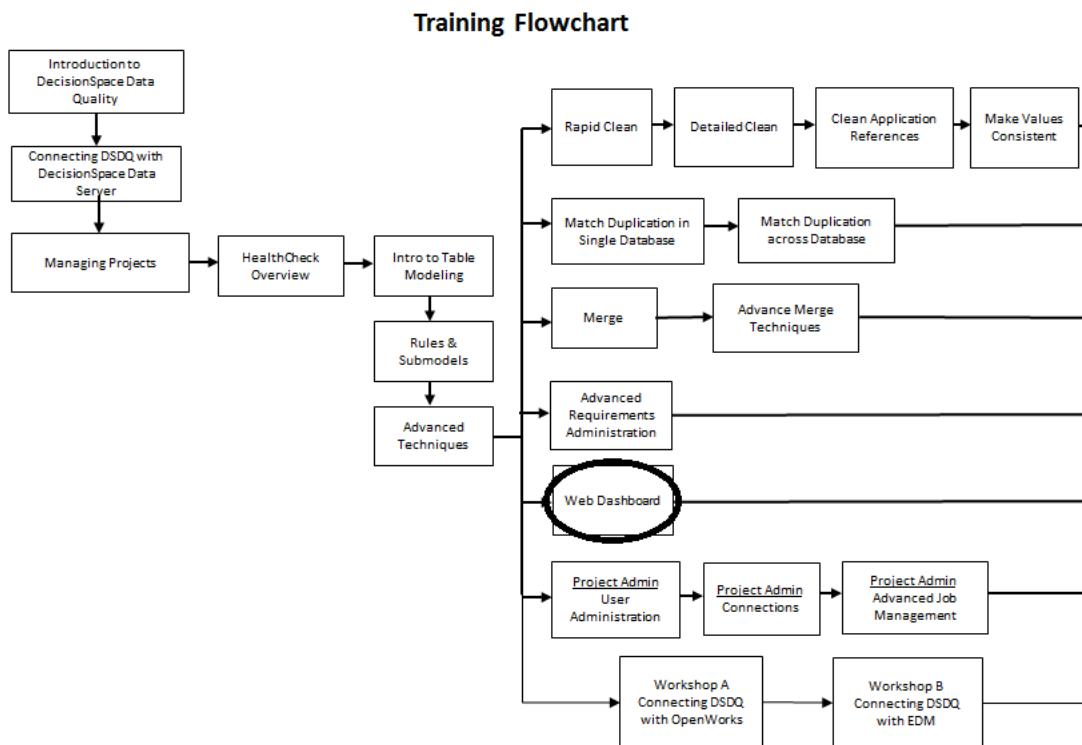
# Chapter Overview

---

In this chapter, you will learn about:

- Submodel analysis
- Grouping Master Data information
- Updating Statistics on the Web Dashboard

Topics covered in each chapter are outlined in the following illustration. Those specific to the current chapter will be circled in black for your reference:



## Submodel Analysis

---

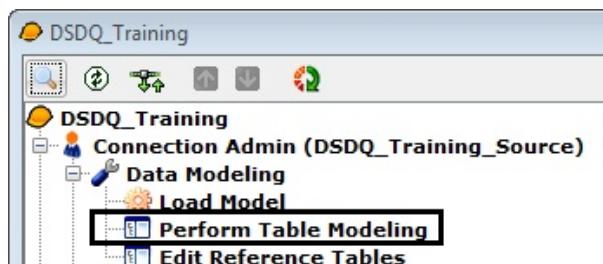
A submodel is a core data structure within the Data Quality application that is used by all Activities. It is created using the Perform Table Modeling task, a submodel is a grouping of data tables within a database and is integral to defining data configurations. Submodels can be used to represent one database table, many database tables, views, synonyms and more. Submodels have a hierarchical structure, where elements are tiered according to how they are linked together. Within submodels, primary keys, report groups, and column ordering can also be set.

### ***Exercise: Publishing Submodels***

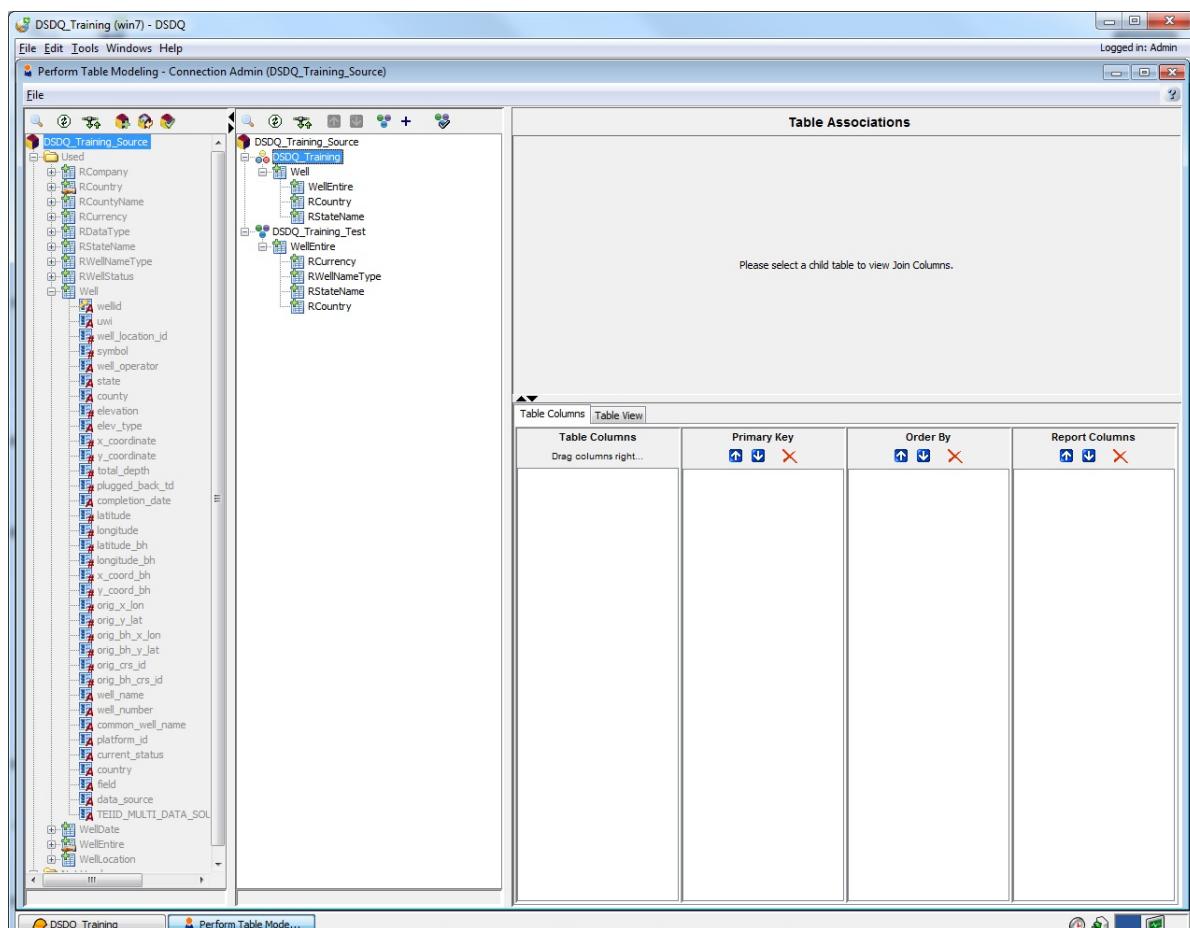
To view information about a DecisionSpace Data Quality submodel, you must publish it to the Web Dashboard. To publish a submodel to the Web Dashboard:

1. Click  on the DecisionSpace Data Quality Tree to expand **Connection Admin (DSDQ\_Training\_Source)**.
2. Click  on the DecisionSpace Data Quality Tree to expand the **Data Modeling** Activity.

3. Double-click the **Perform Table Modeling** Tool.

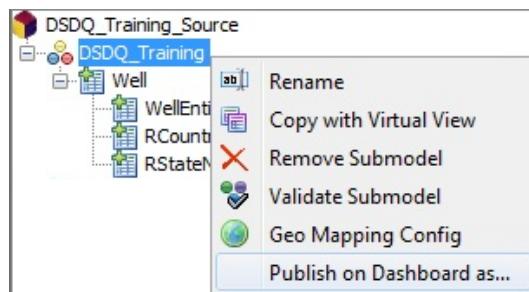


The **Perform Table Modeling - Connection Admin** window appears.



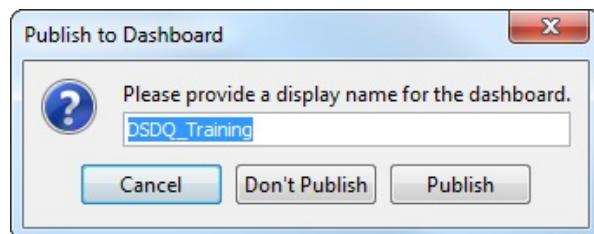
4. From the right tree, right-click **DSDQ\_Training** from the Submodel Listing Tree and select **Publish on Dashboard as...**

from the pop-up menu.



The **Publish to Dashboard** dialog box opens.

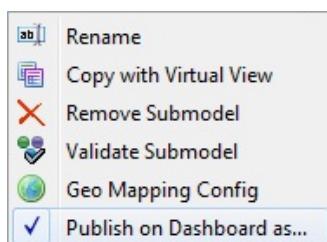
5. Enter **DSDQ\_Training** in the **Publish to Dashboard** dialog box and click **Publish**.



The sub-model is published to the Web Dashboard.

6. To confirm that the submodel has been published to the Web Dashboard, right-click **DSDQ\_training** on the Submodel Listing Tree.

A checkmark appears on the right side of the **Publish on dashboard as...** option.



7. Select **File > Exit** from the menu bar to close the **Perform Table Modeling** window.

## Grouping Master Data Information

---

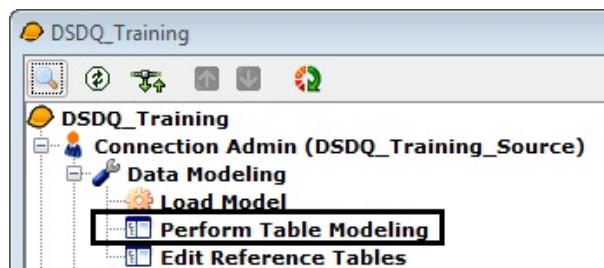
You group master data information into Column Display Groups for the purposes of displaying it on the Web Dashboard or using it as search criteria within the Web Dashboard.

### **Exercise: Configuring Columns Display Groups**

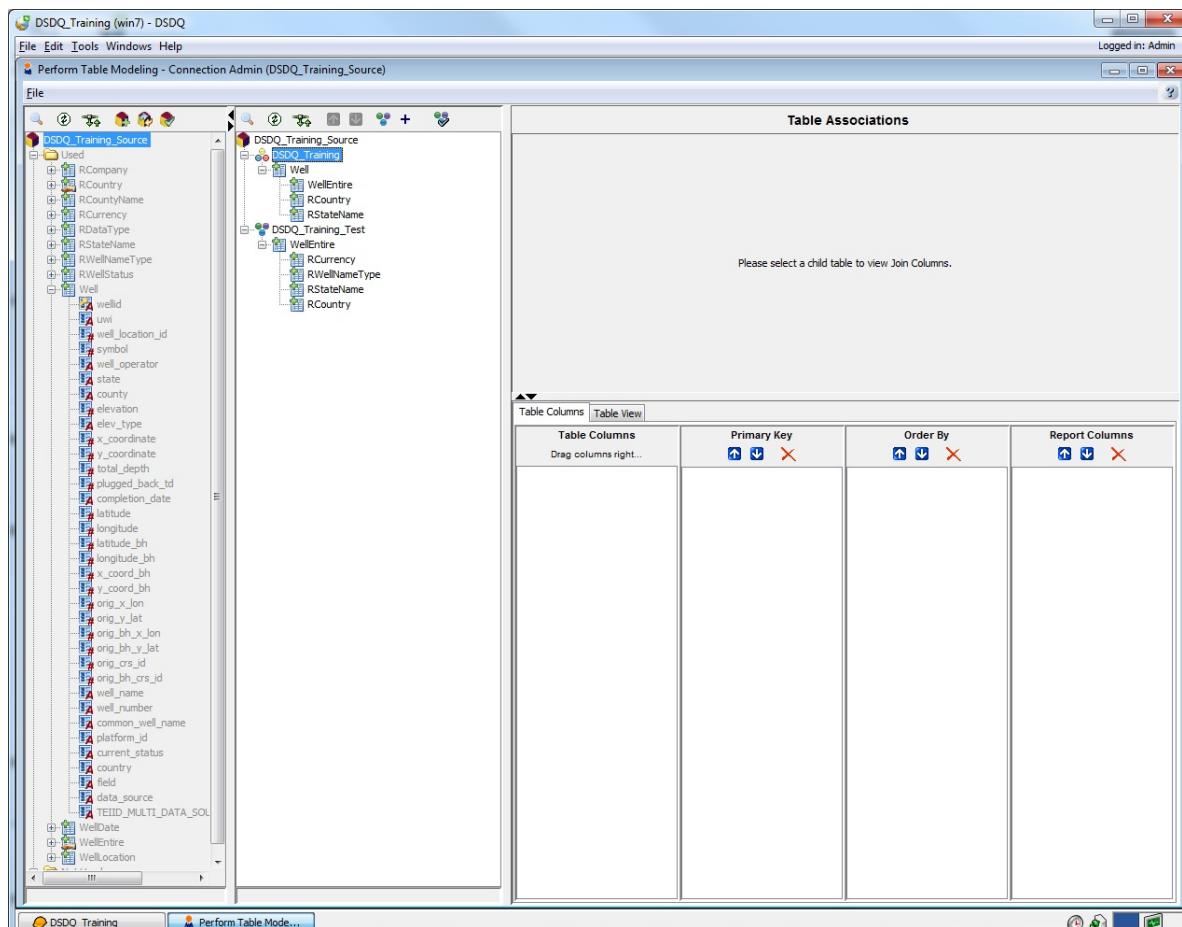
The web dashboard allows users to view master data information through the intranet, grouped by columns display groups.

To configure column display groups:

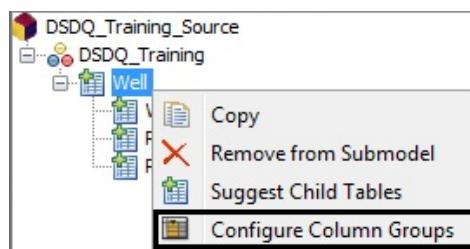
1. Ensure that **Connection Admin (DSDQ\_Training\_Source)** appears expanded on the DecisionSpace Data Quality Tree.
2. Ensure that the **Data Modeling** Activity appears expanded the DecisionSpace Data Quality Tree.
3. Double-click the **Perform Table Modeling** tool.



The Perform Table Modeling window appears.

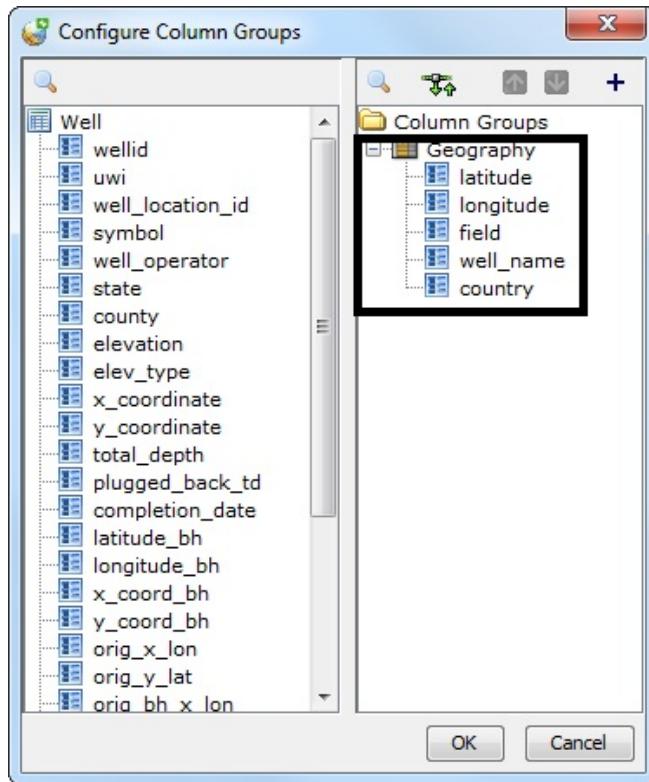


- From the right tree, right-click **Well** from the Submodel Listing Tree and select **Configure Column Groups** from the pop-up menu.



- Enter **Geography** as the name of the column group and click **OK**. The column group is created and displays under Column Groups.
- Select desired columns from the table tree on the left and then drag and drop each column under the **Column Groups** folder on the

right.



#### Note

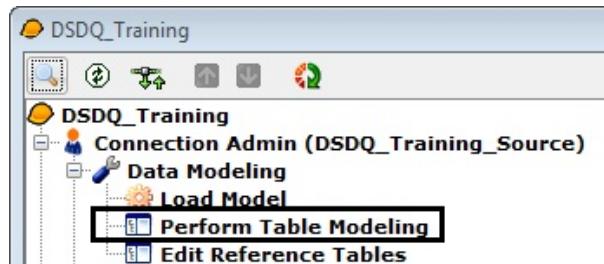
You can only drag and drop one column from the table tree to the **Column Groups** folder at any given time. Repeat step 6 as required for each column.

7. Click **OK** to save the changes and return to the **Perform Table Modeling** window.
8. Select **File > Exit** from the menu bar to close the **Perform Table Modeling** window.

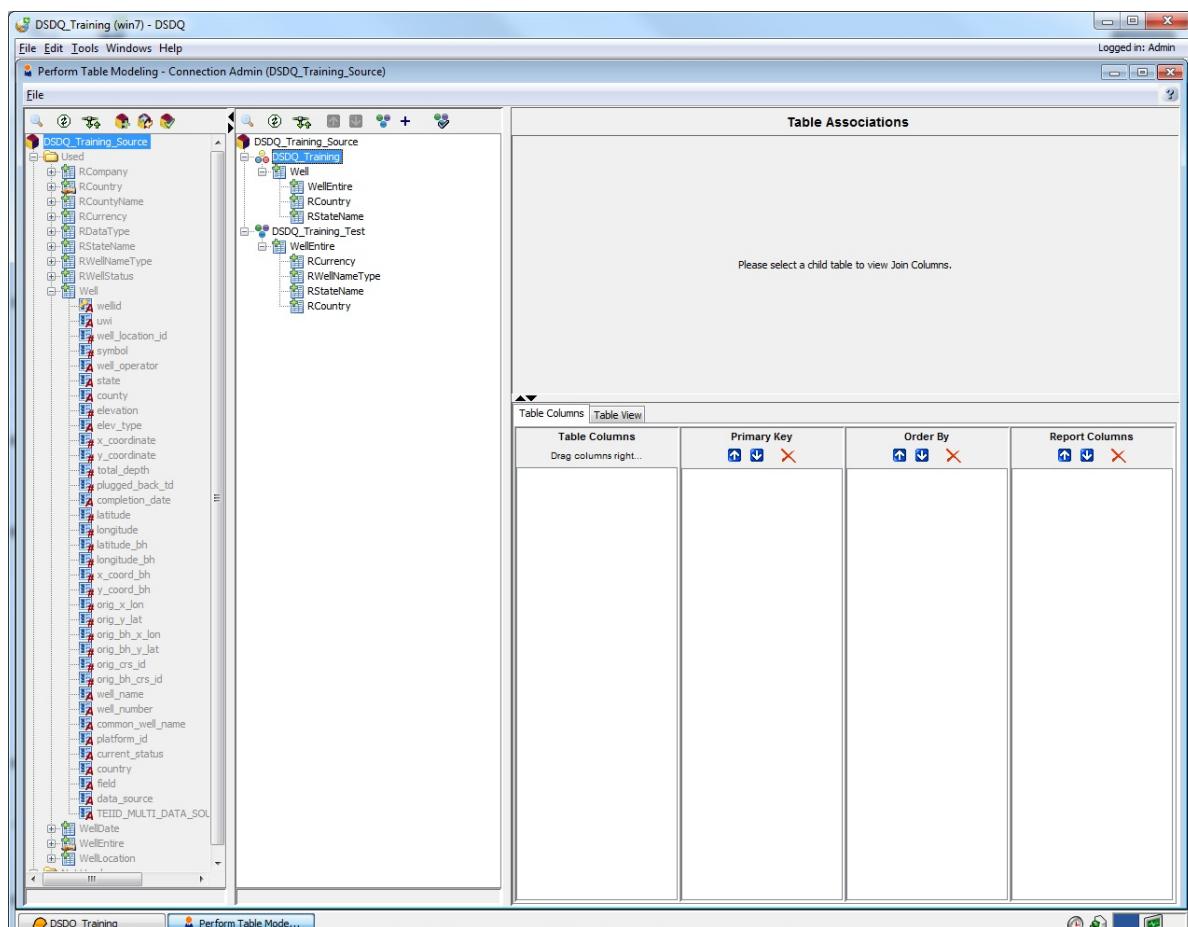
### **Exercise: Specifying a Column Dashboard Name**

The Web Dashboard can be used by not only those with database experience, but anyone interested in viewing data. Due to this, often times a direct column name from a database can be seen as confusing or unnecessary. In order to make information more concise when viewing the dashboard master data, every column can have a dashboard column name that is used instead of the database column name. To set a column dashboard name:

1. Ensure that **Connection Admin (DSDQ\_Training\_Source)** appears expanded on the DecisionSpace Data Quality Tree.
2. Ensure that the **Data Modeling** Activity appears expanded on the DecisionSpace Data Quality Tree.
3. Double-click the **Perform Table Modeling** tool



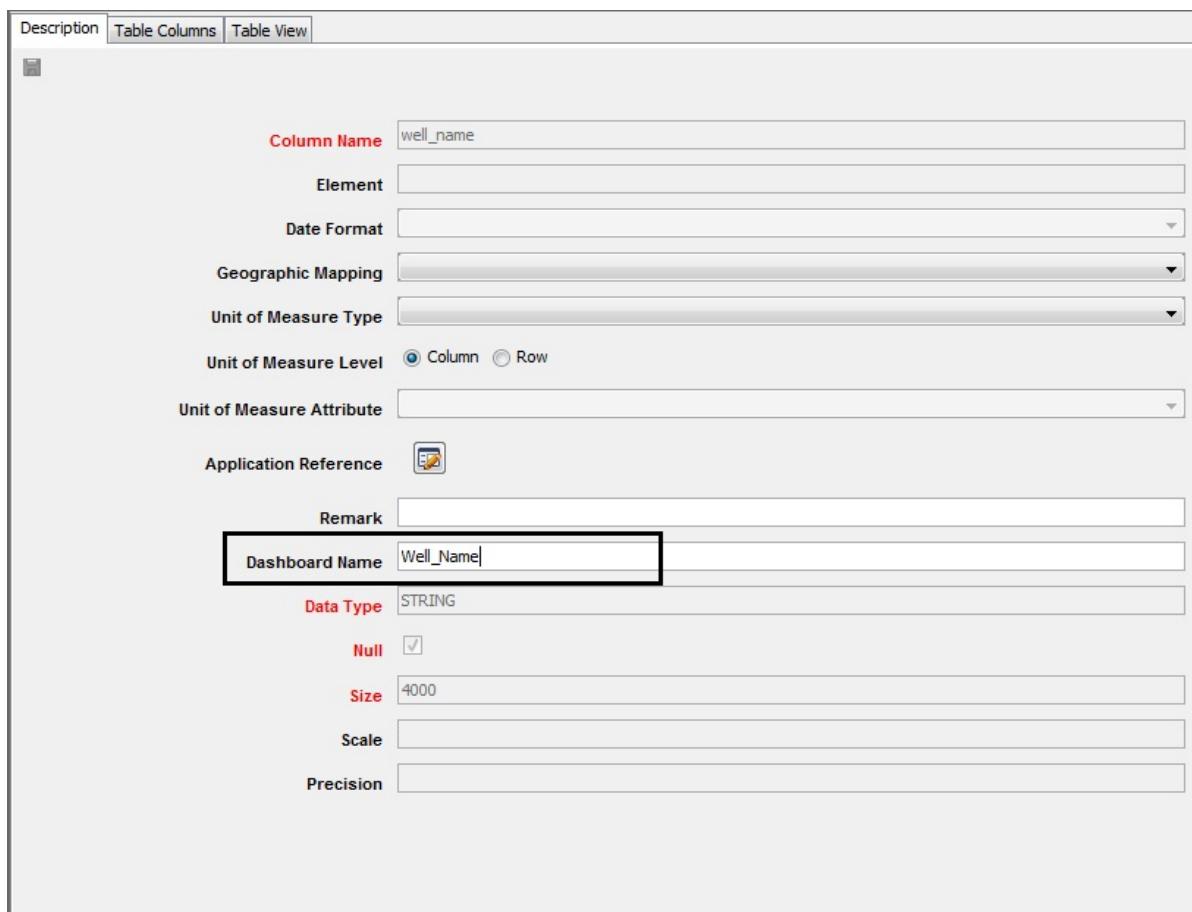
The **Perform Table Modeling** window appears.



4. Click  to expand the **Well** table from the **Data Model Tree** and select the **well\_name** column.



5. Select the **Description** tab from the **Model View** area on the right side of the **Perform Table Modeling** window.
6. On the **Description** tab, enter **Well\_Name** in the **Dashboard Name** field.



The screenshot shows the 'Perform Table Modeling' window with the 'Description' tab selected. The 'Table Columns' and 'Table View' tabs are also visible at the top. The 'well\_name' column is being configured. The 'Dashboard Name' field is highlighted with a black border and contains the value 'Well\_Name'. Other fields include 'Column Name' (well\_name), 'Element', 'Date Format', 'Geographic Mapping', 'Unit of Measure Type', 'Unit of Measure Level' (radio buttons for Column and Row), 'Unit of Measure Attribute', 'Application Reference' (with a browse icon), 'Remark', 'Data Type' (STRING), 'Null' (checkbox checked), 'Size' (4000), 'Scale', and 'Precision'.

7. Click  to save the specified column dashboard name.

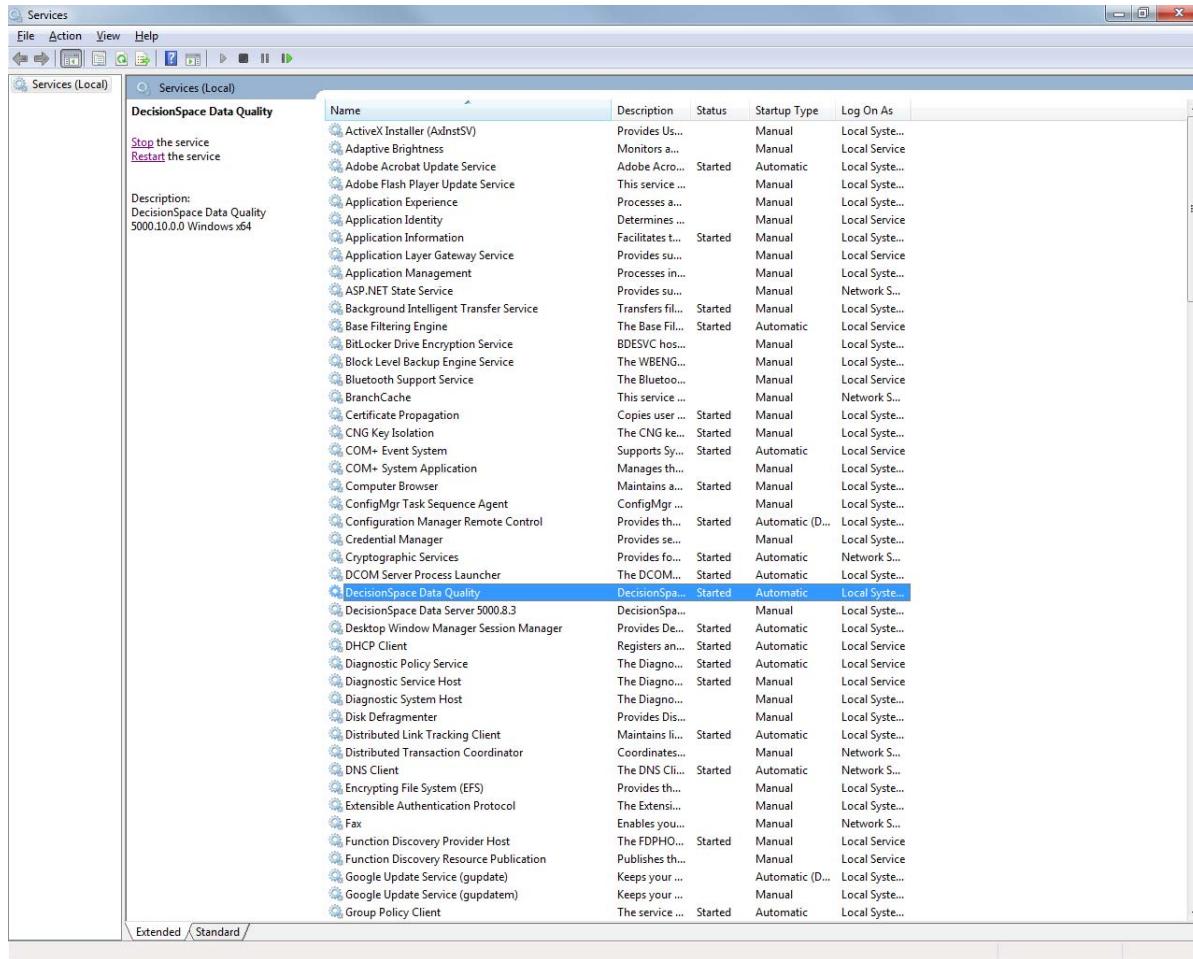
8. Select **File > Exit** from the menu bar to close the **Perform Table Modeling** window.

In order to immediately review any changes to the web dashboard, the DecisionSpace Data Quality Server must be refreshed.

## Updating Statistics on the Web Dashboard

In earlier exercises of this chapter, you published a submodel, configured column display groups and specified a dashboard name for a column. This section includes information about viewing these changes on the Web Dashboard. To refresh data on the Web Dashboard:

1. Click  on the Taskbar of your computer.
2. Enter **Services** in the **Search Programs and files** field of the Start menu  
The **Services** window appears.



3. Select **DecisionSpace Data Quality** from the **Name** column.
4. Click **Restart** to restart the DecisionSpace Data Quality service.

5. Select **File > Exit** to close the **Services** window.
6. Enter **http://localhost:8091** in the address bar of the web browser. The **Please wait. Your browser will be redirected when ready** message window appears.



**DSDQ is currently refreshing dashboard reports.**

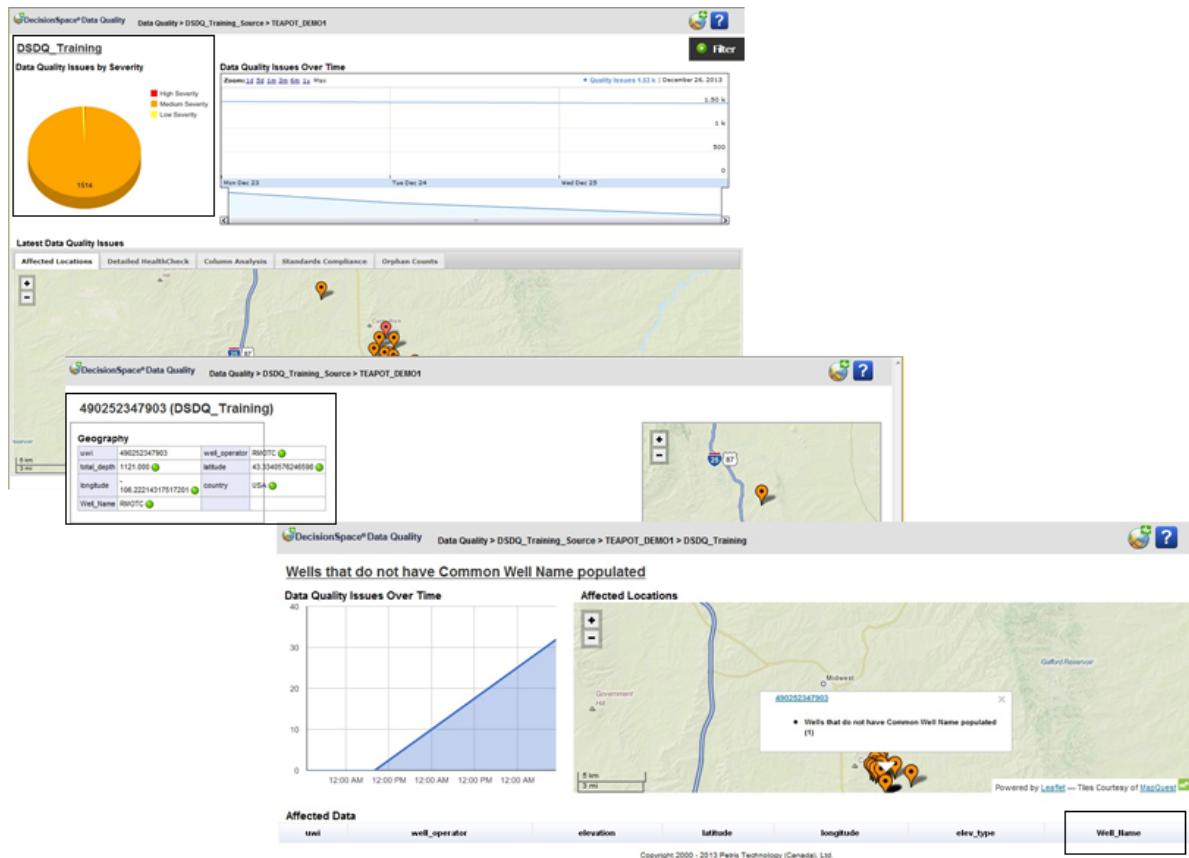
**Please wait. Your browser will be redirected when ready.**

The **Web Dashboard** is launched in your web browser.



7. Select **DSDQ\_Training\_Source** and then **DSDQ\_Training** from the Data Quality Web Dashboard to display the published submodel.
8. Select the **Detailed HealthCheck** tab and then a desired **Requirement** from the **Latest Data Quality Issues** area of the

Web Dashboard to view the Column Display Group and the dashboard name for the column..



# **Chapter 10**

# ***Project Administration***

DecisionSpace Data Quality requires users to have authentic user names and passwords to login to the application. The administrator user (created while installing the application) can create additional users when required.

The Advanced Scheduling tool enables the users to create scheduled events that are comprised of existing jobs that run in a sequence at a specified date and time. Multiple events (consisting of more than one jobs) can also be scheduled.

The Job Administrator tool allows the user to easily manage current job groups and review recently run job groups. This is helpful in determining problems (via real-time warnings), which results in enhanced application performance.

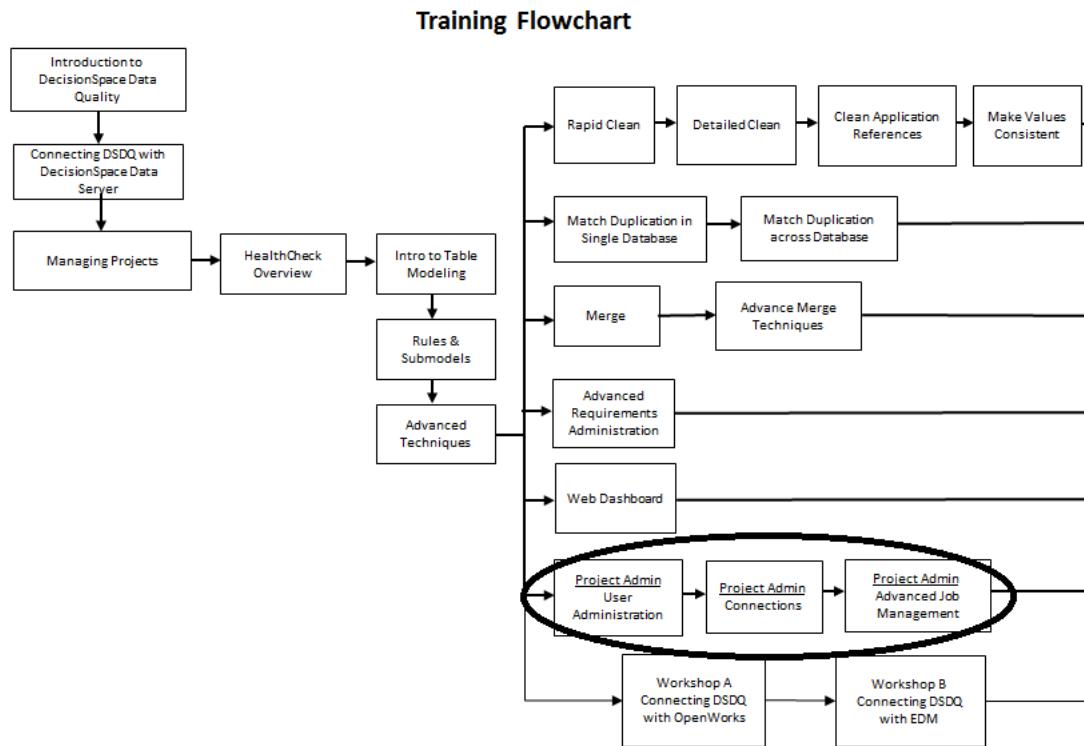
A DSDQ Administrator can also manage data owner and workspace connections, in order to provide access to the data sources.

# Chapter Overview

In this chapter, you will learn about:

- Managing users
- Setting up and configuring scheduled events
- Managing Jobs
- Managing Connections

Topics covered in each chapter are outlined in the following illustration. Those specific to the current chapter will be circled in black for your reference:



## Roles & Permissions

Users in the Data Quality application can be assigned the following roles:

- **Data Custodian:** A Data Custodian is responsible for ensuring safe transport and storage of data as well as maintaining the data infrastructure and business rules. This includes setting up and configuring roles, creating associations between columns and tables and creating connections to data stores for **Data Stewards**.
- **Data Steward:** A Data Steward is responsible for managing content in data stores (i.e. running jobs, confirming data matches and consolidating information) and controlling any modifications made to it. This user deals with the daily governance of a company's information and is often the subject matter expert on the governed data.

Roles And Permission	Data Custodian	Data Steward
Project Administration	execute	read
Project Tools	execute	locked
Unit of Measure Aliases	execute	read
Test Data	execute	read
Rapid HealthCheck	execute	read
Detailed HealthCheck	execute	read
Clean Application References	execute	read
Make Values Consistent	execute	read
Detailed Match	execute	read
Manage Duplication	execute	read
Setup and Manage Registry	execute	read
Setup and Manage Alias Set	execute	read
MasterSet HealthCheck	execute	read
Manage Master Records	execute	read
Merge Setup	execute	locked
Merge	execute	read
Advanced Scheduling	execute	read
Job Administrator	execute	read
Requirements Administrator	execute	locked

<b>Roles And Permission</b>	<b>Data Custodian</b>	<b>Data Steward</b>
Reference Data Administrator	execute	locked
Unit of Measure Administrator	execute	locked
Regular Expression Helper	execute	execute
Manage Users	Admin task	Admin task
Manage Project Access	Admin task	Admin task
Manage Connection Access	Admin task	Admin task

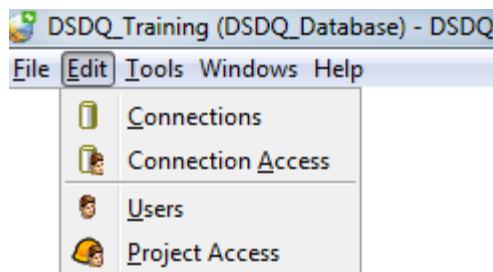
## Managing Users

The Manage Users window is used to create, copy, edit and remove a user. This functionality is only available to users with Administrator rights.

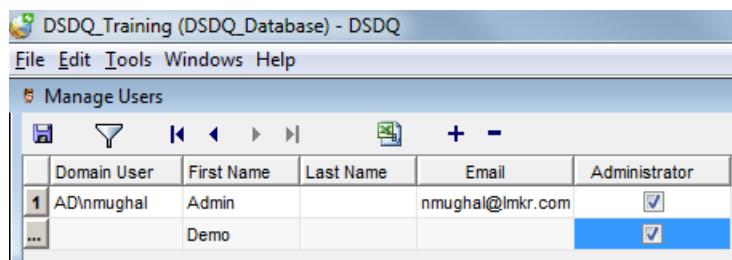
### **Exercise: Creating a User**

To create a user:

1. Select **Edit > Users** from the DecisionSpace Data Quality menu bar.

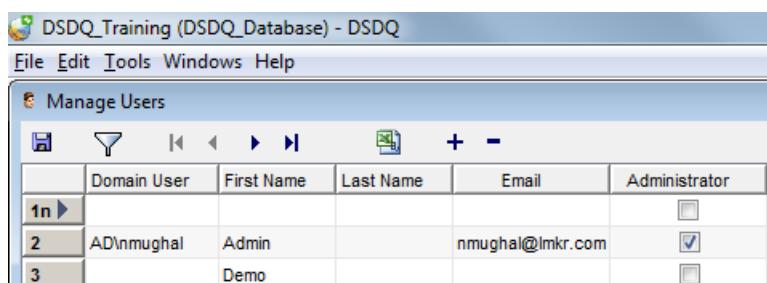


The **Manage Users** window appears.



2. Click **+**.

A new row is added to the table and the letter 'n' appears next to the record number.



3. Enter **Student** in the **User Name** field.

4. Enter **Student** in the **Password** field.
5. Enter **Student** in the **First Name** field.
6. Enter **Student** in the **Last Name** field.

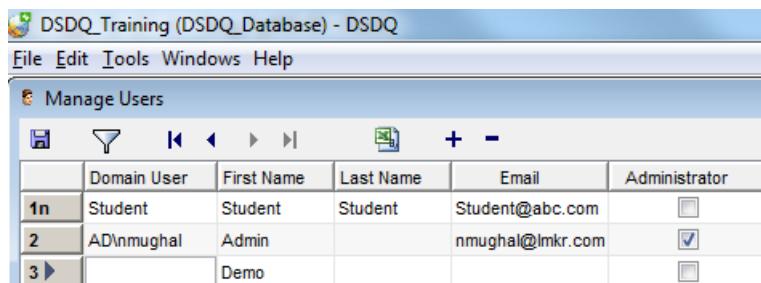
**Note**

The **User Name** and **Password** fields are case sensitive. **First Name**, **Last Name**, and **Email** fields are optional.

**Note**

Only select the **Administrator** check box if the user needs administrative rights.

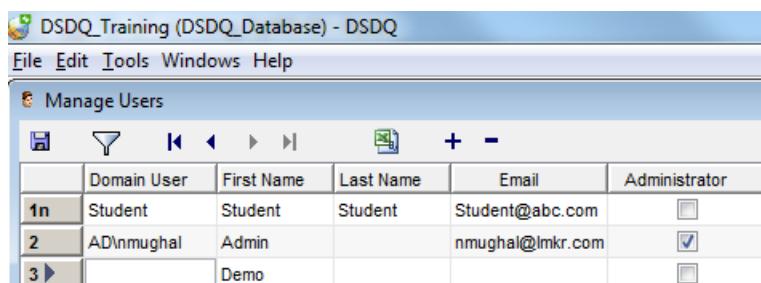
7. Click .



### **Exercise: Editing a User**

To edit a user:

1. Select **Edit > Users** from the DecisionSpace Data Quality menu bar. The **Manage Users** window appears.



2. Click the field of your choice to edit the information provided in it. The selected field is highlighted.

3. Click  to update the record.

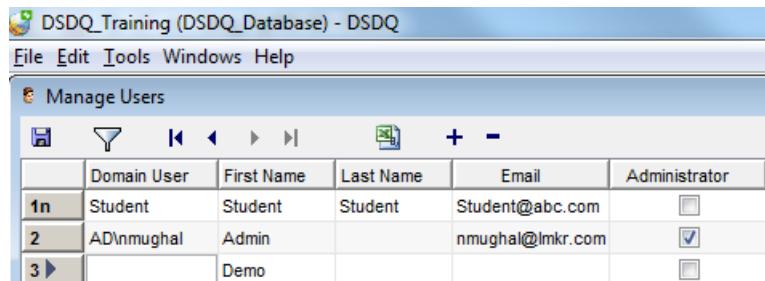
**Note**

If you do not click Save to update the record, you will be prompted to save your changes when closing the **Manage Users** window.

## Exercise: Removing a User

To remove a user:

1. Select **Edit > Users** from the DecisionSpace Data Quality menu bar. The **Manage Users** window appears.



2. Select the user that you want to delete. In this case, select **Student** from the **User Name** field.  
An arrow appears next to the record number.
3. Click .
4. Click  to save your changes.



**Note**

If a user has been given access to a specific project, a **Delete Selected User** warning dialog box appears.

5. Click **Yes** to confirm deletion or **No** to cancel it.

## **Advanced Scheduling**

---

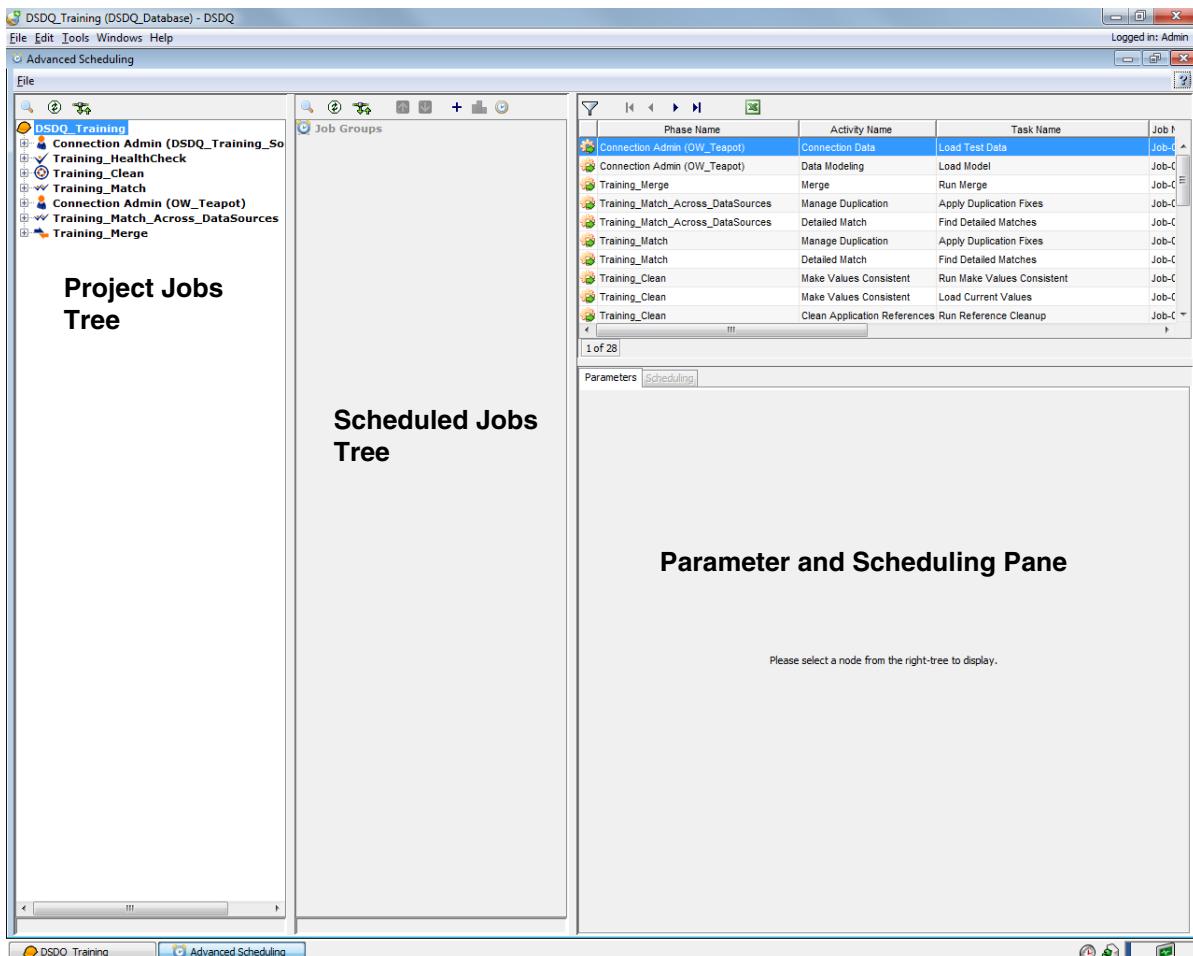
The **Advanced Scheduling** tool is used to set up and configure scheduled events (that consist of current jobs) to run in sequence. You can use this tool to schedule multiple events for multiple jobs.

### ***Exercise: Adding a Scheduled Event***

The **Scheduled Jobs Tree** lists all the jobs that the user has created with the associated Task, Activity, Phase, and Project. Selecting the project name will display all the jobs under that project. On selecting a job, details of that job are displayed in the **Parameter and Scheduling Pane** under the **Parameters** tab. This area cannot be edited from the Advanced Scheduling Tool. Edits can only be made from the Project window.

To add a scheduled event:

1. Click **Tools > Advanced Scheduling** from the DecisionSpace Data Quality menu bar.



The **Advanced Scheduling** window appears listing all the tables and columns in the database. The left pane displays the **Project Jobs Tree** and the center pane displays the **Scheduled Jobs Tree**. Both trees are color coded and the active tree has a white background. You can create a scheduled event to run existing jobs in a specific order at a specified date and time. **The Advanced Scheduling tool** is project specific and a project needs to be opened prior to accessing it.

2. Right-click **Job Groups** in the **Scheduled Jobs Tree** and select **Add** from the pop-up menu. Alternatively, click **+** on the

**Scheduled Jobs tree toolbar.**

The **New Job Group** dialog box appears.



3. Enter **Schedule Job 01** in the **New Job Group** dialog box.

4. Click **OK**.

The job is added and displays under **Job Groups** in the **Scheduled Jobs Tree**.

**Note**

To remove a **Job Group** from the **Scheduled Jobs Tree**, right-click the desired Job Group and select **Remove** from the pop-up menu. A confirmation dialog box will appear. Click **Yes** to remove the Job Group or **No** to keep it. To duplicate a Job Group, right-click the desired Job Group and select **Duplicate** from the pop-up menu.

### **Exercise: Adding Jobs to Events**

To add a job to events:

1. Click **+** on the DecisionSpace Data Quality Tree to expand the **Training\_HealthCheck** Phase.
2. Click **+** to expand the **Rapid HealthCheck** Activity.
3. Click **+** to expand the **Profile Data Using SQL Query** Task.
4. Drag **Job-01** from the Project Jobs tree in the left pane to **Schedule Job 01** event in the Scheduled Jobs tree in the center pane.  
The selected job is added to the Event.

5. Optionally, change the order of jobs within the event by using the Up Arrow & Down Arrow buttons.

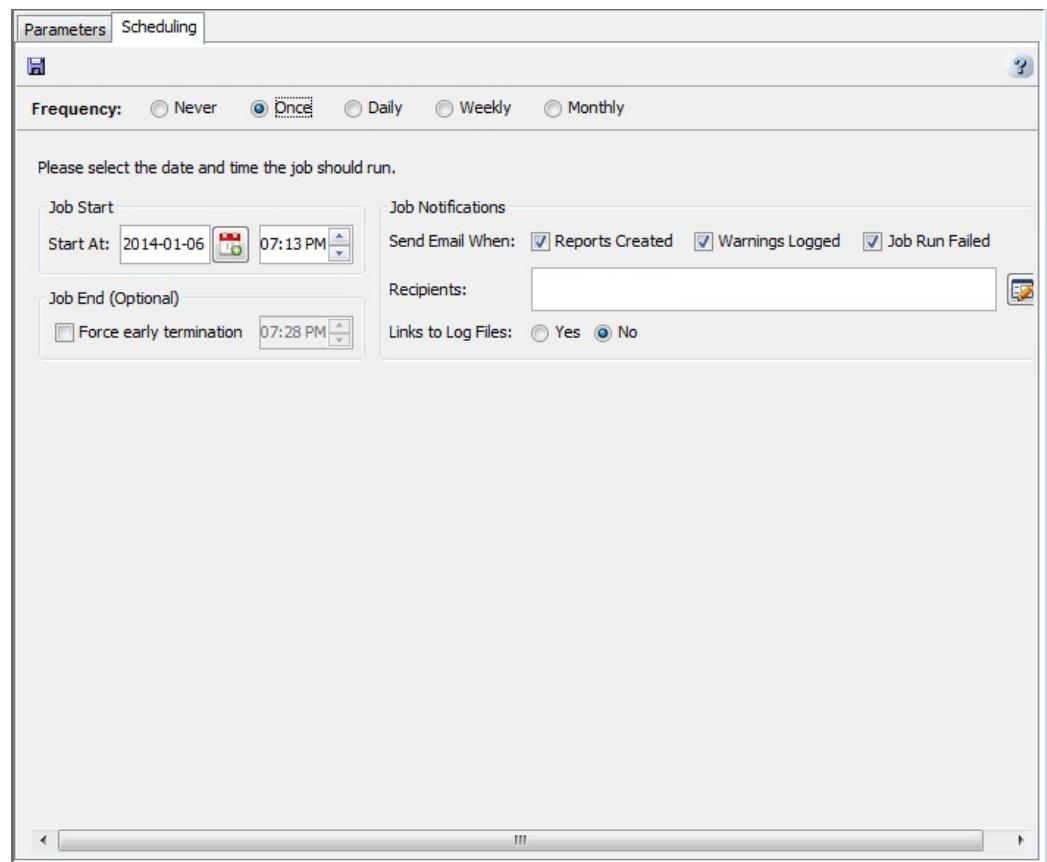
**Note**

To remove a Job from the Job Group, right-click the desired job and select **Remove** from the **Pop-up** menu. A confirmation dialog box will appear. Click **Yes** to remove the Job or **No** to cancel the remove operation.

## Exercise: Scheduling Events

To schedule an event:

1. Select **Job 01** created in the last exercise.  
The **Scheduling** tab automatically displays scheduling options. Events can be scheduled just like jobs. You can also schedule an event to run multiple jobs at a specified schedule in a specified order.



2. Select the **Once** option from the **Frequency** field.
3. Select the date and time when you want the job to start from the **Job Start** field.
4. Select all the check boxes in the **Send Email When** field.
5. Enter your email address in the **Recipients** field.  
An email will be sent to you when the job is completed.
6. Select the **No** option from the **Links to Log Files** field.

7. Click .

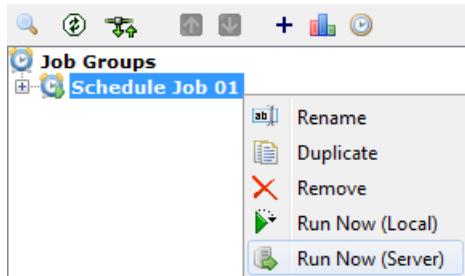
**Note**

The Data Quality Server must be installed, correctly configured, and running in order to schedule events.

### **Exercise: Running a Job Group on the Server**

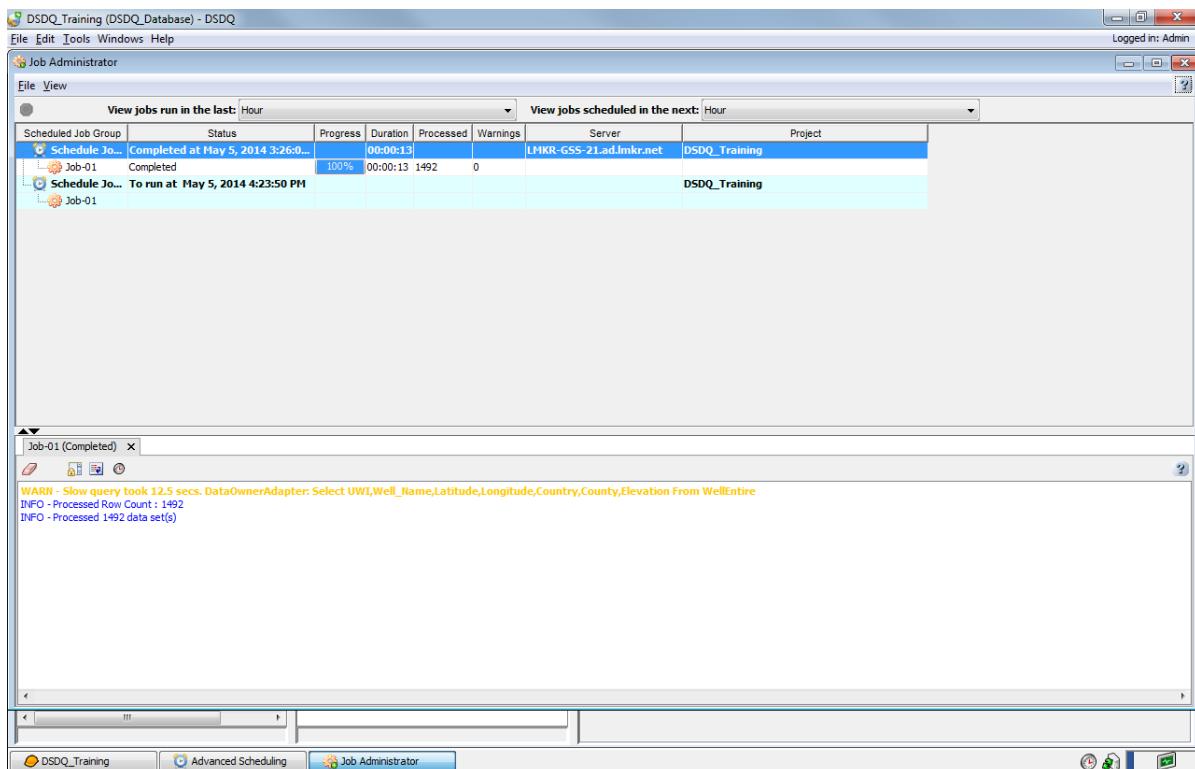
To run a job group on the server:

1. Right-click **Job-01** and select **Run Now (Server)** from the pop-up menu.



The **Job Administrator** window appears. The Data Quality application sends the selected job group to the server to be run, and the **Job**

**Administrator** window shows you the progress of the running job.



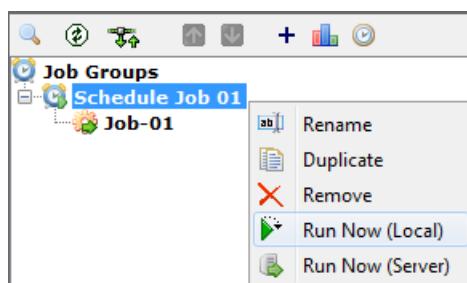
#### Note

The Data Quality Server must be installed, correctly configured, and running in order to run Job Groups.

## Exercise: Run Job Group Locally

To run a job group locally:

1. Right-click **Job-01** and select **Run Now (Local)** from the pop-up menu.



The **Job Administrator** window appears. The Data Quality application sends the selected job group to the local machine run,

and the **Job Administrator** window shows you the progress of the running job.

Scheduled Job Group	Status	Progress	Duration	Processed	Warnings	Server	Project
test	Failed at Dec 27, 2013 8:29:20 PM	00:00:00			0	LMKR-...	DSDQ_Training
Job-01	Failed	00:00:00			0	LMKR-...	DSDQ_Training
test	Failed at Dec 27, 2013 8:40:24 PM	00:00:00			0	LMKR-...	DSDQ_Training
Job-01	Failed	00:00:00			0	LMKR-...	DSDQ_Training
Job-01	Failed at Dec 27, 2013 8:40:24 PM	00:00:00			0	LMKR-...	DSDQ_Training
Job-01	Failed	00:00:00			0	LMKR-...	DSDQ_Training
test	Failed at Dec 27, 2013 8:40:24 PM	00:00:00			0	LMKR-...	DSDQ_Training
Job-01	Failed	00:00:00			0	LMKR-...	DSDQ_Training
Job-01	Failed at Dec 27, 2013 8:40:44 PM	00:00:00			0	LMKR-...	DSDQ_Training
Job-01	Failed	00:00:00			0	LMKR-...	DSDQ_Training
Job-01	Failed at Dec 27, 2013 8:52:12 PM	00:00:00			0	LMKR-...	DSDQ_Training
Job-01	Failed	00:00:00			0	LMKR-...	DSDQ_Training
Job-01	Failed at Dec 27, 2013 8:54:30 PM	00:00:00			0	LMKR-...	DSDQ_Training
Job-01	Failed	00:00:00			0		
Job-01	To run at Dec 28, 2013 10:00:00 AM						
Job-01	To run at Dec 28, 2013 12:27:33 PM						
Job-01	To run at Dec 29, 2013 10:00:00 PM						
Job-01							

#### Note

Job Groups can also be run locally by clicking the Run Now (Local)  button in the **Parameters** tab for a Job Group.

## Viewing Job Group Charts

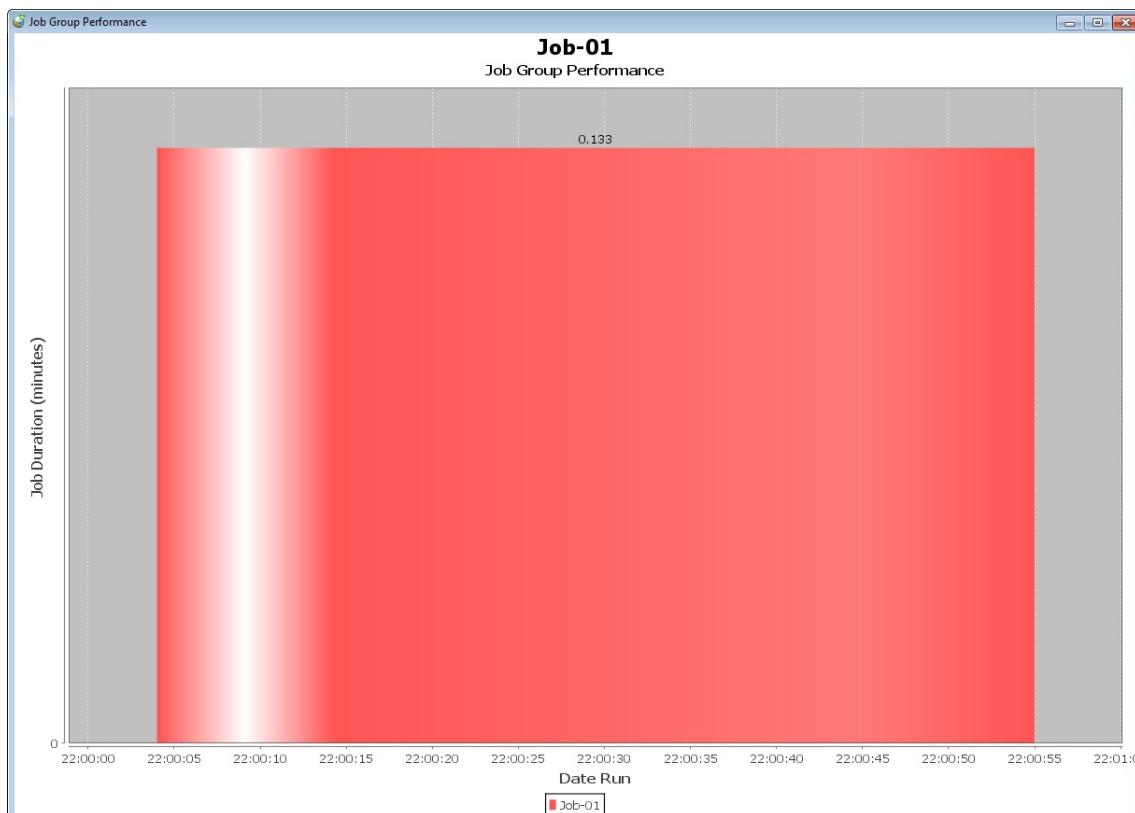
---

The **Job Group Charts** is used to view the performance of individual Jobs within Job Groups as well as the performance of Job Groups.

### **Exercise: Viewing Job Group Performance**

To view job group performance:

1. Select a Job Group and click on the **View Job Group Performance**  button located on the **Scheduled Jobs Tree** menu bar.  
The Job Group Performance chart displays.



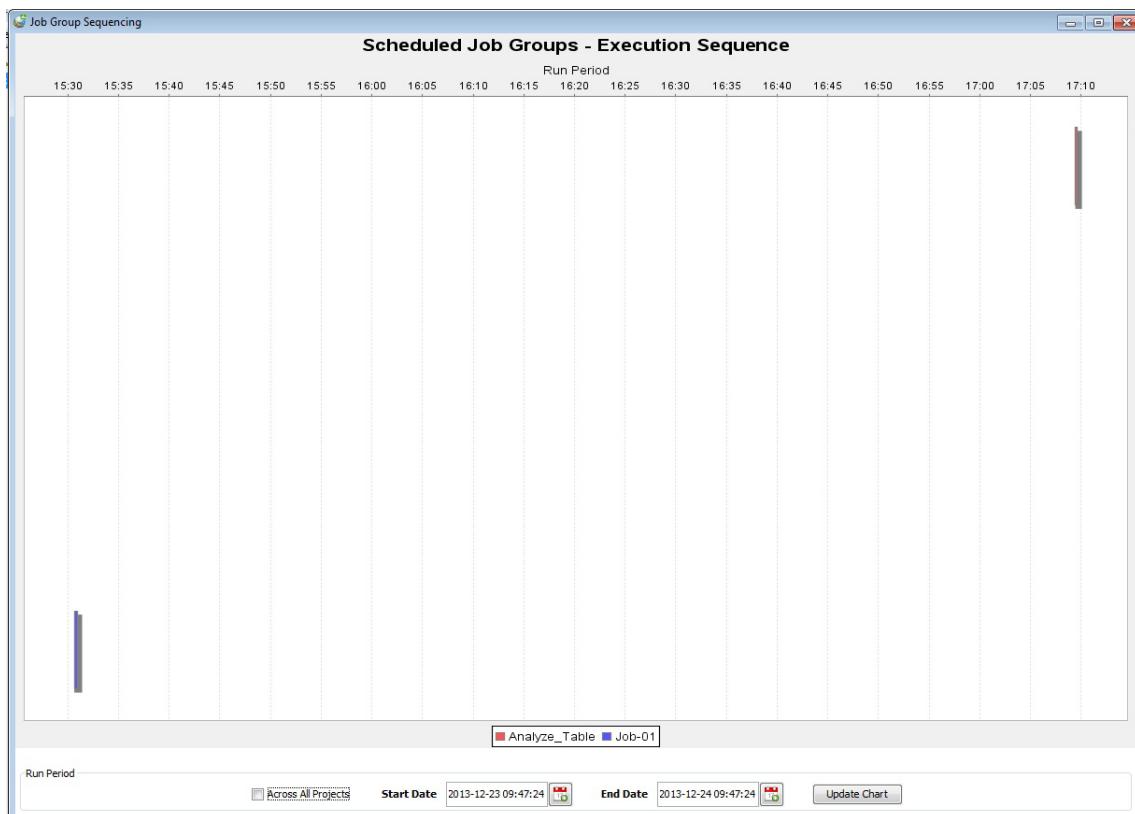
#### Note

Bars on the chart represents the time when the selected job group was run, with each job within the selected job group denoted by a unique color.

## Exercise: Viewing Job Group Sequencing

To view job group sequencing:

1. Click the **View Job Group**  button located on the Scheduled Jobs Tree menu bar.  
The **Job Group Sequencing chart** appears.



---

## Managing Jobs

---

Jobs scheduled in DecisionSpace Data Quality are managed through the **Job Administrator** tool. This tool is used to easily manage running job groups and review recently run job groups. This helps in determining if there are any problems with the job group itself (via the real-time response of warnings) and any application performance issues that may be caused by the running job groups.

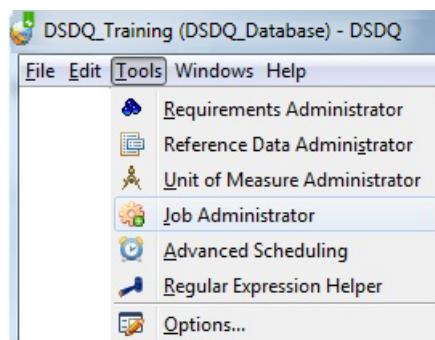
Jobs have the following three states:

- **Running:** The Job is being processed right now.
- **Completed:** The Job has been processed and results have been produced.
- **Not Responding:** The Job did not process properly and has produced an incomplete Job result.

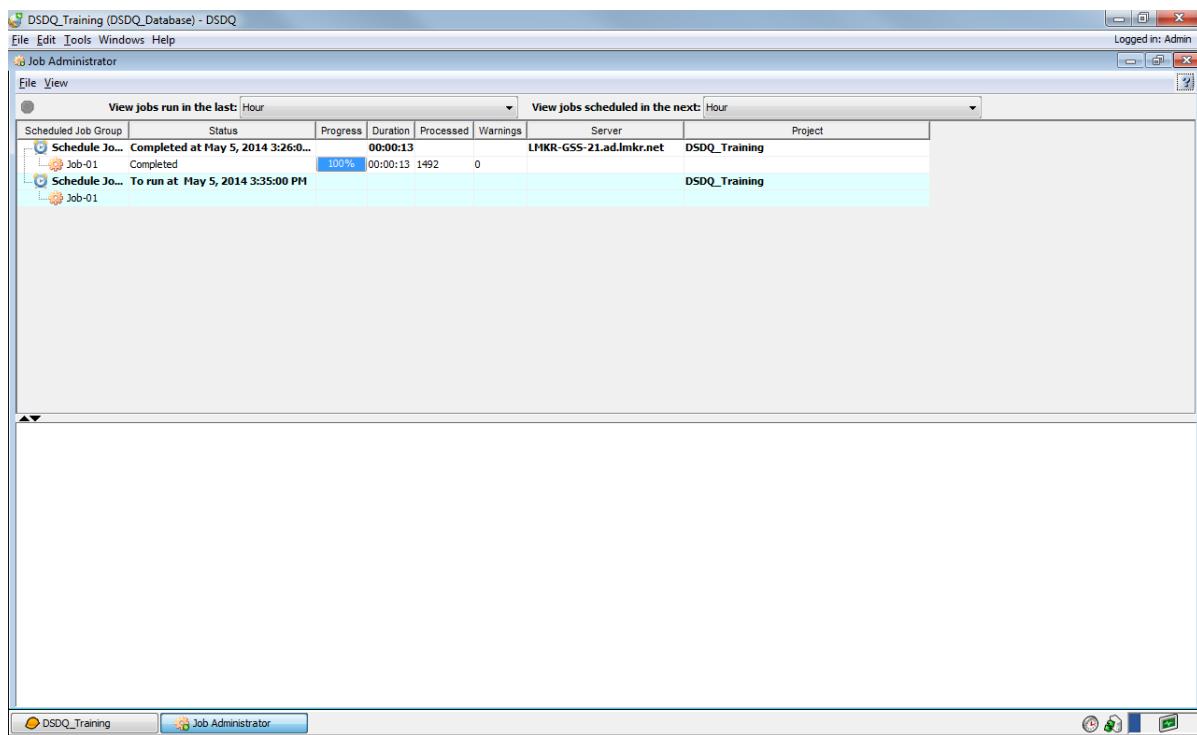
## Exercise: Opening Job Administrator

To open Job Administrator:

1. Click Tools > Job Administrator from the DecisionSpace Data Quality menu bar.



The **Job Administrator** window appears.



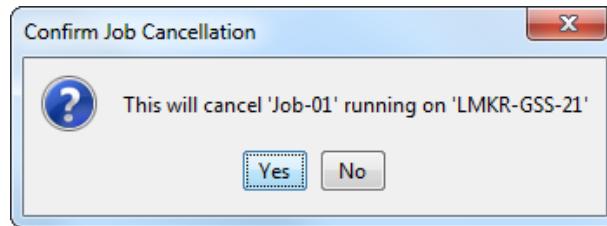
### Note

The progress card percentage is calculated based on the last time the job was run. If the job has never been run before, or parameters in the job have changed, the progress bar may have discrepancies.

## **Exercise: Stopping Current Running Jobs**

To stop a running job:

1. Select the record for the target job or job group.  
A **Stop**  button becomes available on the menu.
2. Click the **Stop**  button.  
The **Confirm Job Cancellation** dialog box appears.



3. Click **Yes** to confirm cancellation of the Job.  
The status of the job changes to **Stopping** and moves to the **recently run jobs** window.

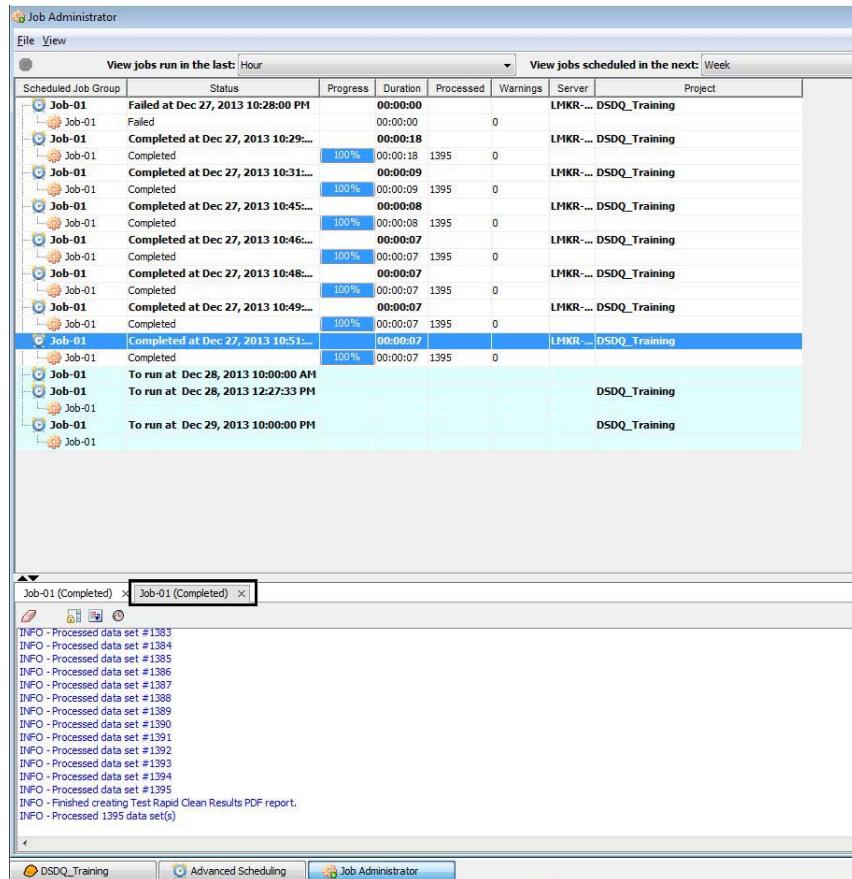
**Note**

Users with access to other projects have access (across the network) to the current running jobs and recently run jobs in that specific project.

## Exercise: Viewing Server Logs

To view server log:

- From the **Job Administrator** window, select the tab of **Job-01(Completed)**.



The bottom window is populated with log information for the job.

- Click the **Wrap Lines** button to align the text to the visible window and to carry the text over to the next line if it is too long.

### Note

To close the completed Job logs, click **View > Close Completed Job Logs** on the Data Quality menu bar of the **Job Administrator** window.

---

---

## **Managing Connections**

---

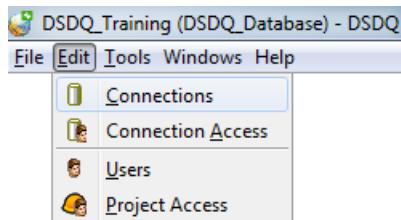
Every connection consists of three Source Dataset elements - Data Source, Workspace and Data Model. The application reads data from the data source. The Workspace is the schema that the application results are written to. The data model is the location to store detailed information about the data source tables and columns, including table relationships and element assignments.

Administrative users have rights to manage connections.

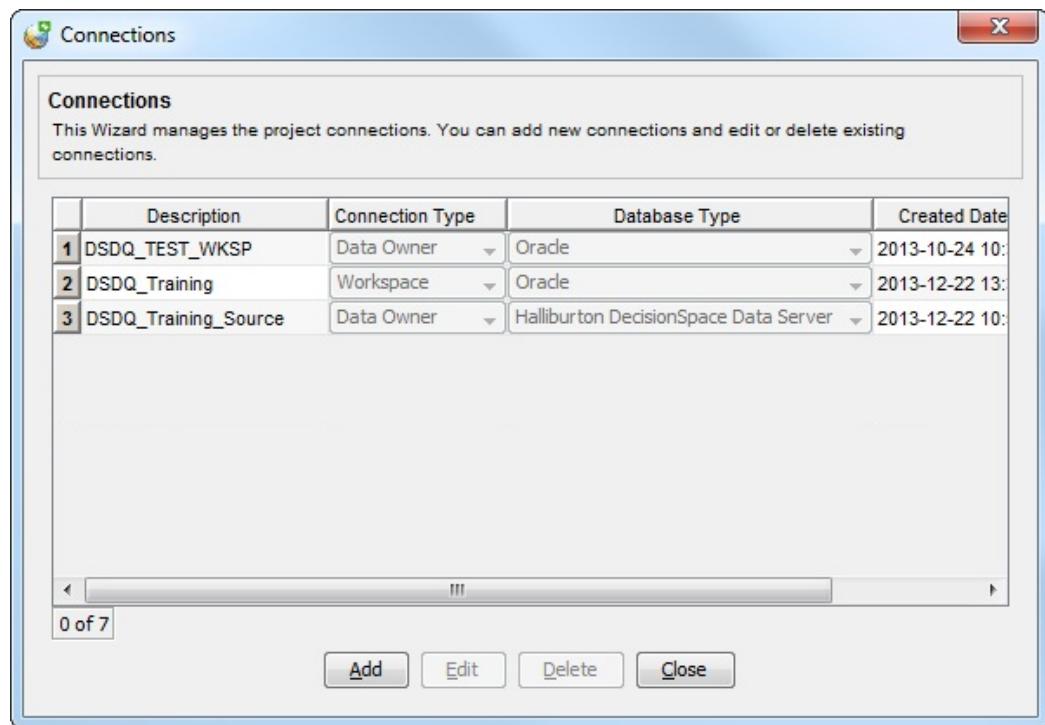
## Exercise: Editing a Data Owner Connection

To edit a data owner connection:

1. Select **Edit > Connections** from the DecisionSpace Data Quality menu bar.

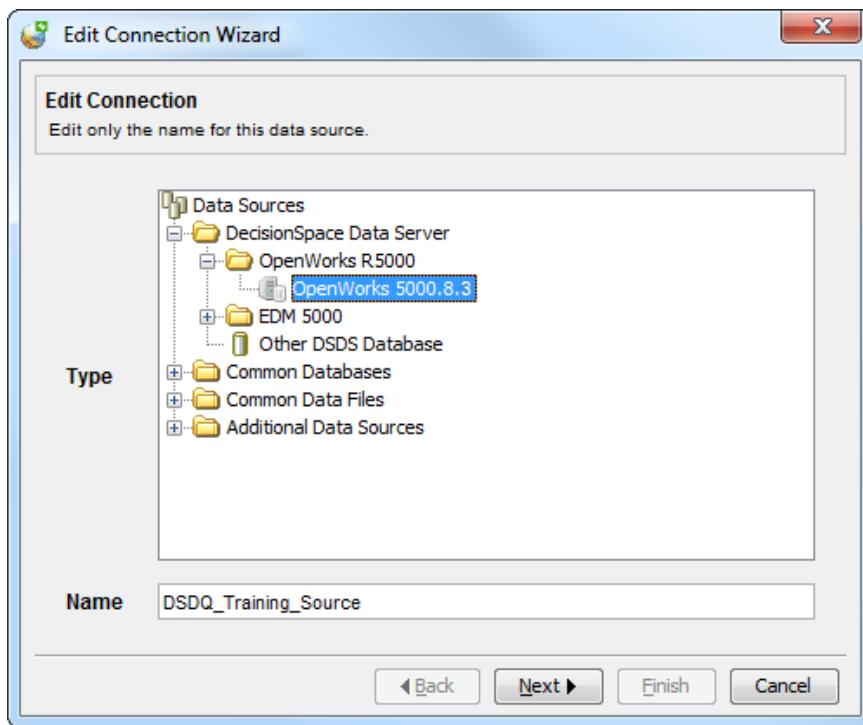


The **Connections** window appears.



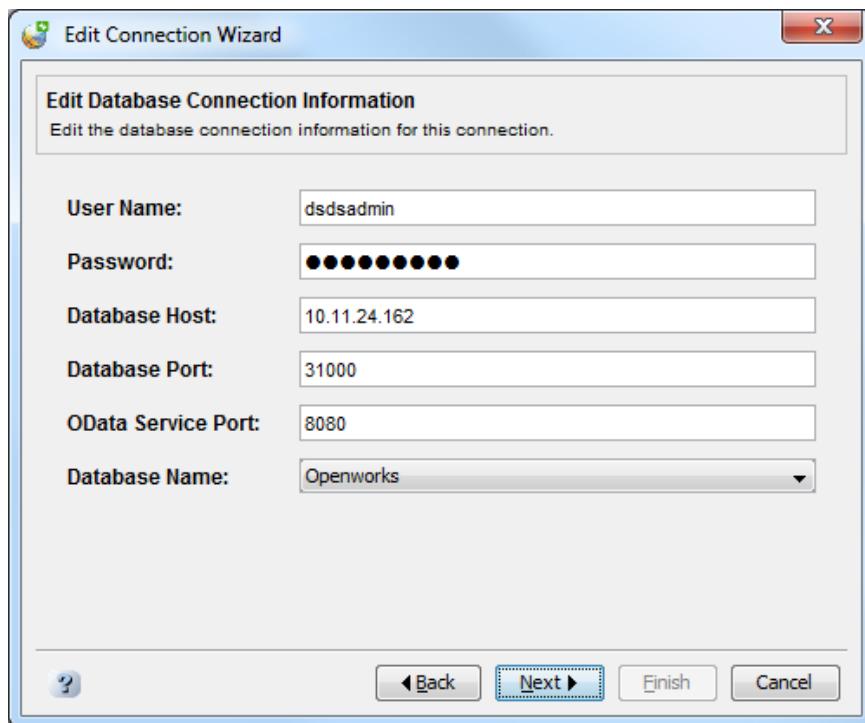
2. Select **DSDQ\_Training\_Source** and click **Edit**.

The **Edit Connection** window appears.



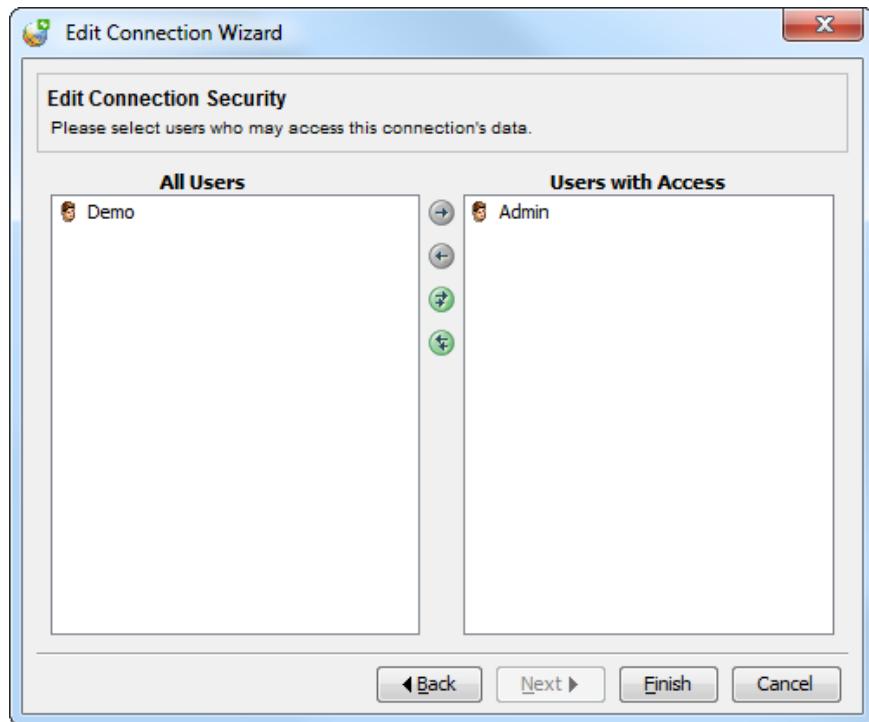
3. Enter DSDQ\_Training\_Source1 in the **Name** field and click **Next**.

The **Edit Database Connection Information** window appears.



4. Click **Next**.

The **Edit Connection Security** window appears.

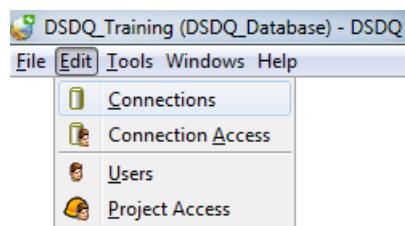
5. Edit the users who have access to the Connection and click **Finish**.**Note**

System Administrators can edit all information for connections.

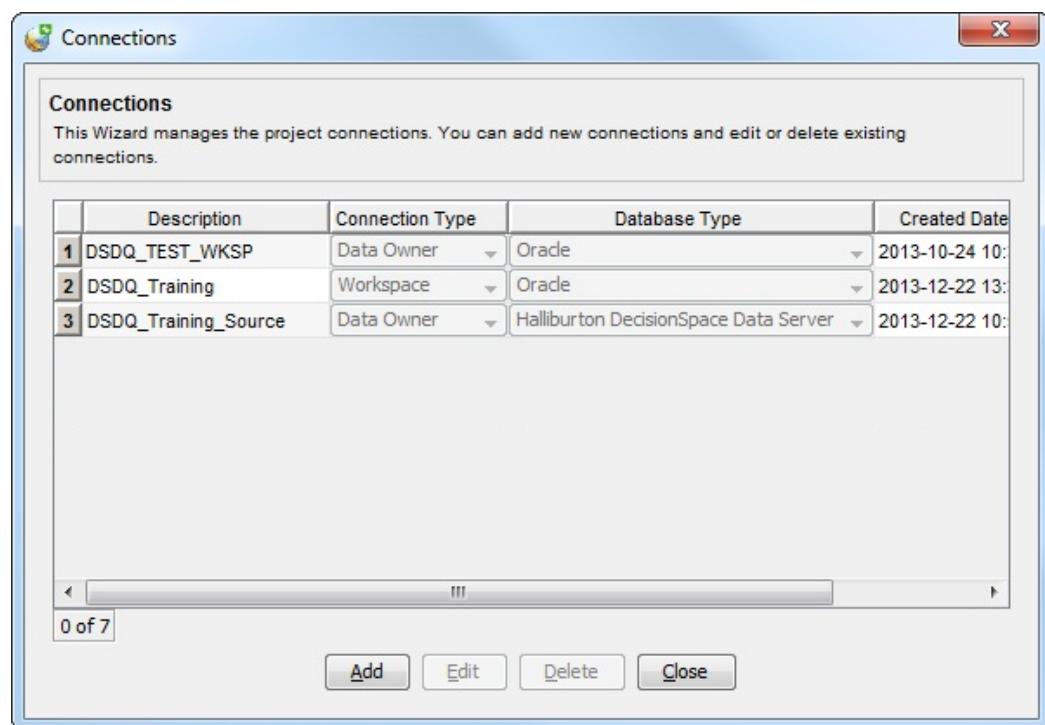
## Exercise: Editing a Workspace Connection

To edit a data owner connection:

1. Select **Edit > Connections** from the DecisionSpace Data Quality menu bar.

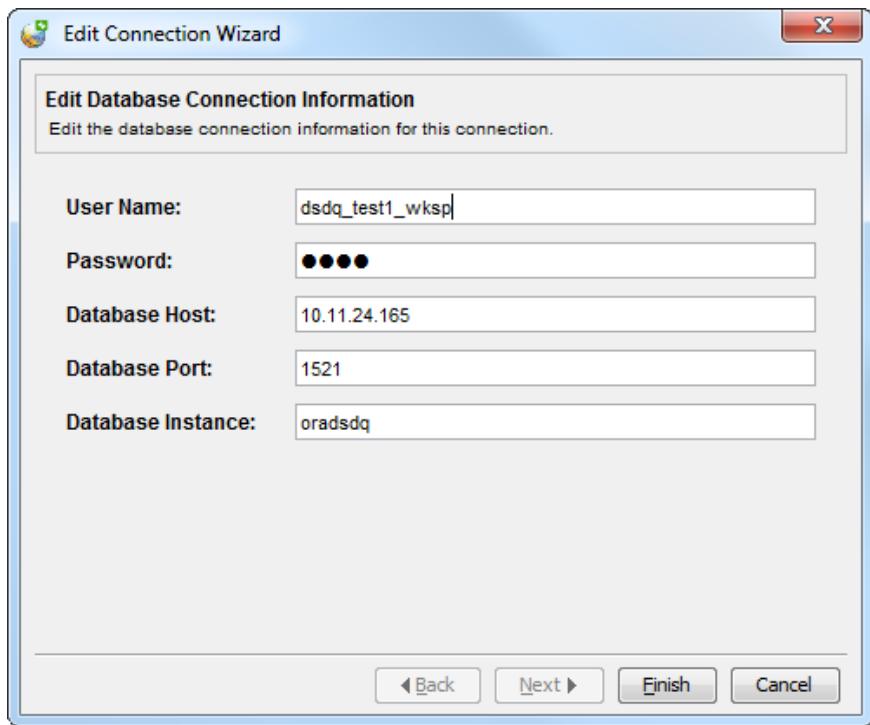


The **Connections** window appears.



2. Select **DSDQ\_Training** and click **Edit**.

The **Edit Database Connection Information** window appears.



3. Edit information for the Workspace Connection as per your requirement.
4. Click **Finish**.

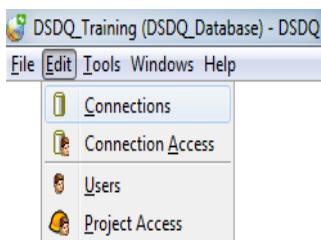
**Note**

Since editing a Connection has an effect on all existing Projects that use the Connection; it is advised to only keep the Project window open while editing the Connection.

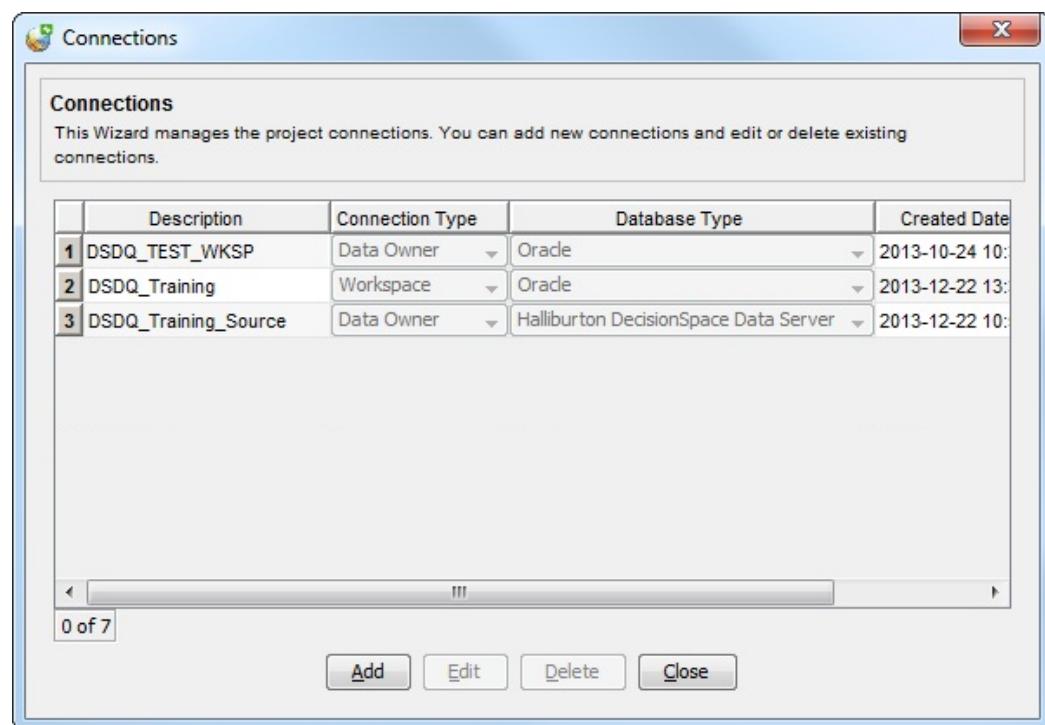
## Exercise: Removing a Data Owner/Workspace Connection

To remove a connection:

1. Select **Edit > Connections** from the DecisionSpace Data Quality menu bar.



The **Connections** window appears.



2. Select **DSDQ\_Training** and click **Delete**.

The selected connection will be deleted from the database.

## **Workshop A**

# ***Connecting DSDQ to an OpenWorks Data Source***

This section of the DecisionSpace Data Quality training manual aims at walking you through the process of:

- Accessing data stored in an OpenWorks data source
- Profiling data to identify the full spectrum of data quality issues
- Addressing all such data issues by means of quality control queries and a repeatable cleansing methodology
- Validating corrections made to the data in the OpenWorks data source, and
- Finally, viewing the data quality results in the DecisionSpace Data Quality Web Dashboard

---

## Overview

---

In this workshop, you will perform the following exercises to evaluate, clean and standardize data from an OpenWorks data source in DecisionSpace Data Quality.

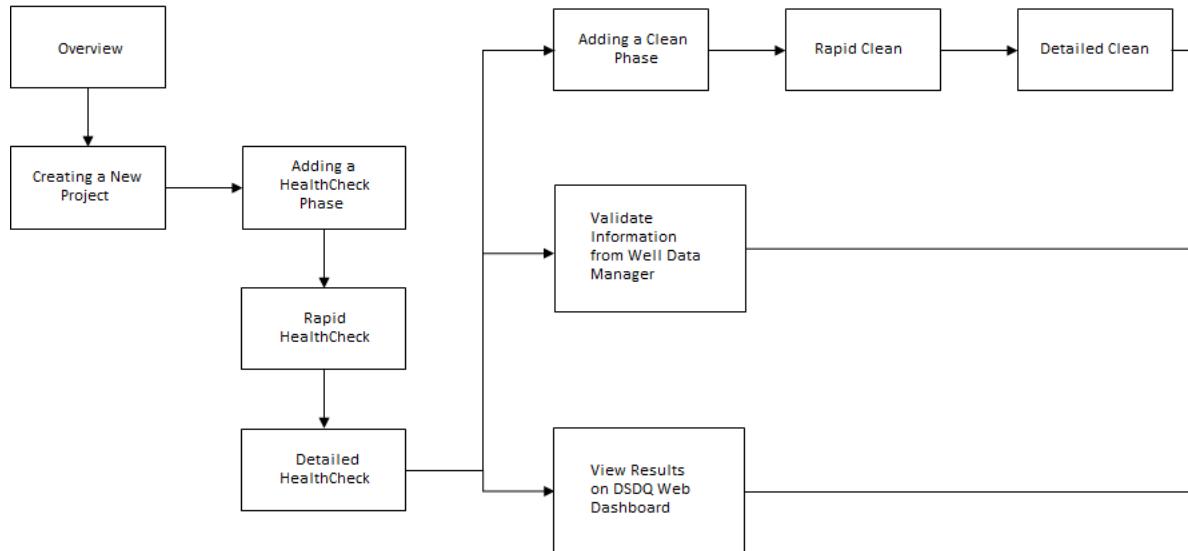
- Creating a New Project
- Evaluating data using the HealthCheck Phase
  - a) Rapid HealthCheck
    - Running Table Analysis on All Tables
    - Running Column Analysis on Columns
    - Running Table Analysis on Modeled Tables
    - Running Column Analysis on Modeled Tables
  - b) Detailed HealthCheck
    - Configuring the Detailed HealthCheck Tool
    - Running the Detailed HealthCheck Task
- Resolving data quality issues using the Clean Phase
  - a) Adding a Clean Phase
  - b) Rapid Clean
    - Configuring the Rapid Clean Tool
    - Running the Test Rapid Clean Task
    - Running the Rapid Clean Task
  - c) Detailed Clean
    - Configuring the Detailed Clean Tool

— Running the Detailed Clean Task

- Validating corrected data in the OpenWorks data source
- Viewing the data quality results in the DecisionSpace Data Quality Web Dashboard

The purpose of this workshop is to reinforce what you have learnt in previous sections of this manual in a single workflow. Topics covered in each section of the workflow are outlined in the following illustration:

**Workflow for Connecting DecisionSpace Data Quality to an OpenWorks Data Source**



## Creating a New Project

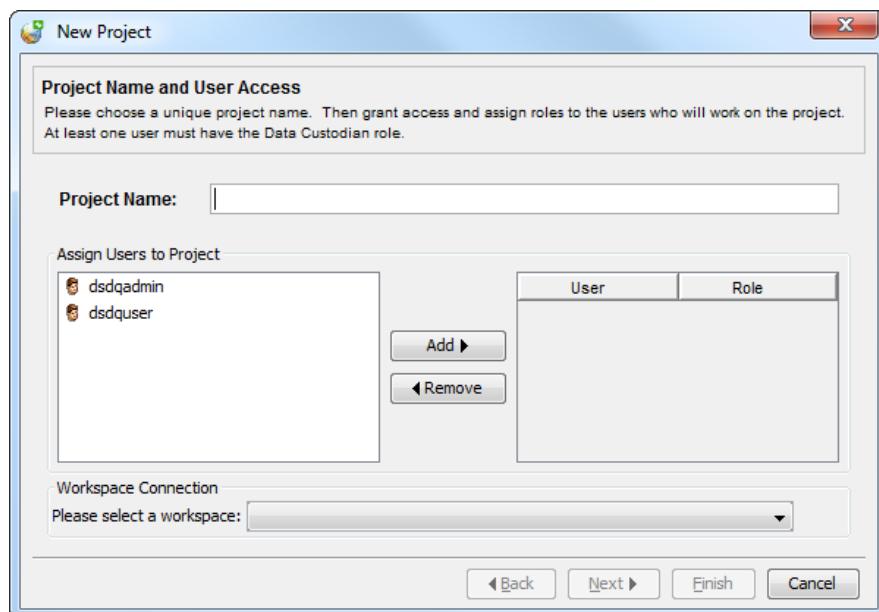
---

A DSDQ project comprises of all the Phases, Activities, Tools etc. During this process, you will assign users to the project and their roles while they work on it; select a Workspace Connection (the database where results will be written), the desired Phase and a Source Connection (data source that the application reads from).

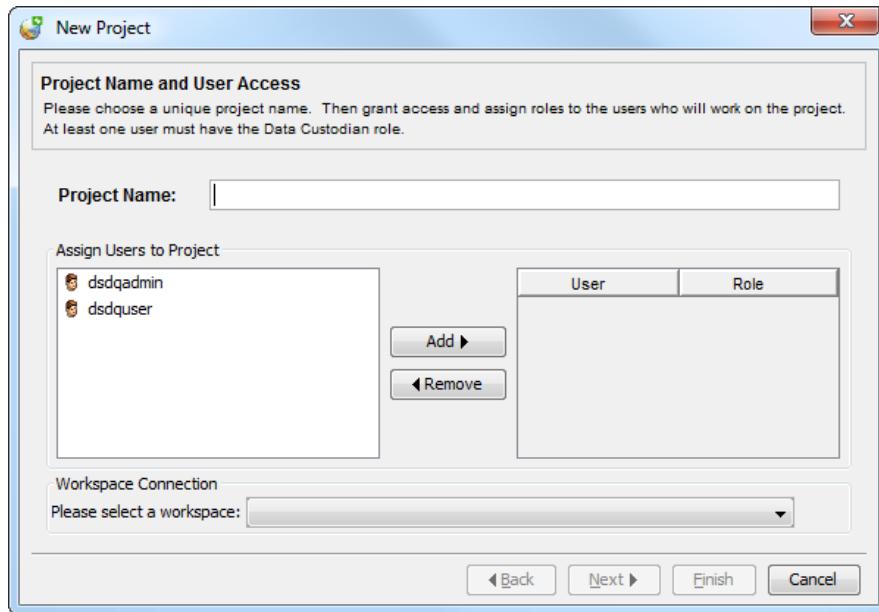
A new Project can be created:

- When the software is initially installed and a project does not exist.
- By selecting **New Project** from the File menu.
- By clicking the **New** button in the **Open an Existing Project** window.

In all instances, the **New Project** window appears displaying all available users.



1. Select **File > New Project** from the menu bar on the **DSDQ Project** window.  
The **New Project** window appears.



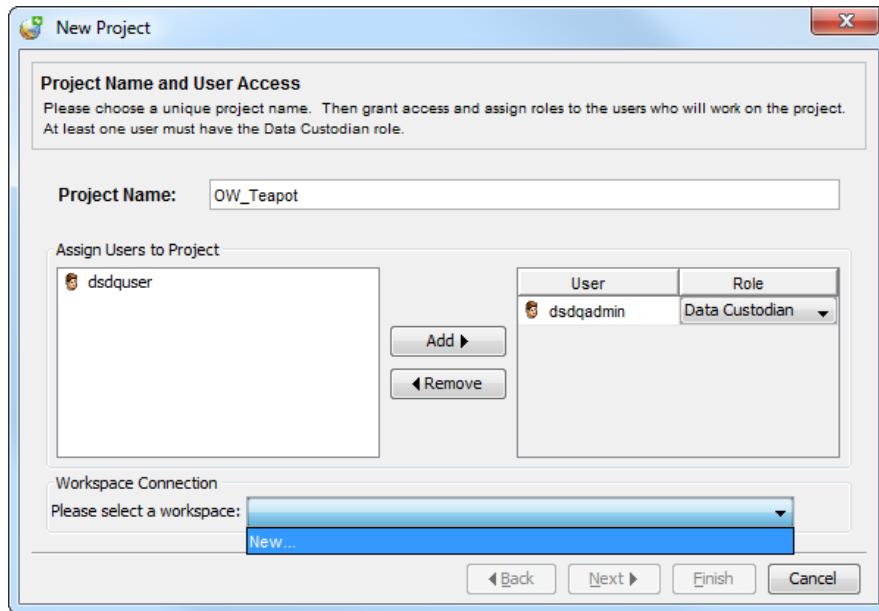
2. Enter **OW\_Teapot** in the **Project Name** field.

**Note**

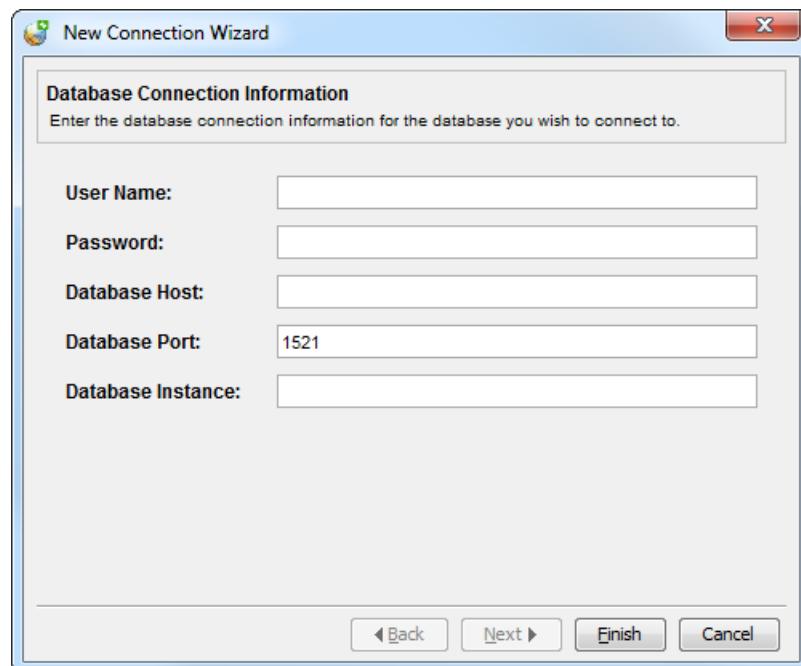
You will be using **OW\_Teapot** throughout this workflow.

3. Select **dsdqadmin** from the **Assign Users to Project** group box.

4. Click the **Add**  button to assign project access to the selected user.



5. Select **Data Custodian** from the **Role** drop-down list.
6. Select **New...** from the **Please select a workspace** drop-down list.  
**The New Connection Wizard - Database Connection Information** window appears.

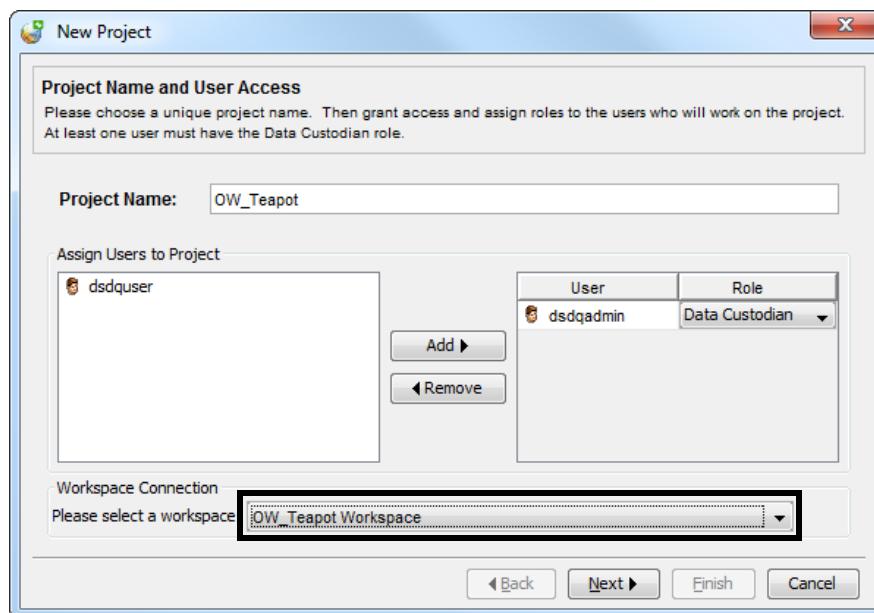


7. Enter **DSDQ\_OW\_Teapot\_WKSP** in the **User Name** field.

**Note**

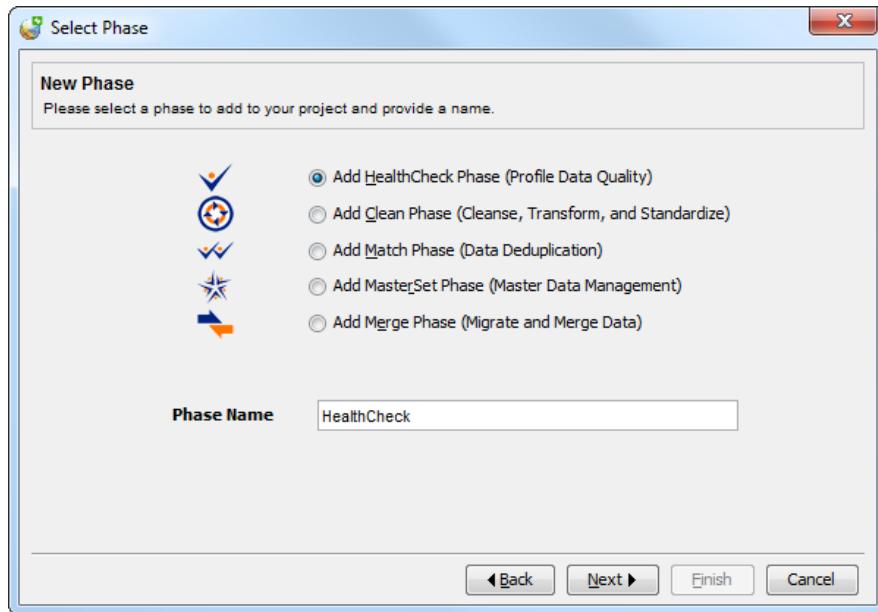
Make sure that the user and database connection have already been setup in Oracle (*reference: chapter 1, DSDQ Training Manual*).

8. Enter **DSDQ** in the **Password** field.
9. Enter **localhost** in the **Database Host** field.  
The **Database Port** is set to **1521** by default. If DecisionSpace Data Server connects to a different port, this number will need to be updated.
10. Enter **oradssdq10** in the **Database Instance** field.
11. Click **Finish**.  
You will notice that the newly created workspace **OW\_Teapot\_Workspace** is populated in the **Please select a workspace** drop-down list on the **New Project** window. \

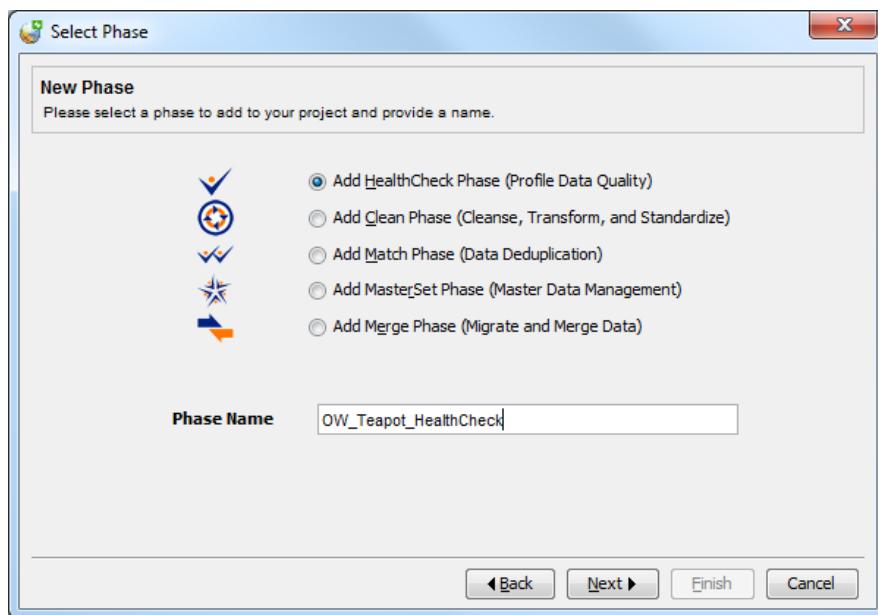


12. Click **Next** to continue.

The **Select Phase** window appears with the **Add HealthCheck Phase (Profile Data Quality)** option selected by default.

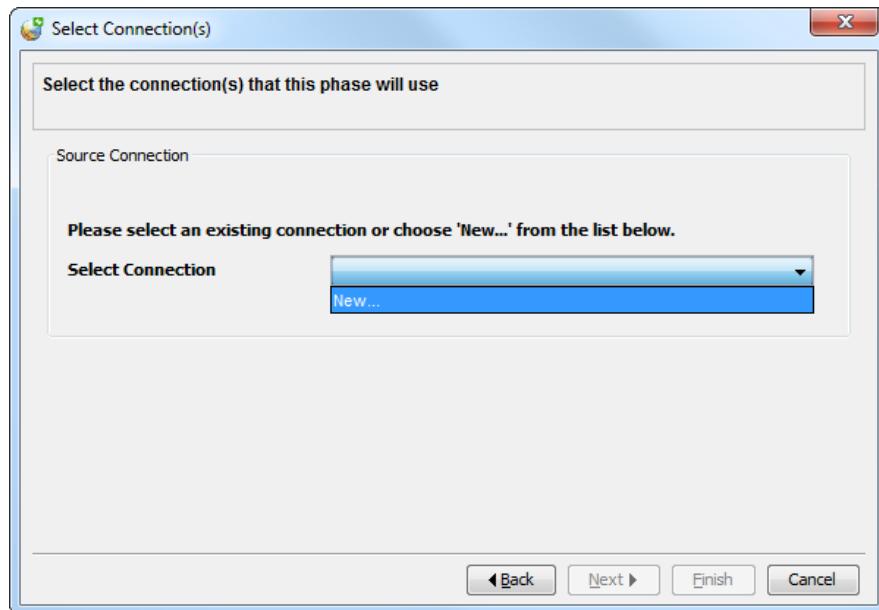


13. Enter **OW\_Teapot\_HealthCheck** in the **Phase Name** field.

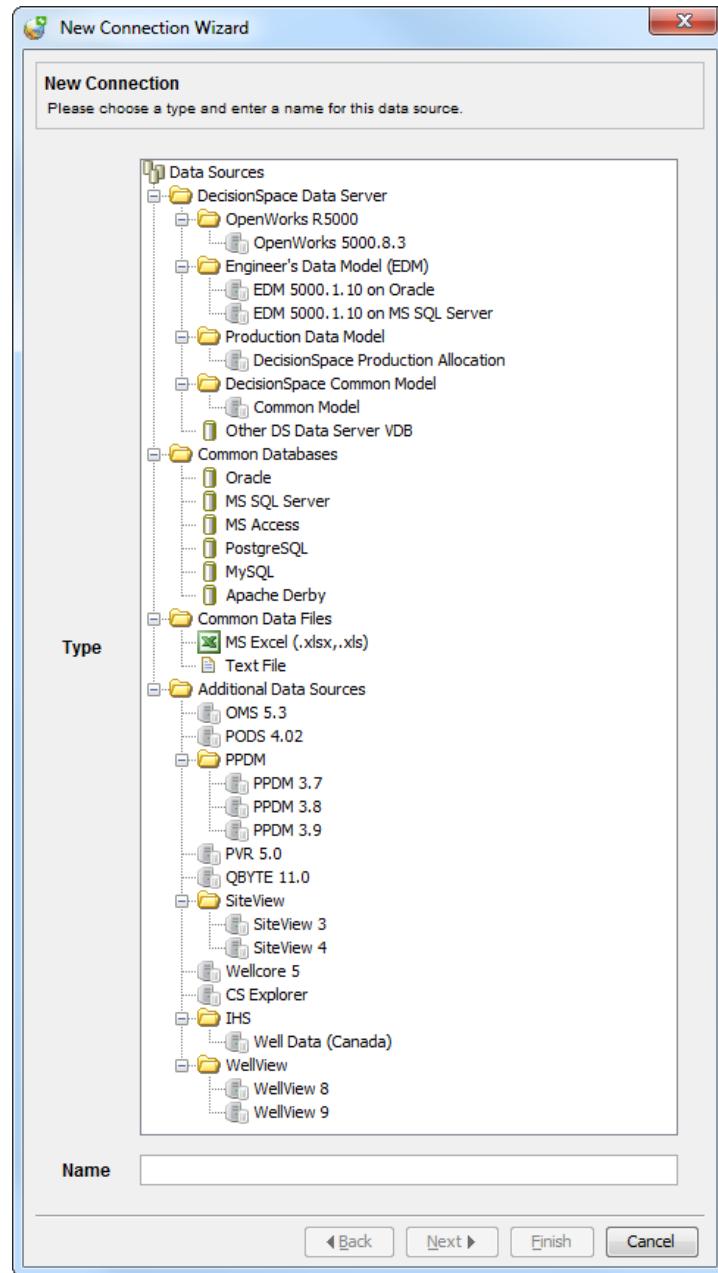


14. Click **Next** to continue.

The **Select Connection(s)** window appears.

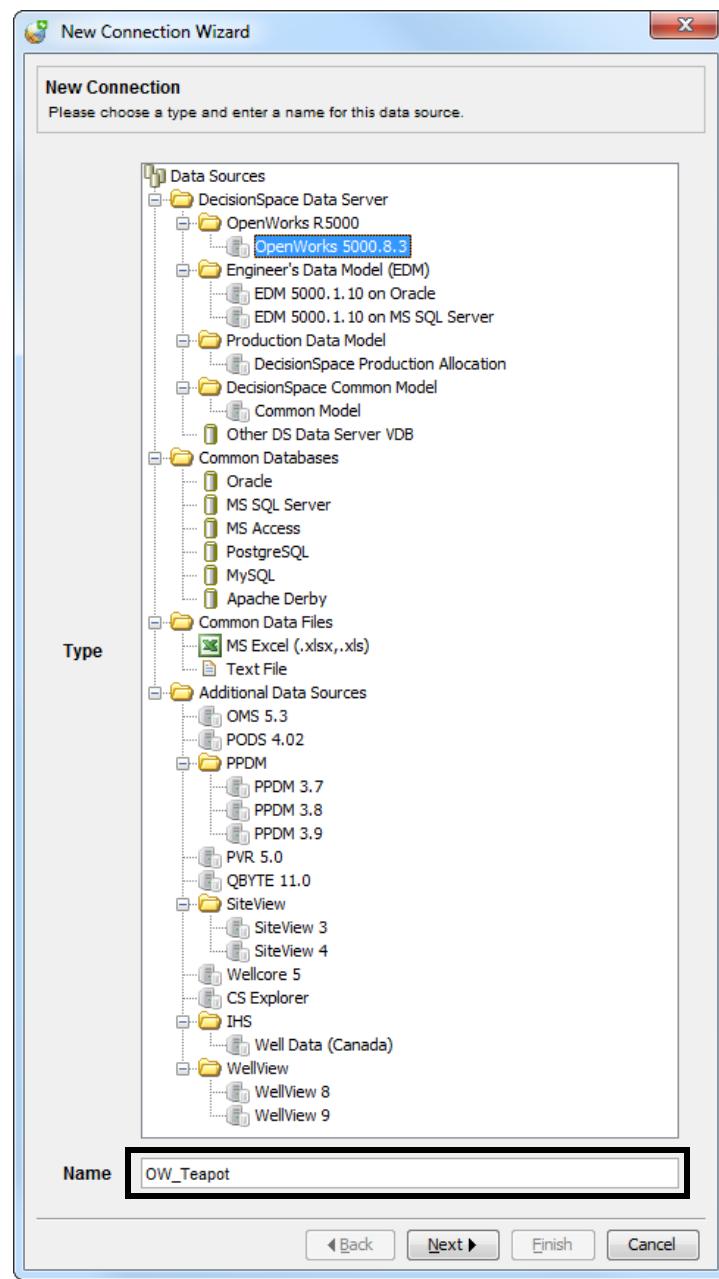


15. Select **New...** from the **Select Connection** drop-down list.  
The **New Connection Wizard - New Connection** window appears.



16. Select **OpenWorks 5000.8.3** from the Connection Type tree.

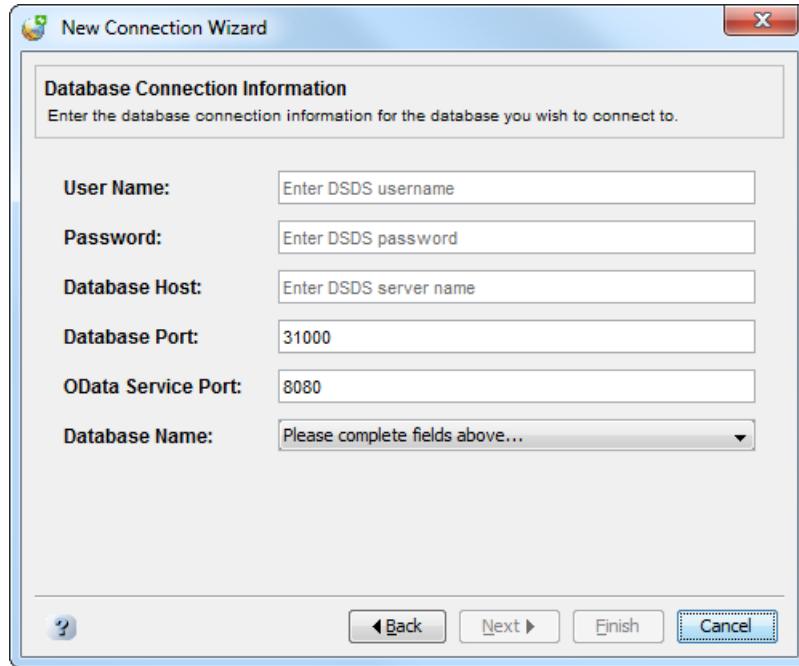
17. Enter **OW\_Teapot** in the **Name** field.



18. Click **Next** to continue.

#### The New Connection Wizard -Database Connection

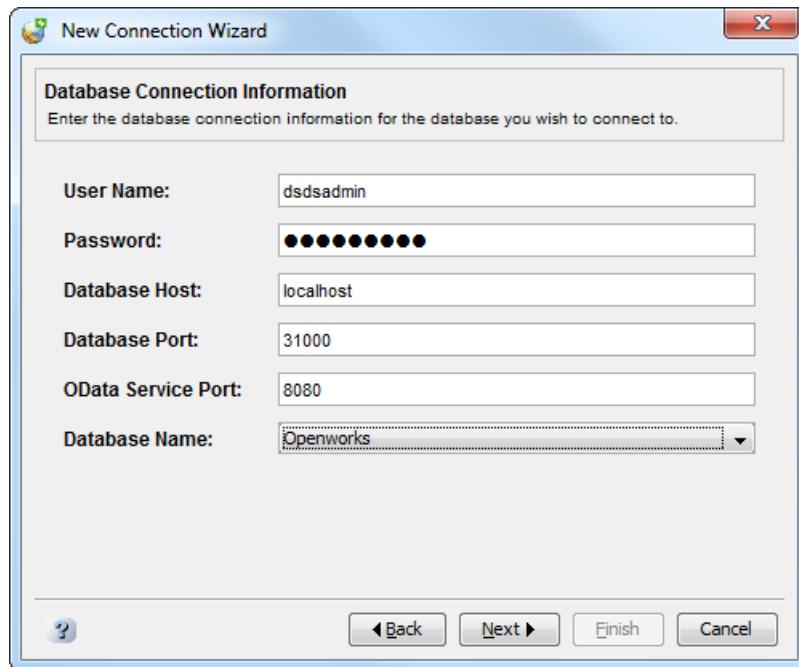
**Information** window appears displaying database connection information for DecisionSpace Data Server.



19. Enter **dsdsadmin** in the **User Name** field.
20. Enter **dsdsadmin** in the **Password** field.
21. Enter **localhost** in the **Database Host** field.
22. Accept the default **Database Port** value of **31000**.

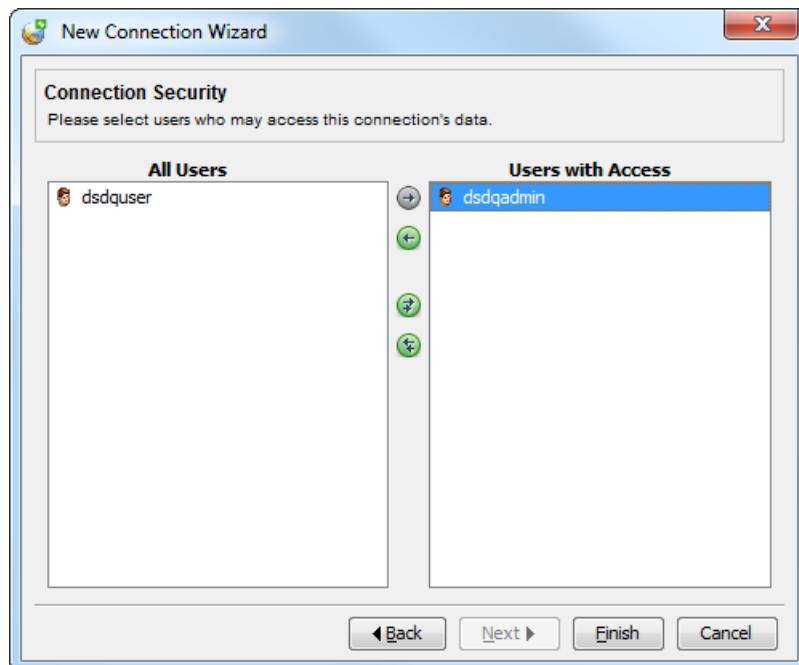
23. Accept the default **OData Service Port** value of **8080**.

The **Database Name** drop-down list auto-populates with the OpenWorks data source option.



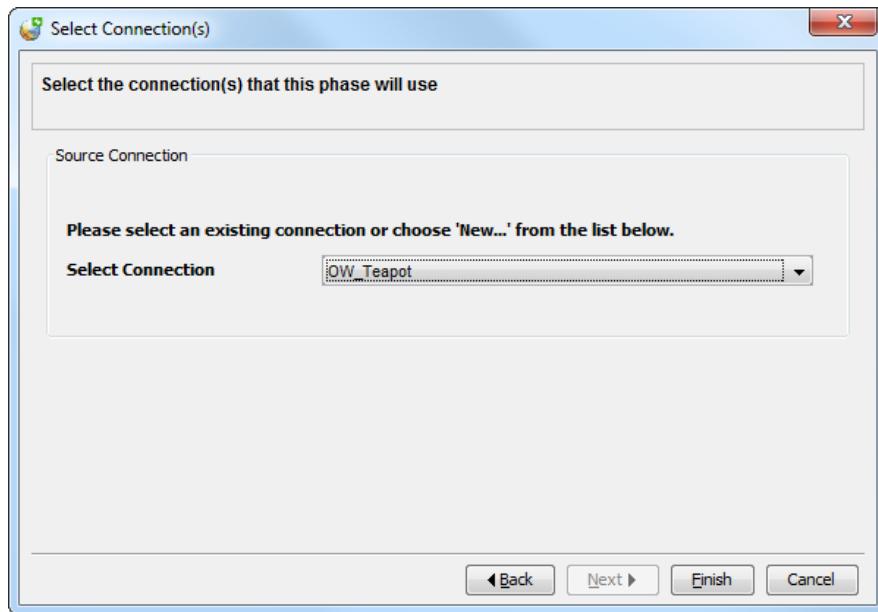
24. Click **Next** to continue.

The **New Connection Wizard - Connection Security** window appears.



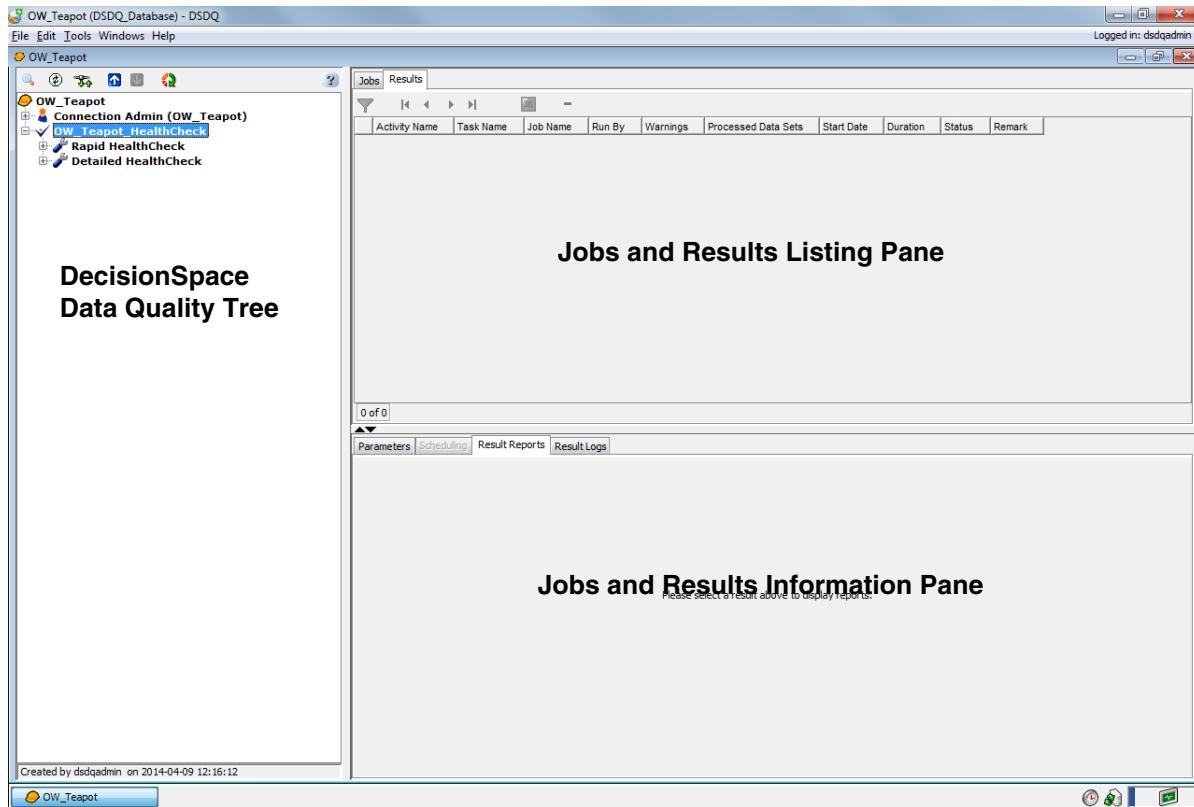
25. Click **Finish**.

The **Select Connection(s)** window appears.



26. Click **Finish**.

The **HealthCheck** Phase is created and displays in the DecisionSpace Data Quality Project window.



## Evaluating Data using the HealthCheck Phase

The **HealthCheck** Phase assists in evaluating the "where", "what" and "why" issues in your valuable data assets.

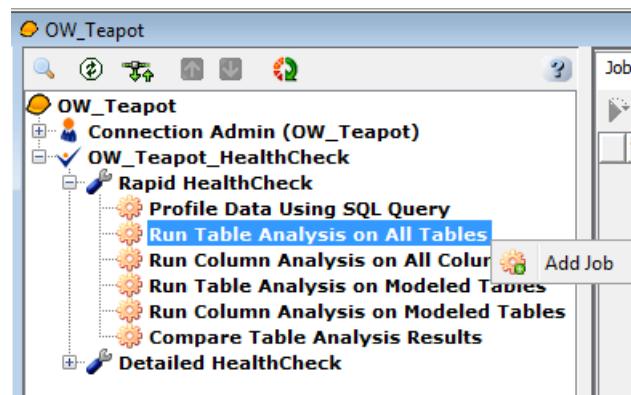
### Rapid HealthCheck Activity

The **Rapid HealthCheck** Activity provides a quick look at the volume and quality of the data.

#### Exercise: Running Table Analysis on All Tables Task

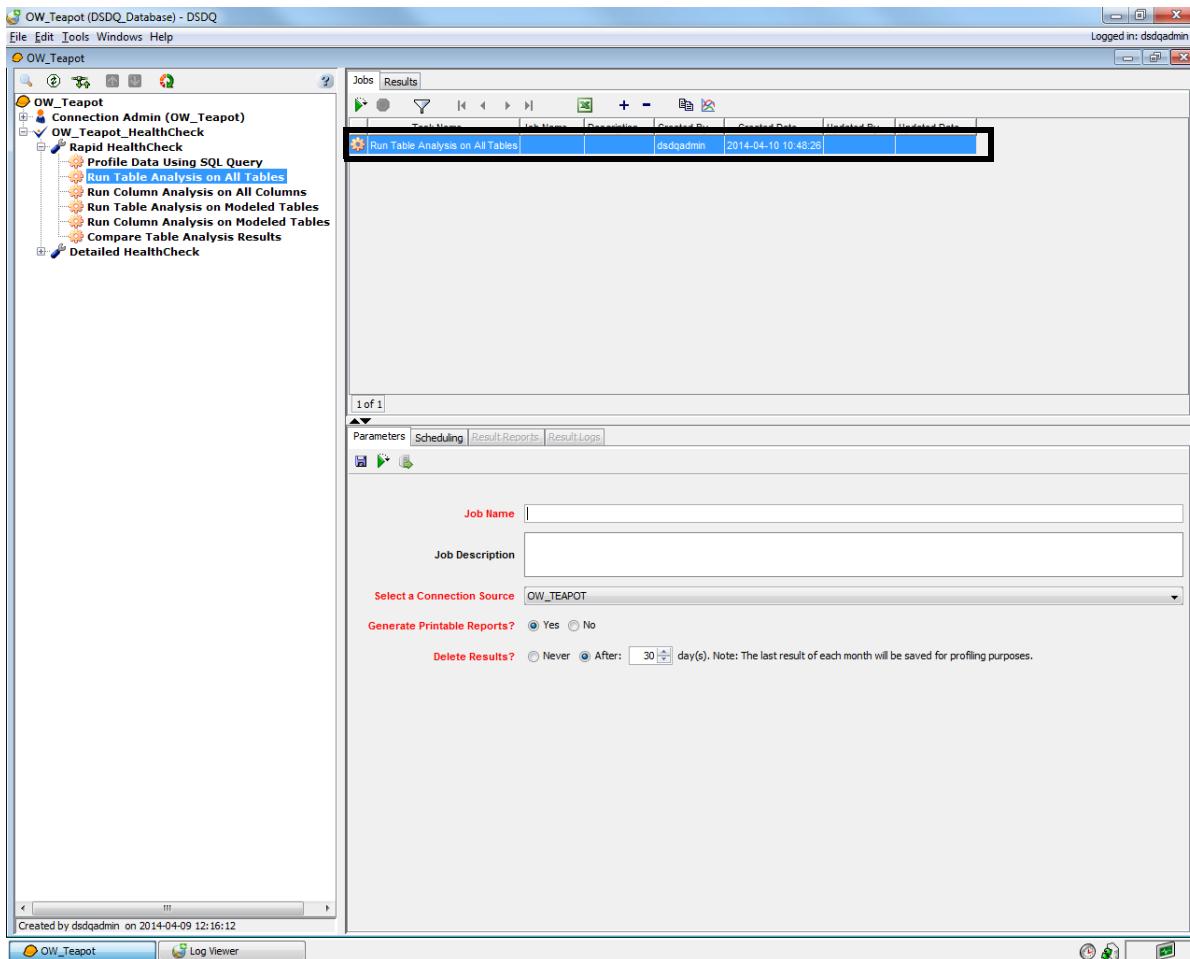
The **Run Table Analysis on All Tables** Task is used to analyze all the tables for issues and inconsistencies. In this particular exercise, we will analyze the tables and count the number of rows in them. Rows are counted when values are entered in them. To run Table Analysis on all tables:

1. Click  to expand the **Rapid HealthCheck** Activity on the DecisionSpace Data Quality Tree.
2. Double-click the **Run Table Analysis on All Tables** Task or right-click the **Run Table Analysis on All Tables** Task and select **Add Job** from the pop-up menu.



A new job is initiated and displays on the **Job and Results**

**Information Pane** on the right side of the DecisionSpace Data Quality Project window.

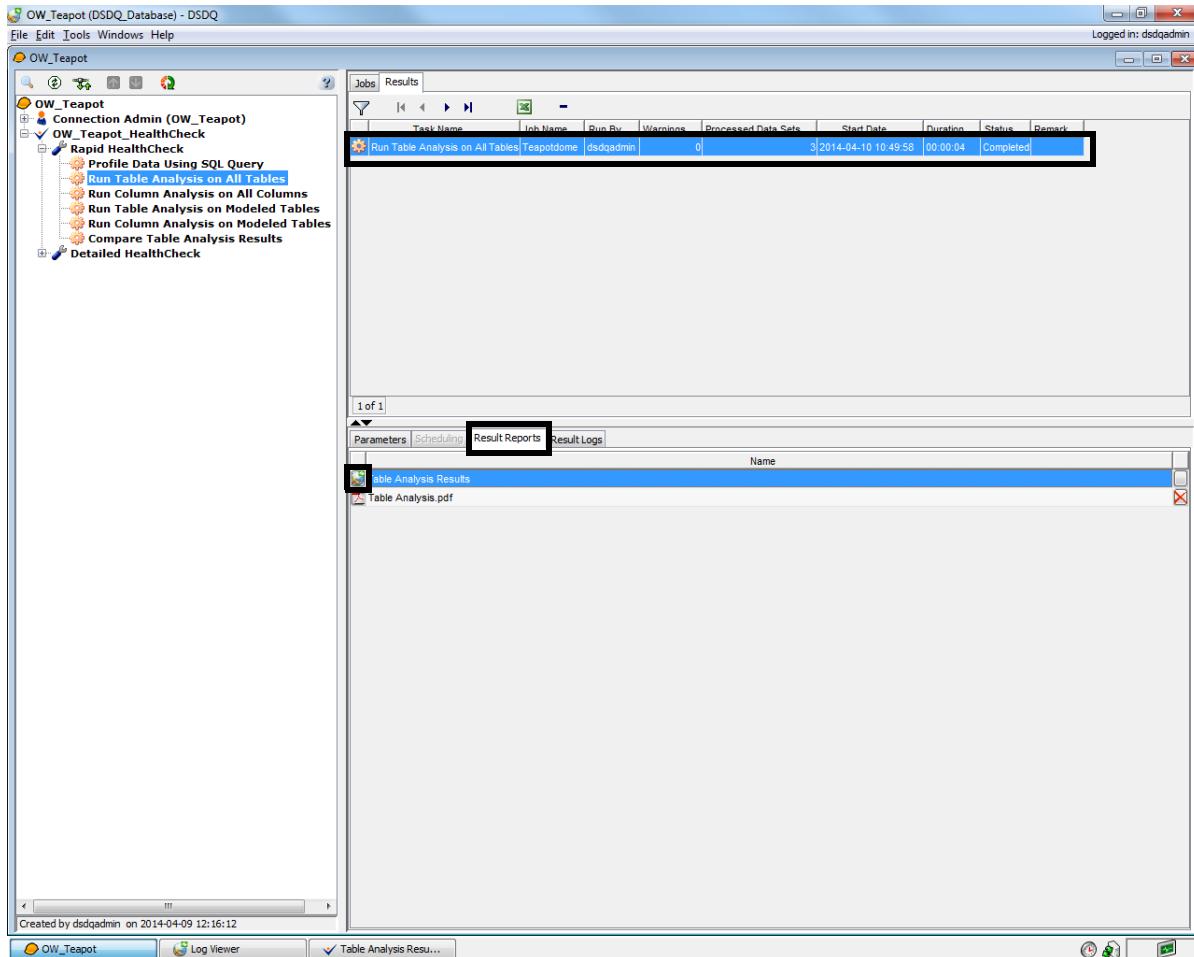


3. Enter **Teapotdome** in the **Job Name** field.
  4. Enter **Table Analysis on All Tables of OW\_Teapot** in the **Job Description** field.
  5. Select **OW\_Teapot** from the **Select a Connection Source** drop-down list.
  6. Select the **Yes** option for **Generate Printable Reports**.
  7. Select the **After** option for **Delete Results**. Leave the number of days as **30**.
  8. Click to save changes in the **Parameters** tab.
  9. Click to run the job.
- The **Run Table Analysis on All Tables** Task runs and displays

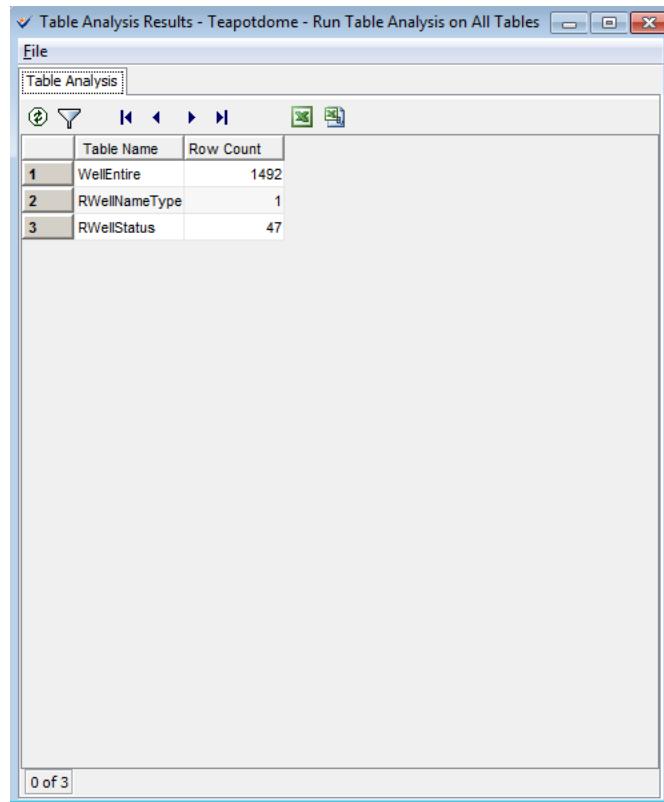
results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

10. Select the **Results** tab.

The **Jobs and Results Listing Pane** displays a list of results.



11. Click the **Open Basic View Frame**  button on the **Result Reports** tab to display the **Run Table Analysis on All Tables** Task results in the **Basic View Frame** window.



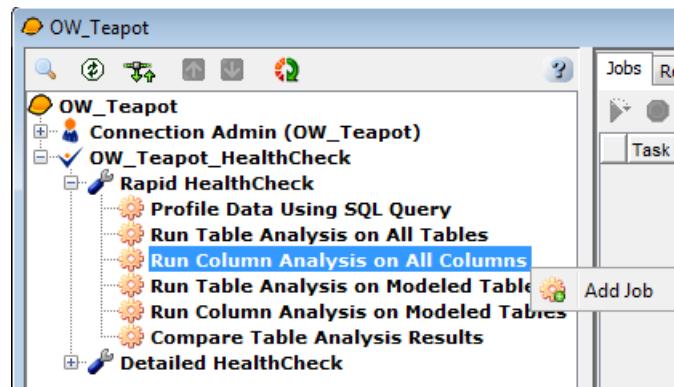
12. Select **File > Exit** to close the **Basic View Frame** window.

### Exercise: Running Column Analysis on All Columns Task

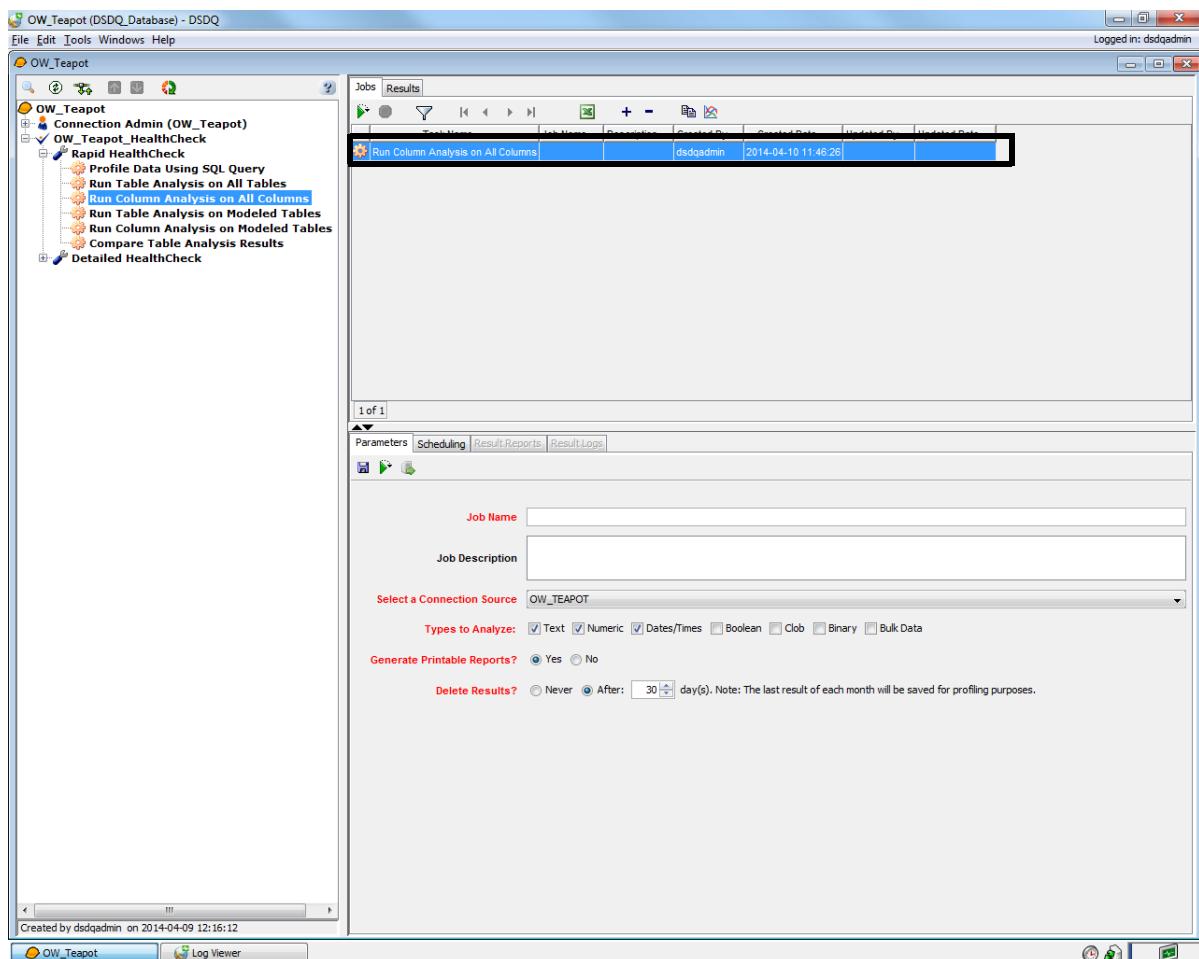
The **Run Column Analysis on All Columns** Task offers basic data profiling by checking the number of rows, not null values & unique values, percentage of row populated, minimum & maximum values, number of values with mixed cases, number of non-printable characters, and number of preceding, trailing & double white spaces parameters within a column. To run the Column Analysis on all columns:

1. Double-click the **Run Column Analysis on All Columns** Task or right-click the **Run Column Analysis on All Columns** Task and

select **Add Job** from the pop-up menu.

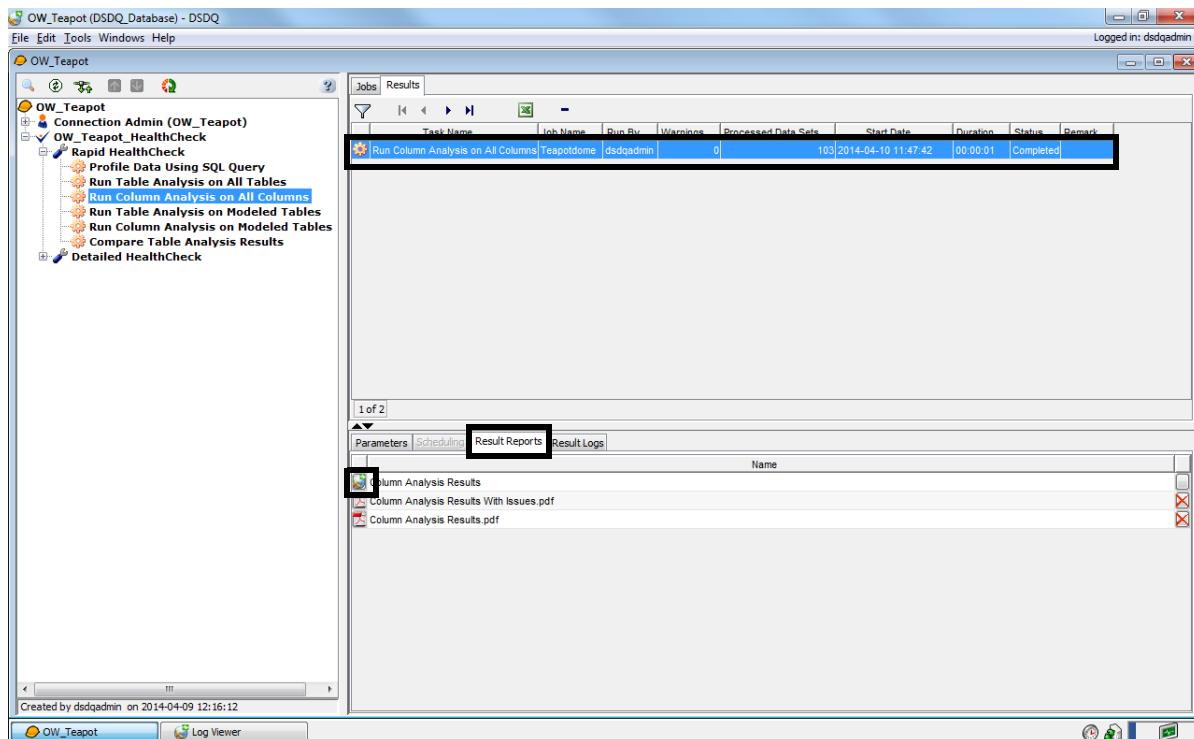


A new job is initiated and displays on the **Job and Results Information Pane** on the right side of the DecisionSpace Data Quality Project window.



2. Enter **Teapotdome** in the **Job Name** field.

3. Enter **Column Analysis on All Columns of OW\_Teapot** in the **Job Description** field.
4. Select **OW\_Teapot** from the **Select a Connection Source** drop-down list.
5. Select all the options for **Types to Analyze**.
6. Select the **Yes** option for **Generate Printable Reports**.
7. Select the **After** option for **Delete Results**. Leave the number of days as **30**.
8. Click  to save changes in the **Parameters** tab.
9. Click  to run the job.  
The **Run Column Analysis on All Columns** Task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.
10. Select the **Results** tab.  
The **Jobs and Results Listing Pane** displays a list of results.



**11. Click the Open Basic View Frame  button on the Result Reports tab to display the Run Column Analysis on All Columns Task results in the Basic View Frame window.**

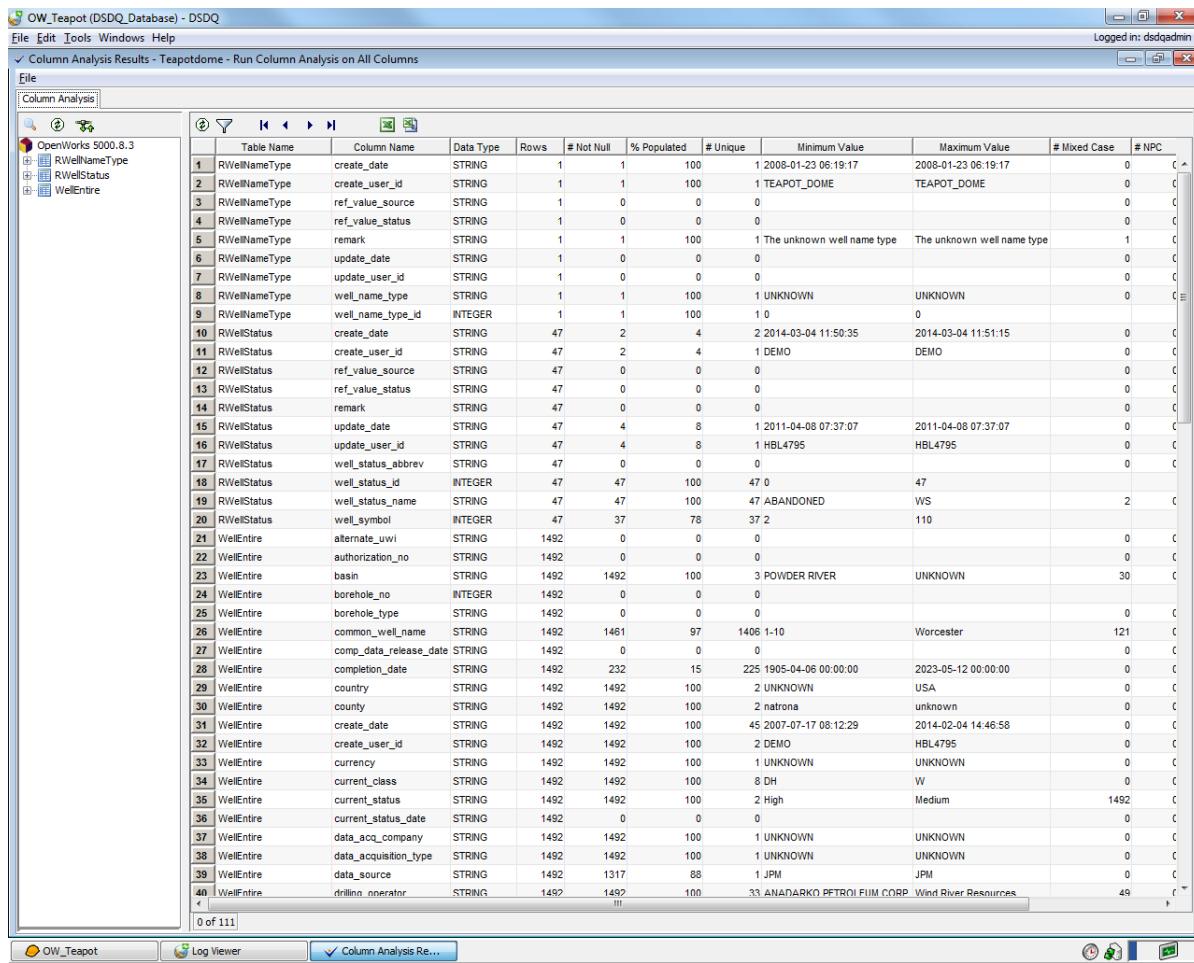


Table Name	Column Name	Data Type	Rows	# Not Null	% Populated	# Unique	Minimum Value	Maximum Value	# Mixed Case	# NPC
1 RWellNameType	create_date	STRING	1	1	100	1	2008-01-23 06:19:17	2008-01-23 06:19:17	0	C
2 RWellNameType	create_user_id	STRING	1	1	100	1	TEAPOT_DOME	TEAPOT_DOME	0	C
3 RWellNameType	ref_value_source	STRING	1	0	0	0			0	C
4 RWellNameType	ref_value_status	STRING	1	0	0	0			0	C
5 RWellNameType	remark	STRING	1	1	100	1	The unknown well name type	The unknown well name type	1	C
6 RWellNameType	update_date	STRING	1	0	0	0			0	C
7 RWellNameType	update_user_id	STRING	1	0	0	0			0	C
8 RWellNameType	well_name_type	STRING	1	1	100	1	UNKNOWN	UNKNOWN	0	C
9 RWellStatus	well_name_type_id	INTEGER	1	1	100	1	0	0	0	C
10 RWellStatus	create_date	STRING	47	2	4	2	2014-03-04 11:50:35	2014-03-04 11:51:15	0	C
11 RWellStatus	create_user_id	STRING	47	2	4	1	DEMO	DEMO	0	C
12 RWellStatus	ref_value_source	STRING	47	0	0	0			0	C
13 RWellStatus	ref_value_status	STRING	47	0	0	0			0	C
14 RWellStatus	remark	STRING	47	0	0	0			0	C
15 RWellStatus	update_date	STRING	47	4	8	1	2011-04-08 07:37:07	2011-04-08 07:37:07	0	C
16 RWellStatus	update_user_id	STRING	47	4	8	1	HBL4795	HBL4795	0	C
17 RWellStatus	well_status_abbrv	STRING	47	0	0	0			0	C
18 RWellStatus	well_status_id	INTEGER	47	47	100	47	0	47	0	C
19 RWellStatus	well_status_name	STRING	47	47	100	47	ABANDONED	WS	2	C
20 RWellStatus	well_symbol	INTEGER	47	37	78	37	2	110		C
21 WellEntire	alternate_uwi	STRING	1492	0	0	0			0	C
22 WellEntire	authorization_no	STRING	1492	0	0	0			0	C
23 WellEntire	basin	STRING	1492	1492	100	3	POWDER RIVER	UNKNOWN	30	C
24 WellEntire	borehole_no	INTEGER	1492	0	0	0			0	C
25 WellEntire	borehole_type	STRING	1492	0	0	0			0	C
26 WellEntire	common_well_name	STRING	1492	1481	97	1406	1-10	Worcester	121	C
27 WellEntire	comp_data_release_date	STRING	1492	0	0	0			0	C
28 WellEntire	completion_date	STRING	1492	232	15	225	1905-04-06 00:00:00	2023-05-12 00:00:00	0	C
29 WellEntire	country	STRING	1492	1492	100	2	UNKNOWN	USA	0	C
30 WellEntire	county	STRING	1492	1492	100	2	natrona	unknown	0	C
31 WellEntire	create_date	STRING	1492	1492	100	45	2007-07-17 08:12:29	2014-02-04 14:46:58	0	C
32 WellEntire	create_user_id	STRING	1492	1492	100	2	DEMO	HBL4795	0	C
33 WellEntire	currency	STRING	1492	1492	100	1	UNKNOWN	UNKNOWN	0	C
34 WellEntire	current_class	STRING	1492	1492	100	8	DH	W	0	C
35 WellEntire	current_status	STRING	1492	1492	100	2	High	Medium	1492	C
36 WellEntire	current_status_date	STRING	1492	0	0	0			0	C
37 WellEntire	data_acq_company	STRING	1492	1492	100	1	UNKNOWN	UNKNOWN	0	C
38 WellEntire	data_acquisition_type	STRING	1492	1492	100	1	UNKNOWN	UNKNOWN	0	C
39 WellEntire	data_source	STRING	1492	1317	88	1	JPM	JPM	0	C
40 WellEntire	drilling_operator	STRING	1492	1492	100	33	ANADARKO PETROLIUM CORP	Wind River Resources	49	C

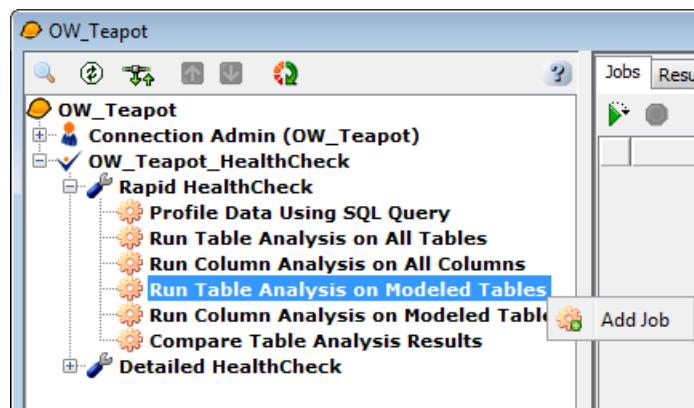
**12. Select File > Exit to close the Basic View Frame window.**

### Exercise: Running Table Analysis on Modeled Tables Task

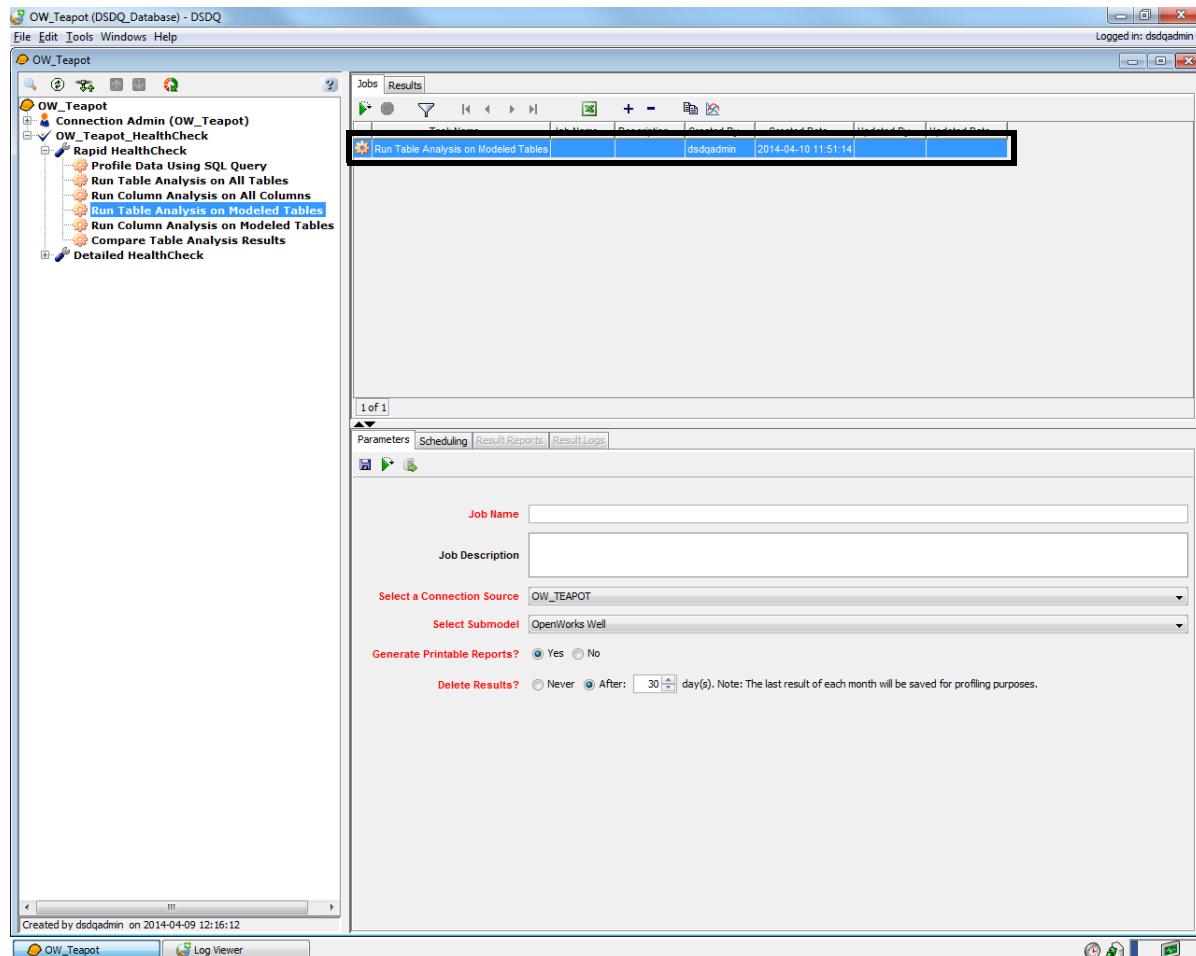
The **Run Table Analysis on Modeled Tables** Task runs only on tables that have been modeled in the **Perform Table Modeling** Tool (*reference: DecisionSpace Data Quality Training Manual, Chapter 4: Data Evaluation, Perform Table Modeling*). To run Table Analysis on all the modeled tables:

1. Double-click the **Run Table Analysis on Modeled Tables** Task or right-click the **Run Table Analysis on Modeled Tables** Task and

select **Add Job** from the pop-up menu.

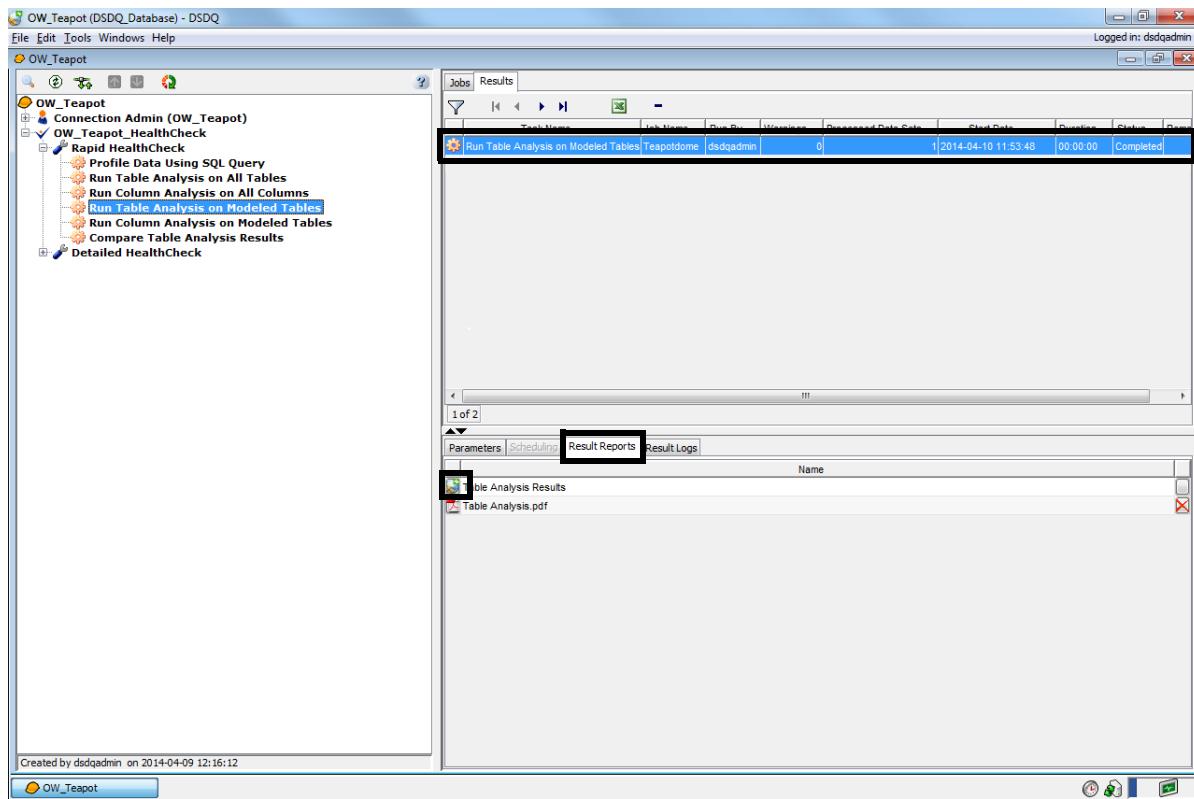


A new job is initiated and it displays on the **Job and Results Information Pane** on the right side of the DecisionSpace Data Quality Project window.

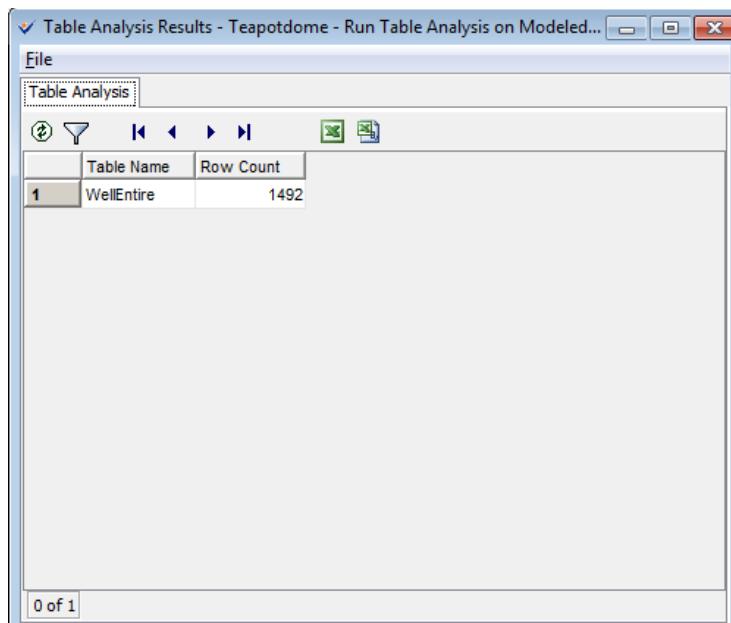


2. Enter **Teapotdome** in the **Job Name** field.

3. Enter **Table Analysis on Modeled Tables of OW\_Teapot** in the **Job Description** field.
4. Select **OW\_Teapot** from the **Select a Connection Source** drop-down list.
5. Select **OpenWorks Well** from the **Select Submodel** drop-down list.
6. Select the **Yes** option for **Generate Printable Reports**.
7. Select the **After** option for **Delete Results**. Leave the number of days as **30**.
8. Click  to save changes in the **Parameters** tab.
9. Click  to run the job.  
The **Run Table Analysis on Modeled Tables** Task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.
10. Select the **Results** tab.  
The **Jobs and Results Listing Pane** displays a list of results.



11. Click the **Open Basic View Frame**  button on the **Result Reports** tab to display the **Run Table Analysis on Modeled Tables** Task results in the **Basic View Frame** window.



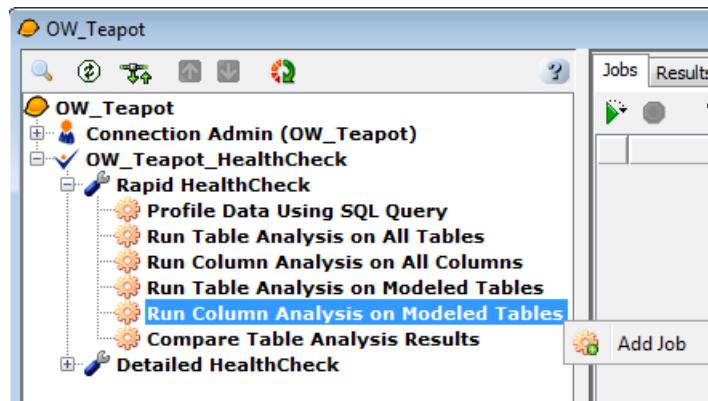
12. Select **File > Exit** to close the **Basic View Frame** window.

### Exercise: Running Column Analysis on Modeled Tables Task

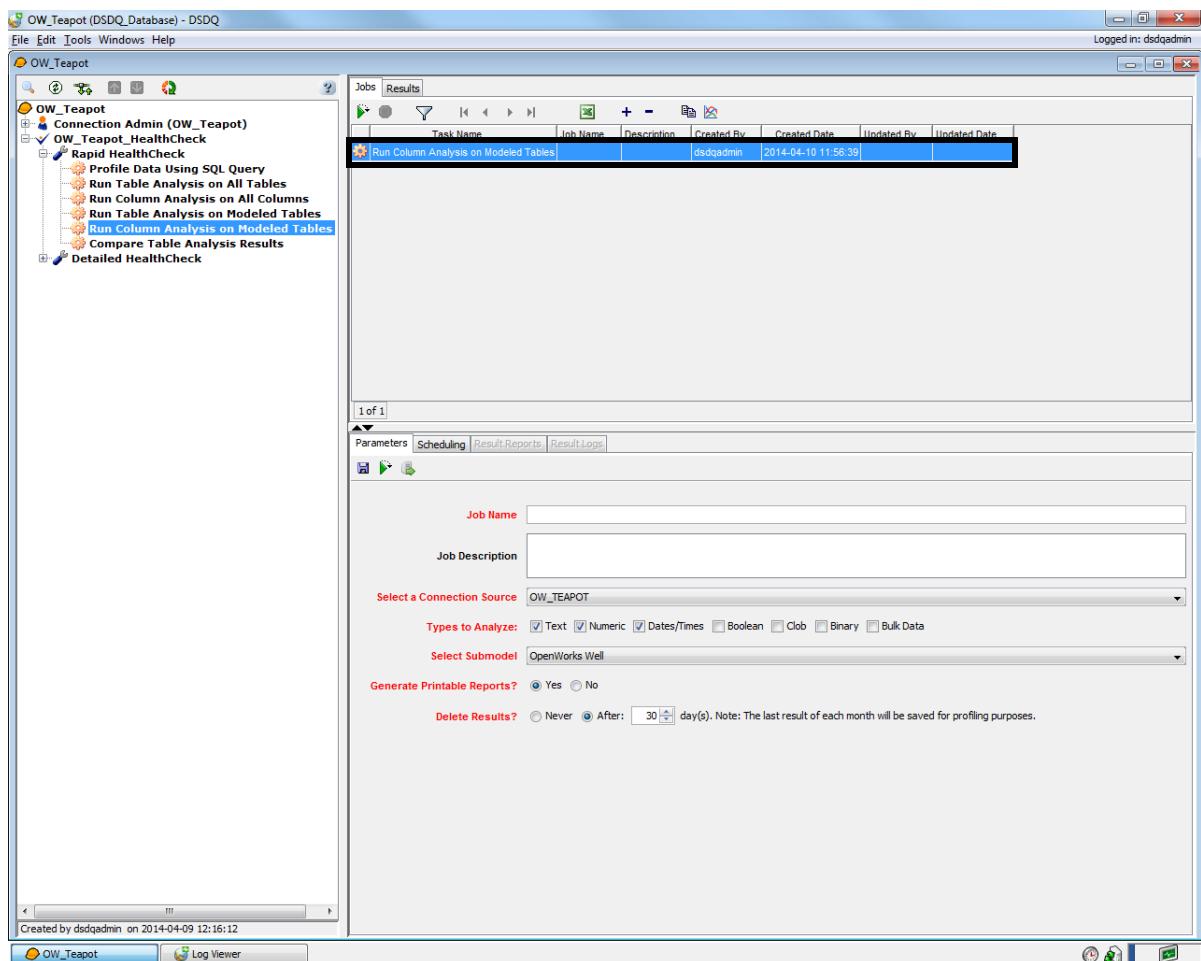
The **Run Column Analysis on Modeled Tables** Task runs only on tables that have been modeled in the **Perform Table Modeling** Tool (*reference: DecisionSpace Data Quality Training Manual, Chapter 4: Data Evaluation, Perform Table Modeling*). To run Column Analysis on all the modeled tables:

1. Double-click the **Run Column Analysis on Modeled Tables** Task or right-click the **Run Column Analysis on Modeled Tables** Task

and select **Add Job** from the pop-up menu.|



A new job is initiated and it displays on the **Job and Results Information Pane** on the right side of the DecisionSpace Data Quality Project window.

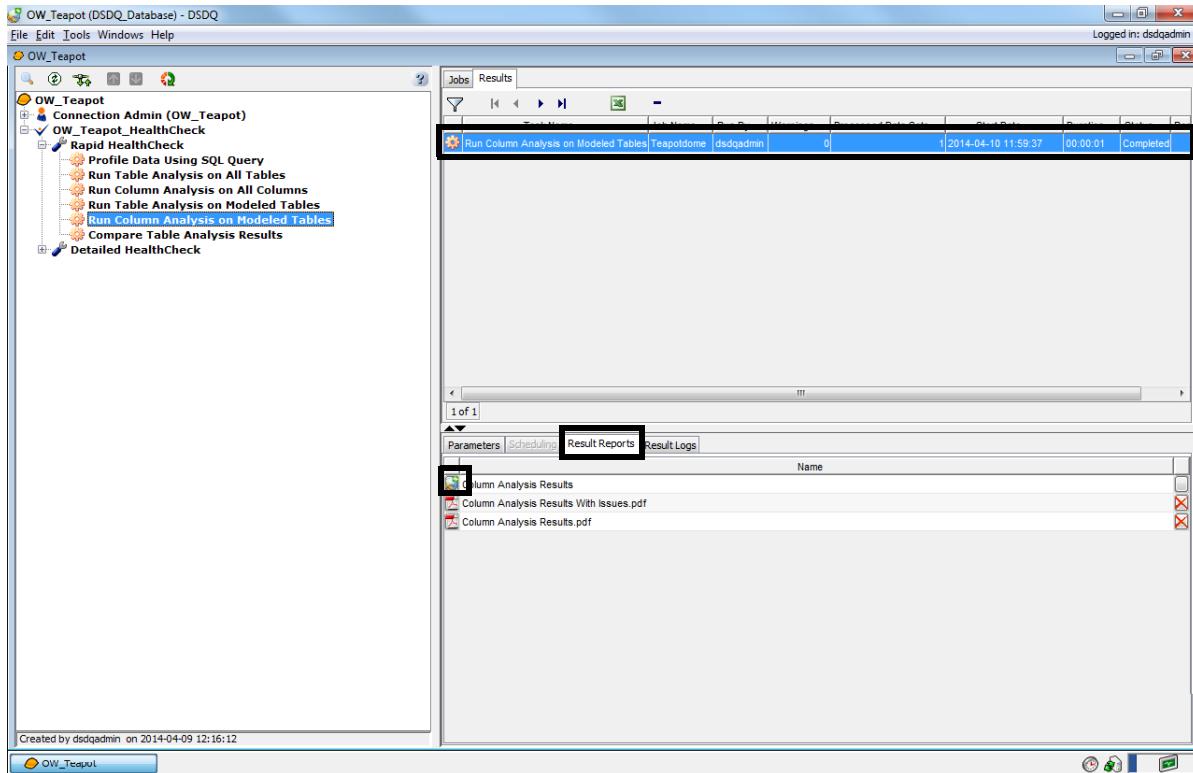


2. Enter **Teapotdome** in the **Job Name** field.

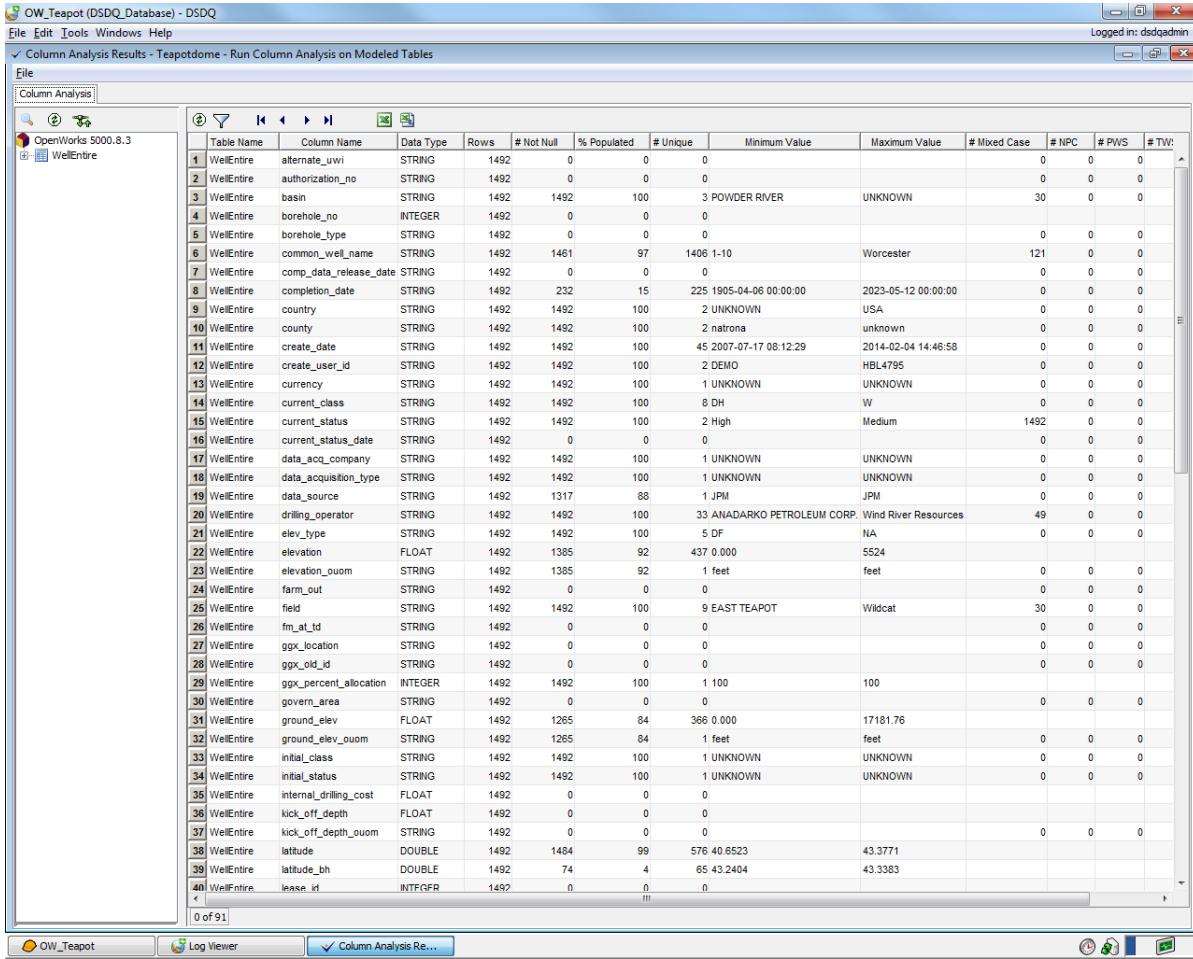
3. Enter **Column Analysis on Modeled Tables of OW\_Teapot** in the **Job Description** field.
4. Select **OW\_Teapot** from the **Select a Connection Source** drop-down list.  
The **Text**, **Numeric** and **Dates/Times** options are selected by default for **Types to Analyze**.
5. Select **OpenWorks Well** from the **Select Submodel** drop-down list.
6. Select the **Yes** option for **Generate Printable Reports**.
7. Select the **After** option for **Delete Results**. Leave the number of days as **30**.
8. Click  to save changes in the **Parameters** tab.
9. Click  to run the job.  
The **Run Column Analysis on Modeled Tables** Task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

10. Select the **Results** tab.

The **Jobs and Results Listing Pane** displays a list of results.



11. Click the **Open Basic View Frame**  button on the Result Reports tab to display the Column Analysis on Modeled Tables Task results in the **Basic View Frame** window.



The screenshot shows the 'Column Analysis' results for the 'WellEntire' table. The grid includes the following columns:

Table Name	Column Name	Data Type	Rows	# Not Null	% Populated	# Unique	Minimum Value	Maximum Value	# Mixed Case	# NPC	# PWS	# TWL
WellEntire	alternate_uwi	STRING	1492	0	0	0			0	0	0	
WellEntire	authorization_no	STRING	1492	0	0	0			0	0	0	
WellEntire	basin	STRING	1492	1492	100	3 POWDER RIVER	UNKNOWN		30	0	0	
WellEntire	borehole_no	INTEGER	1492	0	0	0			0	0	0	
WellEntire	borehole_type	STRING	1492	0	0	0			0	0	0	
WellEntire	common_well_name	STRING	1492	1461	97	1406 1-10	Worcester		121	0	0	
WellEntire	comp_data_release_date	STRING	1492	0	0	0			0	0	0	
WellEntire	completion_date	STRING	1492	232	15	225 1905-04-06 00:00:00	2023-05-12 00:00:00		0	0	0	
WellEntire	country	STRING	1492	1492	100	2 UNKNOWN	USA		0	0	0	
WellEntire	county	STRING	1492	1492	100	2 natrona	unknown		0	0	0	
WellEntire	create_date	STRING	1492	1492	100	45 2007-07-17 08:12:29	2014-02-04 14:46:58		0	0	0	
WellEntire	create_user_id	STRING	1492	1492	100	2 DEMO	HBL4795		0	0	0	
WellEntire	currency	STRING	1492	1492	100	1 UNKNOWN	UNKNOWN		0	0	0	
WellEntire	current_class	STRING	1492	1492	100	8 DH	W		0	0	0	
WellEntire	current_status	STRING	1492	1492	100	2 High	Medium		1492	0	0	
WellEntire	current_status_date	STRING	1492	0	0	0			0	0	0	
WellEntire	data_acq_company	STRING	1492	1492	100	1 UNKNOWN	UNKNOWN		0	0	0	
WellEntire	data_acquisition_type	STRING	1492	1492	100	1 UNKNOWN	UNKNOWN		0	0	0	
WellEntire	data_source	STRING	1492	1317	88	1 JPM	JPM		0	0	0	
WellEntire	drilling_operator	STRING	1492	1492	100	33 ANADARKO PETROLEUM CORP, Wind River Resources			49	0	0	
WellEntire	elev_type	STRING	1492	1492	100	5 DF	NA		0	0	0	
WellEntire	elevation	FLOAT	1492	1385	92	437.0000	5524					
WellEntire	elevation_ouom	STRING	1492	1385	92	1 feet	feet		0	0	0	
WellEntire	farm_out	STRING	1492	0	0	0			0	0	0	
WellEntire	field	STRING	1492	1492	100	9 EAST TEAPOT	Wildcat		30	0	0	
WellEntire	fm_at_td	STRING	1492	0	0	0			0	0	0	
WellEntire	ggx_location	STRING	1492	0	0	0			0	0	0	
WellEntire	ggx_old_id	STRING	1492	0	0	0			0	0	0	
WellEntire	ggx_percent_allocation	INTEGER	1492	1492	100	1 100	100					
WellEntire	govern_area	STRING	1492	0	0	0			0	0	0	
WellEntire	ground_elev	FLOAT	1492	1265	84	366.0000	17181.76					
WellEntire	ground_elev_ouom	STRING	1492	1265	84	1 feet			0	0	0	
WellEntire	initial_class	STRING	1492	1492	100	1 UNKNOWN	UNKNOWN		0	0	0	
WellEntire	initial_status	STRING	1492	1492	100	1 UNKNOWN	UNKNOWN		0	0	0	
WellEntire	internal_drilling_cost	FLOAT	1492	0	0	0						
WellEntire	kick_off_depth	FLOAT	1492	0	0	0			0	0	0	
WellEntire	kick_off_depth_ouom	STRING	1492	0	0	0						
WellEntire	latitude	DOUBLE	1492	1484	99	576.40.6523	43.3771					
WellEntire	latitude_bh	DOUBLE	1492	74	4	65.43.2404	43.3383					
WellEntire	lease_id	INTEGER	1492	0	0	0						

12. Select **File > Exit** to close the **Basic View Frame** window.

## Detailed HealthCheck Activity

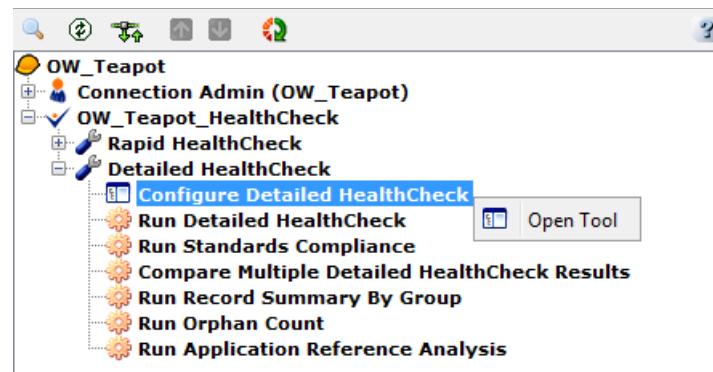
The **Detailed HealthCheck** Activity allows you to run business rules against the dataset to identify data problems.

### Exercise: Configuring the Detailed HealthCheck Tool

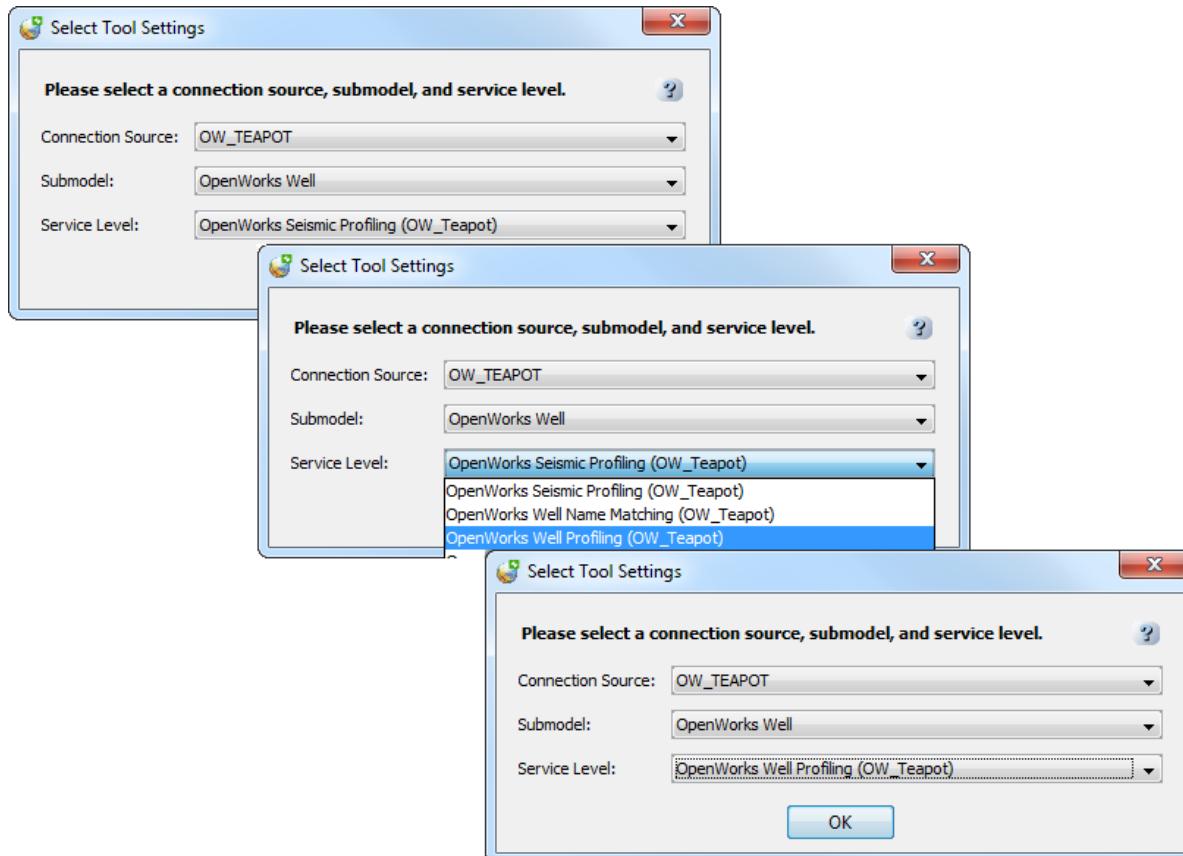
The **Configure Detailed HealthCheck** Tool configures service levels for testing prior to running the **Run Detailed HealthCheck** Task. You can select which requirements are to be enabled/disabled in the service level. You can also select subset of the total data to be used when testing a service level. A service level containing HealthCheck requirements

must exist in the DSDQ project prior to opening the **Configure Detailed HealthCheck** Tool. To configure the Detailed HealthCheck tool:

1. Click  to expand the **Detailed HealthCheck** Activity.
2. Double-click the **Configure Detailed HealthCheck** Tool or right-click the **Configure Detailed HealthCheck** Tool and select **Open Tool** from the pop-up menu.

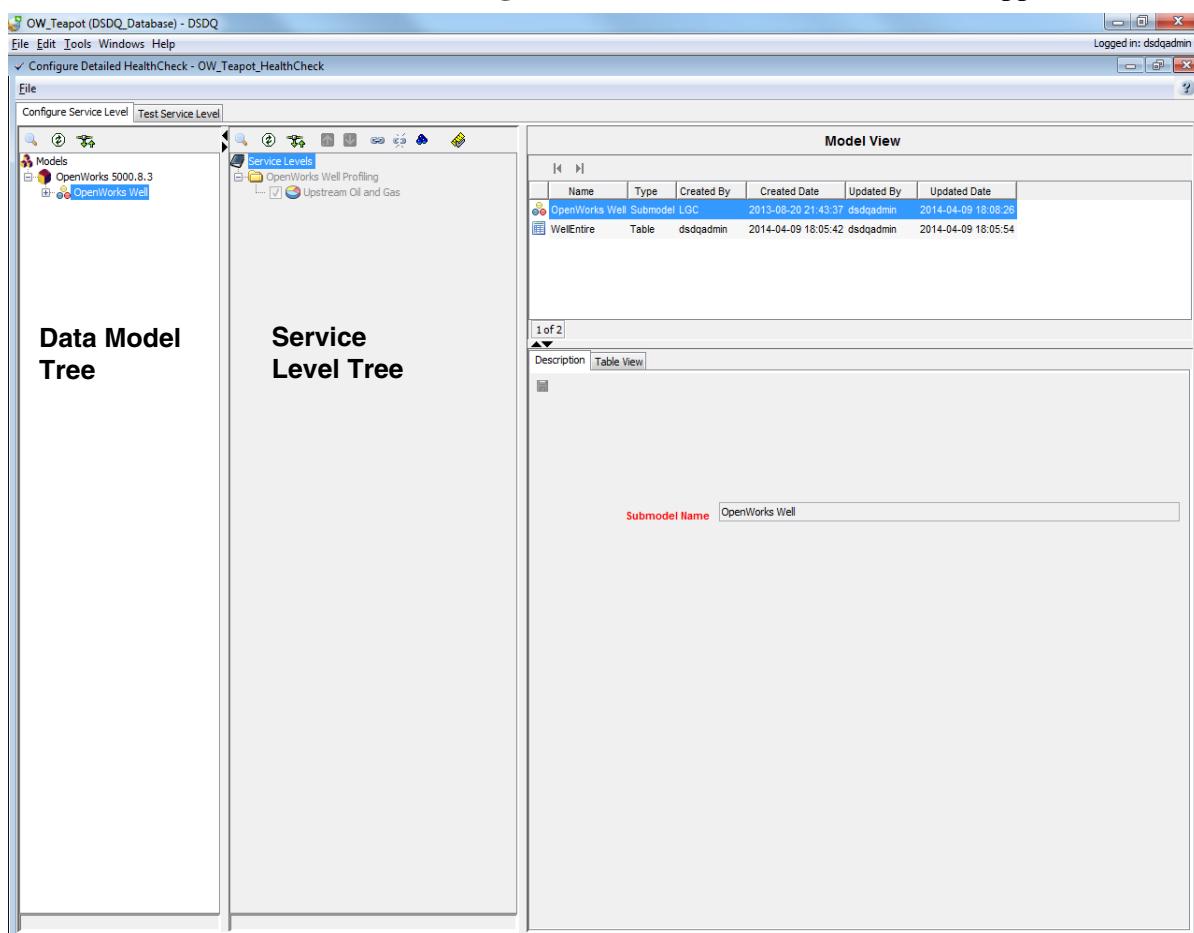


The **Select Tool Settings** window appears:



3. The **Connection Source** drop-down list is set to **OW\_Teapot** by default.
4. The **Submodel** drop-down list is set to **OpenWorks Well** by default.
5. Select **OpenWorks Well Profiling (OW\_Teapot)** from the **Service Level** drop-down list.
6. Click **OK**.

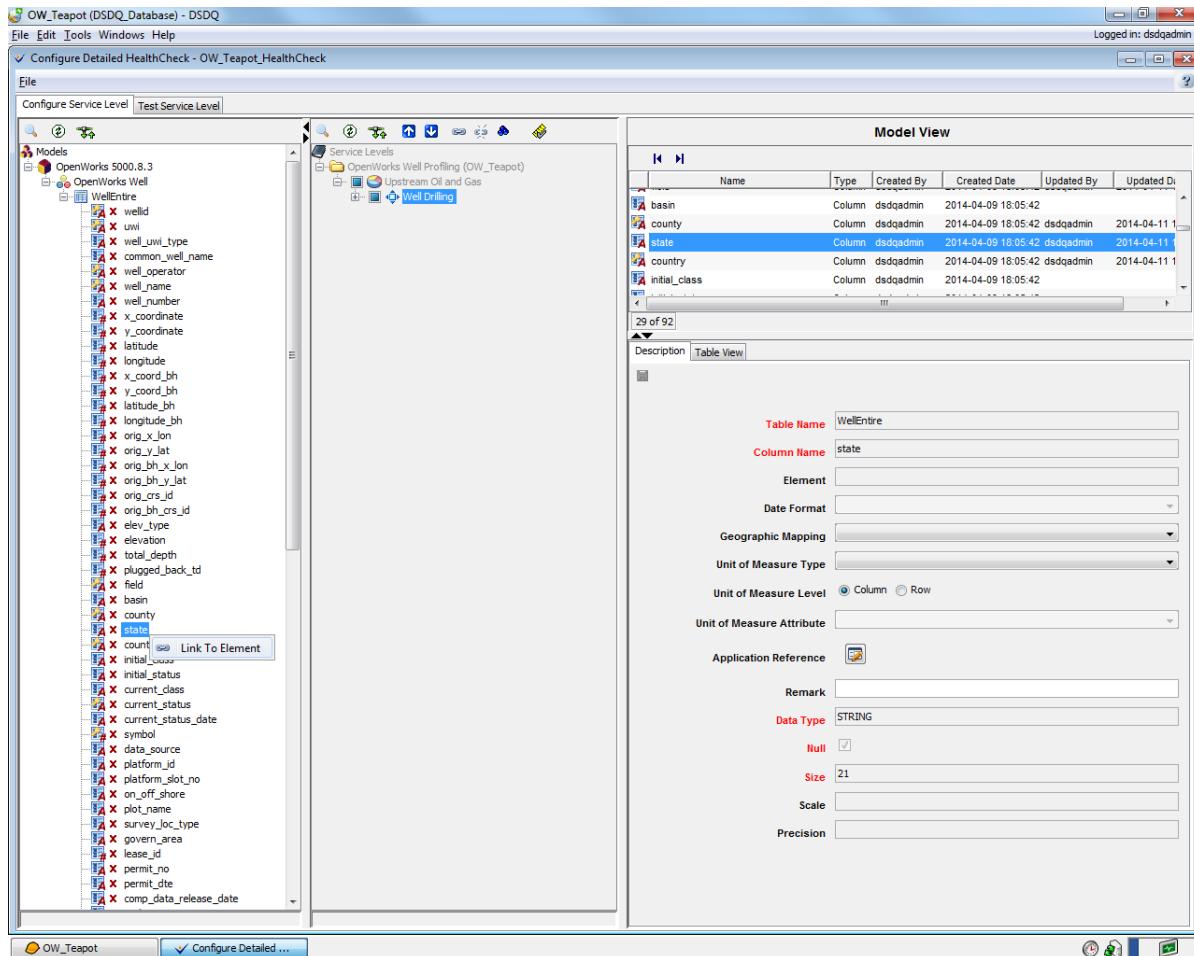
The **Configure Detailed HealthCheck** window appears.



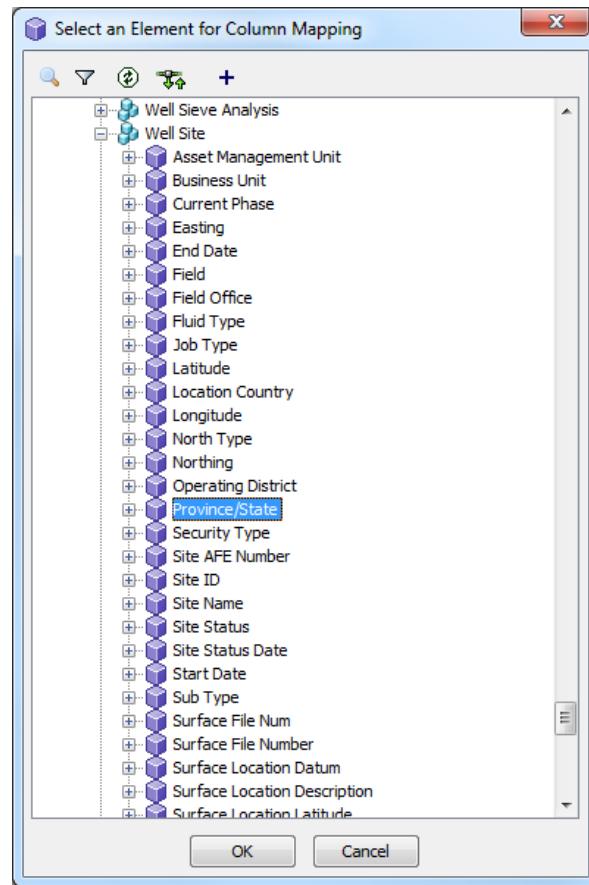
The Data Model Tree populates with the submodel selected in step 4. The Service Level Tree populates with the service level selected in step 5.

7. Click to expand the **OpenWorks Well** submodel in the DataModel Tree.
8. Click to expand the **WellEntire** table.

9. Right-click the **State** column and select **Link to Element** from the pop-up menu.



The **Select an Element for Column Mapping** window appears.



10. Select the **Province/State** element.

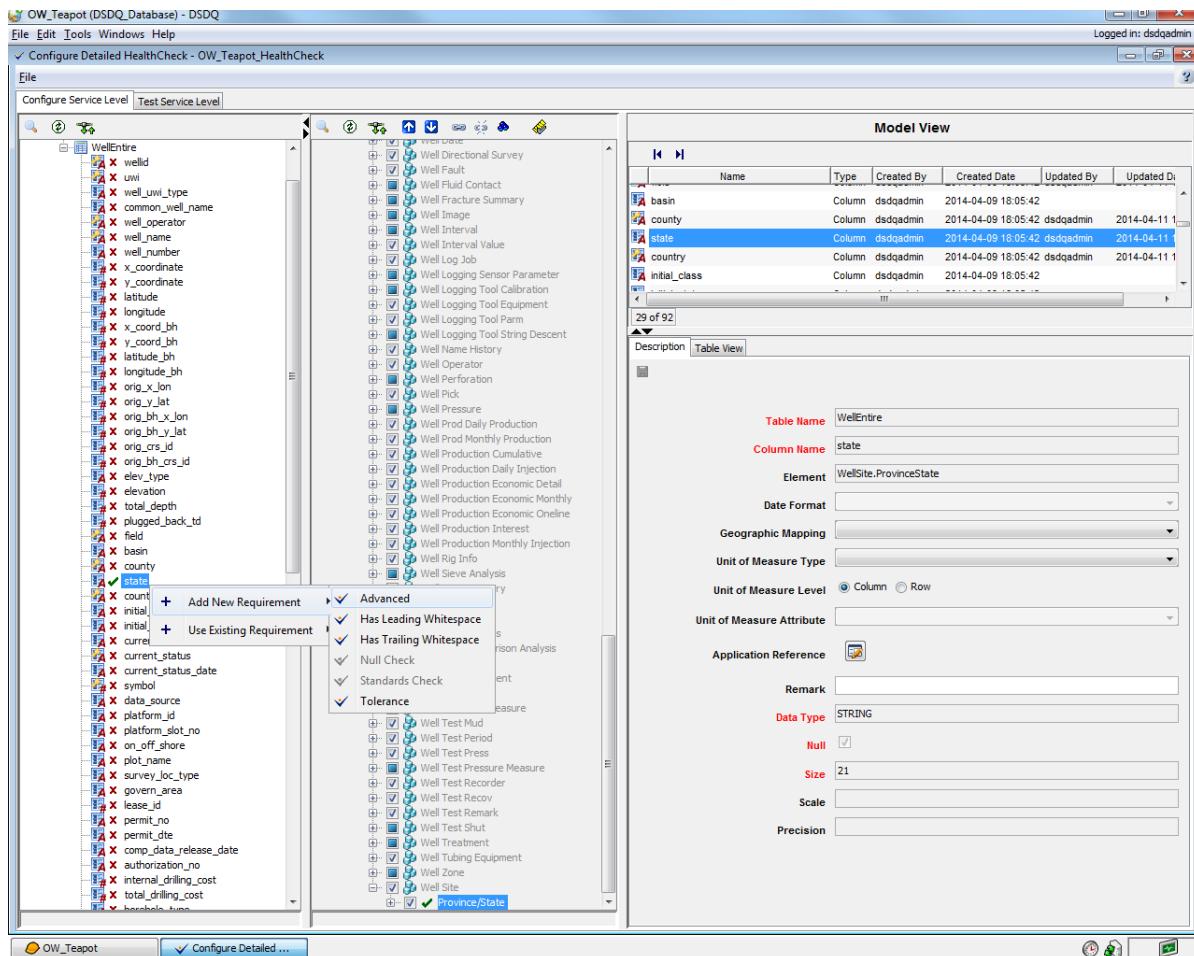
11. Click **OK**.

The **State** column and the **Province/State** element are associated with each other. A green check mark appears adjacent to the column and element that have just been associated. Only one column from the same table can be linked to the same element.

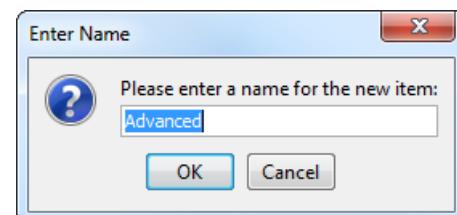
However, it is possible to link different tables' columns to the same element.

Name	Type	Created By	Created Date	Updated By	Updated Date
Province/State	Element	dsdqadmin	2014-04-11 09:47:47		
Convert State to Upper Case	Requirement	dsdqadmin	2014-04-11 10:17:09	dsdqadmin	2014-04-11 10:21:43
Null	Requirement	dsdqadmin	2014-04-11 10:29:38	dsdqadmin	2014-04-11 10:29:39
Standards Check	Requirement	dsdqadmin	2014-04-11 10:29:43	dsdqadmin	2014-04-11 10:29:45

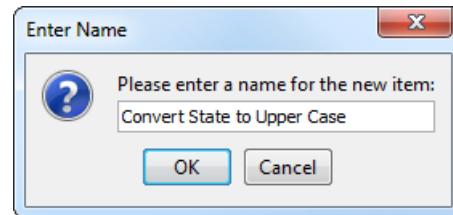
12. Right-click the **State** column in the Data Model Tree and select **Add New Requirement > Advanced** from the pop-up menu.



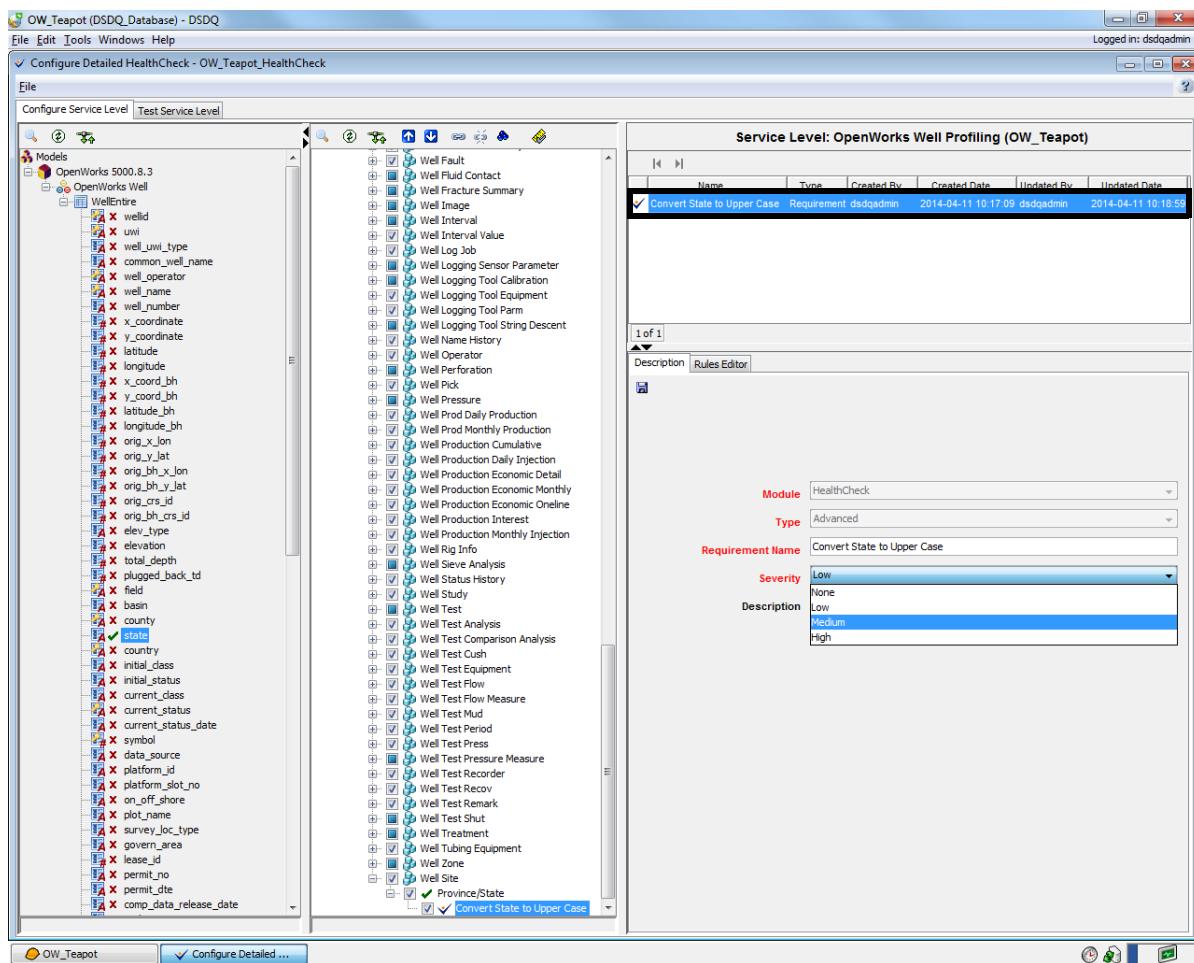
The Enter Name dialog box appears.



**13. Enter Convert State to Upper Case in the Please enter a name for the new item field.**



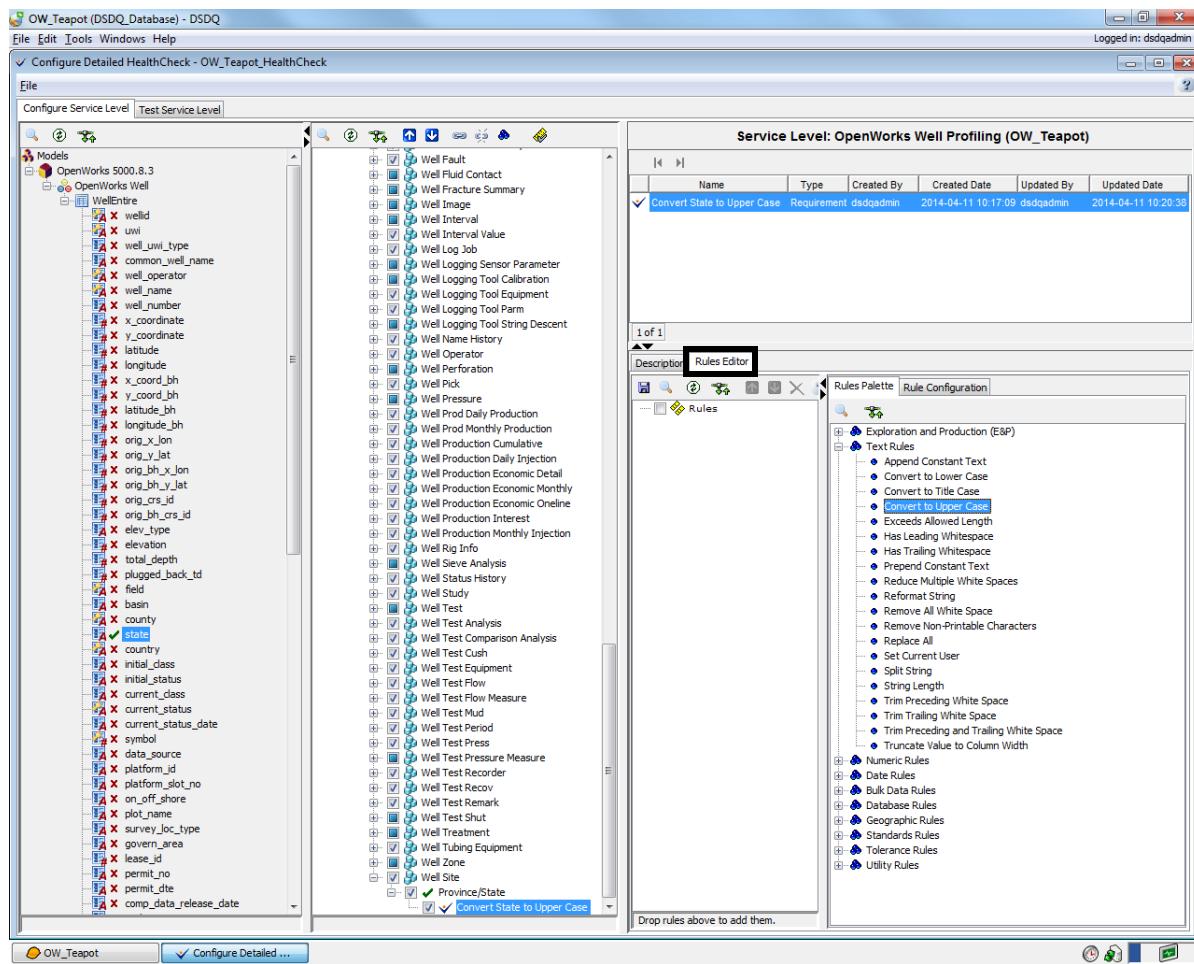
**14. Click OK to add the requirement to the selected column. The Convert State to Upper Case requirement is added to the State element.**



**15. The Module and Type fields are disabled by default.**

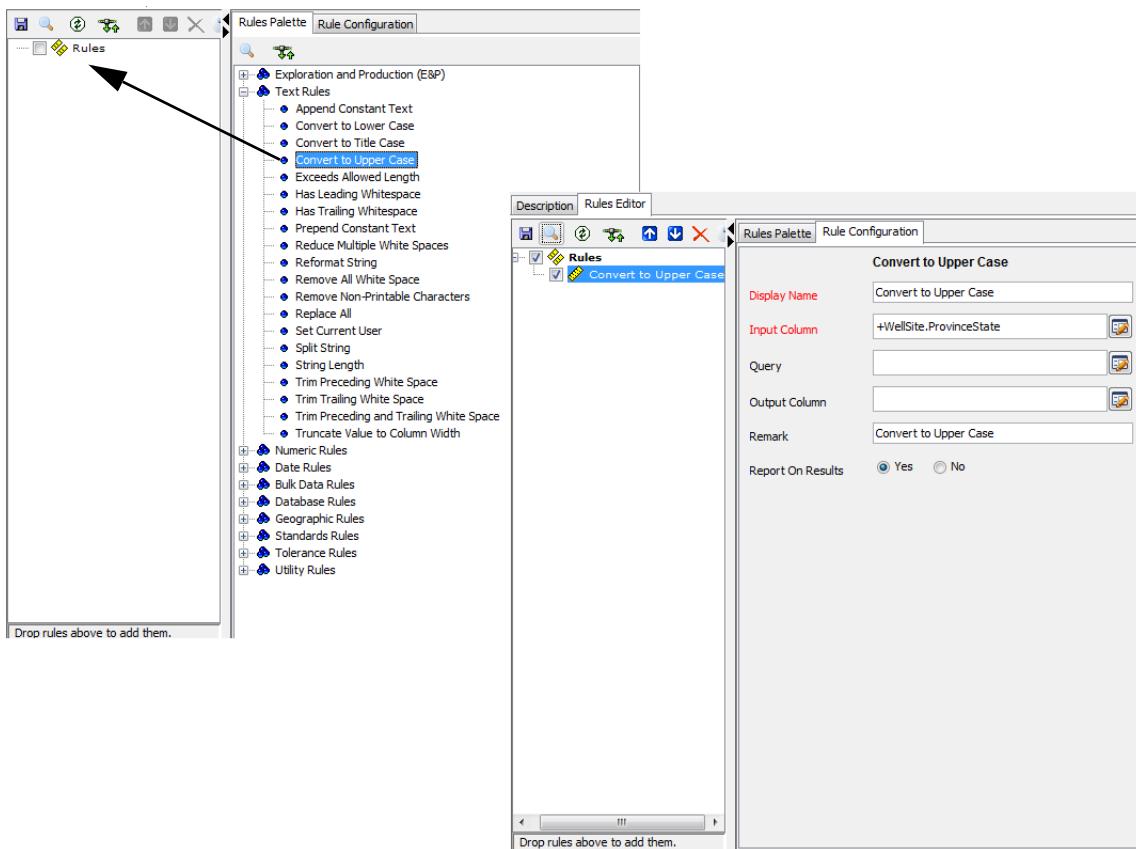
**16. The Requirement name field populates with the name assigned in step 13.**

17. Select **Medium** from the **Severity** drop-down list.
18. Enter **Advance Rule for State** column in the **Description** field.
19. Click  to save changes in the **Description** tab.
20. Select the **Rules Editor** tab adjacent to the **Description** tab.



21. Click  to expand the **Text Rules** tree in the **Rules Palette** tab.

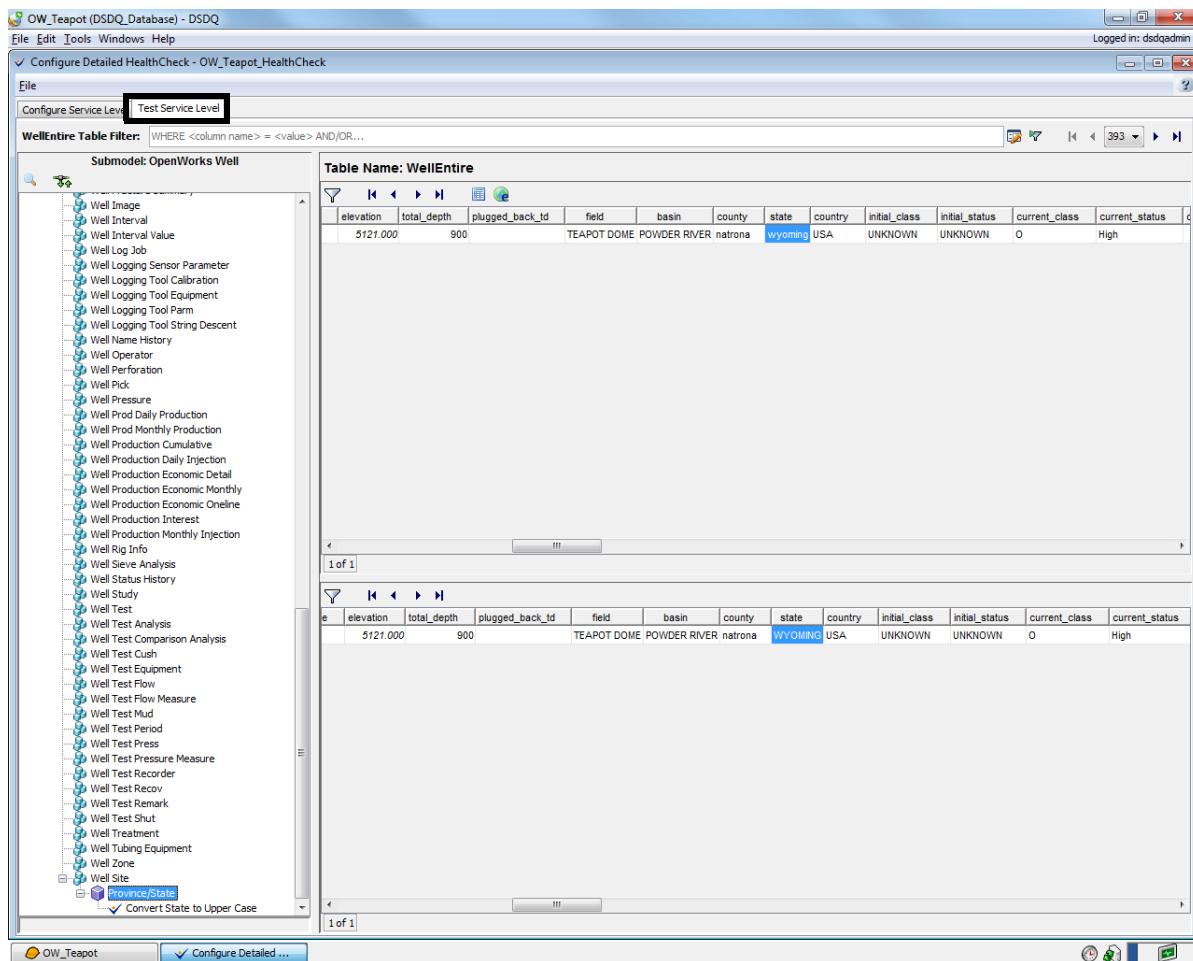
22. Drag and drop the **Convert to Upper Case** rule onto the **Rule** area.



23. Click to save changes in the **Rules Editor** tab.

24. Select the **Test Service Level** tab.

The test is automatically executed for the first record of the test data subset.



By looking at the columns that have been changed and temporary columns, you can verify that the behavior of the service level is correct prior to running the **Run Detailed HealthCheck Task**.

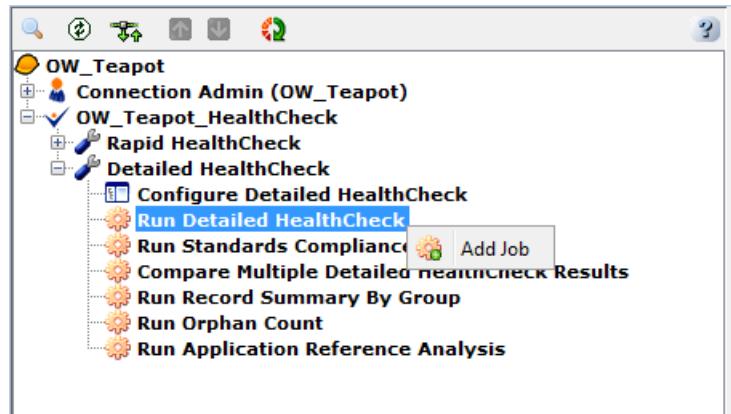
25. Click the **Next Data Set** button to test the next record.
26. Repeat step **25** to test all records.
27. Select **File > Exit** to close the **Configure Detailed HealthCheck** window.

### Exercise: Running the Detailed HealthCheck Task

The **Run Detailed HealthCheck** Task generates results for the requirements that are enabled in the service level. Prior to executing the **Run Detailed HealthCheck** Task, ensure that columns in the specified

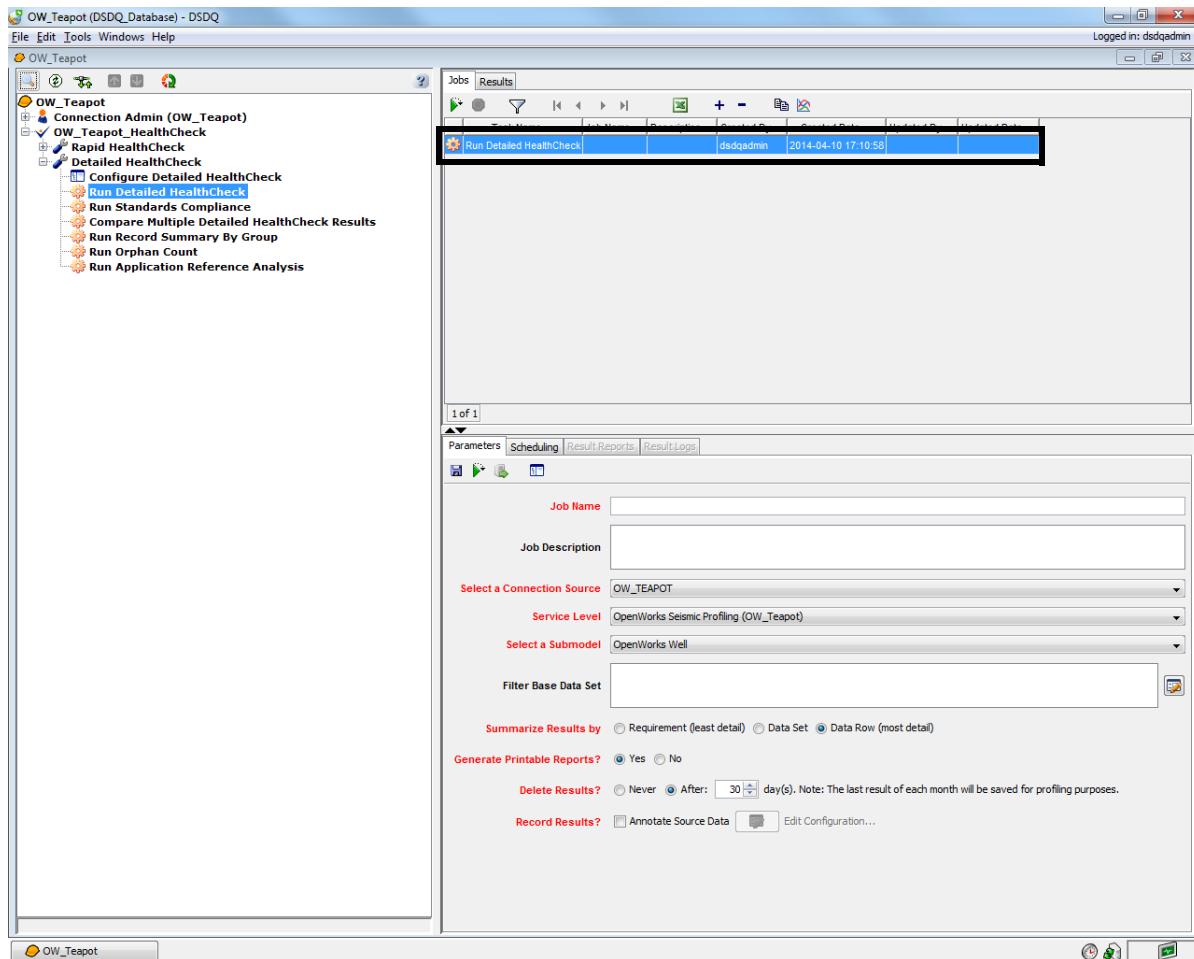
submodel table have been assigned elements from the desired service level. To run the Detailed HealthCheck task:

1. Double-click the **Run Detailed HealthCheck** Task or right-click the **Run Detailed HealthCheck** Task and select **Add Job** from the pop-up menu.



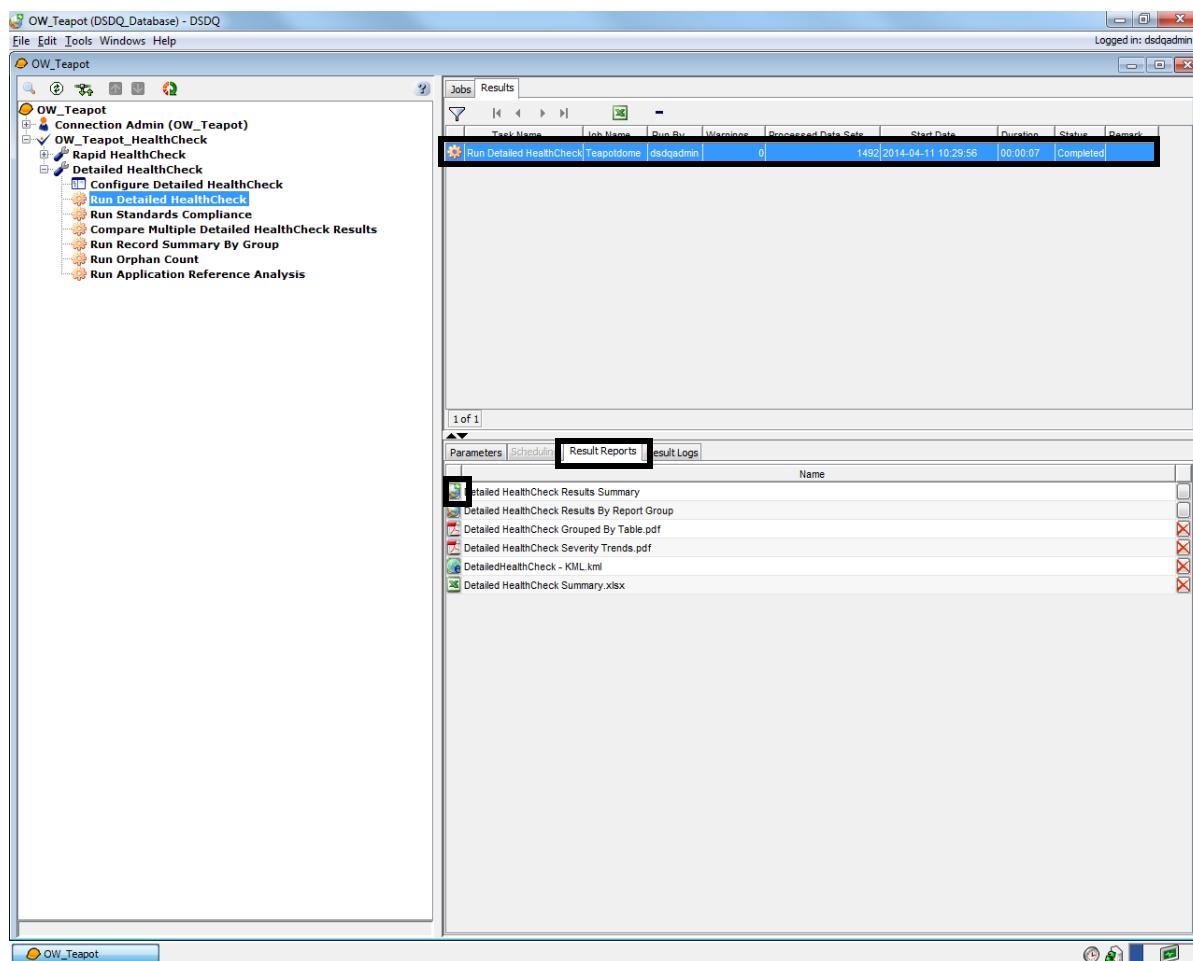
A new job is initiated and it displays on the **Job and Results**

**Information Pane** on the right side of the DecisionSpace Data Quality Project window.

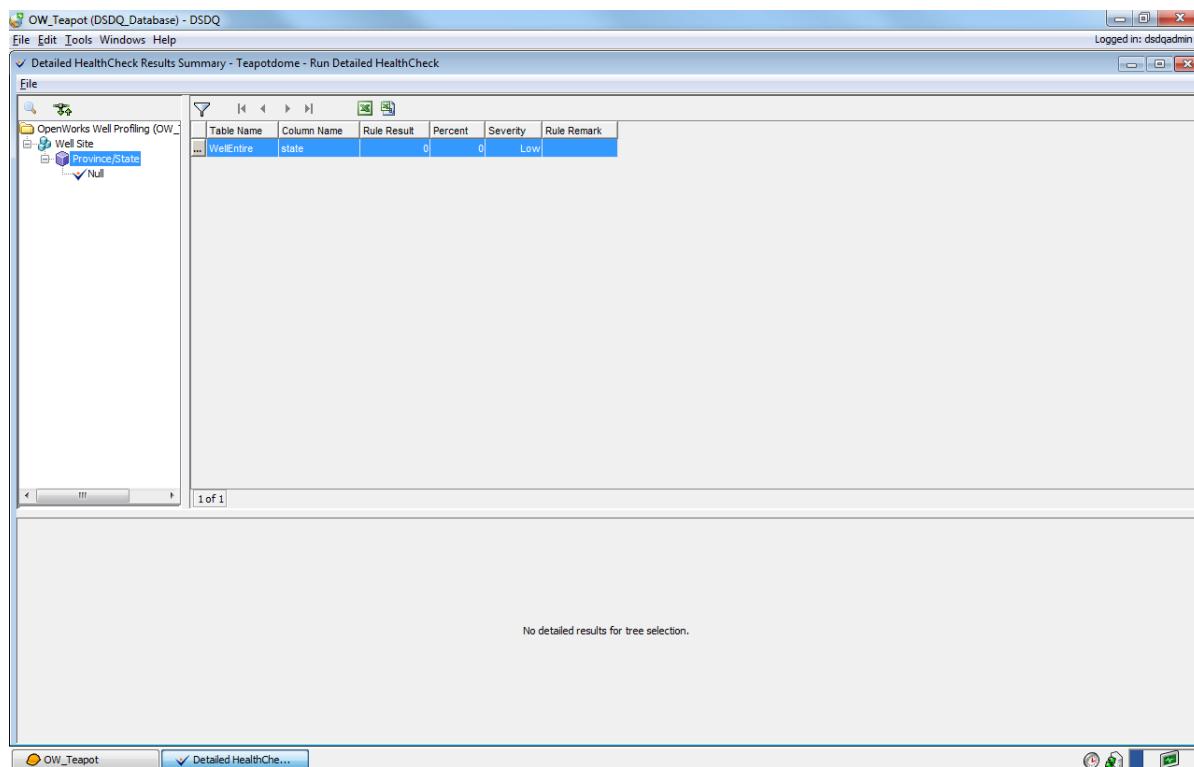


2. Enter **Teapotdome** in the **Job Name** field.
3. Enter **Detailed HealthCheck** for **OW\_Teapot** in the **Job Description** field.
4. Select **OW\_TEAPOT** from the **Select a Connection Source** drop-down list.
5. Select **OpenWorks Well Profiling (OW\_Teapot)** from the **Service Level** drop-down list.
6. Select **OpenWorks Well** from the **Select a Submodel** drop-down list.
7. Do not select the filter for **Filter Base Data Set**.

8. Select the **Data Row (most detail)** option for **Summarize Results by**.
9. Select the **Yes** option for **Generate Printable Reports**.
10. Select the **After** option for **Delete Results**. Leave the number of days as **30**.
11. Do not select the check box for **Record Results**.
12. Click  to save changes in the **Parameters** tab.
13. Click  to run the job.  
The **Run Detailed HealthCheck** Task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.
14. Select the **Results** tab.  
The **Jobs and Results Listing Pane** displays a list of results.



15. Click the **Open Basic View Frame**  button on the **Result Reports** tab to display the **Run Detailed HealthCheck** Task results in the **Basic View Frame** window.



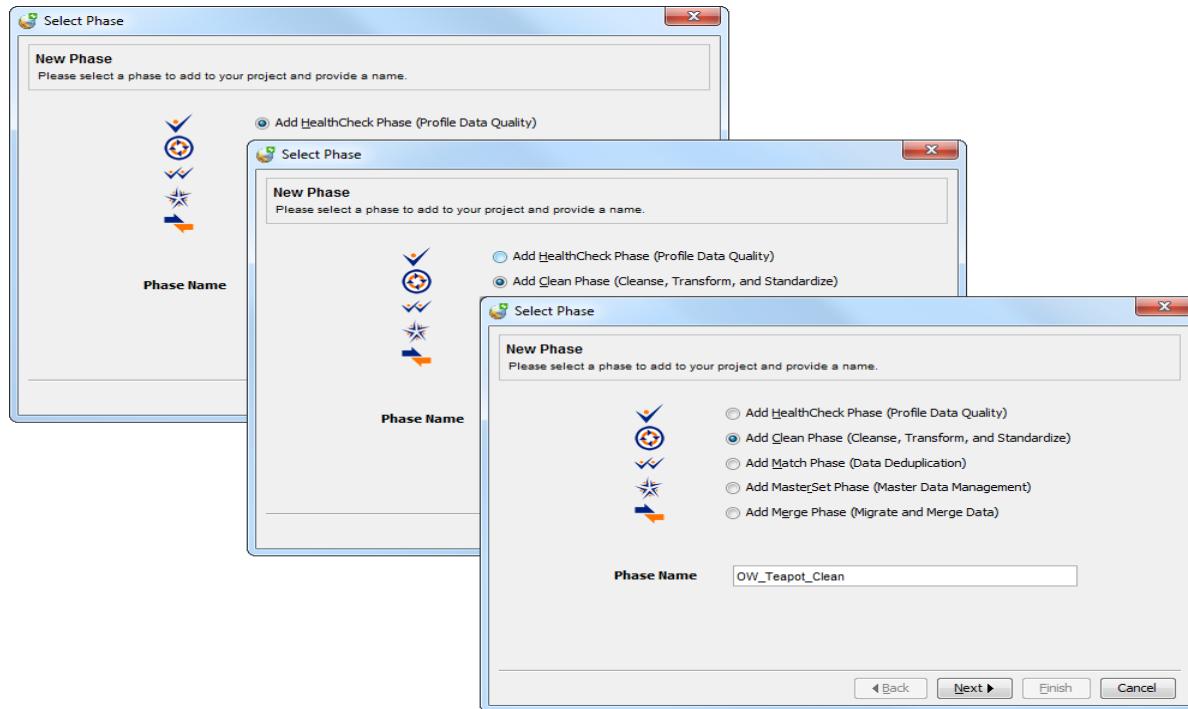
16. Select **File > Exit** to close the **Basic View Frame** window.

# Resolving Data Quality Issues using the Clean Phase

## Adding a Clean Phase

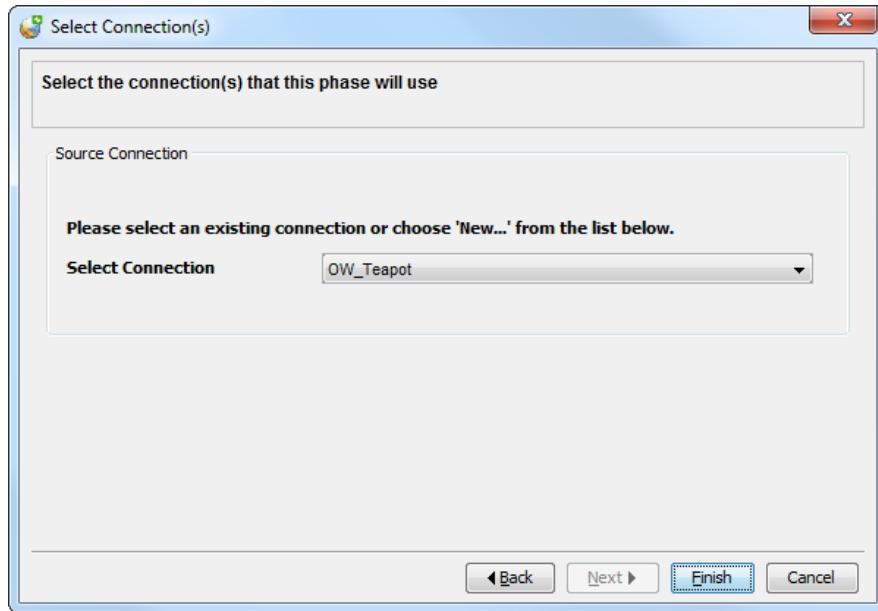
To add a Clean Phase:

1. Click the **Add New Phase** button on the Project Toolbar. The **Select Phase** window appears with the **Add HealthCheck Phase (Report on Data Quality Profiling)** option selected by default.



2. Select the **Add Clean Phase (Cleanse, Transform, and Standardize)** option.
3. Enter **OW\_Teapot\_Clean** in the **Phase Name** field.

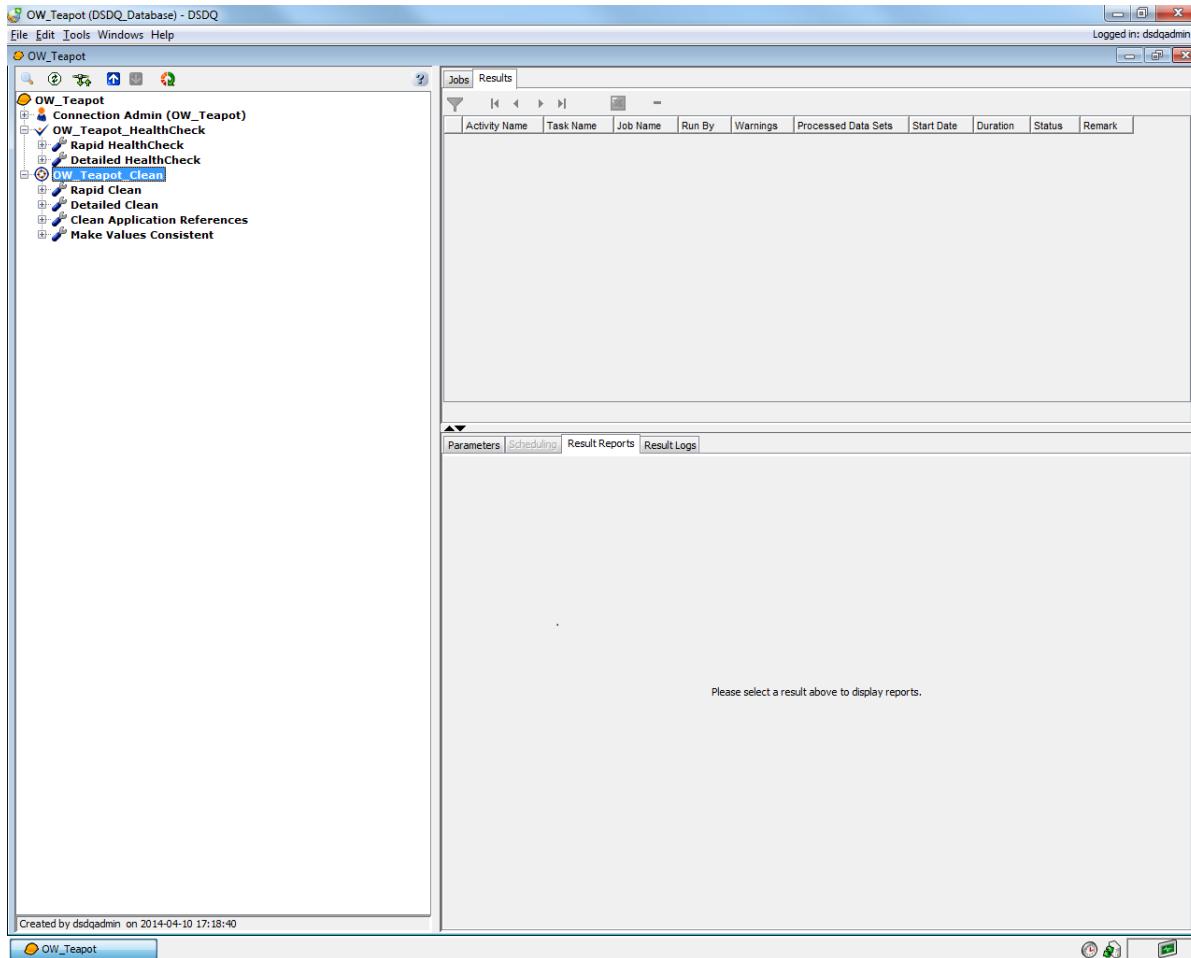
4. Click **Next** to continue.  
The **Select Connection(s)** window appears.



5. Select **OW\_Teapot** from the **Select Connection** drop-down list.

**6. Click Finish.**

The **Clean** Phase is created and displays in the DecisionSpace Data Quality Project window.



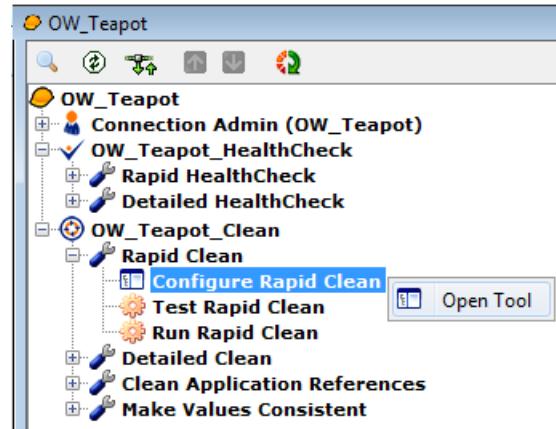
## Rapid Clean Activity

The **Rapid Clean** Activity cleans out data issues in selected submodel table columns.

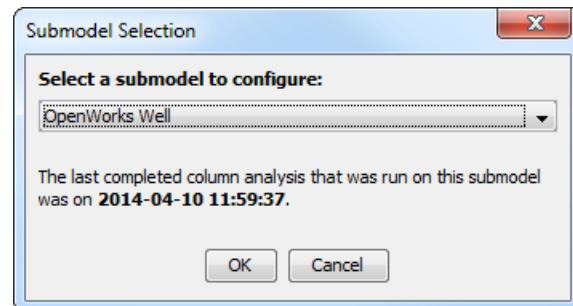
### Exercise: Configuring the Rapid Clean Tool

Prior to running the **Configure Rapid Clean Tool**, ensure that the **Run Column Analysis on Modeled Table** Task has been run. To configure the Rapid Clean Tool:

1. Click  to expand the **Rapid Clean** Activity.
2. Double-click the **Configure Rapid Clean** Tool or right-click the **Configure Rapid Clean** Tool and select **Open Tool** from the pop-up menu.



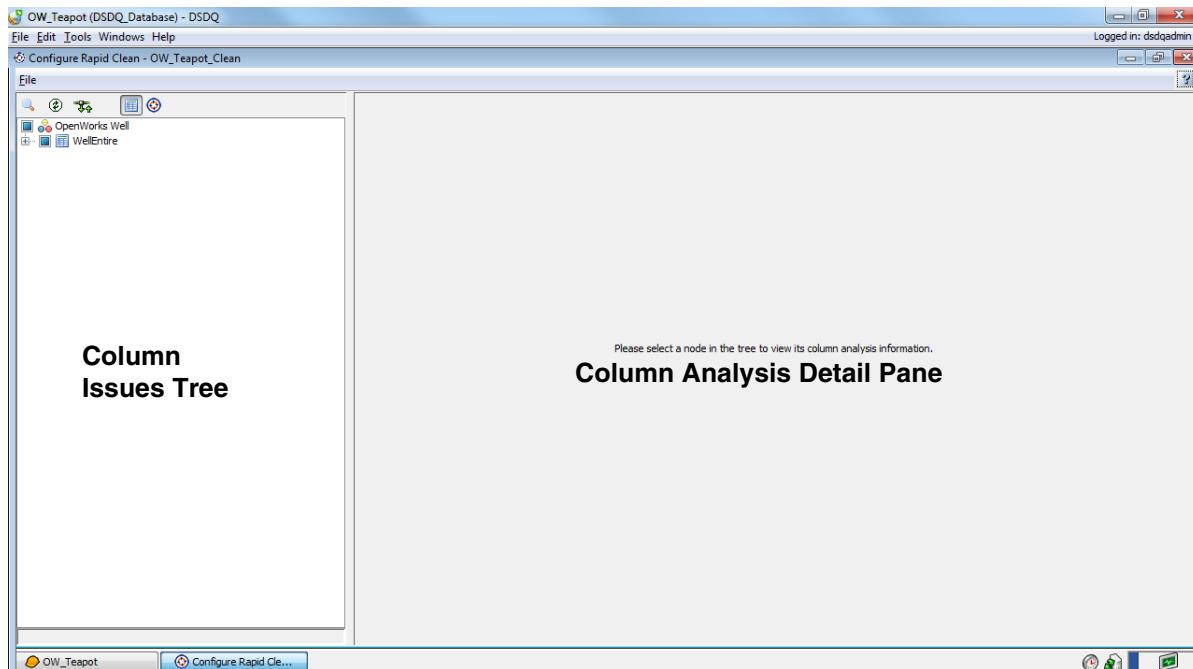
The **Submodel Selection** dialog box appears.



3. Select **OpenWorks Well** from the **Select a submodel** to configure drop-down list.

- Click **OK** to continue.

The **Configure Rapid Clean - OW\_Teapot\_Clean** window appears.



- Click **+** to expand the **WellEntire** table.

- Click **+** to expand the **Mixed Case** issues.

- Click the **State** column.

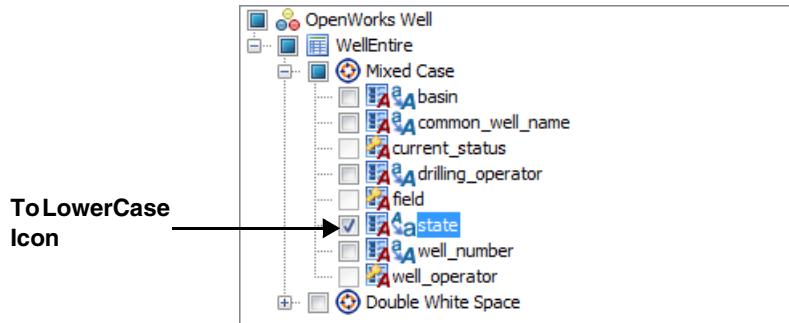
Any issue(s) for the **WellEntire** table highlight in the **Column Analysis Details Pane**.

**Table ToUpperCase Icon Issue**

Table Name	Column Name	Data Type	Rows	# Not Null	% Populated	# Unique	# Mixed Case	# NPC	# PWS	# TWS	# DI
WellEntire	basin	STRING	1492	1492	100	3	30	0	0	0	
WellEntire	current_status	STRING	1492	1492	100	2	1492	0	0	0	
WellEntire	drilling_operator	STRING	1492	1492	100	33	49	0	0	0	
WellEntire	field	STRING	1492	1492	100	9	50	0	0	0	
WellEntire	remark	STRING	1308	1308	87	41	0	0	0	0	
WellEntire	state	STRING	1492	1492	100	2	1492	0	0	0	
WellEntire	well_number	STRING	1380	92	1254	131	0	0	0	0	
WellEntire	wel_operator	STRING	1492	1492	100	22	114	0	0	0	

- Select the check box adjacent to the **State** column.

9. Click the **To Uppercase**  icon adjacent to the **State** column. The **To Uppercase**  icon changes to the **To LowerCase**  icon.

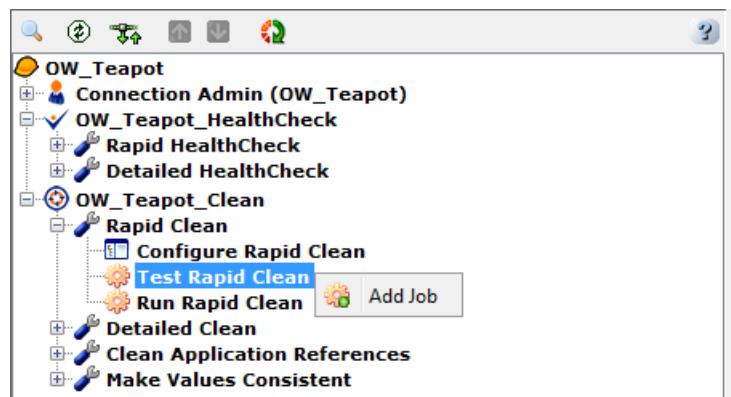


10. Select **File > Exit** to close the **Configure Rapid Clean** window.

### Exercise: Running the Test Rapid Clean Task

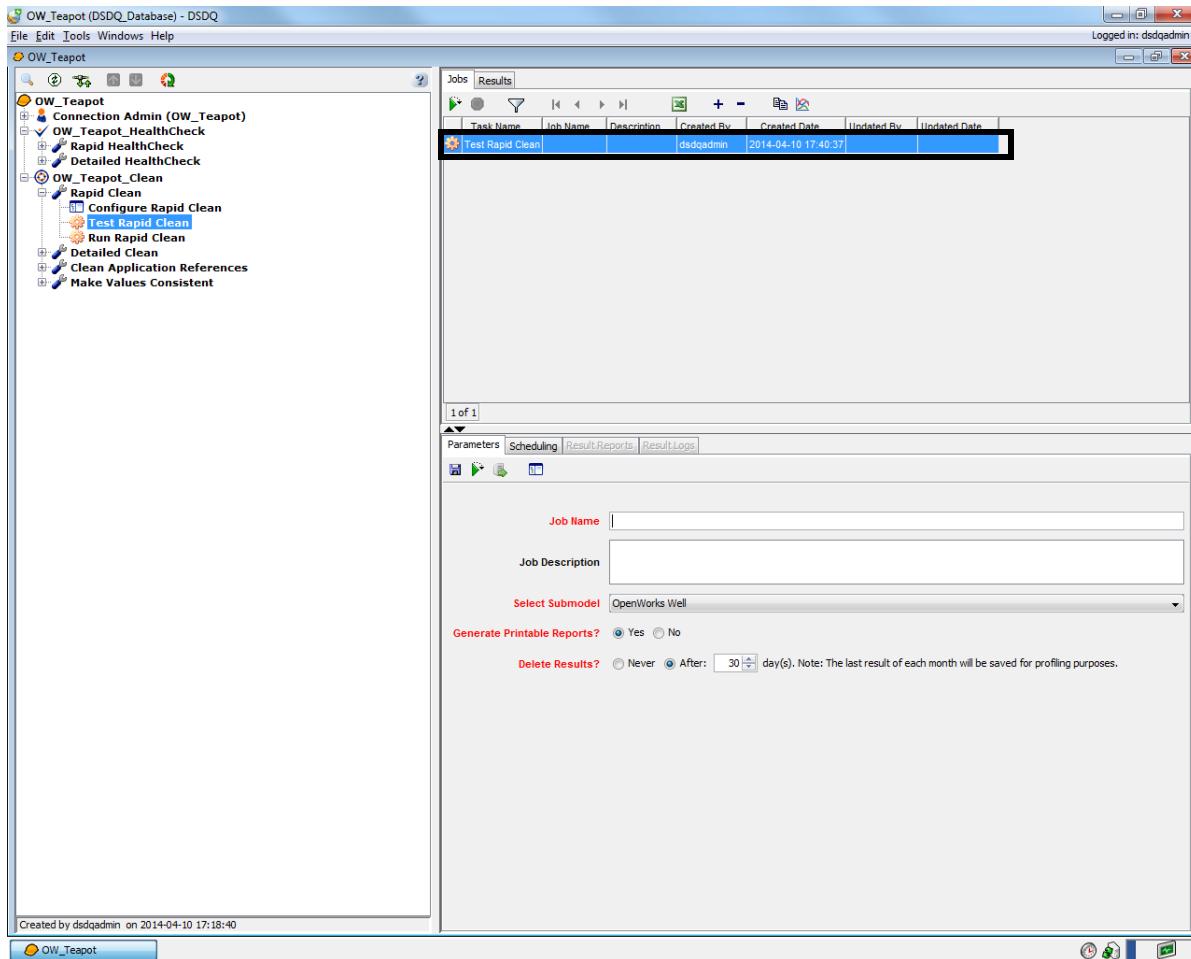
After issues have been selected to be cleaned using the **Configure Rapid Clean Tool**, the **Test Rapid Clean Task** is run to make sure that the expected results are seen before running the **Run Rapid Clean Task** to fix the entire dataset in the submodel. To run the **Test Rapid Clean Task**:

1. Double-click the **Test Rapid Clean Task** or right-click the **Test Rapid Clean Task** and select **Add Job** from the pop-up menu.



A new job is initiated and it displays on the **Job and Results**

**Information Pane** on the right side of the DecisionSpace Data Quality Project window.



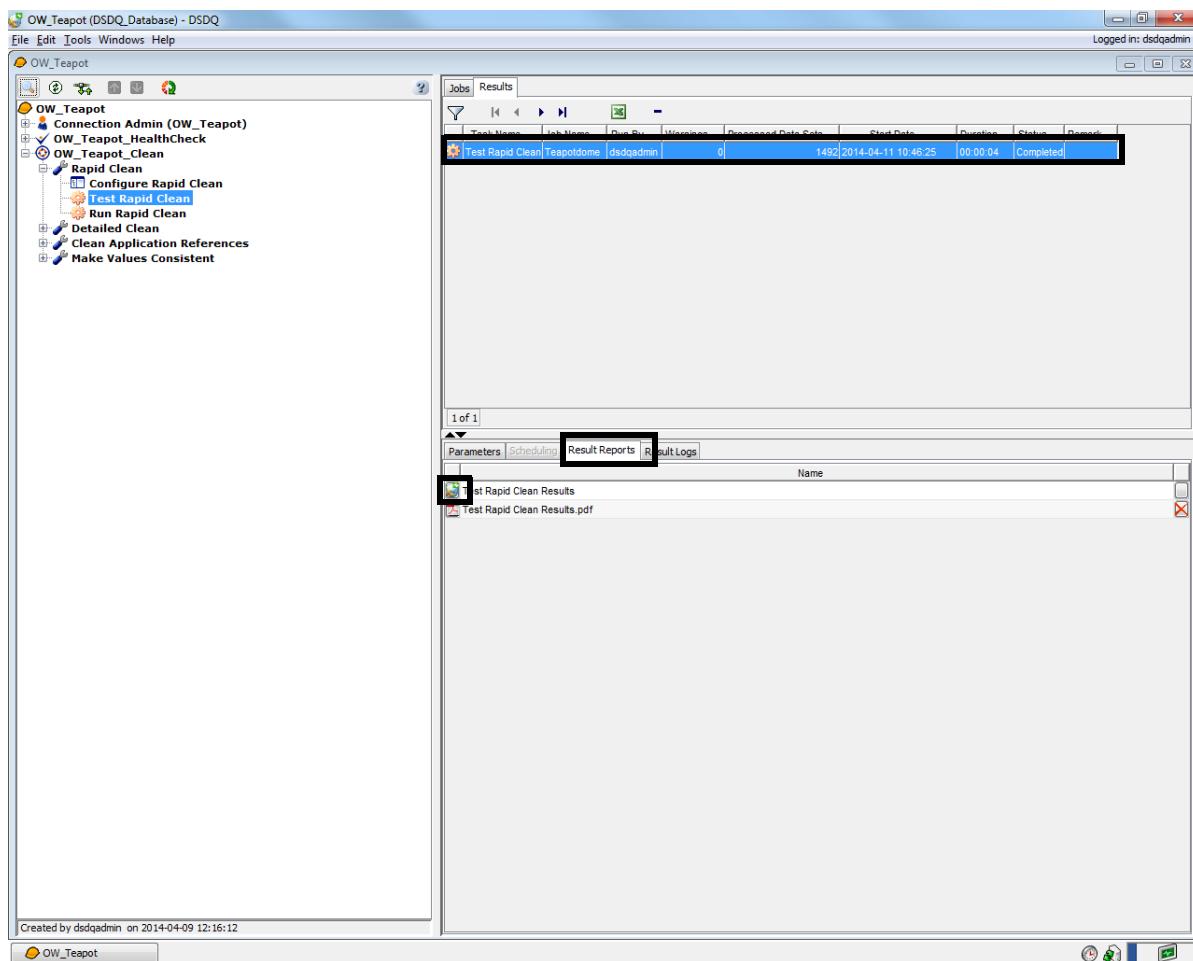
2. Enter **Teapotdome** in the **Job Name** field.
3. Enter **OW\_Teapot State Name Conversion** in the **Job Description** field.
4. Select **OpenWorks Well** from the **Select Submodel** drop-down list.
5. Select the **Yes** option for **Generate Printable Reports**.
6. Select the **After** option for **Delete Results**. Leave the number of days as **30**.
7. Click to save changes in the **Parameters** tab.

8. Click  to run the job.

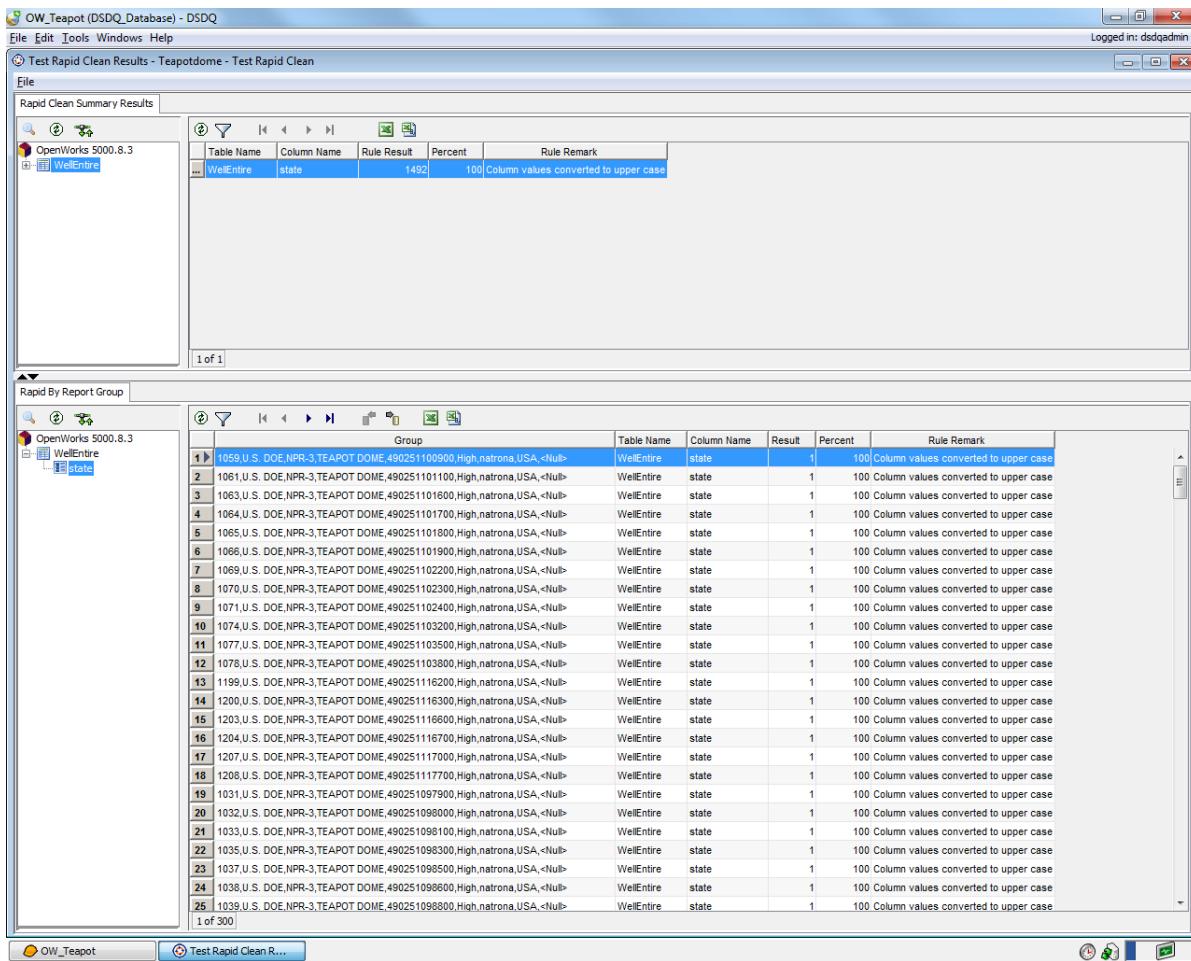
The **Test Rapid Clean** Task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

9. Select the **Results** tab.

The **Job and Results Listing Pane** displays the list of results.



10. Click the **Open Basic View Frame**  button on the **Result Reports** tab to display the **Test Rapid Clean** Task results in the **Basic View Frame** window.

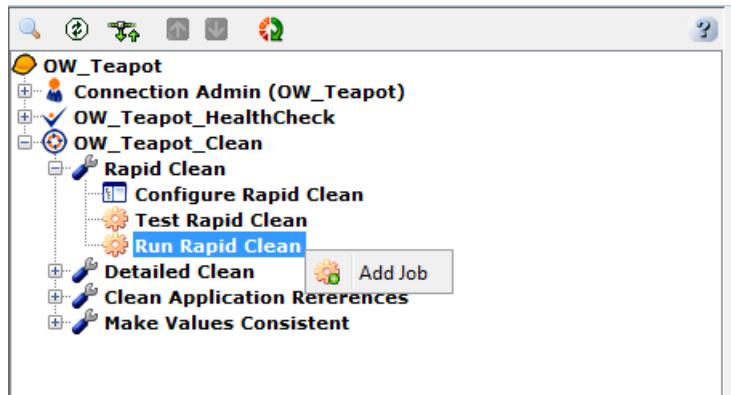


11. Select **File > Exit** to close the **Basic View Frame** window.

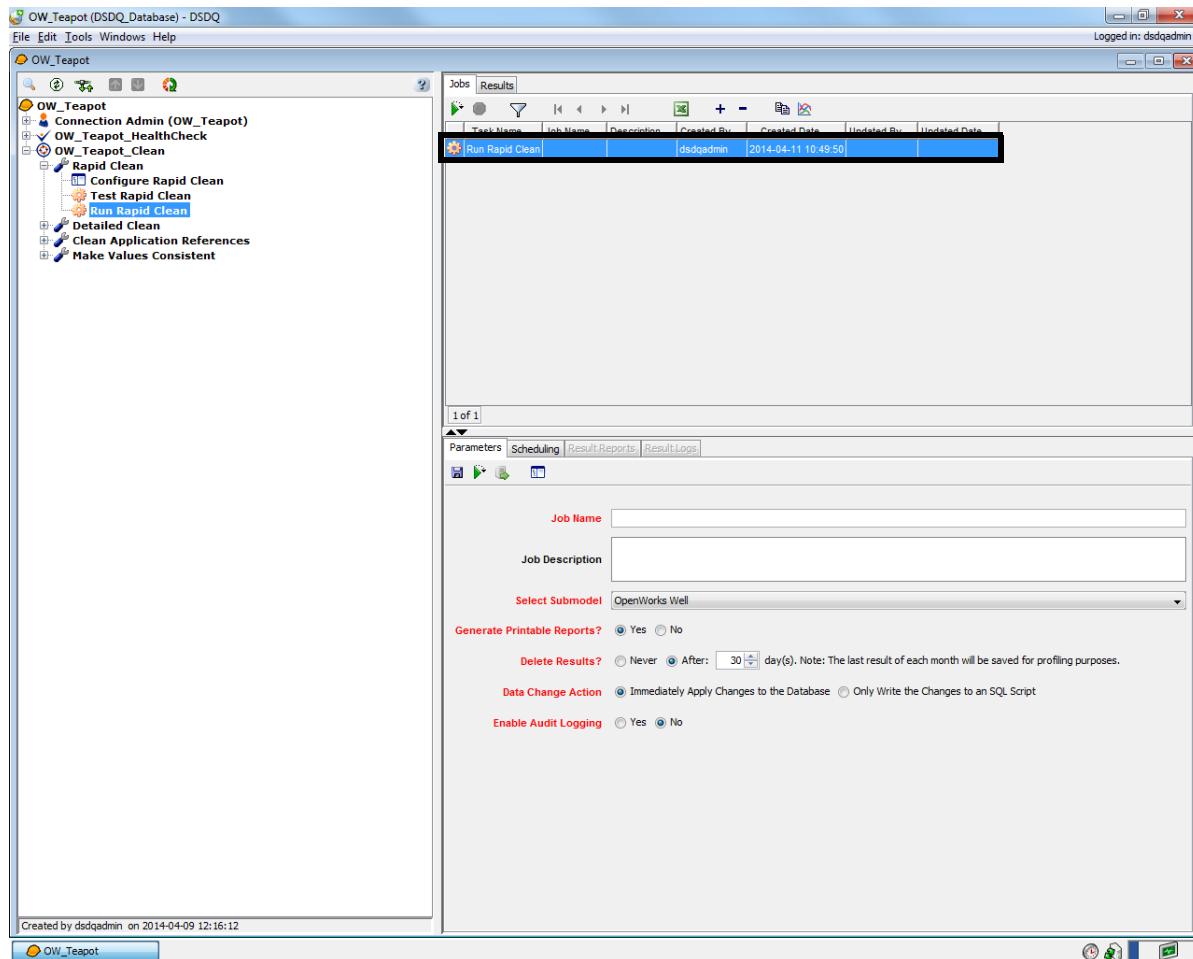
### Exercise: Running the Rapid Clean Task

The **Run Rapid Clean** Task fixes the issues that were selected in the **Configure Rapid Clean** Tool for the specific submodel. To run the Rapid Clean task:

1. Double-click the **Run Rapid Clean** Task or right-click the **Run Rapid Clean** Task and select **Add Job** from the pop-up menu.



A new job is initiated and it displays on the **Job and Results Information Pane**.



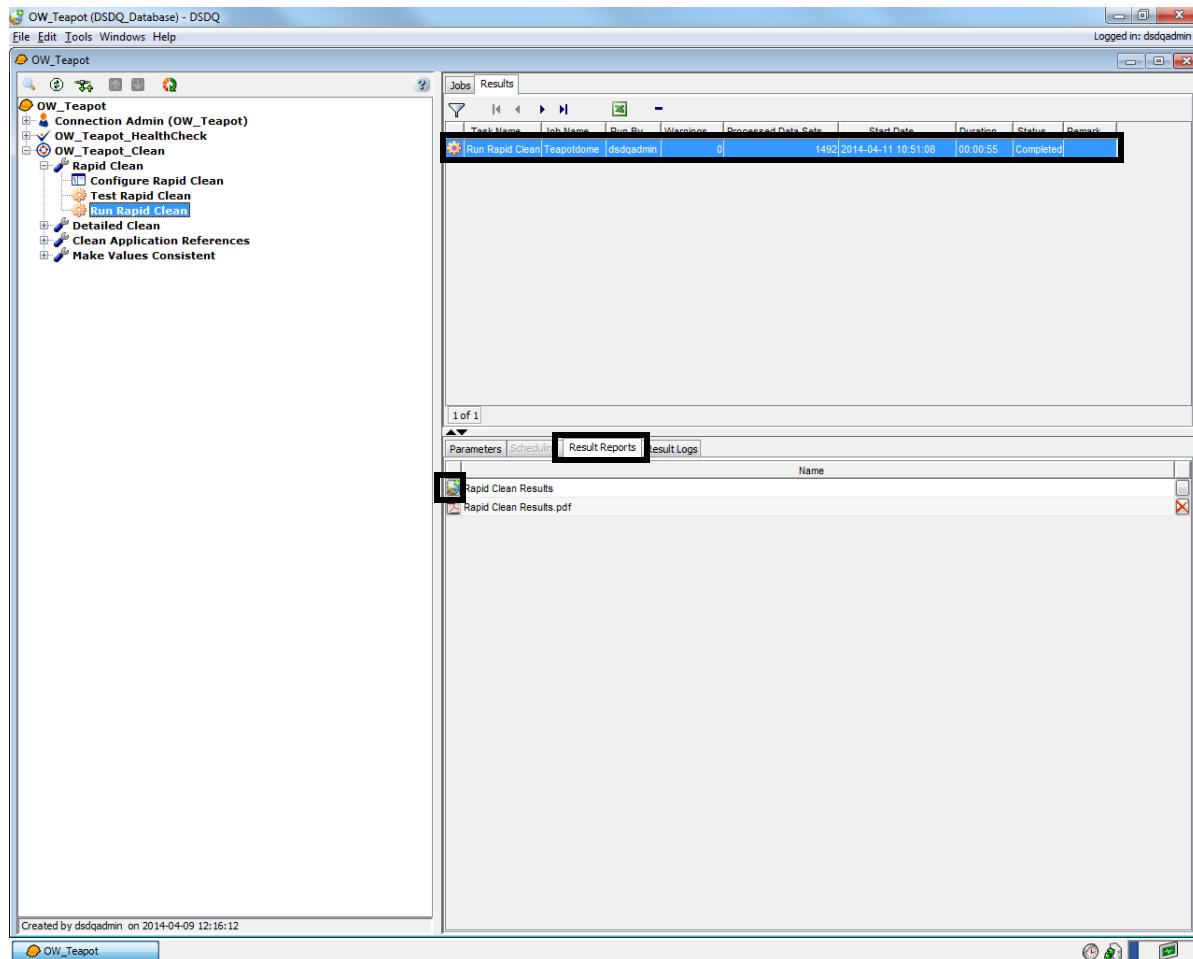
2. Enter **Teapotdome** in the **Job Name** field.
3. Enter **OW\_Teapot Case Convert** in the **Job Description** field.

4. Select **OpenWorks Well** from the **Select Submodel** drop-down list.
5. Select the **Yes** option for **Generate Printable Reports**.
6. Select the **After** option for **Delete Results**. Leave the number of days as **30**.
7. Select the **Immediately Apply Changes to the Database** option for **Data Change Action**.
8. Select the **No** option for **Enable Audit Logging**.
9. Click  to save changes in the **Parameters** tab.
10. Click  to run the job.

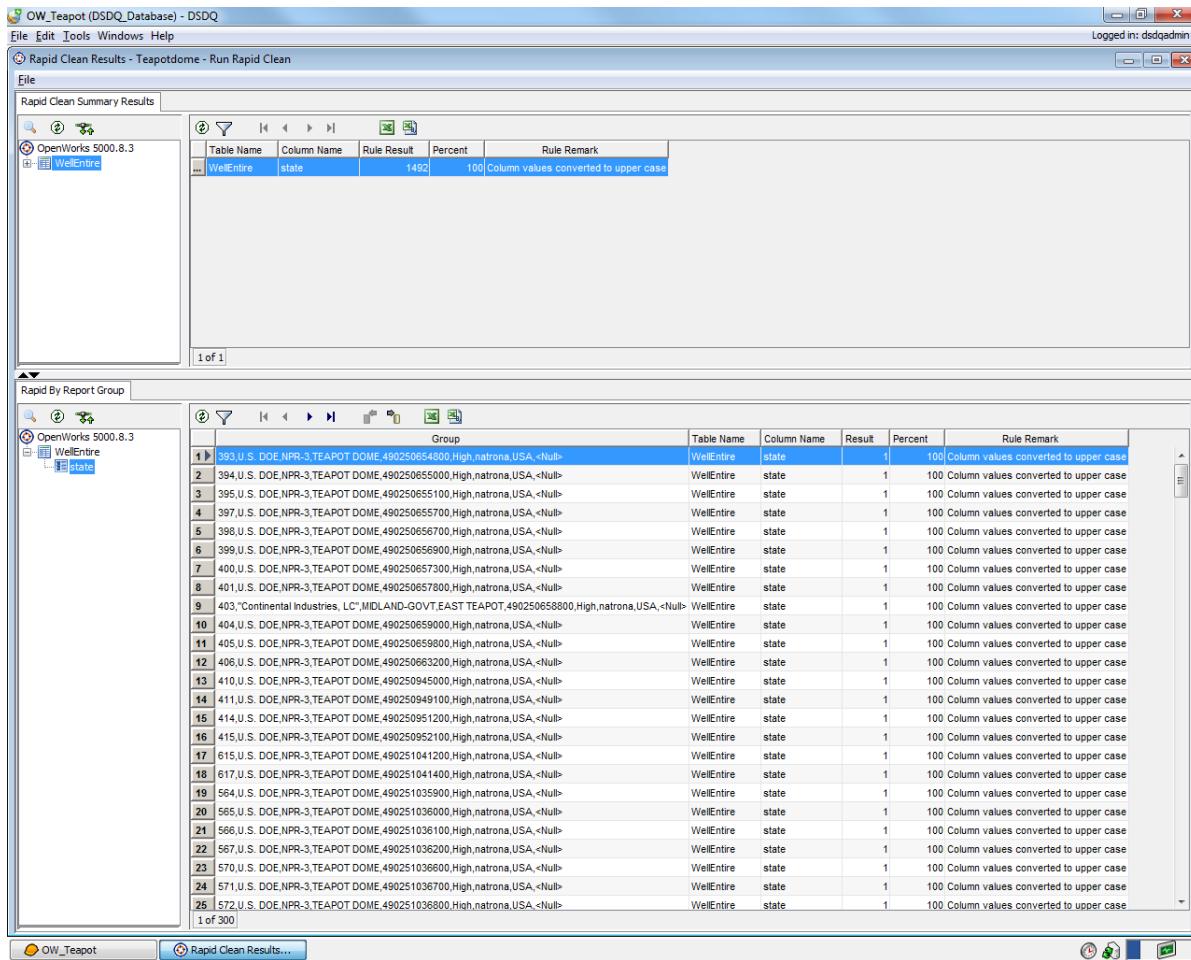
The **Run Rapid Clean Task** runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

11. Select the **Results** tab.

The **Job and Results Listing Pane** displays a list of results.



12. Click the **Open Basic View Frame**  button on the **Result Reports** tab to display the **Run Rapid Clean** Task results in the **Basic View Frame** window.



13. Select **File > Exit** to close the **Basic View Frame** window.

## Detailed Clean Activity

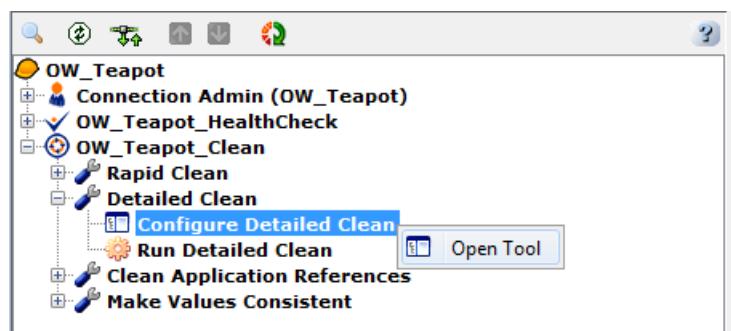
The **Detailed Clean** Activity helps in assigning columns to the clean requirements and testing service levels.

### Exercise: Configuring the Detailed Clean Tool

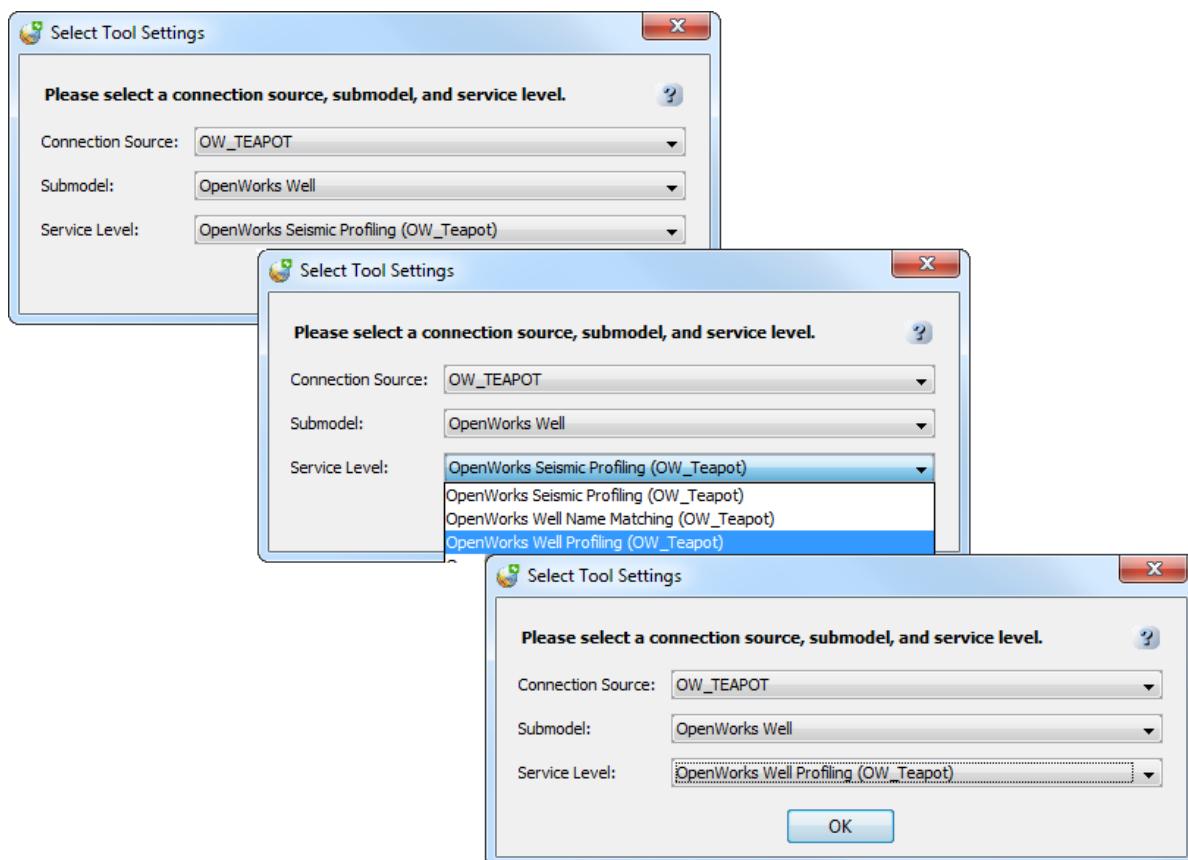
The **Configure Detailed Clean Tool** is used to configure service levels for testing prior to running the **Run Detailed Clean Task**. You can select which requirements in the service level to enable/disable, and when testing a service level, what subset of the total data to use. A

service level containing clean requirements must exist prior to running the **Configure Detailed Clean** Tool.

1. Double-click the **Configure Detailed Clean** Tool or right-click the **Configure Detailed Clean** Tool and select **Open Tool** from the pop-up menu.

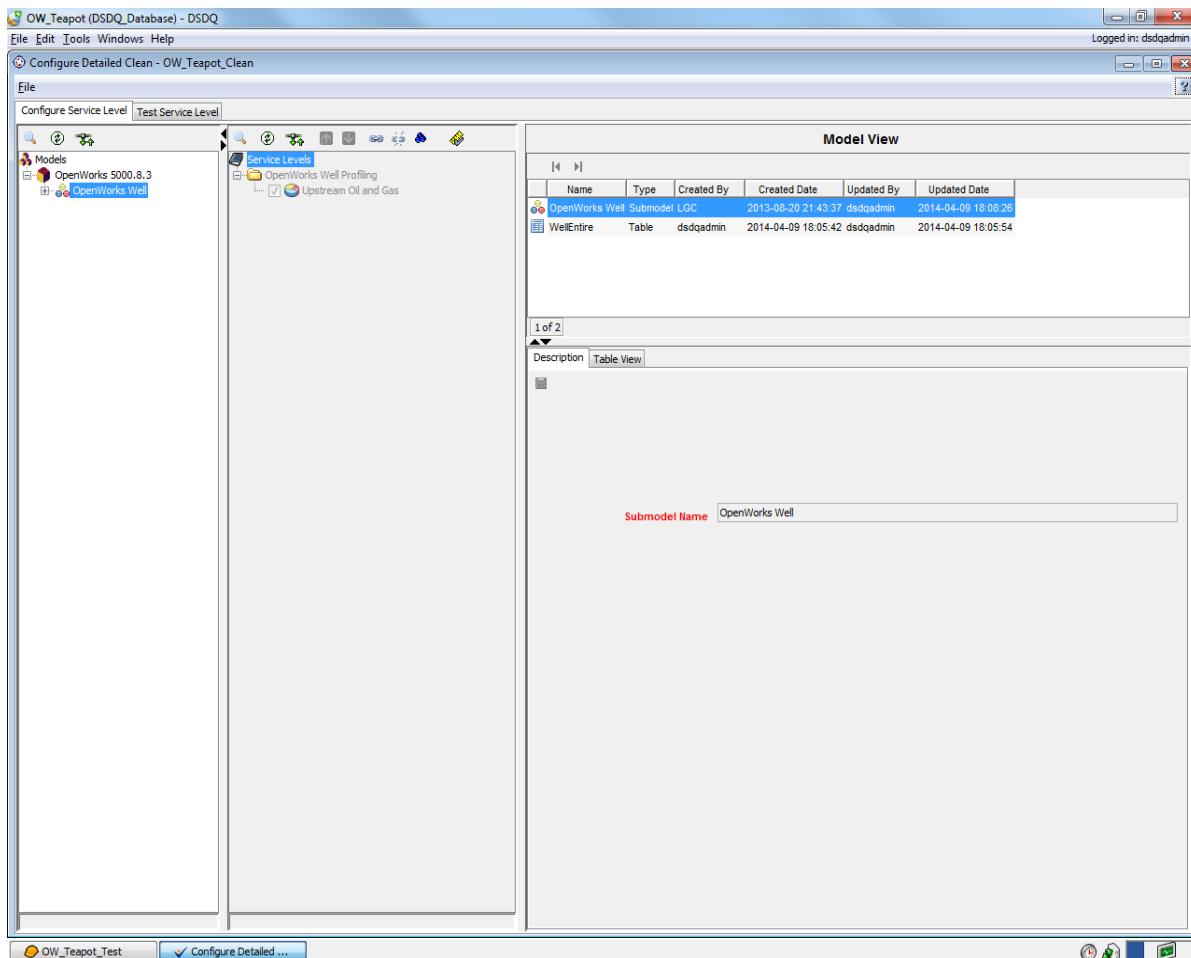


The **Select Tool Settings** window appears:



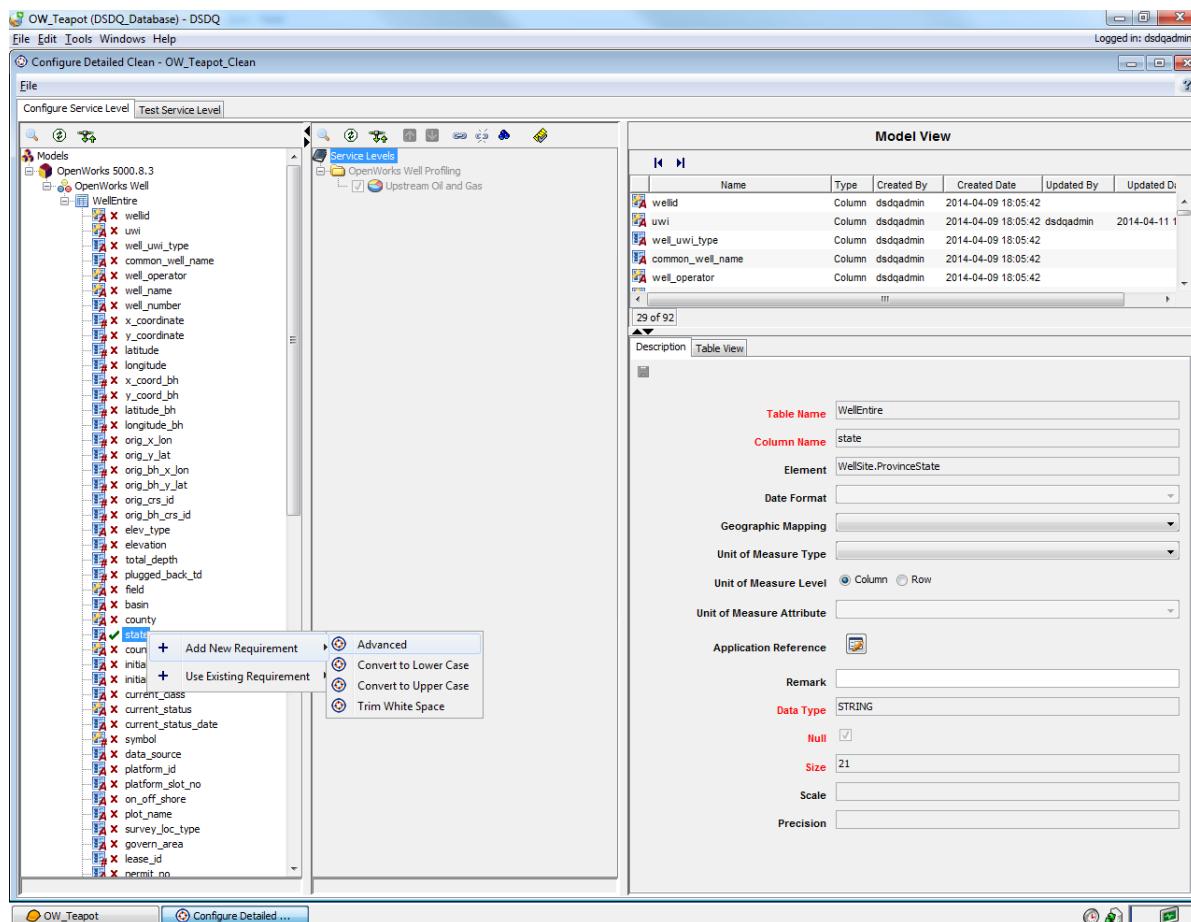
2. The **Connection Source** drop-down list is set to **OW\_Teapot** by default.

3. The **Submodel** drop-down list is set to **OpenWorks Well** by default.
  4. Select **OpenWorks Well Profiling (OW\_Teapot)** from the **Service Level** drop-down list.
  5. Click **OK**.
- The **Configure Detailed Clean - OW\_Teapot\_Clean** window appears.

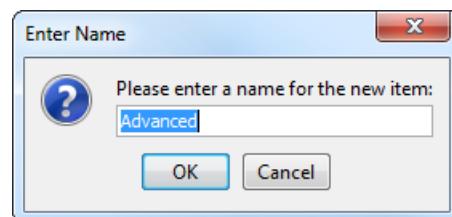


6. Click to expand the **OpenWorks Well** submodel in the Data Model Tree.
7. Click to expand the **WellEntire** table in the Data Model Tree.

8. Right-click the **State** column in the Data Model Tree and select **Add New Requirement > Advanced** from the pop-up menu.



The **Enter Name** dialog box appears.



9. Enter **Convert State to Title Case** in the **Please enter a name for the new item** dialog box.

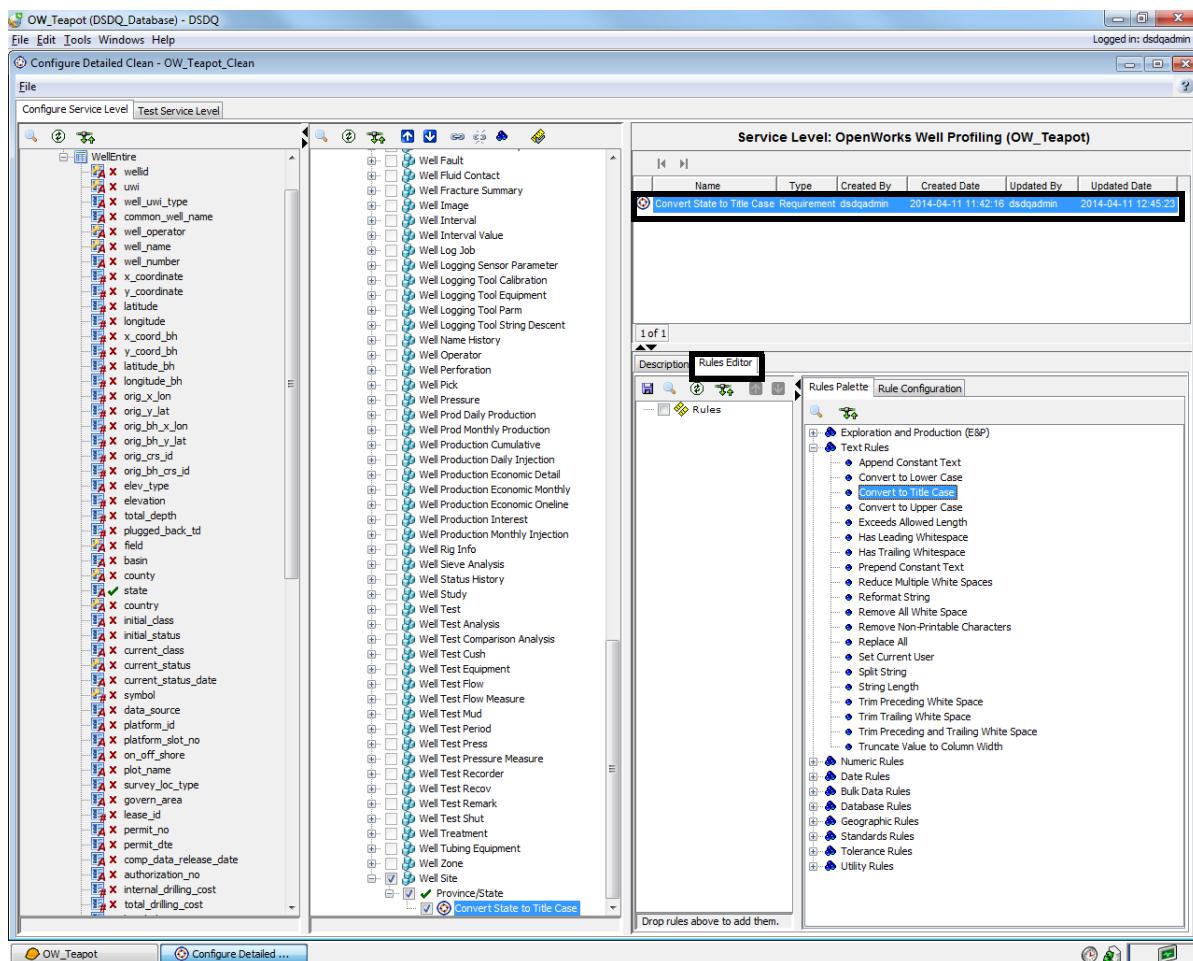


**10. Click OK.**

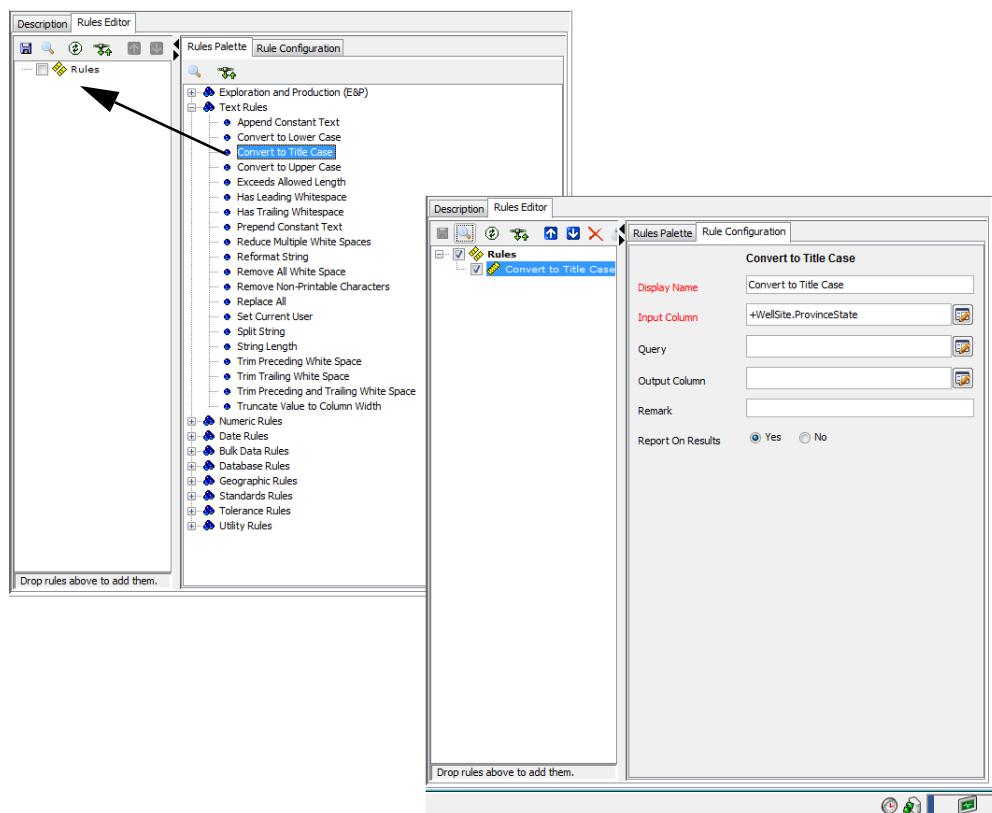
The **Convert State to Tile Case** requirement is added to the **State** column and displays in the **Model View Pane**.

**11. Select the Rules Editor tab adjacent to the Description tab.**

**12. Click  to expand the Text Rules in the Rule Palette tab.**



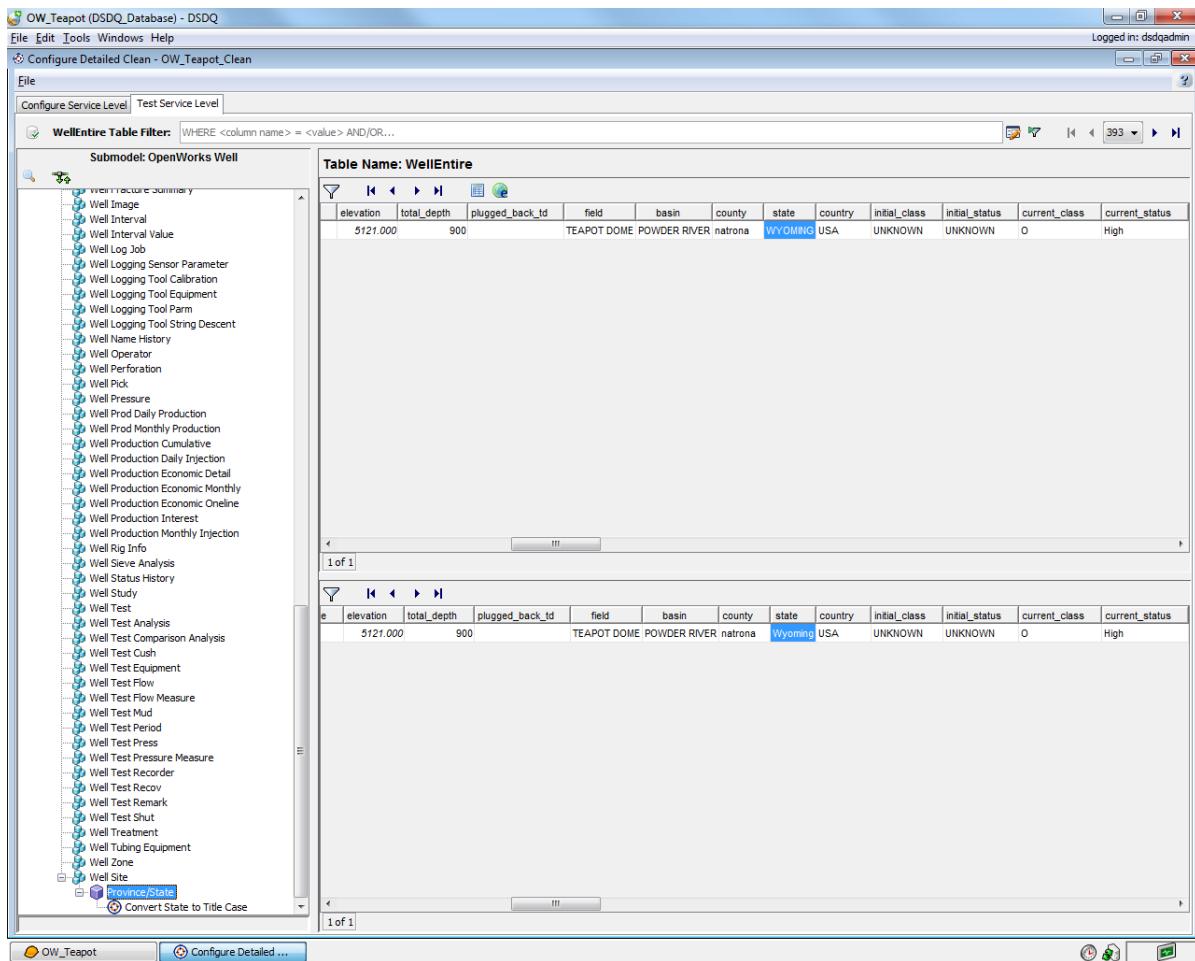
13. Drag and drop the **Convert to Title Case** onto the **Rules** area.



14. Click to save changes in the **Rule Editor** tab.

15. Select the **Test Service Level** tab.

The test is automatically executed for the first record of the test data subset.



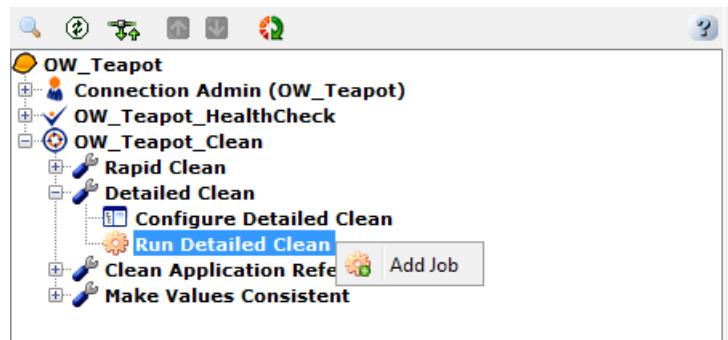
By looking at the columns that have been changed and temporary columns, you can verify that the behavior of the service level is correct prior to running the **Run Detailed Clean Task**.

16. Click the **Next Data Set**  button to test the next record.
17. Repeat step **16** to test all records.
18. Select **File > Exit** to close the **Configure Detailed Clean** window.

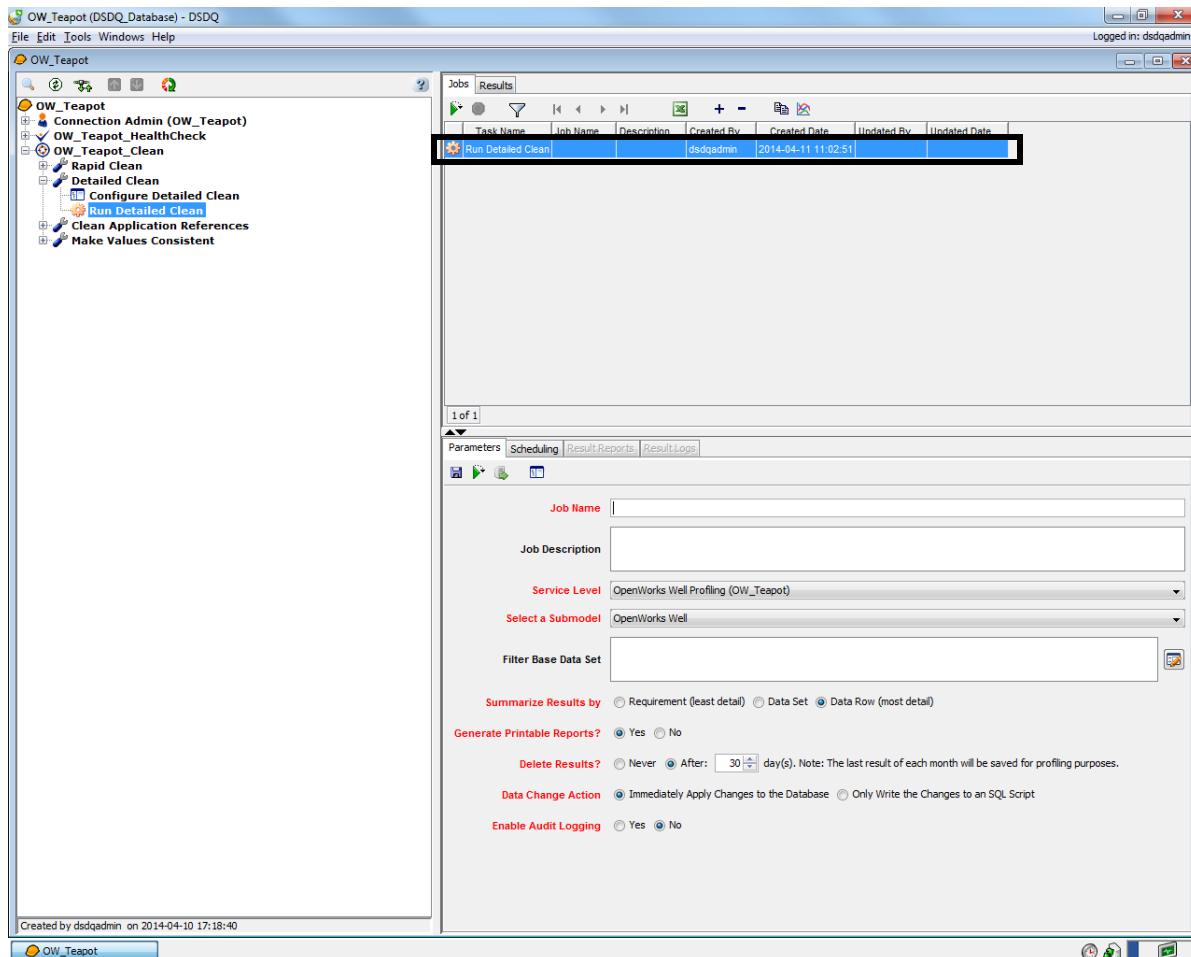
### Exercise: Running the Detailed Clean Task

The **Run Detailed Clean** Task performs cleansing of entire data and updates the actual data. To run the Detailed Clean task:

1. Double-click the **Run Detailed Clean** Task on the DecisionSpace Data Quality Tree or right-click the **Run Detailed Clean** Task and select **Add Job** from the pop-up menu.



A new job is initiated and it displays on the **Jobs and Results Listing Pane**.



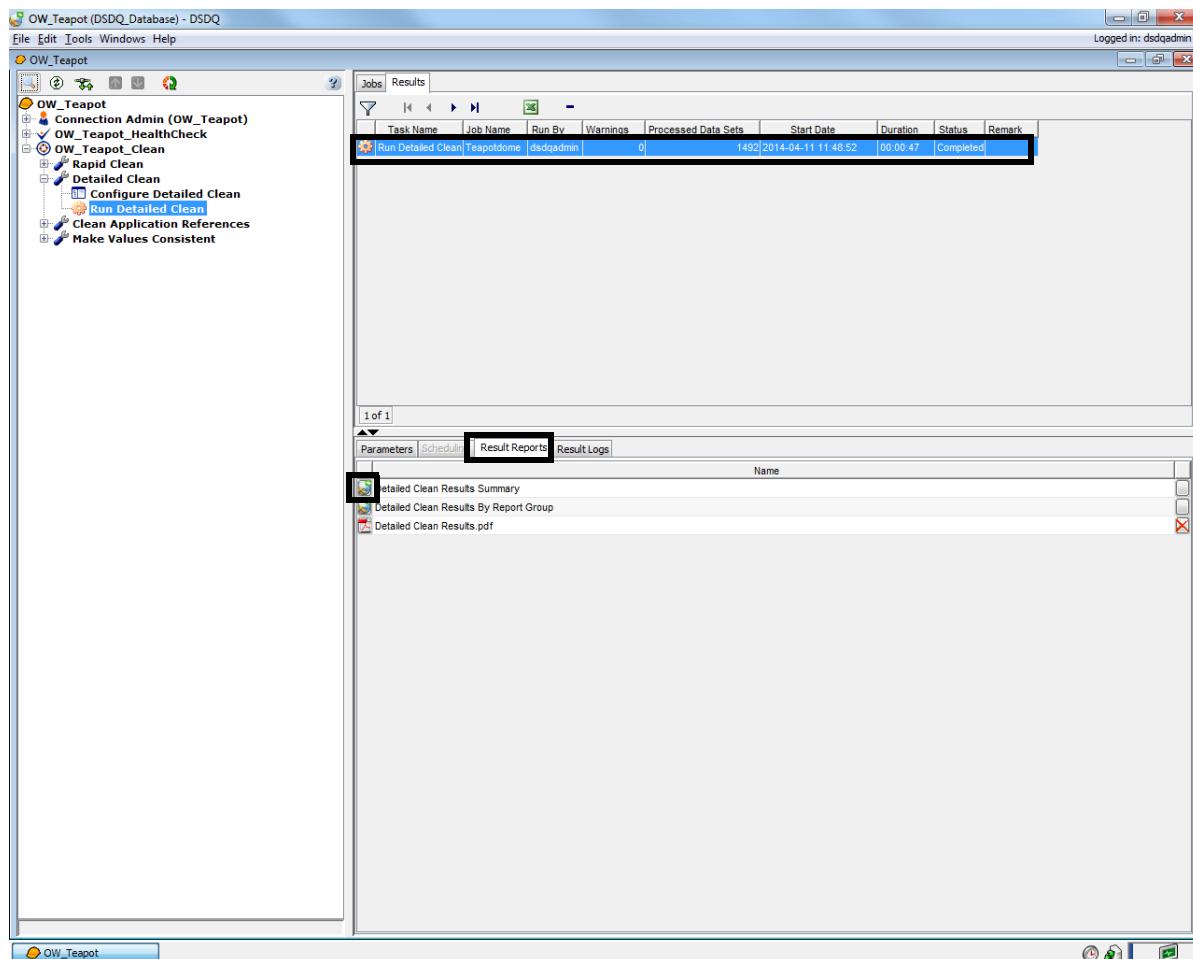
2. Enter **Teapotdome** in the **Job Name** field.

3. Enter **Run Detailed Clean for OW\_Teapot State to Title Case** in the **Job Description** field.
4. Select **OpenWorks Well Profiling (OW\_Teapot)** from the **Service Level** drop-down list.
5. Select **OpenWorks Well** from the **Select Submodel** drop-down list.
6. Do not set the filter for **Filter Base Data Set**.
7. Select the **Data Row (most detail)** option for **Summarize Results by**.
8. Select the **Yes** option for **Generate Printable Reports**.
9. Select the **After** option for **Delete Results**. Leave the number of days as **30**.
10. Select the **Immediately Apply Changes to the Database** option for **Data Change Action**.
11. Select the **No** option for **Enable Audit Logging**.
12. Click  to save changes in the **Parameters** tab.
13. Click  to run the job.

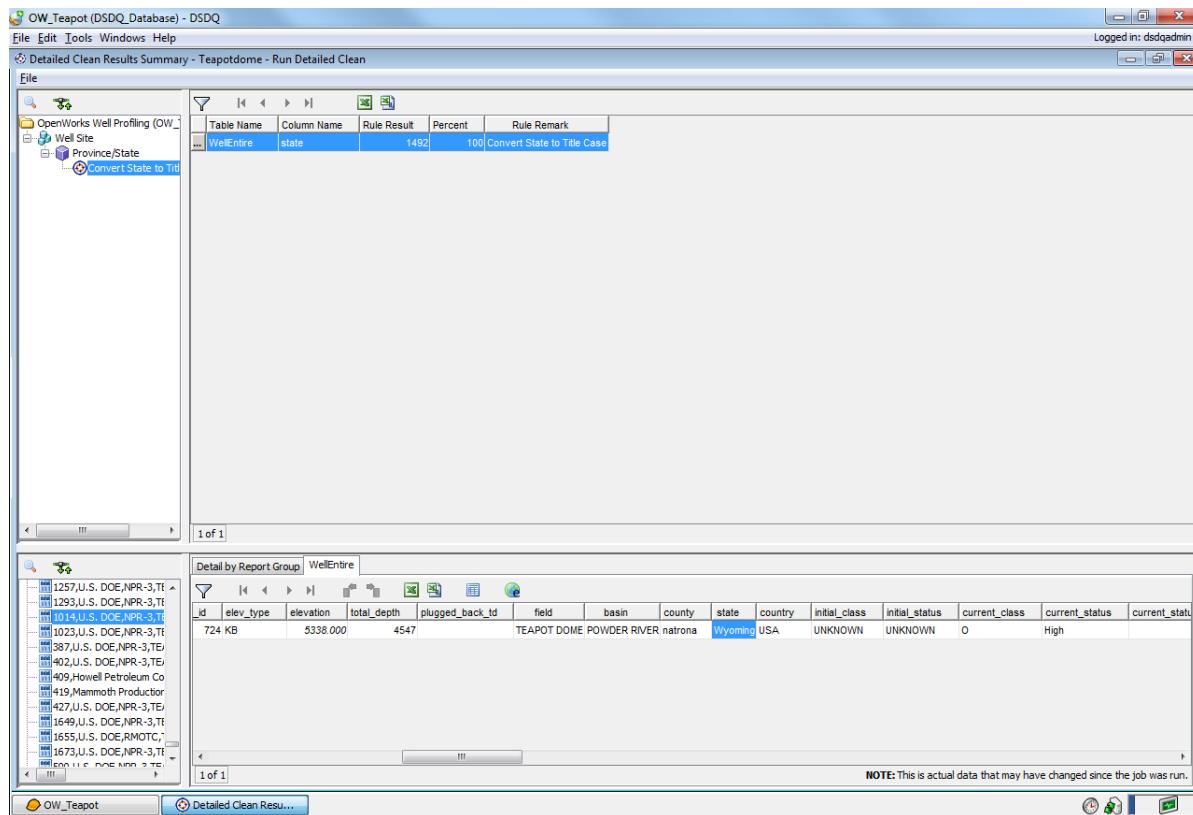
The **Run Detailed Clean** Task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

14. Select the **Results** tab.

The **Jobs and Results Listing Pane** displays a list of results.



15. Click the **Open Basic View Frame**  button on the **Result Reports** tab to display the **Run Detailed Clean** Task results in the **Basic View Frame** window.



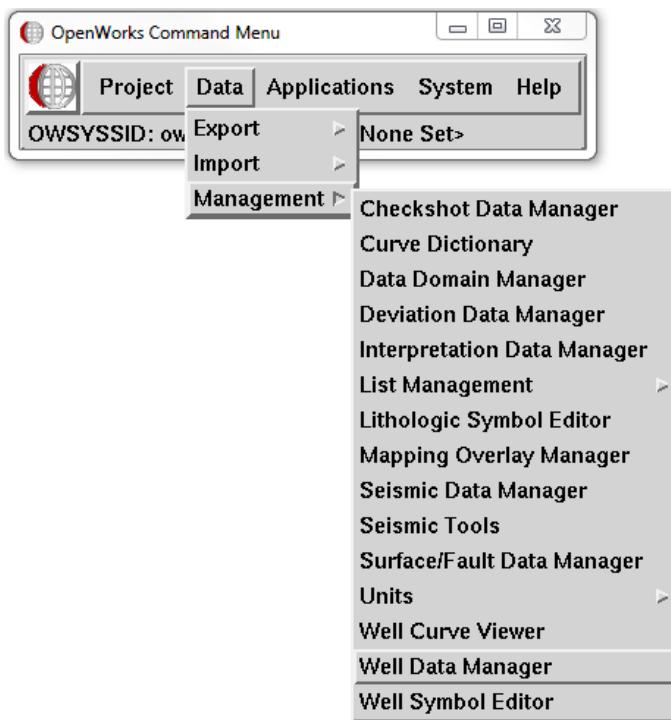
16. Select **File > Exit** to close the **Basic View Frame** window.

## Validating Corrected Data in the OpenWorks Data Source

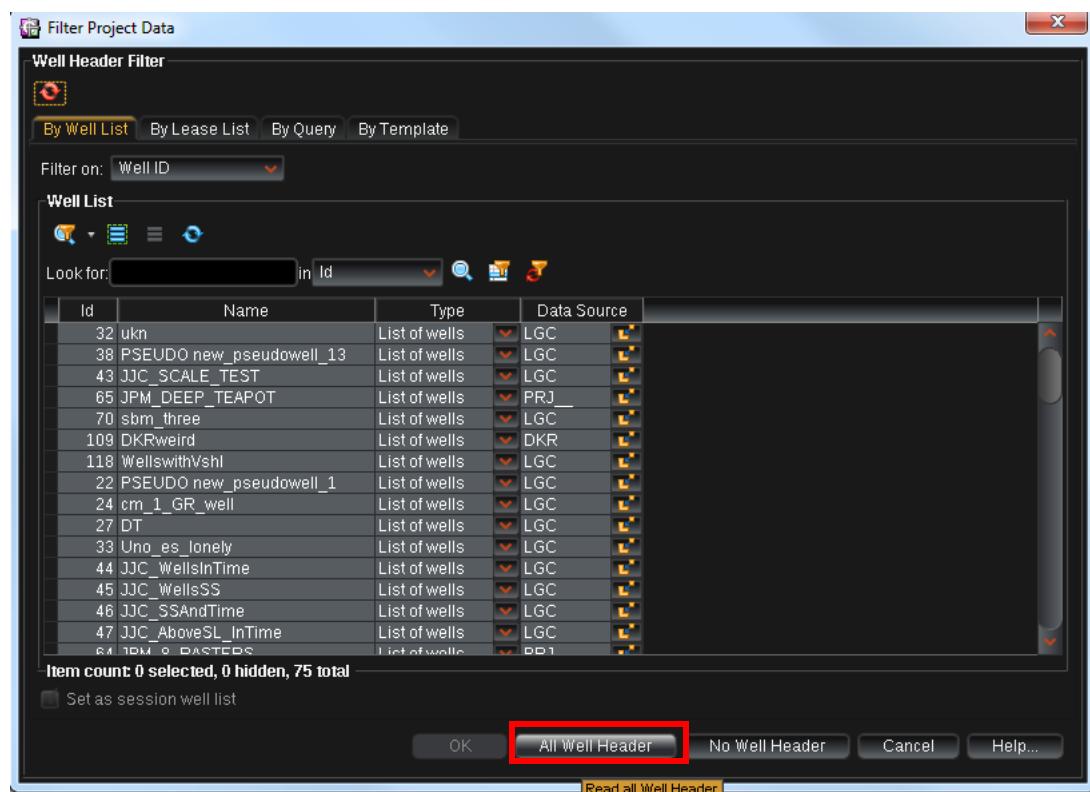
---

To validate information from the Well Data Manager of the OpenWorks data source:

1. Double-click the OpenWorks desktop icon to launch the OpenWorks application.
2. Select **Data > Management > Well Data Manager** from the command menu.

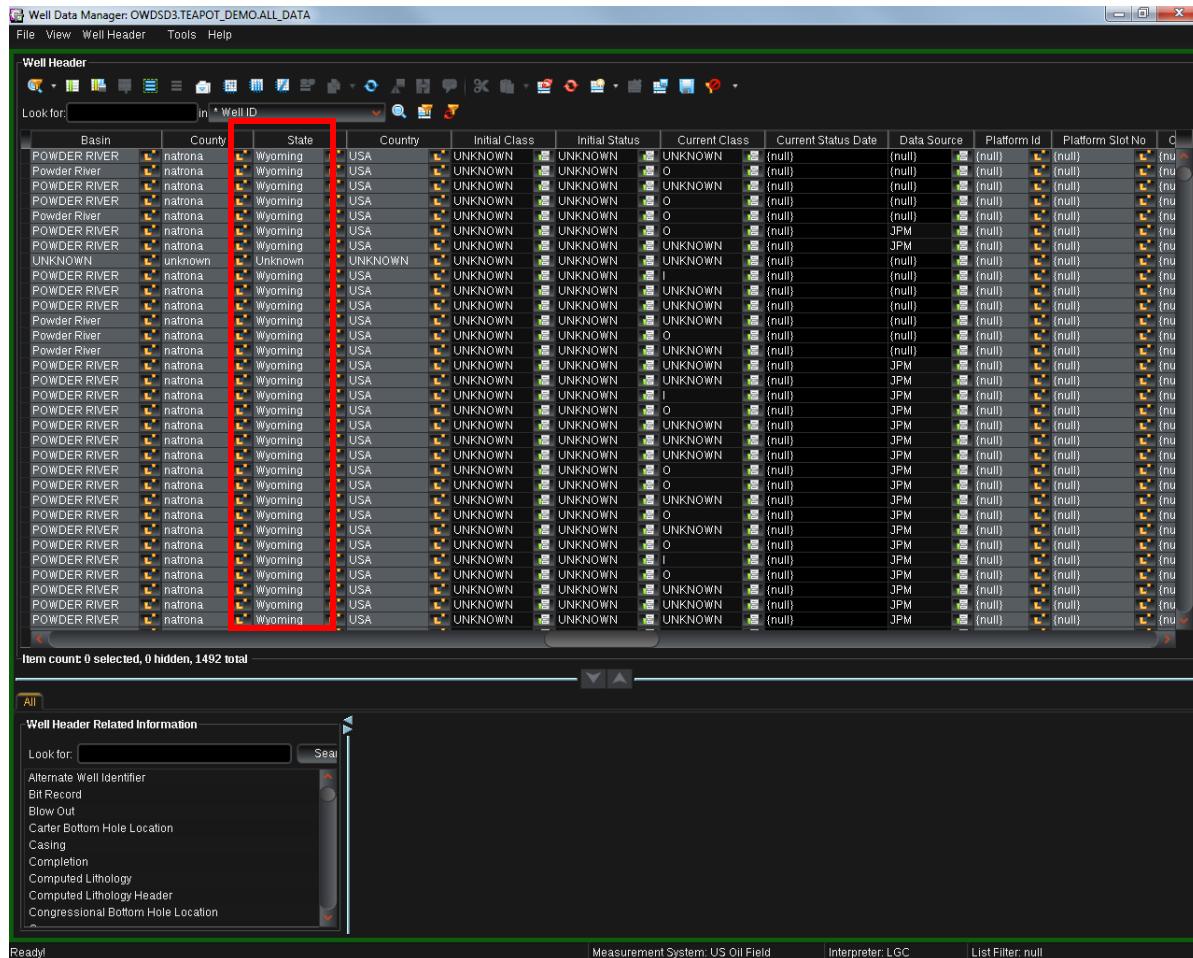


The **Filter Project Data** window appears.



3. Click the **All Well Header** button.

The **Well Data Manager** window appears displaying the **State** column data in Title Case.

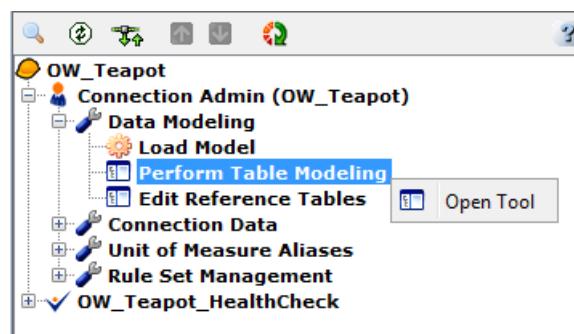


4. Select **File > Exit** to close the **Well Data Manager** window.

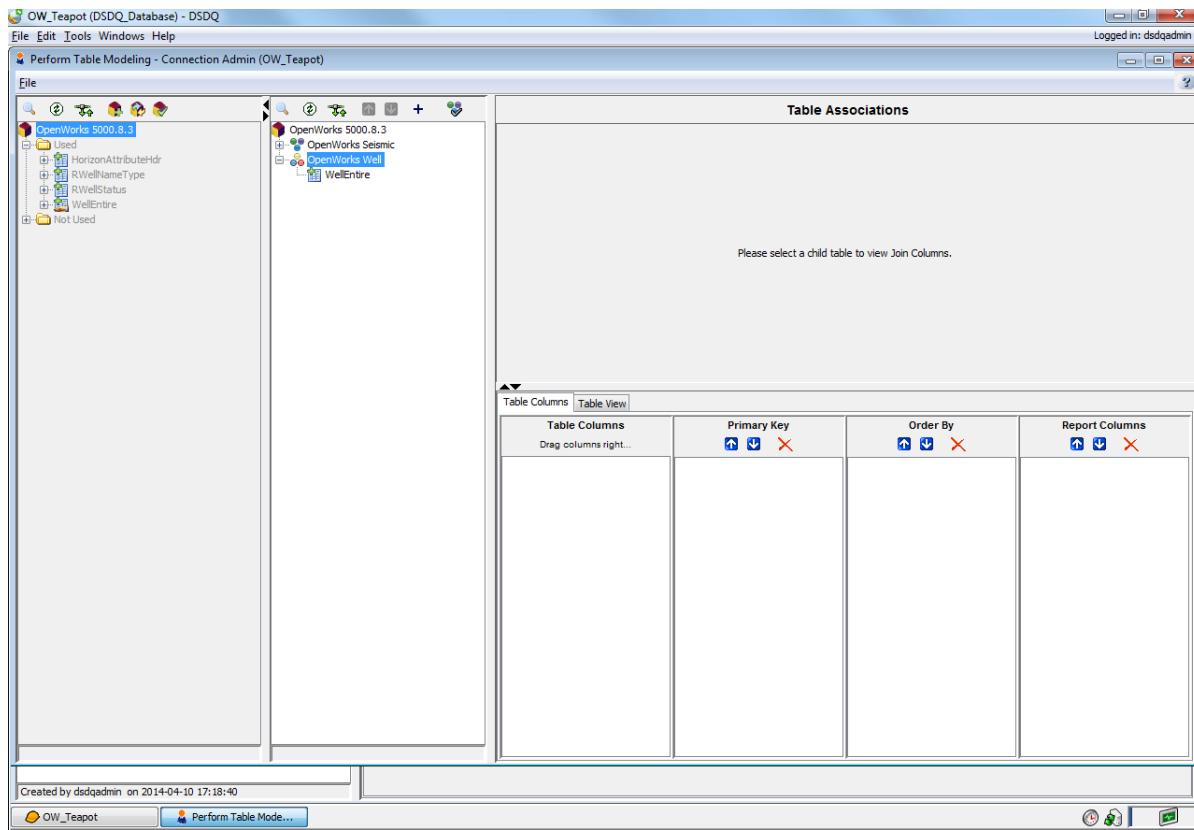
## Viewing Results on the DSDQ Web Dashboard

To view information about a DecisionSpace Data Quality submodel, you publish it to the Web Dashboard. To publish a submodel to the Web Dashboard:

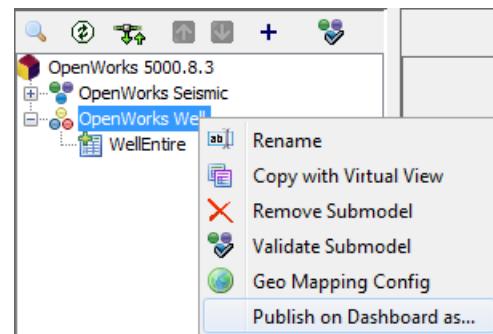
1. Click  on the DecisionSpace Data Quality Tree to expand **Connection Admin (OW\_Teapot)**.
2. Click  on the **Data Modeling** Activity.
3. Double-click the **Perform Table Modeling** Tool on the DecisionSpace Data Quality Tree or right-click the **Perform Table Modeling** Tool and select **Open Tool** from the pop-up menu.



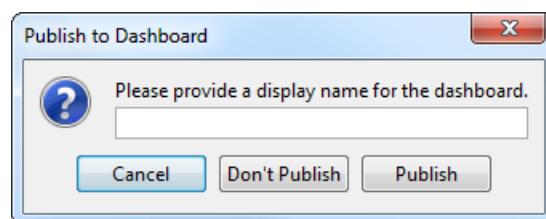
The **Perform Table Modeling - Connection Admin** window appears.



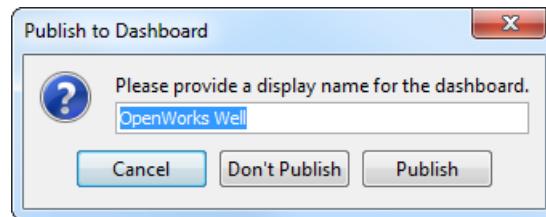
- Right-click the **OpenWorks Well** submodel from the Submodel Listing Tree and select the **Publish on Dashboard as...** option from the pop-up menu.



The **Publish to Dashboard** dialog box appears.

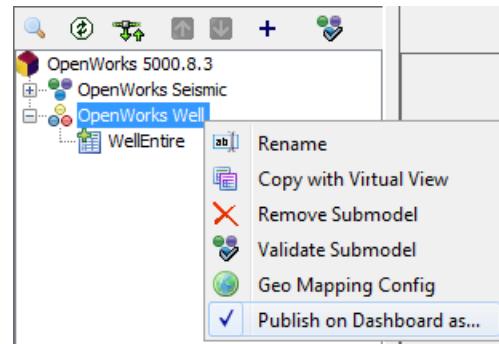


5. Enter **OpenWorks Well** in the **Please Provide a display name for the dashboard** field.



The **OpenWorks Well** submodel is published to the Web Dashboard.

6. To confirm that the submodel has been published to the Web Dashboard, right-click the **OpenWorks Well** submodel on the Submodel Listing Tree. A checkmark appears on the right side of the **Publish on dashboard as...** option.



7. Select **File > Exit** from the menu bar to close the **Perform Table Modeling** window.
8. Enter **http://localhost:8091** in the address bar of the web browser. The **Please wait. Your browser will be redirected when ready.**



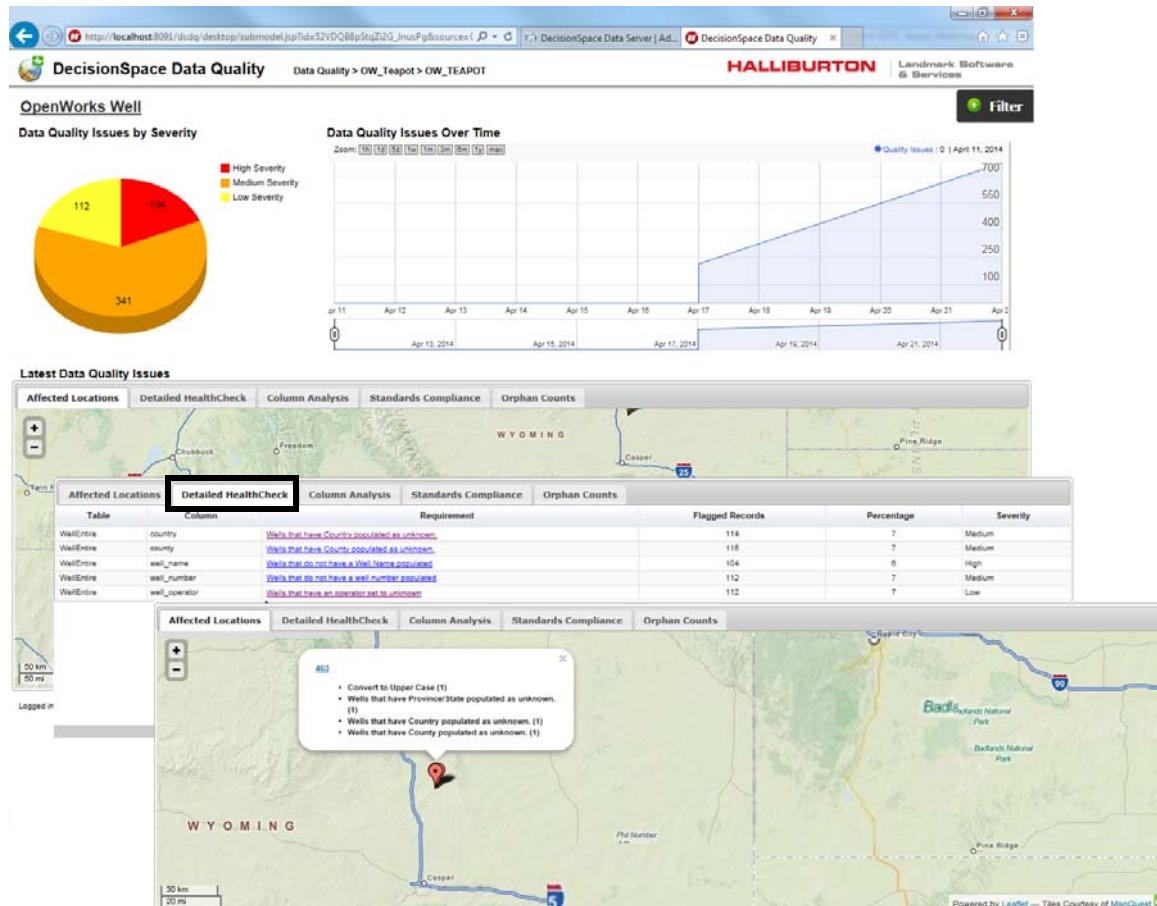
message displays in the web browser.

The Web Dashboard is launched in the web browser.



9. Select **OW\_Teapot** from the Data Quality Web Dashboard to display the published submodel.

10. Select the **Detailed HealthCheck** tab and then a desired **Requirement** from the **Latest Data Quality Issues** area of the Web Dashboard to view the Column Display Group and the dashboard name for the column.



# ***Connecting DSDQ to an EDM Data Source***

This section of the DecisionSpace Data Quality training manual aims at walking you through the process of:

- Accessing data stored in an EDM data source
- Profiling data to identify the full spectrum of data quality issues
- Addressing all such data issues by means of quality control queries and a repeatable cleansing methodology
- Validating corrections made to the data in the EDM data source, and
- Finally, viewing the data quality results in the DecisionSpace Data Quality Web Dashboard

---

## Overview

---

In this workshop, you will perform the following exercises to evaluate, clean, and standardize data from an EDM data source in DecisionSpace Data Quality:

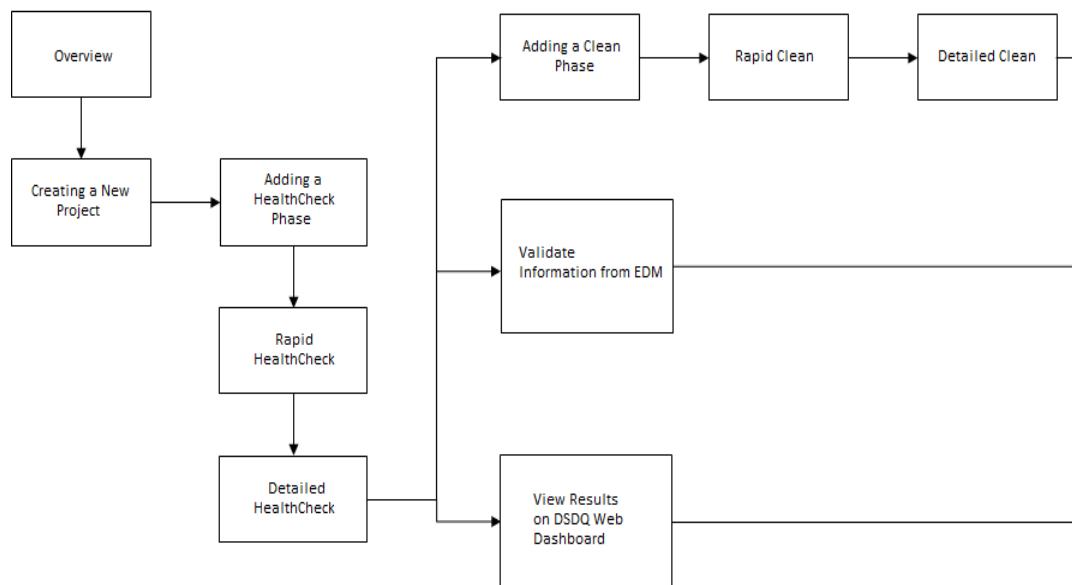
- Creating a New Project
- Evaluating data using the HealthCheck Phase
  - a) Rapid HealthCheck
    - Running Table Analysis on All Tables
    - Running Column Analysis on Columns
    - Running Table Analysis on Modeled Tables
    - Running Column Analysis on Modeled Tables
  - b) Detailed HealthCheck
    - Configuring the Detailed HealthCheck Tool
    - Running the Detailed HealthCheck Task
- Resolving data quality issues using the Clean Phase
  - a) Adding a Clean Phase
  - b) Rapid Clean
    - Configuring the Rapid Clean Tool
    - Running the Test Rapid Clean Task
    - Running the Rapid Clean Task
  - c) Detailed Clean
    - Configuring the Detailed Clean Tool

### — Running the Detailed Clean Task

- Validating corrected data in the EDM data source
- Viewing the data quality results in the DecisionSpace Data Quality Web Dashboard

The purpose of this workshop is to reinforce what you have learnt in previous sections of this manual in a single workflow. Topics covered in each section of the workflow are outlined in the following illustration:

### Workflow for Connecting DSDQ to an EDM Data Source



## Creating a New Project

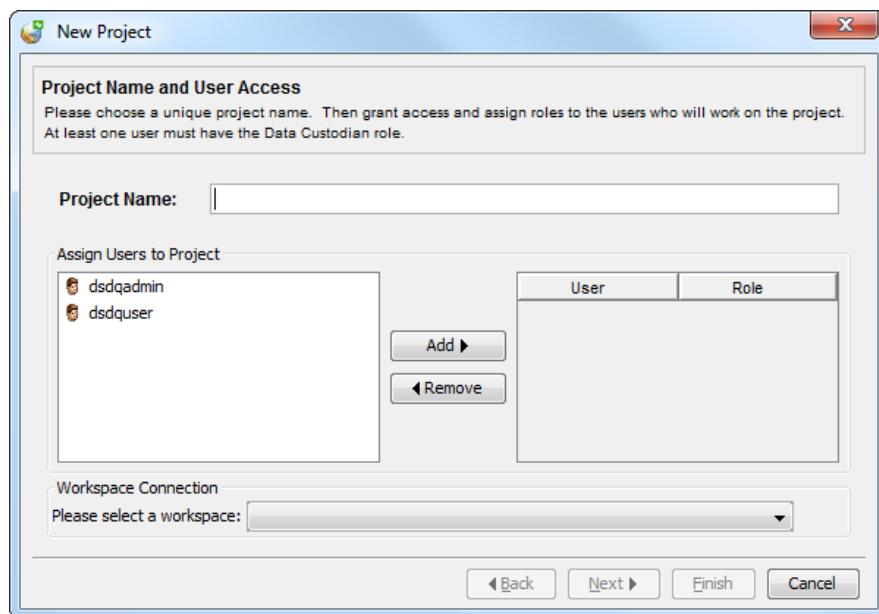
---

A DSDQ project comprises of all the Phases, Activities, Tools etc. During this process, you will assign users to the project and their roles while they work on it; select a Workspace Connection (the database where results will be written), the desired Phase and a Source Connection (data source that the application reads from).

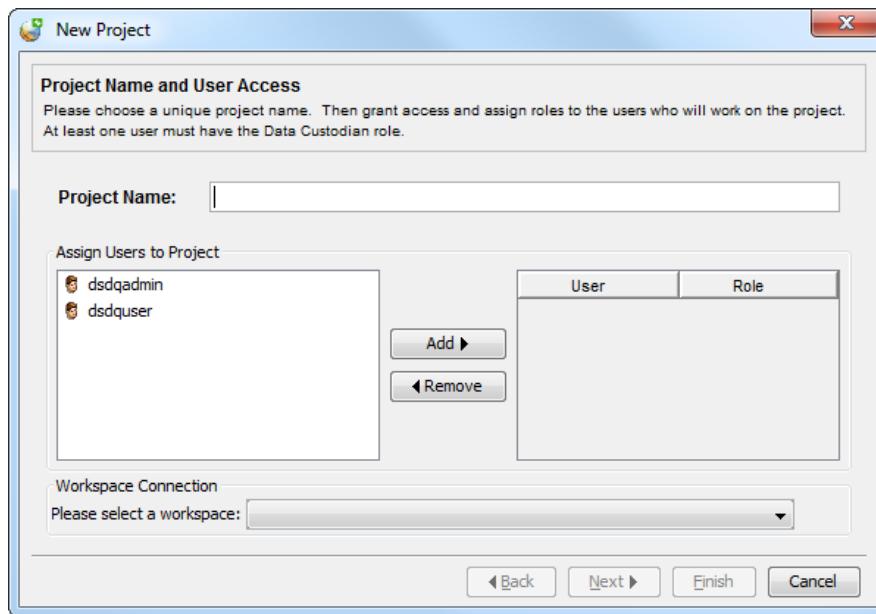
A new Project can be created:

- When the software is initially installed and a project does not exist.
- By selecting **New Project** from the File menu.
- By clicking the **New** button in the **Open an Existing Project** window.

In all instances, the **New Project** window appears displaying all available users.



1. Select **File > New Project** from the menu bar on the **DSDQ Project** window.  
The **New Project** window appears.



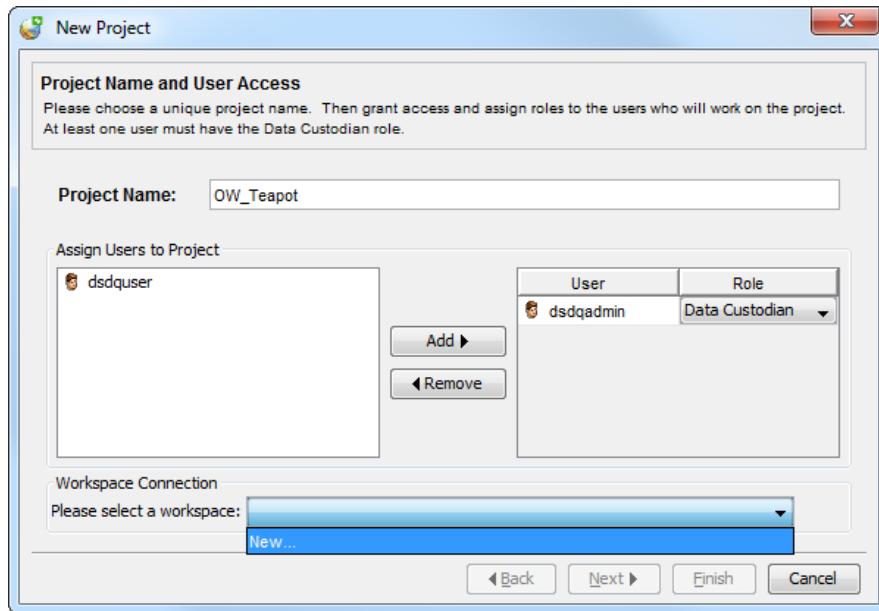
2. Enter **EDM\_Teapot** in the **Project Name** field.

**Note**

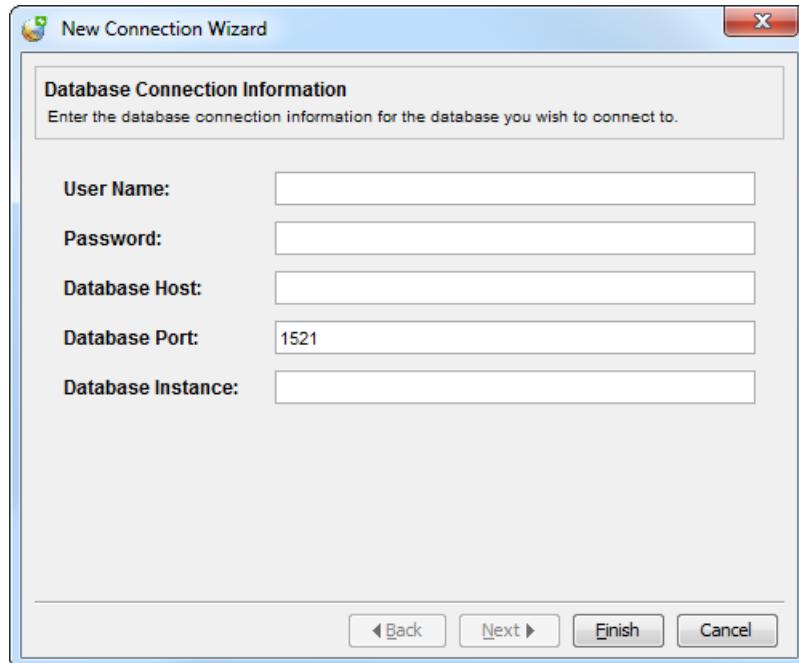
You will be using **EDM\_Teapot** throughout this workflow.

3. Select **dsdqadmin** from the **Assign Users to Project** group box.

4. Click the **Add**  button to assign project access to the selected user.



5. Select **Data Custodian** from the **Role** drop-down list.
6. Select **New...** from the **Please select a workspace** drop-down list.  
**The New Connection Wizard - Database Connection Information** window appears.

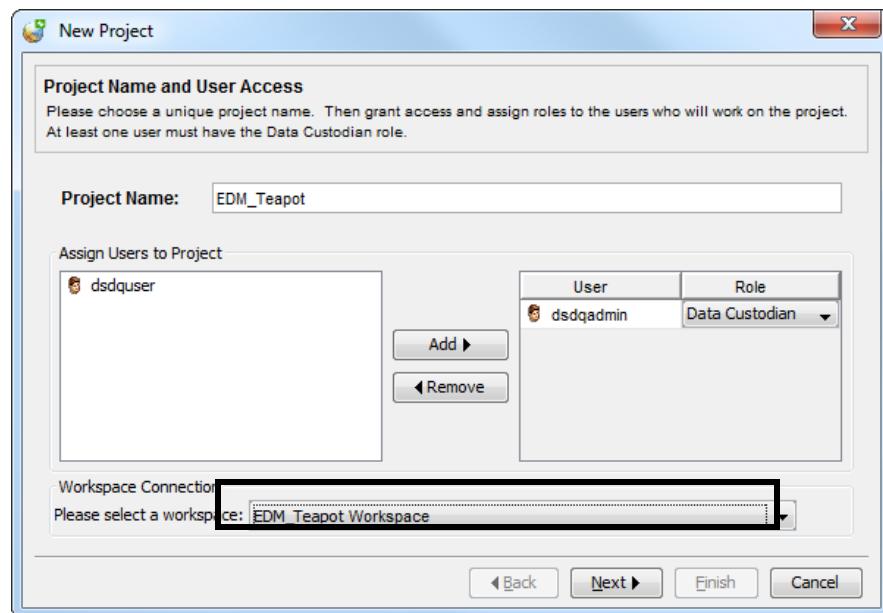


7. Enter **DSDQ\_EDM\_Teapot\_WKSP** in the **User Name** field.

**Note**

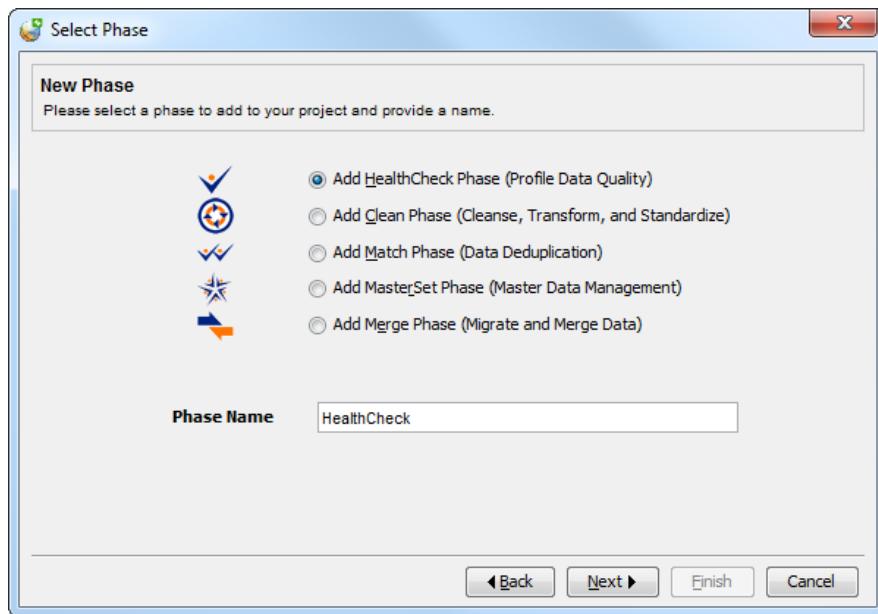
Make sure that the user and database connection have already been setup in Oracle (*reference: chapter 1, DSDQ Training Manual*).

8. Enter **DSDQ** in the **Password** field.
9. Enter **localhost** in the **Database Host** field.  
The **Database Port** is set to **1521** by default. If DecisionSpace Data Server connects to a different port, this number will need to be updated.
10. Enter **oradsdq10** in the **Database Instance** field.
11. Click **Finish**.  
You will notice that the newly created workspace **EDM\_Teapot\_Workspace** is populated in the **Please select a workspace** drop-down list on the **New Project** window.

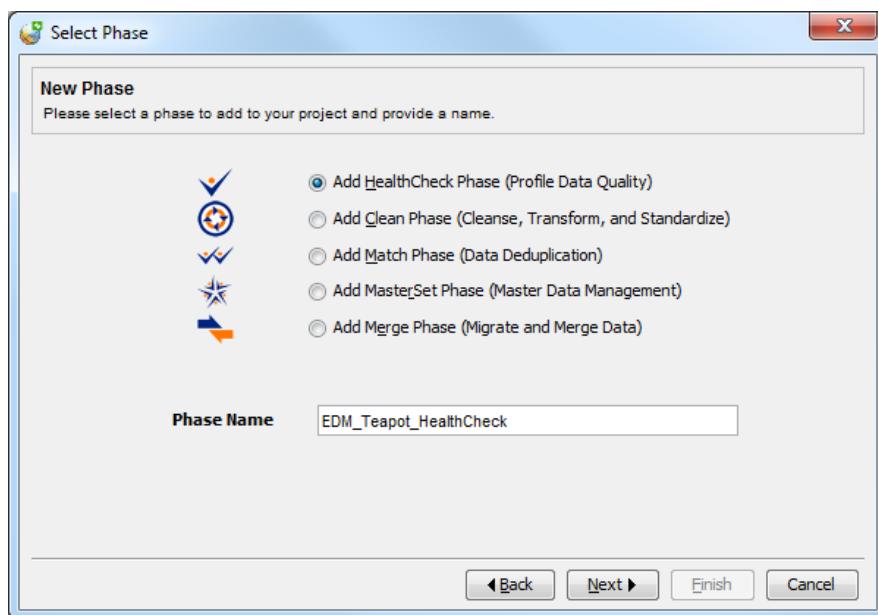


12. Click **Next** to continue.

The **Select Phase** window appears with the **Add HealthCheck Phase (Profile Data Quality)** option selected by default.

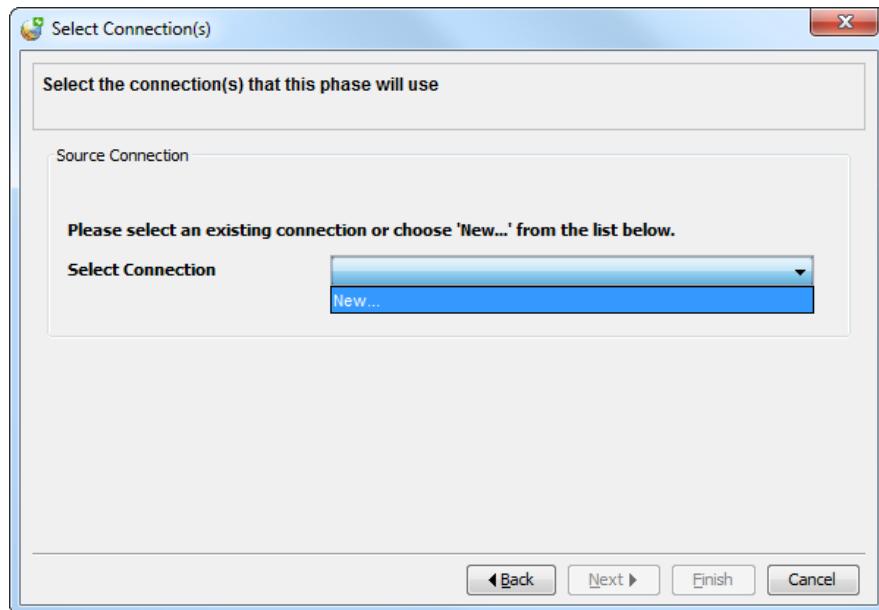


13. Enter **EDM\_Teapot\_HealthCheck** in the **Phase Name** field.

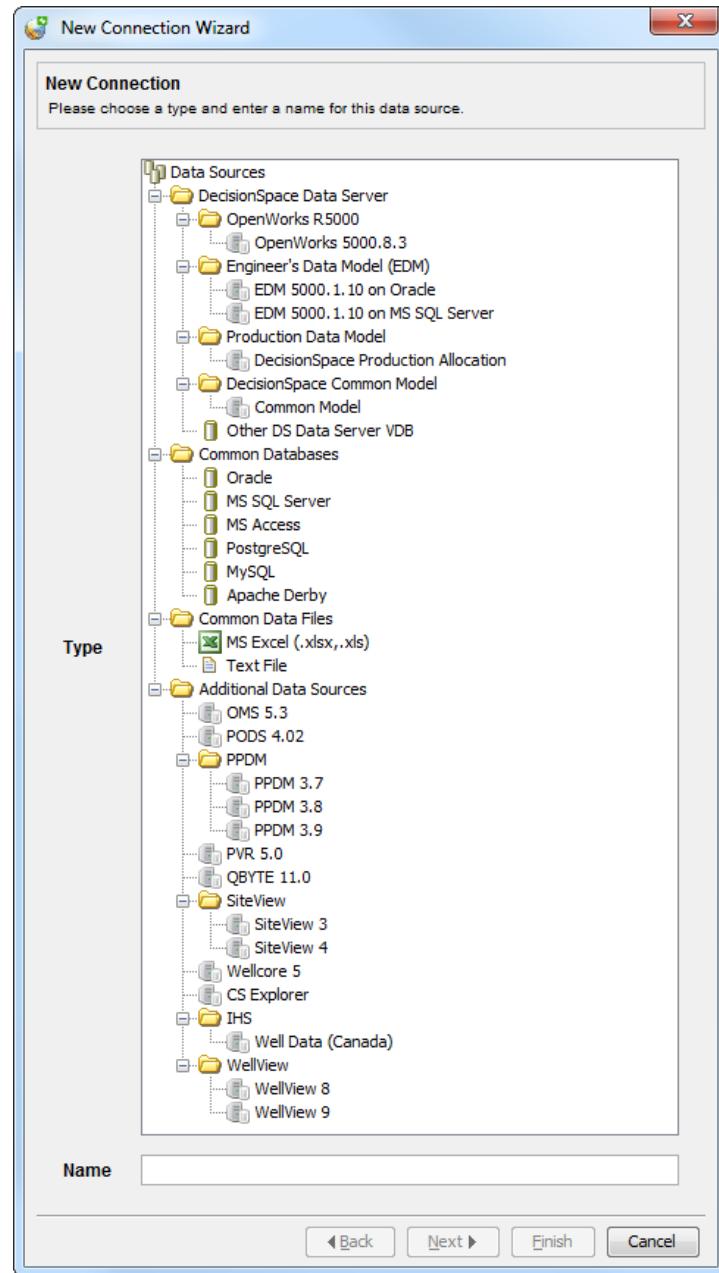


14. Click **Next** to continue.

The **Select Connection(s)** window appears.

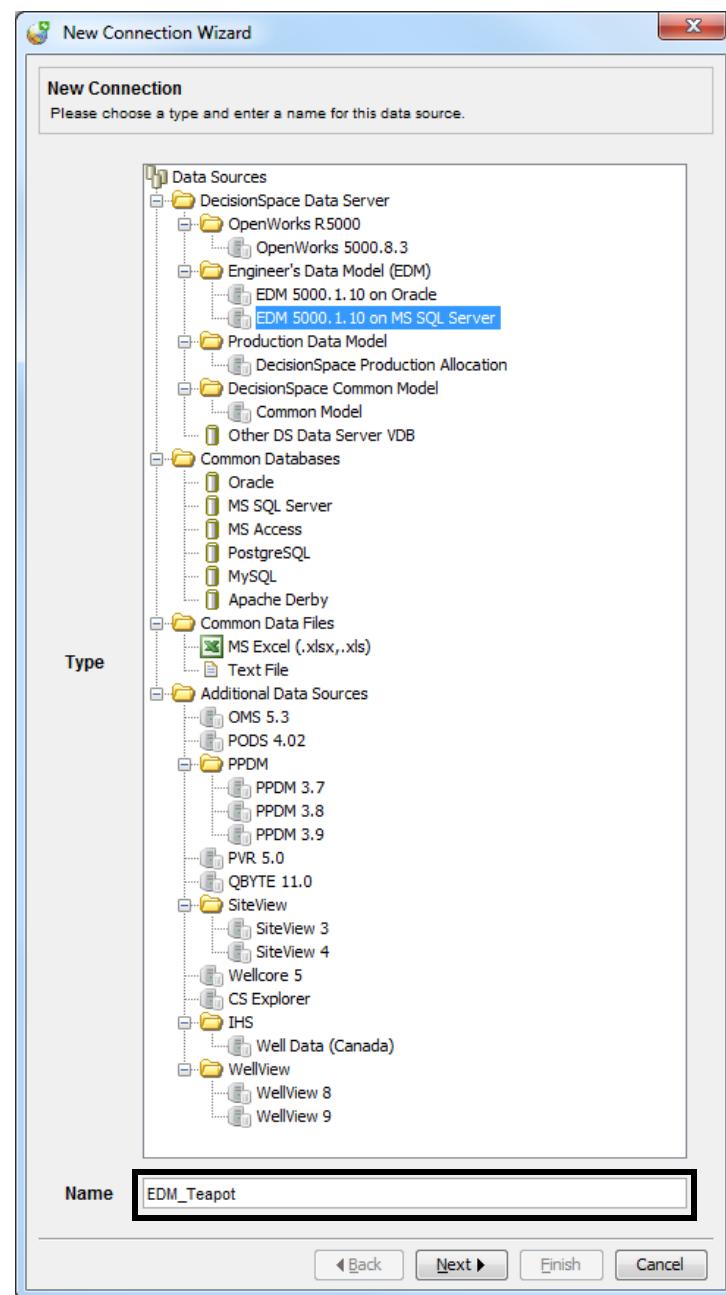


15. Select **New...** from the **Select Connection** drop-down list.  
The **New Connection Wizard - New Connection** window appears.



16. Select **EDM 5000.1.10 on MS SQL Server** from the Connection Type tree.

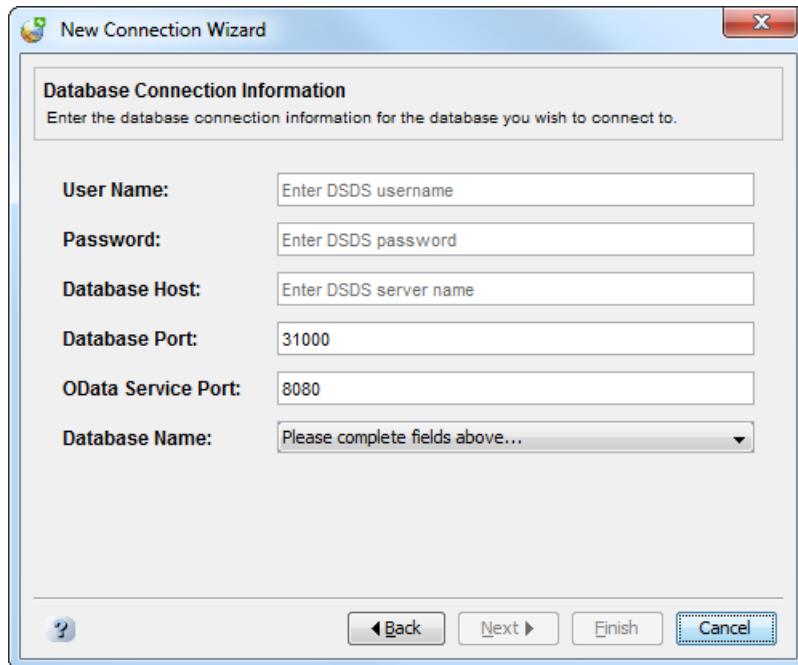
17. Enter **EDM\_Teapot** in the **Name** field.



18. Click **Next** to continue.

**The New Connection Wizard - Database Connection**

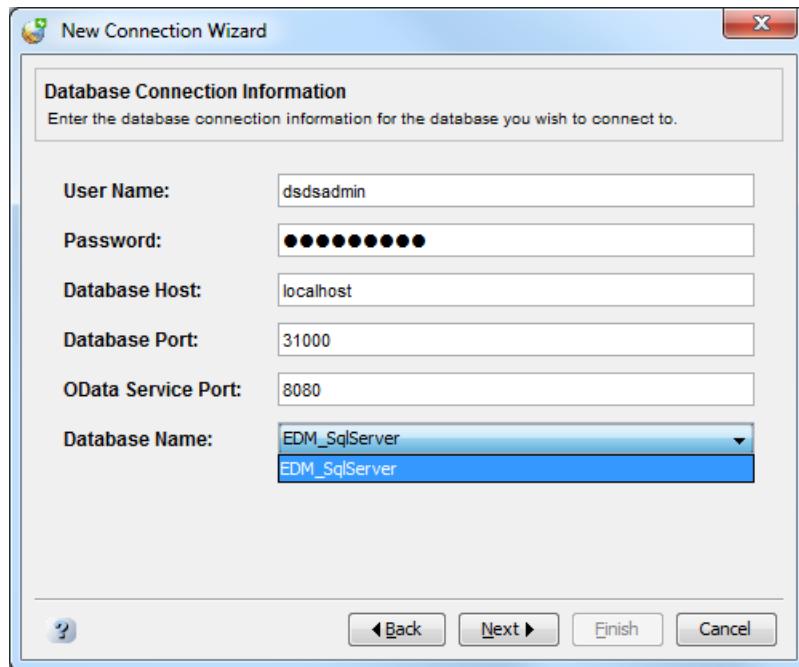
**Information** window appears displaying database connection information for DecisionSpace Data Server.



19. Enter **dsdsadmin** in the **User Name** field.
20. Enter **dsdsadmin** in the **Password** field.
21. Enter **localhost** in the **Database Host** field.
22. Accept the default **Database Port** value of **31000**.

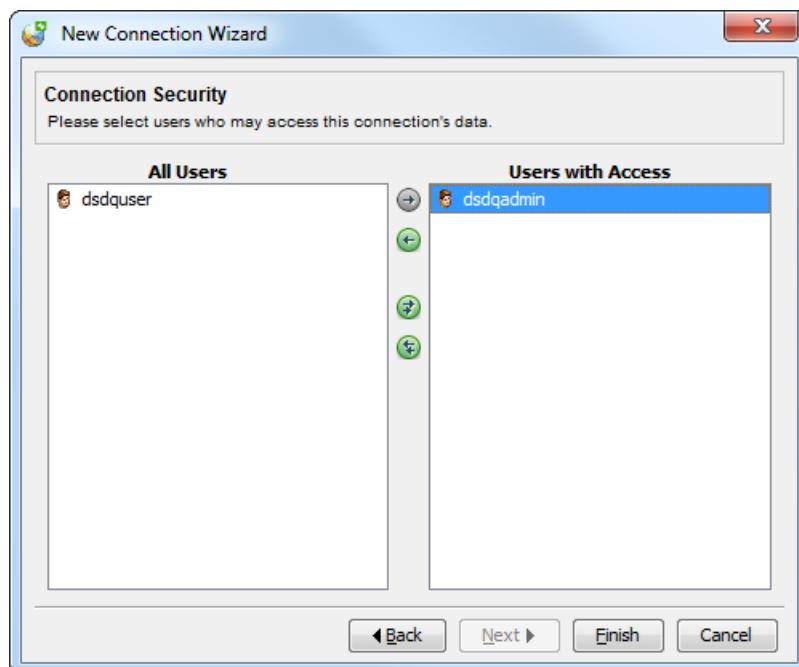
23. Accept the default **OData Service Port** value of **8080**.

The **Database Name** drop-down list auto-populates with the EDM data source option.



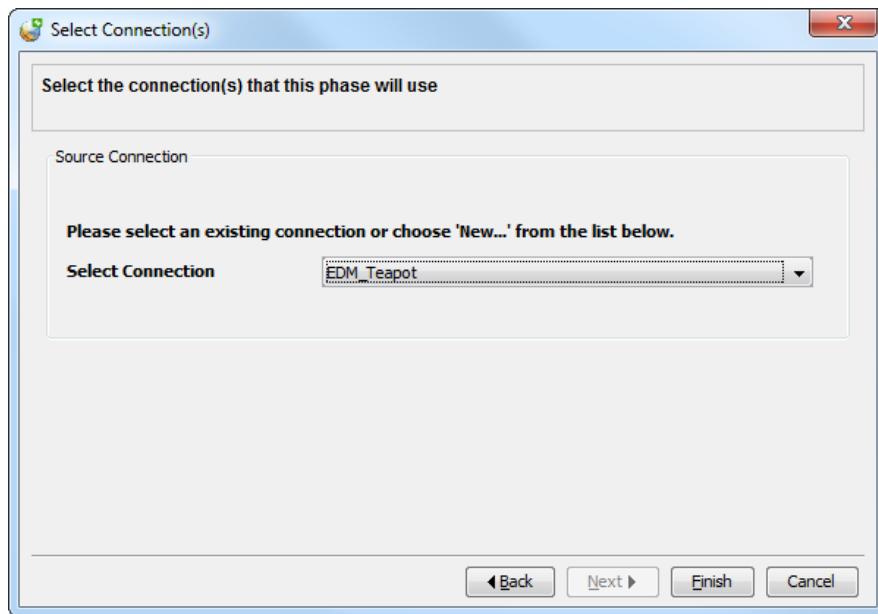
24. Click **Next** to continue.

The **New Connection Wizard - Connection Security** window appears.



25. Click **Finish**.

The **Select Connection(s)** window appears.



26. Click **Finish**.

The **HealthCheck** Phase is added and displayed in the DecisionSpace Data Quality tree

A screenshot of the DecisionSpace Data Quality application. The title bar says "EDM\_Teapot (DSDQ\_Database) - DSDQ" and "Logged in: dsdqadmin".  
**DecisionSpace Data Quality Tree:** On the left, there's a tree view under "EDM\_Teapot" with nodes: "Connection Admin (EDM\_Teapot)" and "EDM\_Teapot\_HealthCheck".  
**Jobs and Results Listing Pane:** This pane has tabs for "Jobs" and "Results". It shows a table header with columns: Phase Name, Activity Name, Task Name, Job Name, Run By, Warnings, Processed Data Sets, Start Date, Duration, Status, Remark. Below the header, it says "0 of 0".  
**Jobs and Results Information Pane:** This large pane at the bottom displays the message: "Please select a result above to display reports." It also has tabs for "Parameters", "Scheduling", "Result Reports", and "Result Logs".  
At the bottom of the application window, there's a toolbar with icons for Save, Undo, Redo, Cut, Copy, Paste, and others.

## Evaluating Data using the HealthCheck Phase

The **HealthCheck** Phase assists in evaluating the "where", "what" and "why" issues in your valuable data assets.

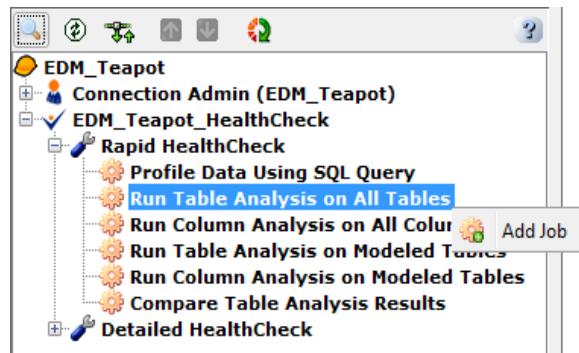
### Rapid HealthCheck Activity

The **Rapid HealthCheck** Activity provides a quick look at the volume and quality of the data.

#### Exercise: Running Table Analysis on All Tables Task

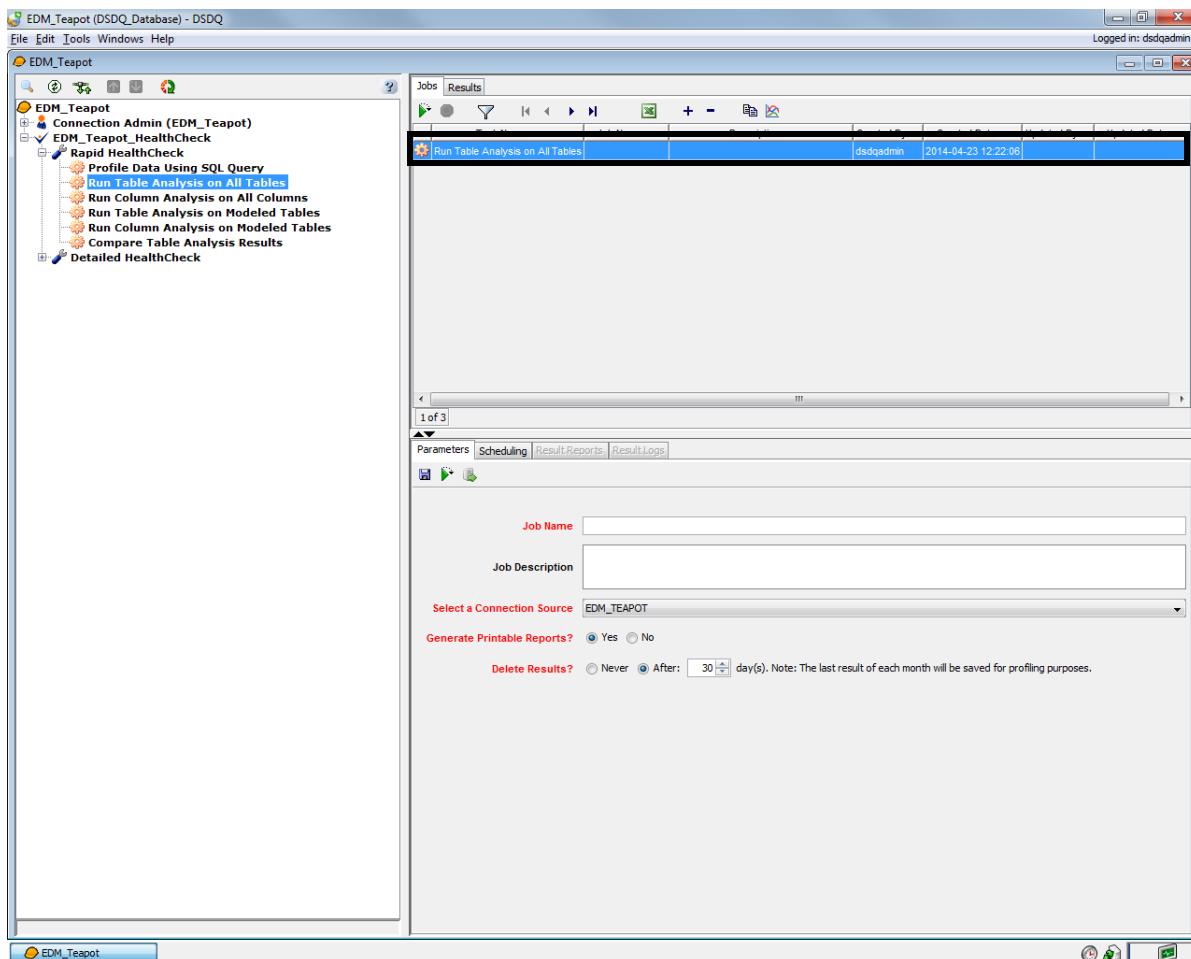
The **Run Table Analysis on All Tables** Task is used to analyze all the tables for issues and inconsistencies. In this particular exercise, we will analyze the tables and count the number of rows in them. Rows are counted when values are entered in them. To run Table Analysis on all tables:

1. Click  to expand the **Rapid HealthCheck** Activity on the DecisionSpace Data Quality Tree.
2. Double-click the **Run Table Analysis on All Tables** Task or right-click the **Run Table Analysis on All Tables** Task and select **Add Job** from the pop-up menu.



A new job is initiated and displays in the **Job and Results**

**Information Pane** on the right side of the DecisionSpace Data Quality Project window.

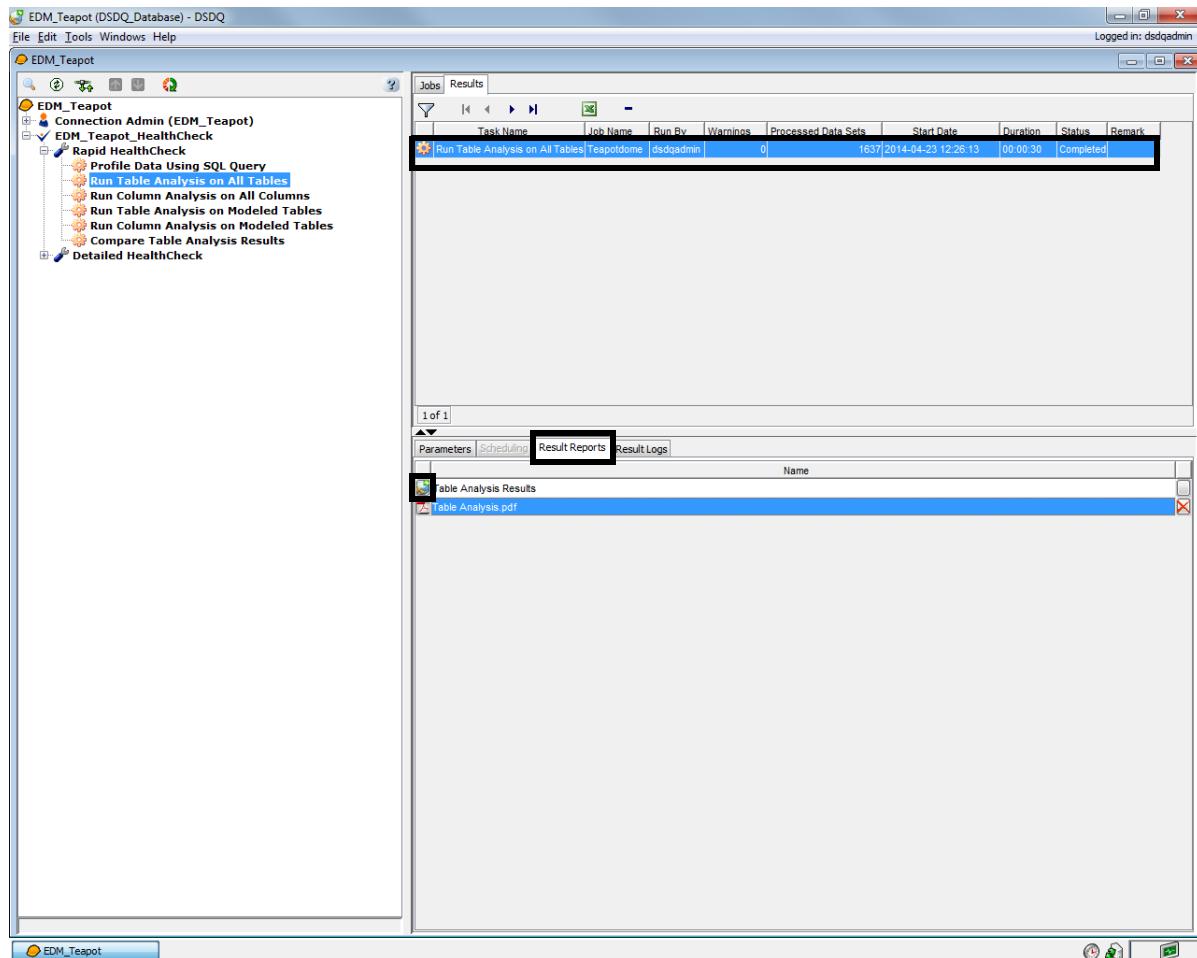


3. Enter **Teapotdome** in the **Job Name** field.
  4. Enter **Table Analysis on All Tables of EDM\_Teapot** in the **Job Description** field.
  5. Select **EDM\_Teapot** from the **Select a Connection Source** drop-down list.
  6. Select the **Yes** option for **Generate Printable Reports**.
  7. Select the **After** option for **Delete Results**. Leave the number of days as **30**.
  8. Click to save changes in the **Parameters** tab.
  9. Click to run the job.
- The **Run Table Analysis on All Tables** Task runs and displays

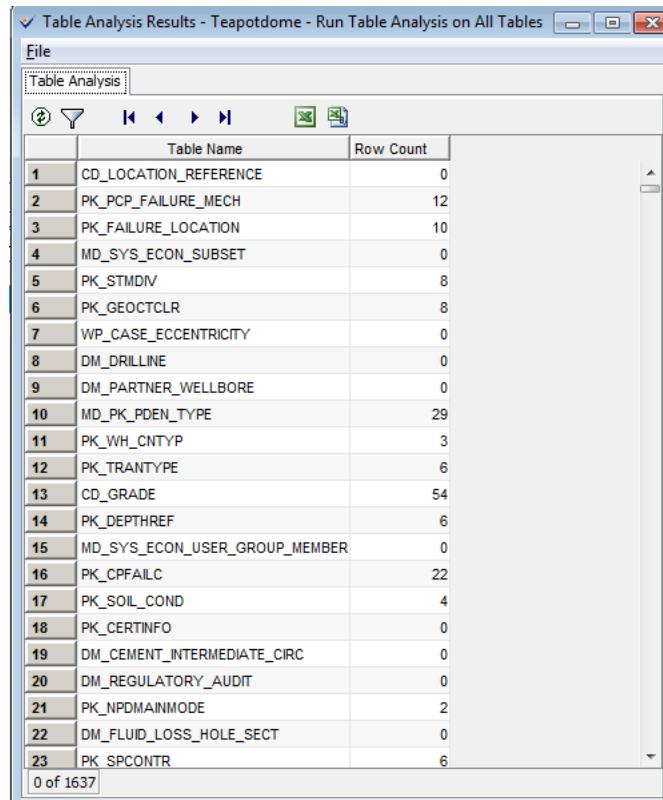
results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

10. Select the **Results** tab.

The **Jobs and Results Listing Pane** displays a list of results.



11. Click the **Open Basic View Frame**  button on the **Result Reports** tab to display the **Run Table Analysis on All Tables** Task results in the **Basic View Frame** window.



The screenshot shows a software interface titled "Table Analysis Results - Teapotdome - Run Table Analysis on All Tables". The main area is a table titled "Table Analysis" with columns "Table Name" and "Row Count". The table lists 23 tables, each numbered from 1 to 23. The data is as follows:

	Table Name	Row Count
1	CD_LOCATION_REFERENCE	0
2	PK_PCP_FAILURE_MECH	12
3	PK_FAILURE_LOCATION	10
4	MD_SYS_ECON_SUBSET	0
5	PK_STMDIV	8
6	PK_GEOCTCLR	8
7	WP_CASE_ECCENTRICITY	0
8	DM_DRILLINE	0
9	DM_PARTNER_WELLBORE	0
10	MD_PK_PDEN_TYPE	29
11	PK_WH_CNTYP	3
12	PK_TRANTYPE	6
13	CD_GRADE	54
14	PK_DEPTHREF	6
15	MD_SYS_ECON_USER_GROUP_MEMBER	0
16	PK_CPFALC	22
17	PK_SOIL_COND	4
18	PK_CERTINFO	0
19	DM_CEMENT_INTERMEDIATE_CIRC	0
20	DM_REGULATORY_AUDIT	0
21	PK_NPDMAINMODE	2
22	DM_FLUID_LOSS_HOLE_SECT	0
23	PK_SPCONTR	6

At the bottom left of the table, it says "0 of 1637".

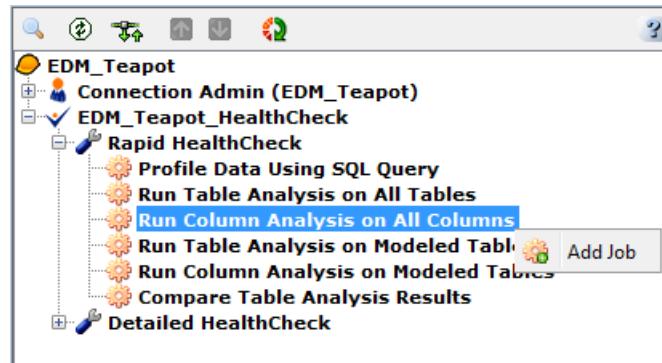
12. Select **File > Exit** to close the **Basic View Frame** window.

### Exercise: Running Column Analysis on All Columns Task

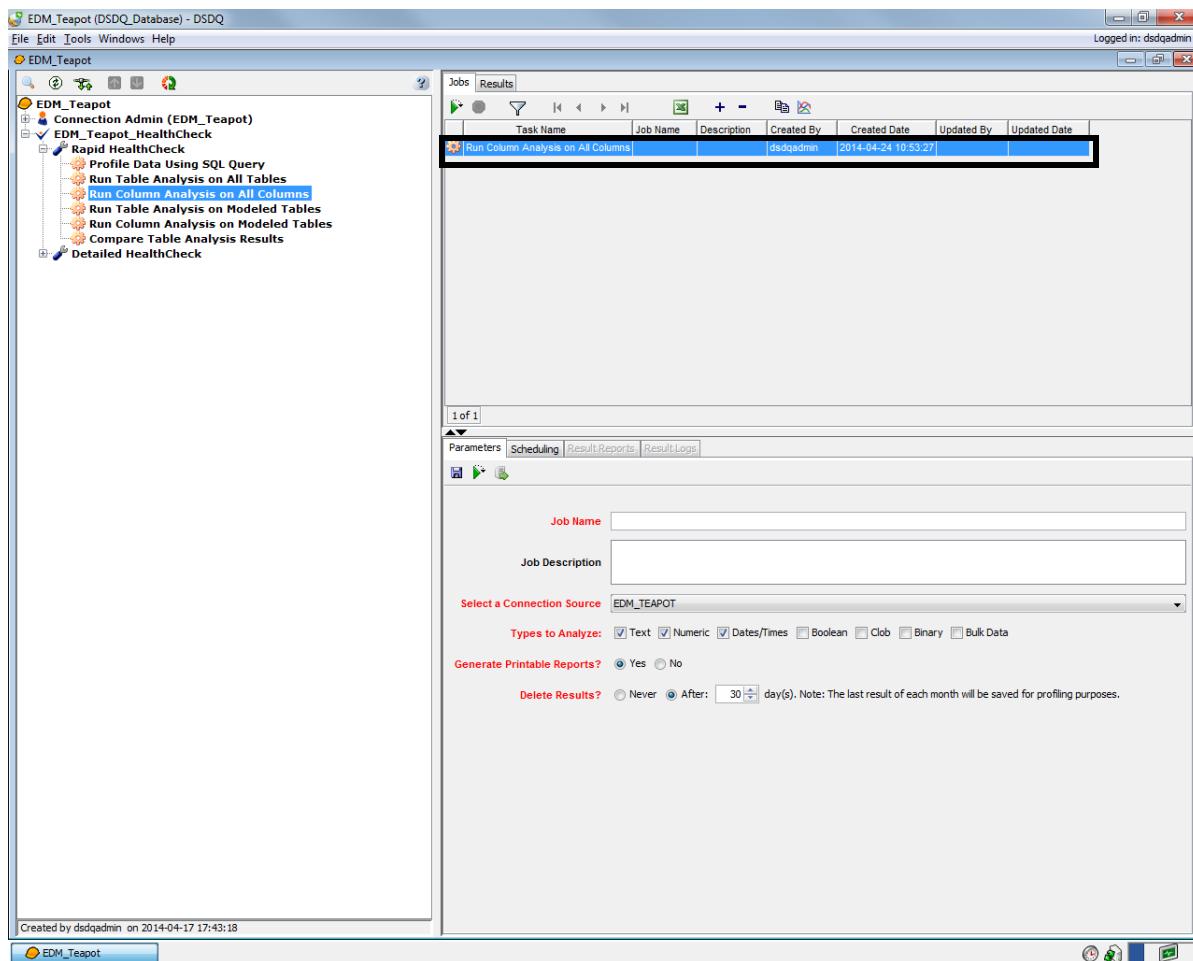
The **Run Column Analysis on All Columns** Task offers basic data profiling by checking the number of rows, number of not null values, unique values, percentage of row populated, minimum values, maximum values, number of values with mixed cases, number of non-printable characters, number of preceding, trailing and double white spaces parameters within a column. To run the Column Analysis on all columns:

1. Double-click the **Run Column Analysis on All Columns** Task or right-click the **Run Column Analysis on All Columns** Task and

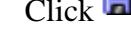
select **Add Job** from the pop-up menu.

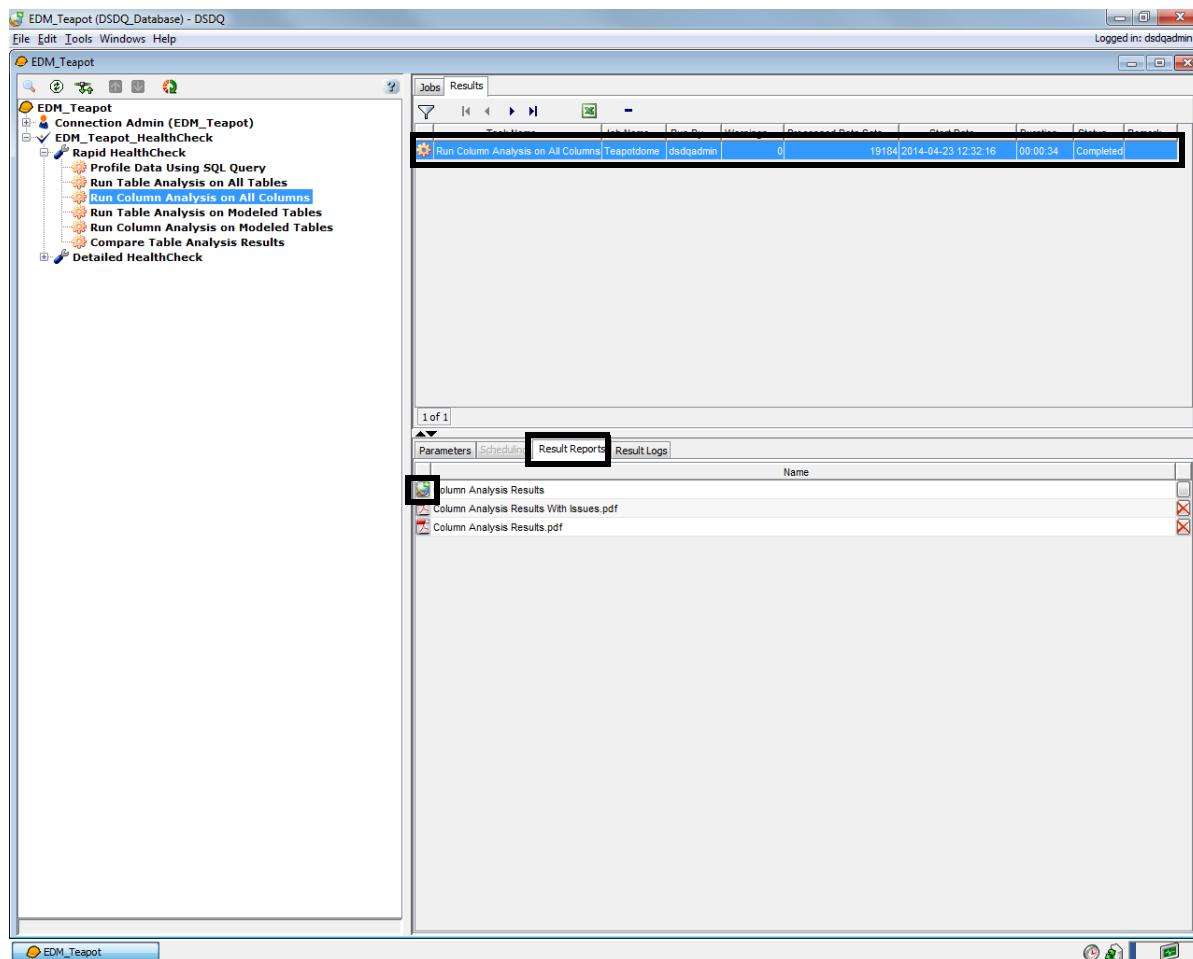


A new job is initiated and displays on the **Job and Results Information Pane** on the right side of the DecisionSpace Data Quality Project window.



2. Enter **Teapotdome** in the **Job Name** field.
3. Enter **Column Analysis on All Columns of EDM\_Teapot** in the **Job Description** field.

4. Select **EDM\_Teapot** from the **Select a Connection Source** drop-down list.
5. Select all the options for **Types to Analyze**.
6. Select the **Yes** option for **Generate Printable Reports**.
7. Select the **After** option for **Delete Results**. Leave the number of days as **30**.
8. Click  to save changes in the **Parameters** tab.
9. Click  to run the job.  
The **Run Column Analysis on All Columns** Task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.
10. Select the **Results** tab.  
The **Jobs and Results Listing Pane** displays a list of results.



11. Click the **Open Basic View Frame**  button on the **Result Reports** tab to display the **Run Column Analysis on All Columns** Task results in the **Basic View Frame** window.

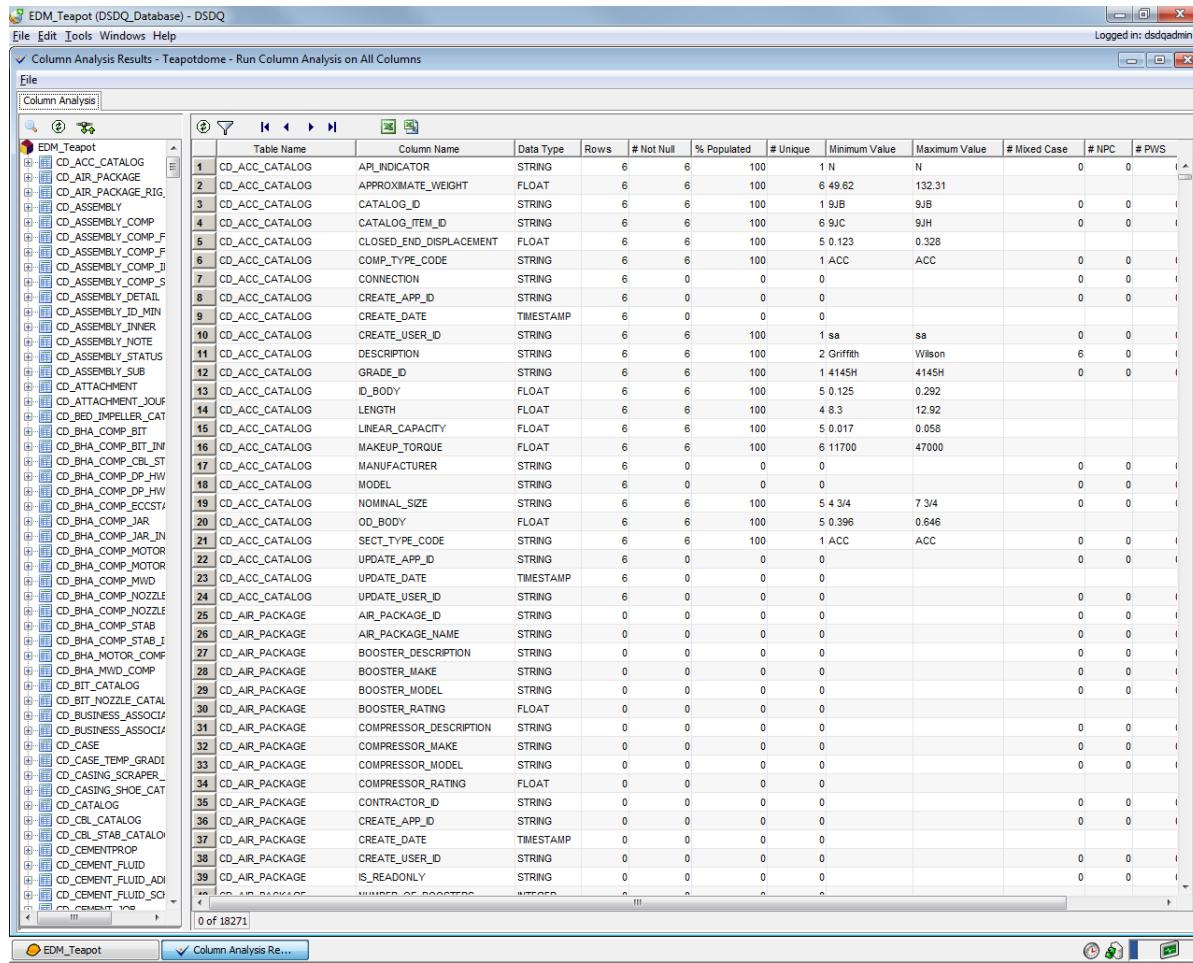


Table Name	Column Name	Data Type	Rows	# Not Null	% Populated	# Unique	Minimum Value	Maximum Value	# Mixed Case	# NPC	# PWS
CD_ACC_CATALOG	APLINDICATOR	STRING	6	6	100	1 N	N		0	0	
CD_ACC_CATALOG	APPROXIMATE_WEIGHT	FLOAT	6	6	100	6 49.62	132.31				
CD_ACC_CATALOG	CATALOG_ID	STRING	6	6	100	1 9JB	9JB	0	0		
CD_ACC_CATALOG	CATALOG_ITEM_ID	STRING	6	6	100	6 9JC	9JH	0	0		
CD_ACC_CATALOG	CLOSED_END_DISPLACEMENT	FLOAT	6	6	100	5 0.123	0.328				
CD_ACC_CATALOG	COMP_TYPE_CODE	STRING	6	6	100	1 ACC	ACC	0	0		
CD_ACC_CATALOG	CONNECTION	STRING	6	0	0	0		0	0		
CD_ACC_CATALOG	CREATE_APP_ID	STRING	6	0	0	0		0	0		
CD_ACC_CATALOG	CREATE_DATE	TIMESTAMP	6	0	0	0					
CD_ACC_CATALOG	CREATE_USER_ID	STRING	6	6	100	1 ss	ss	0	0		
CD_ACC_CATALOG	DESCRIPTION	STRING	6	6	100	2 Griffith	Wilson	6	0		
CD_ACC_CATALOG	GRADE_ID	STRING	6	6	100	1 4145H	4145H	0	0		
CD_ACC_CATALOG	ID_BODY	FLOAT	6	6	100	5 0.125	0.292				
CD_ACC_CATALOG	LENGTH	FLOAT	6	6	100	4 8.3	12.92				
CD_ACC_CATALOG	LINEAR_CAPACITY	FLOAT	6	6	100	5 0.017	0.058				
CD_ACC_CATALOG	MAKEUP_TORQUE	FLOAT	6	6	100	6 11700	47000				
CD_ACC_CATALOG	MANUFACTURER	STRING	6	0	0	0		0	0		
CD_ACC_CATALOG	MODEL	STRING	6	0	0	0		0	0		
CD_ACC_CATALOG	NOMINAL_SIZE	STRING	6	6	100	5 4 3/4	7 3/4	0	0		
CD_ACC_CATALOG	OD_BODY	FLOAT	6	6	100	5 0.396	0.646				
CD_ACC_CATALOG	SECT_TYPE_CODE	STRING	6	6	100	1 ACC	ACC	0	0		
CD_ACC_CATALOG	UPDATE_APP_ID	STRING	6	0	0	0		0	0		
CD_ACC_CATALOG	UPDATE_DATE	TIMESTAMP	6	0	0	0					
CD_ACC_CATALOG	UPDATE_USER_ID	STRING	6	0	0	0		0	0		
CD_AIR_PACKAGE	AIR_PACKAGE_ID	STRING	0	0	0	0		0	0		
CD_AIR_PACKAGE	AIR_PACKAGE_NAME	STRING	0	0	0	0		0	0		
CD_AIR_PACKAGE	BOOSTER_DESCRIPTION	STRING	0	0	0	0		0	0		
CD_AIR_PACKAGE	BOOSTER_MAKE	STRING	0	0	0	0		0	0		
CD_AIR_PACKAGE	BOOSTER_MODEL	STRING	0	0	0	0		0	0		
CD_AIR_PACKAGE	BOOSTER_RATING	FLOAT	0	0	0	0					
CD_AIR_PACKAGE	COMPRESSOR_DESCRIPTION	STRING	0	0	0	0		0	0		
CD_AIR_PACKAGE	COMPRESSOR_MAKE	STRING	0	0	0	0		0	0		
CD_AIR_PACKAGE	COMPRESSOR_MODEL	STRING	0	0	0	0		0	0		
CD_AIR_PACKAGE	COMPRESSOR_RATING	FLOAT	0	0	0	0					
CD_AIR_PACKAGE	CONTRACTOR_ID	STRING	0	0	0	0		0	0		
CD_AIR_PACKAGE	CREATE_APP_ID	STRING	0	0	0	0		0	0		
CD_AIR_PACKAGE	CREATE_DATE	TIMESTAMP	0	0	0	0					
CD_AIR_PACKAGE	CREATE_USER_ID	STRING	0	0	0	0		0	0		
CD_AIR_PACKAGE	IS_READONLY	STRING	0	0	0	0		0	0		
CD_CBL_CATALOG	NUMBER_OF_BOOSTERS	INTEGER	0	0	0	0					

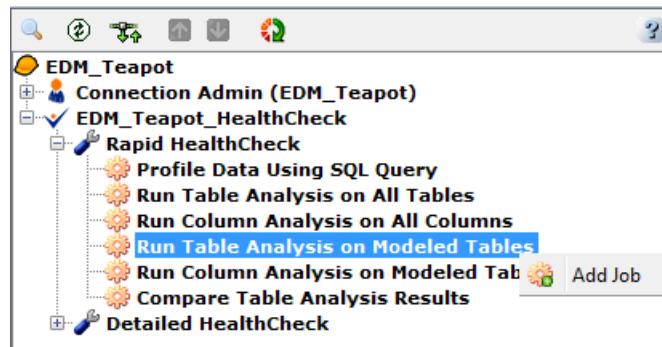
12. Select **File > Exit** to close the **Basic View Frame** window.

### Exercise: Running Table Analysis on Modeled Tables Task

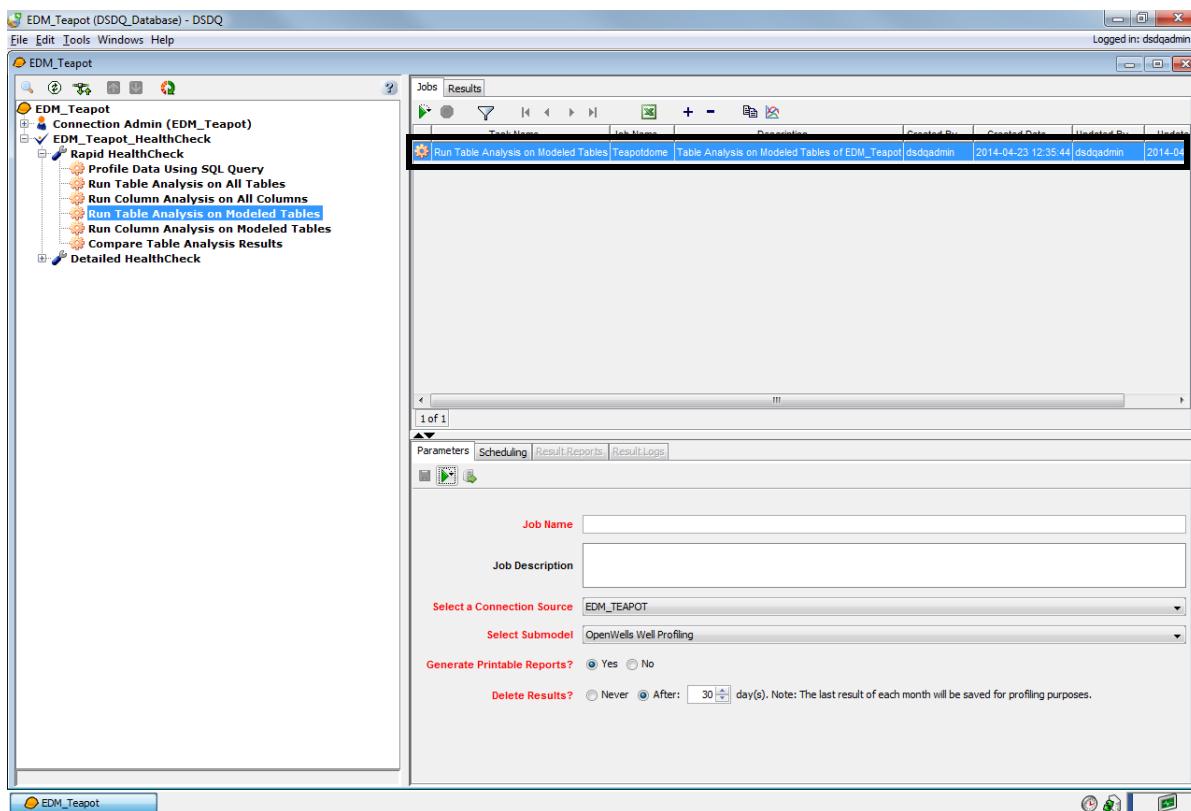
The **Run Table Analysis on Modeled Tables** Task runs only on tables that have been modeled in the **Perform Table Modeling** Tool (*reference: DecisionSpace Data Quality Training Manual, Chapter 4: Data Evaluation, Perform Table Modeling*). To run Table Analysis on all the modeled tables:

1. Double-click the **Run Table Analysis on Modeled Tables** Task or right-click the **Run Table Analysis on Modeled Tables** Task and

select **Add Job** from the pop-up menu.

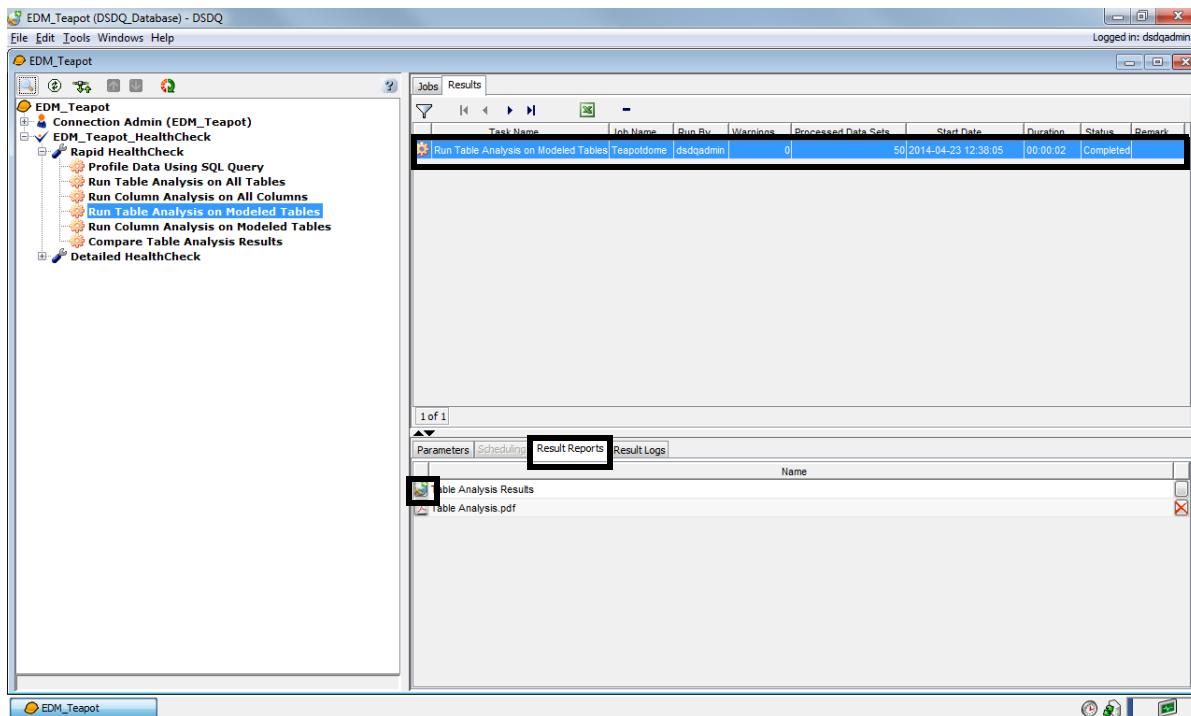


A new job is initiated and it displays on the **Job and Results Information Pane** on the right side of the DecisionSpace Data Quality Project window.

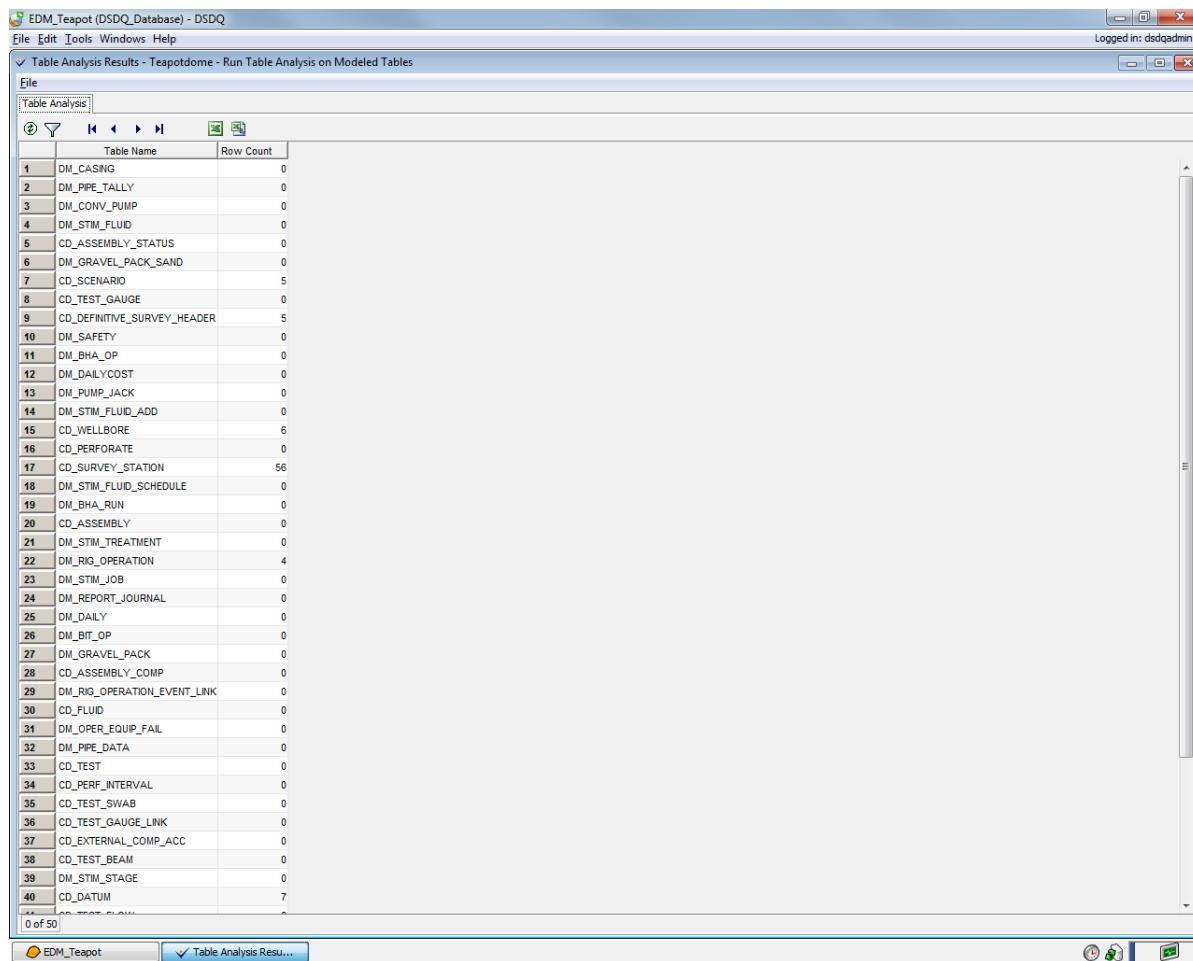


2. Enter **Teapotdome** in the **Job Name** field.
3. Enter **Table Analysis on Modeled Tables of EDM\_Teapot** in the **Job Description** field.
4. Select **EDM\_Teapot** from the **Select a Connection Source** drop-down list.

5. Select **OpenWells Well Profiling** from the **Select Submodel** drop-down list.
6. Select the **Yes** option for **Generate Printable Reports**.
7. Select the **After** option for **Delete Results**. Leave the number of days as **30**.
8. Click  to save changes in the **Parameters** tab.
9. Click  to run the job.  
The **Run Table Analysis on Modeled Tables** Task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.
10. Select the **Results** tab.  
The **Jobs and Results Listing Pane** displays a list of results.



11. Click the **Open Basic View Frame**  button on the **Result Reports** tab to display the **Run Table Analysis on Modeled Tables** Task results in the **Basic View Frame** window.



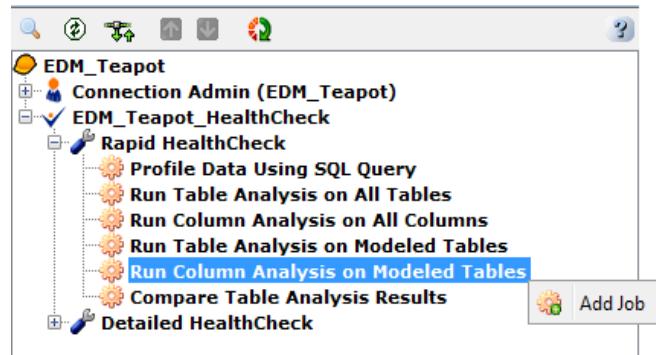
12. Select **File > Exit** to close the **Basic View Frame** window.

### Exercise: Running Column Analysis on Modeled Tables Task

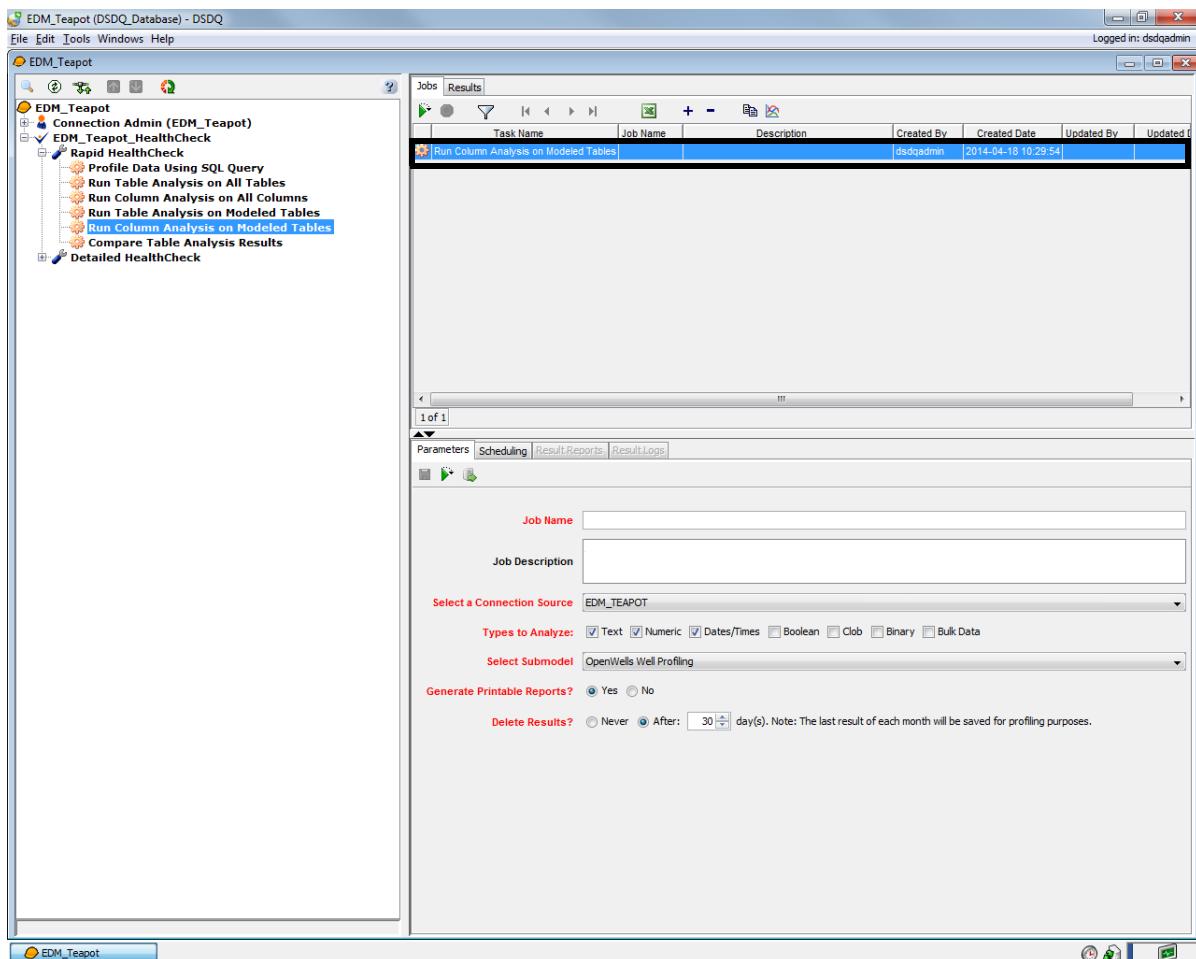
The **Run Column Analysis on Modeled Tables** Task runs only on tables that have been modeled in the **Perform Table Modeling** Tool (reference: *DecisionSpace Data Quality Training Manual, Chapter 4: Data Evaluation, Perform Table Modeling*). To run Column Analysis on all the modeled tables:

1. Double-click the **Run Column Analysis on Modeled Tables** Task or right-click the **Run Column Analysis on Modeled Tables** Task

and select **Add Job** from the pop-up menu.



A new job is initiated and it displays on the **Job and Results Information Pane** on the right side of the DecisionSpace Data Quality Project window.



2. Enter **Teapotdome** in the **Job Name** field.
3. Enter **Column Analysis on Modeled Tables of EDM\_Teapot** in the **Job Description** field.

4. Select **EDM\_Teapot** from the **Select a Connection Source** drop-down list.

The **Text**, **Numeric** and **Dates/Times** options are selected by default for **Types to Analyze**.

5. Select **OpenWells Well Profiling** from the **Select Submodel** drop-down list.

6. Select the **Yes** option for **Generate Printable Reports**.

7. Select the **After** option for **Delete Results**. Leave the number of days as **30**.

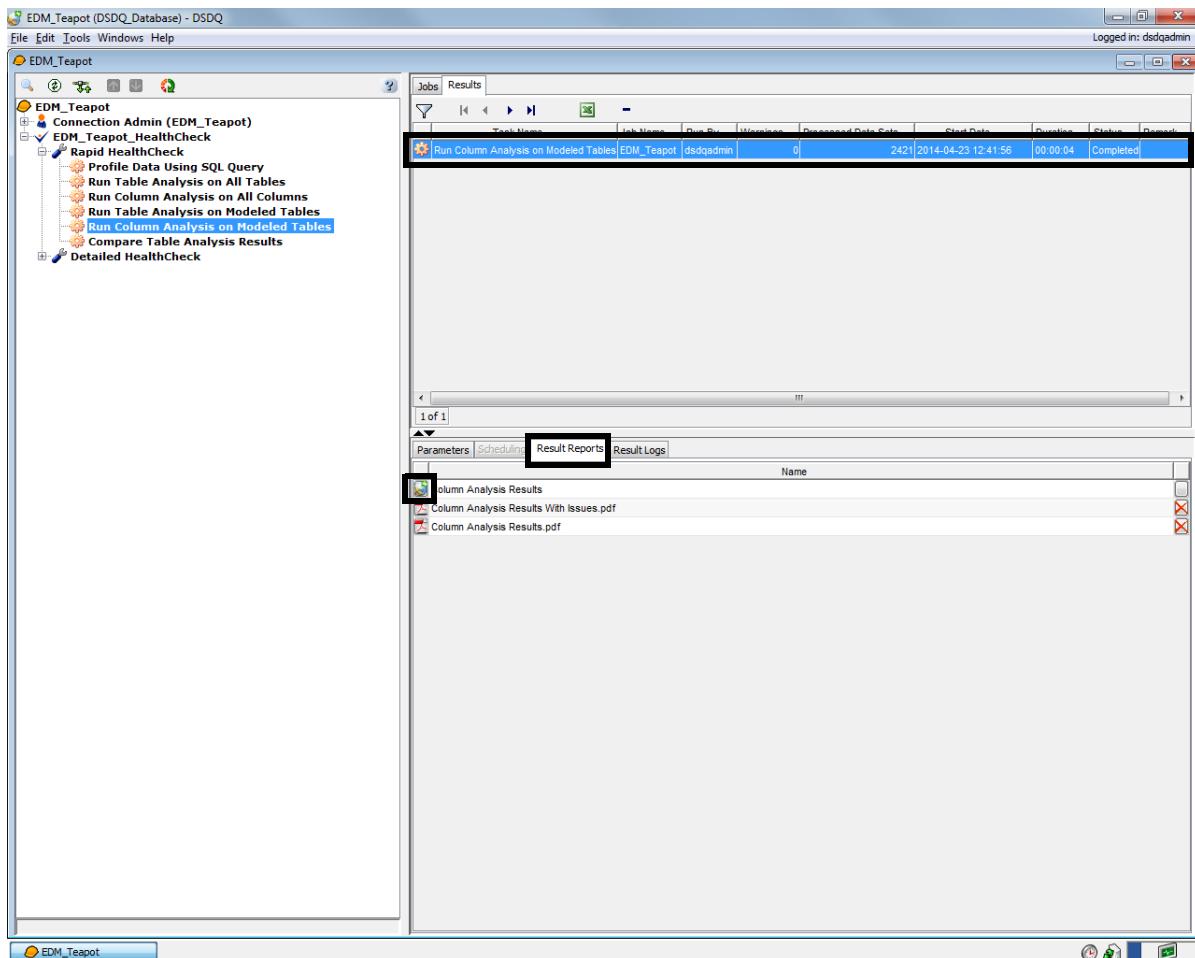
8. Click  to save changes in the **Parameters** tab.

9. Click  to run the job.

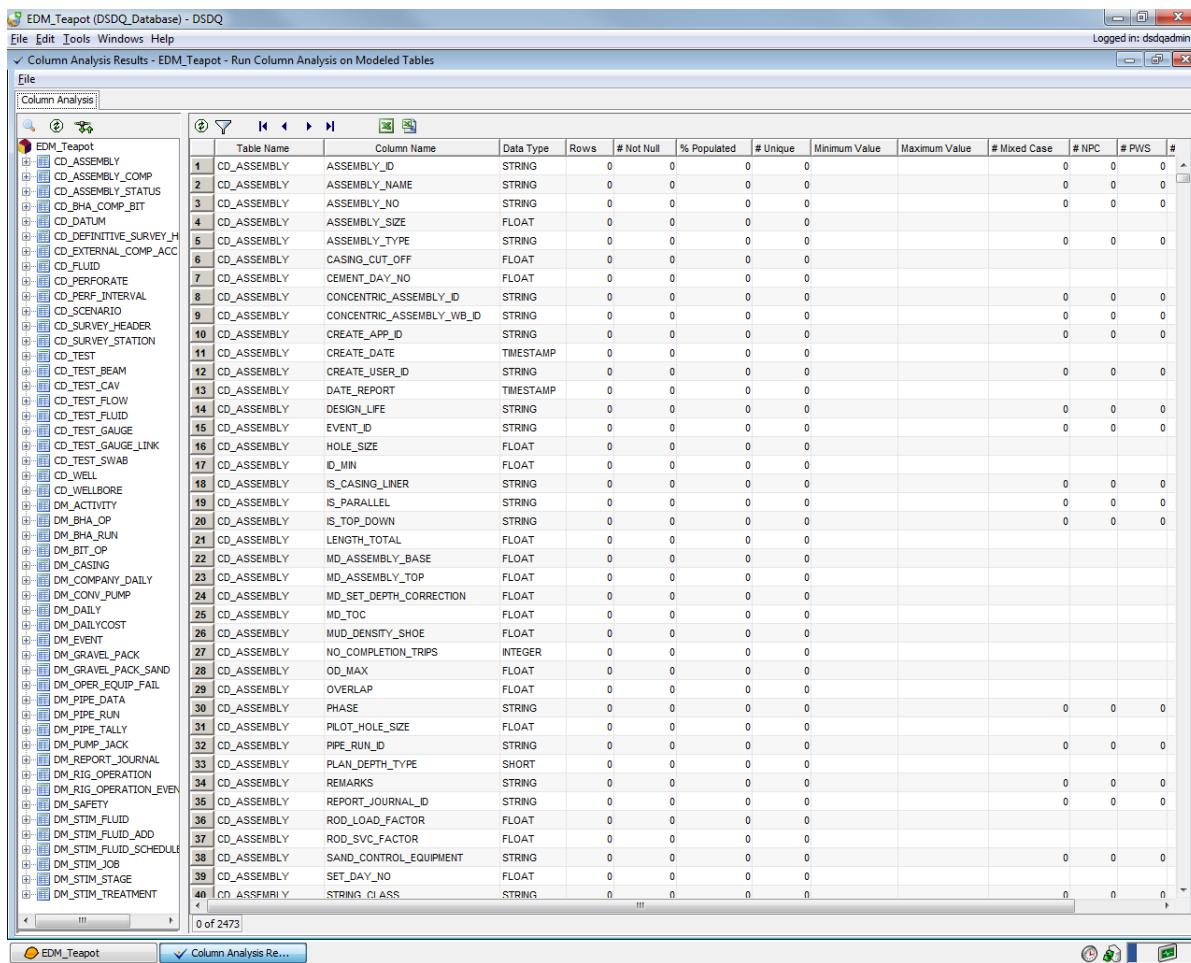
The **Run Column Analysis on Modeled Tables** Task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

## 10. Select the **Results** tab.

The **Jobs and Results Listing Pane** displays a list of results.



11. Click the **Open Basic View Frame**  button on the Result Reports tab to display the Column Analysis on Modeled Tables Task results in the **Basic View Frame** window.



The screenshot shows the 'Column Analysis' results for the 'EDM\_Teapot' database. The table has the following structure:

	Table Name	Column Name	Data Type	Rows	# Not Null	% Populated	# Unique	Minimum Value	Maximum Value	# Mixed Case	# NPC	# PWS	# Errors
1	CD_ASSEMBLY	ASSEMBLY_ID	STRING	0	0	0	0			0	0	0	
2	CD_ASSEMBLY	ASSEMBLY_NAME	STRING	0	0	0	0			0	0	0	
3	CD_ASSEMBLY	ASSEMBLY_NO	STRING	0	0	0	0			0	0	0	
4	CD_ASSEMBLY	ASSEMBLY_SIZE	FLOAT	0	0	0	0			0	0	0	
5	CD_ASSEMBLY	ASSEMBLY_TYPE	STRING	0	0	0	0			0	0	0	
6	CD_ASSEMBLY	CASING_CUT_OFF	FLOAT	0	0	0	0			0	0	0	
7	CD_ASSEMBLY	CEMENT_DAY_NO	FLOAT	0	0	0	0			0	0	0	
8	CD_ASSEMBLY	CONCENTRIC_ASSEMBLY_ID	STRING	0	0	0	0			0	0	0	
9	CD_ASSEMBLY	CONCENTRIC_ASSEMBLY_WB_ID	STRING	0	0	0	0			0	0	0	
10	CD_ASSEMBLY	CREATE_APP_ID	STRING	0	0	0	0			0	0	0	
11	CD_ASSEMBLY	CREATE_DATE	TIMESTAMP	0	0	0	0			0	0	0	
12	CD_ASSEMBLY	CREATE_USER_ID	STRING	0	0	0	0			0	0	0	
13	CD_ASSEMBLY	DATE_REPORT	TIMESTAMP	0	0	0	0			0	0	0	
14	CD_ASSEMBLY	DESIGN_LIFE	STRING	0	0	0	0			0	0	0	
15	CD_ASSEMBLY	EVENT_ID	STRING	0	0	0	0			0	0	0	
16	CD_ASSEMBLY	HOLE_SIZE	FLOAT	0	0	0	0			0	0	0	
17	CD_ASSEMBLY	ID_MIN	FLOAT	0	0	0	0			0	0	0	
18	CD_ASSEMBLY	IS_CASING_LINER	STRING	0	0	0	0			0	0	0	
19	CD_ASSEMBLY	IS_PARALLEL	STRING	0	0	0	0			0	0	0	
20	CD_ASSEMBLY	IS_TOP_DOWN	STRING	0	0	0	0			0	0	0	
21	CD_ASSEMBLY	LENGTH_TOTAL	FLOAT	0	0	0	0			0	0	0	
22	CD_ASSEMBLY	MD_ASSEMBLY_BASE	FLOAT	0	0	0	0			0	0	0	
23	CD_ASSEMBLY	MD_ASSEMBLY_TOP	FLOAT	0	0	0	0			0	0	0	
24	CD_ASSEMBLY	MD_SET_DEPTH_CORRECTION	FLOAT	0	0	0	0			0	0	0	
25	CD_ASSEMBLY	MD_TOC	FLOAT	0	0	0	0			0	0	0	
26	CD_ASSEMBLY	MUD_DENSITY_SHOE	FLOAT	0	0	0	0			0	0	0	
27	CD_ASSEMBLY	NO_COMPLETION_TRIPS	INTEGER	0	0	0	0			0	0	0	
28	CD_ASSEMBLY	OD_MAX	FLOAT	0	0	0	0			0	0	0	
29	CD_ASSEMBLY	OVERLAP	FLOAT	0	0	0	0			0	0	0	
30	CD_ASSEMBLY	PHASE	STRING	0	0	0	0			0	0	0	
31	CD_ASSEMBLY	PILOT_HOLE_SIZE	FLOAT	0	0	0	0			0	0	0	
32	CD_ASSEMBLY	PIPE_RUN_ID	STRING	0	0	0	0			0	0	0	
33	CD_ASSEMBLY	PLAN_DEPTH_TYPE	SHORT	0	0	0	0			0	0	0	
34	CD_ASSEMBLY	REMARKS	STRING	0	0	0	0			0	0	0	
35	CD_ASSEMBLY	REPORT_JOURNAL_ID	STRING	0	0	0	0			0	0	0	
36	CD_ASSEMBLY	ROD_LOAD_FACTOR	FLOAT	0	0	0	0			0	0	0	
37	CD_ASSEMBLY	ROD_SVC_FACTOR	FLOAT	0	0	0	0			0	0	0	
38	CD_ASSEMBLY	SAND_CONTROL_EQUIPMENT	STRING	0	0	0	0			0	0	0	
39	CD_ASSEMBLY	SET_DAY_NO	FLOAT	0	0	0	0			0	0	0	
40	CD_ASSEMBLY	STRING_CLASS	STRING	0	0	0	0			0	0	0	

12. Select **File > Exit** to close the **Basic View Frame** window.

## Detailed HealthCheck Activity

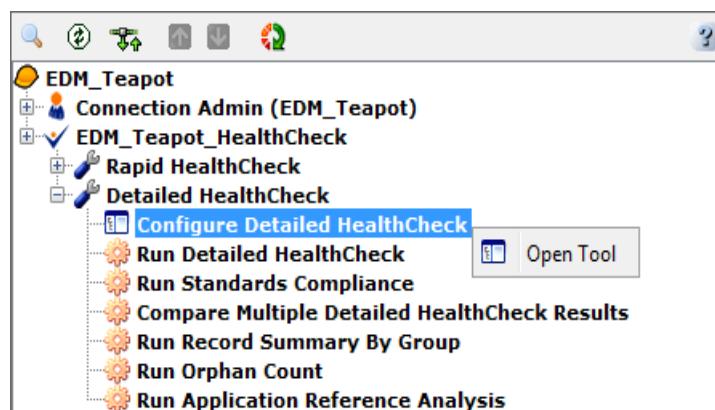
The **Detailed HealthCheck Activity** allows you to run business rules against the dataset to identify data problems.

### Exercise: Configuring the Detailed HealthCheck Tool

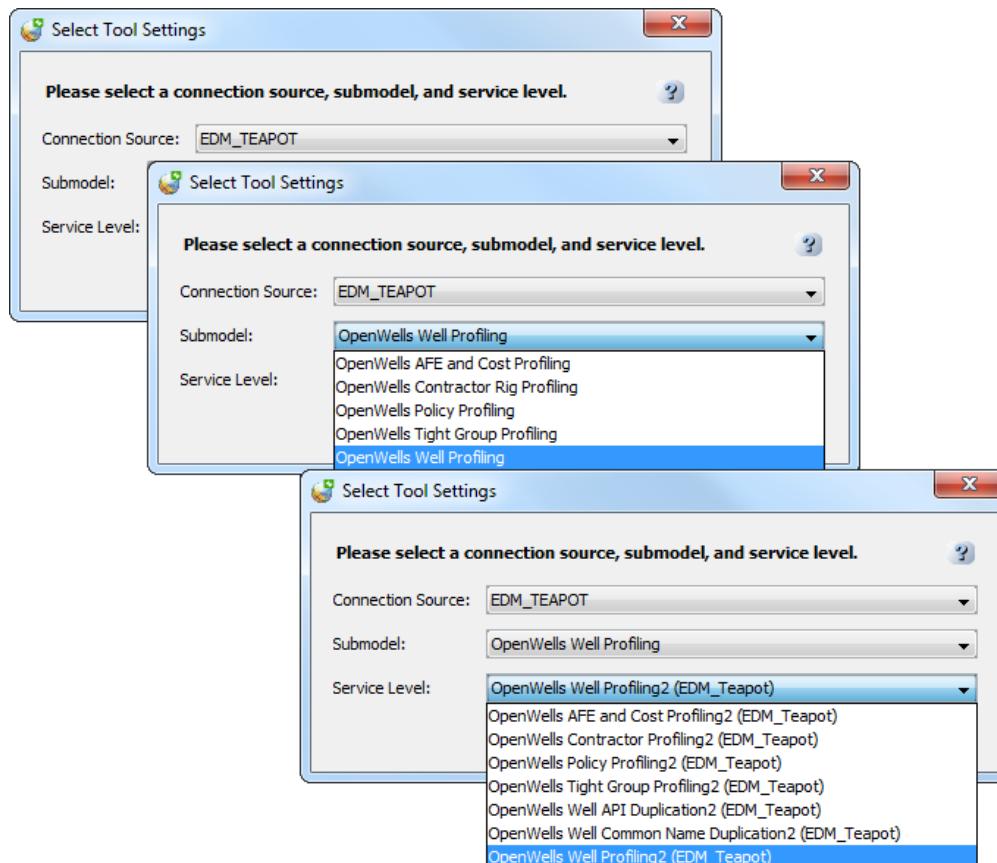
The **Configure Detailed HealthCheck Tool** configures service levels for testing prior to running the **Run Detailed HealthCheck Task**. You can select which requirements are to be enabled/disabled in the service level. You can also select subset of the total data to be used when testing a service level. A service level containing HealthCheck requirements

must exist in the DSDQ project prior to opening the **Configure Detailed HealthCheck** Tool. To configure the Detailed HealthCheck tool:

1. Click  to expand the **Detailed HealthCheck** Activity.
2. Double-click the **Configure Detailed HealthCheck** Tool or right-click the **Configure Detailed HealthCheck** Tool and select **Open Tool** from the pop-up menu.

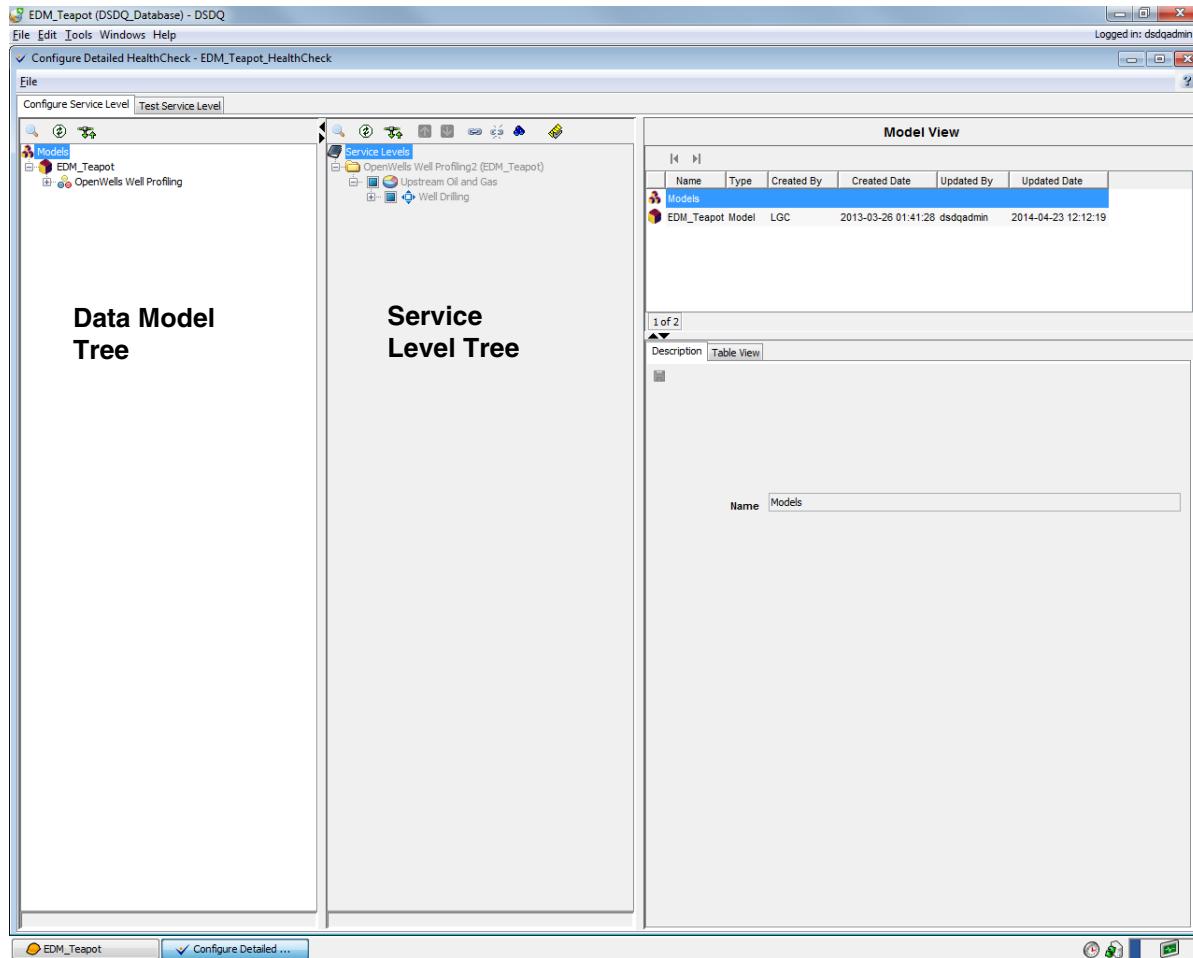


The **Select Tool Settings** window appears.



3. The **Connection Source** drop-down list is set to **EDM\_Teapot** by default.
4. Select **OpenWells Well Profiling** from the **Submodel** drop-down list.
5. Select **OpenWells Well Profiling2 (EDM\_Teapot)** from the **Service Level** drop-down list.
6. Click **OK**.

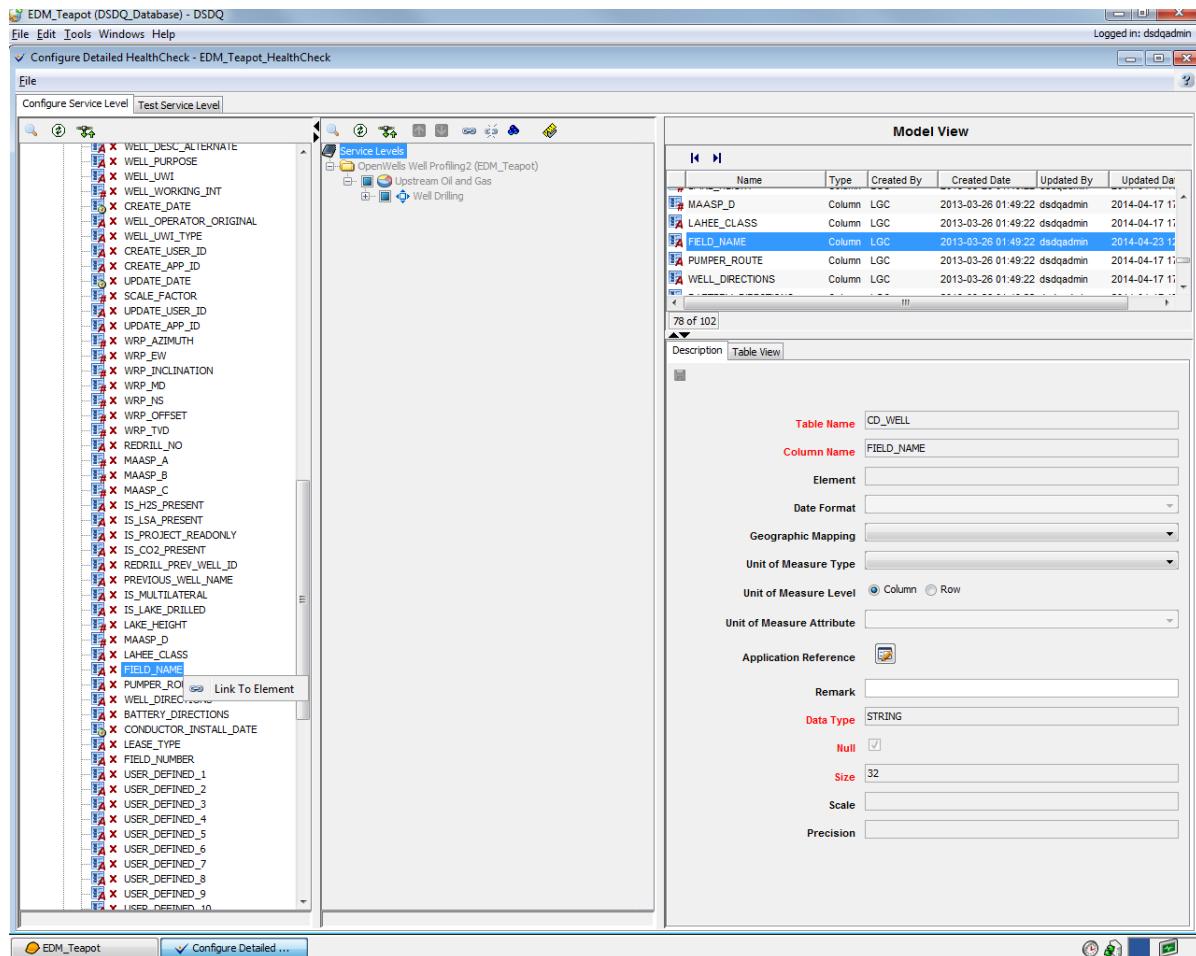
The **Configure Detailed HealthCheck** window appears.



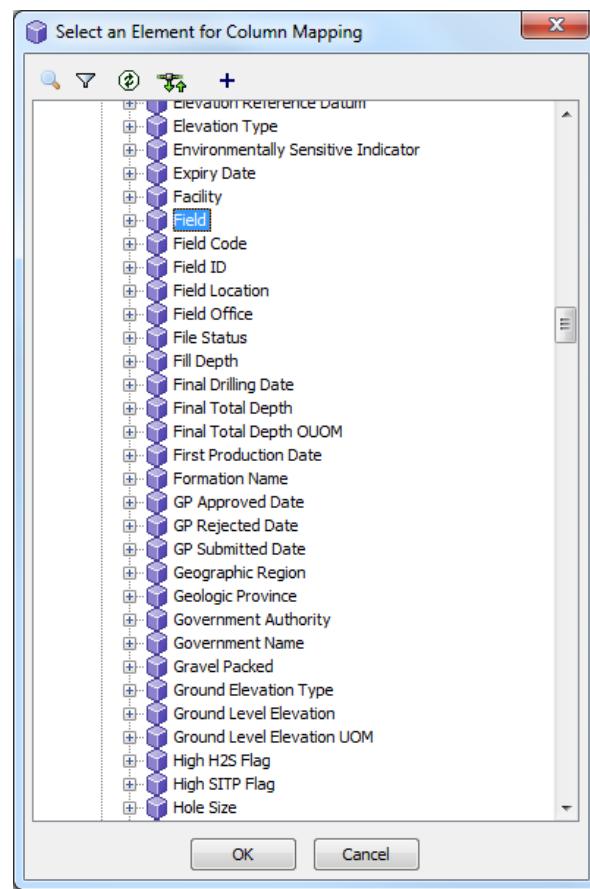
The Data Model Tree populates with the submodel selected in step 4. The Service Level Tree populates with the service level selected in step 5.

7. Click to expand the **OpenWells Well Profiling** submodel in the DataModel Tree.
8. Click to expand the **CD\_Well** table.

9. Right-click the **FIELD\_NAME** column and select **Link to Element** from the pop-up menu.



The **Select an Element for Column Mapping** window appears.

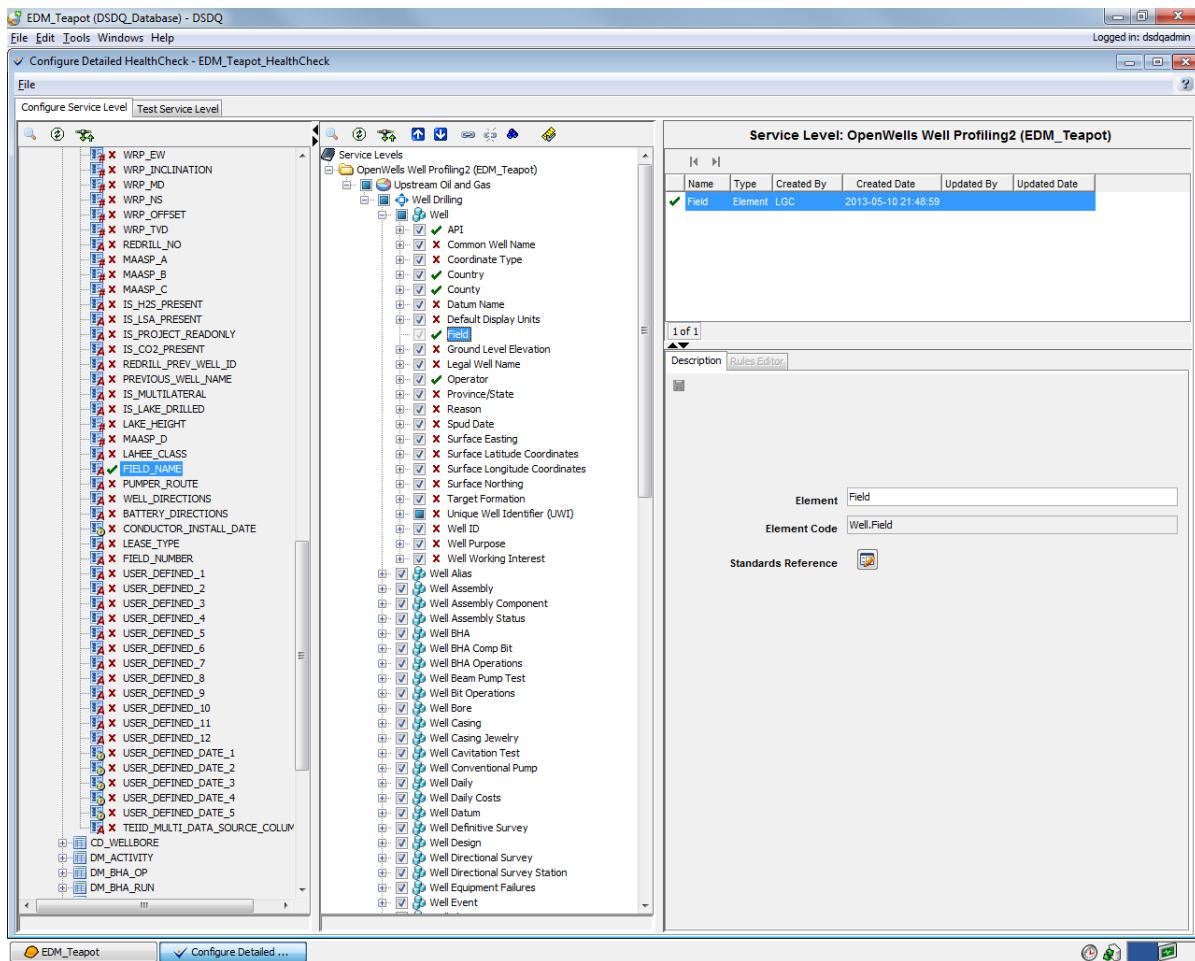


10. Select the **Field** element.

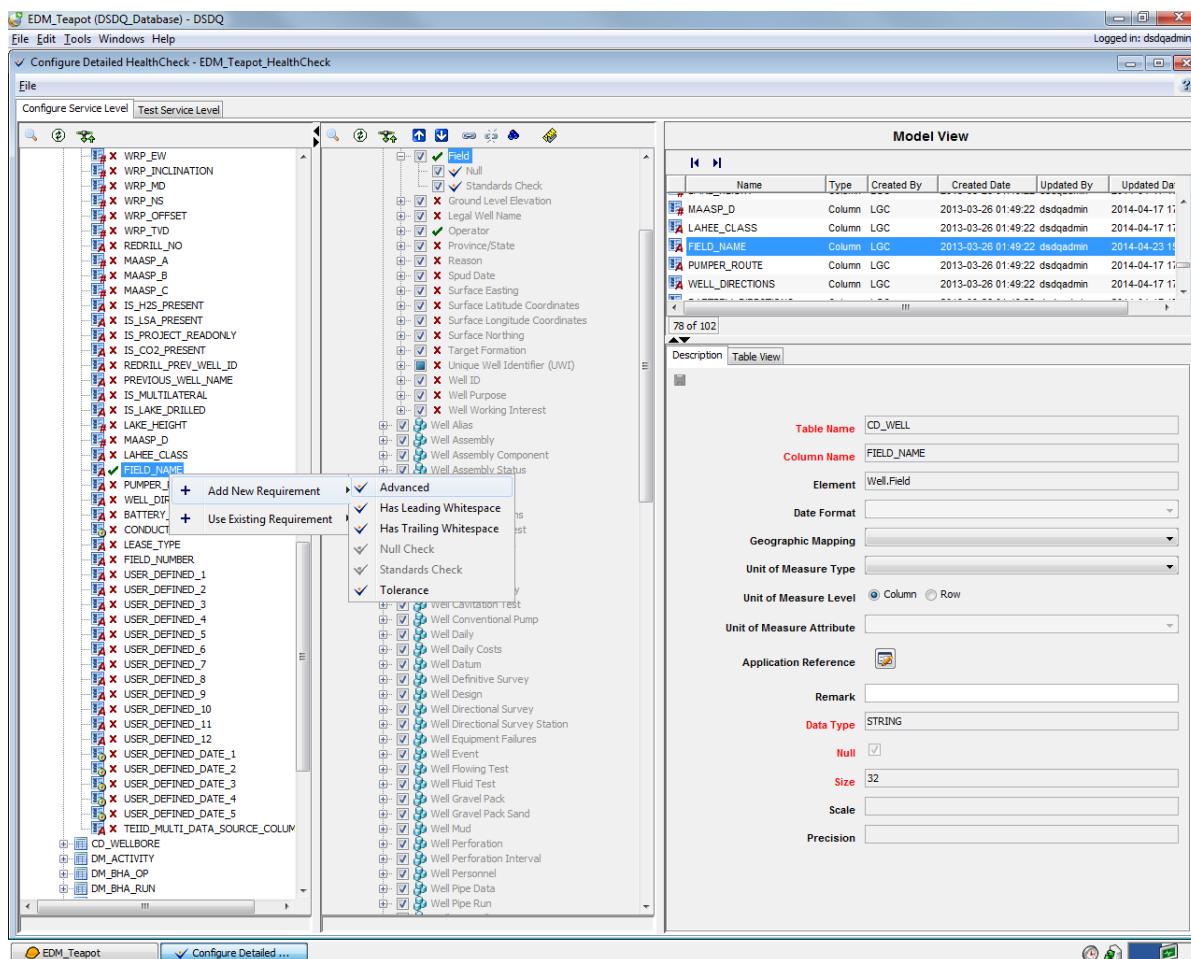
11. Click **OK**.

The **FIELD\_NAME** column and the **Field** element are associated with each other. A green check mark appears adjacent to the column and element that have just been associated. Only one column from the same table can be linked to the same element.

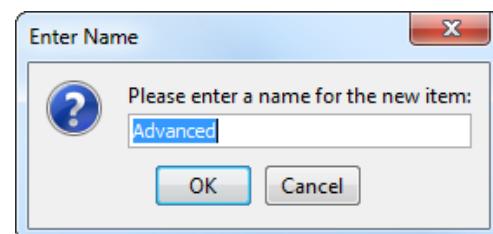
However, it is possible to link different tables' columns to the same element.



12. Right-click the **FIELD\_NAME** column in the Data Model Tree and select **Add New Requirement > Advanced** from the pop-up menu.



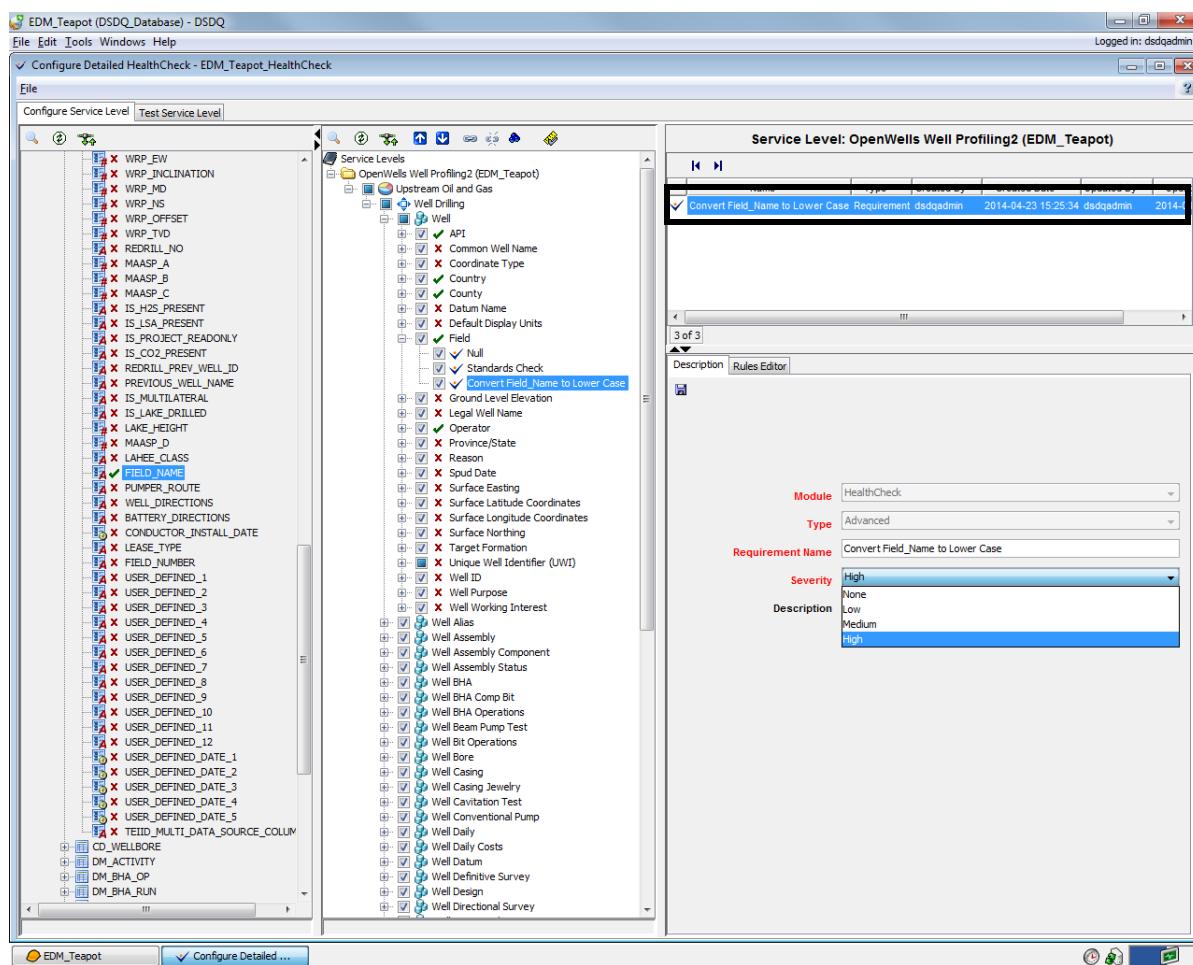
The Enter Name dialog box appears.



13. Enter **Convert Field\_Name to Lower Case** in the **Please enter a name for the new item** field.

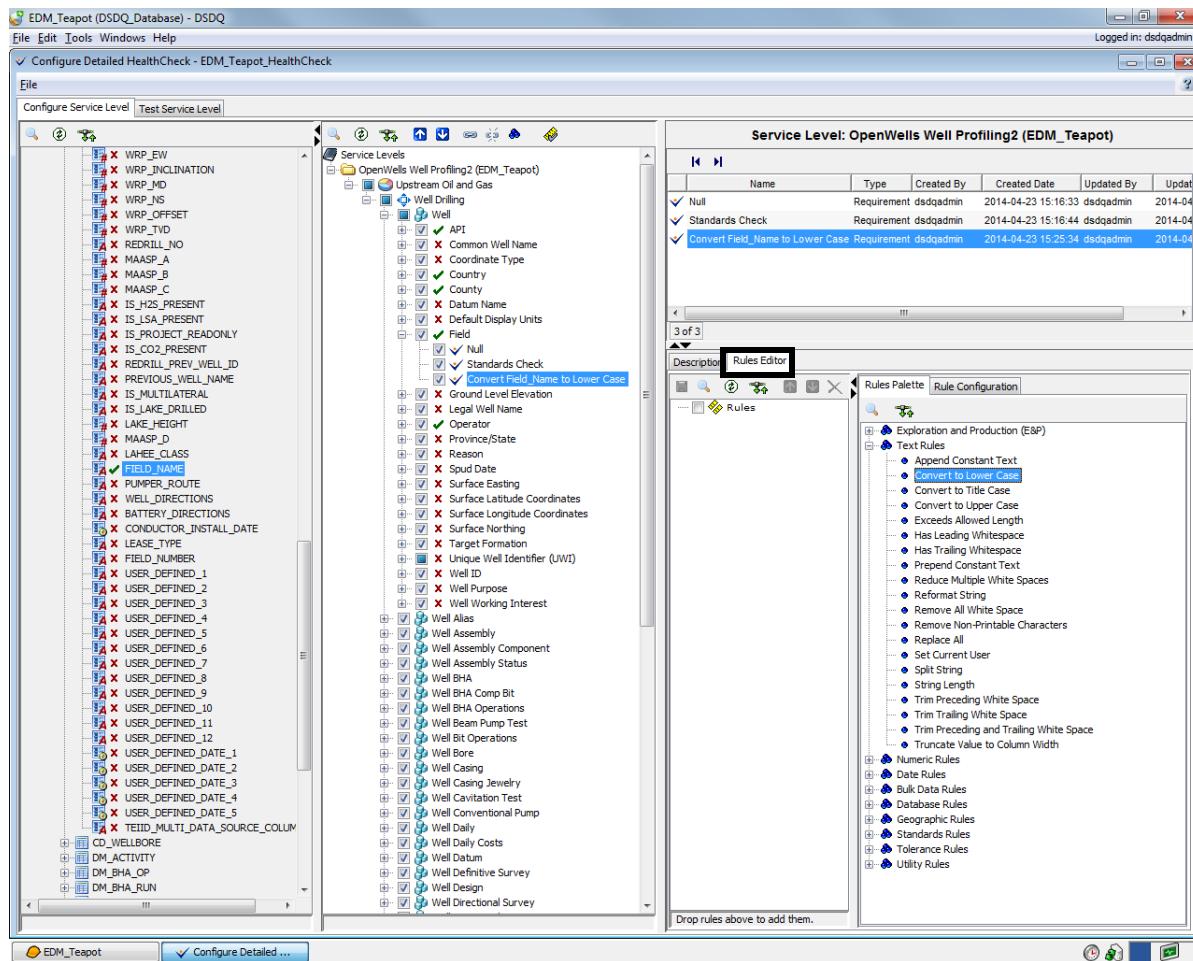


14. Click **OK** to add the requirement to the selected column.  
 The **Convert Field\_Name to Lower Case** requirement is added to the **Field** element.



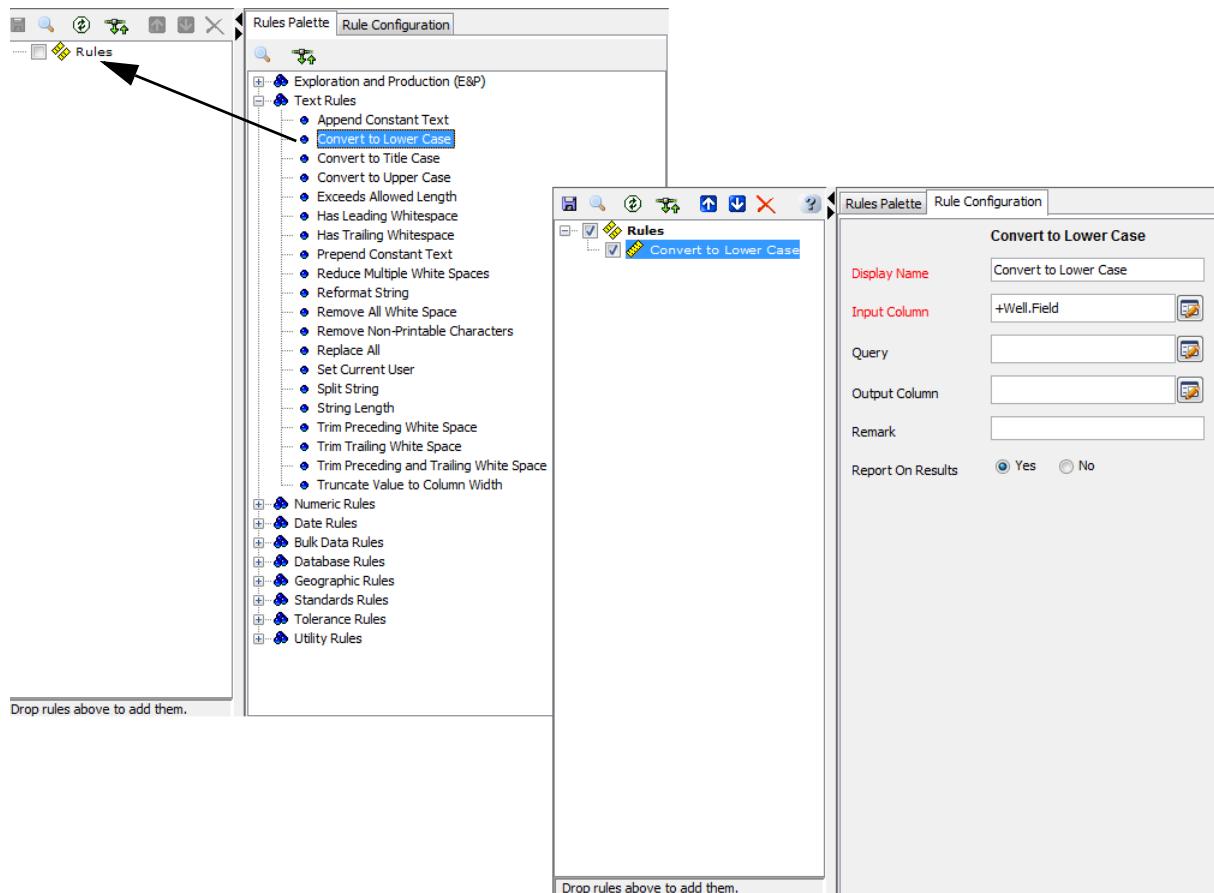
15. The **Module** and **Type** fields are disabled by default.
16. The **Requirement name** field populates with the name assigned in step **13**.
17. Select **High** from the **Severity** drop-down list.
18. Enter **Advance Rule for FIELD\_NAME column** in the **Description** field.
19. Click to save changes in the **Description** tab.
20. Repeat step **12** to **19** to add the **Has Leading White Spaces** requirement.

21. Select the **Rules Editor** tab adjacent to the **Description** tab.



22. Click to expand the **Text Rules** tree in the **Rules Palette** tab.

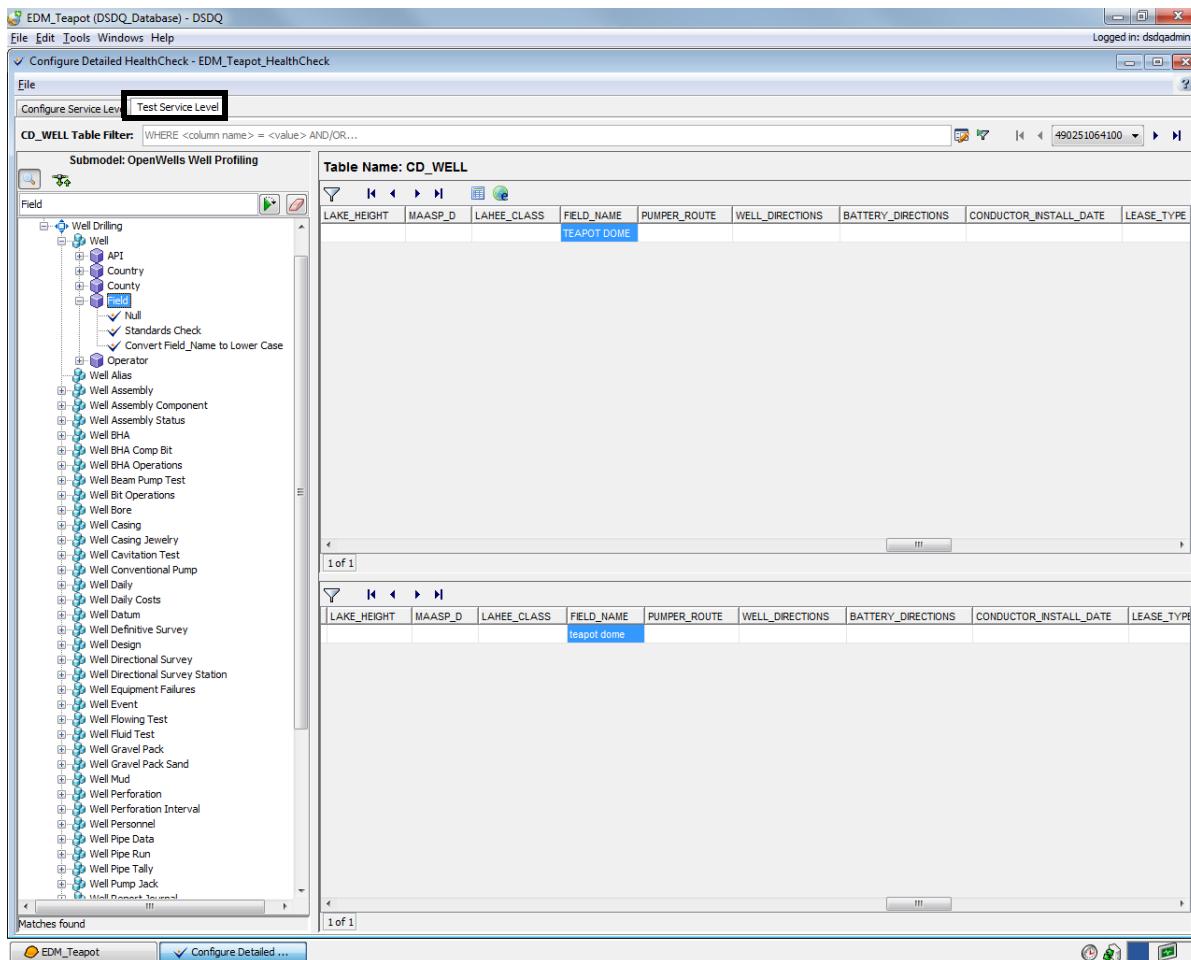
23. Drag and drop the **Convert to Lower Case** rule onto the **Rule** area.



24. Click to save changes in the **Rules Editor** tab.

25. Select the **Test Service Level** tab.

The test is automatically executed for the first record of the test data subset.



By looking at the columns that have been changed and temporary columns, you can verify that the behavior of the service level is correct prior to running the **Run Detailed HealthCheck** Task.

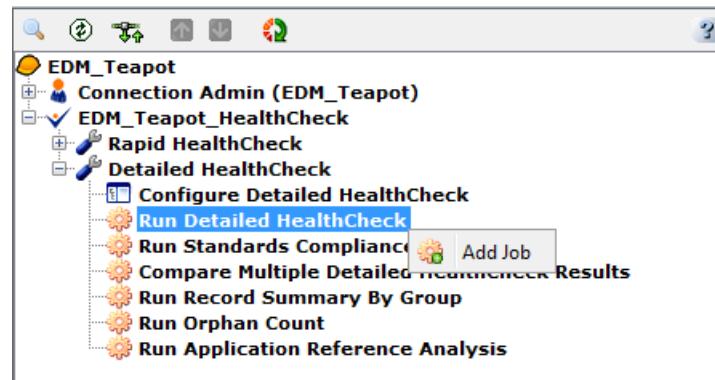
26. Click the **Next Data Set** button to test the next record.
27. Repeat step **26** to test all records.
28. Select **File > Exit** to close the **Configure Detailed HealthCheck** window.

### Exercise: Running the Detailed HealthCheck Task

The **Run Detailed HealthCheck** Task generates results for the requirements that are enabled in the service level. Prior to executing the **Run Detailed HealthCheck** Task, ensure that columns in the specified

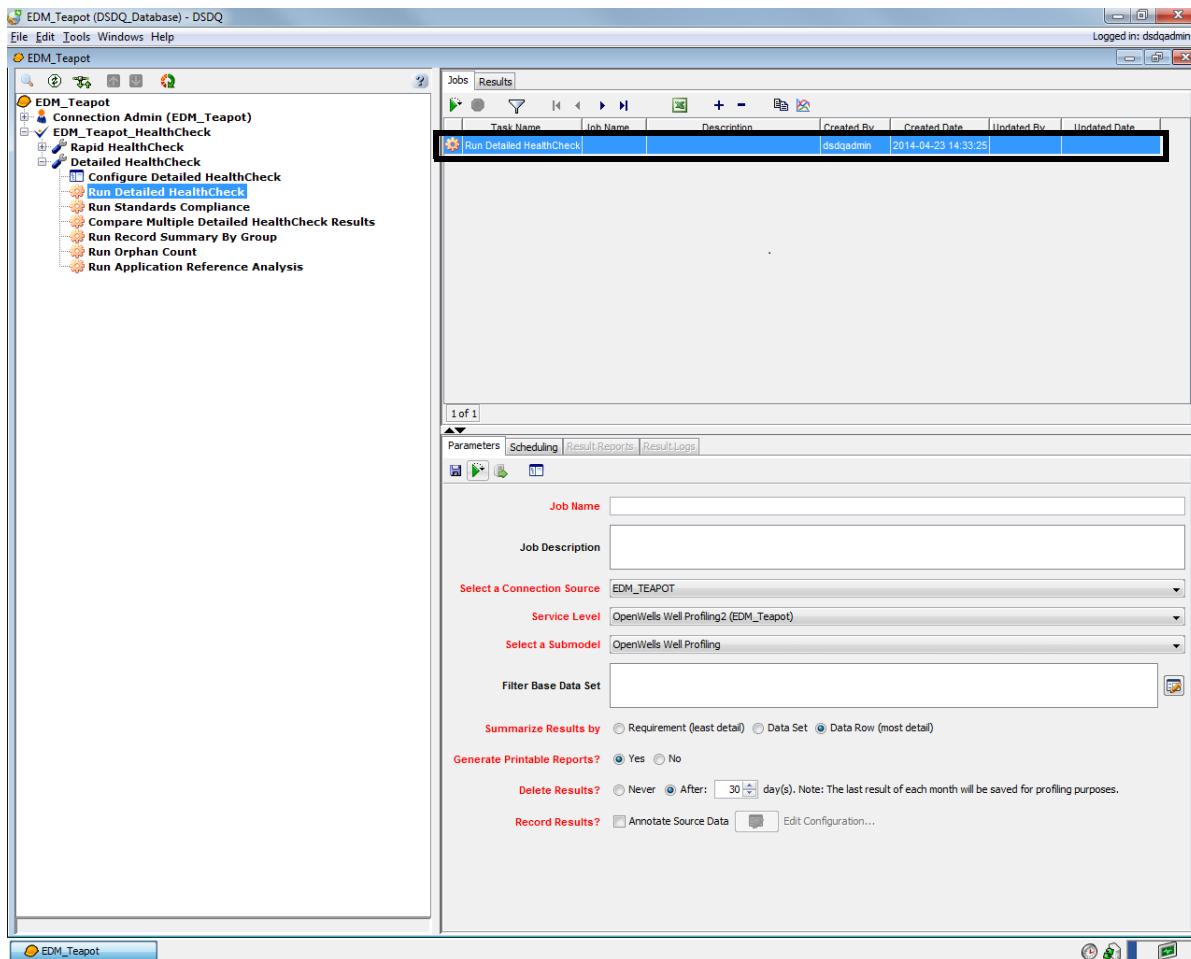
submodel table have been assigned elements from the desired service level. To run the Detailed HealthCheck task:

1. Double-click the **Run Detailed HealthCheck** Task or right-click the **Run Detailed HealthCheck** Task and select **Add Job** from the pop-up menu.



A new job is initiated and it displays on the **Job and Results**

**Information Pane** on the right side of the DecisionSpace Data Quality Project window.



2. Enter **Teapotdome** in the **Job Name** field.
3. Enter **Detailed HealthCheck** for **EDM\_Teapot** in the **Job Description** field.
4. Select **EDM\_TEAPOT** from the **Select a Connection Source** drop-down list.
5. Select **OpenWells Well Profiling2 (EDM\_Teapot)** from the **Service Level** drop-down list.
6. Select **OpenWells Well Profiling** from the **Select a Submodel** drop-down list.
7. Do not select the filter for **Filter Base Data Set**.

8. Select the **Data Row (most detail)** option for **Summarize Results by**.

9. Select the **Yes** option for **Generate Printable Reports**.

10. Select the **After** option for **Delete Results**. Leave the number of days as **30**.

11. Do not select the check box for **Record Results**.

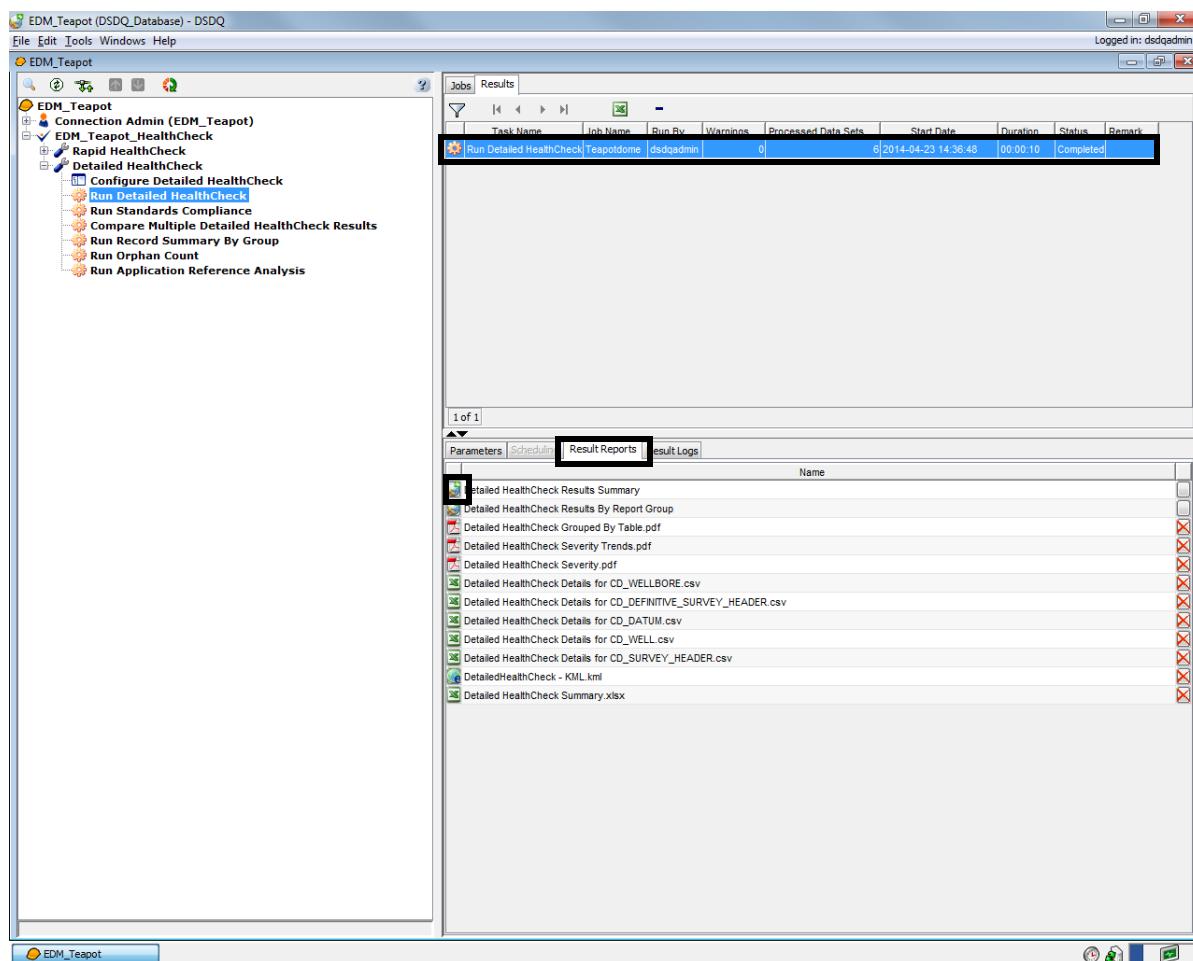
12. Click to save changes in the **Parameters** tab.

13. Click to run the job.

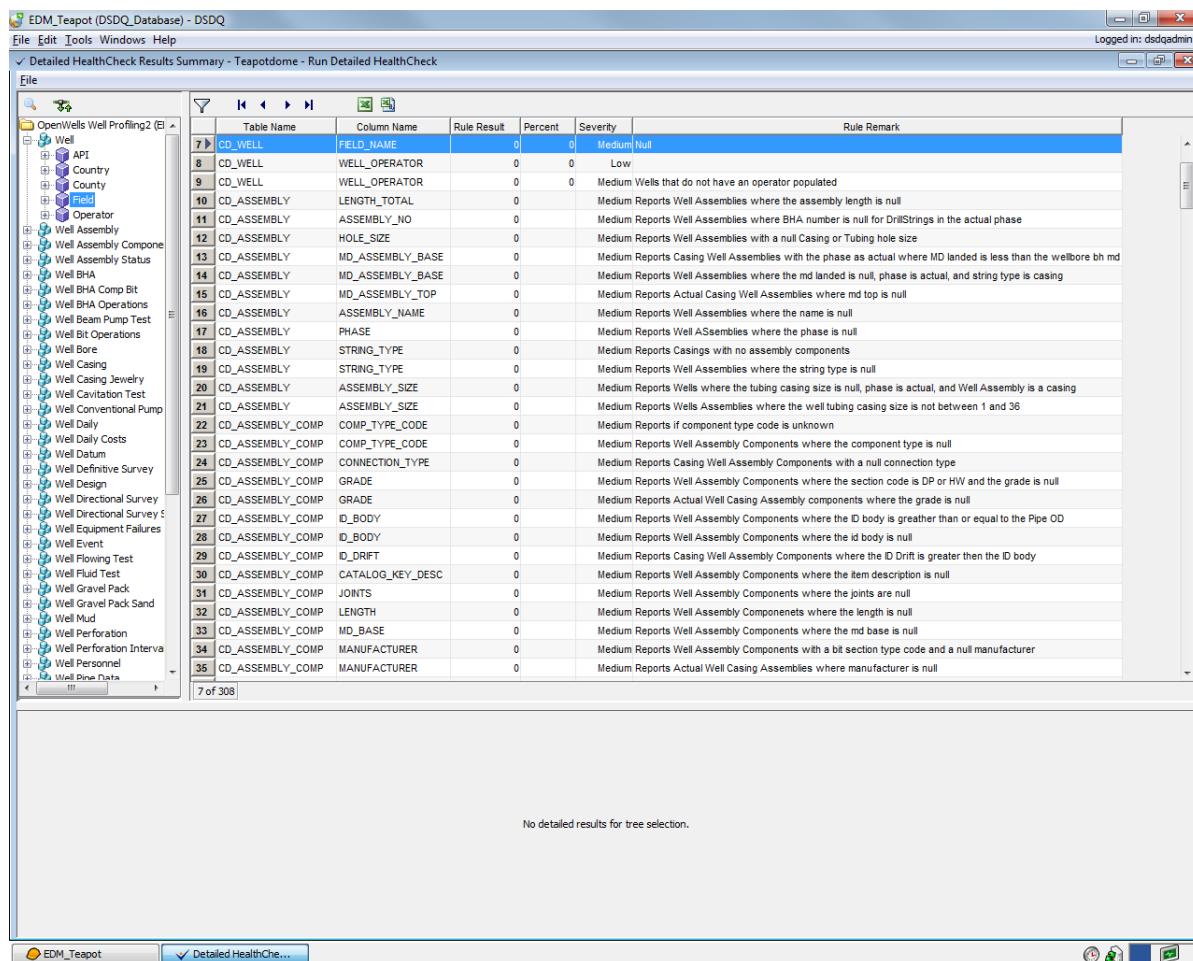
The **Run Detailed HealthCheck** Task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

14. Select the **Results** tab.

The **Jobs and Results Listing Pane** displays a list of results.



15. Click the **Open Basic View Frame**  button on the **Result Reports** tab to display the **Run Detailed HealthCheck** Task results in the **Basic View Frame** window.



The screenshot shows the EDM Teapot (DSDQ\_Database) - DSDQ application window. The title bar indicates the database is 'EDM\_Teapot (DSDQ\_Database) - DSDQ'. The menu bar includes File, Edit, Tools, Windows, Help, and a status message 'Logged in: dsdqadmin'. The main area is titled 'Detailed HealthCheck Results Summary - Teapotdome - Run Detailed HealthCheck'. On the left, there is a tree view of well components under 'Well Assembly Components'. The right side contains a table with 35 rows of health check results. The columns are: Table Name, Column Name, Rule Result, Percent, Severity, and Rule Remark. The first row is highlighted in blue, indicating it is selected. The 'Rule Remark' column provides detailed descriptions for each rule.

Table Name	Column Name	Rule Result	Percent	Severity	Rule Remark
CD_WELL	FIELD_NAME	0	0	Medium Null	
CD_WELL	WELL_OPERATOR	0	0	Low	Medium Wells that do not have an operator populated
CD_WELL	WELL_OPERATOR	0	0	Medium	Reports Well Assemblies where the assembly length is null
CD_ASSEMBLY	LENGTH_TOTAL	0	0	Medium	Reports Well Assemblies where BHA number is null for DrillStrings in the actual phase
CD_ASSEMBLY	ASSEMBLY_NO	0	0	Medium	Reports Well Assemblies with a null Casing or Tubing hole size
CD_ASSEMBLY	HOLE_SIZE	0	0	Medium	Reports Casing Well Assemblies with the phase as actual where MD landed is less than the wellbore bh md
CD_ASSEMBLY	MD_ASSEMBLY_BASE	0	0	Medium	Reports Well Assemblies where the md landed is null, phase is actual, and string type is casing
CD_ASSEMBLY	MD_ASSEMBLY_BASE	0	0	Medium	Reports Actual Casing Well Assemblies where md top is null
CD_ASSEMBLY	MD_ASSEMBLY_TOP	0	0	Medium	Reports Well Assemblies where the name is null
CD_ASSEMBLY	ASSEMBLY_NAME	0	0	Medium	Reports Well Assemblies where the phase is null
CD_ASSEMBLY	PHASE	0	0	Medium	Reports Casings with no assembly components
CD_ASSEMBLY	STRING_TYPE	0	0	Medium	Reports Well Assemblies where the string type is null
CD_ASSEMBLY	STRING_TYPE	0	0	Medium	Reports Wells where the tubing casing size is null, phase is actual, and Well Assembly is a casing
CD_ASSEMBLY	ASSEMBLY_SIZE	0	0	Medium	Reports Wells Assemblies where the well tubing casing size is not between 1 and 36
CD_ASSEMBLY_COMP	COMP_TYPE_CODE	0	0	Medium	Reports if component type code is unknown
CD_ASSEMBLY_COMP	COMP_TYPE_CODE	0	0	Medium	Reports Well Assembly Components where the component type is null
CD_ASSEMBLY_COMP	CONNECTION_TYPE	0	0	Medium	Reports Casing Well Assembly Components with a null connection type
CD_ASSEMBLY_COMP	GRADE	0	0	Medium	Reports Well Assembly Components where the section code is DP or HW and the grade is null
CD_ASSEMBLY_COMP	GRADE	0	0	Medium	Reports Actual Well Casing Assembly Components where the grade is null
CD_ASSEMBLY_COMP	ID_BODY	0	0	Medium	Reports Well Assembly Components where the ID body is greater than or equal to the Pipe OD
CD_ASSEMBLY_COMP	ID_BODY	0	0	Medium	Reports Well Assembly Components where the id body is null
CD_ASSEMBLY_COMP	ID_DRIFT	0	0	Medium	Reports Casing Well Assembly Components where the ID Drift is greater than the ID body
CD_ASSEMBLY_COMP	CATALOG_KEY_DESC	0	0	Medium	Reports Well Assembly Components where the item description is null
CD_ASSEMBLY_COMP	JOINTS	0	0	Medium	Reports Well Assembly Components where the joints are null
CD_ASSEMBLY_COMP	LENGTH	0	0	Medium	Reports Well Assembly Components where the length is null
CD_ASSEMBLY_COMP	MD_BASE	0	0	Medium	Reports Well Assembly Components where the md base is null
CD_ASSEMBLY_COMP	MANUFACTURER	0	0	Medium	Reports Well Assembly Components with a bit section type code and a null manufacturer
CD_ASSEMBLY_COMP	MANUFACTURER	0	0	Medium	Reports Actual Well Casing Assemblies where manufacturer is null

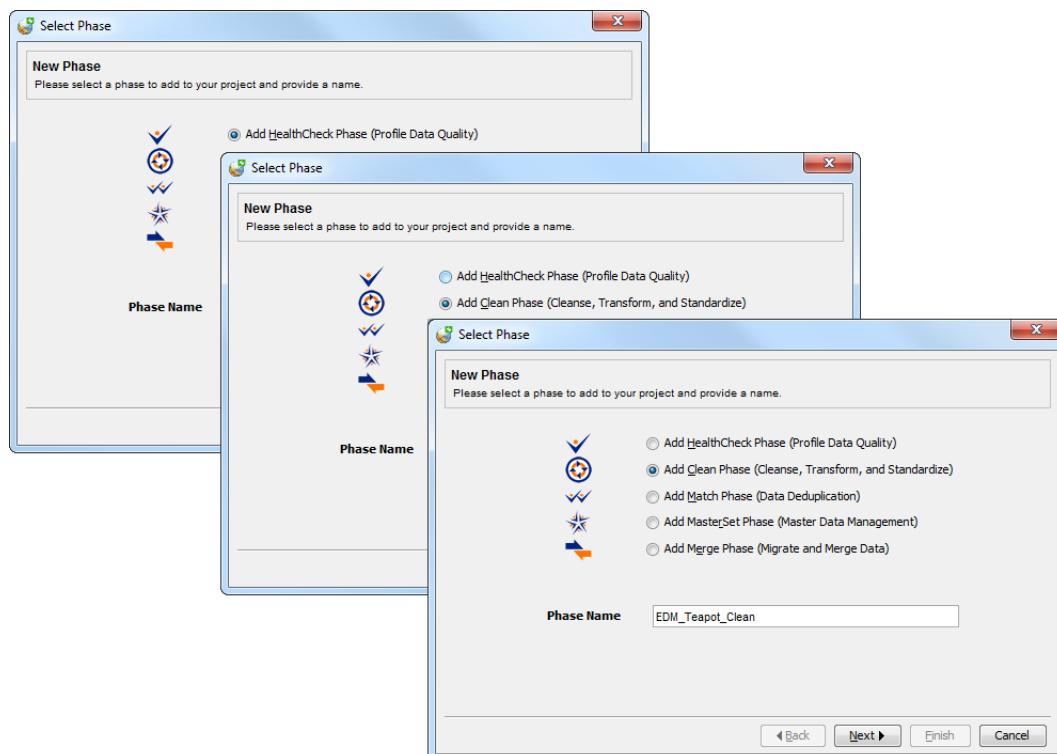
16. Select **File > Exit** to close the **Basic View Frame** window.

# Resolving Data Quality Issues using the Clean Phase

## *Adding a Clean Phase*

To add a Clean Phase:

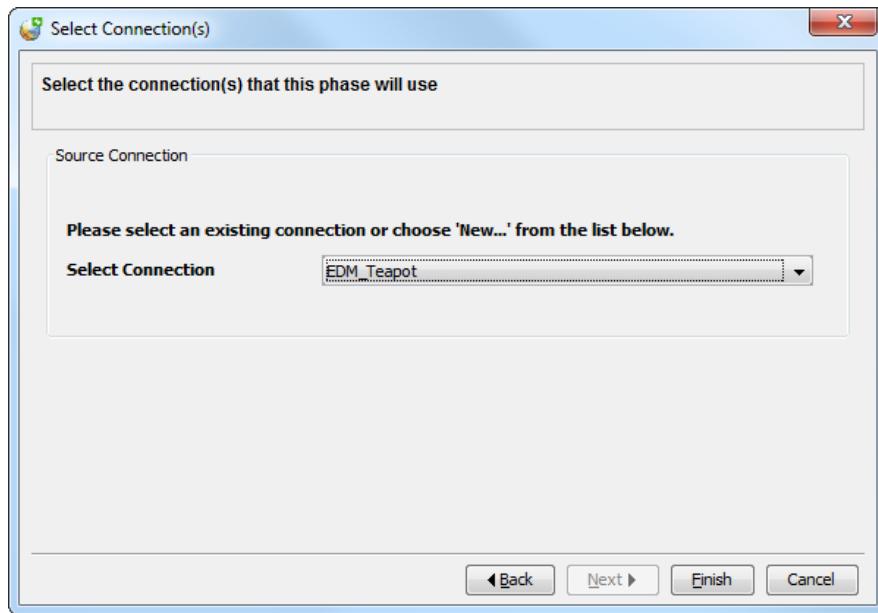
1. Click the **Add New Phase** button on the Project Toolbar. The **Select Phase** window appears with the **Add HealthCheck Phase (Report on Data Quality Profiling)** option selected by default.



2. Select the **Add Clean Phase (Cleanse, Transform, and Standardize)** option.
3. Enter **EDM\_Teapot\_Clean** in the **Phase Name** field.

4. Click **Next** to continue.

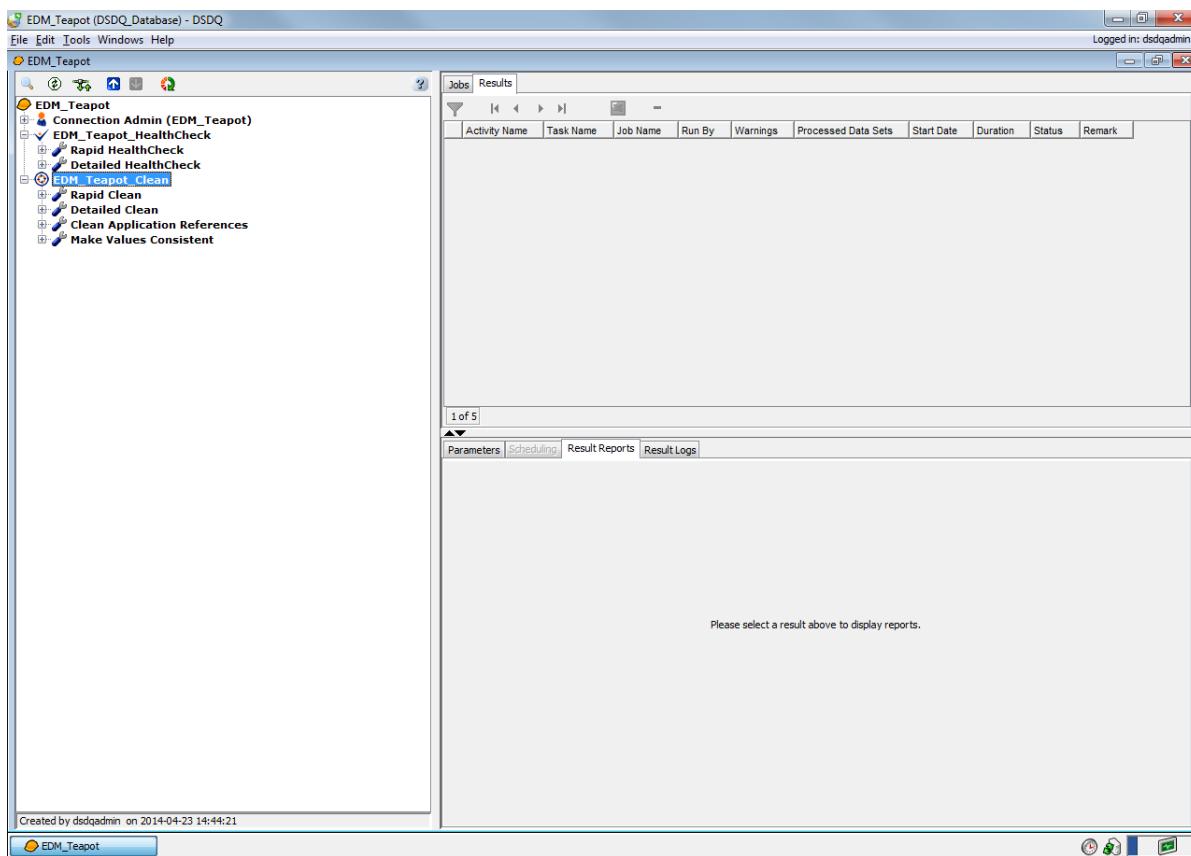
The **Select Connection(s)** window appears.



5. Select **EDM\_Teapot** from the **Select Connection** drop-down list.

## 6. Click **Finish**.

The **Clean** Phase is created and displays in the DecisionSpace Data Quality Project window.



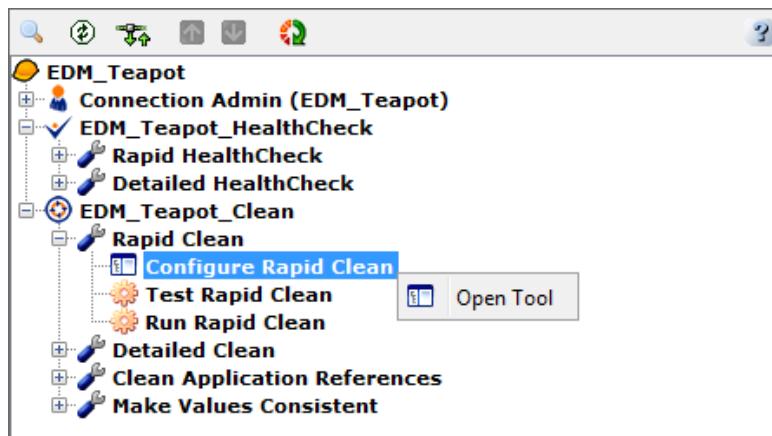
## Rapid Clean Activity

The **Rapid Clean** Activity cleans out data issues in columns of selected submodel tables.

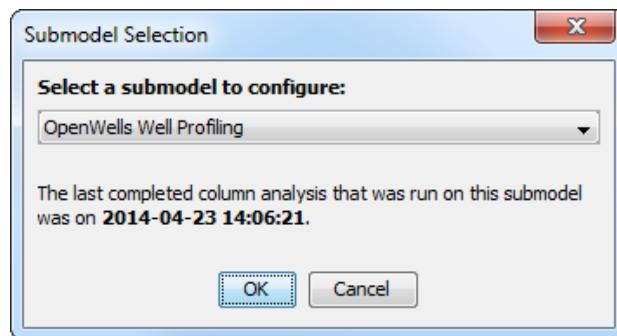
### Exercise: Configuring the Rapid Clean Tool

Prior to running the **Configure Rapid Clean** Tool, ensure that the **Run Column Analysis on Modeled Table** Task has been run. To configure the Rapid Clean Tool:

1. Click to expand the **Rapid Clean** Activity.
2. Double-click the **Configure Rapid Clean** Tool or right-click the **Configure Rapid Clean** Tool and select **Open Tool** from the pop-up menu



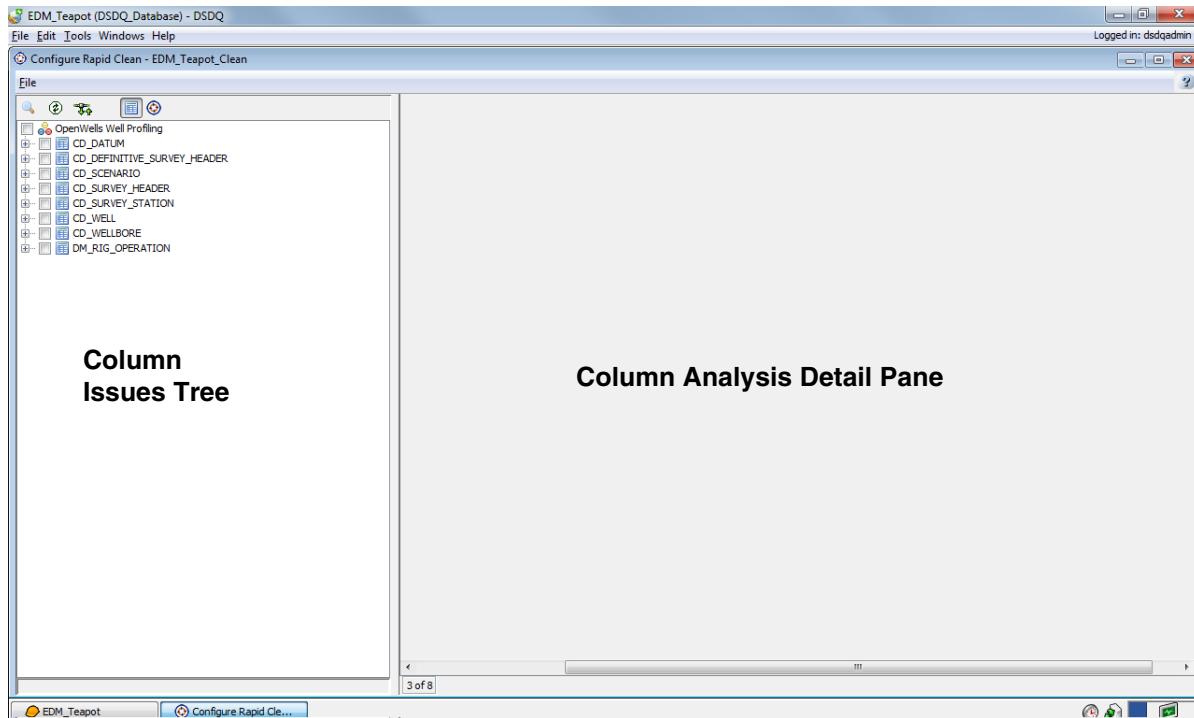
The **Submodel Selection** dialog box appears.



3. Select **OpenWells Well Profiling** from the **Select a submodel to configure** drop-down list.

4. Click **OK** to continue.

The **Configure Rapid Clean - EDM\_Teapot\_Clean** window appears.



5. Click  to expand the **CD\_WELL** table.

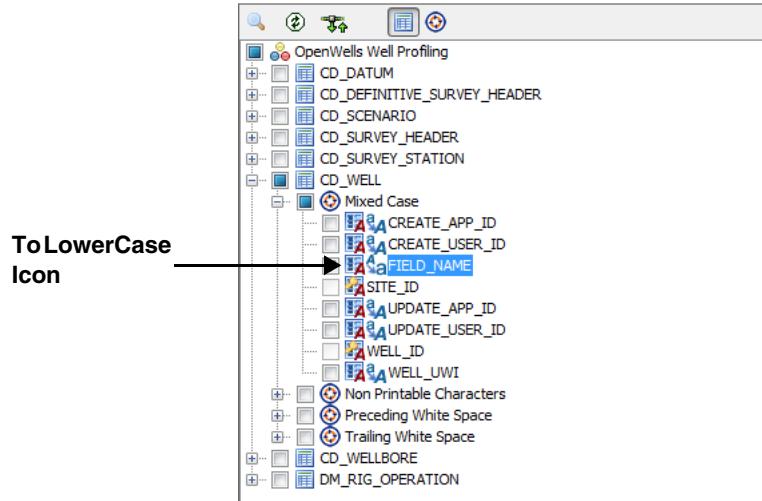
6. Click  to expand the **Mixed Case** issues.

7. Click the **FIELD\_NAME** column.

Any issue(s) for the **CD\_WELL** table highlight in the **Column Analysis Details Pane**.

Data Type	Rows	# Not Null	% Populated	# Unique	# Mixed Case	# NPC	# PWS	# TWS	# DWS	Minimum Value	Maximum Value
STRING	6	6	100	1	6	0	0	0	0	0 OpenWells	OpenWells
STRING	6	6	100	2	6	0	0	0	0	0 RDIVxvdx122(ed...)	RDIVxvdx139
STRING	6	6	100	3	1	0	2	0	0	0 TEAPOT.DOME	TEAPOT.DOME
STRING	6	6	100	1	6	0	0	0	0	0 EDH0H20H6Z	EDH0H20H6Z
STRING	6	6	100	1	6	0	0	0	0	0 OpenWells	OpenWells
STRING	6	6	100	1	6	0	0	0	0	0 ADInmughal(edm)	ADInmughal(edm)
STRING	6	6	100	6	6	0	0	0	0	0 1LKx2Q9t6C	wk05NAAb6
STRING	6	6	100	6	1	0	0	0	0	0 490250633800	Site #3

8. Select the check box adjacent to the **FIELD\_NAME** column.
9. Click the **To Uppercase**  icon adjacent to the **FIELD\_NAME** column.  
The **To Uppercase**  icon changes to the **To LowerCase**  icon.

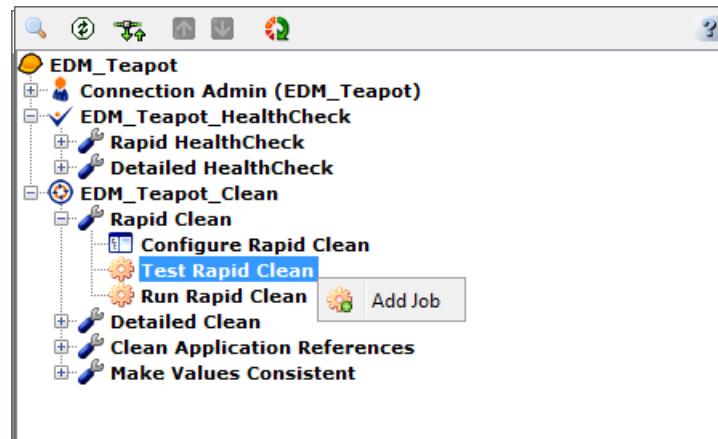


10. Select the **Preceding White Spaces** requirement.
11. Select **File > Exit** to close the **Configure Rapid Clean** window.

### Exercise: Running the Test Rapid Clean Task

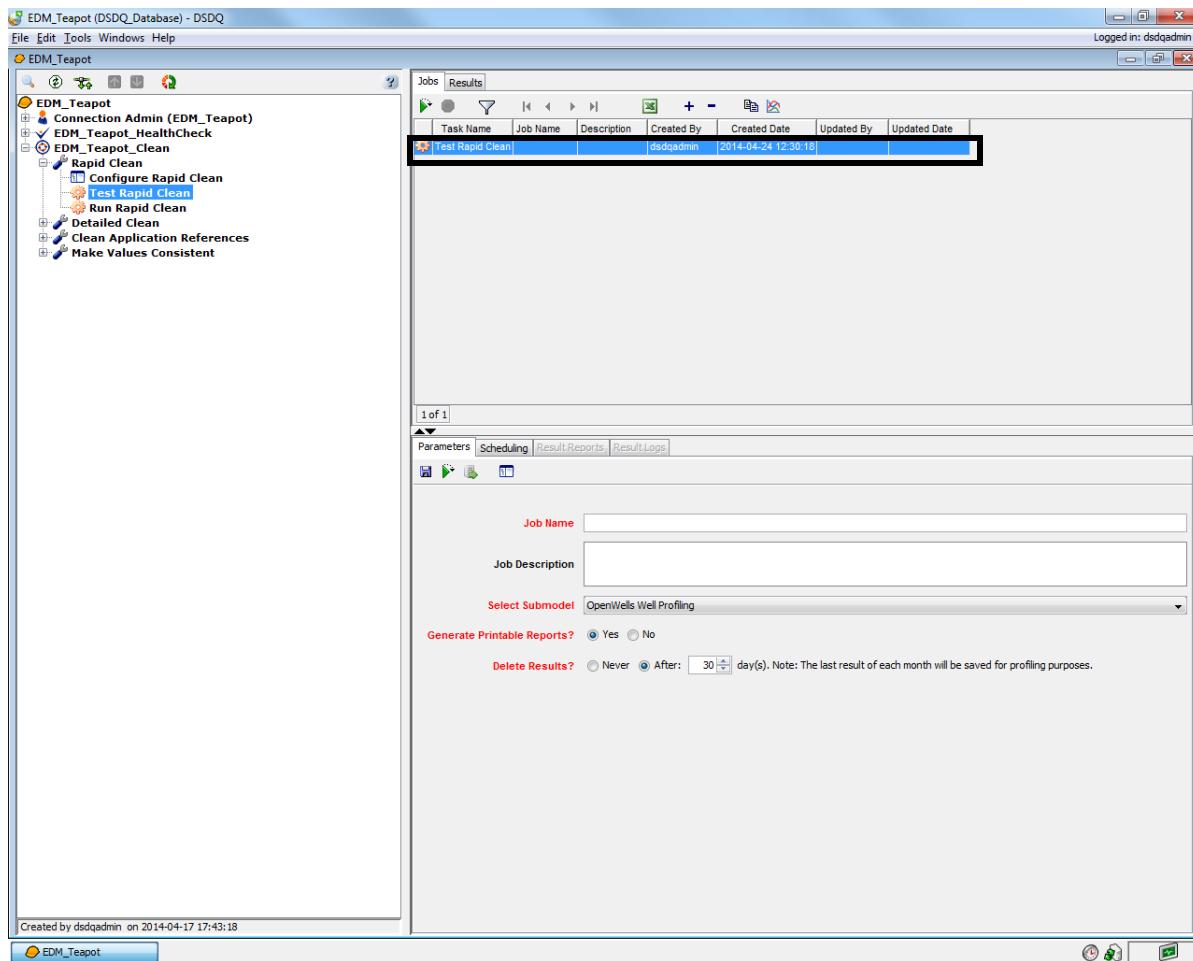
After issues have been selected to be cleaned using the **Configure Rapid Clean Tool**, the **Test Rapid Clean Task** is run to make sure that the expected results are seen before running the **Run Rapid Clean Task** to fix the entire dataset in the submodel. To run the **Test Rapid Clean Task**:

1. Double-click the **Test Rapid Clean** Task or right-click the **Test Rapid Clean** Task and select **Add Job** from the pop-up menu.



A new job is initiated and it displays on the **Job and Results**

**Information Pane** on the right side of the DecisionSpace Data Quality Project window.



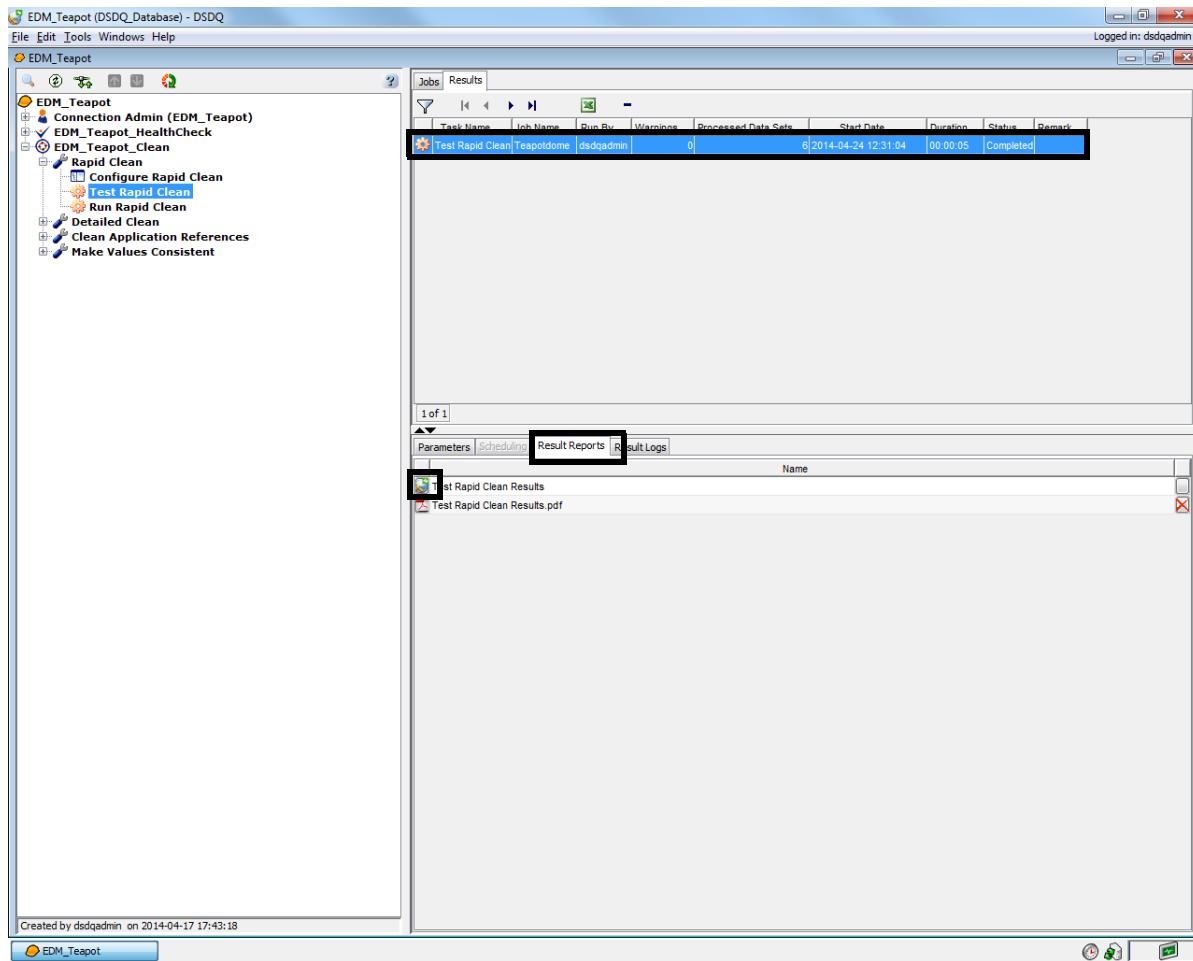
2. Enter **Teapotdome** in the **Job Name** field.
3. Enter **EDM\_Teapot FIELD\_NAME Name Conversion** in the **Job Description** field.
4. Select **OpenWells Well Profiling** from the **Select Submodel** drop-down list.
5. Select the **Yes** option for **Generate Printable Reports**.
6. Select the **After** option for **Delete Results**. Leave the number of days as **30**.
7. Click to save changes in the **Parameters** tab.

8. Click  to run the job.

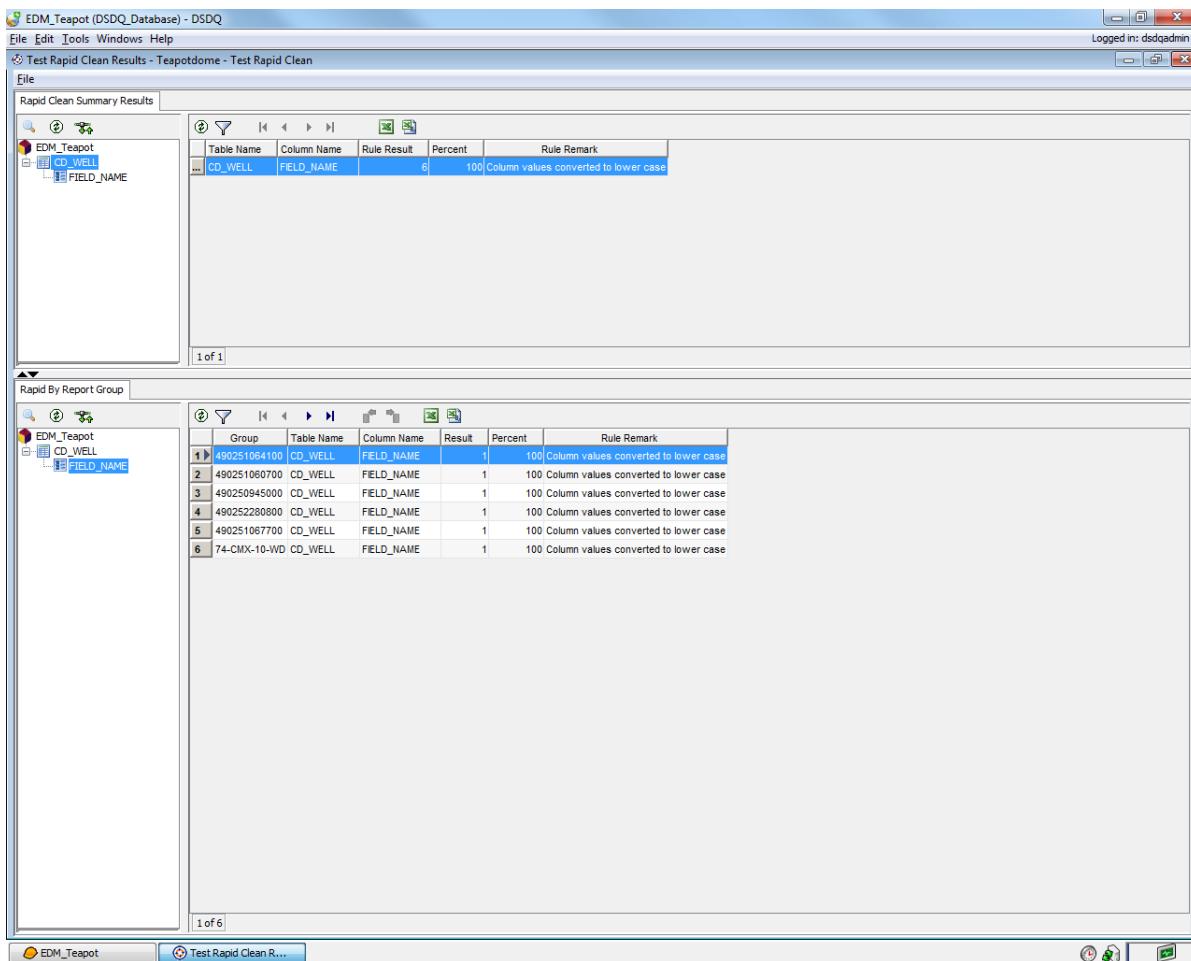
The **Test Rapid Clean** Task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

9. Select the **Results** tab.

The **Job and Results Listing Pane** displays the list of results.



10. Click the **Open Basic View Frame**  button on the **Result Reports** tab to display the **Test Rapid Clean** Task results in the **Basic View Frame** window.

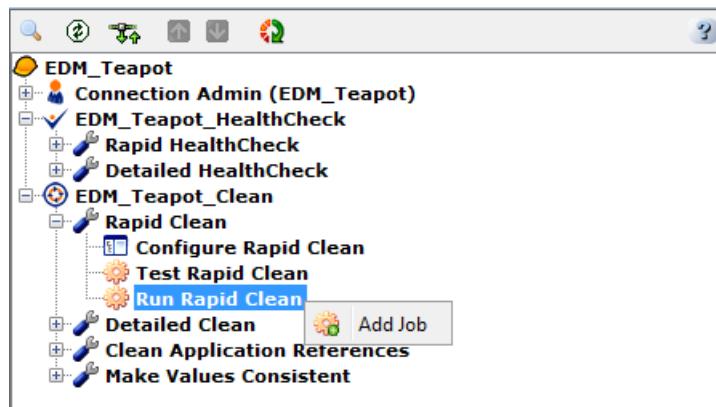


11. Select **File > Exit** to close the **Basic View Frame** window.

### Exercise: Running the Rapid Clean Task

The **Run Rapid Clean** Task fixes the issues that were selected in the **Configure Rapid Clean** Tool for the specific submodel. To run the Rapid Clean task:

1. Double-click the **Run Rapid Clean** Task or right-click the **Run Rapid Clean** Task and select **Add Job** from the pop-up menu.



A new job is initiated and it displays on the **Job and Results Information Pane**.

The screenshot shows the EDM Teapot interface with the 'Job and Results' pane open. In the 'Jobs' tab, there is a single row of data:

Task Name	Job Name	Description	Created By	Created Date	Updated By	Updated Date
Run Rapid Clean			dsdqadmin	2014-04-23 16:43:33		

Below the table, there are several configuration tabs: Parameters, Scheduling, Result Reports, and Result Logs. The 'Parameters' tab is active, showing fields for 'Job Name' (set to 'Teapotdome'), 'Job Description' (set to 'EDM\_Teapot Case Convert'), 'Select Submodel' (set to 'OpenWells Well Profiling'), and other options like 'Generate Printable Reports' (Yes), 'Delete Results?' (After 30 days), 'Data Change Action' (Immediately Apply Changes to the Database), and 'Enable Audit Logging' (No).

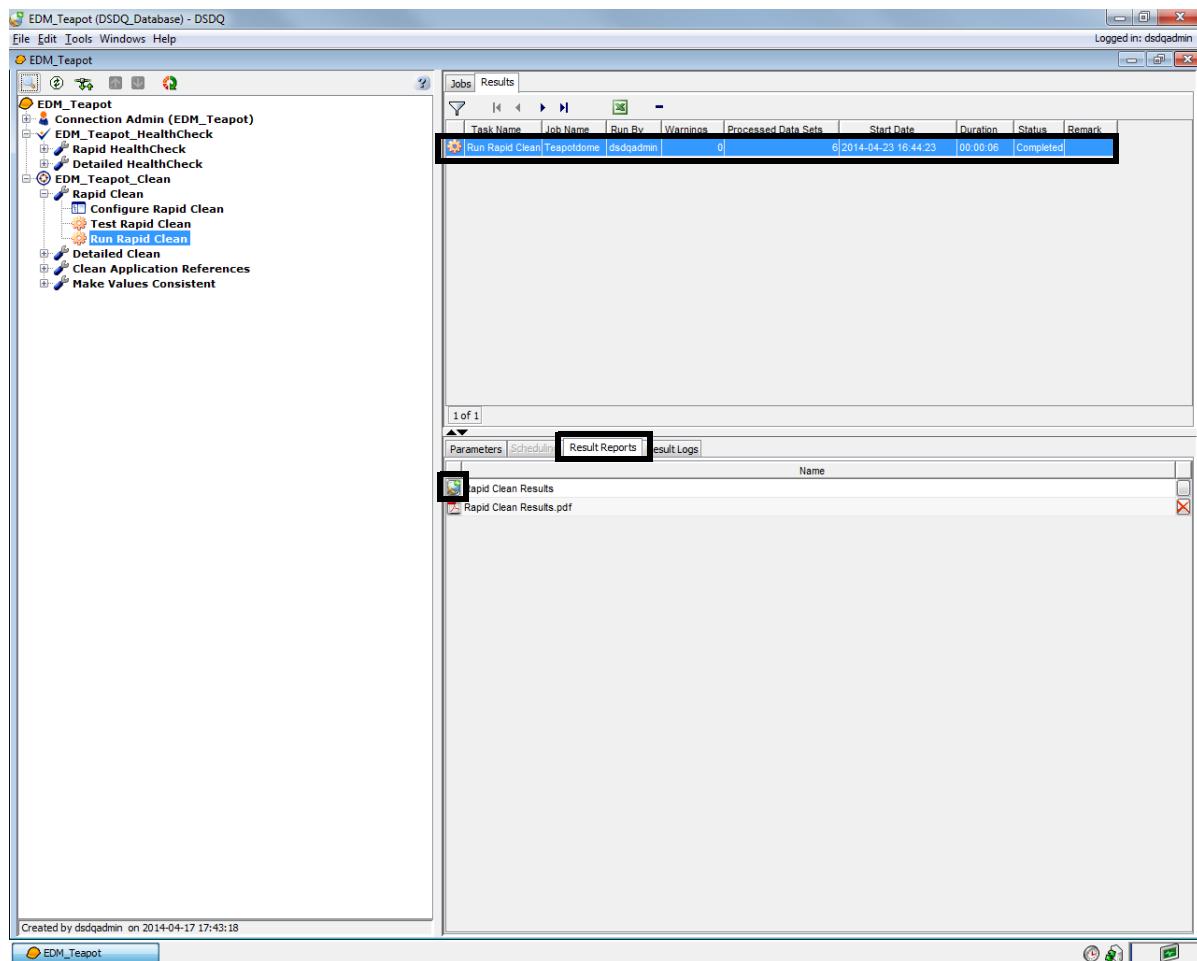
2. Enter **Teapotdome** in the **Job Name** field.
3. Enter **EDM\_Teapot Case Convert** in the **Job Description** field.

4. Select **OpenWells Well Profiling** from the **Select Submodel** drop-down list.
5. Select the **Yes** option for **Generate Printable Reports**.
6. Select the **After** option for **Delete Results**. Leave the number of days as **30**.
7. Select the **Immediately Apply Changes to the Database** option for **Data Change Action**.
8. Select the **No** option for **Enable Audit Logging**.
9. Click  to save changes in the **Parameters** tab.
10. Click  to run the job.

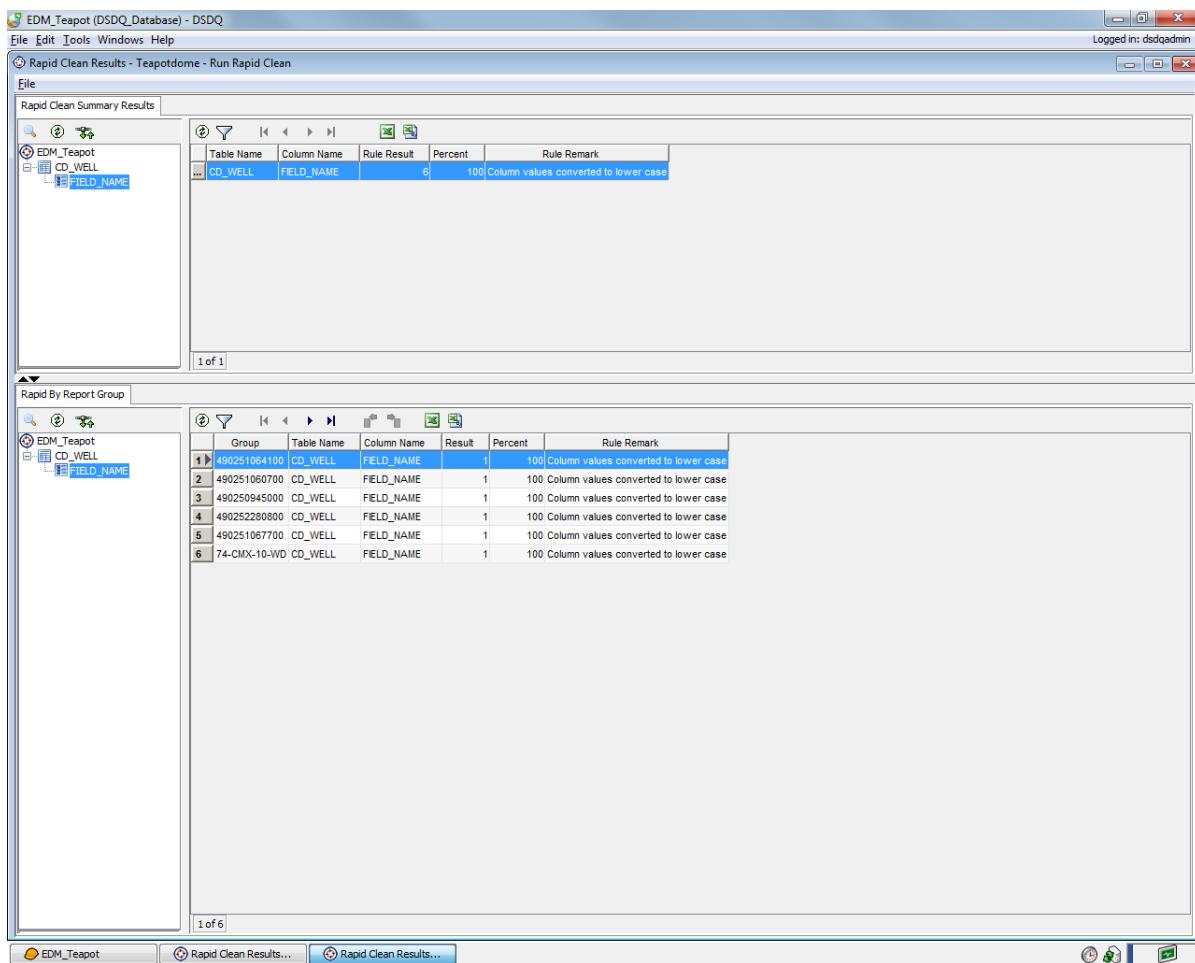
The **Run Rapid Clean Task** runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

## 11. Select the **Results** tab.

The **Job and Results Listing Pane** displays a list of results.



12. Click the **Open Basic View Frame**  button on the **Result Reports** tab to display the **Run Rapid Clean** Task results in the **Basic View Frame** window.



13. Select **File > Exit** to close the **Basic View Frame** window.

## Detailed Clean Activity

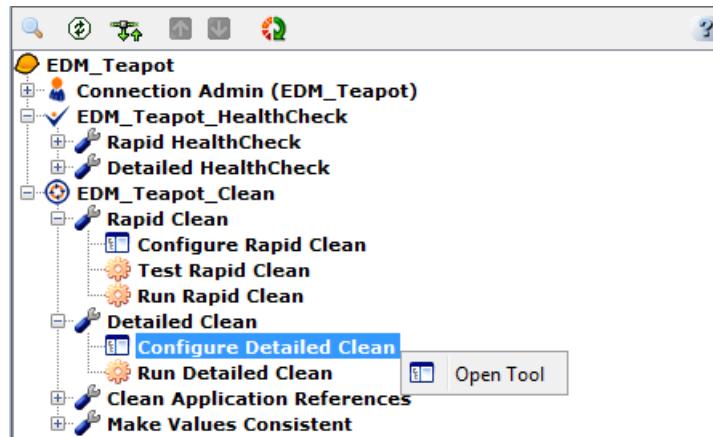
The **Detailed Clean** Activity helps in assigning columns to the clean requirements and testing service levels.

### Exercise: Configuring the Detailed Clean Tool

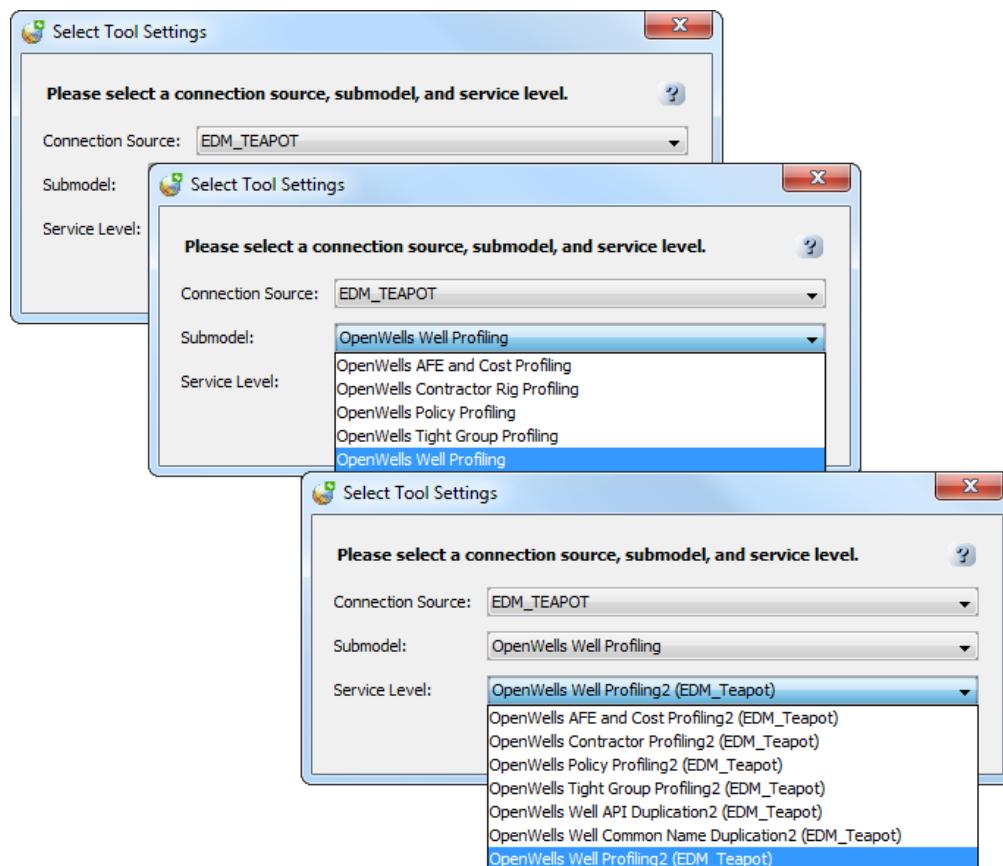
The **Configure Detailed Clean** Tool is used to configure service levels for testing prior to running the **Run Detailed Clean** Task. You can select which requirements in the service level to enable/disable, and when testing a service level, what subset of the total data to use. A

service level containing clean requirements must exist prior to running the **Configure Detailed Clean** Tool.

1. Double-click the **Configure Detailed Clean** Tool or right-click the **Configure Detailed Clean** Tool and select **Open Tool** from the pop-up menu.

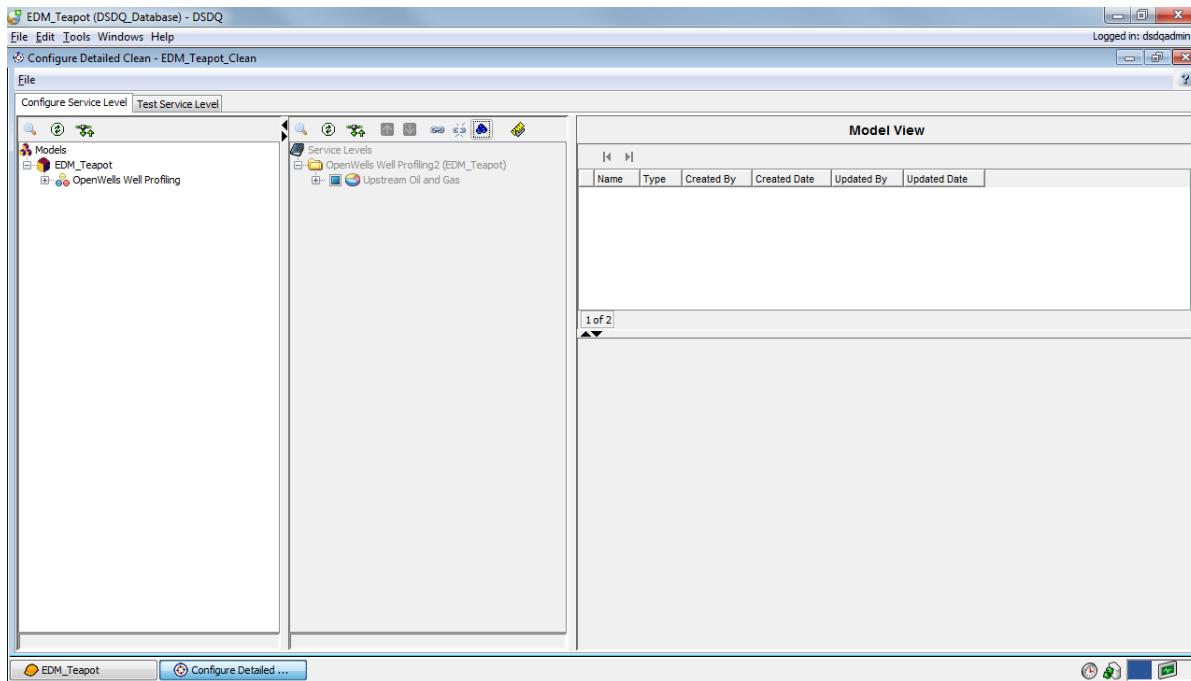


The **Select Tool Settings** window appears.

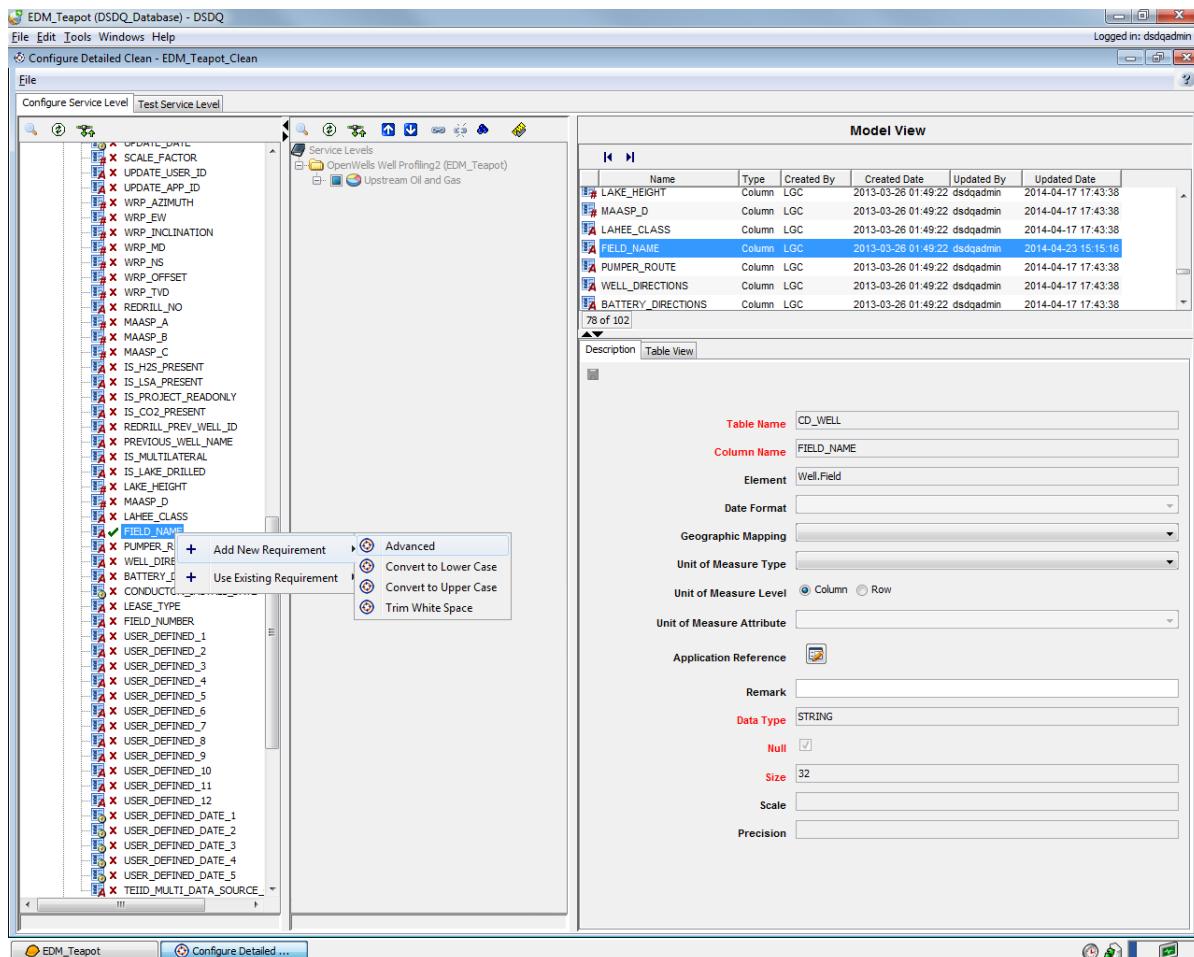


2. The **Connection Source** drop-down list is set to **EDM\_Teapot** by default.
3. Select **OpenWells Well Profiling** from the **Submodel** drop-down list.
4. Select **OpenWells Well Profiling2 (EDM\_Teapot)** from the **Service Level** drop-down list.
5. Click **OK**.

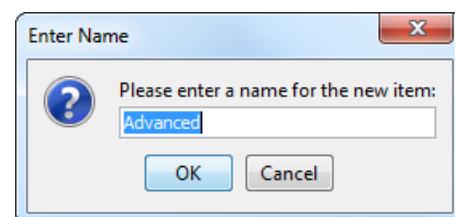
The **Configure Detailed Clean - EDM\_Teapot\_Clean** window appears.



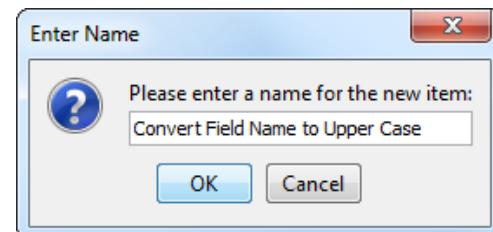
6. Click to expand the **OpenWells Well Profiling** submodel in the Data Model Tree.
7. Click to expand the **CD\_WELL** table in the Data Model Tree.
8. Right-click the **FIELD\_NAME** column in the Data Model Tree and select **Add New Requirement > Advanced** from the pop-up menu.



The Enter Name dialog box appears.



9. Enter **Convert Field Name to Upper Case** in the **Please enter a name for the new item** dialog box.

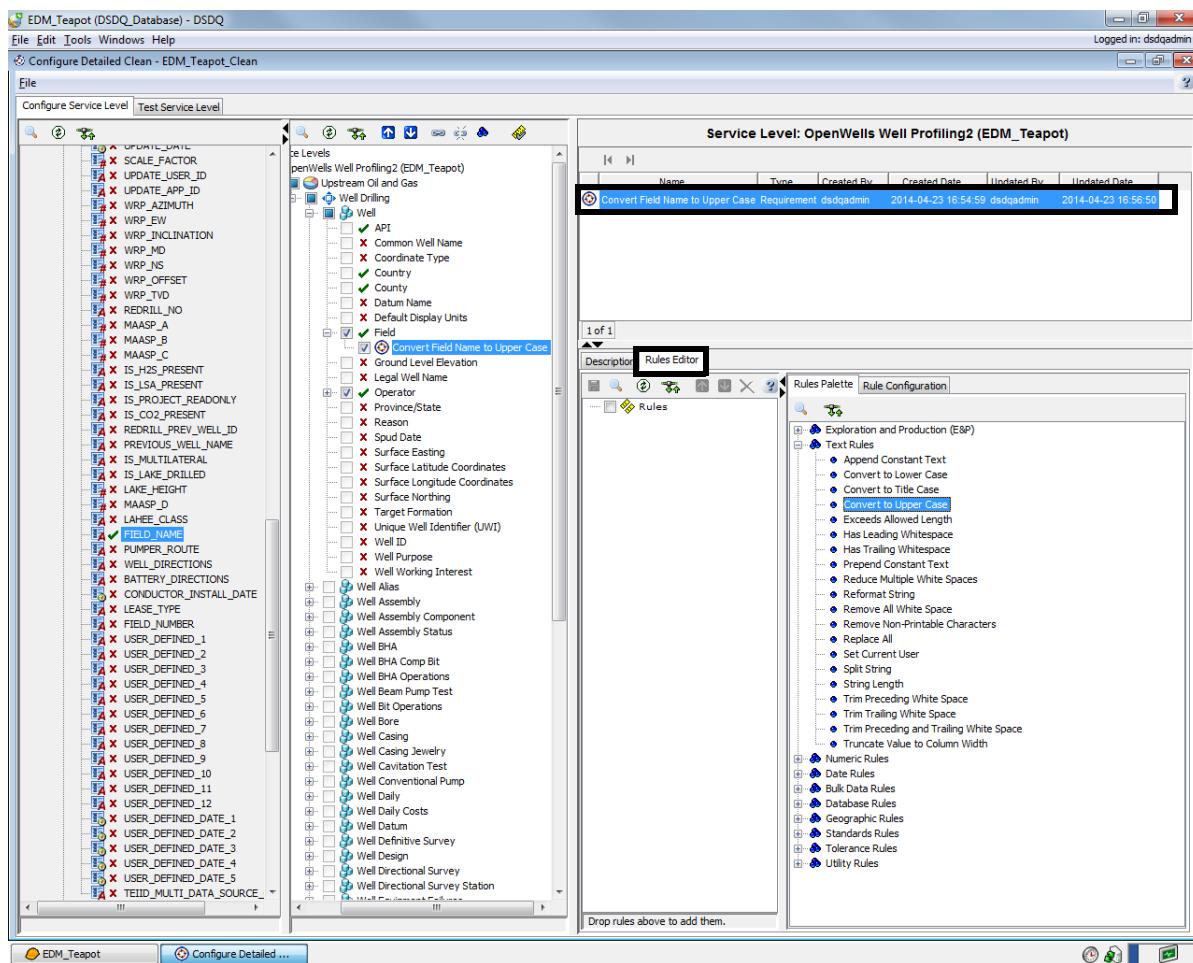


10. Click **OK**.

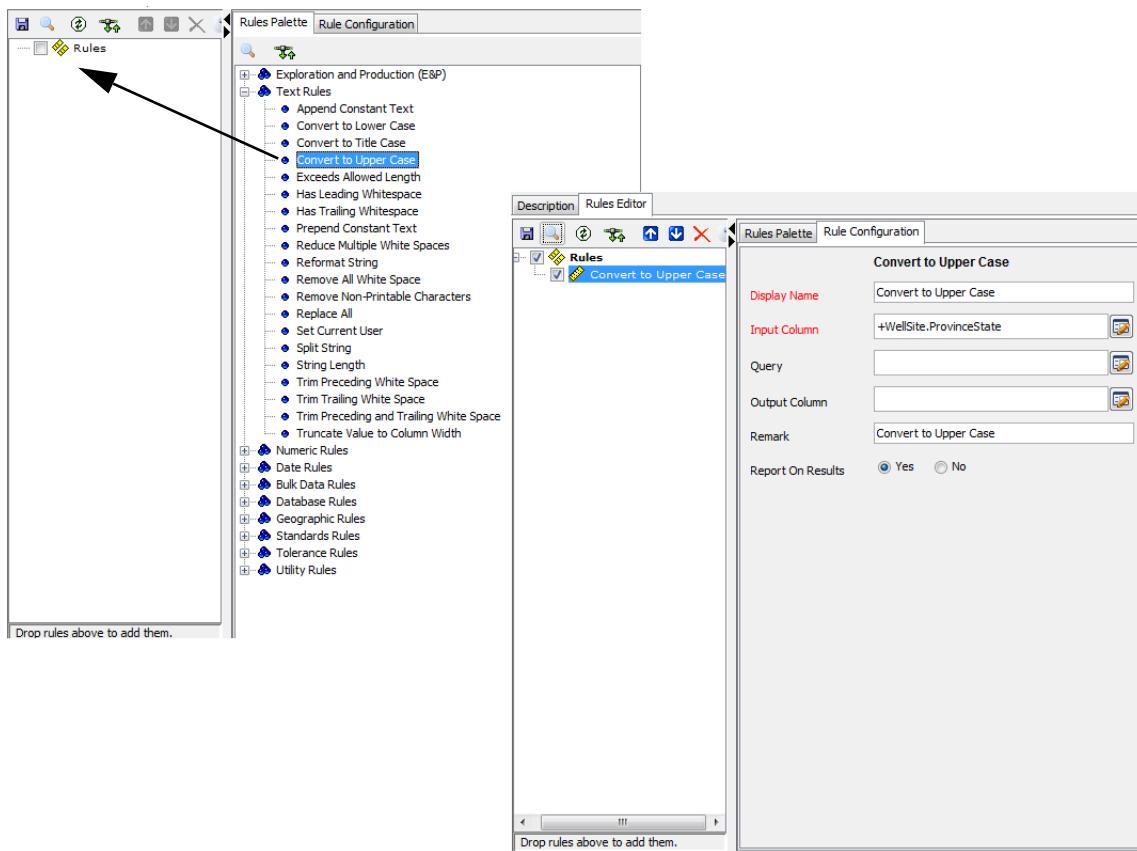
The **Convert Field Name to Upper Case** requirement is added to the **FIELD\_NAME** column and displays in the **Model View Pane**.

11. Select the **Rules Editor** tab adjacent to the **Description** tab.

12. Click to expand the **Text Rules** in the **Rule Palette** tab.



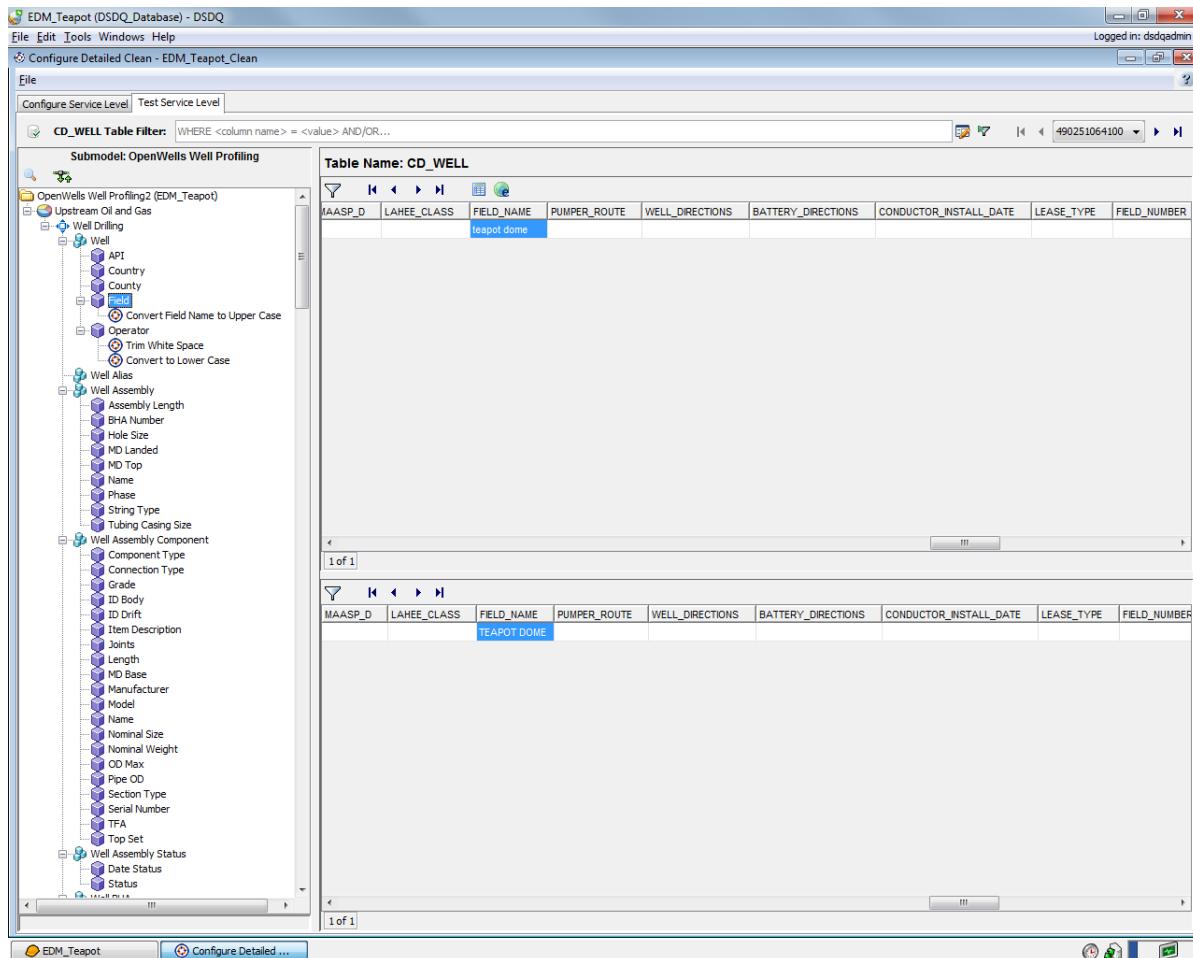
13. Drag and drop the **Convert to Upper Case** onto the **Rules** area.



14. Click to save changes in the **Rule Editor** tab.

15. Select the **Test Service Level** tab.

The test is automatically executed for the first record of the test data subset.



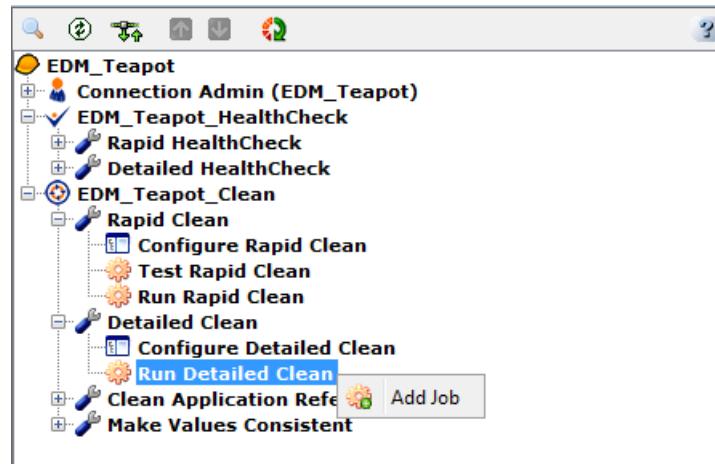
By looking at the columns that have been changed and temporary columns, you can verify that the behavior of the service level is correct prior to running the **Run Detailed Clean Task**.

16. Click the **Next Data Set** button to test the next record.
17. Repeat step **16** to test all records.
18. Select **File > Exit** to close the **Configure Detailed Clean** window.

### Exercise: Running the Detailed Clean Task

The **Run Detailed Clean** Task performs cleansing of entire data and updates the actual data. To run the Detailed Clean task:

1. Double-click the **Run Detailed Clean** Task on the DecisionSpace Data Quality Tree or right-click the **Run Detailed Clean** Task and select **Add Job** from the pop-up menu.



A new job is initiated and it displays on the **Jobs and Results Listing Pane**.

The screenshot shows the DecisionSpace interface with the 'Jobs' tab selected in the 'Jobs and Results' pane. A single job entry is listed:

Task Name	Job Name	Description	Created By	Created Date	Updated By	Updated Date
Run Detailed Clean	dsdqadmin		dsdqadmin	2014-04-23 16:59:38		

The main pane shows the 'EDM\_Teapot' tree structure. The 'Job Details' pane on the right contains the following configuration fields:

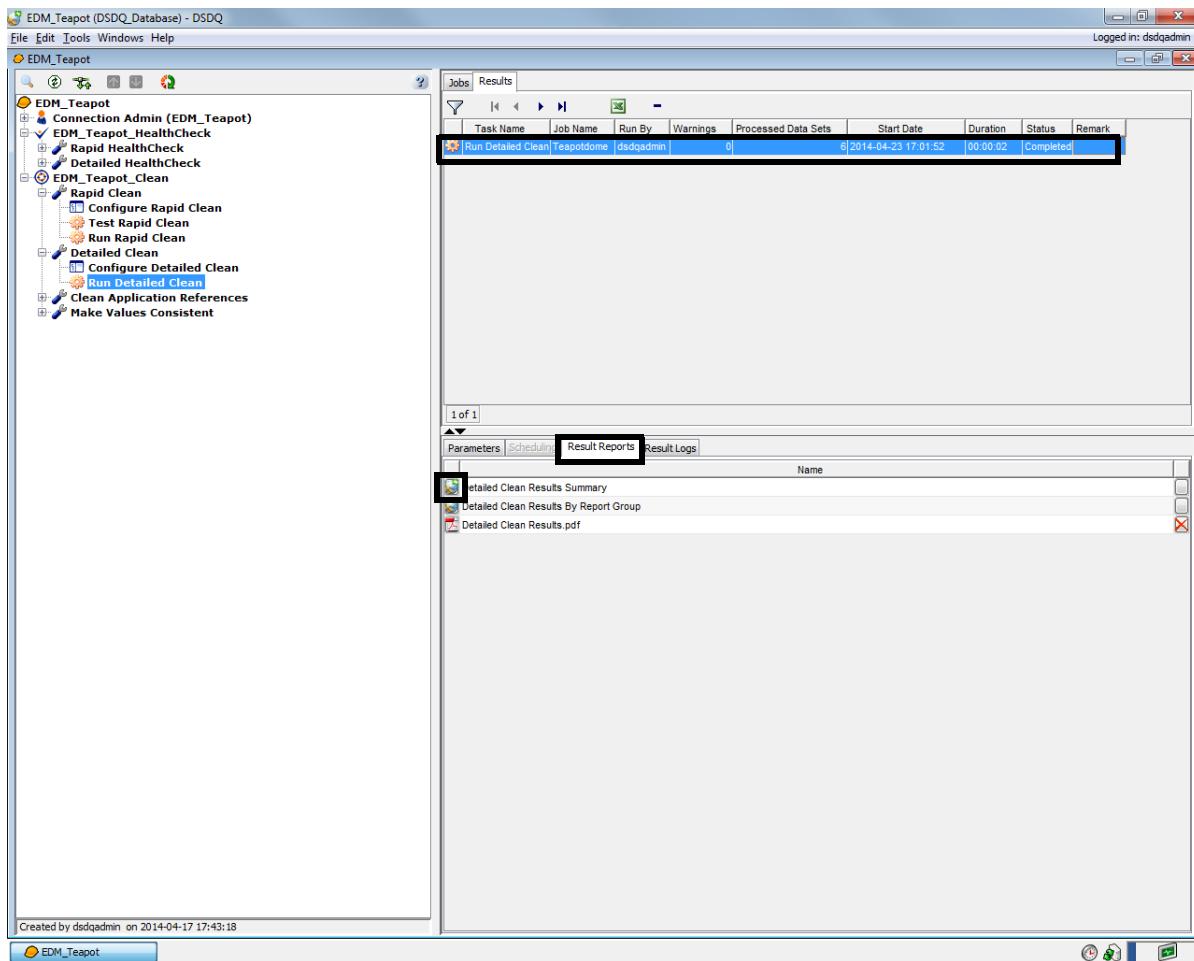
- Job Name:**
- Job Description:**
- Service Level:** OpenWells Well Profiling2 (EDM\_Teapot)
- Select a Submodel:** OpenWells Well Profiling
- Filter Base Data Set:**
- Summarize Results by:**  Requirement (least detail)  Data Set  Data Row (most detail)
- Generate Printable Reports?**  Yes  No
- Delete Results?**  Never  After: 30 day(s). Note: The last result of each month will be saved for profiling purposes.
- Data Change Action:**  Immediately Apply Changes to the Database  Only Write the Changes to an SQL Script
- Enable Audit Logging:**  Yes  No

2. Enter **Teapotdome** in the **Job Name** field.

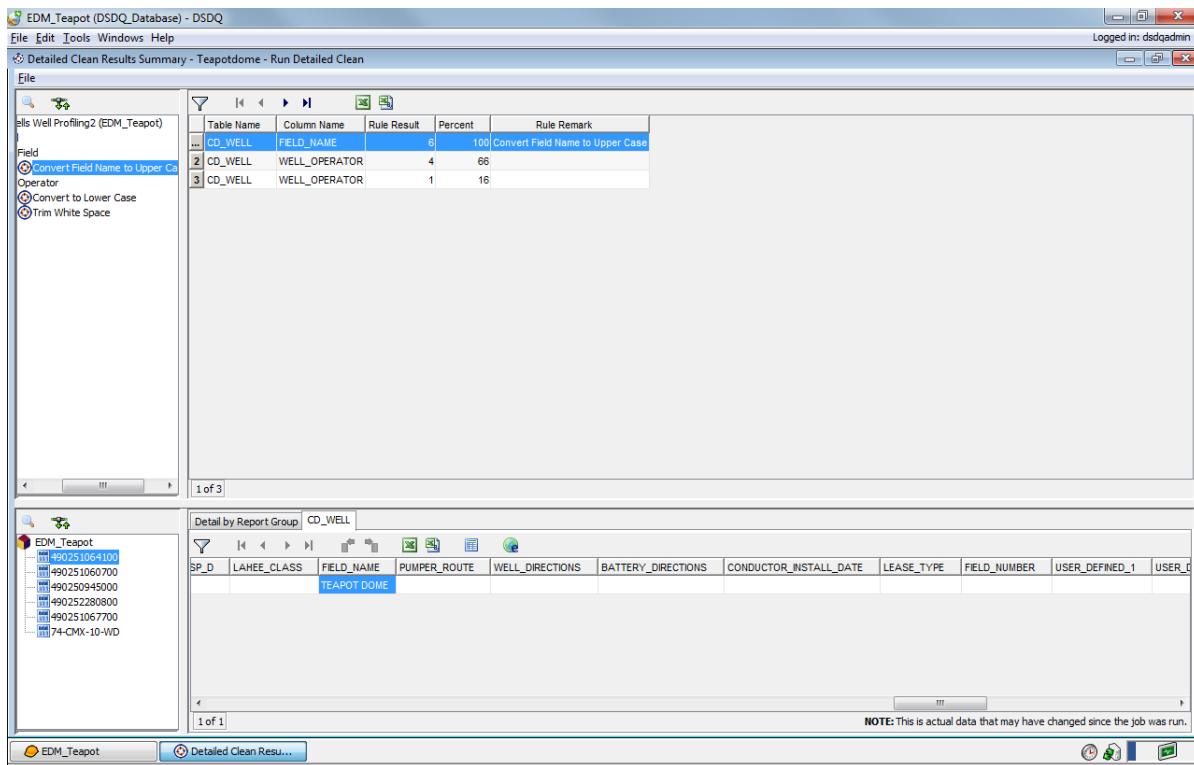
3. Enter **Run Detailed Clean for EDM\_Teapot State to Upper Case** in the **Job Description** field.
4. Select **OpenWells Well Profiling2 (EDM\_Teapot)** from the **Service Level** drop-down list.
5. Select **OpenWells Well Profiling** from the **Select a Submodel** drop-down list.
6. Do not set the filter for **Filter Base Data Set**.
7. Select the **Data Row (most detail)** option for **Summarize Results by**.
8. Select the **Yes** option for **Generate Printable Reports**.
9. Select the **After** option for **Delete Results**. Leave the number of days as **30**.
10. Select the **Immediately Apply Changes to the Database** option for **Data Change Action**.
11. Select the **No** option for **Enable Audit Logging**.
12. Click  to save changes in the **Parameters** tab.
13. Click  to run the job.  
The **Run Detailed Clean** Task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

**14. Select the Results tab.**

The **Jobs and Results Listing Pane** displays a list of results.



15. Click the **Open Basic View Frame**  button on the **Result Reports** tab to display the **Run Detailed Clean** Task results in the **Basic View Frame** window.

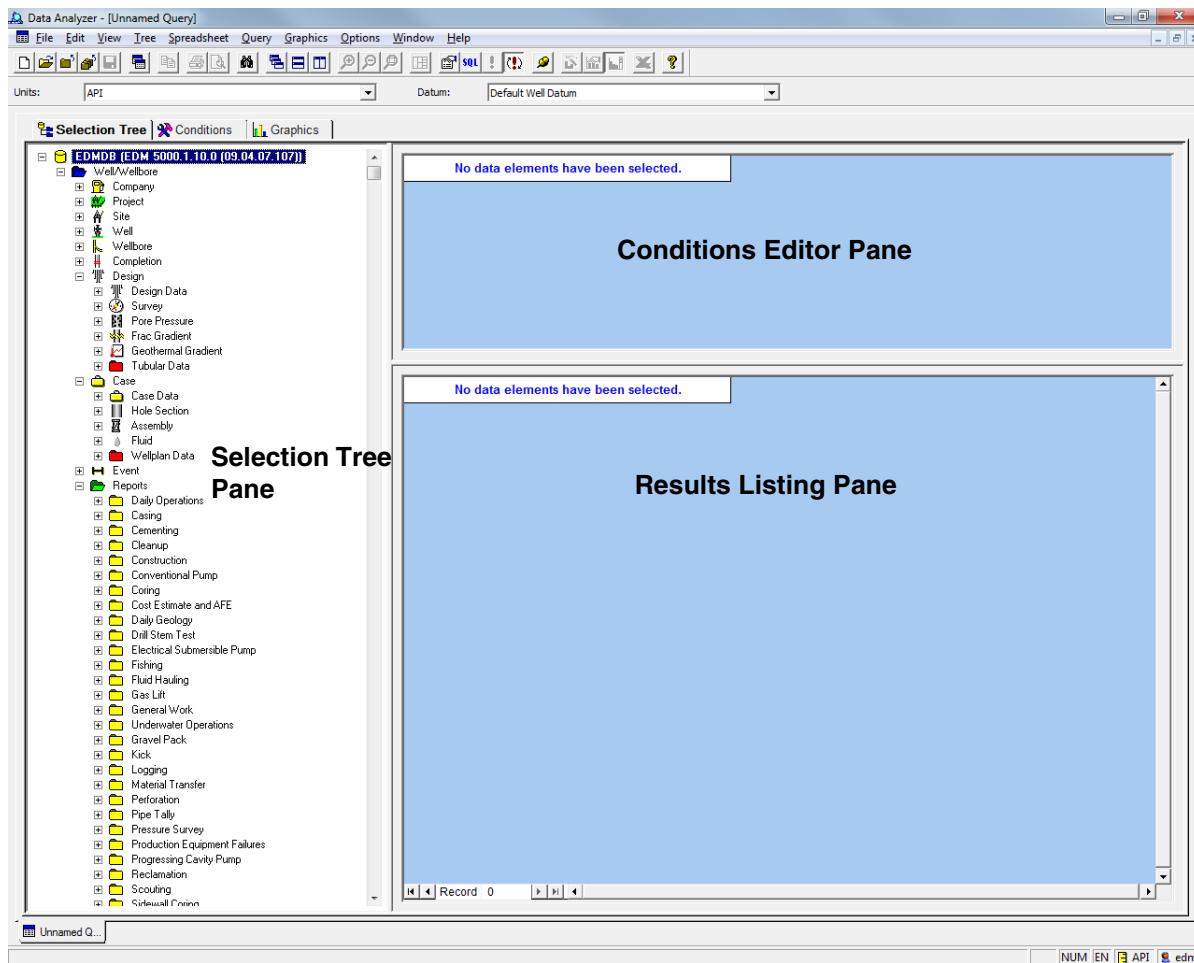


16. Select **File > Exit** to close the **Basic View Frame** window.

## Validating Corrected Data in the EDM Data Source

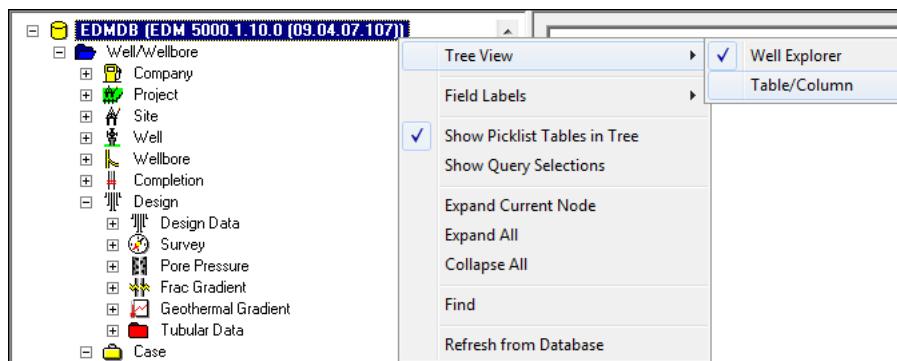
To validate information from the Data Analyzer of the EDM data source:

1. Double-click the Data Analyzer desktop icon to launch the Data Analyzer application.
- The **Data Analyzer** window appears.

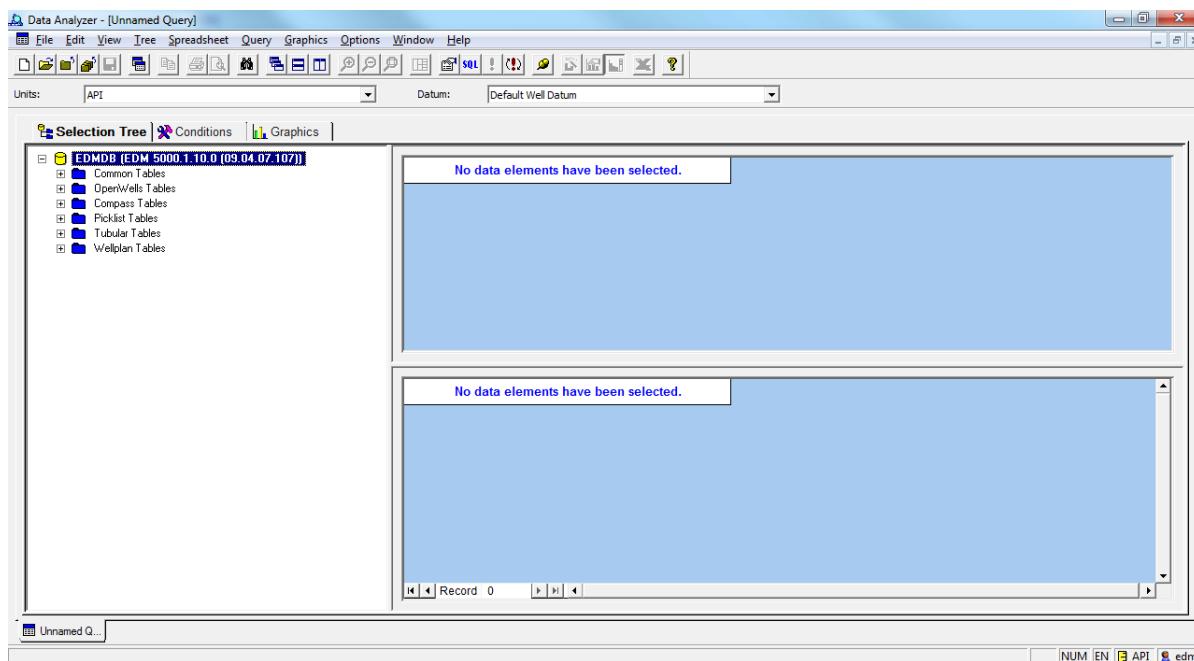


2. Right-click the **EDMDB (EDM 5000.1.10.0 (09.04.07.107))** data source in the **Selection Tree Pane** and select **Tree View > Table/**

**Column** from the pop-up menu.



A list of tables displays in the **Selection Tree Pane**.



3. Click to expand the **Common Table** from the **Selection Tree Pane**.
4. Click to expand the **CD\_WELL** table.
5. Select the **Field name** element.  
The selected elements display on the **Conditions Editor Pane**.

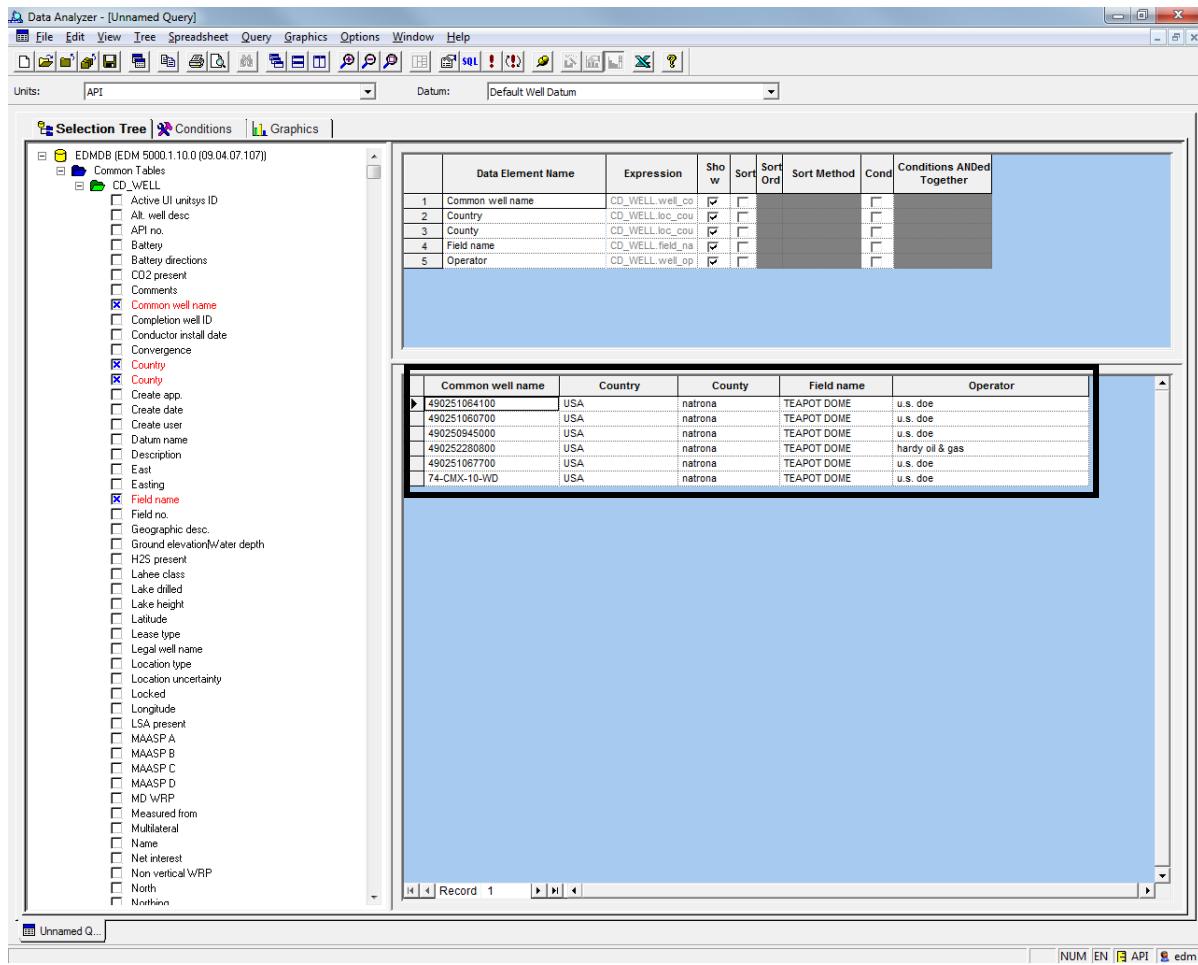
6. Repeat step 5 to select other elements.

The screenshot shows the Data Analyzer software interface with the following details:

- Menu Bar:** File, Edit, View, Tree, Spreadsheet, Query, Graphics, Options, Window, Help.
- Toolbar:** Includes icons for Open, Save, Print, Copy, Paste, Find, Replace, and various graphics tools.
- Units:** API
- Datum:** Default Well Datum
- Selection Tree:** Shows the EDMDB (EDM 5000.1.10.0 (09.04.07.107)) structure. Under the CD\_WELL folder, several items are checked, including Common well name, County, and Field name.
- Grid View:** A table showing selected data elements:
 

	Data Element Name	Expression	Show	Sort	Sort Ord	Sort Method	Cond	Conditions ANDed Together
1	Common well name	CD_WELL.well_cd	<input checked="" type="checkbox"/>					
2	Country	CD_WELL.lncl_cou	<input checked="" type="checkbox"/>					
3	County	CD_WELL.loc_cou	<input checked="" type="checkbox"/>					
4	Field name	CD_WELL.field_na	<input checked="" type="checkbox"/>					
5	Operator	CD_WELL.well_op	<input checked="" type="checkbox"/>					
- Message Bar:** Query needs to be Run.
- Status Bar:** Record 1
- Bottom Navigation:** NUM | EN | API | edm

7. Click the **Run Query**  icon from the icon bar.  
The results display on the **Results Listing Pane**.



The screenshot shows the Data Analyzer application interface. The top menu bar includes File, Edit, View, Tree, Spreadsheet, Query, Graphics, Options, Window, and Help. The toolbar contains various icons for file operations, search, and analysis. The left pane is the Selection Tree, showing a hierarchy of tables and fields under 'EDMDB (EDM 5000.1.10.0 (09.04.07.107))' and 'CD\_WELL'. A specific field, 'Common well name', is selected and highlighted in red. The middle pane displays the Query Definition, which includes a table with columns: Data Element Name, Expression, Show, Sort, Sort Order, Sort Method, Condition, and Conditions ANDed Together. The table lists five rows corresponding to the selected fields. The bottom pane is the Results Listing, showing a table with columns: Common well name, Country, County, Field name, and Operator. It displays several records, with the first record highlighted.

	Data Element Name	Expression	Show	Sort	Sort Ord	Sort Method	Condition	Conditions ANDed Together
1	Common well name	CD_WELL.well_na	<input checked="" type="checkbox"/>	<input type="checkbox"/>			<input type="checkbox"/>	<input type="checkbox"/>
2	Country	CD_WELL.loc_cou	<input checked="" type="checkbox"/>	<input type="checkbox"/>			<input type="checkbox"/>	<input type="checkbox"/>
3	County	CD_WELL.loc_cou	<input checked="" type="checkbox"/>	<input type="checkbox"/>			<input type="checkbox"/>	<input type="checkbox"/>
4	Field name	CD_WELL.field_na	<input checked="" type="checkbox"/>	<input type="checkbox"/>			<input type="checkbox"/>	<input type="checkbox"/>
5	Operator	CD_WELL.well_op	<input checked="" type="checkbox"/>	<input type="checkbox"/>			<input type="checkbox"/>	<input type="checkbox"/>

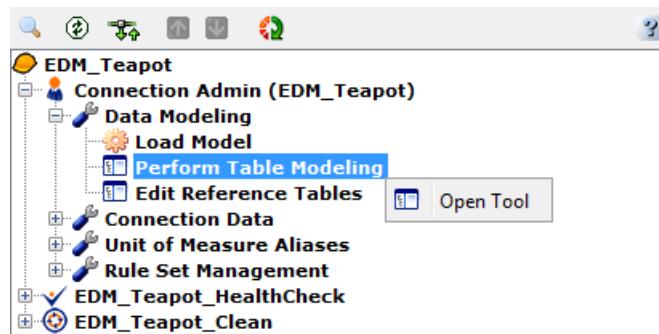
Common well name	Country	County	Field name	Operator
490251064100	USA	natrona	TEAPOT DOME	u.s. doe
490251060700	USA	natrona	TEAPOT DOME	u.s. doe
490250945000	USA	natrona	TEAPOT DOME	u.s. doe
490252280800	USA	natrona	TEAPOT DOME	hardy oil & gas
490251067700	USA	natrona	TEAPOT DOME	u.s. doe
74-CMX-10-WD	USA	natrona	TEAPOT DOME	u.s. doe

8. Select **File > Exit** to close the Data Analyzer application.

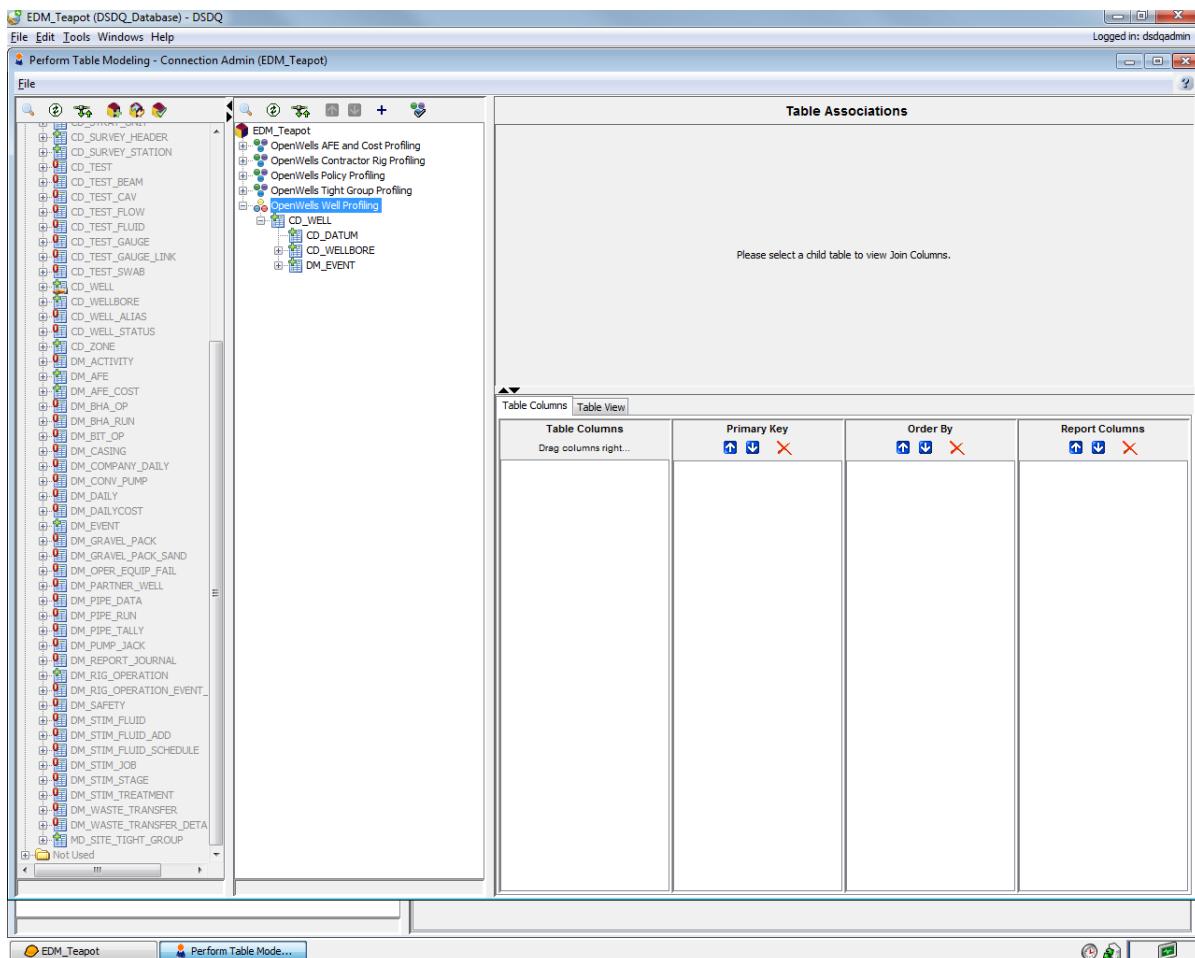
## Viewing Results on the DSDQ Web Dashboard

To view information about a DecisionSpace Data Quality submodel, you publish it to the Web Dashboard. To publish a submodel to the Web Dashboard:

1. Click  on the DecisionSpace Data Quality Tree to expand **Connection Admin (EDM\_Teapot)**.
2. Click  on the **Data Modeling** Activity.
3. Double-click the **Perform Table Modeling** Tool on the DecisionSpace Data Quality Tree or right-click the **Perform Table Modeling** Tool and select **Open Tool** from the pop-up menu.

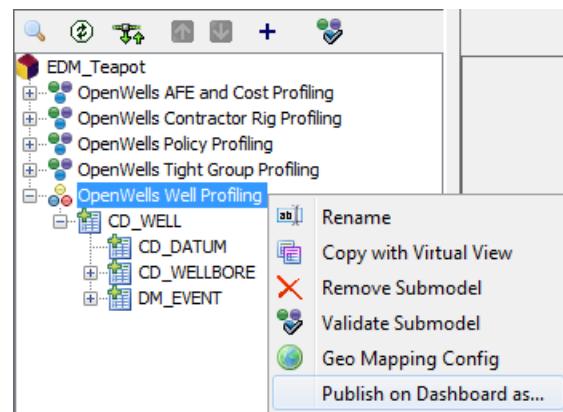


The Perform Table Modeling - Connection Admin window appears.

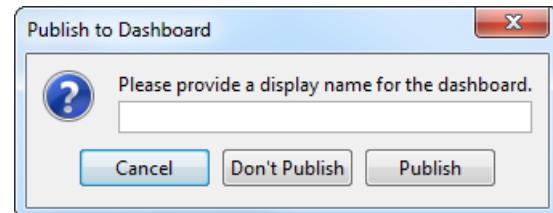


4. Right-click the **OpenWells Well Profiling** submodel from the Submodel Listing Tree and select the **Publish on Dashboard as...**

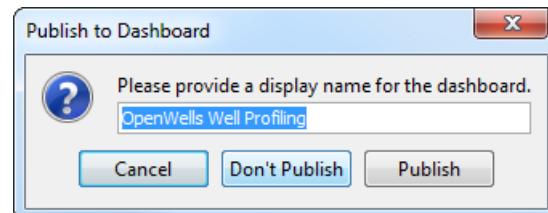
option from the pop-up menu.



The **Publish to Dashboard** dialog box appears.



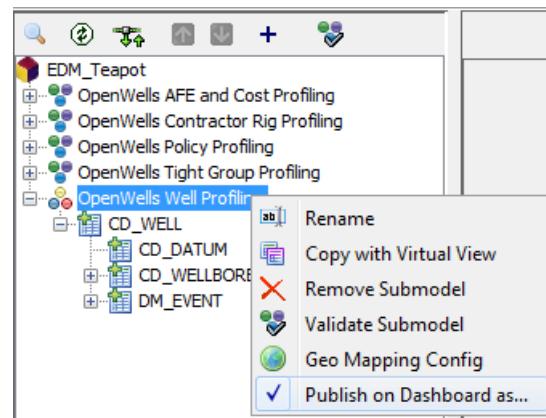
- Enter **OpenWells Well Profiling** in the **Please Provide a display name for the dashboard** field.



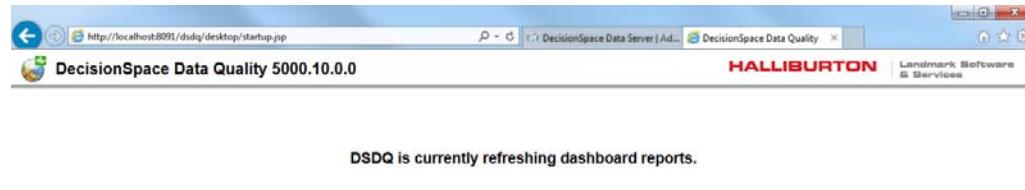
The **OpenWells Well Profiling** submodel is published to the Web Dashboard.

- To confirm that the submodel has been published, right-click the **OpenWell Well Profiling** submodel on the Submodel Listing Tree.

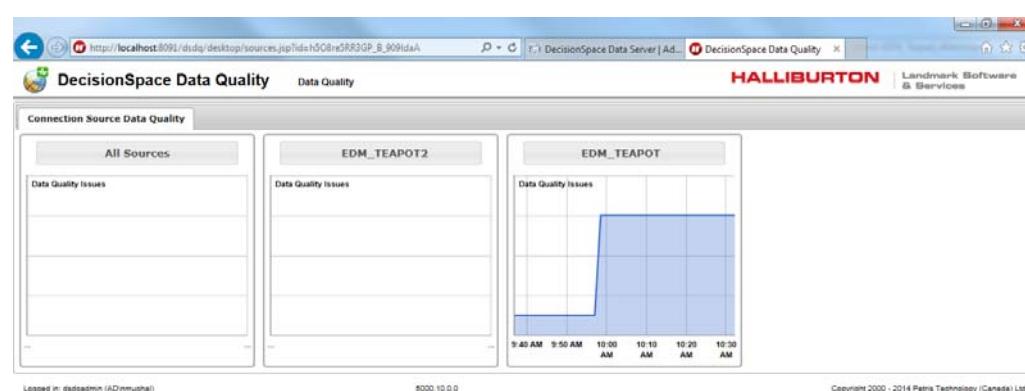
A checkmark appears on the right side of the **Publish on dashboard as...** option.



7. Select **File > Exit** from the menu bar to close the **Perform Table Modeling** window.
8. Enter **http://localhost:8091** in the address bar of the web browser. The **Please wait. Your browser will be redirected when ready.**



message displays in the web browser.  
The **Web Dashboard** is launched in the web browser.



9. Select **EDM\_Teapot** from the Data Quality Web Dashboard to display the published submodel.

10. Select the **Detailed HealthCheck** tab and then a desired **Requirement** from the **Latest Data Quality Issues** area of the Web Dashboard to view the Column Display Group and the dashboard name for the column.

