
Data Quality Management
using
DecisionSpace® Data Quality
Release 5000.10

© 2014 Halliburton

HALLIBURTON | Landmark Software
& Services

**© 2014 Halliburton
All Rights Reserved**

Information in this document is subject to change without notice. No part of this document may be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without the express written permission of Halliburton. Unauthorized reproduction or distribution of this documentation, or any portion of it, may result in severe civil and criminal penalties, and will be prosecuted to the maximum extent possible under law.

Halliburton may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. The furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

Halliburton | Landmark Software & Services

2107 CityWest Blvd, Building 2, Houston, Texas 77042-3051, USA

P.O. Box 42806, Houston, Texas 77242, USA

Phone: 713-839-2000

FAX: 713-839-2015

Internet: www.halliburton.com/landmark

Trademark Notice

3D Drill View, 3D Drill View KM, 3D Surveillance, 3DFS, 3DView, Active Field Surveillance, Active Reservoir 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, 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 Geosciences, 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 FrameworkBuilder, 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, For Production, FrameBuilder, Frameworks to Fill, FZAP!, GeoAtlas, GeoDataLoad, GeoGraphix, GeoGraphix Exploration System, Geologic Interpretation Component, Geometric Kernel, GeoProbe, GeoProbe GI DataServer, GeoSmith, GES, GES97, GesFull, GESXplorer, GMIAplus, GMI Imager, Grid3D, GRIDGENR, H. Clean, Handheld Field Operator, HHFO, High Science Simplified, Horizon Generation, I² Enterprise, iDIMS, iEnergy, Infrastructure, iNotes, Iso Core, IsoMap, iWellFile, KnowledgeSource, Landmark (*as a 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 a 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, 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 3DPSDM, 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, SEISINFO, 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 Training Manual

Chapter 1: Introduction	1-1
Course Objectives	1-2
Chapter Overview	1-3
DecisionSpace Data Quality Software: Introduction	1-4
Users of the Data Quality Application	1-5
DecisionSpace Data Quality Project Window	1-6
DecisionSpace Data Quality Components	1-7
DecisionSpace Data Quality Installation	1-8
Hardware Requirements	1-8
Application Database Storage	1-9
Application Licensing	1-10
Installation Steps	1-11
Setting up Database Storage for Oracle 10/11	1-12
Setting up Database Storage for MS SQL 2008/2012	1-13
Installing DecisionSpace Data Quality on a Server System	1-13
Deploying DecisionSpace Data Quality to Workstations - Automatic Deployment	1-23
Deploying DecisionSpace Data Quality to Workstations - Manual Deployment	1-23
Starting the Data Quality Application	1-24
To Start DSDQ from a Shared Network Drive	1-25
For Assistance	1-27
Using the User Guide	1-27
Contacting Landmark Customer Support	1-28
Support via Web Portal	1-28
Technical Assistance Centers	1-29
Regional Offices	1-29

Chapter 2: Connecting DecisionSpace Data Quality (DSDQ) with DecisionSpace Data Server (DSDS)	2-1
Chapter Overview	2-2
DecisionSpace Data Server: An Introduction.....	2-3
DecisionSpace Data Server Layout.....	2-6
DecisionSpace Data Server Installation.....	2-9
System Requirements	2-9
Installing DecisionSpace Data Server.....	2-11
Starting the DecisionSpace Data Server Service.....	2-21
On Windows:	2-21
On Linux:	2-22
On both platforms:	2-22
Exercise: Stopping the DecisionSpace Data Server Service.....	2-23
On Windows:	2-23
On Linux:	2-23
Post-Installation Procedures	2-24
Exercise: Adding a Data Source	2-24
Testing the Connection to the Data Source	2-28
Generating Virtual Databases (VDBs)	2-29
Exercise: Creating Database Connections	2-30
Exercise: Connecting to DSDS using Web Browser, Excel (powerpivot).....	2-34
Connecting DecisionSpace Data Quality with DecisionSpace Data Server	2-42
Exercise: Creating Connections	2-42
DecisionSpace Data Server Quick Start Connections	2-47
Exercise: Creating New DecisionSpace Data Server Quick Start Connections ...	2-47

Chapter 3: Managing Projects	3-1
Chapter Overview	3-2
Creating a Project	3-3
Exercise: Creating a Project	3-5
Working with Projects	3-9
Exercise: Opening an Existing Project	3-10
Exercise: Exporting a Project	3-10
Exercise: Deleting a Project	3-12
Exercise: Importing a Project	3-13
 Chapter 4: Data Evaluation	4-1
Chapter Overview	4-2
Perform Table Modeling	4-3
Exercise: Loading a Data Model	4-3
Exercise: Perform Table Modeling	4-6
Exercise: Editing a Reference Table	4-18
Exercise: Loading Test Data	4-21
Exercise: Selecting Test Data	4-23
Data Evaluation in DecisionSpace Data Quality	4-26
Evaluating Data Volume and Quality	4-28
Exercise: Profiling Data Using SQL Query	4-28
Exercise: Running Table Analysis on All Tables	4-33
Exercise: Running Column Analysis on All Columns	4-35
Exercise: Running Table Analysis on Modeled Tables	4-38
Exercise: Running Column Analysis on Modeled Tables	4-41
Exercise: Comparing Table Analysis Results	4-44
Identifying Data Issues	4-48
Exercise: Configuring the Detailed HealthCheck Tool	4-48
Exercise: Running Detailed HealthCheck	4-57
Exercise: Running Standards Compliance	4-61
Exercise: Comparing Multiple Detailed HealthCheck Results	4-64

Exercise: Running Record Summary by Group	4-68
Exercise: Running Orphan Count	4-71
Exercise: Running Application Reference Analysis	4-74
<i>Chapter 5: Data Cleansing and Standardization</i>	5-1
Chapter Overview	5-2
Resolving Data Issues.....	5-3
Exercise: Adding a Clean Phase	5-3
Rapid Clean Activity.....	5-5
Exercise: Configuring the Rapid Clean Tool	5-5
Exercise: Running the Test Rapid Clean Task	5-7
Exercise: Running the Rapid Clean Task	5-12
Detailed Clean Activity	5-17
Exercise: Configuring the Detailed Clean Tool	5-17
Exercise: Running the Detailed Clean Task	5-24
Cleaning Application References.....	5-27
Exercise: Loading Application References.....	5-27
Exercise: Selecting Application Reference Values	5-30
Exercise: Running Reference Cleanup	5-33
Making Values Consistent	5-37
Exercise: Configuring Make Values Consistent.....	5-37
Exercise: Loading Current Values	5-42
Exercise: Selecting Consistent Values	5-45
Exercise: Running Make Values Consistent.....	5-51
<i>Chapter 6: Managing Data Duplication in DecisionSpace Data Quality</i>	6-1
Chapter Overview	6-2
Data Duplication	6-3
Exercise: Adding a Match Phase for a Single Data Source	6-3

Detailed Match for a Single Data Source	6-6
Exercise: Configuring the Detailed Match for a Single Data Source	6-6
Exercise: Finding Detailed Matches within a Single Data Source	6-15
Managing Duplication for a Single Data Source	6-20
Exercise: Configuring Duplication Fixes for a Single Data Source	6-20
Exercise: Applying Duplication Fixes for a Single Data Source	6-25
Detailed Match across Data Sources.....	6-29
Exercise: Adding a Match Phase across Data Sources	6-29
Exercise: Configuring Detailed Match across Data Sources	6-31
Exercise: Finding Detailed Matches across Data Sources	6-39
Manage Duplication across Data Sources	6-43
Exercise: Configuring Duplication Fixes across Data Sources	6-43
Exercise: Applying Duplication Fixes across Data Sources.....	6-48

<i>Chapter 7: Data Consolidation in DecisionSpace Data Quality</i>	7-1
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	7-16
Merge Flags for Target Columns	7-16
Using Merge	7-19
Exercise: Running the Merge Tool.....	7-19
Exercise: Modifying Data while Merging	7-23

<i>Chapter 8: Requirements Administrator in DecisionSpace Data Quality</i>	8-1
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
 <i>Chapter 9: Configuring the Data Quality Web Dashboard</i> ...	9-1
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
 Chapter 10: Project Administration.....	 10-1
Chapter Overview	10-2
Roles & Permissions.....	10-3
Managing Users	10-5
Exercise: Creating a User	10-5
Exercise: Editing a User	10-6
Exercise: Removing a User.....	10-7
Advanced Scheduling	10-8
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
Viewing Job Group Charts.....	10-16
Exercise: Viewing Job Group Performance	10-16
Exercise: Viewing Job Group Sequencing	10-17
Managing Jobs	10-18
Exercise: Opening Job Administrator.....	10-19
Exercise: Stopping Current Running Jobs	10-20
Exercise: Viewing Server Logs	10-21
Managing Connections.....	10-22
Exercise: Editing a Data Owner Connection	10-23
Exercise: Editing a Workspace Connection	10-26
Exercise: Removing a Data Owner/Workspace Connection	10-28

Workshop A: Connecting DSDQ to an OpenWorks Data Source

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

Workshop B: Connecting DSDQ to an EDM Data Source ... 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 1

Introduction

Welcome to the Data Quality Management using DecisionSpace® Data Quality class. This course introduces you to the DecisionSpace Data Quality software that helps you evaluate, correlate, correct, monitor and synchronize data across the enterprise.

It covers the functionality and use of the modules within the Data Quality application so that you can manage projects, users, data source connections and the rules repository to:

- Identify the full spectrum of data issues including ambiguous, incomplete, inaccurate, inconsistent or missing data, and
- Clean the data before it impacts end user workflows or results in poor decision making

Course Objectives

This manual is designed to be used with the DecisionSpace Data Quality application. Refer to the exercises during class and use the manual for reference when you return to your workplace.

Briefly, this manual covers the following topics:

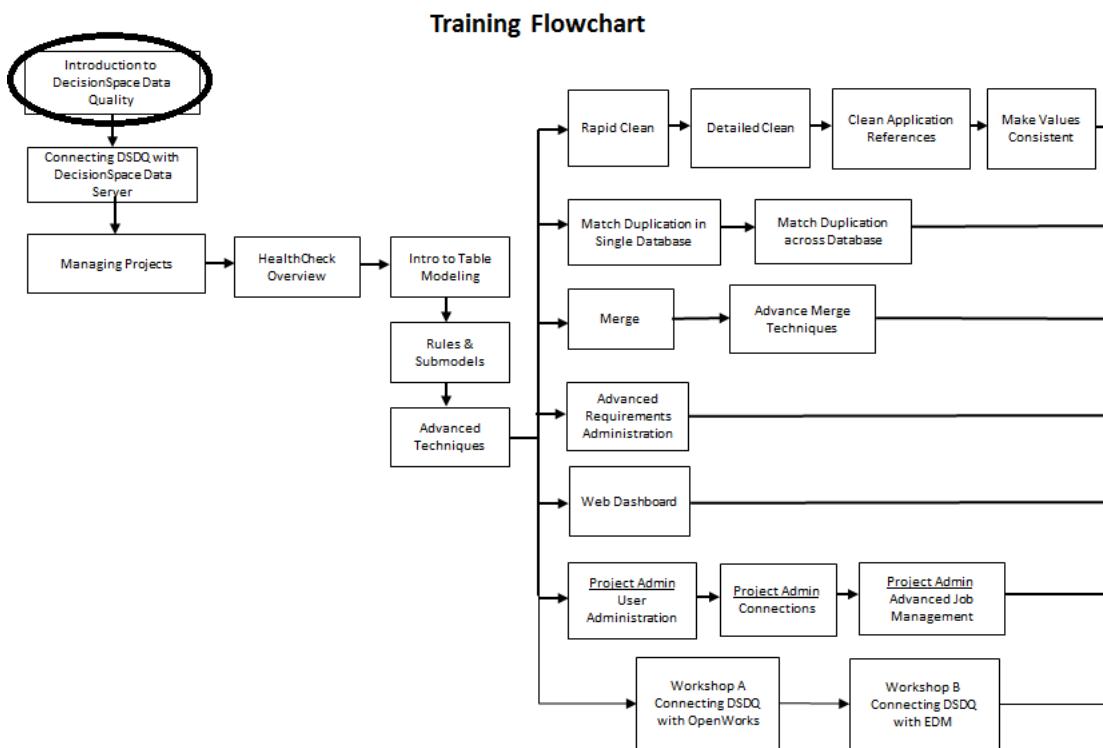
- An overview of the DecisionSpace Data Quality (DSDQ) application
- Connecting the Data Quality application with DecisionSpace Data Server
- Managing projects in the Data Quality application
- Using the Data Quality Phases i.e. HealthCheck, Clean, Match & Merge
- Managing the Data Quality Rules Repository
- Using the Data Quality Web Dashboard for analyzing the data quality assurance process
- Managing users and projects

Chapter Overview

In this chapter, you will learn about:

- The DecisionSpace Data Quality software including the Data Quality Project window and components
- Installing the Data Quality application
- Starting the DecisionSpace Data Quality software

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:



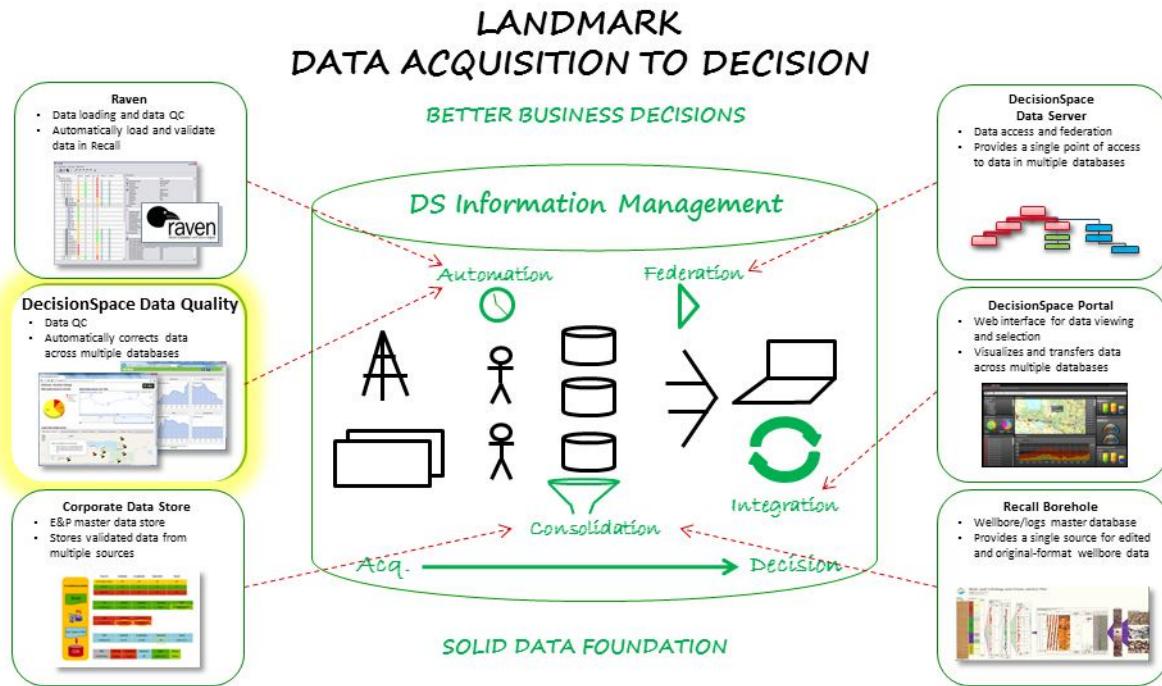
DecisionSpace Data Quality Software: Introduction

The growing number of new data types from complex resources plays coupled with a lack of skilled data management experts to ensure availability of quality data is consistently driving E&P companies to automate their data quality management efforts. To create long term, sustainable value, the upstream E&P industry is looking for a solution that:

- Addresses data quality issues across disparate data stores
- Promptly identifies and corrects data quality issues helping in decreasing the risk associated with bad quality data
- Resolves data quality issues with reusable, repeatable and automated quality assurance processes

The DecisionSpace Data Quality software helps you deal with this avalanche of data by offering data quality tools designed to evaluate, correlate, correct, monitor and synchronize data across the enterprise. Thus, enabling you to:

- Quickly assess the health of your data
- Automate and schedule data quality jobs for perpetual monitoring and continuous data improvement
- Removing data bottlenecks that hinder a project's progress
- Communicate data quality improvements over time to management and end users



Users of the Data Quality Application

The Data Quality application is intended for users who need help with:

- Data profiling prior to consolidations and clean-ups. Includes data types such as seismic, well, drilling, land and production
- Ongoing data quality management
- Data assessments conducted post-merger and acquisition activities
- Creating a business case to justify data quality and ROI to senior management
- Confirming operated/non-operated interest wells against large public repositories
- Developing and maintaining a cross-master well list across data stores and unique well identifier (UWI) changes
- Identifying and correcting duplicate records during the merger of two companies

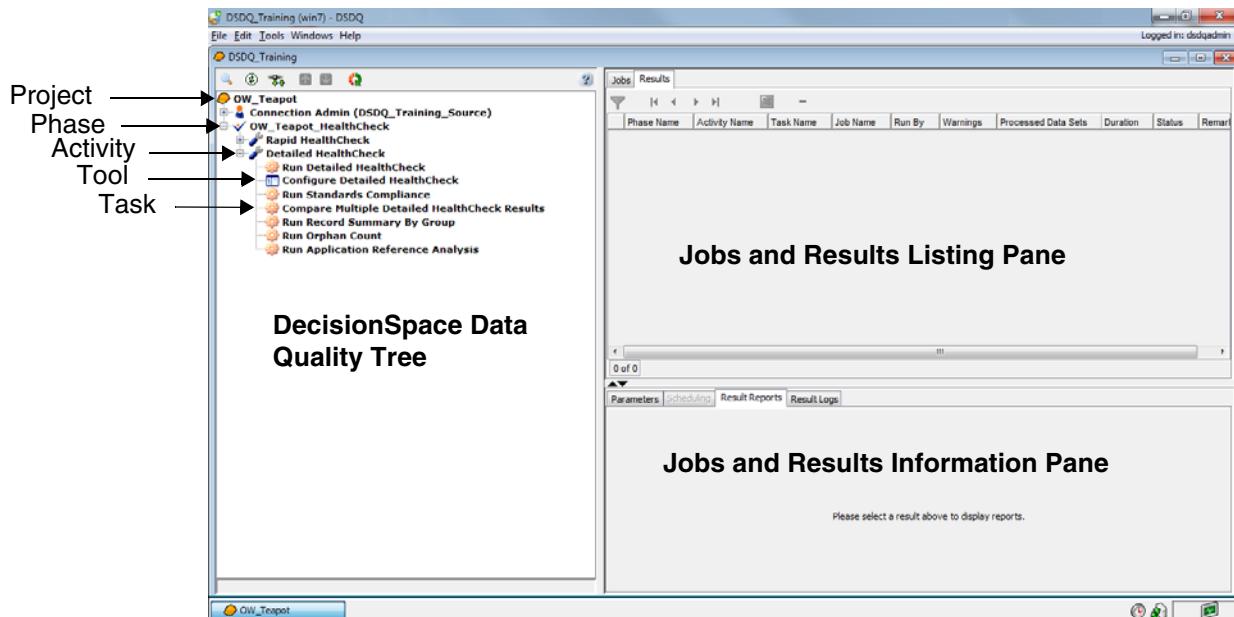
- Matching well data, land records, facilities information and production data in one “view”
- Maintaining corporate standards for reference purposes

DecisionSpace Data Quality Project Window

DecisionSpace Data Quality is a straightforward, yet powerful tool for managing data quality. Part of the ease of use of this application is that all of the work to be performed is controlled from one window, the DecisionSpace Data Quality Project Window. This window allows you to maneuver through the numerous functions of the application. It is divided into 3 panes:

- The DecisionSpace Data Quality Tree
- The Jobs and Results Listings Pane
- The Jobs and Results Information Pane

The Project Window can be maximized, minimized, or resized according to your preference



DecisionSpace Data Quality Components

Before using the Data Quality application, it is important to understand the following components. A brief description of each component is given in the table below:

Component	Description
Project	A project contains all the Phases, Activities, Tools, Tasks, Jobs, and Results for a defined area of work. This component works as a Master folder that holds all the files created by you and the application. A project is identified in the application by the hard hat  icon.
Phase	A Phase is composed of a set of Activities that are designed to achieve a particular set of objectives, which in turn contains Tools and/or Tasks. Phases are added to the project by you as the Project is created. An icon is specified in the application for each type of Phase (i.e. HealthCheck Phase  , Clean Phase  , Match Phase  and Merge Phase ).
Activity	An Activity is a group of Tools and/or Tasks that are steps toward reaching a specific goal. When you add a Phase to a Project, the Activities are saved to the Project. Activities do not require any action on your part, but instead hold all the Tools and Tasks that you may perform. An activity is identified in the application by a wrench  icon.
Tool	A Tool can be thought of as a mini-application. When a Tool is activated, an application window opens, allowing for data manipulation and/or configuration. The information supplied by the user is then used within the application. A Tool is identified in the application by the window  icon.
Task	The purpose of a Task is to create jobs, provide reports on the data to the user, and/or to set-up the data for a specific processing operation. A Job in a Task may be run after it has been configured via the relevant Tool. The user creates a job with specific parameters within the Task before running it. A Task is identified in the application by the gear  icon.
Job	Jobs are tasks with specific Parameters. These specifications can make jobs unique and generate different results. An existing job can be run by selecting it and clicking on the Run Selected Job  button in the Jobs toolbar. A new job can be created, and saved and run at the same time by clicking on the Run Job  button in the Parameters toolbar.
View	View is a window that is an interactive display of results generated by a task. Views are accessed in the Result Reports tab in the Results Information pane and are launched by clicking the Basic View Frame  icon once a result is selected.

Component	Description
Submodel	A Submodel is a core data structure within the Data Quality application that is used by all Activities. Created in Perform Table Modeling , a Submodel is a grouping of data tables within a database, identified by the three circle  icon.
Service Level	A Service Level is a grouping of Data Quality rules that can be configured for specific tasks, projects, or even intended for a single submodel. Identified by the book  icon, Service Levels are used directly with Submodels to set up and run rules against the Application Model.

At the project window level, you can perform an action either by right-clicking an item or by using the toolbar. At the DecisionSpace Data Quality Tree level, there are some actions that can be performed by right-clicking the component to display the pop-up menu. Double-clicking a tool or task opens the tool or initiates the process of job creation for a task.

DecisionSpace Data Quality Installation

This section outlines the requirements and procedures for installing DecisionSpace Data Quality on a server system and deploying it to workstations.

Hardware Requirements

The table below describes the minimum and recommended hardware configurations for running the DecisionSpace Data Quality application:

Workstation	Recommendation
Processor and Memory	<ul style="list-style-type: none"> 64-bit CPU (AMD64 or Intel 64) with 12 GB of RAM. (Recommended) 64-bit CPU (AMD64 or Intel 64) with 6 GB of RAM. (Minimum)
Hard Drive Space	<ul style="list-style-type: none"> Free space on the hard drive for application and TEMP directory files: 10 GB
Windows operating systems	<ul style="list-style-type: none"> Windows 7 (64-bit)

Workstation	Recommendation
Server	
Processor and Memory	<ul style="list-style-type: none"> 64-bit CPU (AMD64 or Intel 64) with 12 GB of RAM. (Recommended) 64-bit CPU (AMD64 or Intel 64) with 6 GB of RAM. (Minimum)
Hard Drive Space	<ul style="list-style-type: none"> Free space on the hard drive for application and TEMP directory files: 20 GB
Windows operating systems	<ul style="list-style-type: none"> Server 2008 (Certified) (64-bit)
Database	
Application data storage of projects and results can be held in either:	<ul style="list-style-type: none"> PostgreSQL: Bundled with DecisionSpace® Data Quality installer, requires no database administrator configuration Oracle 10/11: Storage and schemas configured by your database administrator Microsoft SQL Server 2008/2012: Storage and schemas configured by your database administrator
Storage space required depends on size and amount of data sources connected to by DecisionSpace Data Quality	<ul style="list-style-type: none"> A good guideline is to allocate 100 GB of tablespace for production systems initially, with room to grow as more projects and connections are added.

DecisionSpace Data Quality Server offers a web dashboard, web deployment of DecisionSpace Data Quality desktop client, and runs scheduled jobs in unattended mode.

Application Database Storage

A new feature in DecisionSpace 5000.10.0 is automatic data storage using the Postgres database. During the DecisionSpace Data Quality installation, you only need to select Postgres Database Server from list of components to install.

Postgres automatically installs and is configured within the main DecisionSpace Data Quality installation's 'database' folder. Optionally, during installation, one can specify a different folder to hold Postgres data. For example, this is useful if the installation of the DecisionSpace Data Quality application occurs on the C: drive, while the much larger

D: drive should be used for DecisionSpace Data Quality application data storage.

Note

Automatic Storage Using Postgres:

If automatic data storage using Postgres is selected, then there is no need to manually configure application storage in Oracle or MS SQL Server as described in the next two section.

Alternatively, for those who prefer to manage DecisionSpace Data Quality Application data storage manually, database administrators will need to create a minimum of three empty data schemas in either Oracle or MS SQL Server:

- **DSDQ_MASTER** is the DecisionSpace Data Quality application configuration repository and requires at least **10 GB** of table space.
- **DSDQ_REFERENCE** holds the standards tables that Landmark provides with DecisionSpace Data Quality that can be used as a reference when cleaning the user database. These references can also be modified by the user. It is recommended that you make available at least **1 GB** of table space.
- Workspace schemas, often named **DSDQ_TEST_WKSP**, are used to capture the project results. DecisionSpace Data Quality software creates the necessary tables in this workspace schema to capture results. The user has the option to keep the results from each of the tasks, so this database can grow quite large. It is recommended to make available at least **25 GB** of table space for each workspace; however, **additional space may be required** depending on the data retention policies set by DecisionSpace Data Quality users. A separate workspace should be created for every new DecisionSpace Data Quality project.

The names above are a guideline only. You may use any database user names to match your local IT policies. You will be prompted for the names you use in the DecisionSpace Data Quality Setup Wizard and Desktop applications.

Application Licensing

DecisionSpace Data Quality requires an existing Landmark License Application Manager (LAM) to be configured. For more information

about setting up an LAM, please refer to the files and documentation found on the Landmark Software Manager site.

Once the LAM is configured, all systems running DecisionSpace Data Quality require a **LM_LICENSE_FILE** environment variable to be configured. This environment variable will direct DecisionSpace Data Quality to the location of the LAM license file. The **LM_LICENSE_FILE** environment variable should point to the LAM server, using the configured port, which is 2013 by default.

To configure a new environment variable:

1. Select **Start > Control Panel** from the Windows taskbar and then Systems and Security from the Control Panel window.
2. Select System.
3. Click the **Advanced system settings** option under **Control Panel Home**.
The **System Properties** dialog appears with the **Advanced** tab selected by default.
4. Click the **Environment Variables...** button.
5. Click the **New...** button on the **System Variables** group box.
6. For **Variable name**, enter LM_LICENSE_FILE.

For **Variable value**, enter a value according to the configured LAM Server.

E.g. **2013@LocalServer**, where **2013** is the LAM configured port, and **LocalServer** is the server name that the LAM is configured on.

Installation Steps

In order to complete the installation of DecisionSpace Data Quality, you will need to:

- Set up Database Storage on either Oracle or MS SQL Server (For this exercise, we will use Oracle 11 database)
- Install DSDQ on a server system, alongside an existing DecisionSpace Data Server installation

- Deploy DSDQ to workstations, either automatically or manually

Setting up Database Storage for Oracle 10/11

To setup application database storage for Oracle, you will need to create new **DSDQ_MASTER**, **DSDQ_<PROJ>_WKSP** and **DSDQ_REFERENCE** users and grant access. A sample script is provided in the *docs/sql/directory* for Oracle databases.

1. Log in to the Oracle database as an administrator to ensure that you have the permission required to create new users (using a tool like SQLPlus).
2. From the command prompt, execute the following command in order to connect to Oracle:

```
sqlplus /nolog
```

3. From sqlplus, execute the following command to connect as sysdba:

```
SQL>connect / as sysdba;
```

4. Execute the following command to create the table space:

```
CREATE TABLESPACE DSDQ DATAFILE  
'c:\app\student\oradata\DSDQ\dsdq.dbf' SIZE 1024M  
AUTOEXTEND ON MAXSIZE unlimited EXTENT MANAGEMENT  
LOCAL UNIFORM SIZE 4M ONLINE;
```

5. Execute the following commands to create the **DSDQ_MASTER**, **DSDQ_REFERENCE** and **DSDQ_WORKSPACE** users and grant access:

```
SQL> create user DSDQ_MASTER identified by landmark  
default tablespace DSDQ temporary tablespace temp  
account unlock;
```

```
SQL> create user DSDQ_REFERENCE identified by landmark  
default tablespace DSDQ temporary tablespace temp  
account unlock;
```

```
SQL> create user DSDQ_TEST_WKSP identified by landmark  
default tablespace DSDQ temporary tablespace temp  
account unlock;
```

```
SQL> grant connect, resource to DSDQ_MASTER;
```

```
SQL> grant connect, resource to DSDQ_REFERENCE;  
  
SQL> grant connect, resource to DSDQ_TEST_WKSP;  
  
SQL> exit
```

Note

It is necessary to set all options by default in each step for this exercise.

Setting up Database Storage for MS SQL 2008/2012

Note

This procedure is for your reference. For the purpose of this training, please skip this section.

To setup application database storage for MS SQL, you will need to create new DSDQ_MASTER, DSDQ_TEST_WKSP and DSDQ_REFERENCE users and grant access.

1. Log in to MS SQL as an administrator to ensure that you have the permission required to create new users (using a tool like SQL Server Management Studio).
2. Right-click on the Databases folder and select New Database... from the drop-down menu.
The New Databases window appears.
3. Specify a Database Name and Owner for each user that you create.
4. Repeat steps 2 and 3 to create all three users.

Note

The Owner will be associated with the User parameter in the Create Connections window within DecisionSpace Data Quality.

Installing DecisionSpace Data Quality on a Server System

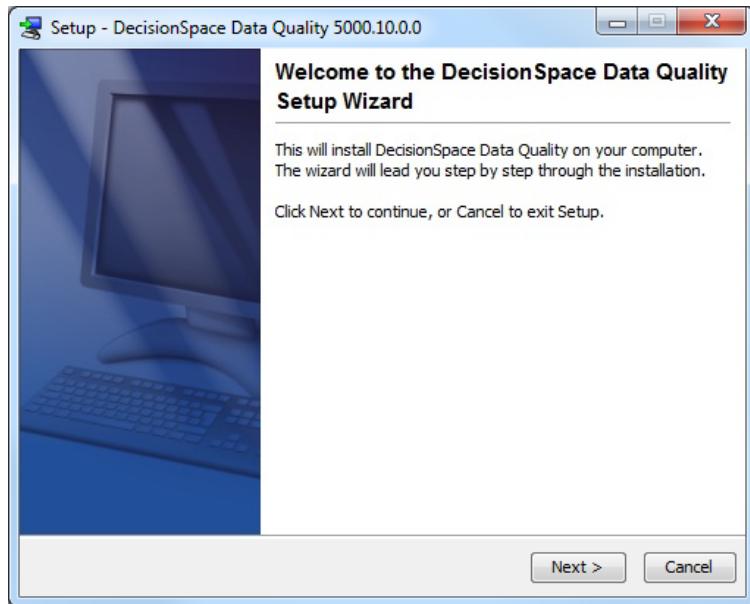
There are two types of DecisionSpace Data Quality installers (Windows example provided):

- DSDQ_Server_Win64_5000_10_0.exe: 64-bit installer for 64-bit Windows operating systems (for server installation).

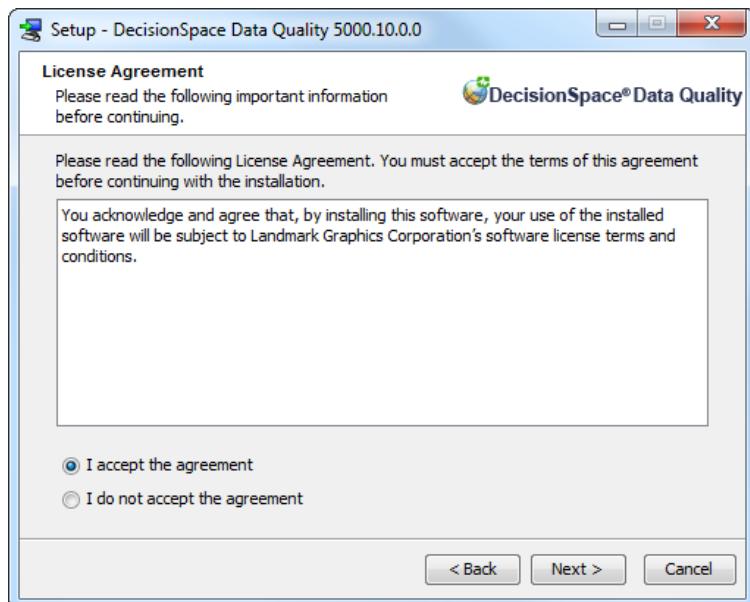
- DSDQ_Client_Win64_5000_10_0_0.exe: 64-bit installer for 64-bit Windows operating systems (for client installation).

1. Double-click the provided installer to launch the **DecisionSpace Data Quality Setup Wizard**.

The DecisionSpace Data Quality Setup Wizard screen appears.

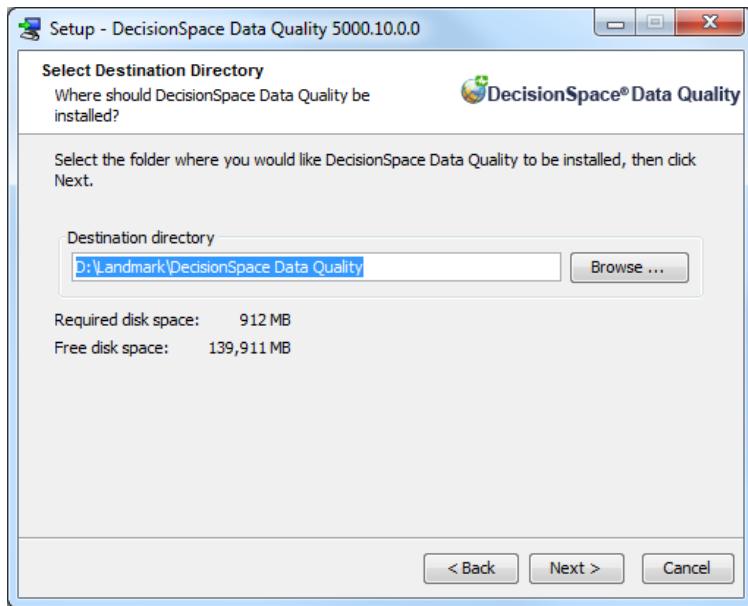


2. Click **Next** to continue.

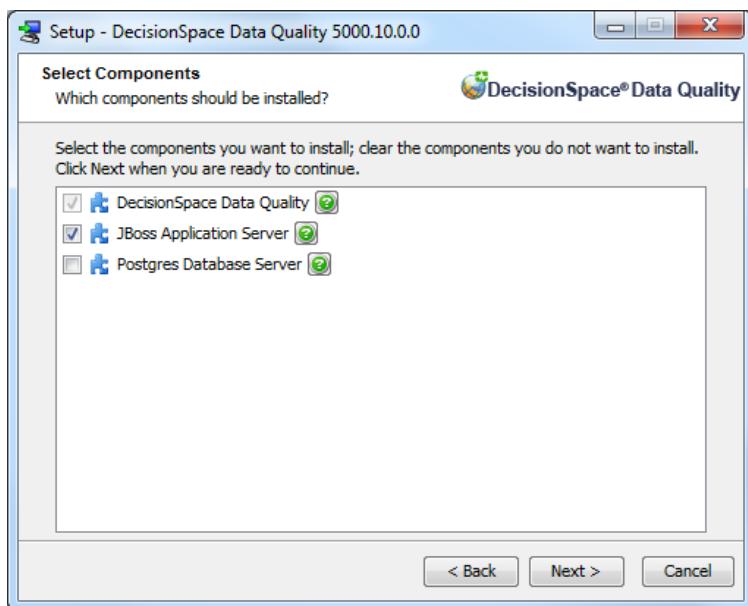


3. Review the software license agreement information given in the **License Agreement** screen and select the **I accept the agreement** option.

4. Click **Next** to continue.



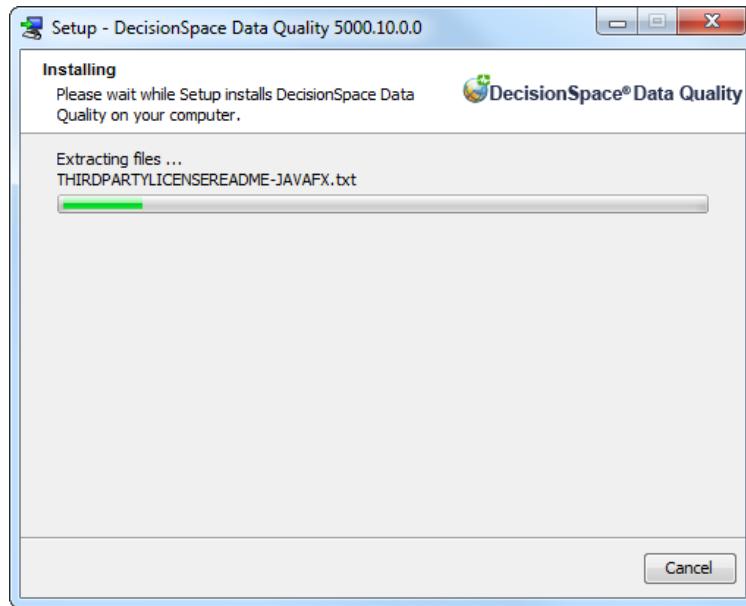
5. Select the destination directory where you would like DSDQ to be installed by clicking the **Browse...** button. If the specified directory does not exist, it will be created.
6. Click **Next** to continue.



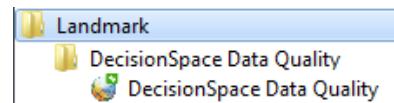
7. Select the following component:
 - **DecisionSpace Data Quality**

- **JBoss Application Server**

8. Click **Next** to install DecisionSpace Data Quality on your computer.

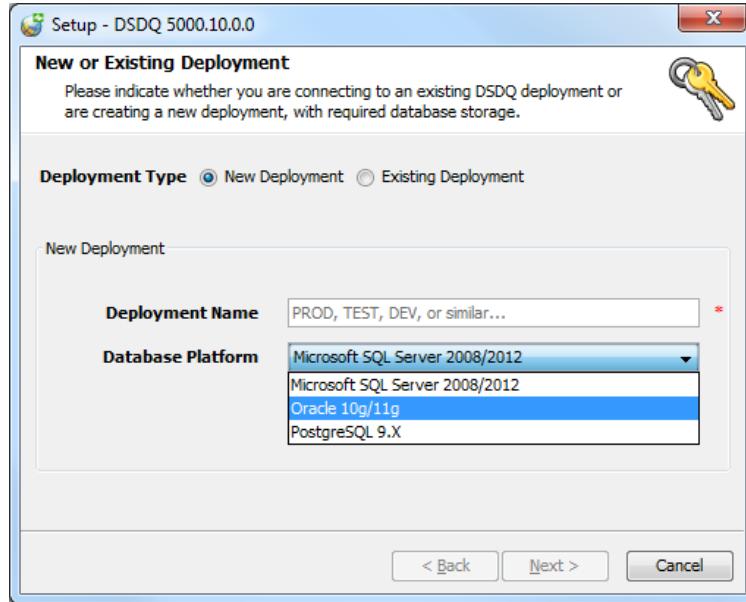


9. Once the installation process is complete, the setup process will begin.
An application shortcut will be automatically created on your desktop along with a *DecisionSpace Data Quality* program group under *Landmark* in the Windows Start Menu.

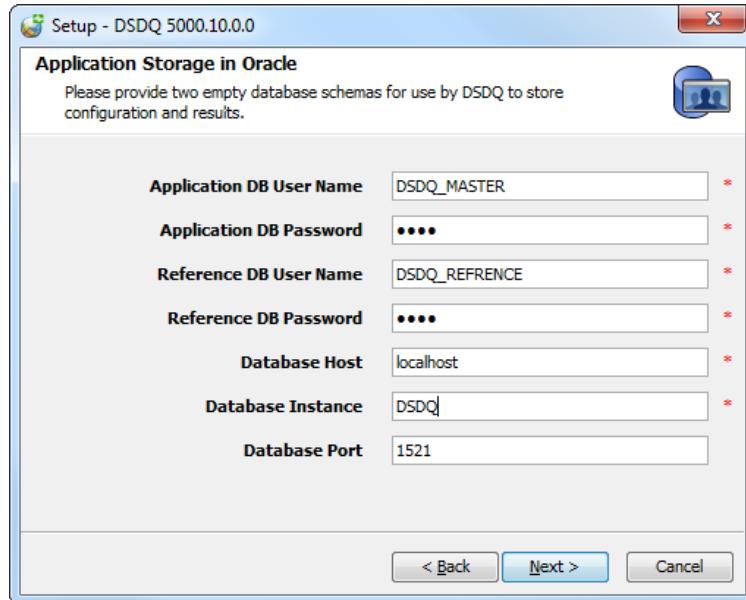


10. Continue with the application setup process by entering the following information in the **New or Existing Deployment** screen:
 - Select **New Deployment** for **Deployment Type**.
 - Enter a name in the **Deployment Name** field.

- Select **Oracle 10g/11g** from the **Database Platform** drop down list.



- Click **Next** to continue.
- Enter the **User Name** and **Password** to connect to the DSDQ_MASTER and DSDQ_REFERENCE databases and then the **Data Host** and **Instance** information (modify the **Database Port** if required).



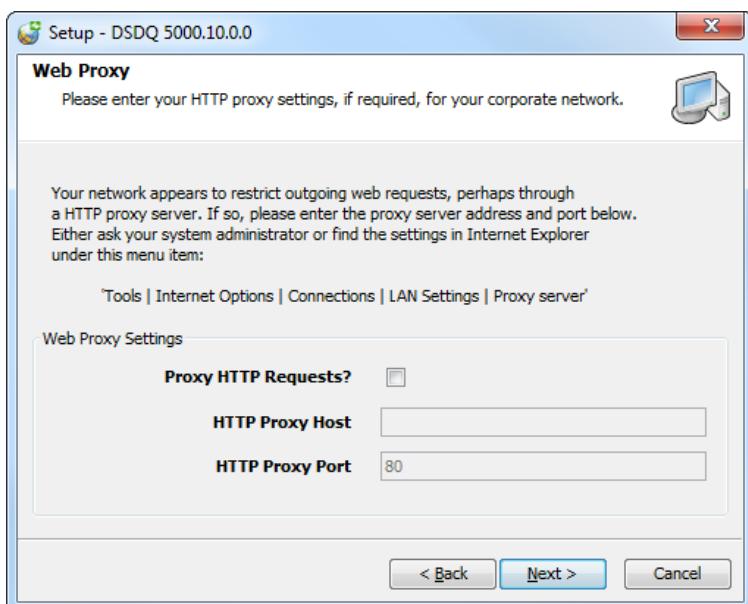
- Click **Next** to continue.
The application will attempt to connect to the databases. If this

attempt is successful, the specified databases will be installed and the setup process will continue.

Note

There must be empty database schemas created solely for the use of DecisionSpace Data Quality.

14. Enter the **HTTP Proxy Host** and **HTTP Proxy Port** in case your network restricts outgoing web requests, perhaps through a HTTP proxy server. This will allow for error reports to be delivered to Halliburton Technical Support.



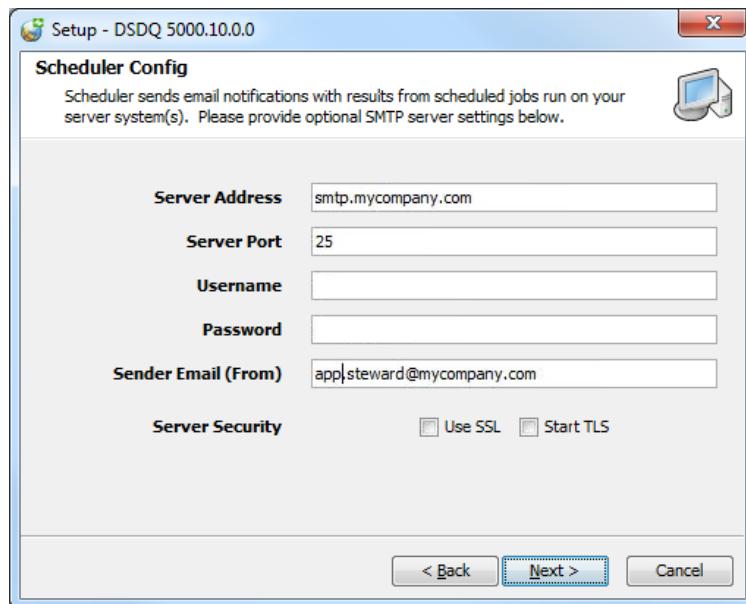
Note

The DecisionSpace Data Quality Setup Web Proxy window will be enabled if during the set up stage, DecisionSpace Data Quality is unable to access an external web server directly.

15. Click **Next** to continue.
The application will try to connect to the proxy server and if successful, the **Scheduler config** window appears.
16. Enter the **Server Address** and **Server Port** if you have setup the DSDQ Server package and you are planning to schedule jobs to run and wish to receive email notifications regarding the scheduled jobs.

17. Specify the email address or email alias of your local DecisionSpace Data Quality application steward in the **Sender Email (From)** field.

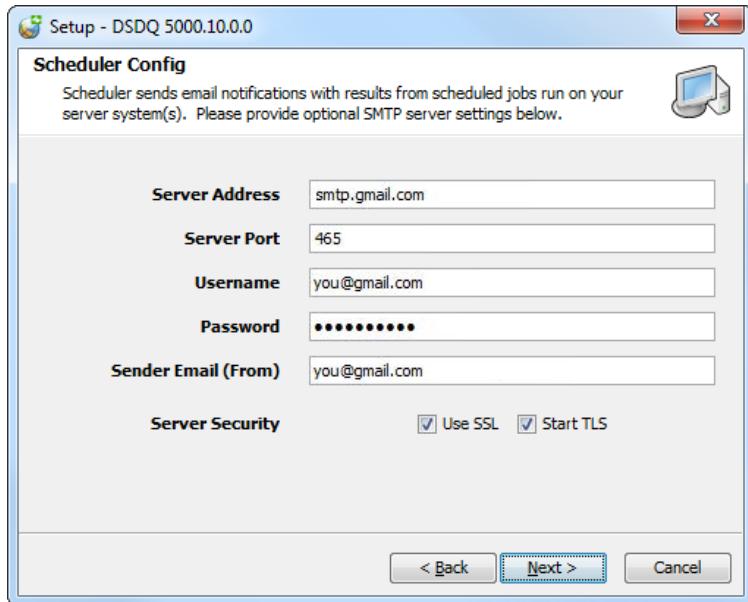
This address will be used in the From: field of email notifications sent to users in order to control who they reply to with their questions.



18. Optionally, if your mail server requires authentication in order to send email, please enter the SMTP **Username** and **Password**.

19. If your mail server requires additional connection security, select **Use SSL** and **Start TLS** check boxes as appropriate.

Shown below are some example settings for sending to a secure SMTP server.

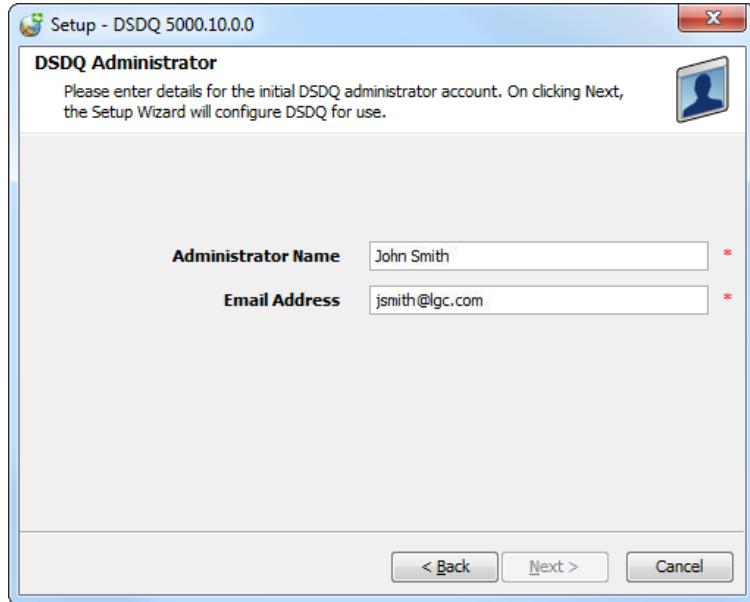


20. Leave all options by default and click **Next** to continue.
The Setup Wizard attempts to connect to the Mail Server. The **DSDQ Administrator** window appears.
21. Create a DecisionSpace Data Quality login account by specifying an **Administrator Name** and **Email Address** in the **DSDQ Administrator** window.

Note

This initial setup will allow the application to email information about scheduled jobs to specified users.

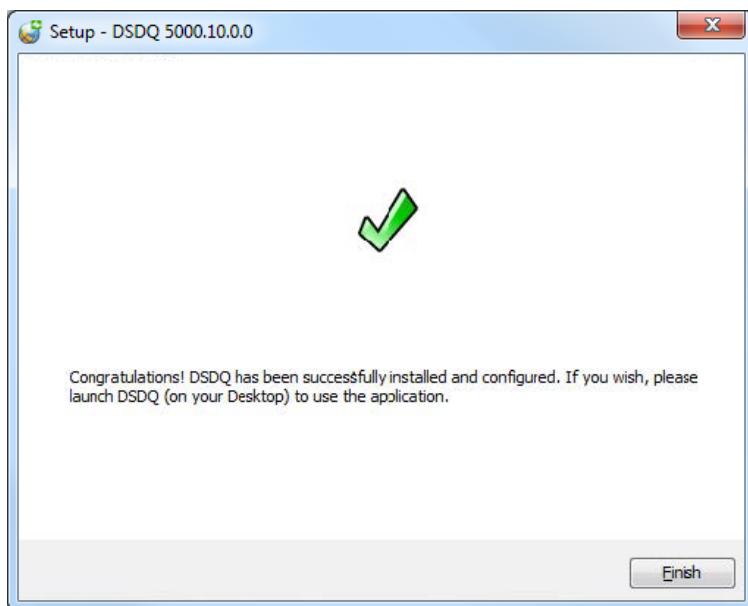
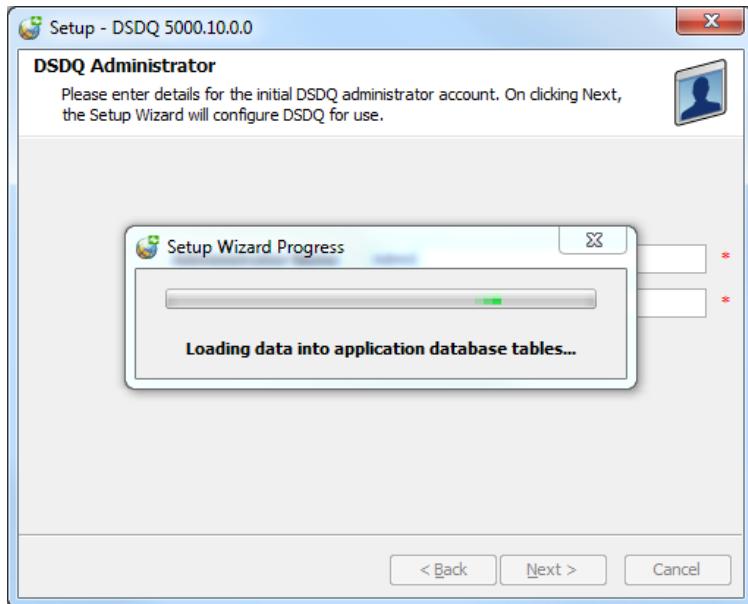
The system administrator account will be created and used for logging in DSDQ.



22. Click **Next** to continue.

The System Account will be validated, DSDQ_MASTER and

DSDQ_REFERENCE data will be loaded and the application will be registered. Please be patient as this step may take a few minutes.



23. Click **Finish** to complete the DecisionSpace Data Quality setup process.

Deploying DecisionSpace Data Quality to Workstations - Automatic Deployment

DecisionSpace Data Quality automatic deployment allows you to launch the DecisionSpace Data Quality Client application from a web page hosted on your intranet, your desktop or from the Start menu. As a pre-requisite to automatic deployment, your computer must have Java version 7 or higher installed on it. For information on installing Java, please visit www.java.com.

The primary benefit of automatic deployment is that updates applied to the DecisionSpace Data Quality Server system are automatically provided to workstation users the next time they run DecisionSpace Data Quality Client.

To automatically deploy DecisionSpace Data Quality:

1. Ensure that the Server installation has been completed successfully.
2. Enter the web server address and port in the address bar of your web browser. E.g. `http://ServerMachine:8091`
3. Click on the icon button to launch the **DecisionSpace Data Quality Desktop Client** on the desktop.
The **Opening dsdq.jnlp** window appears.
4. Select **Open with Java™ Web Start Launcher** (default) and click **OK**.
The DSDQ application will be launched and a shortcut will be created in the *All Programs* list on the Start menu.

Deploying DecisionSpace Data Quality to Workstations - Manual Deployment

This option is for IT departments who wish to script their own deployment of DecisionSpace Data Quality Client for their users. In this scenario, subsequent DecisionSpace Data Quality Server updates are not automatically pushed to workstations. Therefore, system administrators are responsible for keeping all deployed clients up-to-date. Manual deployment of DecisionSpace Data Quality includes the correct version of Java, negating the need to install Java 7 or later.

To manually deploy DecisionSpace Data Quality:

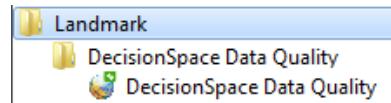
1. Ensure that the Server installation has been completed successfully.
2. Install the DecisionSpace Data Quality application on the client machine.
3. Cancel the DecisionSpace Data Quality Setup: License Key window that will automatically open up.
4. Copy and paste the dvconnections file (dvconnections.xml) from the server to the "..\Local DSDQ install\conf\DataVera" directory.
5. Run the DSDQ executable to start the application.

Starting the Data Quality Application

When the Data Quality application is installed on your computer, a shortcut to the application may be placed on your desktop. The relevant program group, Data Quality, also gets set up under the Windows Start menu giving you access to the Data Quality application.

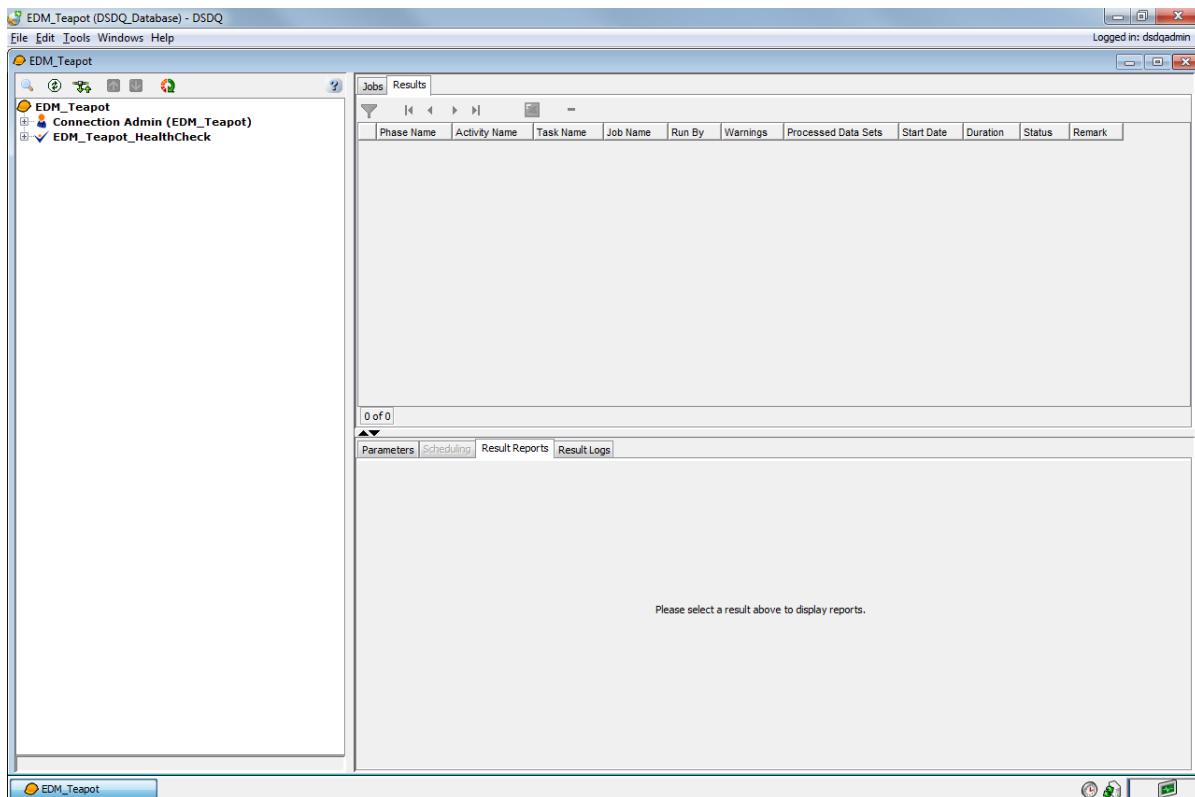
To start the Data Quality application:

1. Double-click the Data Quality icon  on the desktop or select **Start > All Programs > Landmark > DecisionSpace Data Quality > DecisionSpace Data Quality** from the Windows taskbar.

**Note**

If a shortcut is not created on your desktop already, you can create one by right-clicking the DSDQ.exe file and selecting Send to > Desktop (create shortcut) from the pop-up menu. The DSDQ.exe file is located in the 'local installed folder' directory.

The Data Quality Project Window displays.



To Start DSDQ from a Shared Network Drive

If the Data Quality application is set up on a shared network drive, you can access it from your machine (i.e. the client machine) in the following way:

1. Browse to the network drive that the Data Quality application is installed on. (Consult your system administrator if you do not know the location.)
2. Select the network path where '..\Installed DecisionSpace Data Quality' directory.
3. Right-click on the DSDQ.exe file and select **Create Shortcut** to create a shortcut and dragging the DSDQ shortcut to local desktop.

For Assistance

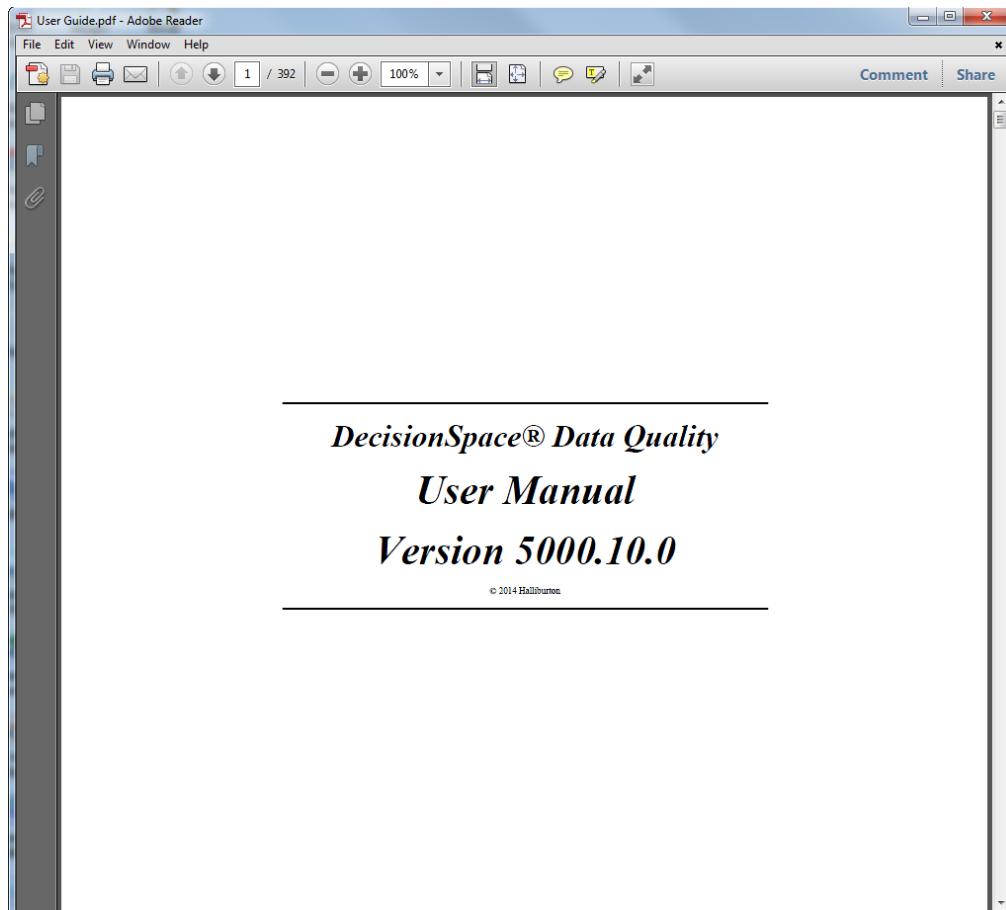
The Data Quality software comes with a user guide that provides an overview of the Data Quality software as well as step-by-step procedures for using the application. For additional explanation from one of Landmark's support specialists, call or e-mail the Landmark help desk.

Using the User Guide

There are multiple ways to access the Data Quality User Guide:

Click the **Help**  button wherever available for help. This brings up more information for the relevant window via the user guide.

- Access the User Guide by selecting **View Help** from the **Help** menu.



Contacting Landmark Customer Support

Landmark software operates Technical Assistance Centers (TACs) in Australia, the United Kingdom, and the United States. Additional support is also provided through regional support offices around the world.

- [Support via Web Portal](#)
- [Technical Assistance Centers](#)
- [Regional Offices](#)

Support via Web Portal

Support information is always available on the Landmark Customer Support internet page. You can also submit a support request directly to Landmark Customer Support through the Landmark Customer Support Portal:

<http://www.landmarksoftware.com/Pages/ContactSupport.aspx>

To request support in the Landmark Customer Support Portal:

1. In the **PIN** and **Password** text boxes in the Please Sign In area, enter your registered personal identification number and password.
2. Click the **Sign In** button.
3. In the Case & Defect Information area, click the **Create a New Case** link.
4. In the **Create Case** area, fill in the necessary information. Provide details about your technical concern, including any error messages, the workflow steps where the problem occurred, and attachments of screen shots that display the problem. To help understand the concern, you can also attach other files too, such as example data files.
5. Click the **Submit** button. A support analyst in the nearest Technical Assistance Center will respond to your request.

Technical Assistance Centers

Asia, Pacific	61-8-9481-4488 (Perth, Australia)
8:00 am - 5:00 pm Local Time	Toll Free 1-800-448-488
Monday-Friday, excluding holidays	Fax: 61-8-9481-1580 Email: apsupport@lge.com
Europe, Africa, Middle East	44-1372-868686 (Leatherhead, UK)
9:00 am - 5:30 pm Local Time	Fax: 44-1224-723260 (Aberdeen, UK)
Monday - Friday, excluding holidays	Fax: 44-1372-868601 (Leatherhead, UK) Email: support@lge.com
Latin America	713-839-3405 (Houston, TX, USA)
(Spanish, Portuguese, English)	Fax: 713-839-3646
7:00 am - 5:00 pm Local Time	Email: soporte@lge.com
North America	713-839-2200 (Houston, TX, USA)
7:30 am - 5:30 pm Central Standard Time	Toll Free 1-877-435-7542 (1-877-HELP-LGC)
Monday - Friday, excluding holidays	Fax: 713-839-2168 Email: support@lge.com

Regional Offices

For contact information for regional offices, see the Contact Support page located at:

<http://css.lgc.com/InfoCenter/index?page=contact§ion=contact>

If problems cannot be resolved at the regional level, an escalation team is called to resolve your incidents quickly.

Chapter 2

Connecting DecisionSpace Data Quality (DSDQ) with DecisionSpace Data Server (DSDS)

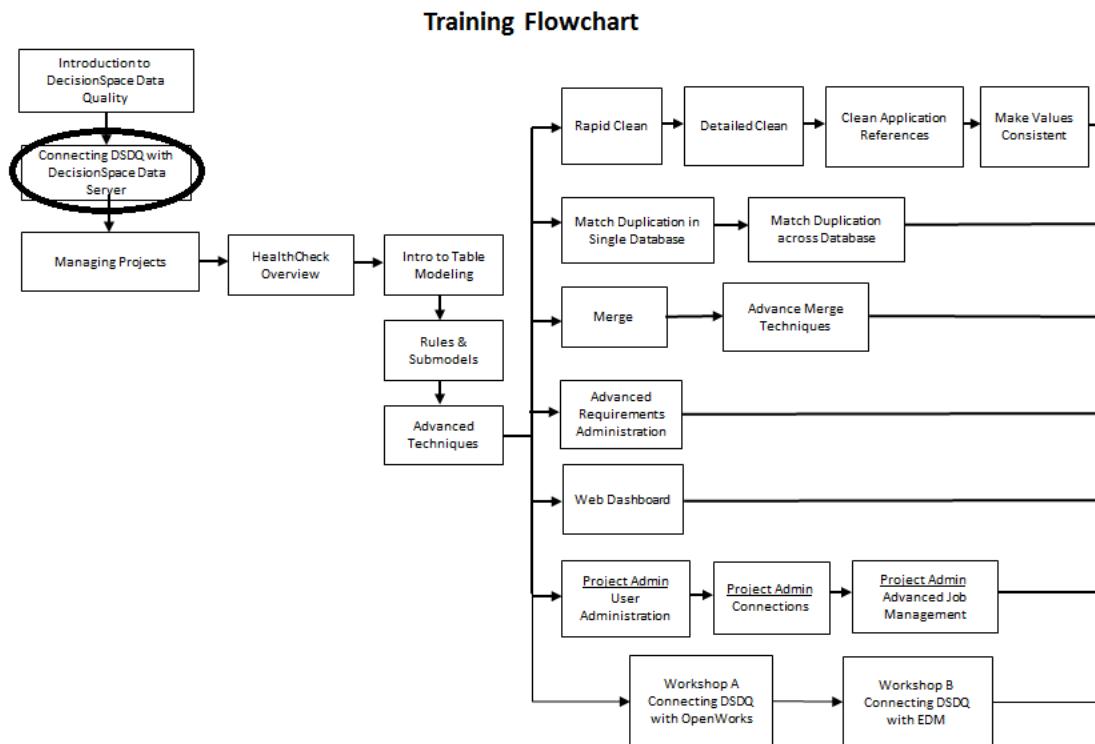
DecisionSpace Data Server allows the Data Quality application to connect to Landmark application databases, such as OpenWorks® and EDM™ and non-Landmark data sources. This enables the Data Quality application to automatically load preconfigured models and rules against these databases.

Chapter Overview

In this chapter, you will learn about:

- The DecisionSpace Data Server software including its advantages and key components
- Installing DecisionSpace Data Server
- Connecting DecisionSpace Data Quality (DSDQ) with DecisionSpace Data Server (DSDS)

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:

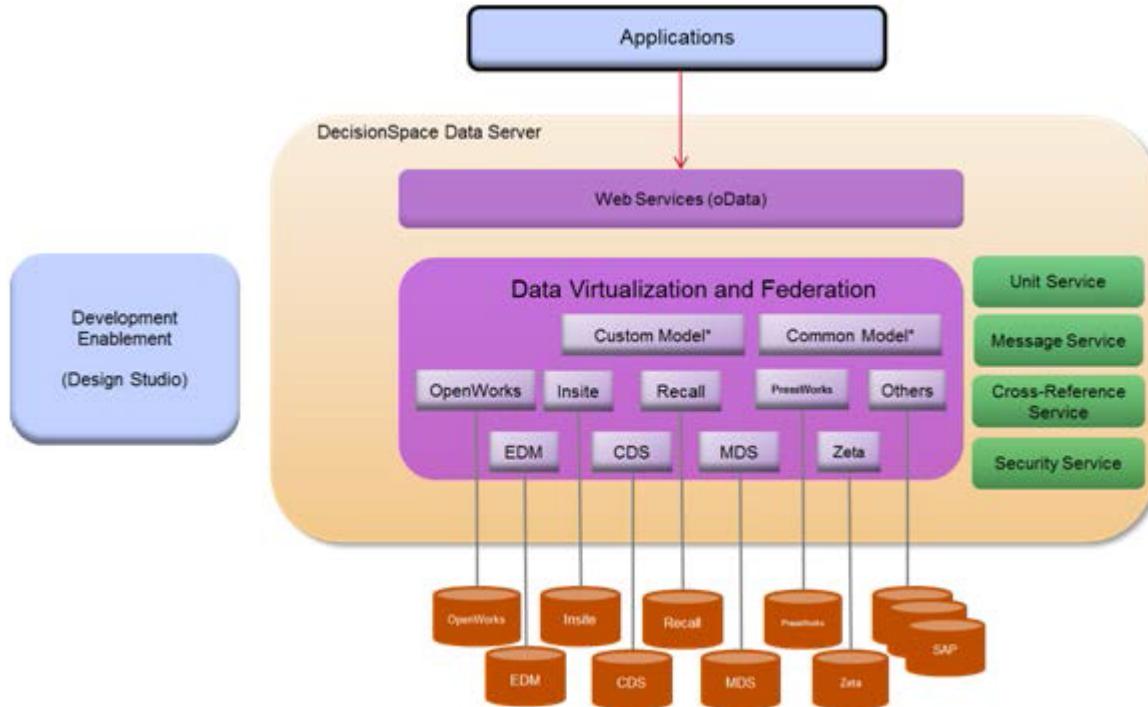


DecisionSpace Data Server: An Introduction

Decision Space Data Server provides applications with a common data access to Landmark and non-Landmark data sources. It enables access to data via services instead of individual development kits for different databases. DecisionSpace Data Server also provides tools for developers and consultants to create connections to additional data sources and expose the data as a service.

It is a simple, yet efficient tool that acts as a bridge between DecisionSpace Data Quality application and various data sources. The following features set DecisionSpace Data Server apart from other applications with similar functionality:

- Enterprise services for Exploration and Production data
- Data integration/virtualization from heterogeneous models and sources
- Data interoperability
- Open standards based web accessible data
- Integration with enterprise security for identity, authentication and authorization

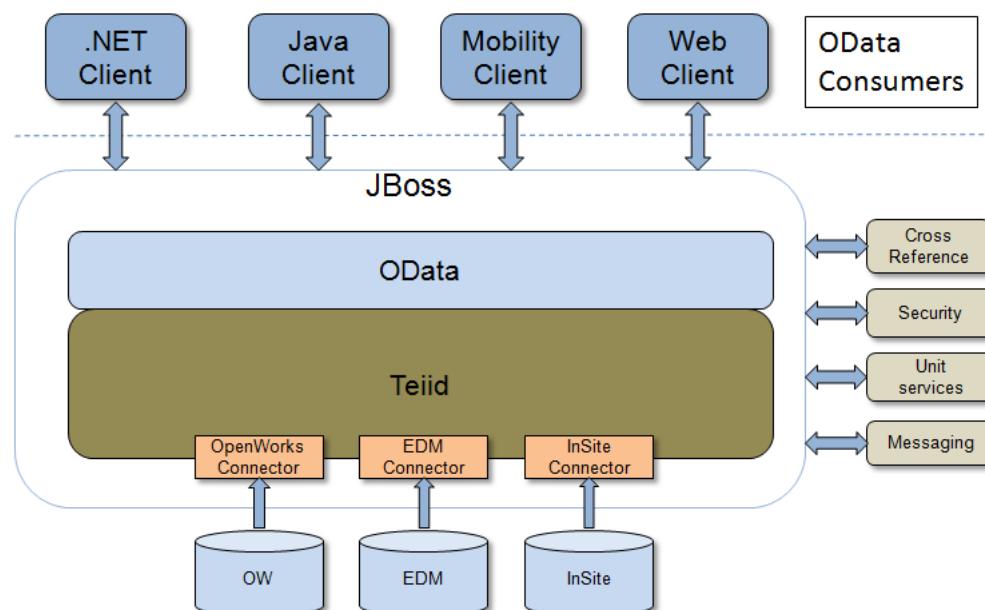


Key components of DecisionSpace Data Server are:

- **JBoss Application Server:** JBoss is a widely used application server. It is cross-platform and open source software written in Java that implements standards from J2EE. JBoss also supports the LDAP login module. To configure single sign on (SSO) using LDAP authentication in DecisionSpace Data Server, refer to the DecisionSpace Data Server System Administration Guide.
- **OData4j:** The Open Data Protocol (OData) is a web protocol for querying and updating data that provides a way to unlock your data by applying and building upon web technologies such as HTTP, Atom Publishing Protocol (AtomPub) and JSON to provide access to information from a variety of applications, services, and stores. OData4j is a new open-source toolkit for building first-class OData producers and first-class OData consumers in Java.
- **Teiid:** Teiid is a data virtualization system that allows applications to use data from multiple, heterogeneous data stores. It is a query engine for joining and unioning data from multiple sources in an optimal manner.
- **OData Consumers:** OData consumers are the clients written in .Net/Java and other scripting languages to remotely access the data

produced by the DecisionSpace data service. The data is published in two formats (ATOMPub and JSON) by service. Consumers can request data in either of the supported formats.

- **Security:** DecisionSpace Data Server can be configured to use several different authentication providers, including file-based, database-based, LDAP-based, or a combination of one or more of these. It can also be configured for single sign on with Active Directory-based authentication service. By default, a file-based security module is used.
- **Messaging:** DecisionSpace Data Server can publish transaction messages to any JMS based messaging broker.
- **Connectors:** Connectors are Teiid plugins that enable access to relational and non-relational data sources. Currently, DecisionSpace Data Server ships with connectors for the following data sources:
 - OpenWorks
 - EDM
 - Insite



DecisionSpace Data Server Layout

DecisionSpace Data Server Console has six tabs, located on the top-left corner of the console window:

Home: This is a default tab. This tab displays summary information about the contents of the other tabs.

The screenshot shows the Home tab of the DecisionSpace Data Server Console. The interface includes a header with the Halliburton logo and a log out link. Below the header are six tabs: Home, Data Sources, Deployment, Status, Measurement System, and Ports. The Home tab is selected. On the left, there are two sections: 'Deployment' (VDB Name: No records found) and 'Data Sources' (Example). On the right, there are three tables: 'Measurement Systems' (SPE Preferred Metric: US Oil Field), 'Ports' (Socket Binding Name: http, https, management-http, management-native; Port: 8080, 8443, 9990, 9999), and a 'Welcome' message for dssadmin.

Data Sources: This tab allows you to create groups in order to manage data sources and then generate VDBs (virtual database connections). To the left you see a tree of default Data Sources and any new additions are listed here. We recommend that you organize the data sources by group.

The screenshot shows the Data Sources tab of the DecisionSpace Data Server Console. The interface includes a header with the Halliburton logo and a log out link. Below the header are six tabs: Home, Data Sources, Deployment, Status, Measurement System, and Ports. The Data Sources tab is selected. On the left, there is a tree view of data sources under 'Data Sources'. Under 'Landmark', there are entries for EDM_TEAPOT, OW_TEAPOT, EDM_test, TEAPOT_Demo1, and Example. Under 'Others', there is a single entry for Example.

Deployment: This tab displays a list of deployed VDBs and their related details and status. This tab also allows the deletion of previously created VDBs or the addition or removal of data sources to an existing VDB.

The screenshot shows the Deployment tab of the DecisionSpace Data Server Console. The interface includes a header with the Halliburton logo and a log out link. Below the header are six tabs: Home, Data Sources, Deployment, Status, Measurement System, and Ports. The Deployment tab is selected. On the left, there is a section titled 'Deployed VDBs'. A table lists one entry: OpenWorks (Version: 5000.8.3, Teiid JDBC Url: jdbc:teiid:OpenWorks@mm://10.11.24.162:31000;version=1, Dynamic: true, Status: ACTIVE). On the right, there are 'Add' and 'Refresh' buttons.

Status: This tab allows you to check the server.log file, helpful for troubleshooting issues.

DecisionSpace Data Server Console 5000.10.0.0

HALLIBURTON | Landmark Software & Services

Welcome ddsadmin! [Log Out](#)

Component	Details
Licensing	Connected to the Licensing server
DSDServer	Active

[Refresh](#)

```

removed in future versions without notice
09:30:44,353 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONRootElementProvider$App' for service type [javax.ws.rs.ext.MessageBodyReader]
09:30:44,355 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONRootElementProvider$General' for service type [javax.ws.rs.ext.MessageBodyReader]
09:30:44,355 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONJAXBElementProvider$App' for service type [javax.ws.rs.ext.MessageBodyReader]
09:30:44,359 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONJAXBElementProvider$General' for service type [javax.ws.rs.ext.MessageBodyReader]
09:30:44,360 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONListElementProvider$App' for service type [javax.ws.rs.ext.MessageBodyReader]
09:30:44,360 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONListElementProvider$General' for service type [javax.ws.rs.ext.MessageBodyReader]
09:30:44,361 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONArrayProvider$App' for service type [javax.ws.rs.ext.MessageBodyReader]
09:30:44,361 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONArrayProvider$General' for service type [javax.ws.rs.ext.MessageBodyReader]
09:30:44,361 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONObjectProvider$App' for service type [javax.ws.rs.ext.MessageBodyReader]
09:30:44,363 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONObjectProvider$General' for service type [javax.ws.rs.ext.MessageBodyReader]
09:30:44,364 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONRootElementProvider$App' for service type [javax.ws.rs.ext.MessageBodyWriter]
09:30:44,365 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONRootElementProvider$General' for service type [javax.ws.rs.ext.MessageBodyWriter]
09:30:44,365 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONJAXBElementProvider$App' for service type [javax.ws.rs.ext.MessageBodyWriter]
09:30:44,366 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONJAXBElementProvider$General' for service type [javax.ws.rs.ext.MessageBodyWriter]
09:30:44,367 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONListElementProvider$App' for service type [javax.ws.rs.ext.MessageBodyWriter]
09:30:44,368 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONListElementProvider$General' for service type [javax.ws.rs.ext.MessageBodyWriter]
09:30:44,368 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONArrayProvider$App' for service type [javax.ws.rs.ext.MessageBodyWriter]
09:30:44,369 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONArrayProvider$General' for service type [javax.ws.rs.ext.MessageBodyWriter]
09:30:44,370 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONObjectProvider$App' for service type [javax.ws.rs.ext.MessageBodyWriter]
09:30:44,372 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONObjectProvider$General' for service type [javax.ws.rs.ext.MessageBodyWriter]
09:30:44,373 WARN [org.jboss.as.server.deployment] (MSC service thread 1-4) JBAS015893: Encountered invalid class name 'com.sun.jersey.json.impl.provider.entity.JSONRootElementWriter'
09:30:45,216 WARN [org.jboss.weld.deployer] (MSC service thread 1-7) JBAS016012: Deployment deployment "ddataserver-console.war" contains CDI annotations but beans.xml was not found.
09:30:51,869 WARN [org.jboss.weld.deployer] (MSC service thread 1-5) JBAS016012: Deployment deployment "ddataserver.war" contains CDI annotations but beans.xml was not found.
10:15:38,053 WARNING [com.lgc.ow.pr.PRJSessionImpl] (JCA PoolFiller) MUID: 418988928: Reopening closed project: TEAPOT_DEMO
11:00:38,124 WARNING [com.lgc.ow.pr.PRJSessionImpl] (JCA PoolFiller) MUID: 418988928: Reopening closed project: TEAPOT_DEMO
11:45:38,215 WARNING [com.lgc.ow.pr.PRJSessionImpl] (JCA PoolFiller) MUID: 418988928: Reopening closed project: TEAPOT_DEMO

```

Measurement System: This tab shows the list of measurement systems with the associated unit types and units. This window allows a user to create custom measurement systems.

DecisionSpace Data Server Console 5000.10.0.0

HALLIBURTON | Landmark Software & Services

Welcome dssadmin! [Log Out](#)

Default Measurement System:

Description:

Measurement Systems: SPE Preferred Metric [New](#) [Edit](#) [Delete](#)

Unit Type	Unit
API	api
API Oil Gravity	degsapi
Abrasive Volume	cubic metres
Abs atmos press	lbs / ft ² abs
Abs pressure	lbs / ft ² abs
Absolute Volume	m ³ _kg
Acceleration	m_sec ²
Acidity	pH
Acou Attenu	Decibels
Acou Impedance	Mayls
Acou Velocity	feet per sec
Acoustic Freq	Hertz
Acoustic vel	metre per sec
Activity time	hours
Additive Cal	feet ³ per deg
Additive mass	pound mass
Additive volume	cubic feet
Agitator setpt	ds/pcf
Air Motor Calib	rev/m ³ /sec
Air Pressure	lbs per sq ft

Ports: This tab shows the port settings used during the initial DecisionSpace Data Server installation. This window allows customizing the port settings, if necessary.

DecisionSpace Data Server Console 5000.10.0.0

HALLIBURTON | Landmark Software & Services

Welcome dssadmin! [Log Out](#)

Socket Binding Group: dsds-sockets

Port Offset: 0 [Edit](#)

Socket Bindings

Socket Name	Effective Port(including offset)	Port	Actions
http	8080	8080	Edit
https	8443	8443	Edit
jacorb	3528	3528	Edit
jacorb-ssl	3529	3529	Edit
jmx-connector-registry	1090	1090	Edit
jmx-connector-server	1091	1091	Edit
management-http	9990	9990	Edit
management-native	9999	9999	Edit
messaging	5445	5445	Edit
messaging-throughput	5455	5455	Edit
osgi-http	8090	8090	Edit
remoting	4447	4447	Edit
teiid-jdbc	31000	31000	Edit
teiid-odbc	35432	35432	Edit
txn-recovery-environment	4712	4712	Edit
txn-status-manager	4713	4713	Edit

DecisionSpace Data Server Installation

This section outlines the requirements and procedures for installing DecisionSpace Data Server on a workstation.

System Requirements

Before you start installing DecisionSpace Data Server, please ensure that your workstation meets the following requirements:

Data Server	
Resource	System Requirements
Operating System	<ul style="list-style-type: none">Windows 2008 R2 Server x64Windows 7 x64RHEL AS 6 (Primary)RHEL AS 5.x (Secondary)
Third-Party Software (packaged with Data Server)	<ul style="list-style-type: none">JBoss Enterprise Application Platform (EAP) ver. 6.0.1 Alpha 1Teiid 8.4.0 FinalOData 3.0JRE 1.6.0 64 BitDecisionSpace Messaging (ActiveMQ 5.7.0)DecisionSpace Data Server Design Studio (Teiid Designer 8.2) available using the Custom install option
Landmark Software Dependencies	<ul style="list-style-type: none">LAM 5000.0.3 (Flex 11.7)Licensed using FlexLM DSDATASERVER ver. 5000.8 which is read from the system environment

Data Source-specific	<ul style="list-style-type: none"> • OpenWorks Connector — Oracle Client 11.2.0.2 Administrative version 64 bit — OpenWorks 5000.8.3 Client • EDM Connector — EDM 5000.1.0 installed with 5000.1.10 Update • Insite Connector — InSite Server — JNBridge Pro 6.1.x (prerequisite: VC++ 2010 redist x86 and .Net 4.0)
Client^a	
JAVA Client	<ul style="list-style-type: none"> • JAVA Client Binding (JRE 1.6 64 Bit)
.NET Client	<ul style="list-style-type: none"> • .NET Client Binding (4.0)
OData Client	<ul style="list-style-type: none"> • OData v3.0
Browsers	
Browsers	<ul style="list-style-type: none"> • Internet Explorer 9 or higher • Firefox 20.x or higher • Google Chrome
General	
Minimum Java Heap Size	<ul style="list-style-type: none"> • 1024
Minimum Memory Size	<ul style="list-style-type: none"> • 2048

a.DecisionSpace Data Server users will write their own OData consumers, which are clients written in .Net/Java and other scripting languages to remotely access the data produced by the DecisionSpace data service.

Installing DecisionSpace Data Server

In this exercise, you will be installing DecisionSpace Data Server 5000.10.0 locally under Windows 7 x64 Platform, during the installation procedure the following third-Party software will be installed:

- JBoss Enterprise Application Platform (EAP) ver. 6.0.1 Alpha 1
- Teiid 8.4.0 Final
- OData 3.0
- JRE 1.6.0 64 Bit
- DecisionSpace Messaging (ActiveMQ 5.7.0)
- DecisionSpace Data Server Design Studio (Teiid Designer 8.2) (available using the Custom install option)

The following Landmark Software Dependencies must be verified in order to run DecisionSpace Data Server after installation:

- LAM 5000.0.3 (Flex 11.7)
- Licensed using FlexLM DSDATASERVER ver. 5000.8(The instructor will tell you where the license file is located)

This procedure must be performed by a domain user with administrator rights on the computer where the application is to be installed. Prior to initiating installation, you must ensure that there is at least 1 GB of space available in the installation directory.

To install the DecisionSpace Data Server software:

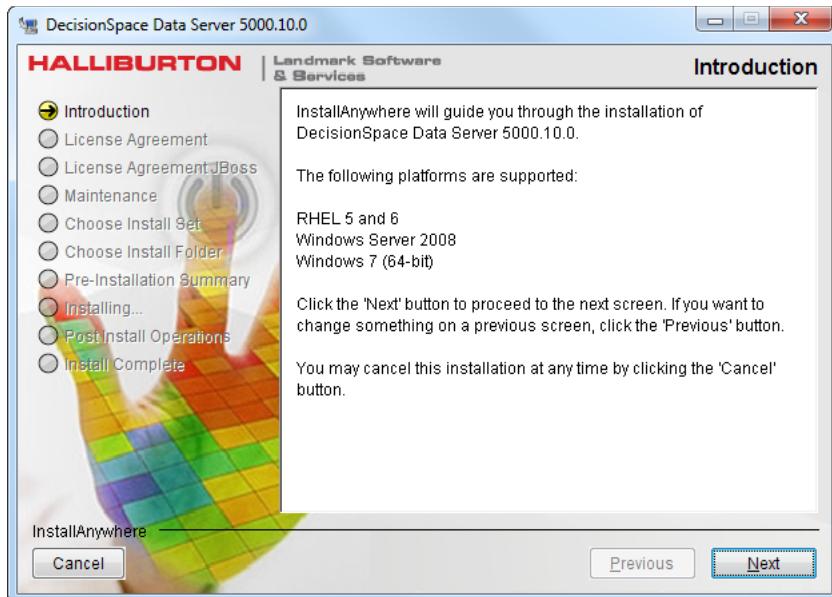
1. Open a terminal session and navigate to the directory where the Data Server installer resides (For example: **\Landmark\DSDataServer5000.10.0.0**).
2. Depending on the operating system on which you have to install the application, enter one of the following commands in the terminal window:
 - **On Windows:** DSDS_5000_10_0_0_Win.exe

- **On Linux:** ./DSDS_5000_10_0_0_RHL.bin

Note

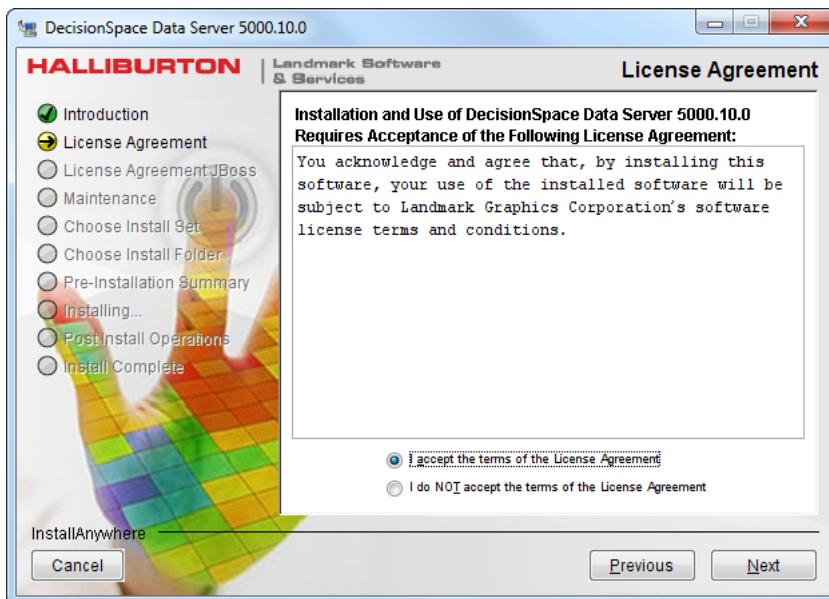
If you are installing DSDS on Linux, please ensure that Java 1.6.0 is loaded and that the \$JAVA_HOME environment variable is set in your environment.

The installer launches and the **Introduction** screen appears.



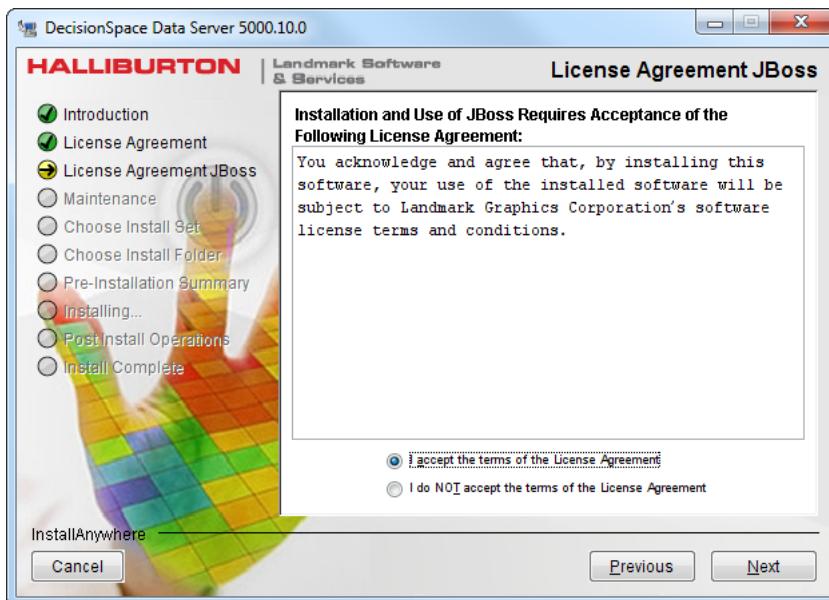
3. Review the information in the **Introduction** screen and click **Next** to continue.

The **License Agreement** screen appears.

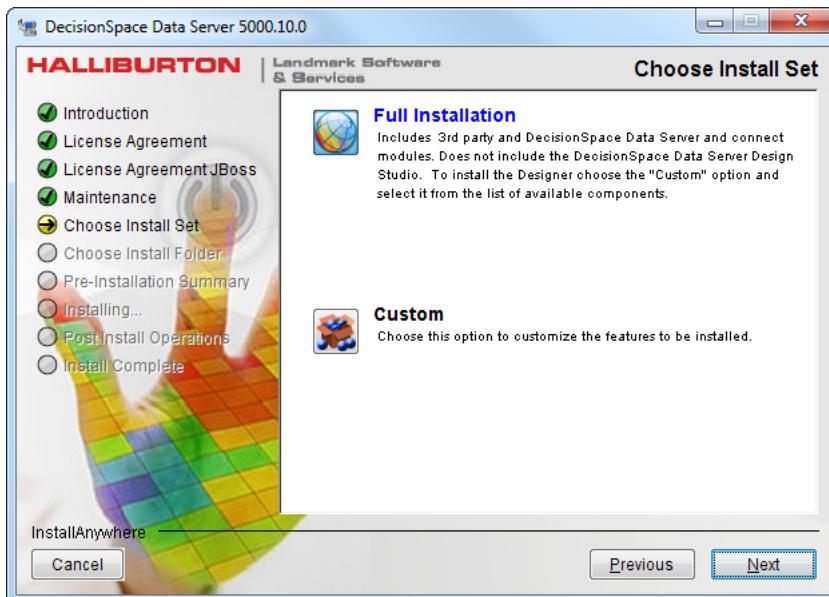


4. Review the software license agreement information given in the **License Agreement** screen and select **I accept the terms of the License Agreement** option.
5. Click **Next** to continue.

The **License Agreement JBoss** screen appears.



6. Review the JBoss software license agreement information given in the **License Agreement JBoss** screen and select **I accept the terms of the License Agreement** option.
7. Click **Next** to continue.
The **Choose Install Set** screen appears.



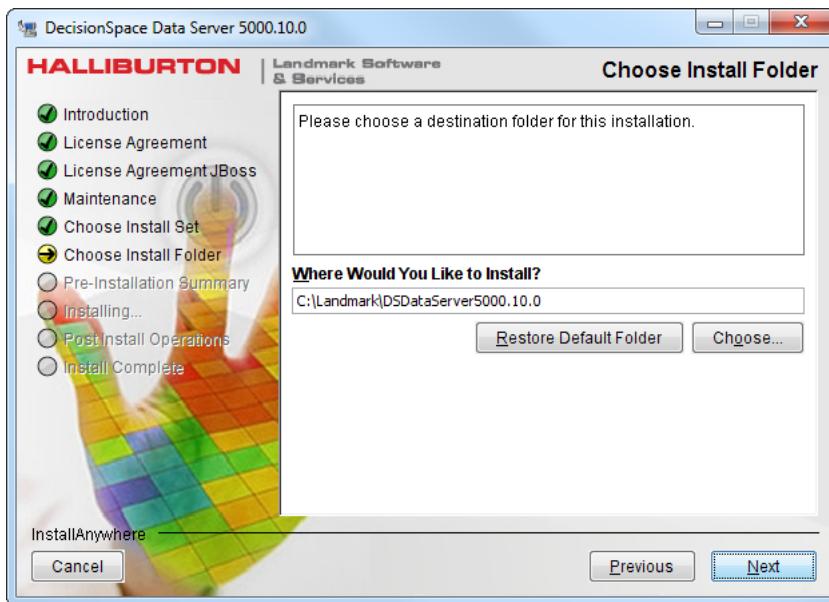
8. Select **Full Installation** to install all software modules required for running DecisionSpace Data Server. These include the Data Server, OpenWorks Connector, EDM Connector, InSite Connector and DecisionSpace Messaging.

Note

Custom installation is selected to install a subset of the components. Currently the options include OpenWorks Connector, EDM Connector, InSite Connector, DecisionSpace Messaging, and DecisionSpace Data Server Design Studio. For more information, refer to the DSDS_5000.10.0_Release_InstallGuide.

9. Click **Next** to continue.

The **Choose Install Folder** screen appears.

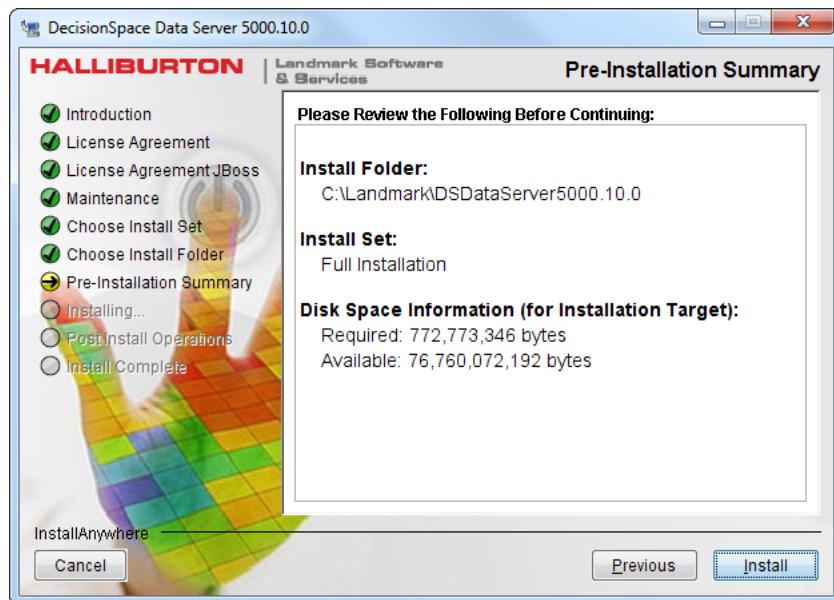


10. Install DSDS at the default location by clicking **Next**.

Note

If you are installing on Linux and you want to install DSDS on a different location than the default: Select "Choose...", navigate to the desired location, select "OK". The "Where Would You Like To Install?" line now shows your selection. Make sure to add"/DSDataServer5000.10.0" (or similar) to the path (for example: /apps/lge/R5000/DSDataServer5000.10.0). The "DSDataServer5000.10.0" folder will be created by the installer.

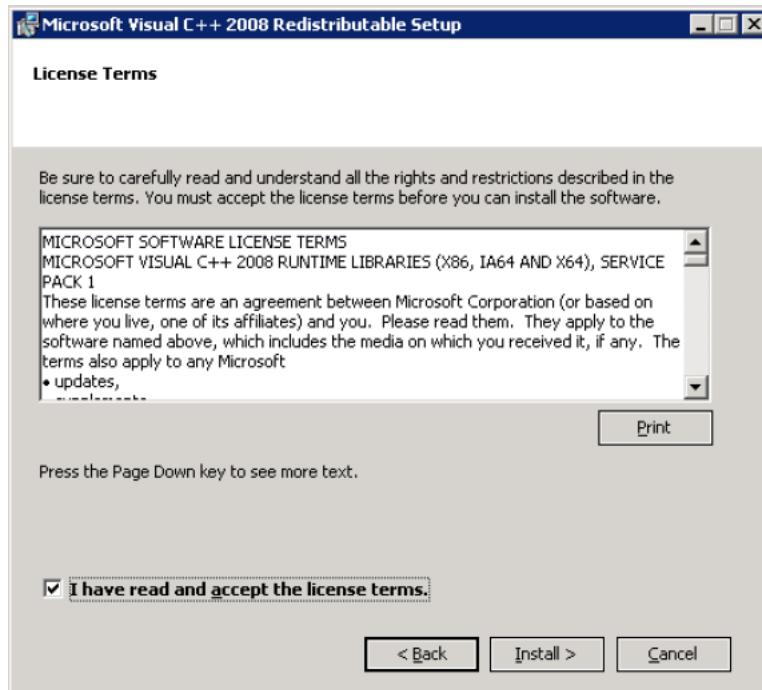
The **Pre-Installation Summary** screen appears.



11. Verify the information shown in the **Pre-Installation Summary** screen (**Required** and **Available** disk space may be different than the illustration above) and click **Install**.

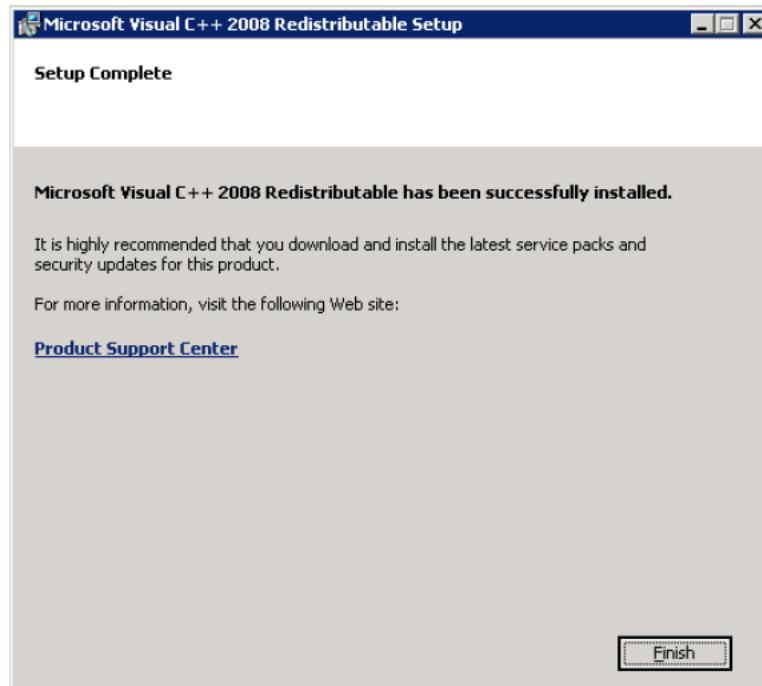
The installer checks if **Microsoft Visual C++ 2008** is installed on the system. If Visual C++ is located, the **License Information** screen appears (go to step 14). If Visual C++ is not located, The **Microsoft Visual C++ 2008 Redistributable Setup** wizard

launches and the **Welcome** screen appears followed by the **License Terms** screen.



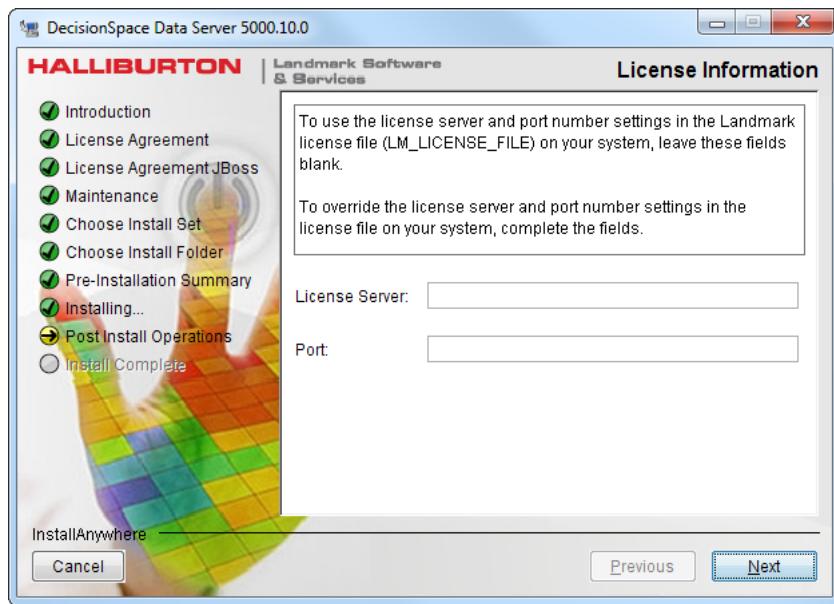
12. Review the software license terms information given in the **License Terms** screen, select **I have read and accept the license terms** option and click **Install**.

The **Setup Complete** screen appears.



13. Click **Finish** to close the **Microsoft Visual C++ 2008 Redistributable Setup** wizard.

The installation process for the DecisionSpace Data Server resumes with the **License Information** screen.



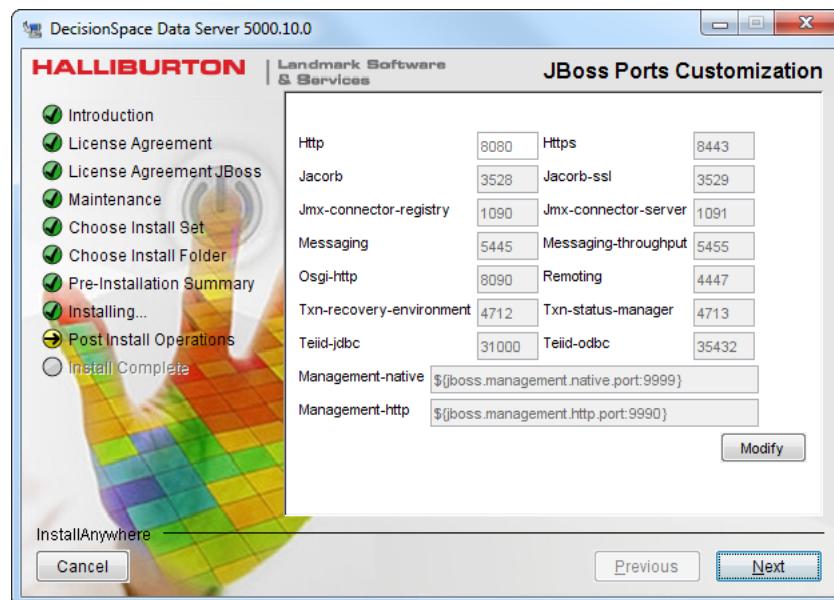
14. As appropriate, perform one of the following procedures:

- To use the license server and port number settings in the Landmark license file (LM_LICENSE_FILE) on your system, leave these fields blank and click **Next**, or
- To override the license server and port number settings in the license file on your system, enter the **License Server** and **Port** number, and then click **Next**.

Note

Your instructor will tell you where the license file is located.

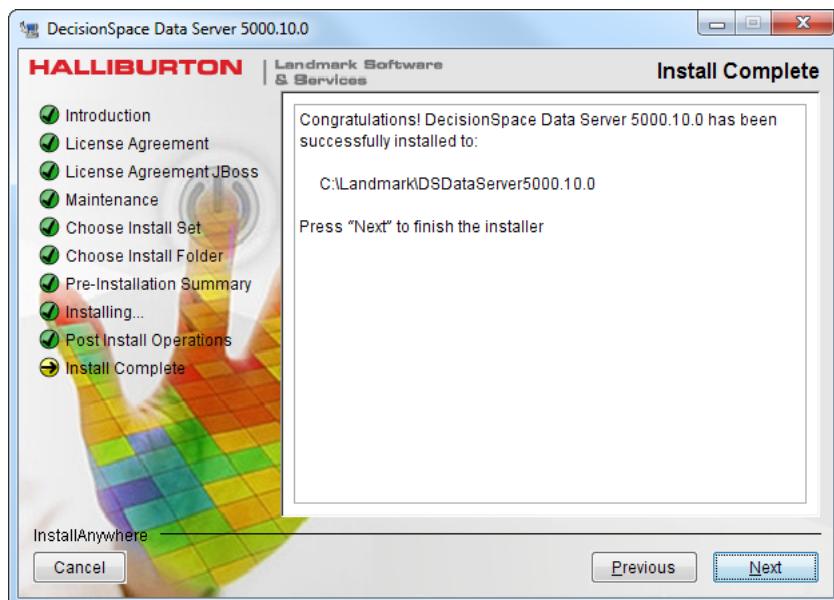
The **JBoss Ports Customization** screen appears.



15. As appropriate, perform one of the following procedures:
 - Click **Next** to accept the default ports, or
 - Click **Modify** to change the ports followed by **Validate Ports** to check the validity of the ports.

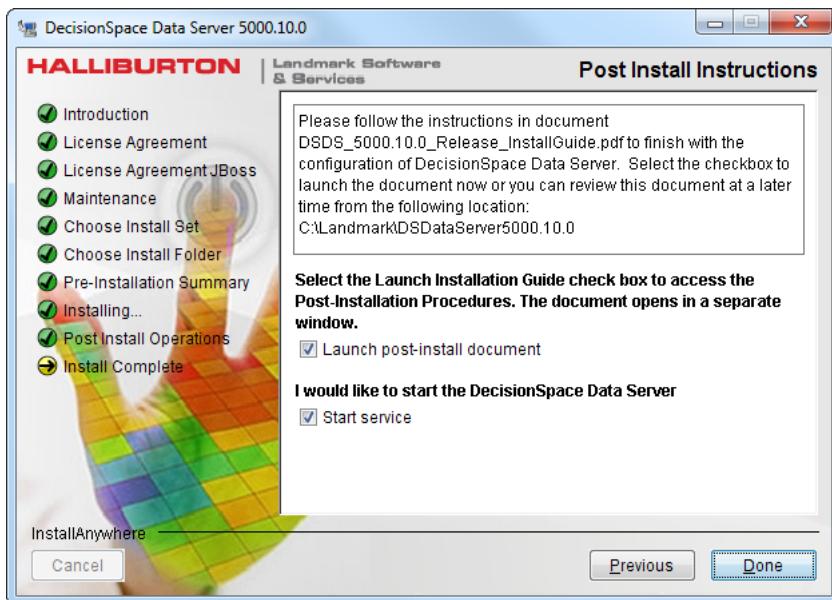
16. Click **Next** to continue.

A progress screen displays during installation. Once the installation is complete, the **Install Complete** screen appears.



17. Click **Next** to continue.

The **Post Install Instructions** screen appears.



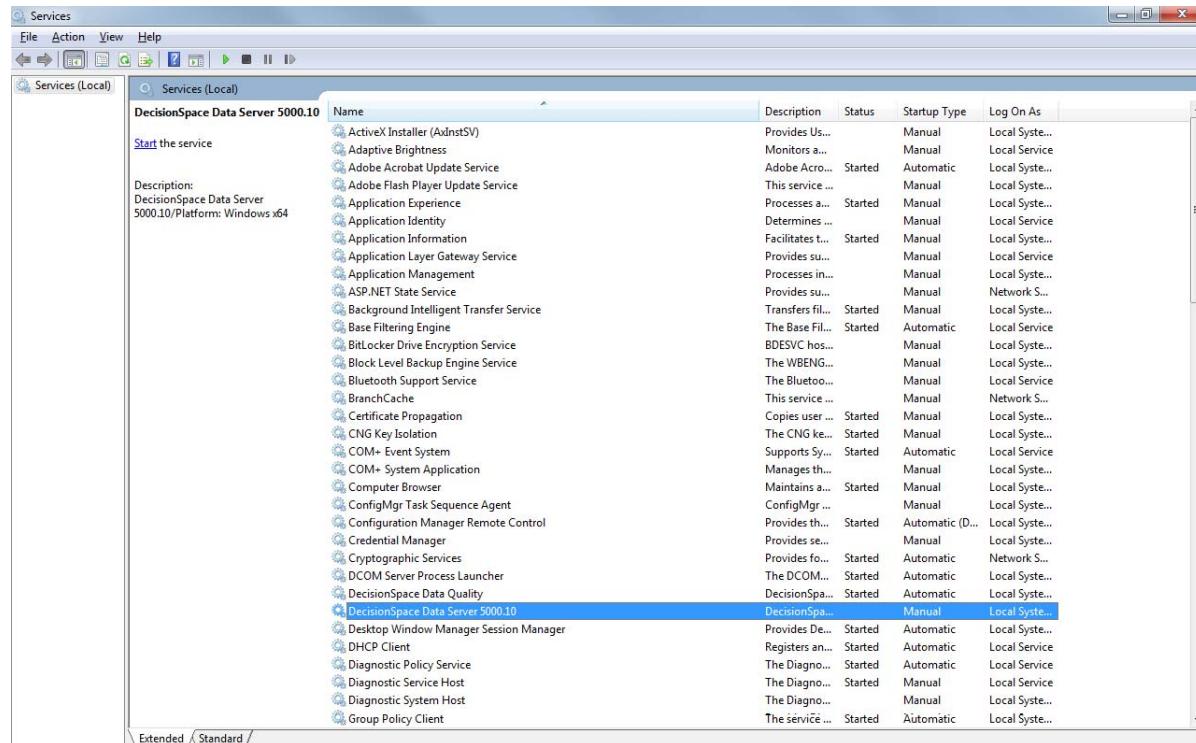
18. Select the **Start service** check-box to start the **JBoss** server.

19. Click **Done** to exit the **installer**.

Starting the DecisionSpace Data Server Service

On Windows:

1. Enter **services** in the Search box on the **Start** menu and then press **<Enter>** on your keyboard.
The **Services** screen appears.



2. Select **DecisionSpace Data Server 5000.10.0**.
3. Click **Start the service** to start the DecisionSpace Data Server service.
4. Double-click on the service to change the **Startup Type** to **Automatic**.

Note

The DSDS Service in Windows can also be launched from a command window, by executing the following script:

```
<DSDS_INSTALL_HOME>
>\ApplicationServer\bin\runDSDS.bat
```

On Linux:

From the terminal, execute the following script:
<DSDS_INSTALL_HOME>/bin/runDSDS.sh

Note

DO NOT close the window if using this method. If you wish the program to run in the background use the ‘&’ sign at the end of the command which will allow you to continue to use the same window for other functions.

On both platforms:

If you are running the **runDSDS** script, wait for a few moments and verify that all JBoss modules are deployed (as shown in the illustration below). Ensure that Teiid VDBs are set to **ACTIVE**.

```
INFO [org.jboss.as.server] <DeploymentScanner-threads - 2> JBAS018559: Deployed "teiid-connector-ws.rar"
INFO [org.jboss.as.server] <DeploymentScanner-threads - 2> JBAS018559: Deployed "teiid-connector-salesforce.rar"
INFO [org.jboss.as.server] <DeploymentScanner-threads - 2> JBAS018559: Deployed "teiid-connector-ldap.rar"
INFO [org.jboss.as.server] <DeploymentScanner-threads - 2> JBAS018559: Deployed "teiid-connector-file.rar"
INFO [org.jboss.as.server] <DeploymentScanner-threads - 2> JBAS018559: Deployed "powerhubjee.ear"
INFO [org.jboss.as.server] <DeploymentScanner-threads - 2> JBAS018559: Deployed "phdic-vdb.xml"
INFO [org.jboss.as.server] <DeploymentScanner-threads - 2> JBAS018559: Deployed "ow-0WDSDS_FLOUNDER-vdb.xml"
INFO [org.jboss.as.server] <DeploymentScanner-threads - 2> JBAS018559: Deployed "edn-DSD690_ORACLE-vdb.xml"
INFO [org.jboss.as.server] <DeploymentScanner-threads - 2> JBAS018559: Deployed "dsdataserver.war"
```

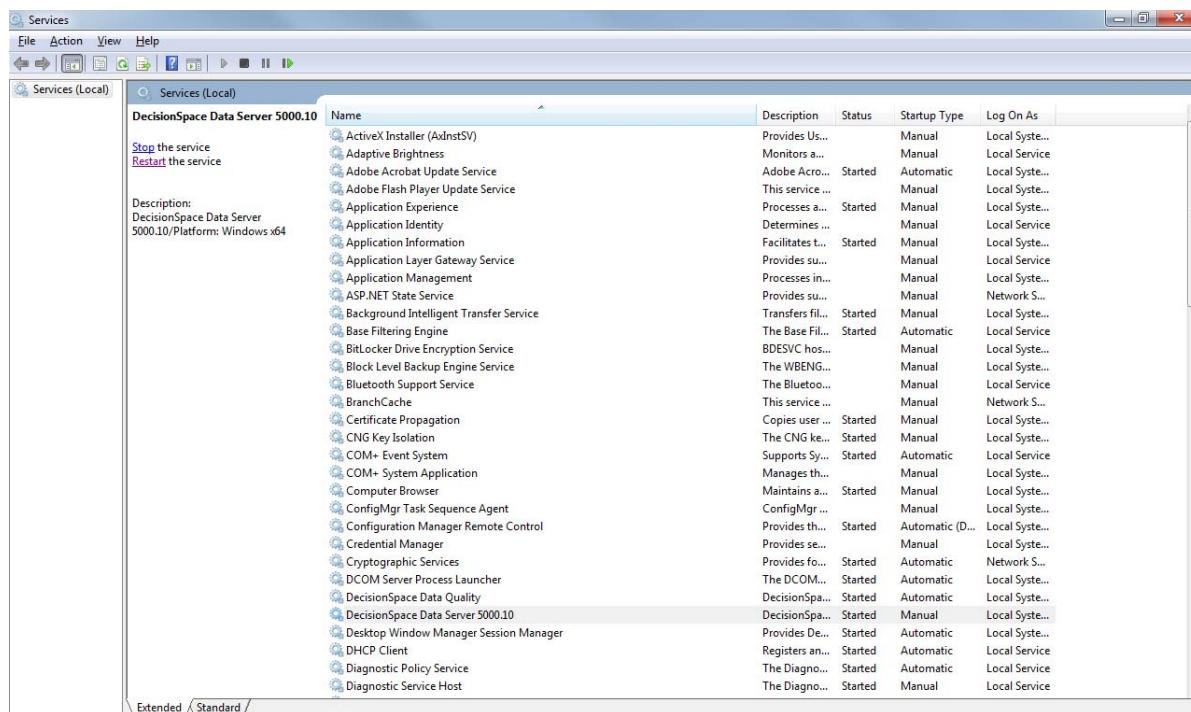
```
INFO [org.teiid.RUNTIME] <teiid-async-threads - 1> TEIID50030 UDB OW.1 model "5000_10" metadata loaded.
INFO [org.teiid.RUNTIME] <teiid-async-threads - 1> TEIID40003 UDB OW.1 is set to ACTIVE
```

Exercise: Stopping the DecisionSpace Data Server Service

To stop the DecisionSpace Data Server service:

On Windows:

1. Enter **services** in the Search box on the **Start** menu and then press **<Enter>** on your keyboard.
The **Services** screen appears.



2. Select **DecisionSpace Data Server 5000.10.0**.
3. Click **Stop the service** to stop the DecisionSpace Data Server service.

On Linux:

Terminate the process by entering **<Ctrl+C>** in the terminal running the service.

Post-Installation Procedures

Once the DecisionSpace Data Server is installed and configured, it can connect to a few internal development data sources (currently shipped as samples). In order to use the services against your own data, the following steps are necessary.

- Add a Data Source
- Test the Connection to the Data Source
- Generating VDBs (Virtual Databases)

Once the DSDS service is running, the DSDS Admin Console can be accessed using one of these methods:

- Start Menu -> All Programs -> Landmark -> DecisionSpace Data Server 5000.10.0 -> Start DecisionSpace Data Server Console
- Open the Web browser and enter the following URL:
 - http://<server_name>:8080/dsdataserver-console, or
 - <http://localhost:8080/dsdataserver-console> (if it is installed locally)

When prompted for credentials, the default user name and password for the console is **dsdsadmin**.

Exercise: Adding a Data Source

Prior to adding a data source and generating your VDB (virtual database), collect the necessary information for the type of data source you want to add. In order to determine what is necessary, highlight the data source type you wish to create and click **Create**. The list of expected parameters is displayed.

For example, if you are adding a **JDBC EDM-SQLServer** data source, you will be asked for the following parameters:

```
jdbc:sqlserver://  
[serverName[instanceName] [:portNumber]] [;property=value[;  
property=value]]
```

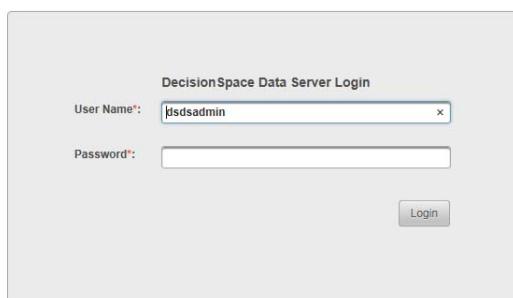
Where:

jdbc:sqlserver://localhost\EDM5000:0;DatabaseName=EDMDB

To add a data source:

1. Open your internet browser and enter the following URL in the address bar: [http://<Server-IP Address: Port>/dsdataserver-console \(e.g. http://localhost:8080/dsdataserver-console\)](http://<Server-IP Address: Port>/dsdataserver-console (e.g. http://localhost:8080/dsdataserver-console)).

The **Authentication Required** window appears.



2. Enter **dsdsadmin** in the **User Name** and **Password** fields and click **Login**.

The DecisionSpace Data Server Console window appears with the **Home** tab selected by default.

DecisionSpace Data Server Console 5000.10.0.0

HALLIBURTON | Landmark Software & Services

Welcome dsdsadmin! [Log Out](#)

Home Data Sources Deployment Status Measurement System Ports

Deployment		Details
VDB Name		No records found.

Data Sources		Details
▼ Data Sources		Example

Measurement Systems		Details
Name		
SPE Preferred Metric		
US Oil Field		

Ports		Details
Socket Binding Name	Port	
http	8080	
https	8443	
management-http	9990	
management-native	9999	

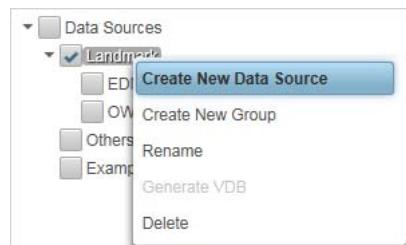
3. Select the **Data Sources** tab.

A list of existing data source types appears on the left side of the DecisionSpace Data Server Console window.

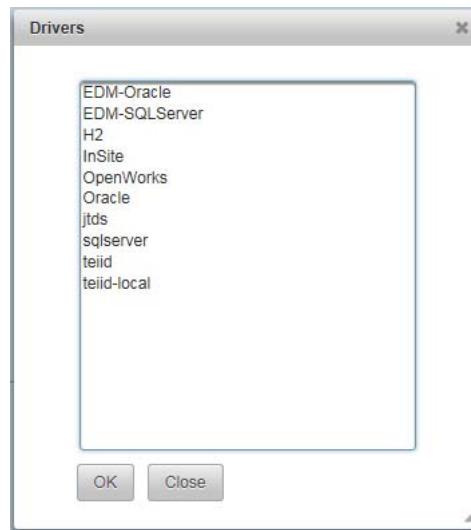


4. Select the **Landmark** data source check-box.

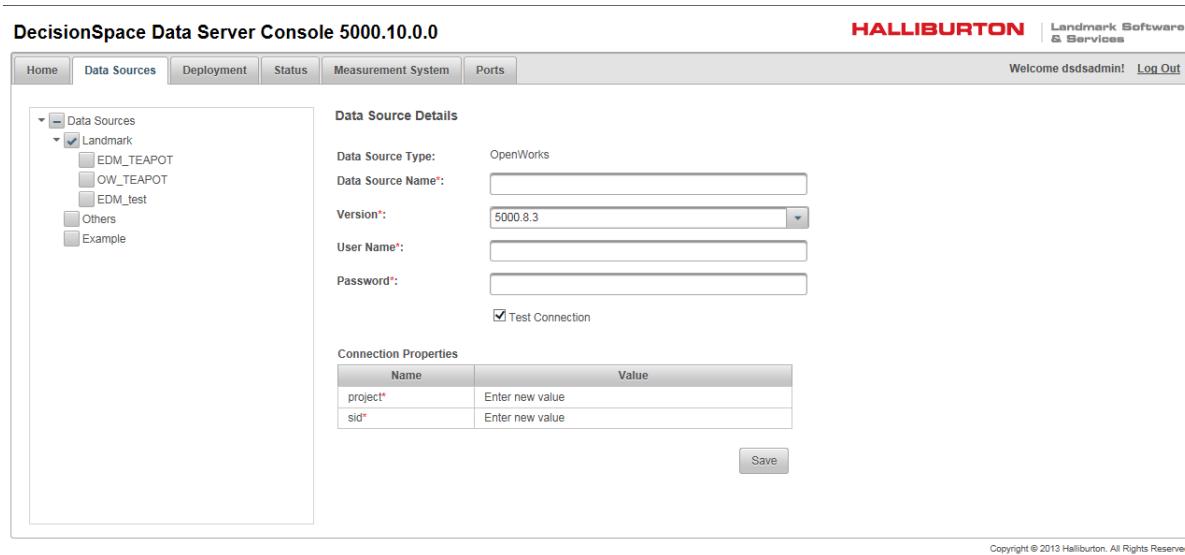
5. Right-click the **Landmark** data source and select **Create New Data Source** from the context menu.



The **Drivers** window appears.

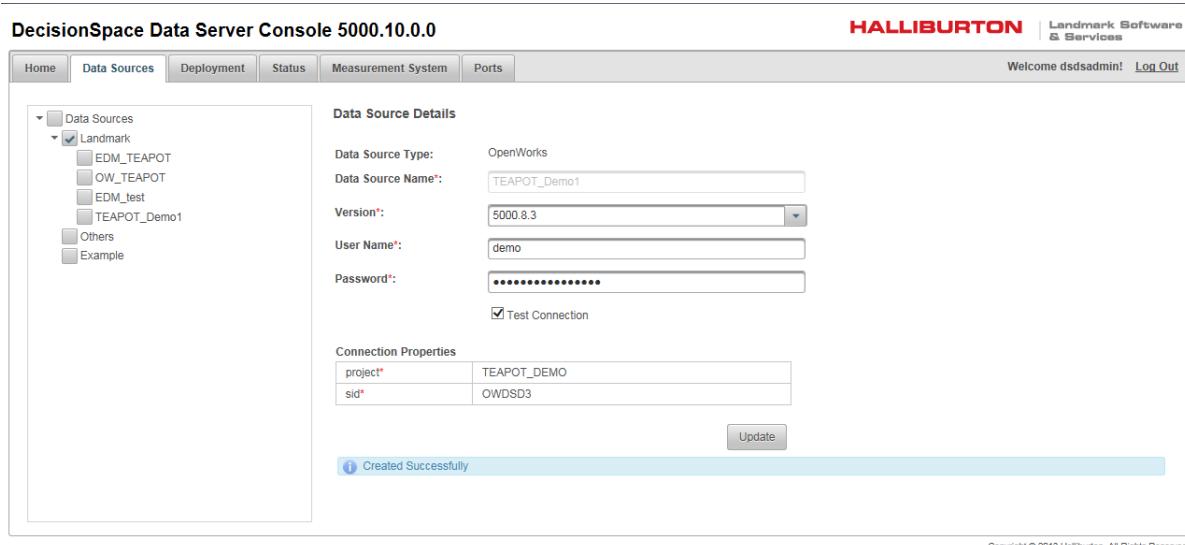


6. Select the **OpenWorks** data source and click **OK**.
The **Data Source Details** window appears.



7. Enter **TEAPOT_Demo1** in the **Data Source Name** field.
8. Enter **demo** and **demo123** in the **Username** and **Password** fields respectively.
9. In the **Connection Properties** group box, enter **TEAPOT_DEMO** in the **project** field and **OWDSD3** in the **sid** field.
10. Click **Save**.

The new data source displays in the tree list on the left side of the DecisionSpace Data Server Console window.



Testing the Connection to the Data Source

To test a connection of the data source that you have just added:

1. Select the **Data Sources** tab.
2. Select the **TEAPOT_Demo1** check-box from the list on the left side of the DecisionSpace Data Server Console.
3. Right-click the **TEAPOT_Demo1** option and select **Test** from the context menu.

A message appears on the DecisionSpace Data Server Console window displaying the status of the connection.

The screenshot shows the 'Data Sources' tab selected in the top navigation bar of the 'DecisionSpace Data Server Console 5000.10.0.0'. On the left, a tree view shows 'Data Sources' expanded, with 'Landmark' selected, revealing sub-options like 'EDM_TEAPOT', 'OW_TEAPOT', 'EDM_test', and 'TEAPOT_Demo1' (which has a checked checkbox). The main panel displays 'Data Source Details' for 'TEAPOT_Demo1', including fields for 'Data Source Type' (OpenWorks), 'Data Source Name' (TEAPOT_Demo1), 'Version' (5000.8.3), 'User Name' (demo), and 'Password' (redacted). A 'Test Connection' checkbox is checked. Below this is a 'Connection Properties' table with rows for 'project*' (TEAPOT_DEMO) and 'sid*' (OWDSD3). At the bottom right is an 'Update' button, and a blue status bar at the bottom indicates 'Connection test is successful.'

Generating Virtual Databases (VDBs)

You can create single and multi-source VDBs. Multi-source VDBs must use the same data source type as you cannot mix data source types. To select two or more data sources, press and hold the <CTRL> key on your keyboard and click on each data source in the **Data Sources** tree to select it.

Note

When generating a VDB for data sources listed under the “Others” section, it is necessary to gather more information about model, group and schema, prior to generating the actual VDB. For more details on the required information, see the “Generate VDB – Others” section below.

To create a single-source VDB for the data source that you have just added i.e. **TEAPOT_DEMO1**.

1. Select the **Data Sources** tab.
2. Select the **TEAPOT_Demo1** check-box from the list on the left side of the DecisionSpace Data Server Console.
3. Right-click the **TEAPOT_Demo1** option and select **Generate VDB** from the context menu.
A message appears on the DecisionSpace Data Server Console displaying the status of the connection



Errors (if any) generated during this process are stored in the **server.log** file. Select the **Status** tab or browse to the **JBOSS_HOME/standalone/log** directory to view this file.

Note

After deployment of a VDB on the server, the data sever will try to connect to the data source(s) selected for VDB generation and load the metadata of those data sources (tables, relations etc). It may take a few moments before the data becomes available. If anything goes wrong during this process, the VDB will not get deployed. To view the full details of a deployment, select the **Deployment** tab.

4. To check if the deployment was successful, click on the **Deployment** tab.

The OpenWorks connection displays in the **Deployment** tab.

VDB Name	Version	Teiid JDBC Url	Dynamic	Status	Actions
OpenWorks	5000.8.3	jdbc:teiid:OpenWorks@mm://10.11.24.162:31000;version=1	true	ACTIVE	Delete Details

5. Click on **Details** to see more details about the deployment of the **VDB**.

Note

If you want to delete a **VDB**, click **Delete**. The **VDB** is deleted along with the connection information.

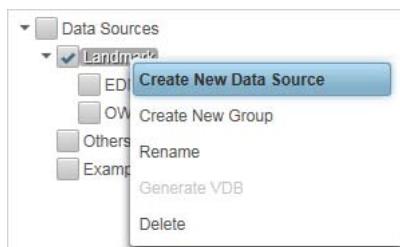
Exercise: Creating Database Connections

The purpose of this exercise is to create, test, generate & deploy the VDB and it also shows deletion of database connections in DSDS. The databases used in this exercise are OpenWorks and EDM:

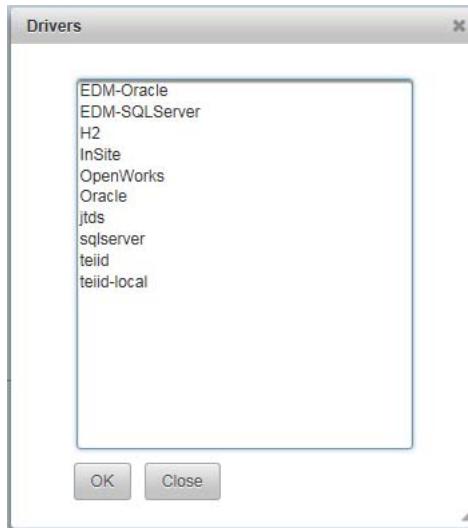
1. Select the **Data Sources** tab of the DecisionSpace Data Server Console.
- A list of existing data source types appears.

2. Select the **Landmark** data source check-box.

3. Right-click the **Landmark** data source and select **Create New Data Source** from the context menu.



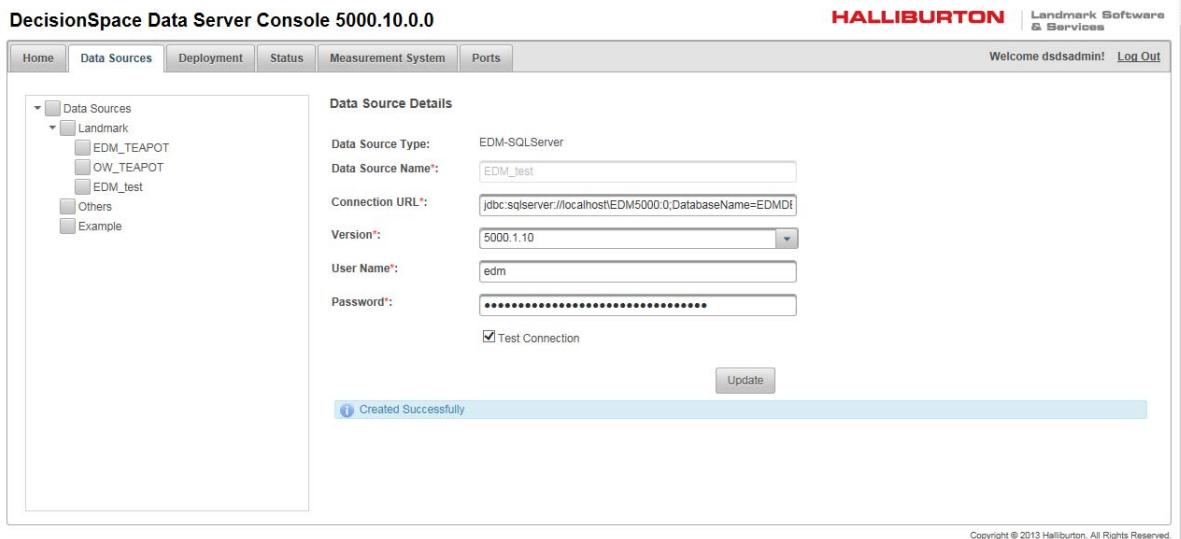
The **Drivers** window appears.



4. Select the **EDM-SQL Server** data source and click **OK**.
The **Data Source Details** window appears.

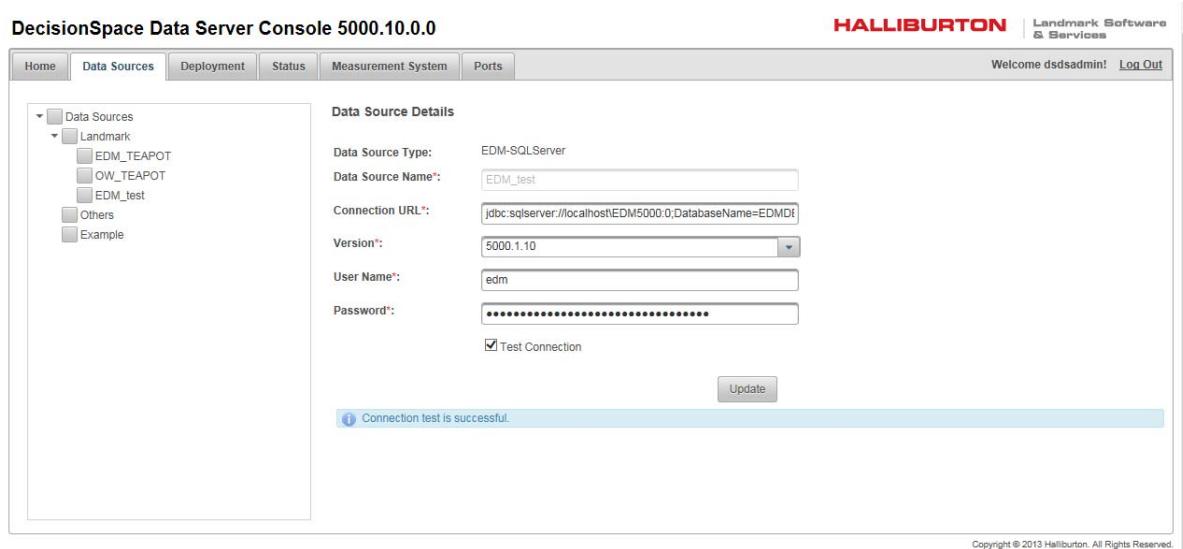
A screenshot of the 'DecisionSpace Data Server Console 5000.10.0.0' application. The top navigation bar includes Home, Data Sources, Deployment, Status, Measurement System, and Ports. The right side shows the 'HALLIBURTON | Landmark Software & Services' logo and a 'Welcome ddsadmin! Log Out' message. The main area has a sidebar with 'Data Sources' expanded, showing Landmark, Others, and Examples. The central panel is titled 'Data Source Details' for the 'Landmark' source. It contains fields for Data Source Type (set to EDM-SQLServer), Data Source Name (empty), Connection URL (empty), Version (set to 5000.1.10), User Name (empty), Password (empty), and a checked 'Test Connection' checkbox. A 'Save' button is at the bottom right. The bottom right corner of the page has a copyright notice: 'Copyright © 2013 Halliburton. All Rights Reserved.'

5. Enter **EDM_test** in the **Data Source Name** field.
6. Enter the following syntax in the **Connection URL** field:
`jdbc:sqlserver://localhost\EDM
5000:0;DatabaseName=EDMDB`
7. Enter **edm** and **Landmark1** in the **Username** and **Password** fields respectively.
8. Click **Save**.
The new data source displays in the tree list on the left side of the DecisionSpace Data Server Console window.

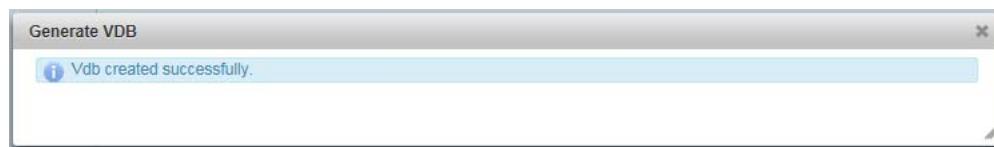


9. Select the **EDM_test** check-box from the list on the left side of the DecisionSpace Data Server Console.
10. Right-click the **EDM_test** data source and select **Test** from the context menu.

A message appears on the DecisionSpace Data Server Console displaying the status of the connection.



11. Select the **EDM_test** check-box from the list on the left side of the DecisionSpace Data Server Console.
12. Right-click the **EDM_test** data source and select **Generate VDB** from the context menu.
A message appears on the DecisionSpace Data Server Console displaying the status of the connection



Errors (if any) generated during this process are stored in the **server.log** file. Select the **Status** tab or browse to the **JBOSS_HOME/standalone/log** directory to view this file.

Note

After deployment of a VDB on the server, the data sever will try to connect to the data source(s) selected for VDB generation and load the metadata of those data sources (tables, relations etc). It may take a few moments before the data becomes available. If anything goes wrong during this process, the VDB will not get deployed. To view the full details of a deployment, select the **Deployment** tab.

13. To check if the deployment was successful, click on the **Deployment** tab.

The screenshot shows the 'DecisionSpace Data Server Console 5000.10.0.0' interface. At the top, there is a navigation bar with tabs: Home, Data Sources, Deployment (which is selected), Status, Measurement System, and Ports. On the right side of the header, it says 'HALLIBURTON Landmark Software & Services' and 'Welcome ddsadmin! Log Out'. Below the header, there is a section titled 'Deployed VDBs' with a table. The table has columns: VDB Name, Version, Teiid JDBC Url, Dynamic, Status, and Actions. There is one row for 'OpenWorks' with version '5000.8.3' and a JDBC URL 'jdbc:teiid:OpenWorks@mm://10.11.24.162:31000;version=1'. The 'Status' column shows 'ACTIVE' and the 'Actions' column has links for 'Delete' and 'Details'. There are also '+ Add' and 'Refresh' buttons at the top of the table.

VDB Name	Version	Teiid JDBC Url	Dynamic	Status	Actions
OpenWorks	5000.8.3	jdbc:teiid:OpenWorks@mm://10.11.24.162:31000;version=1	true	ACTIVE	Delete Details

14. Click on **Details** to see more details about the deployment of the VDB.

Note

If you want to delete a **VDB**, click **Delete**. The **VDB** is deleted along with the connection information.

Exercise: Connecting to DSDS using Web Browser, Excel (powerpivot)

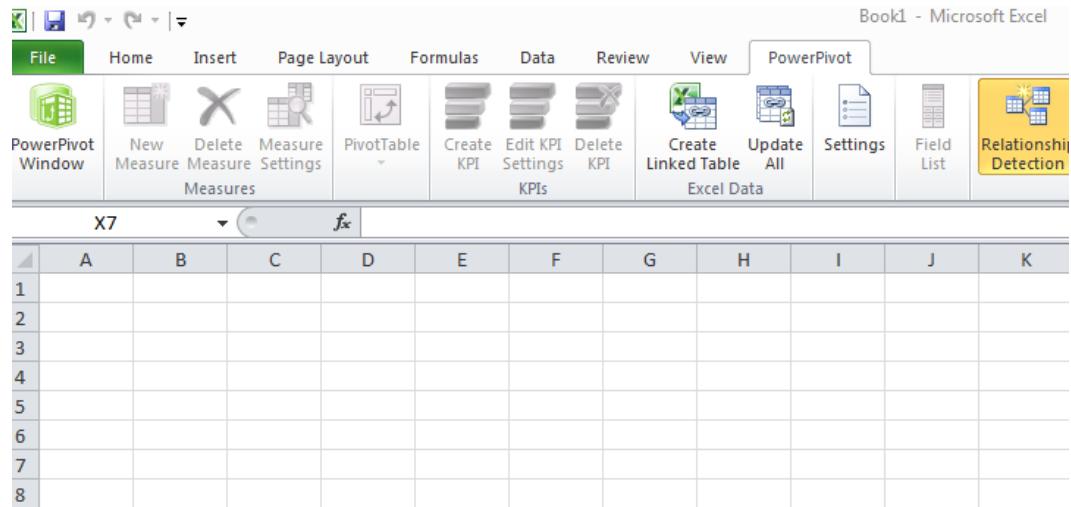
In this exercise you will be using Powerpivot to analyze of the data from OpenWorks and EDM.

Once all the tables are imported you can run the analysis on the data. We will be using the data for the table MD_PK_PDEN_TYPE and the analysis we want to do is to find out the count of PDEN_TYPE.

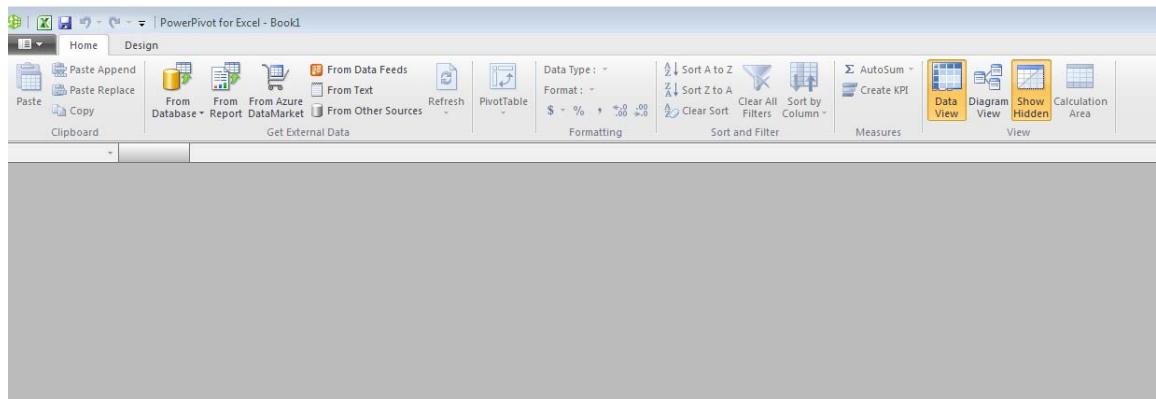
After this exercise you will be able to do different analysis based on requirements.

1. Launch **Excel**.

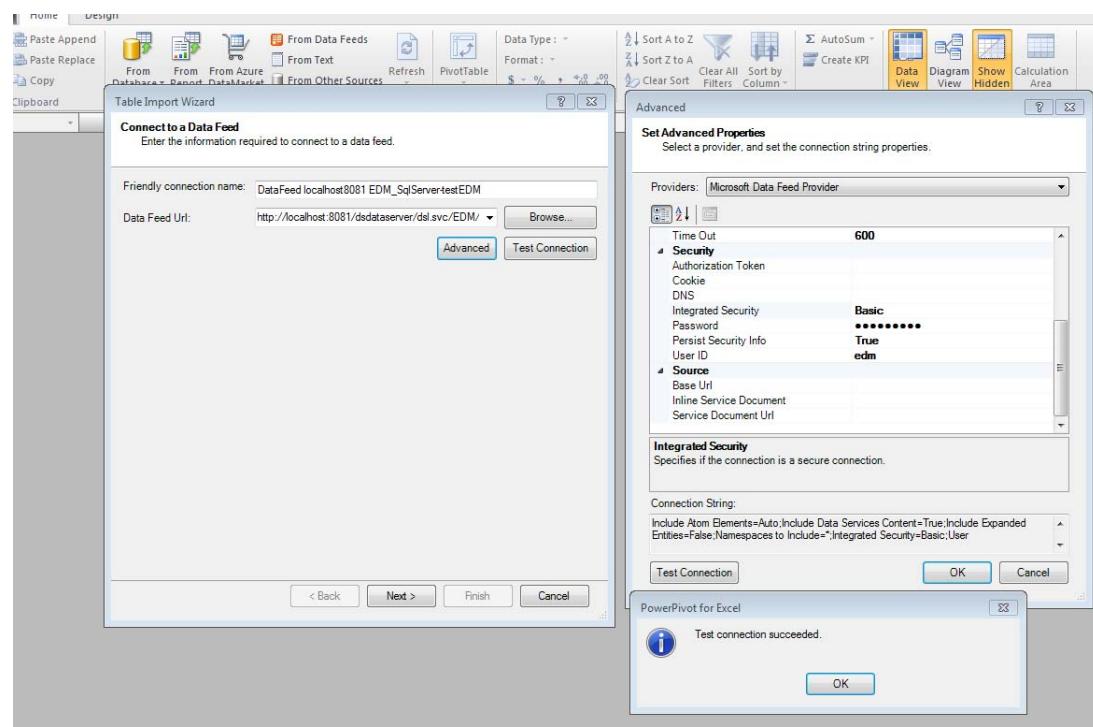
2. Launch **Power Pivot**.



3. Launch the **PowerPivot Window**.

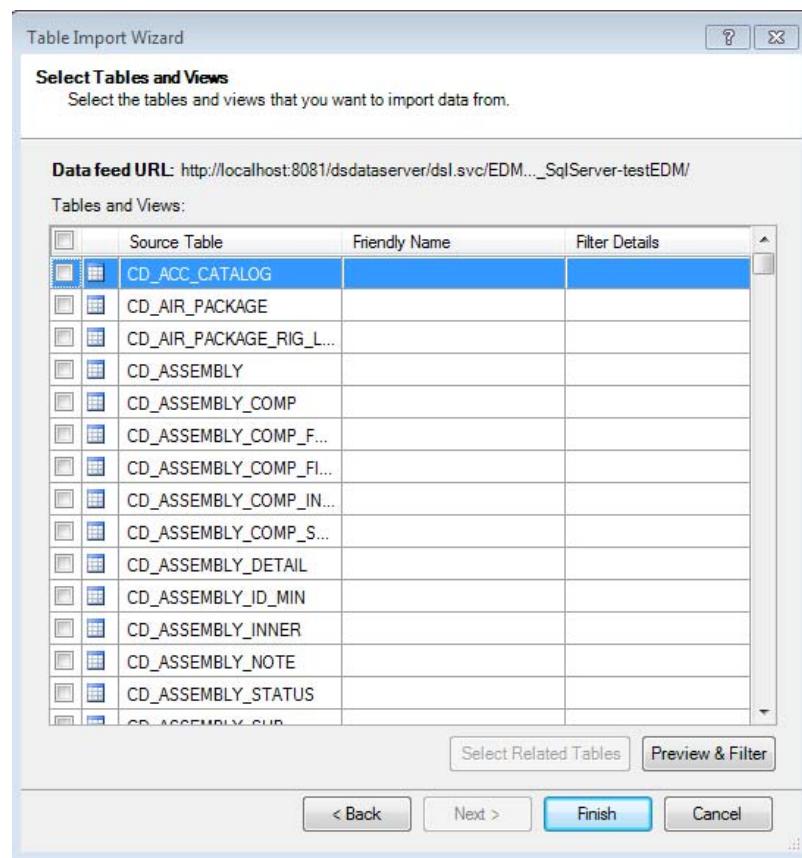


4. Create the connection to EDM data feeds.

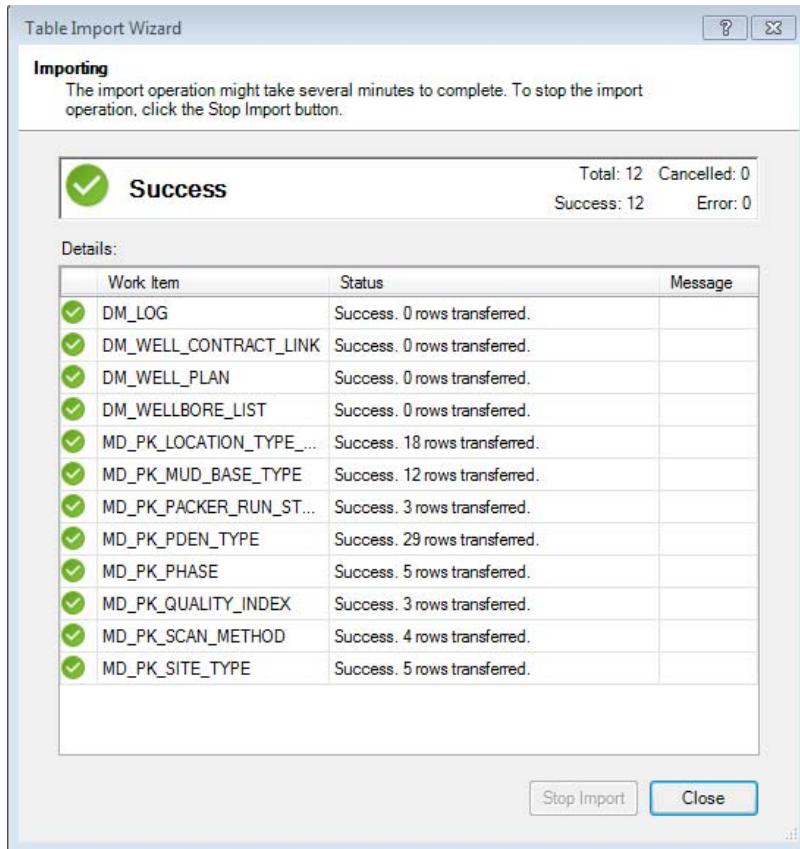


5. Click **OK** for the **PowerPivot for Excel** dialog-box.

6. Click **Next**.

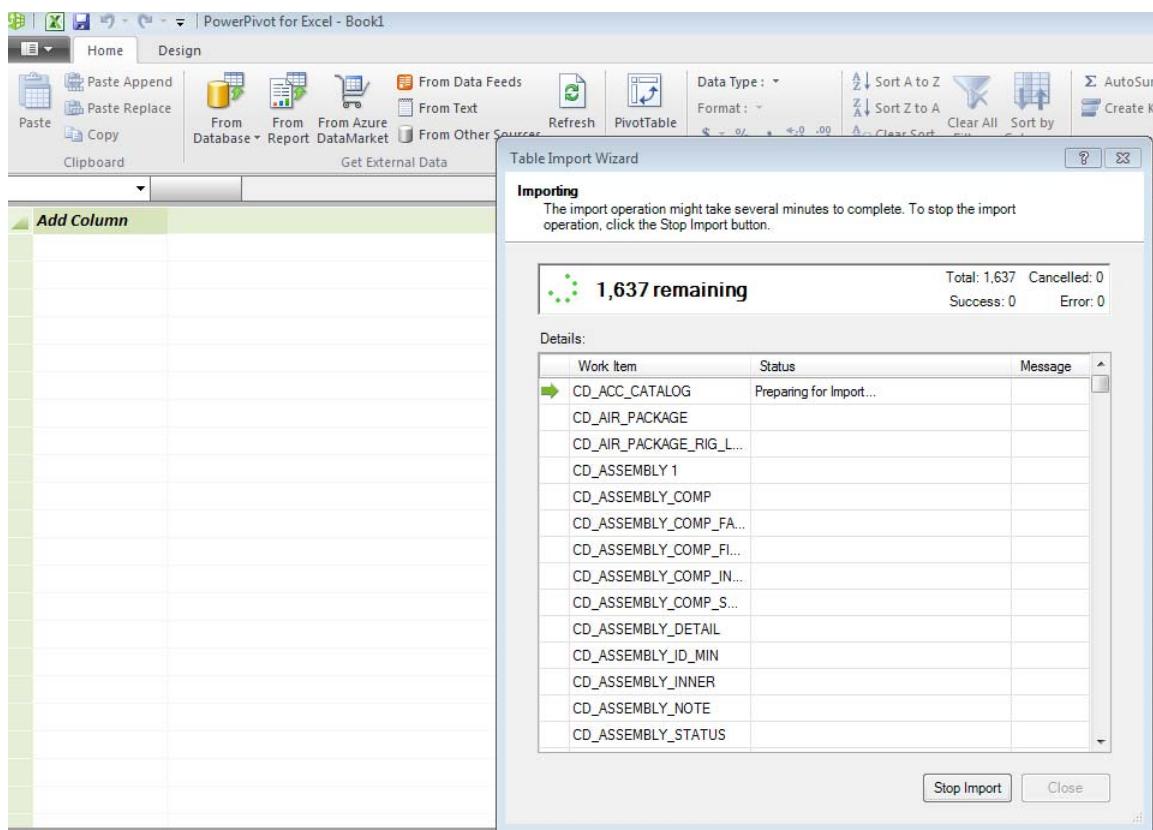


7. Select all the source tables that you want to do analysis on and then click **Finish**.



8. Select all tables.

MS Excel imports all the tables.



- Run the analysis on the imported data.
Data for the table displays.

The screenshot shows a Microsoft Excel window titled "Book1 - Microsoft Excel". The ribbon tabs are visible at the top, including File, Home, Insert, Page Layout, Formulas, Data, Review, View, PowerPivot, Options, and Design. The PowerPivot tab is selected.

The main area displays a PivotTable with the formula `=Count(PDEN_TYPE)` in cell B3, resulting in the value 29. The PivotTable has columns A through F and rows 1 through 28. The "PowerPivot Field List" pane on the right shows various dimensions and measures, such as MD_PK_LOCATION_TYPE_CODE, MD_PK_MUD_BASE_TYPE, and MD_PK_PDEN_TYPE, along with their respective fields like IS_HIDDEN, LOCATION_TYPE_CODE, and PDEN_TYPE.

The "Values" section of the PowerPivot Field List shows the formula `Count of PDEN_TYPE` selected. The "Report Filter" section is currently empty.

The bottom of the screen shows the Excel ribbon with tabs for Sheet4, Sheet1, Sheet2, and Sheet3, and a status bar indicating "Ready" and "100%".

Connecting DecisionSpace Data Quality with DecisionSpace Data Server

Connections in the DecisionSpace Data Quality consist of the following three Source Dataset elements:

- Data Source (Data Owner) - where data is read from
- Workspace - schema where results are written
- Data Model - location where detailed information about data source tables and columns, table relationships, and element assignments are stored

These three elements help connect each phase of the Data Quality software to its own unique configuration. The 'Data Owner' connections can only be deleted from the application, when all the connected phases are deleted. Workspace connections and models are editable and can be updated as and when required.

Note

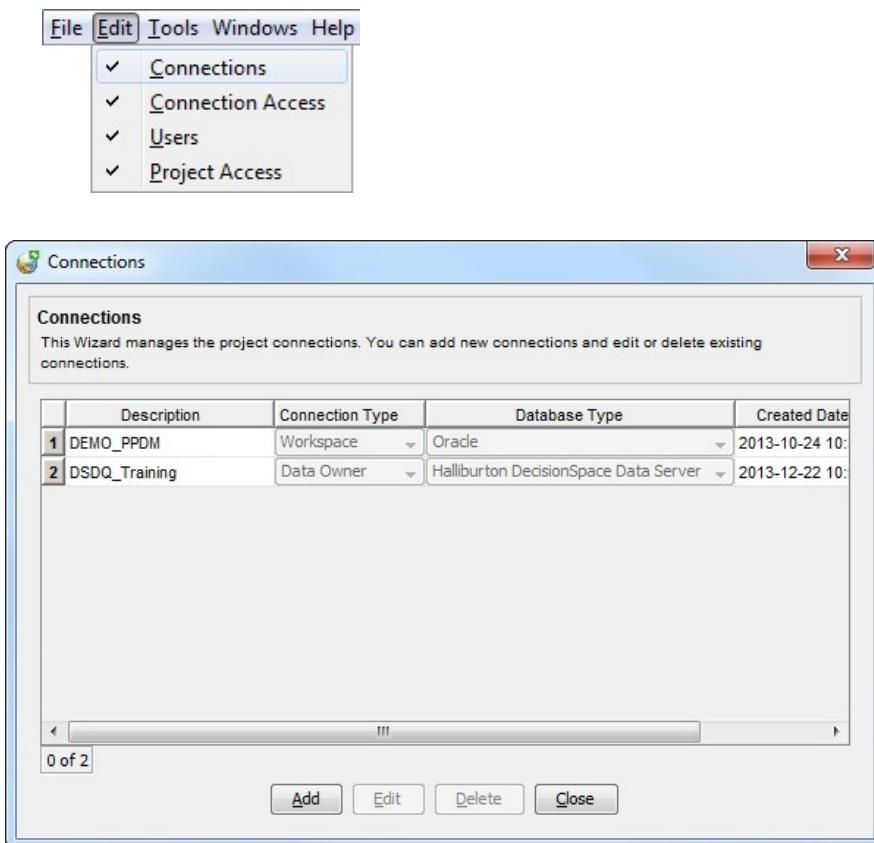
A warning message appears to contact Support if the maximum number of purchased connections is exceeded.

Exercise: Creating Connections

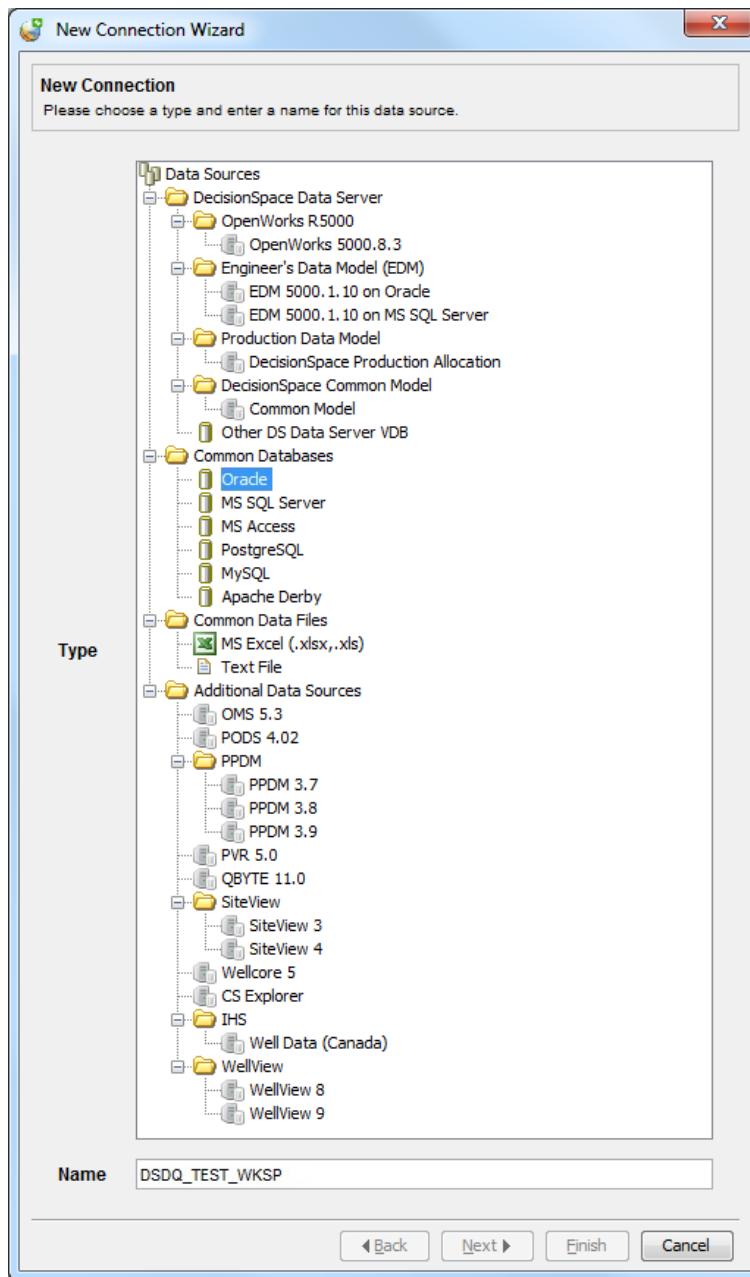
A Data Owner (the data source that the application reads from) must exist prior to reaching this step. This section outlines the process to create a connection between the Data Quality application and the Workspace schema (where results are written) created prior to installing DecisionSpace Data Quality.

To add a new Data Owner connection:

1. Select **Edit > Connections** from the menu bar on the **DSDQ Project Window**.



2. Click **Add** to display the **New Connection** window.



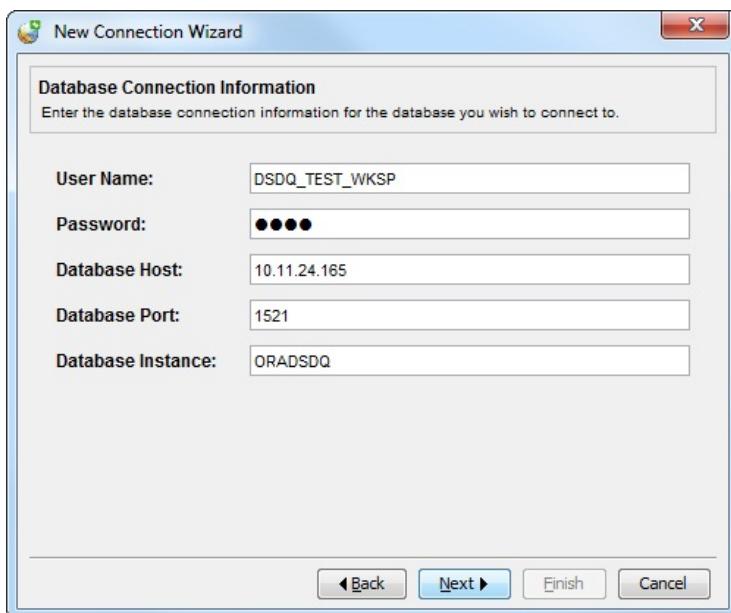
3. Select **Oracle** as the connection type from the **Type** tree list.
 4. Enter **DSDQ_TEST_WKSP** in the **Name** field. This could be a real database name or an alias. This name is used for user reference.

only and appears in the drop-down list after the connection setup is complete.

Note

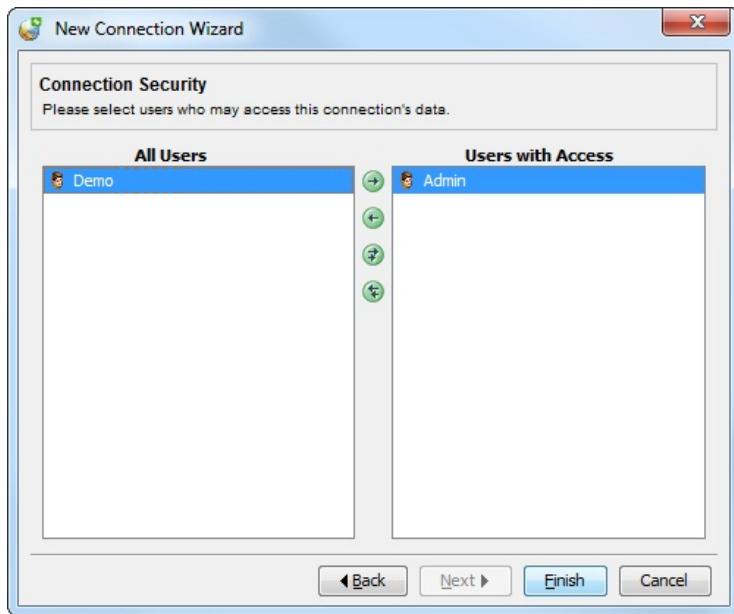
Supported Databases for Connection Type connections include:
DecisionSpace Data Server, Oracle, MS SQL Server, Apache Derby, MySQL,
PostgreSQL, PPDM 3.8, MS Access, .DBF files, .CSV files, and Excel
spreadsheets.

5. Click **Next** to display the **Database Connection Information** window.



6. Enter **DSDQ_TEST_WKSP** as the user name for the database scheme that you are trying to connect to in the **User Name** field.
7. Enter the password used to connect to the database schema in the **Password** field.
8. Enter the host information for the server where the database schema resides in the **Database Host** field.
9. Enter the port number for connection to your data source in the **Database Port** field. The default value is 1521.
10. Enter a name for the database instance in the **Database Instance** field.

11. Click **Next** to display the **Connection Security** window.



12. Select user(s) who should have access to the data source from the **All Users** list and then click to move selected users to the **Users with Access** list.
13. Click **Finish** to complete the process of creating a connection between the Data Quality application and the Workspace schema (where results are written).

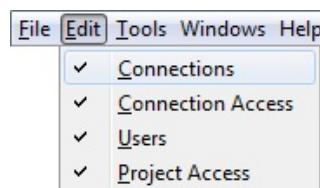
DecisionSpace Data Server Quick Start Connections

DecisionSpace Data Server allows users to connect to Landmark application databases, such as OpenWorks® and EDM™. By connecting to these application databases through the DecisionSpace Data Server, the Data Quality software can automatically load preconfigured models and rules against these databases.

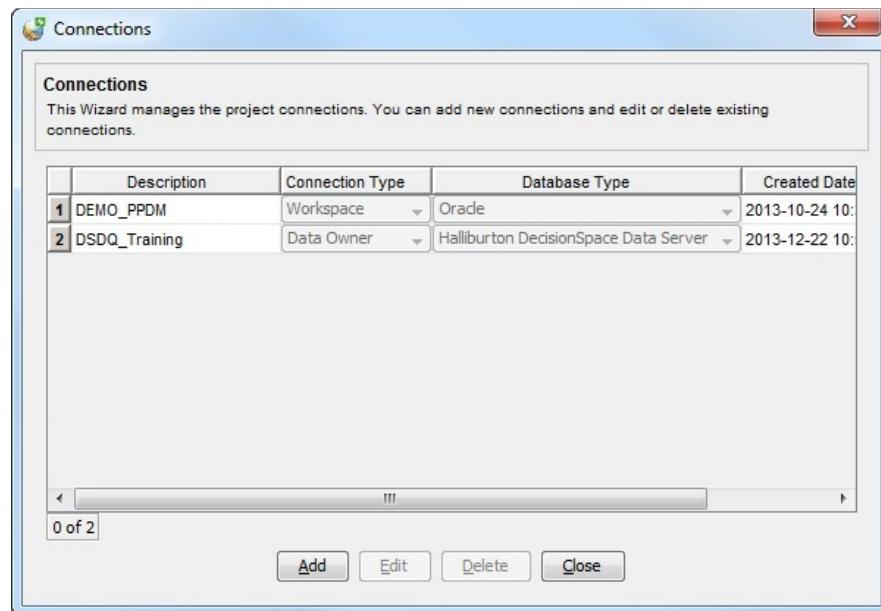
Exercise: Creating New DecisionSpace Data Server Quick Start Connections

To create a DecisionSpace Data Server Quick Start Connection:

1. Select **Edit > Connections** from the menu bar on the **DSDQ Project Window**.

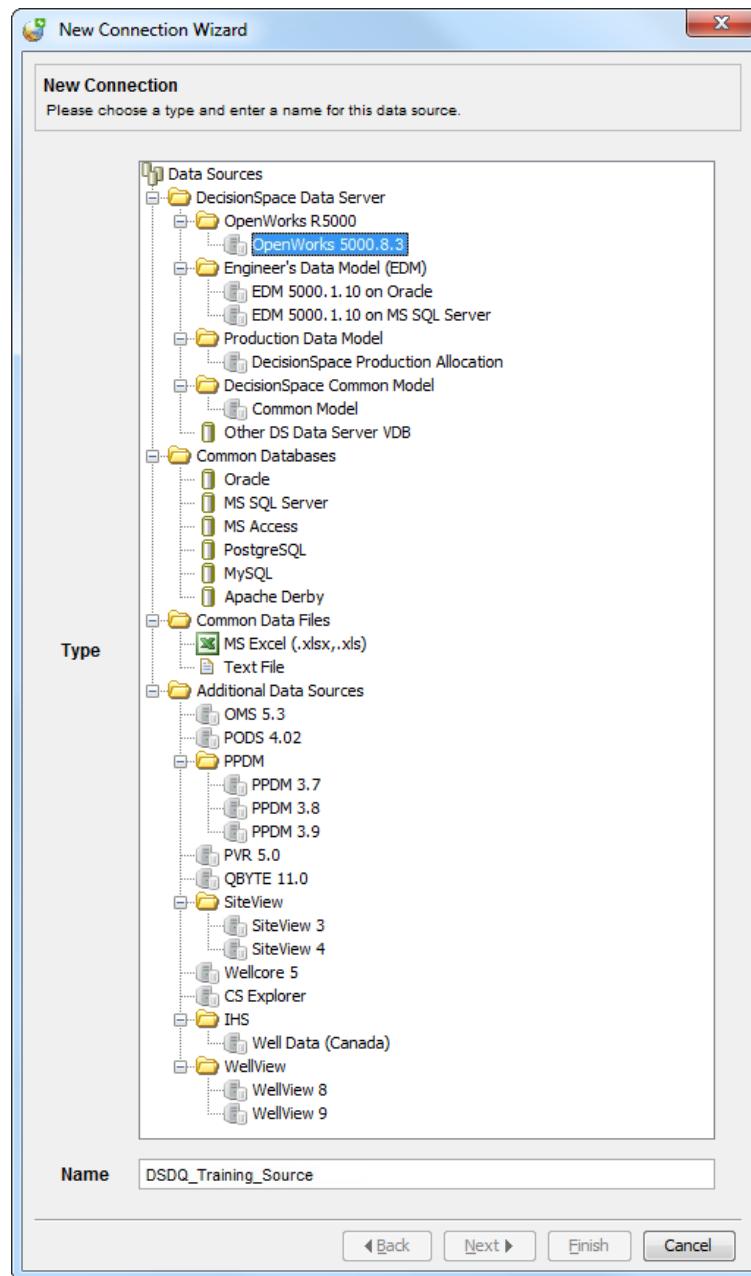


The **Connections** window displays.



2. Click **Add** to display the **New Connection** window.

The **New Connection Wizard** window displays.

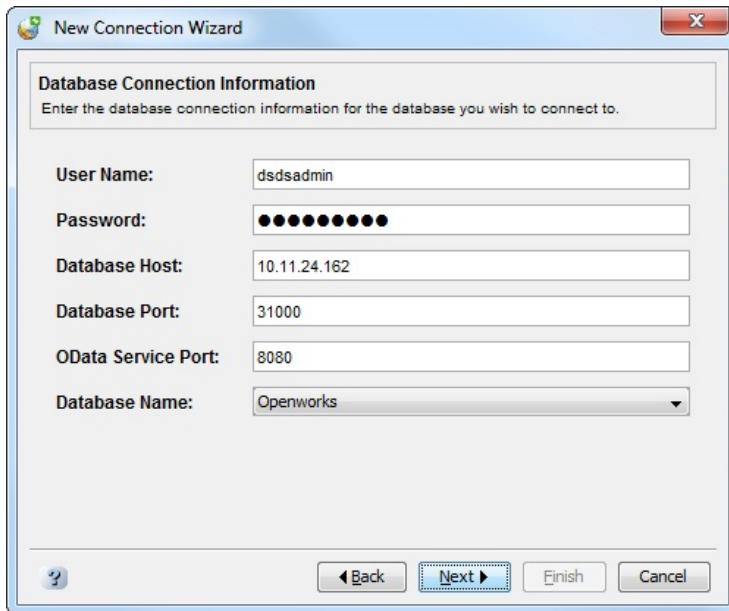


3. Select **OpenWorks 5000.8.3** as the connection type from the **Type** tree list.

Note

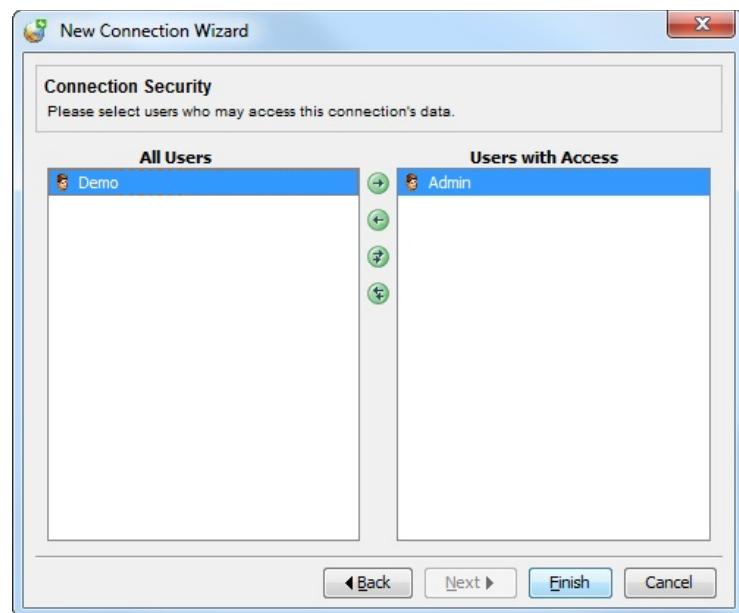
Supported DecisionSpace Data Server Quick Starts are OpenWorks R5000 and OpenWells.

4. Enter **DSDQ_Training_Source** as the name for the connection in the **Name** field. This could be a real database name or an alias. This name is used for user reference only and appears in the drop-down list after the connection setup is complete.
5. Click **Next** to display the **Source Connection Wizard** window.



6. Enter **dsdsadmin** as the user name for the database schema that you are trying to connect to in the **User Name** field.
7. Enter the password used to connect to the database schema in the **Password** field.
8. Enter the host information for the server where the database schema resides in the **Database Host** field.
9. Enter the port number for connection to your data source in the **Database Port** field. The default value is **31000**.
10. Enter the port number that is open for the DecisionSpace Data Server to read the OData stream from in the **OData Service Port** field. The default is 8080.
11. Select the database you want to connect to from the **Database Name** drop-down list.

12. Click **Next** to display the **Connection Security** window.



13. Select user(s) who should have access to the data source from the **All Users list** and then click to move selected users to the **Users with Access** list.
14. Click **Finish** to connect the **OpenWorks 5000.8.3** application database with the Data Quality application.

Chapter 3

Managing Projects

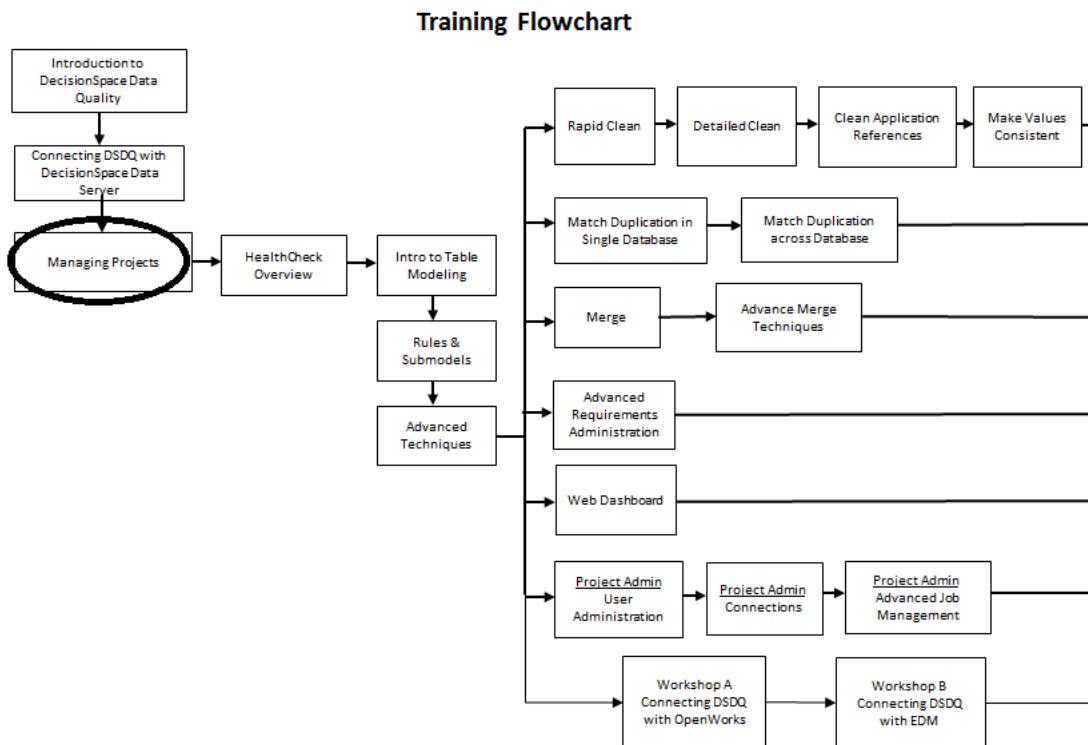
Before you load data that you want processed in DecisionSpace® Data Quality (DSDQ), you must create a project to contain all the Phases, Activities, Tools, Tasks, Jobs and Results associated with the source data. In DSDQ, a Project is a master folder that holds all the files created by end-users and the application while processing data. Rules to profile, audit and validate data quality (i.e. Phases) can only be defined once a project has been created. A Project in DSDQ is identified by a hard hat  icon.

Chapter Overview

In this chapter, you will learn about:

- Creating a Project
- Managing Projects (i.e. opening and deleting projects)
- Importing/Exporting Projects

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:

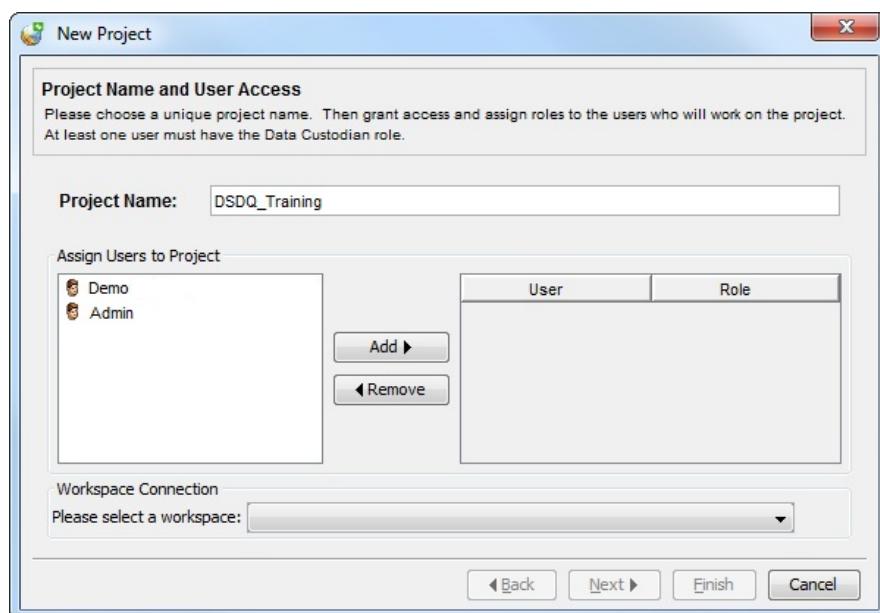


Creating a Project

A project can only be created by a user with **Administrator** rights in DSDQ. It is recommended that you specify a unique name for each project that you create. During this process, you will:

- Assign users to the project and their roles while they work on the project
- Select a Workspace Connection (the database where results will be written)
- Select the desired Phase and source connection (data source that the application reads from)

In all instances, the **New Project** wizard opens displaying all available users.



There are two roles within the Data Quality software-- Data Custodian and Data Steward. These roles carry with them an implicit description based on their roles in an organization:

- **Data Custodian:** Data custodians are responsible for facilitating the safe transport and storage of data. Often working directly in information technology and security, they aim to maintain data infrastructure and business rules. In the Data Quality application,

the **Data Custodian** role relates to responsibilities such as setup and configuration of rules, creating associations between columns and tables, and creating connections to data stores for **Data Stewards**

- **Data Steward:** Data Stewards have the primary responsibility of managing the content in data stores, as well as controlling any modifications to them. They deal with the daily governance of a company's information, and are often subject matter experts on their governed data. Within the Data Quality application, **Data Stewards** are responsible for tasks directly related to the data within data stores, such as running jobs, confirming data matches, and consolidating information. **Data Stewards**, however, do not have access to create new jobs, or modify their configurations

Each role has a set of attached permissions. Permissions are organized by Activities/Tools as they appear in the Data Quality Tree or Tools Menu. Examples of Activities are Project Administration and Project Tools. Examples of Menu Tools are Manage Users and Manage Project Access. Every Activity/Tool can have either **read** or **execute** permissions, or be **locked** to the user. **Execute** defines the ability to create, configure, and run an activity, tool, or menu. **Read** access allows the running of a created job, and extends to allow some work within the results of jobs. Additionally, every **Data Custodian** or **Data Steward** can be granted the **Administrator** role, granting access to Data Quality meta-tasks. The following table summarizes Roles and their Permissions.

Roles and Permissions	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
Rapid Clean	execute	read
Detailed Clean	execute	read
Clean Application References	execute	read
Make Values Consistent	execute	read

Roles and Permissions	Data Custodian	Data Steward
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
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

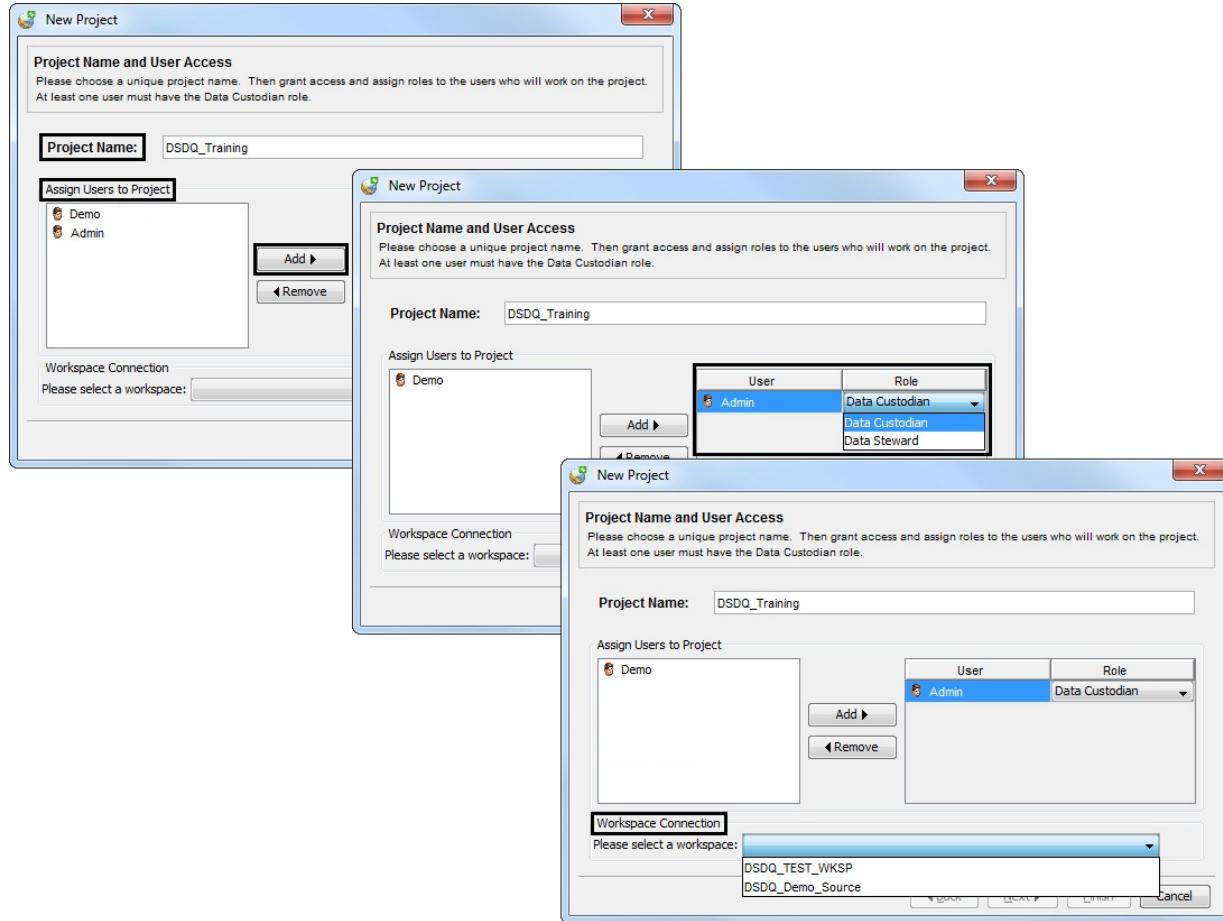
Exercise: Creating a Project

In this exercise, you will be creating a new project in DSDQ, assign user and role under a workspace previously created and add a HealthCheck Phase to the project.

A new project can be created in any of the following three ways:

- When the software is initially installed and no projects exist
- By selecting **New Project** from the **File** menu
- By clicking the **New** button in the **Open an Existing Project** window

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



2. Enter **DSDQ_Training** in the **Project Name** field.

Note

A project name must be less than or equal to 50 characters.

3. Select **Admin** from the **Assign Users to Project** group box. Access to the project will be given to selected users only.
4. Click the **+ Add** button to assign project access to the selected user.
You can only select one user at a time. Repeat steps **3 & 4** for each user that you create. The selected users i.e. **Admin** is added to the box on the right side of the **Assign Users to Project** group box.

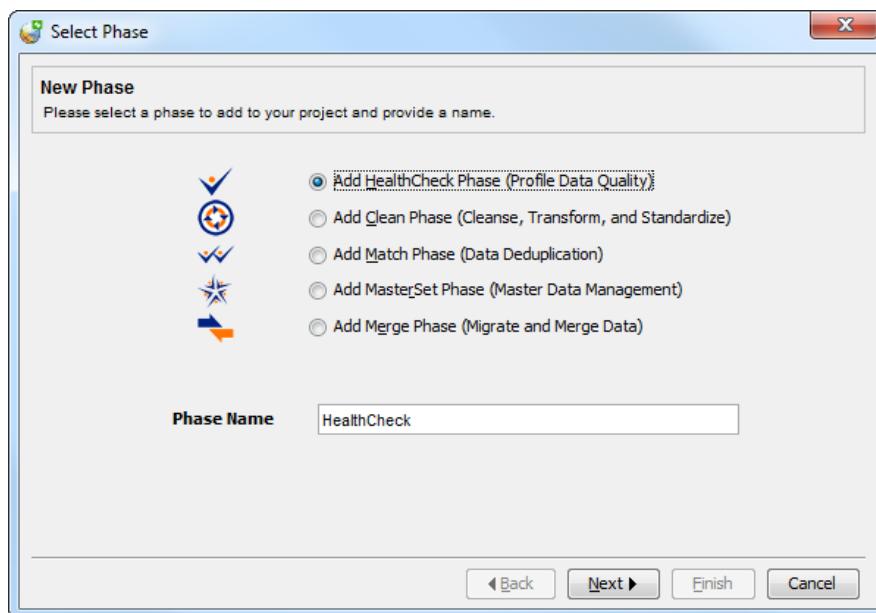
5. Assign the **Admin** user a role in the project by selecting it from the Role drop-down list. Options available for selection include **Data Custodian** and **Data Steward**. Select the **Data Custodian** role for the **Admin** user.
6. Select **DSDQ_TEST_WKSP** from the **Please select a workspace** drop-down list.

Note

A workspace database is a repository where results will be written. Ensure that the workspace database that you connect to this project i.e. **DSDQ_TEST_WKSP** is not used by other Data Quality projects and phases.

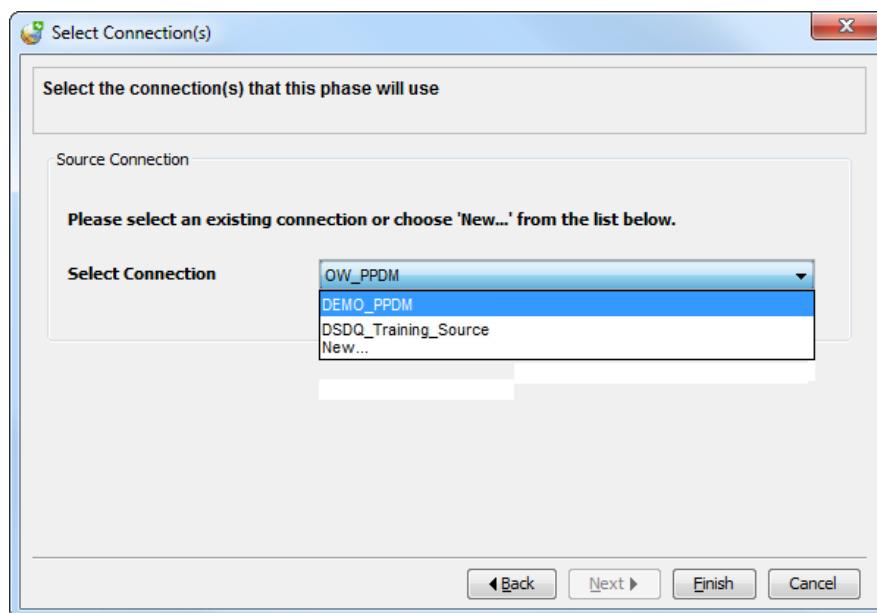
Supported Databases for **Workspace** connection include:

- Oracle
 - MS SQL Server
 - Apache Derby
7. Click **Next** on the **New Project** window to continue. The **Select Phase** window appears with the **Add HealthCheck Phase (Profile Data Quality)** option selected by default.



8. Select the desired phase to add it to your project, if different from the one selected by default.

9. Optionally, change the description of the selected phase by entering a brief narrative for it in the **Phase Name** field. The phase name must be unique, and can be renamed after being created.
10. Click **Next** to continue.
The **Select Connection(s)** window appears.



11. Select **DSDQ_Training_Source** as the data owner connection (the data source that the application reads from) from the **Select Connection** drop-down list.

Note

For more information on data owner connections, refer to Adding a New Data Owner Connection section in Chapter 2. Connecting DecisionSpace Data Quality with DecisionSpace Data Server.

12. Click **Finish** to complete the process of creating a new project in DSDQ. Once created, this project will be added to the Data Quality Tree on the **DSDQ Project Window**.

Note

The Data Quality data model that holds the data source objects for a particular connection is automatically created and given the same name as that of the connection.

Working with Projects

This section provides you information about opening existing projects, deleting unwanted projects and transferring projects to and from the application.

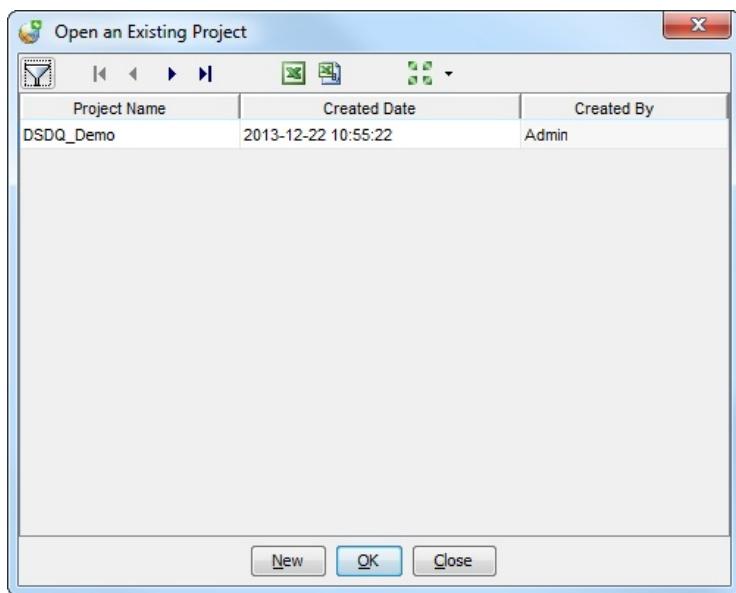
The functionality to import and export projects in DSDQ enables you to:

- Transfer full project configurations between the Data Quality installation
- Ensure a backup mechanism is in place for a project's connection, model and service level configurations

Exercise: Opening an Existing Project

1. Select **File > Open Project** from the menu bar on the **DSDQ Project Window**.

The **Open an Existing Project** window appears with the most recent project i.e. **DSDQ_Training** selected by default.



2. Click **OK** to open it.

Note

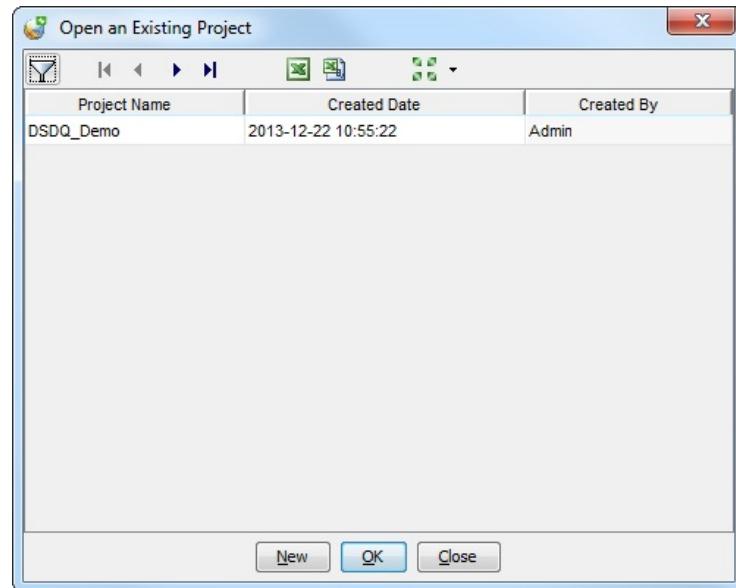
The most recent project will open automatically when you launch the application. In case the project that you want to open is not the one created recently, you will select it from this window and click OK to open it.

Exercise: Exporting a Project

Project Export allows you to transfer full project configuration between new Data Quality installations, as well as providing you with a backup mechanism for your connection, model, and service level configurations.

1. Select **File > Open Project** from the menu bar on the **DSDQ Project Window**.

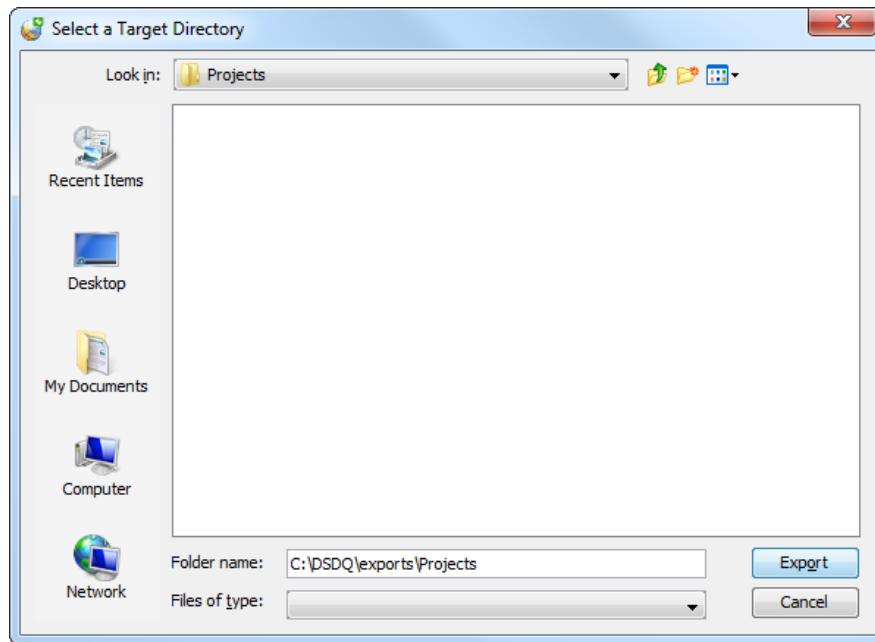
The **Open an Existing Project** window appears with the most recent project i.e. **DSDQ_Training** selected by default.



2. Select the **DSDQ_Training** project from the list.

3. Click the arrow ▾ on the **Import/Export Project** icon ➔ and select the **Export Selected Project(s)** option from the drop-down menu.

The **Select a Target Directory** window appears.



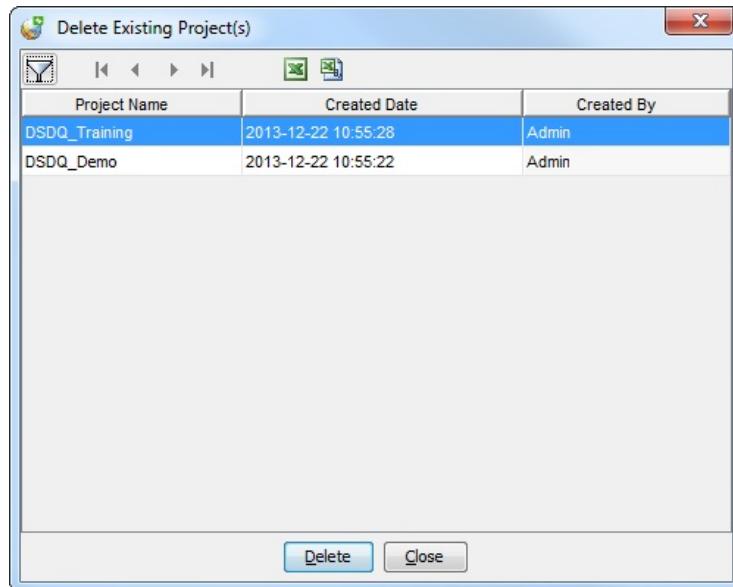
4. Select an export project path and click **Export**.
The selected project is exported at the desired location.

Exercise: Deleting a Project

Deleting a project removes all the data associated with it as well. Before you delete a project, ensure you have the required access privilege in DSDQ to do so i.e. **Administrator** rights.

1. Select **File > Delete Project** from the menu bar on the **DSDQ Project Window**.

The **Delete Existing Project(s)** window appears with the most recent project i.e. **DSDQ_Training** selected by default.

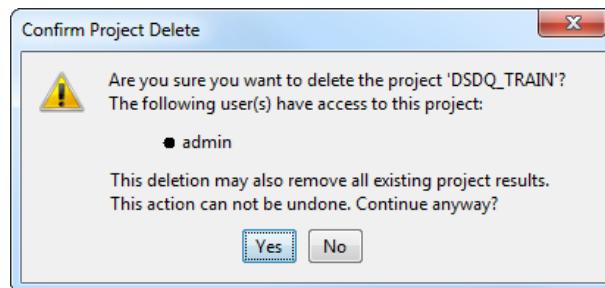


Note

In case the project that you want to delete is not the one created recently, you will select it from this window and then click **Delete** to remove it.

2. Click **Delete**.

The **Confirm Project Delete** message box appears listing all users associated with the project.



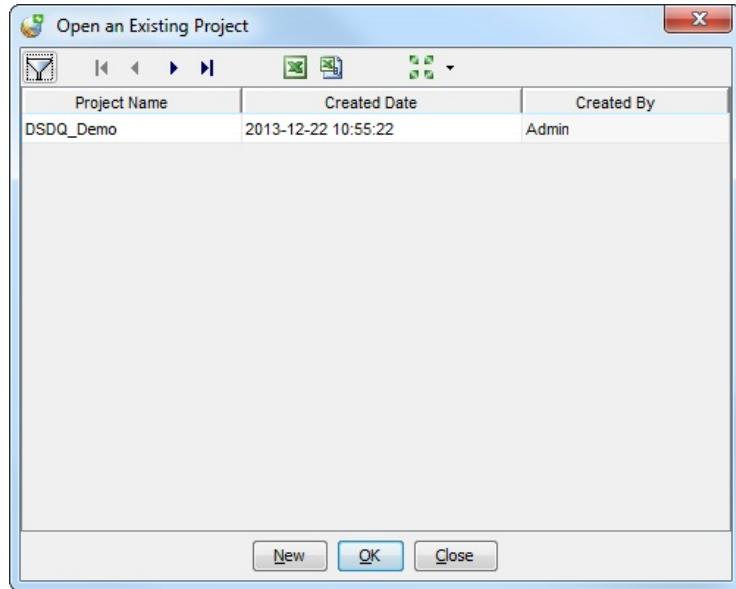
3. Click **Yes** to continue.

Exercise: Importing a Project

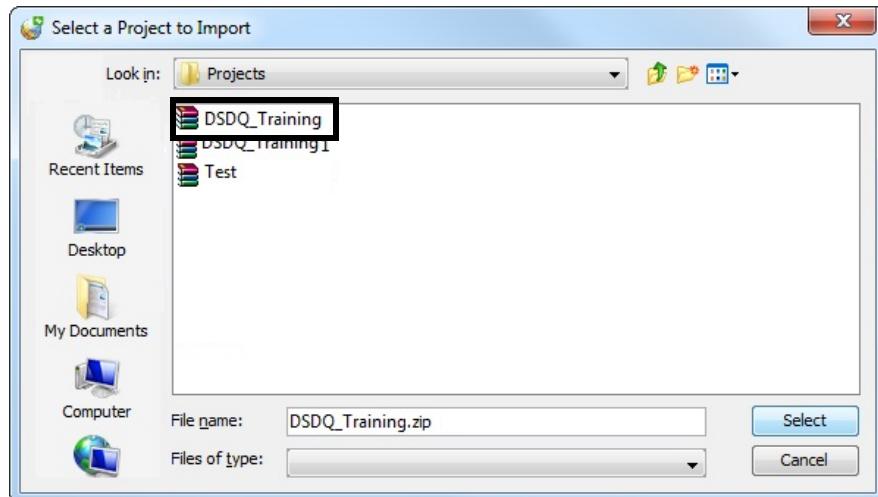
Project Import allows you to load full project configuration on a new Data Quality installation. Before you import a project, ensure you have the required access privilege in DSDQ to do so i.e. **Administrator** rights.

1. Select **File > Open Project** from the menu bar on the **DSDQ Project Window**.

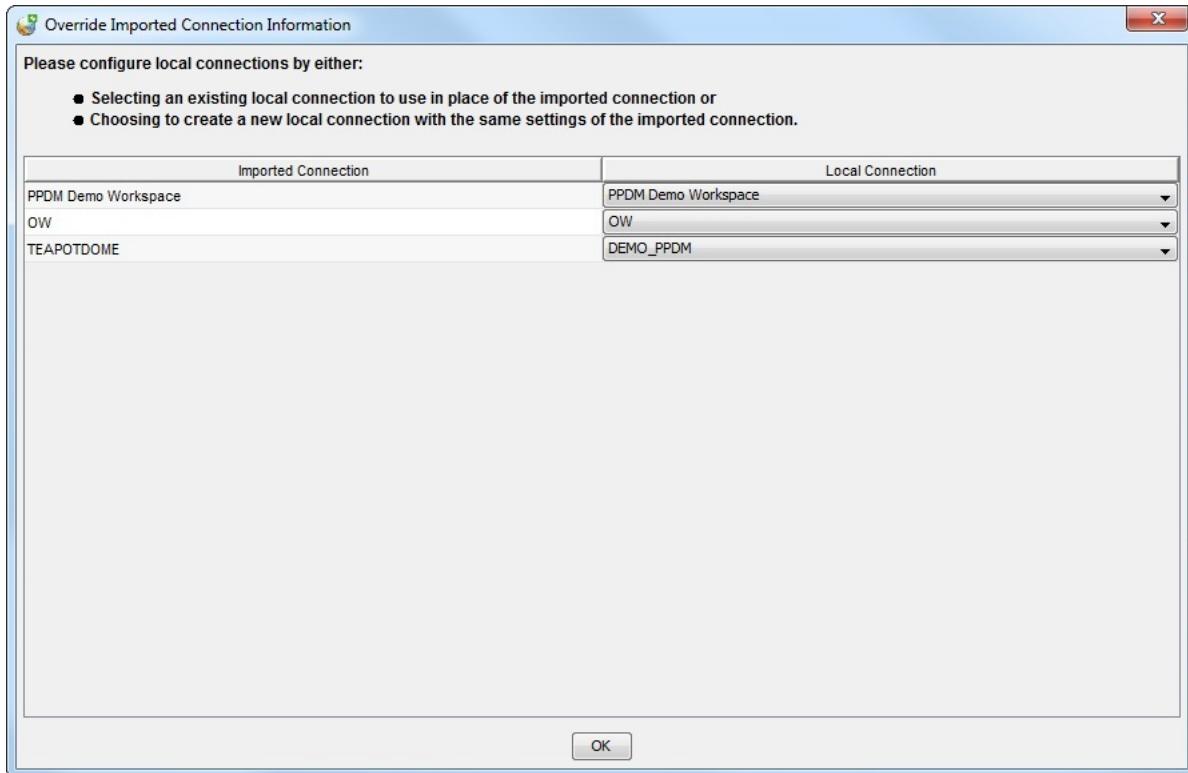
The **Open an Existing Project** window appears with the most recent project i.e. **DSDQ_Training** selected by default.



2. Click the arrow ▾ on the Import/Export Project icon  and select the **Import Project** option from the drop-down menu.
The **Select a Project to Import** window appears.



3. Select the **DSDQ_Training** project file and click **Select** to display the **Override Imported Connection Information** window.



Note

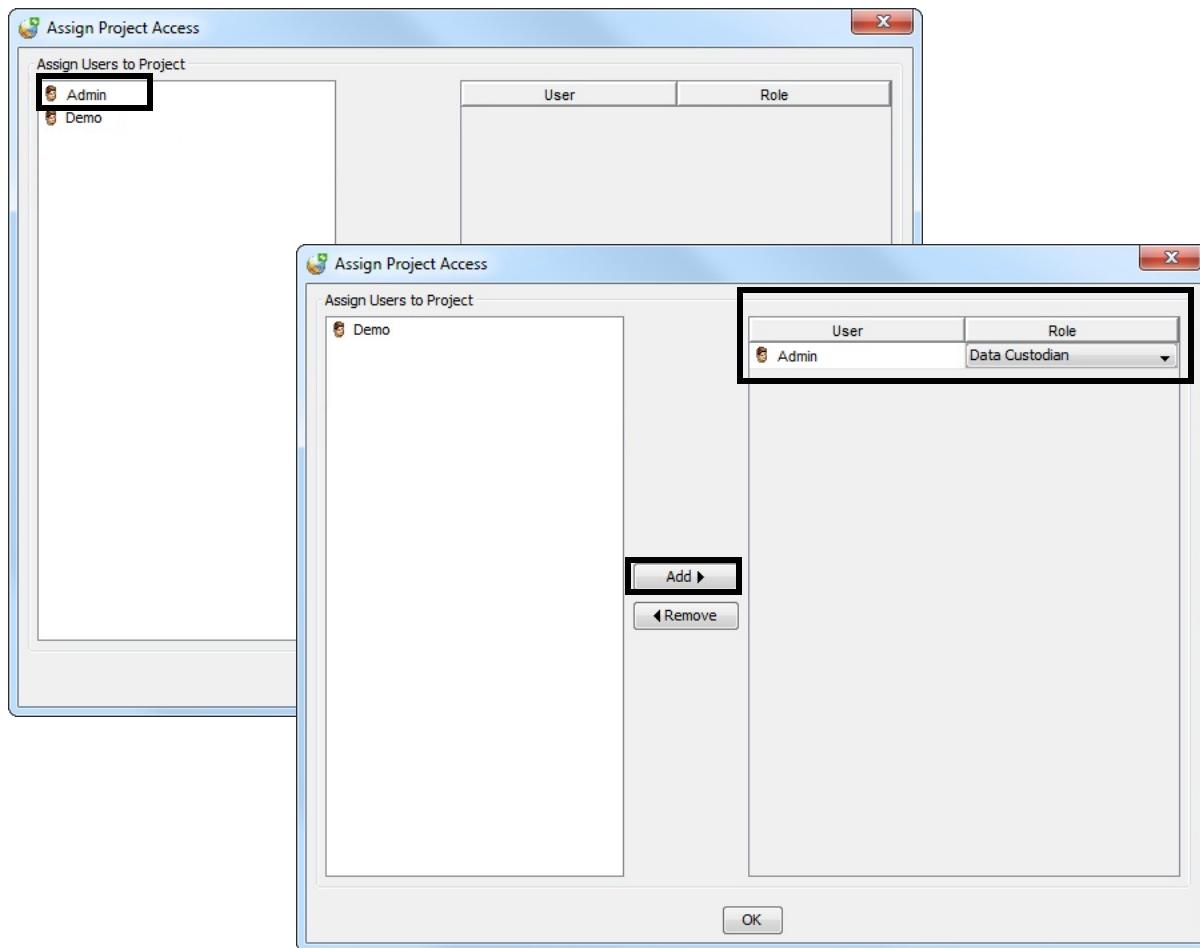
If the selected project export files match a project that already exists in the current Data Quality install, you are prompted if you want to backup that project in order to proceed with the new import process. Clicking Yes causes an export file backup of the existing project to be automatically created in the %DSDQ_HOME%/exports/Projects folder with the same <project name>.1.zip. That project is then deleted from the current Data Quality install and the new one is imported. Clicking No aborts the import process.

Connection information for the imported project is selected from here. Connections under the **Imported Connection** column represent connections coming from the imported project. The **Local Connection** column lists connections available in the current Data Quality install. These can be used by the imported project. By selecting a connection from the **Local Connection** column, once the import process is complete, the imported connection shares connection details with the selected local connection. Select **New...** from the **Local Connection** drop-down list to create a new connection for the imported project. For more information on new connections, refer to Creating Connections section in Chapter 2.

Connecting DecisionSpace Data Quality with DecisionSpace Data Server.

4. Click **OK** to continue the import process.

The **Assign Project Access** window appears. This window allows you to assign the desired Data Quality user access to the imported project. Users given access to an imported project will gain access to all of its connections.



5. Select **Admin** from the **Assign Users to Project** group box. Access to the project will be given to selected users only.
6. Click the **Add ▶** button to assign project access to the selected user.
The selected user i.e. **Admin** is added to the box on the right side of the **Assign Users to Project** group box.

7. Assign the **Admin** user a role in the project by selecting it from the Role drop-down list. Options available for selection include **Data Custodian** and **Data Steward**. Select the **Data Custodian** role for the **Admin** user.
8. Click **OK** to complete the project import process.
Once the project has been successfully imported, an Import Validation Warning may appear. This warning prompts you of possible, noncritical issues encountered during the import process. These issues are stored in the client.log file available at ***Landmark/Decisionspace Data Quality/logs***.

Chapter 4

Data Evaluation

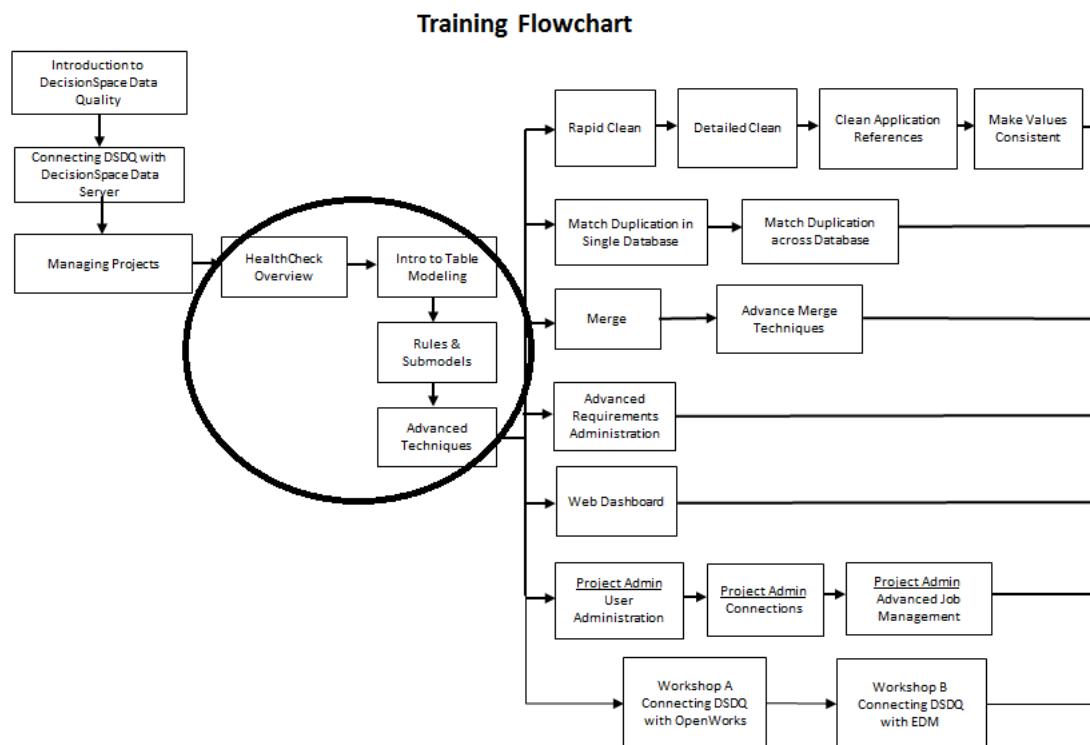
The first step in a successful data quality improvement project begins with data profiling, which uncovers the "where," "what" and "why" problems within your valuable data assets. Data profiling identifies the full spectrum of data issues, including incomplete, inaccurate, inconsistent, missing or ambiguous information. HealthCheck provides a comprehensive look at the quality of the source dataset. The **HealthCheck** Phase is composed of the **Rapid HealthCheck** and **Detailed HealthCheck** Activities. The series of reports that are produced by these activities assist with the completion of data quality analysis.

Chapter Overview

In this chapter, you will learn about:

- Performing table modeling
- Evaluating data volume and quality
- Identifying data issues

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:



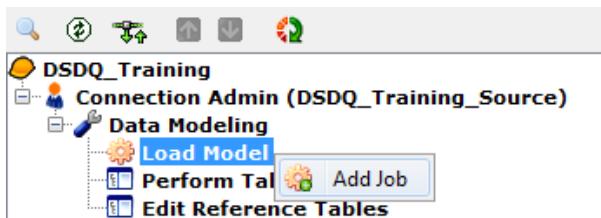
Perform Table Modeling

Table modeling defines the table hierarchy for the source database. All tasks within the Data Quality application are dependent on the information provided within table modeling.

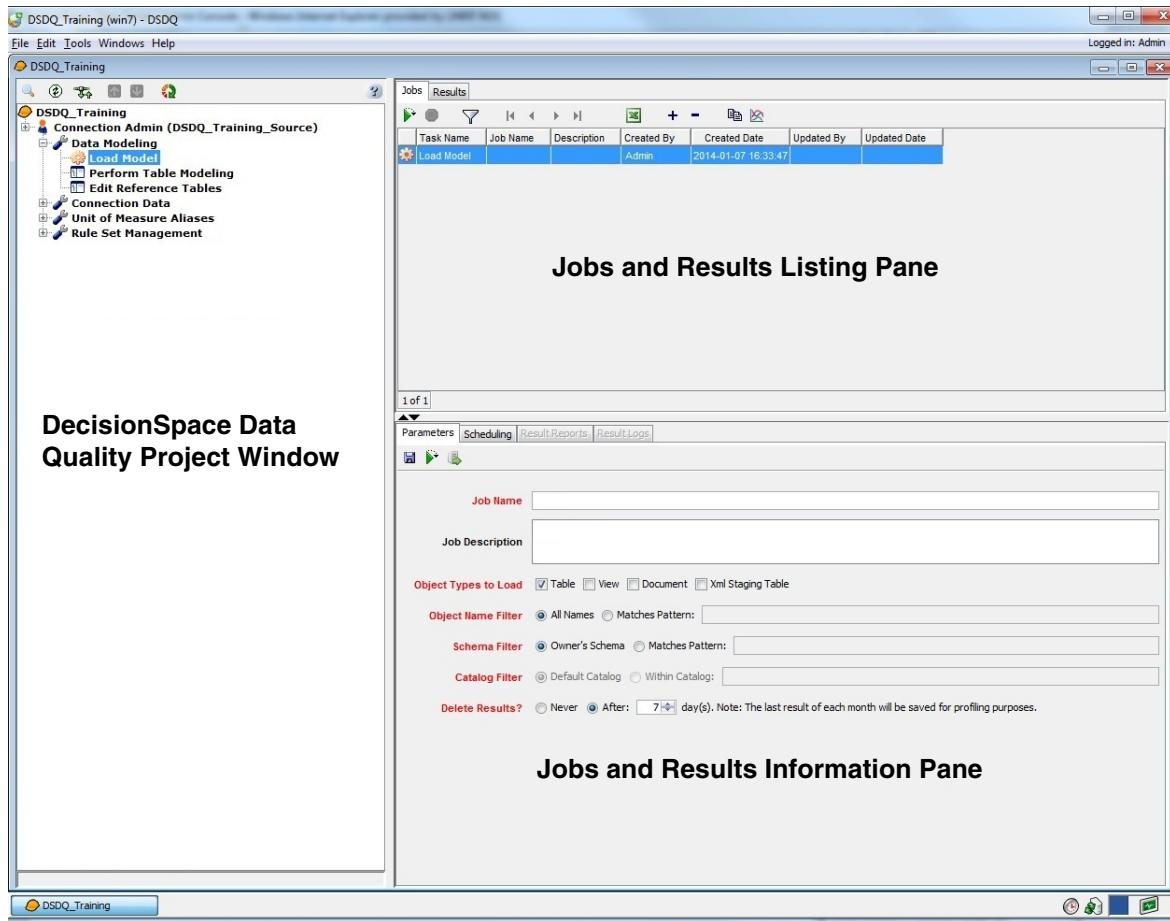
Exercise: Loading a Data Model

To Load a Data Model:

1. Click  on the DecisionSpace Data Quality Tree to expand the **Data Modeling** Activity.
2. Double-click the **Load Model** Task or right-click the **Load Model** Task and select **Add Job** from the pop-up menu.



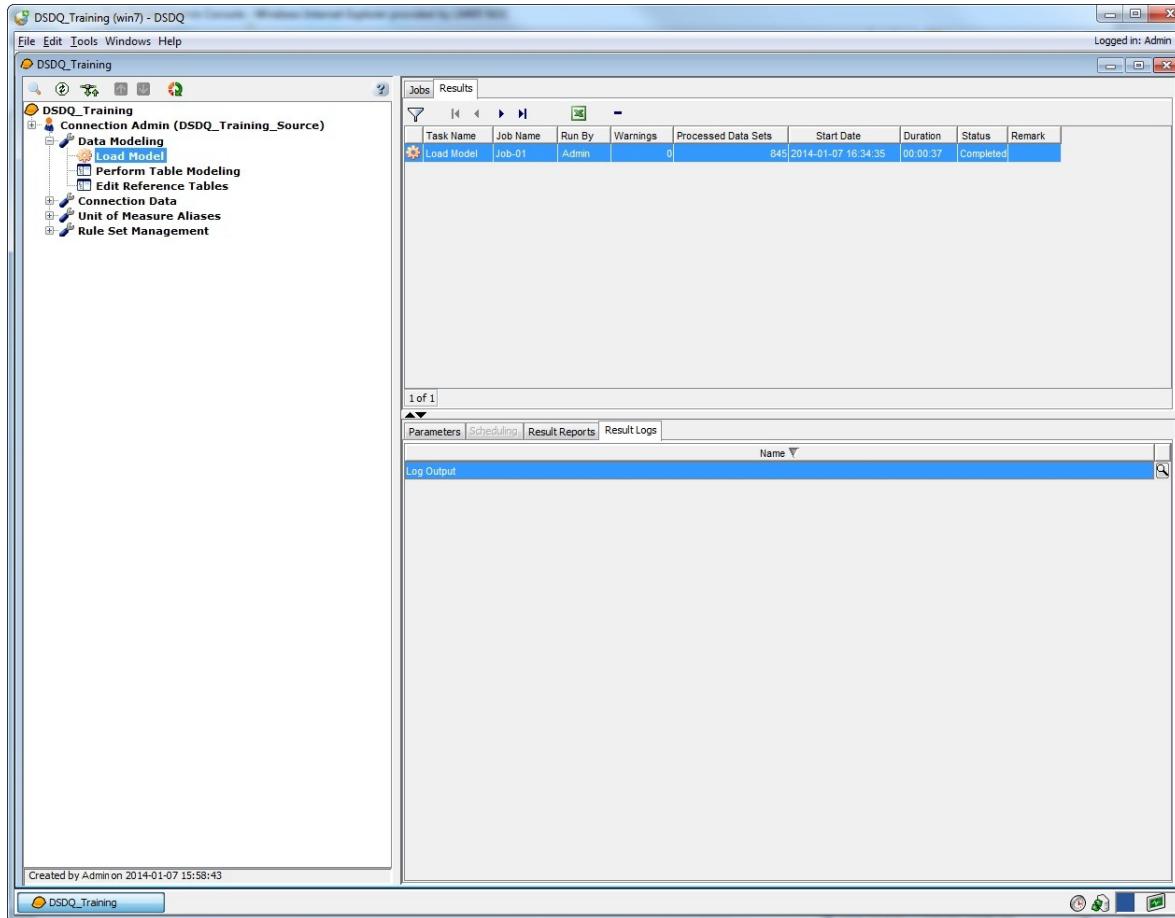
A new job is initiated and displays on the **Jobs and Results Listing Pane**.



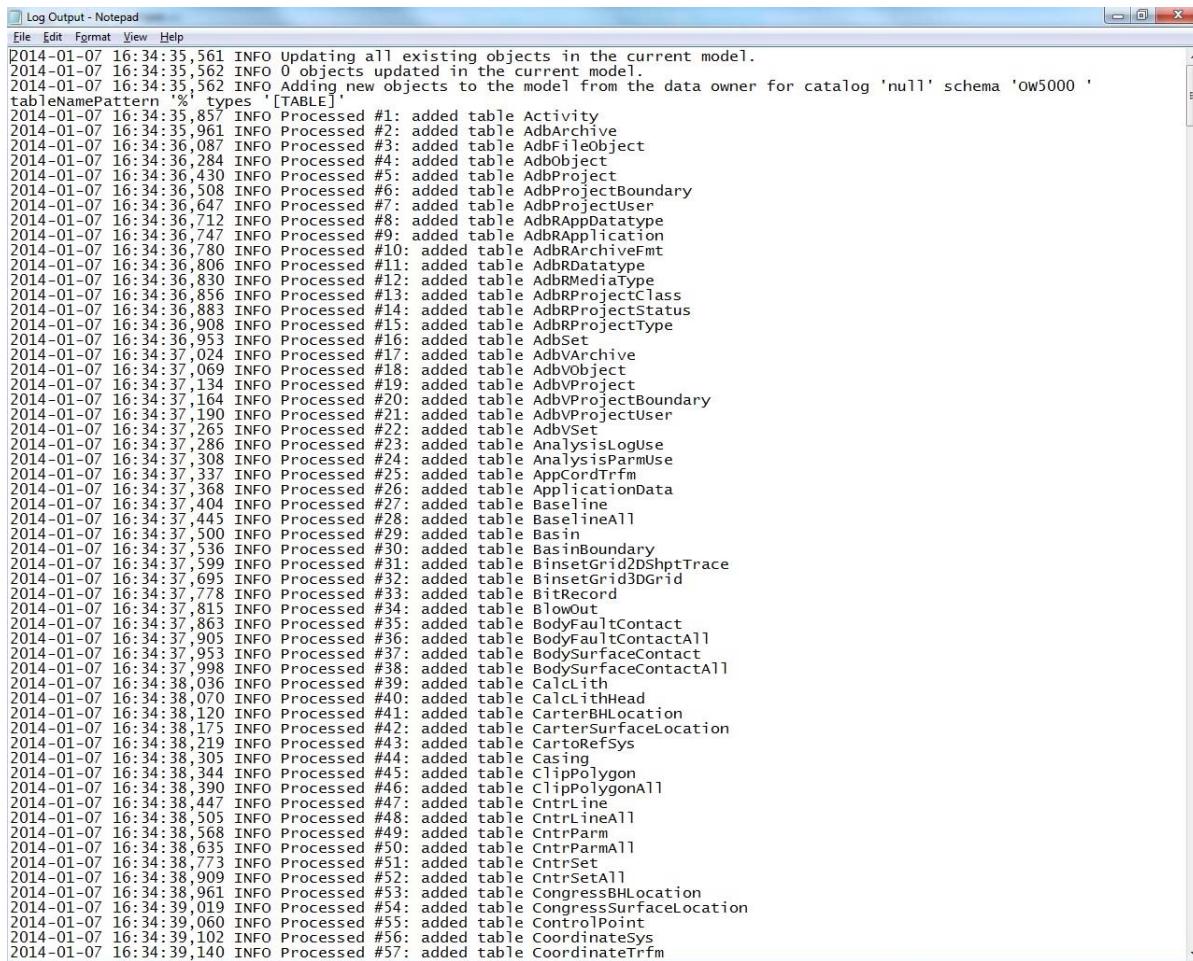
3. Enter **Job-01** in the **Job Name** field.
4. Enter **Load Model** in the **Job Description** field.
5. Select the **Table** option for **Object Types to Load**.
6. Select the **All Names** option for **Object Name Filter**.
7. Select the **Owner's Schema** option for **Schema Filter**. The **Catalog Filter** field will remain unchanged.
8. Select the **After** option for **Delete Results?** Leave the number of days to **7**.
9. Click to save changes in the **Parameters** tab.

10. Click .

The **Load Model** task runs and displays results in the **Result Logs** tab of the **Job and Results Information Pane**.

11. Select the **Results** tab on the **Job and Results Listing Pane** to view the values in the **Result Logs** tab on the **Job and Results Information Pane**.

12. Double-click on Log Output on the Jobs and Results Information Pane to display the Log Output in TXT format.



The screenshot shows a Windows Notepad window titled "Log Output - Notepad". The menu bar includes File, Edit, Format, View, and Help. The content of the window is a log of database operations from January 7, 2014, at 16:34:35. The log entries are as follows:

```

2014-01-07 16:34:35,561 INFO Updating all existing objects in the current model.
2014-01-07 16:34:35,562 INFO 0 objects updated in the current model.
2014-01-07 16:34:35,562 INFO Adding new objects to the model from the data owner for catalog 'null' schema 'ow5000'
tableNamePattern '%_types' [TABLE]
2014-01-07 16:34:35,857 INFO Processed #1: added table Activity
2014-01-07 16:34:35,961 INFO Processed #2: added table AdbArchive
2014-01-07 16:34:36,087 INFO Processed #3: added table AdbFileObject
2014-01-07 16:34:36,284 INFO Processed #4: added table AdboObject
2014-01-07 16:34:36,430 INFO Processed #5: added table AdbProject
2014-01-07 16:34:36,508 INFO Processed #6: added table AdbProjectBoundary
2014-01-07 16:34:36,647 INFO Processed #7: added table AdbProjectUser
2014-01-07 16:34:36,712 INFO Processed #8: added table AdbRAppDatatype
2014-01-07 16:34:36,747 INFO Processed #9: added table AdbRApplication
2014-01-07 16:34:36,780 INFO Processed #10: added table AdbRArchiveFmt
2014-01-07 16:34:36,806 INFO Processed #11: added table AdbRDatatype
2014-01-07 16:34:36,830 INFO Processed #12: added table AdbRMediaType
2014-01-07 16:34:36,856 INFO Processed #13: added table AdbRProjectClass
2014-01-07 16:34:36,883 INFO Processed #14: added table AdbRProjectStatus
2014-01-07 16:34:36,908 INFO Processed #15: added table AdbRProjectType
2014-01-07 16:34:36,953 INFO Processed #16: added table AdbSet
2014-01-07 16:34:37,024 INFO Processed #17: added table AdbArchive
2014-01-07 16:34:37,069 INFO Processed #18: added table AdbVobject
2014-01-07 16:34:37,134 INFO Processed #19: added table AdbVProject
2014-01-07 16:34:37,164 INFO Processed #20: added table AdbVProjectBoundary
2014-01-07 16:34:37,190 INFO Processed #21: added table AdbVProjectUser
2014-01-07 16:34:37,265 INFO Processed #22: added table AdbVSet
2014-01-07 16:34:37,286 INFO Processed #23: added table AnalysisLogUse
2014-01-07 16:34:37,308 INFO Processed #24: added table AnalysisParmUse
2014-01-07 16:34:37,337 INFO Processed #25: added table AppCordTrfm
2014-01-07 16:34:37,368 INFO Processed #26: added table ApplicationData
2014-01-07 16:34:37,404 INFO Processed #27: added table Baseline
2014-01-07 16:34:37,445 INFO Processed #28: added table BaselineAll
2014-01-07 16:34:37,500 INFO Processed #29: added table Basin
2014-01-07 16:34:37,536 INFO Processed #30: added table BasinBoundary
2014-01-07 16:34:37,599 INFO Processed #31: added table BinsetGrid2DShptTrace
2014-01-07 16:34:37,695 INFO Processed #32: added table BinsetGrid3DGrid
2014-01-07 16:34:37,778 INFO Processed #33: added table BitRecord
2014-01-07 16:34:37,815 INFO Processed #34: added table Blowout
2014-01-07 16:34:37,863 INFO Processed #35: added table BodyFaultContact
2014-01-07 16:34:37,905 INFO Processed #36: added table BodyFaultContactAll
2014-01-07 16:34:37,953 INFO Processed #37: added table BodysurfaceContact
2014-01-07 16:34:37,998 INFO Processed #38: added table BodysurfaceContactAll
2014-01-07 16:34:38,036 INFO Processed #39: added table Calclith
2014-01-07 16:34:38,070 INFO Processed #40: added table CalclithHead
2014-01-07 16:34:38,120 INFO Processed #41: added table CarterBHLlocation
2014-01-07 16:34:38,175 INFO Processed #42: added table CarterSurfaceLocation
2014-01-07 16:34:38,219 INFO Processed #43: added table CartoRefSys
2014-01-07 16:34:38,305 INFO Processed #44: added table Casing
2014-01-07 16:34:38,344 INFO Processed #45: added table ClipPolygon
2014-01-07 16:34:38,390 INFO Processed #46: added table ClipPolygonAll
2014-01-07 16:34:38,447 INFO Processed #47: added table CntrLine
2014-01-07 16:34:38,505 INFO Processed #48: added table CntrLineAll
2014-01-07 16:34:38,568 INFO Processed #49: added table CntrParm
2014-01-07 16:34:38,635 INFO Processed #50: added table CntrParmAll
2014-01-07 16:34:38,773 INFO Processed #51: added table CntrSet
2014-01-07 16:34:38,909 INFO Processed #52: added table CntrSetAll
2014-01-07 16:34:38,961 INFO Processed #53: added table CongressBHLlocation
2014-01-07 16:34:39,019 INFO Processed #54: added table CongressSurfaceLocation
2014-01-07 16:34:39,060 INFO Processed #55: added table ControlPoint
2014-01-07 16:34:39,102 INFO Processed #56: added table CoordinateSys
2014-01-07 16:34:39,140 INFO Processed #57: added table CoordinateTrfm

```

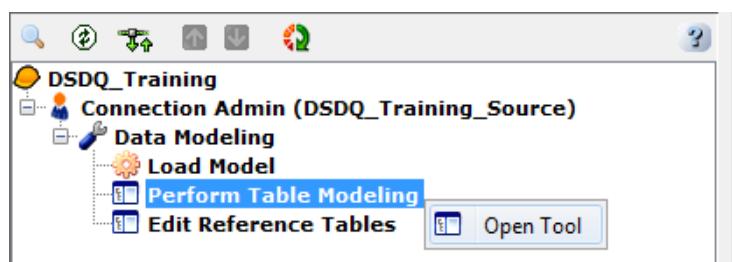
13. Click File > Exit to close the log file.

Exercise: Perform Table Modeling

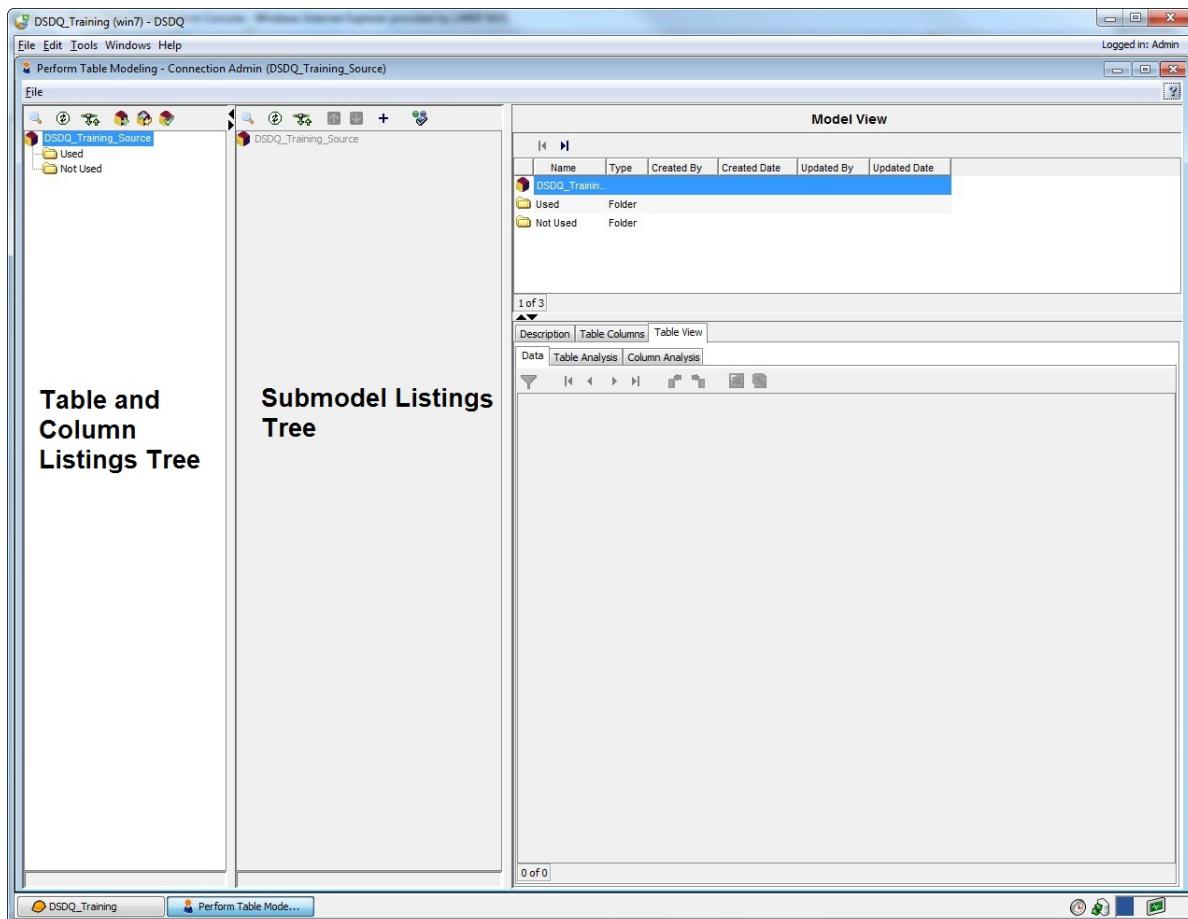
To Perform Table Modeling:

1. Double-click the **Perform Table Modeling** Tool or right-click the **Perform Table Modeling** Tool and select **Open Tool** from the pop-

up menu.

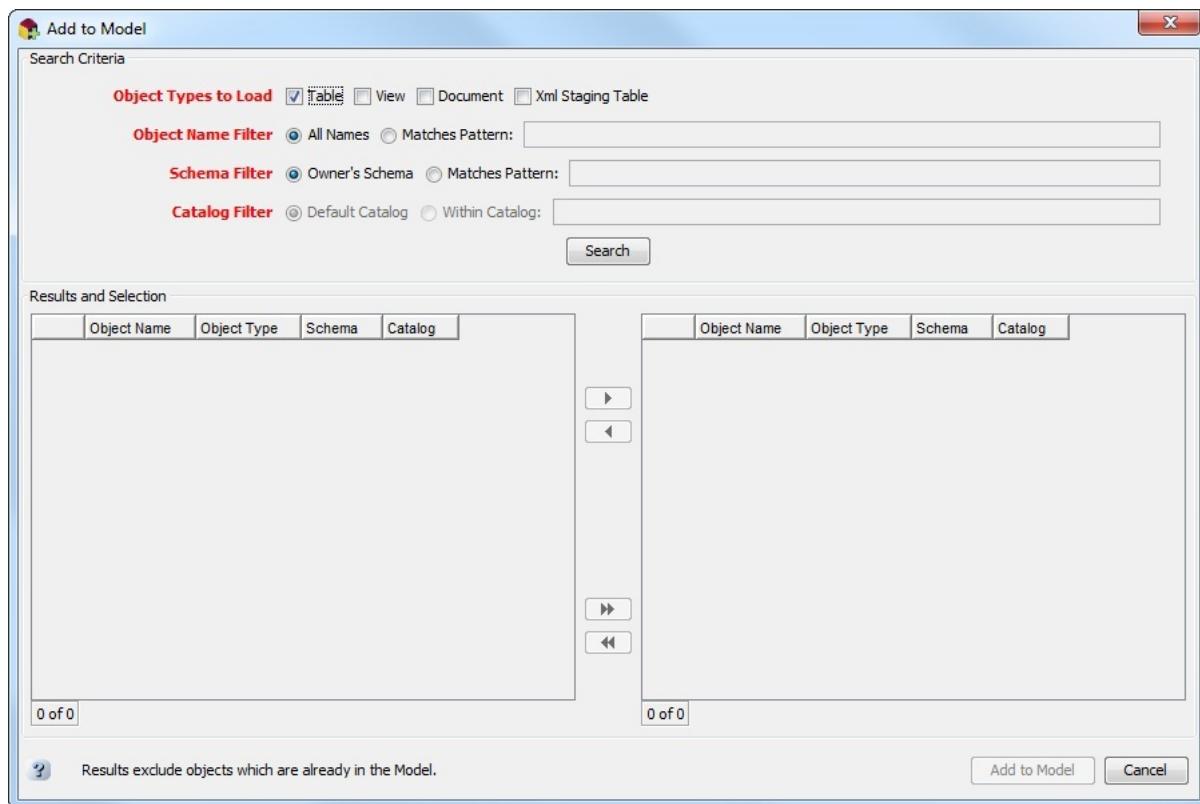


The **Perform Table Modeling** screen appears.



2. Click to add database objects to the model.

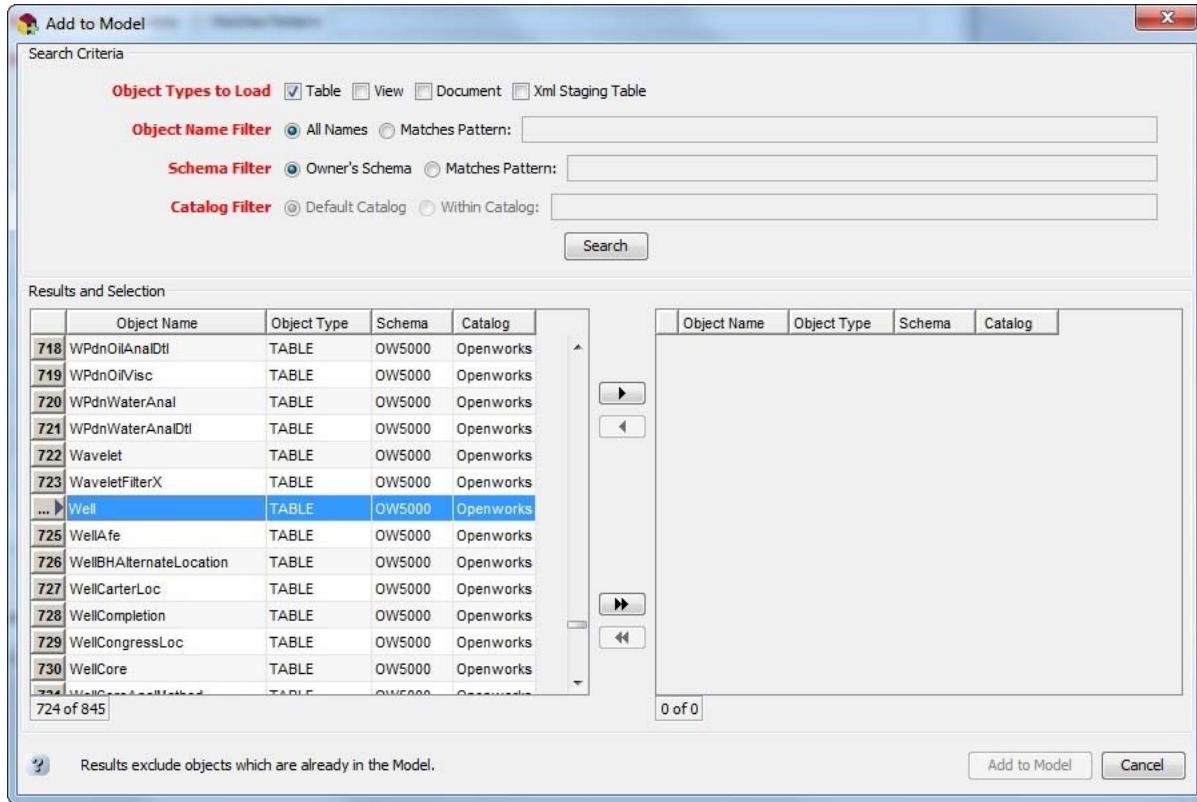
The **Add to Model** window appears.



3. Select the **Table** option for **Object Types to Load**.
4. Select the **All Names** option for **Object Name Filter**.
5. Select the **Owner's Schema** option for **Schema Filter**.
6. The **Catalog Filter** options will be disabled for the current selection.

7. Click Search.

The search results appear based on your search criteria.

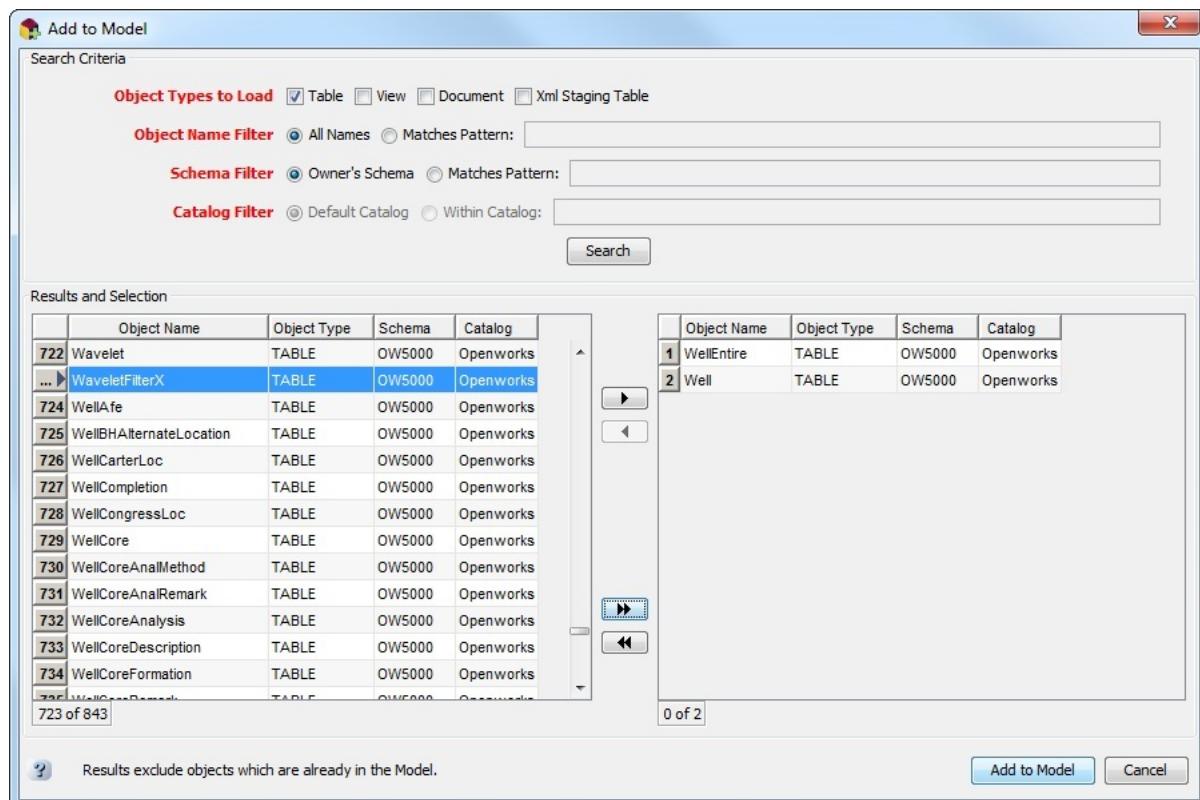


8. Browse through the database objects and select Well and WellEntire.

Note

Press <Ctrl> key, and click on the required objects to select them.

9. Click ➤ to send selected objects to the table on right side.

10. Click **Add to Model** to add selected database objects to the model

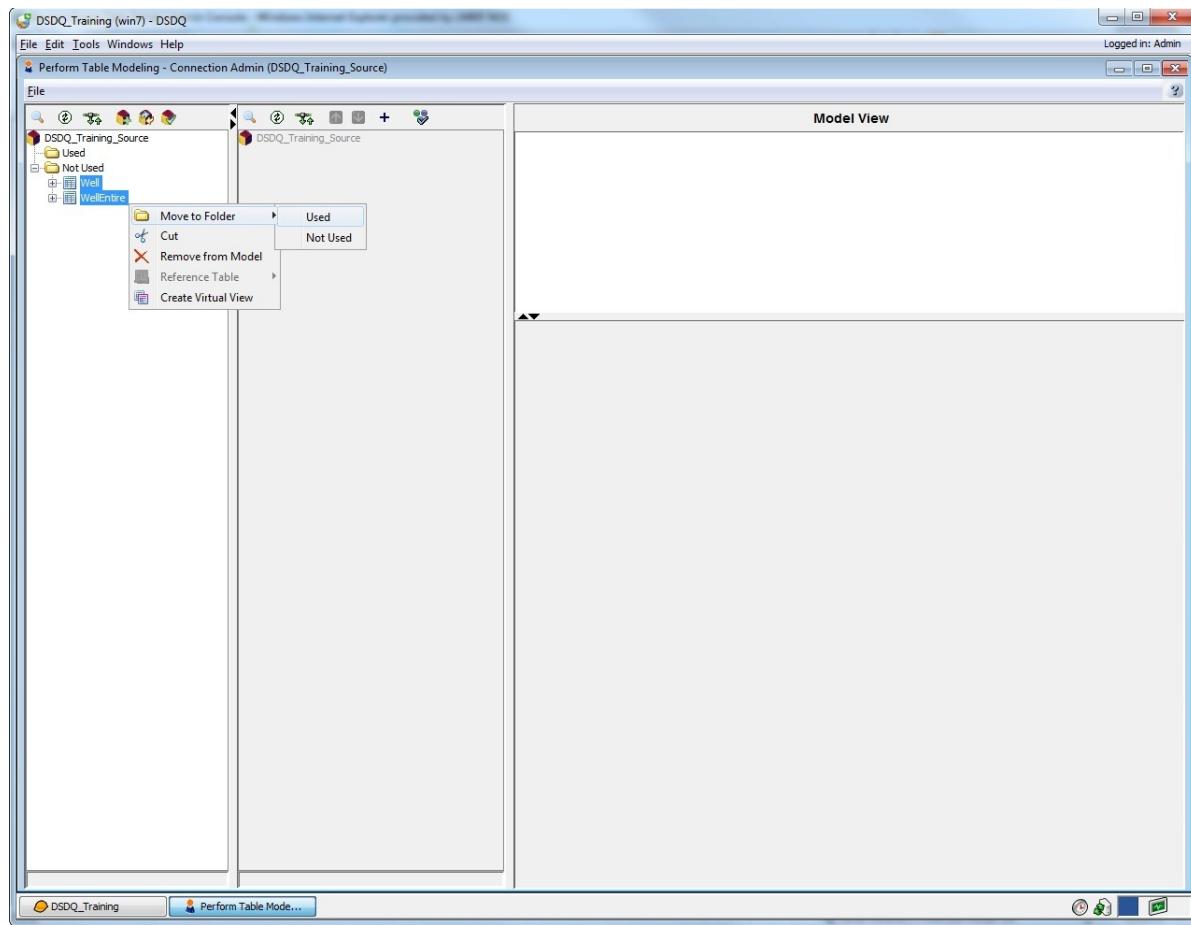
The selected database objects appear under the **Not Used** tree.

The screenshot shows the DSDQ Training software interface. The main window title is "DSDQ_Training (win7) - DSDQ". The menu bar includes File, Edit, Tools, Windows, and Help. A status bar at the bottom right says "Logged in: Admin". The left pane displays a tree view of database objects under "DSDQ_Training_Source": "Used" (containing "Well" and "WellEntire") and "Not Used". The right pane is titled "Model View" and contains a table with the following data:

Name	Type	Created By	Created Date	Updated By	Updated Date
Not Used	Folder				
Well	Table	Admin	2014-01-07 16:45:00		
WellEntire	Table	Admin	2014-01-07 16:45:00		

Below the table, there are tabs for "Description", "Table Columns", and "Table View". The "Table View" tab is selected. At the bottom of the right pane, there are buttons for "Data", "Table Analysis", and "Column Analysis". The status bar at the bottom left shows "0 of 0".

11. Select the database objects and right-click them. Select the **Move to Folder > Used** option from the pop-up menu.



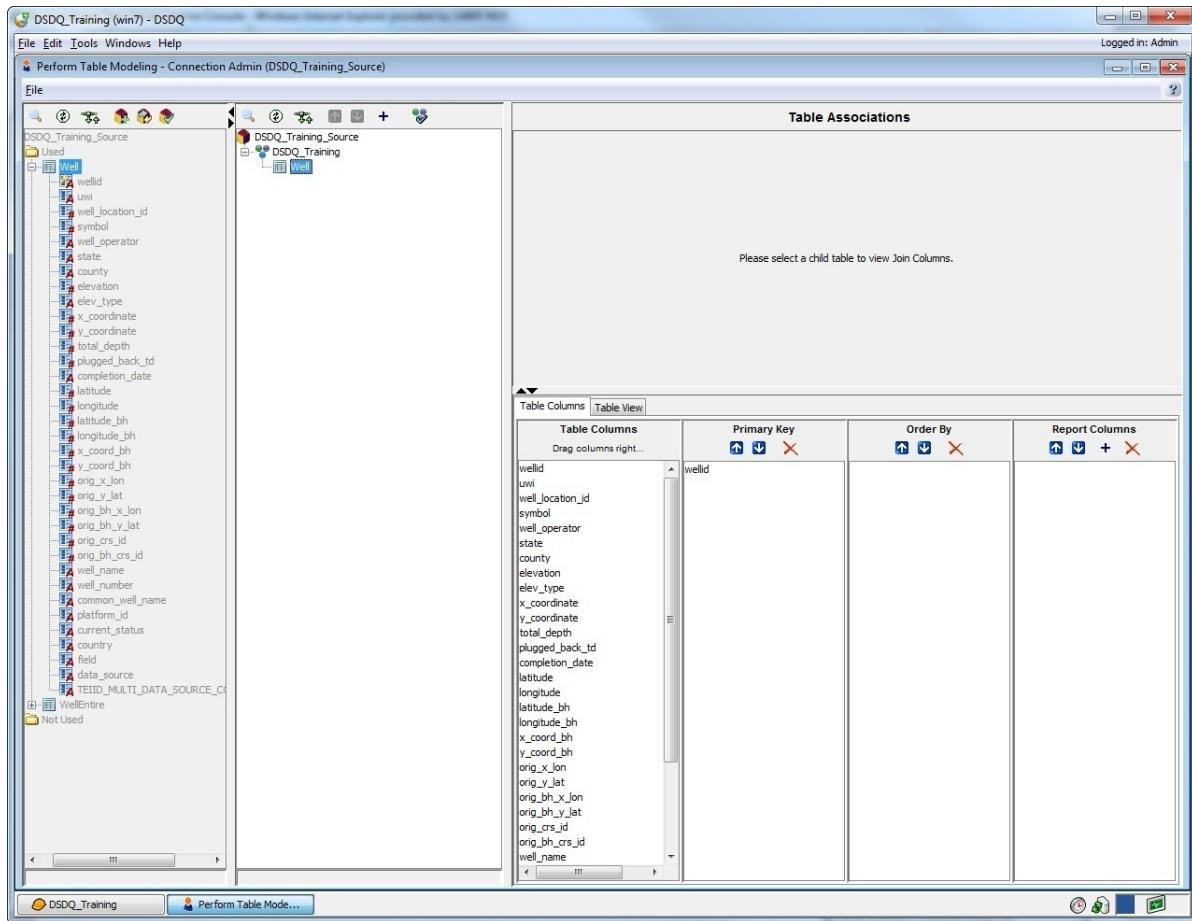
The selected objects now appear under the **Used** tree.

The screenshot shows the DSDQ_Training software interface. The main window title is "DSDQ_Training (win7) - DSDQ". The menu bar includes File, Edit, Tools, Windows, and Help. A toolbar with various icons is visible. On the left, a tree view under "DSDQ_Training_Source" shows "Used" expanded, with "Well" selected. The right side features a "Model View" pane with a table titled "Used". The table has columns: Name, Type, Created By, Created Date, Updated By, and Updated Date. It contains two rows: "Well Table Admin 2014-01-07 16:45:00 Admin 2014-01-07 16:46:21" and "WellEntire Table Admin 2014-01-07 16:45:00 Admin 2014-01-07 16:46:21". Below the table is a large data grid titled "Data" with columns: wellid, uwI, well_location_id, symbol, well_operator, state, county, elevation, and elev_ty. The grid displays 50 rows of well data from 1 to 50. At the bottom of the grid, it says "0 of 50".

12. Click adjacent to **Well** on the **Submodel Listings Pane**.
The **Add Submodel** dialog box appears.



13. Enter **DSDQ_Training** in the **Enter a name for the new submodel** field.

14. Click **OK**.

15. Drag the **Well** table from **Table and Column Listings Tree** to **Submodel Listings Tree**, under the newly created submodel **DSDQ_Training**.

16. From **Table Columns** pane, select the following objects:

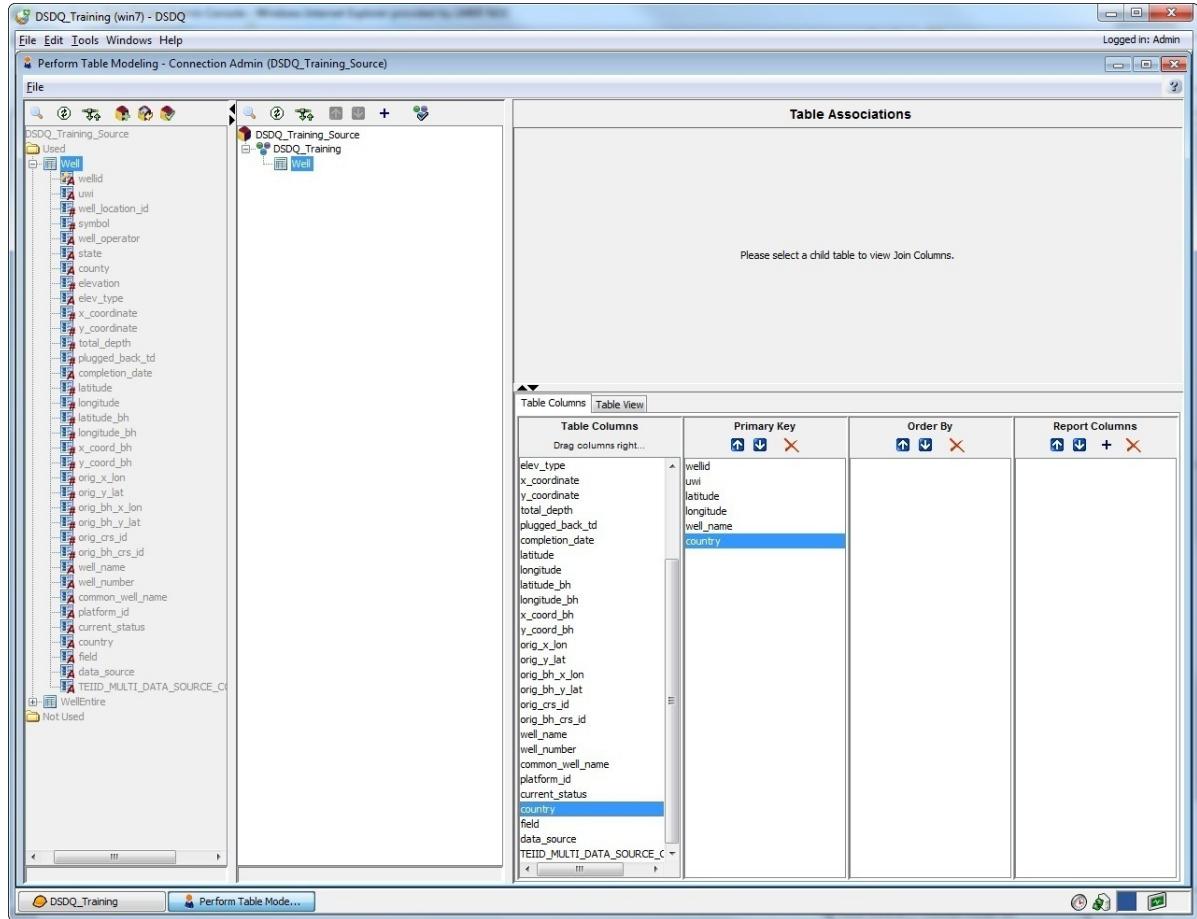
- wellid
- uwi
- latitude
- longitude
- well_name

- country

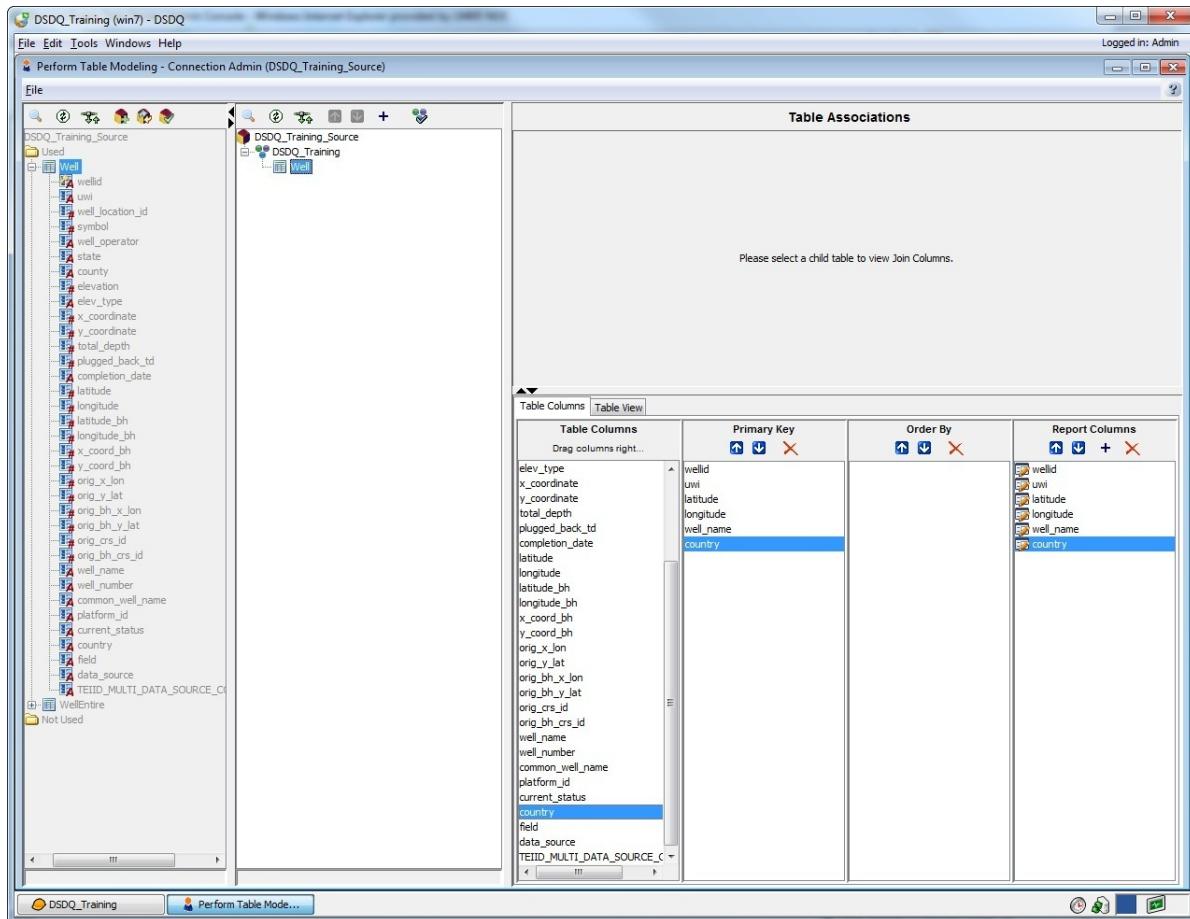
Note

Press <Ctrl> key, and click on the required objects to select them.

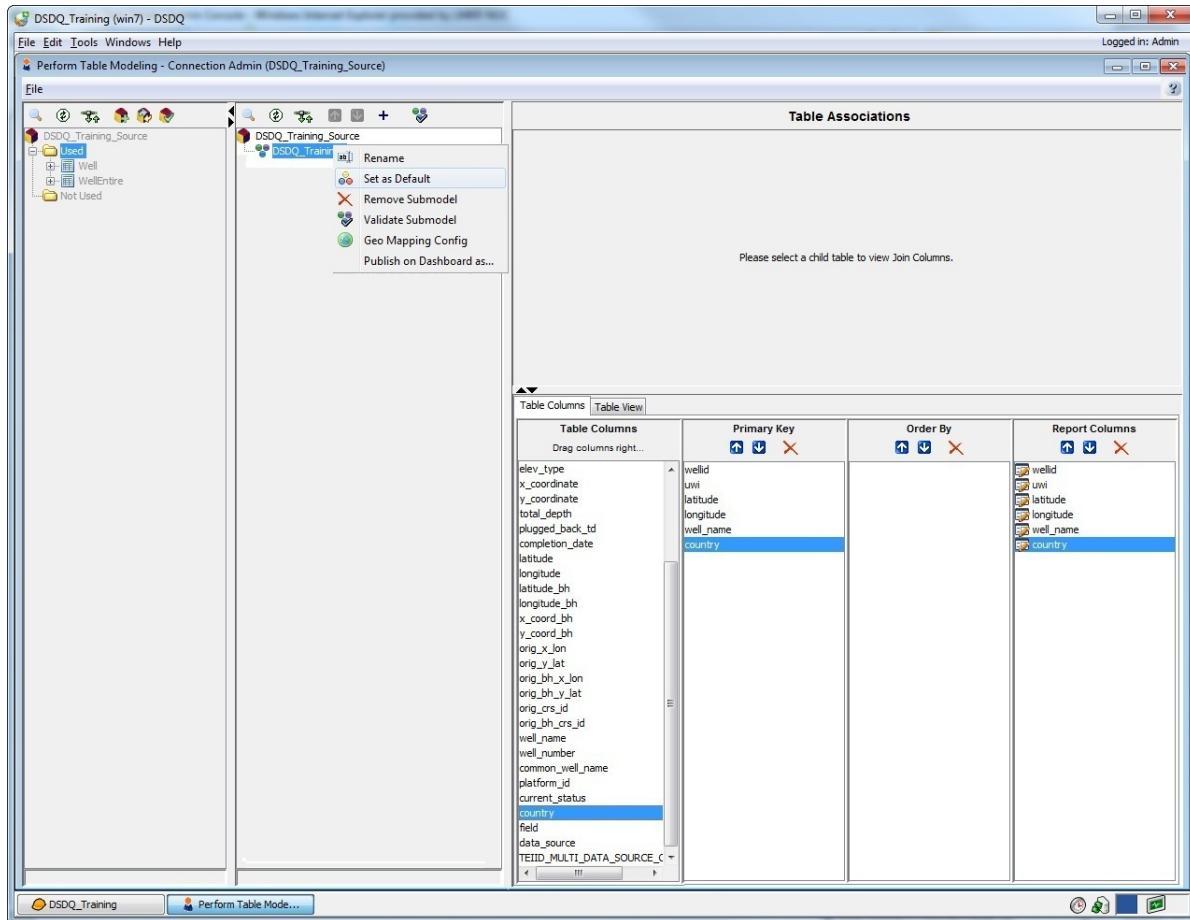
17. Drag all the selected objects to the Primary Key pane.



18. Similarly, move all these objects to the **Report Columns** pane as well.

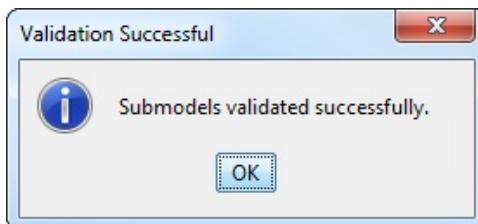


19. Right-click on the **DSDQ_Training** submodel and select **Set as Default** from the pop-up menu.



The submodel's icon will change from to .

20. Click to validate the submodel.

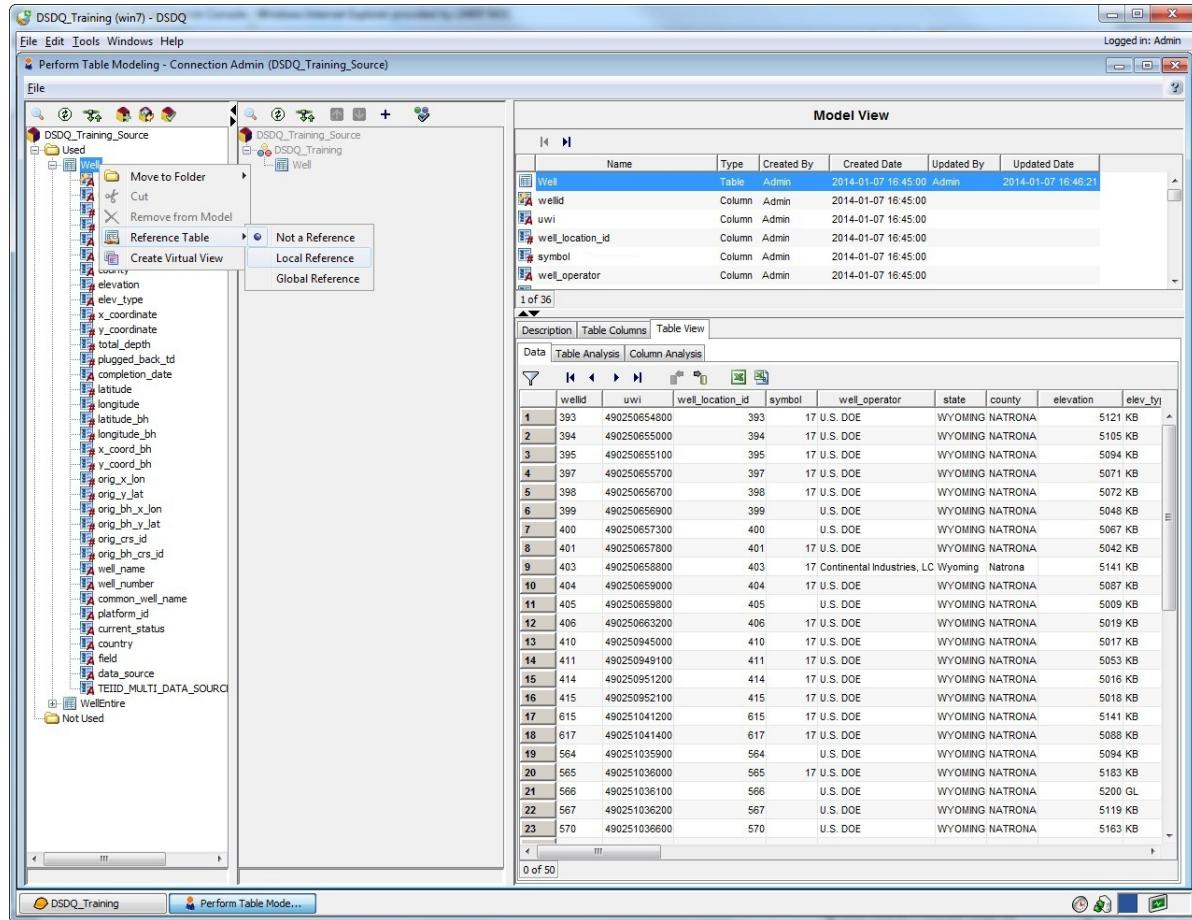


21. Click **OK**.

22. Right-click on the **Well** table in the **Table and Column Listings Tree**, and select **Reference Table > Local Reference** from the

pop-up menu.

The Well database object's icon will change from  to .



23. Select **File > Exit** to close the **Perform Table Modeling** window.

Exercise: Editing a Reference Table

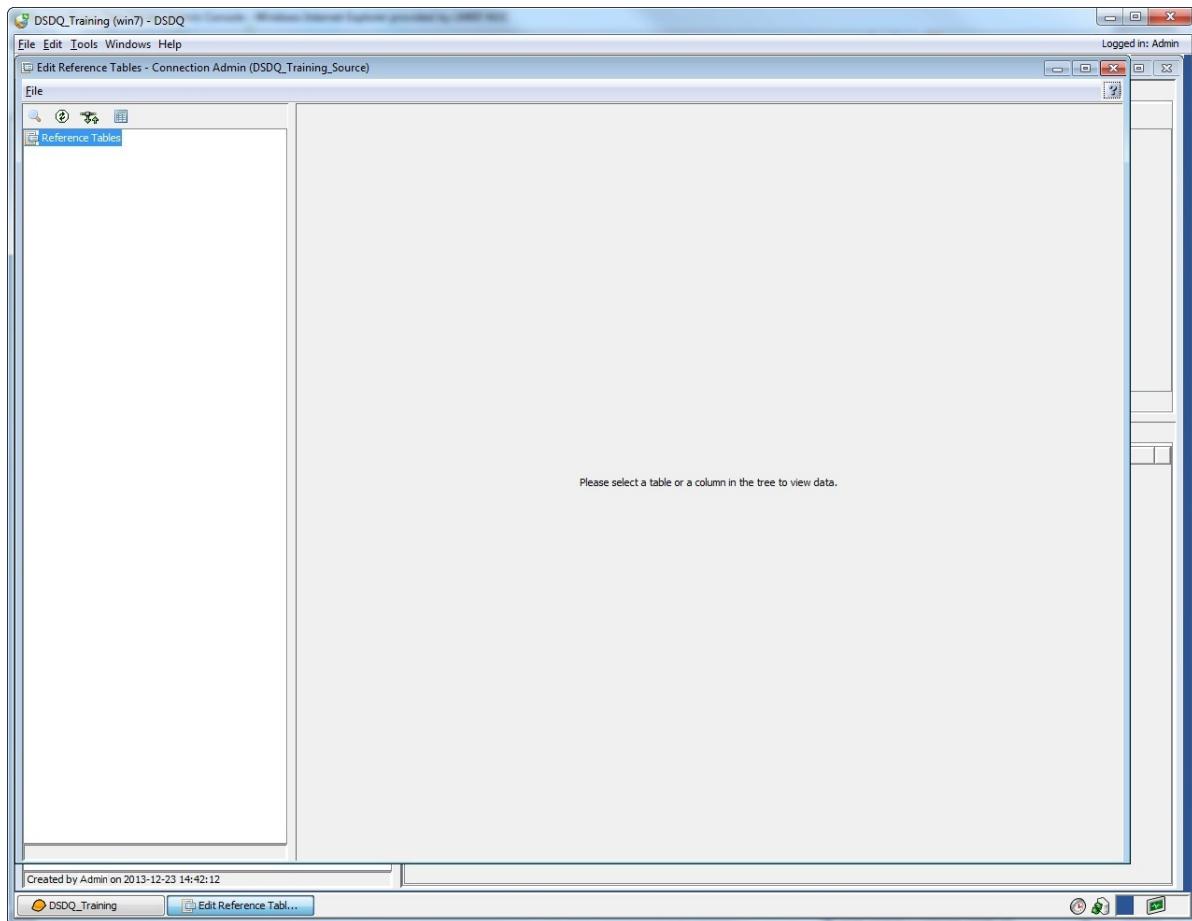
The **Edit Reference Tables** Tool allows the modification of data in application reference tables.

To edit a reference table:

1. Double-click the **Edit Reference Tables** Tool or right-click the **Edit Reference Tables** Tool and select **Open Tool** from the pop-up

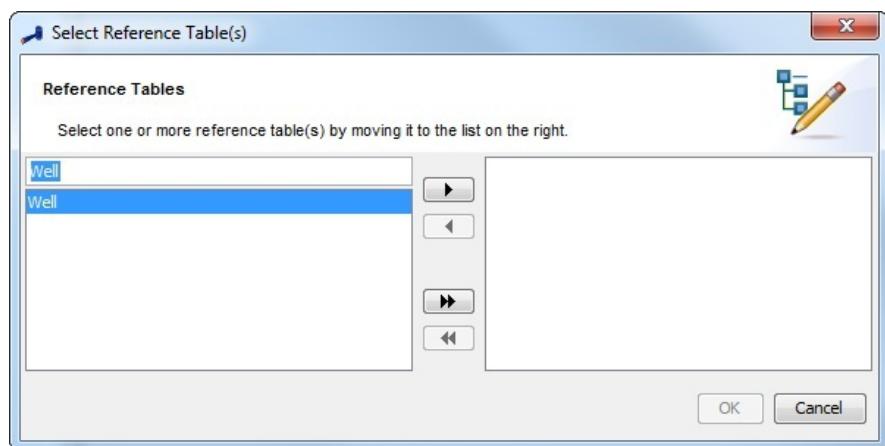
menu.

The **Edit Reference Tables** window appears.

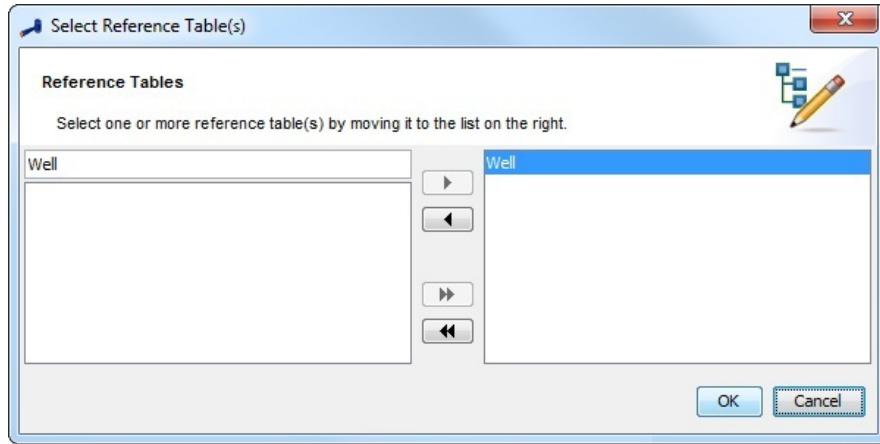


2. Click to select reference tables.

The **Select Reference Tables** dialog box appears.



3. Select the Well table and click ► to move it to the selected tables list.



4. Click OK.

The Edit Reference Tables screen appears.

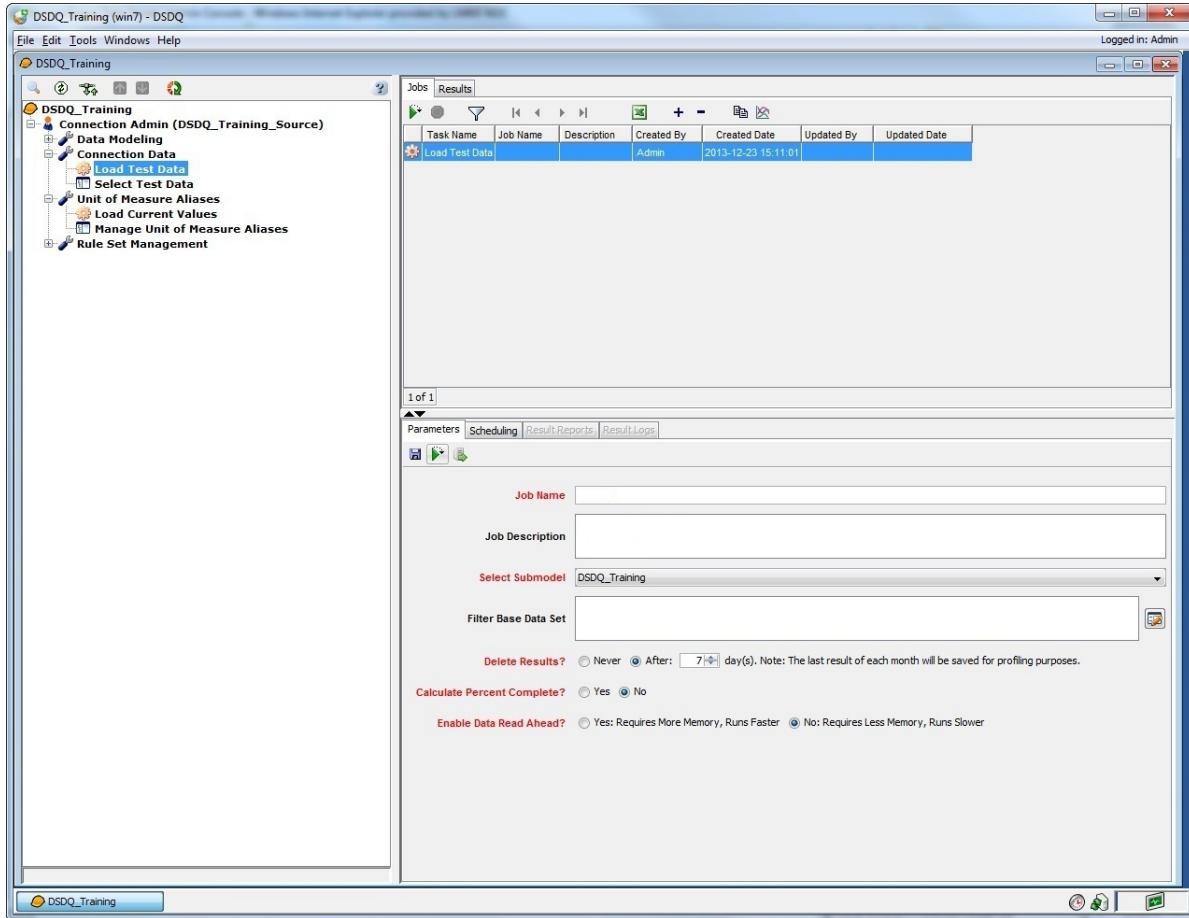
5. Select File > Exit to close the Edit Reference Tables window.

Exercise: Loading Test Data

The **Load Test Data** Task loads the source submodel data into the test data tables.

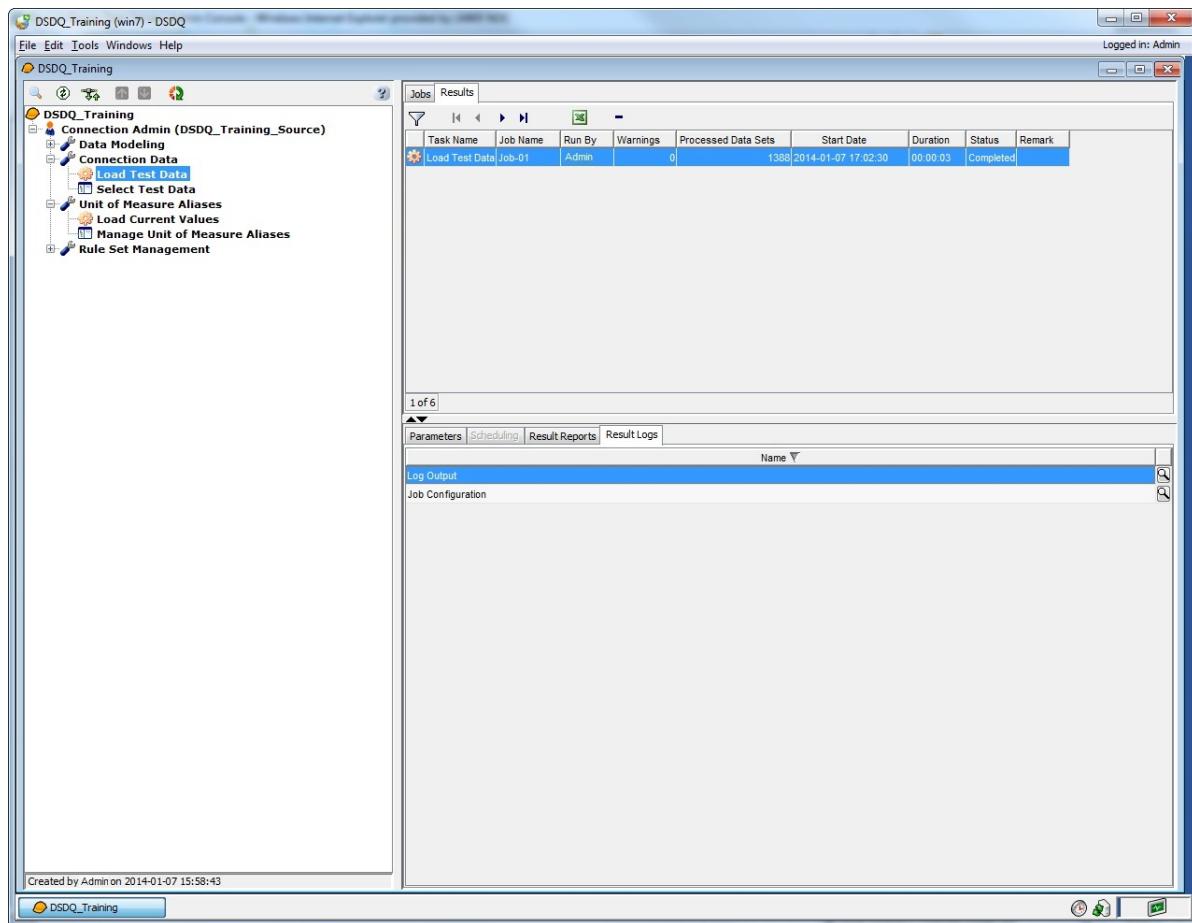
To load test data:

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



2. Enter **Job-01** in the **Job Name** field.
3. Enter **Load Test Data** in the **Job Description** field.
4. Select **DSDQ_Training** from the **Select Submodel** drop-down list.
5. Optionally, set a filter on the data subset.
6. Select the **After** option for **Delete Results?** Change the number of days to **7**.

7. Select the **No** option for **Calculate Percent Complete?**
8. Select the **No** option for **Enable Data Read Ahead?**
9. Click  to save changes in the **Parameters** tab.
10. Click .
11. Select the **Results** tab.
The **Load Test Data** task runs and displays results in the **Result Logs** tab on the **Jobs and Results Information Pane**.



12. Double-click the Log Output file on the Jobs and Results Information Pane to display the Log Output in TXT format.

```

2014-01-07 17:02:30,912 INFO Processed data set #1
2014-01-07 17:02:30,915 INFO Processed data set #2
2014-01-07 17:02:30,917 INFO Processed data set #3
2014-01-07 17:02:30,919 INFO Processed data set #4
2014-01-07 17:02:30,922 INFO Processed data set #5
2014-01-07 17:02:30,924 INFO Processed data set #6
2014-01-07 17:02:30,926 INFO Processed data set #7
2014-01-07 17:02:30,928 INFO Processed data set #8
2014-01-07 17:02:30,930 INFO Processed data set #9
2014-01-07 17:02:30,932 INFO Processed data set #10
2014-01-07 17:02:30,934 INFO Processed data set #11
2014-01-07 17:02:30,936 INFO Processed data set #12
2014-01-07 17:02:30,939 INFO Processed data set #13
2014-01-07 17:02:30,941 INFO Processed data set #14
2014-01-07 17:02:30,943 INFO Processed data set #15
2014-01-07 17:02:30,945 INFO Processed data set #16
2014-01-07 17:02:30,947 INFO Processed data set #17
2014-01-07 17:02:30,949 INFO Processed data set #18
2014-01-07 17:02:30,950 INFO Processed data set #19
2014-01-07 17:02:30,952 INFO Processed data set #20
2014-01-07 17:02:30,954 INFO Processed data set #21
2014-01-07 17:02:30,956 INFO Processed data set #22
2014-01-07 17:02:30,959 INFO Processed data set #23
2014-01-07 17:02:30,961 INFO Processed data set #24
2014-01-07 17:02:30,963 INFO Processed data set #25
2014-01-07 17:02:30,965 INFO Processed data set #26
2014-01-07 17:02:30,967 INFO Processed data set #27
2014-01-07 17:02:30,969 INFO Processed data set #28
2014-01-07 17:02:30,971 INFO Processed data set #29
2014-01-07 17:02:30,973 INFO Processed data set #30
2014-01-07 17:02:30,976 INFO Processed data set #31
2014-01-07 17:02:30,978 INFO Processed data set #32
2014-01-07 17:02:30,980 INFO Processed data set #33
2014-01-07 17:02:30,982 INFO Processed data set #34
2014-01-07 17:02:30,984 INFO Processed data set #35
2014-01-07 17:02:30,987 INFO Processed data set #36
2014-01-07 17:02:30,988 INFO Processed data set #37

```

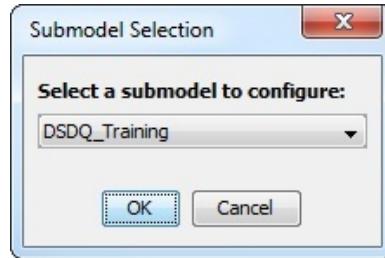
13. Select File > Exit to close the log file.

Exercise: Selecting Test Data

The **Select Test Data** Tool helps in selecting data subset for performing the **HealthCheck** and **Clean** Phases. The **Load Test Data** Task must be run on the submodel prior to running this Tool.

To select test data:

- Double-click the **Select Test Data** Tool or right-click the **Select Test Data** Tool and select **Open Tool** from the pop-up menu. The **Submodel Selection** dialog box appears.



Note

Only Submodels that have been loaded will be displayed in the drop-down list. Refer to **Loading Test Data** for more details.

- Select **DSDQ_Training** from the **Select a submodel to configure** drop-down list.
- Click **OK**.

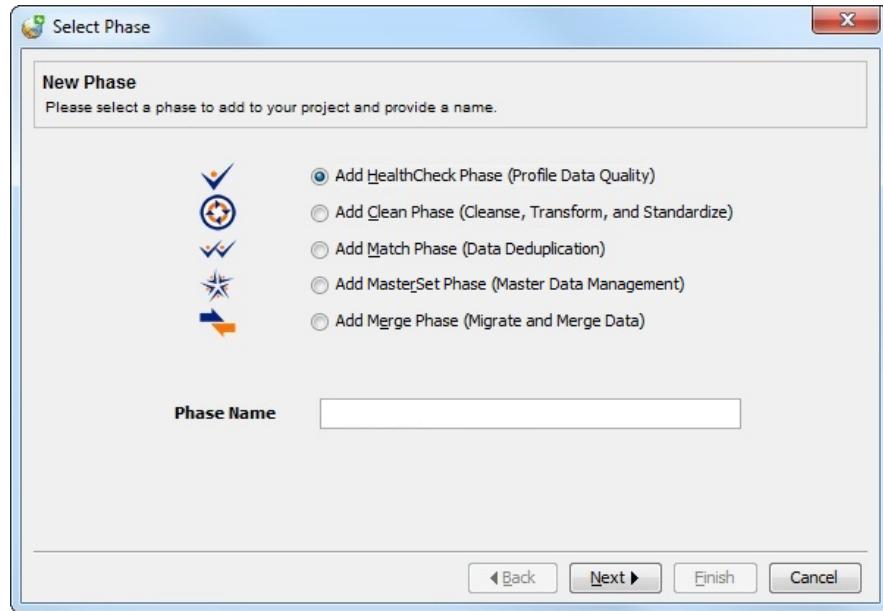
Source Key	HealthCheck?	Clean?	Match?	Merge?	Remark	Created Date	Created By	Updated Date
501u	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		2014-01-07 17:02:32	Admin	2014-01-09 18:...
502u	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		2014-01-07 17:02:32	Admin	2014-01-09 18:...
503u	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		2014-01-07 17:02:32	Admin	2014-01-09 18:...
504u	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		2014-01-07 17:02:32	Admin	2014-01-09 18:...
505u	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		2014-01-07 17:02:32	Admin	2014-01-09 18:...
506u	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		2014-01-07 17:02:32	Admin	2014-01-09 18:...
507u	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		2014-01-07 17:02:32	Admin	2014-01-09 18:...
508u	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		2014-01-07 17:02:33	Admin	2014-01-09 18:...
509	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:32	Admin	
510	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:32	Admin	
511	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:32	Admin	
512	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:32	Admin	
513	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:32	Admin	
514	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:31	Admin	
515	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:32	Admin	
516	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:33	Admin	
517	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:33	Admin	
518	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:31	Admin	
519	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:31	Admin	
520	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:32	Admin	
521	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:31	Admin	
522	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:31	Admin	
523	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:31	Admin	
524	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:31	Admin	
525	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:33	Admin	
526	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:31	Admin	
527	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:31	Admin	
528	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:31	Admin	
529	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:33	Admin	
530	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:31	Admin	
531	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:33	Admin	
532	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:31	Admin	
533	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:33	Admin	
534	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:33	Admin	
535	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:31	Admin	
536	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2014-01-07 17:02:31	Admin	

4. Select the check box for the records that will be used for testing purposes, from the **HealthCheck?**, **Clean?**, **Match?**, and **Merge?** columns. You can select check boxes for multiple phases that will be used during the testing process.
5. Click  to save changes.
6. Click **File > Exit** to close the window.

Data Evaluation in DecisionSpace Data Quality

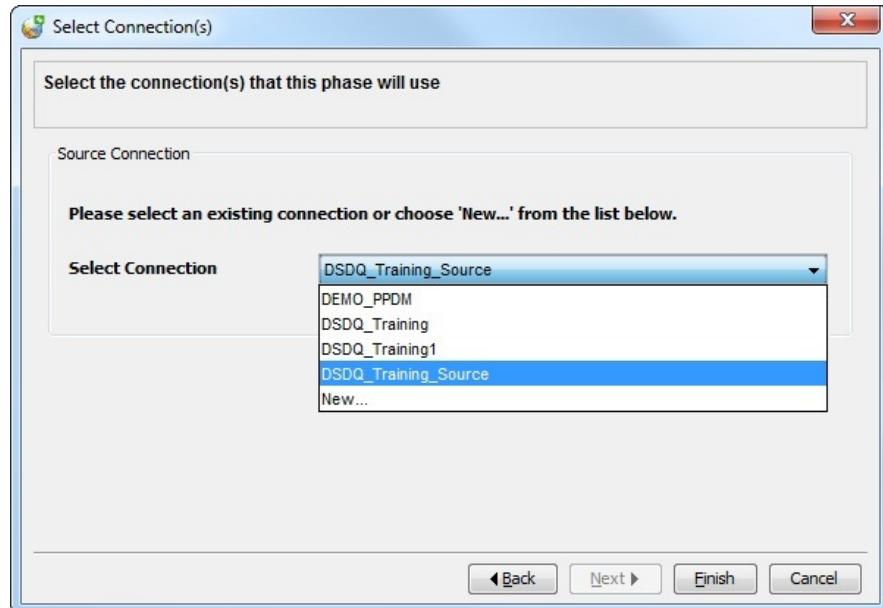
To add a HealthCheck Phase:

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. Enter **Training HealthCheck** in the **Phase Name** field.

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



4. Select **DSDQ_Training_Source** from the **Select Connection** drop-down list.
5. Click **Finish**.
The **HealthCheck** Phase is created and displayed in the **DecisionSpace Data Quality Project Window**.

Evaluating Data Volume and Quality

The **Rapid HealthCheck** Activity provides a quick look at the volume and quality of the data. It is fast to run and does not require a great deal of configuration. The **Run Table Analysis on All Tables** task does a simple row count, which is useful for identifying tables for modeling. The **Run Column Analysis on All Columns** task offers basic data profiling by checking the following parameters within a column:

- “Rows”: Number of rows
- “# Not Null”: Number of not null values
- “% Populated”: Percentage of rows populated
- “# Unique”: Number of unique values
- “Minimum Value”: Minimum value in the column
- “Maximum Value”: Maximum value in the column
- “# Mixed Case”: Number of values with mixed cases
- “# NPC”: Number of Non-Printable Characters
- “# PWS”: Number of Preceding White Spaces
- “# TWS”: Number of Trailing White Spaces
- “# DWS”: Number of Double White Spaces (between words)

Exercise: Profiling Data Using SQL Query

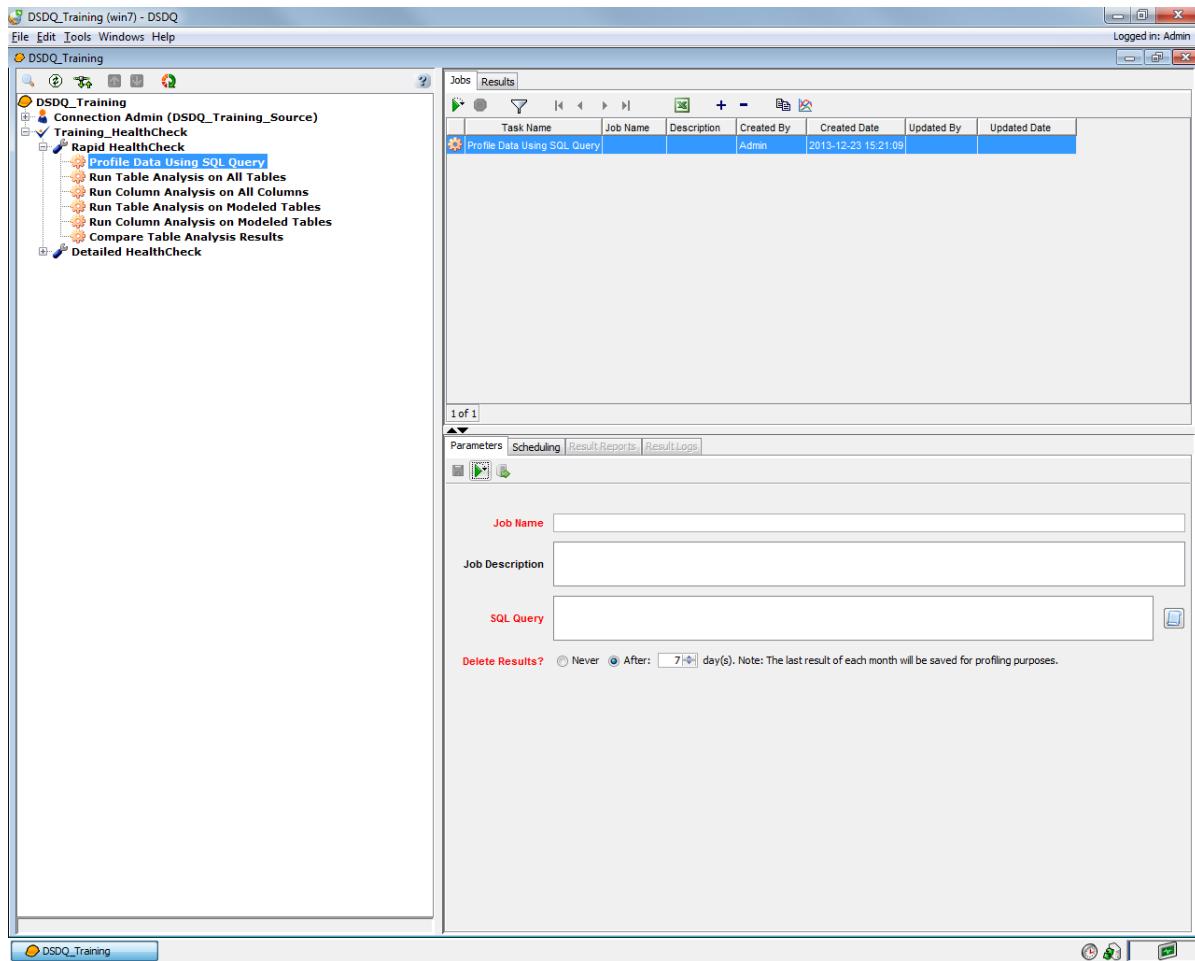
SQL queries can be customized (as per your requirement) and applied to the data source.

To profile data using SQL query:

1. Click  on the DecisionSpace Data Quality Tree to expand the **Rapid HealthCheck** Activity.
2. Double-click the **Profile Data Using SQL Query** Task or right-click the **Profile Data Using SQL Query** Task and select **Add Job**

from the pop-up menu.

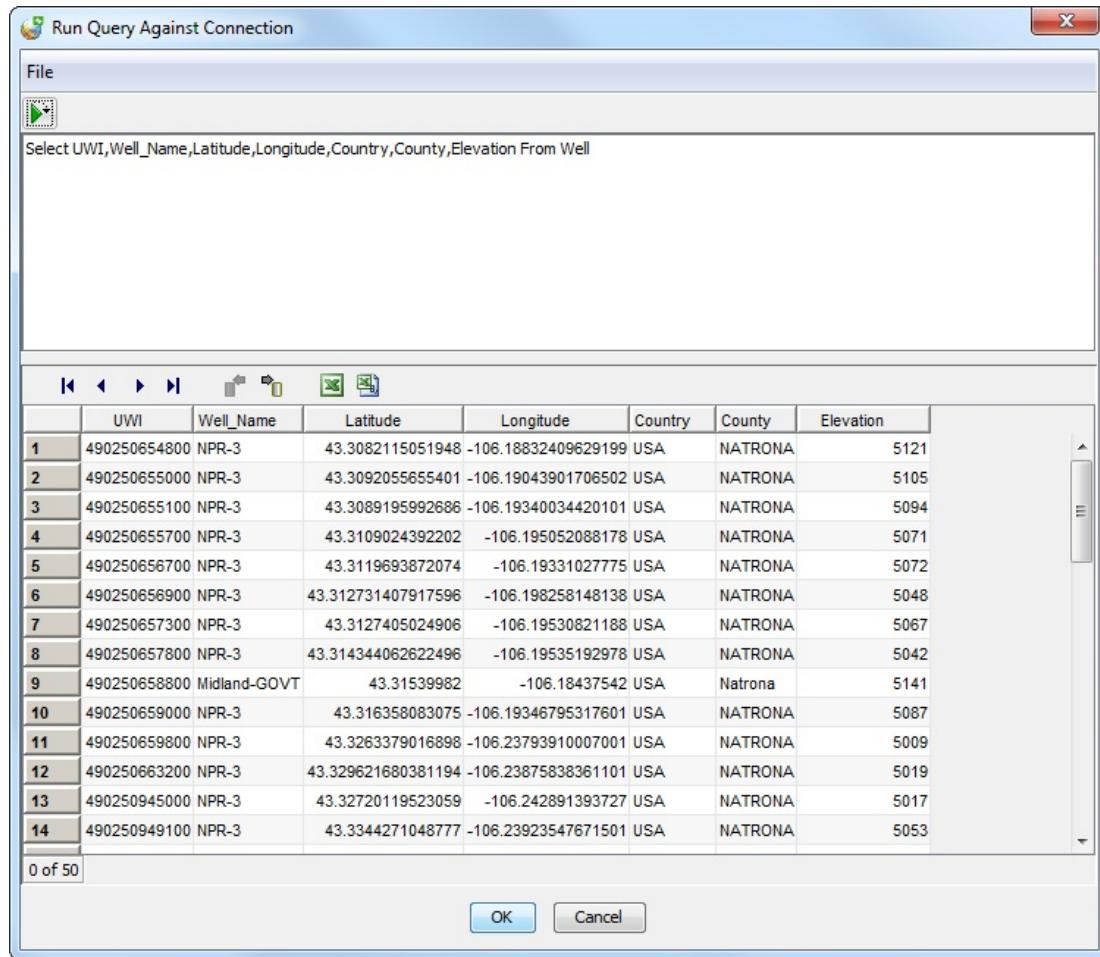
A new job is initiated and displays on the **Job and Results Information Pane** on the right side of the **DecisionSpace Data Quality Project Window**.



3. Enter **Job-01** in the **Job Name** field.
4. Enter **SQL Profiling** in the **Job Description** field.
5. Enter **Select UWI, Well_Name, Latitude, Longitude, Country, County, Elevation From Well** in the **SQL Query** field.
6. Select the **After** option for **Delete Results?** Leave the number of days as **7**.

7. Click  to validate the query.

The **Run Query Against Connection** window appears.



8. You can edit the query in the text area provided.

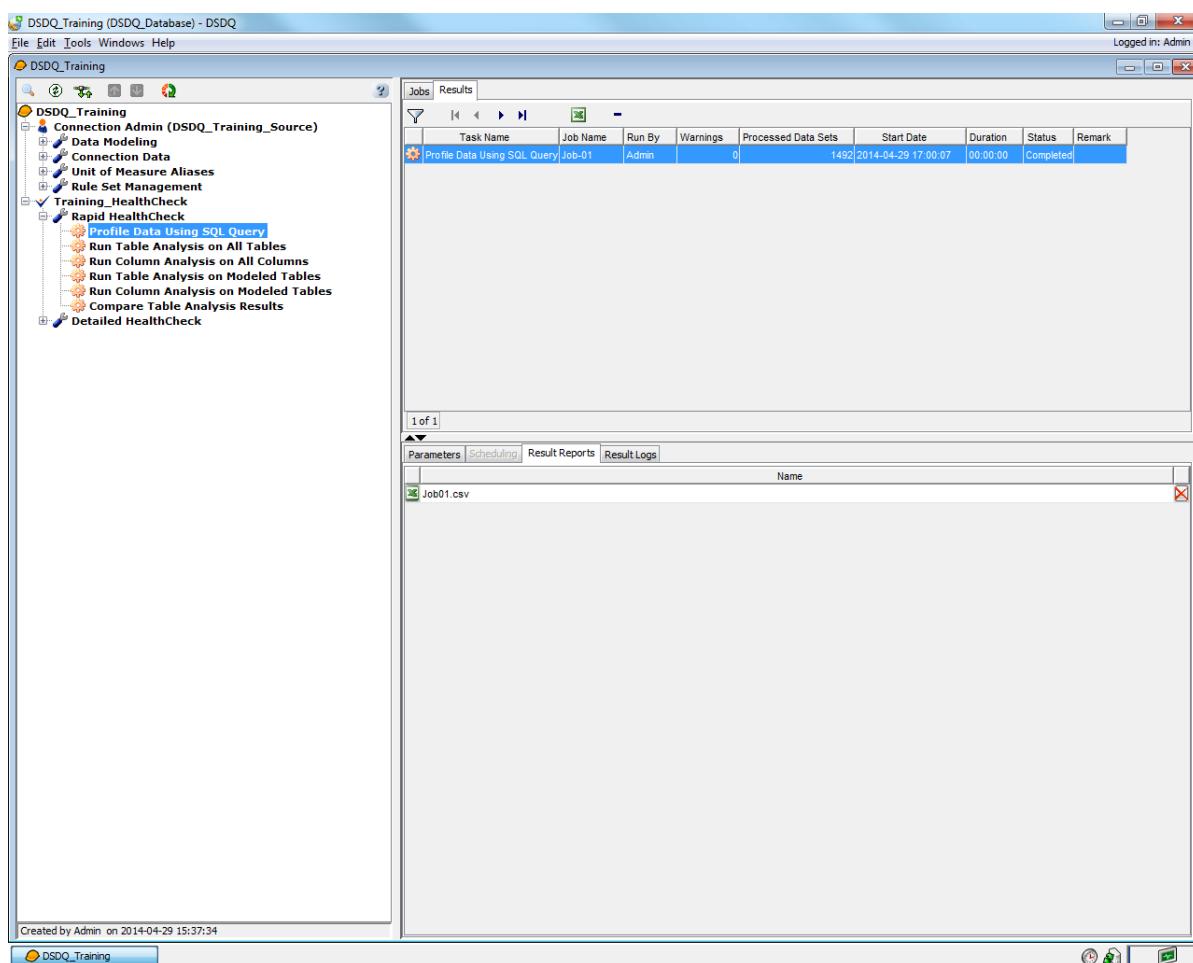
9. Click  to save changes in the **Parameters** tab.

10. Click .

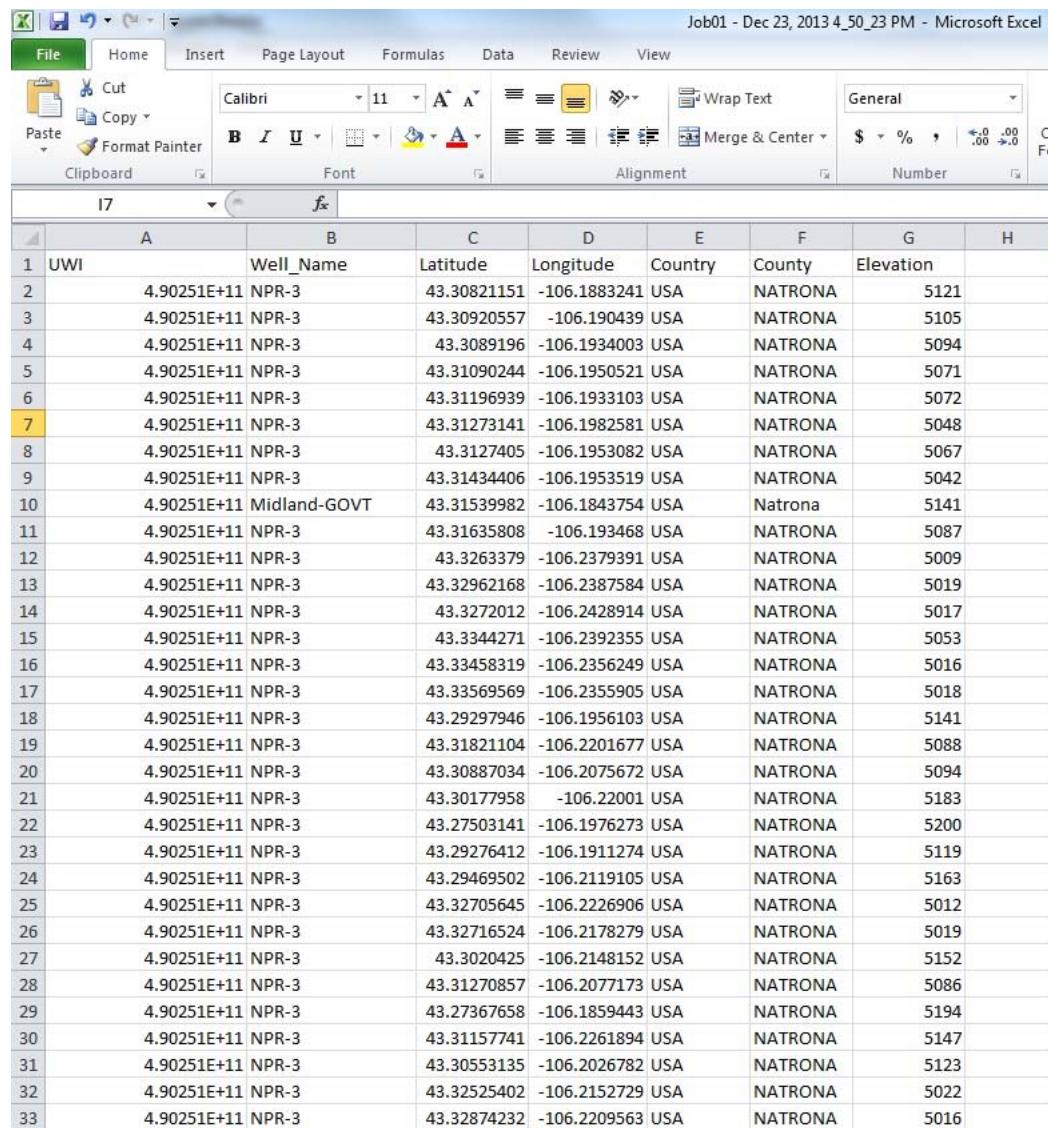
The **Profile Data Using SQL Query** task runs and displays results in the **Result Reports** tab of the **Job and Results Information Pane**.

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

Information Pane.



12. Click  on the **Result Reports** tab on the **Job and Results Information Pane** to display the **Profile Data using SQL Query** results in XLSX format.



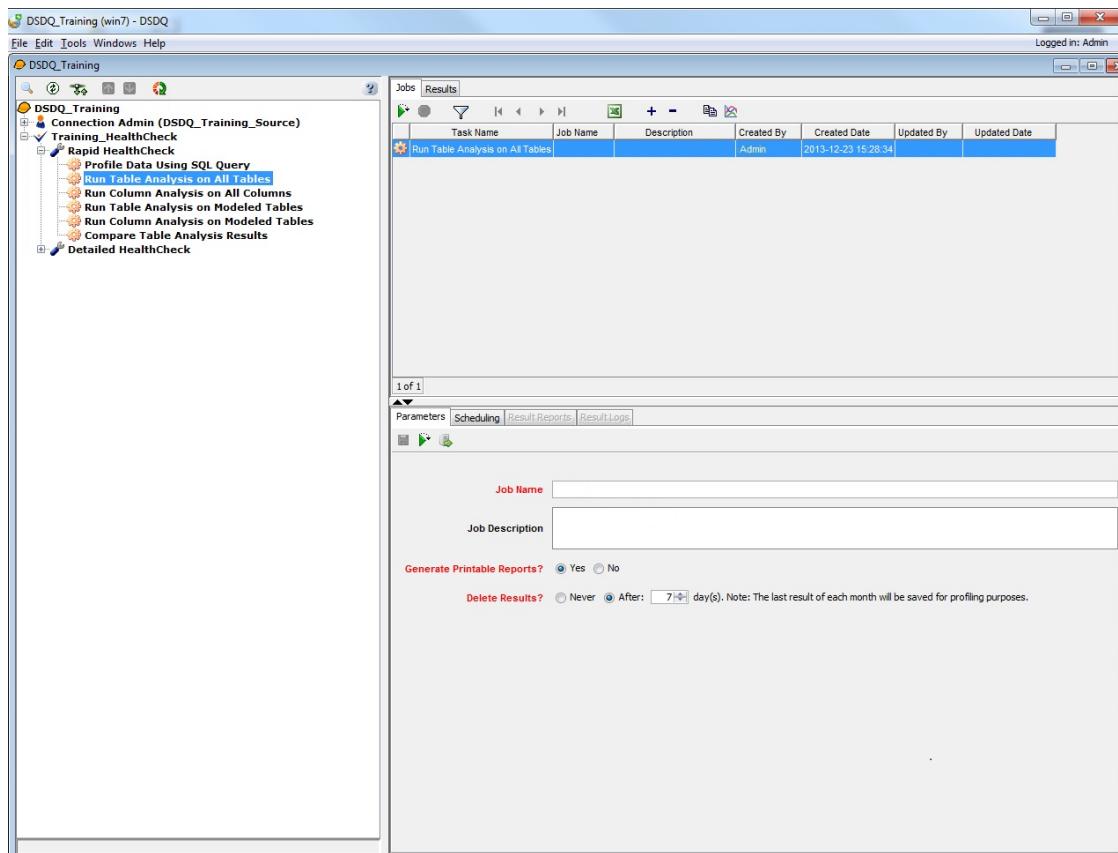
	A	B	C	D	E	F	G	H
1	UWI	Well_Name	Latitude	Longitude	Country	County	Elevation	
2	4.90251E+11	NPR-3	43.30821151	-106.1883241	USA	NATRONA	5121	
3	4.90251E+11	NPR-3	43.30920557	-106.190439	USA	NATRONA	5105	
4	4.90251E+11	NPR-3	43.3089196	-106.1934003	USA	NATRONA	5094	
5	4.90251E+11	NPR-3	43.31090244	-106.1950521	USA	NATRONA	5071	
6	4.90251E+11	NPR-3	43.31196939	-106.1933103	USA	NATRONA	5072	
7	4.90251E+11	NPR-3	43.31273141	-106.1982581	USA	NATRONA	5048	
8	4.90251E+11	NPR-3	43.3127405	-106.1953082	USA	NATRONA	5067	
9	4.90251E+11	NPR-3	43.31434406	-106.1953519	USA	NATRONA	5042	
10	4.90251E+11	Midland-GOVT	43.31539982	-106.1843754	USA	Natrona	5141	
11	4.90251E+11	NPR-3	43.31635808	-106.193468	USA	NATRONA	5087	
12	4.90251E+11	NPR-3	43.3263379	-106.2379391	USA	NATRONA	5009	
13	4.90251E+11	NPR-3	43.32962168	-106.2387584	USA	NATRONA	5019	
14	4.90251E+11	NPR-3	43.3272012	-106.2428914	USA	NATRONA	5017	
15	4.90251E+11	NPR-3	43.3344271	-106.2392355	USA	NATRONA	5053	
16	4.90251E+11	NPR-3	43.33458319	-106.2356249	USA	NATRONA	5016	
17	4.90251E+11	NPR-3	43.33569569	-106.2355905	USA	NATRONA	5018	
18	4.90251E+11	NPR-3	43.29297946	-106.1956103	USA	NATRONA	5141	
19	4.90251E+11	NPR-3	43.31821104	-106.2201677	USA	NATRONA	5088	
20	4.90251E+11	NPR-3	43.30887034	-106.2075672	USA	NATRONA	5094	
21	4.90251E+11	NPR-3	43.30177958	-106.22001	USA	NATRONA	5183	
22	4.90251E+11	NPR-3	43.27503141	-106.1976273	USA	NATRONA	5200	
23	4.90251E+11	NPR-3	43.29276412	-106.1911274	USA	NATRONA	5119	
24	4.90251E+11	NPR-3	43.29469502	-106.2119105	USA	NATRONA	5163	
25	4.90251E+11	NPR-3	43.32705645	-106.2226906	USA	NATRONA	5012	
26	4.90251E+11	NPR-3	43.32716524	-106.2178279	USA	NATRONA	5019	
27	4.90251E+11	NPR-3	43.3020425	-106.2148152	USA	NATRONA	5152	
28	4.90251E+11	NPR-3	43.31270857	-106.2077173	USA	NATRONA	5086	
29	4.90251E+11	NPR-3	43.27367658	-106.1859443	USA	NATRONA	5194	
30	4.90251E+11	NPR-3	43.31157741	-106.2261894	USA	NATRONA	5147	
31	4.90251E+11	NPR-3	43.30553135	-106.2026782	USA	NATRONA	5123	
32	4.90251E+11	NPR-3	43.32525402	-106.2152729	USA	NATRONA	5022	
33	4.90251E+11	NPR-3	43.32874232	-106.2209563	USA	NATRONA	5016	

Exercise: Running Table Analysis on All Tables

This task is used to analyze all the tables for issues and inconsistencies. In this particular exercise, we are analyzing 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. 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**.

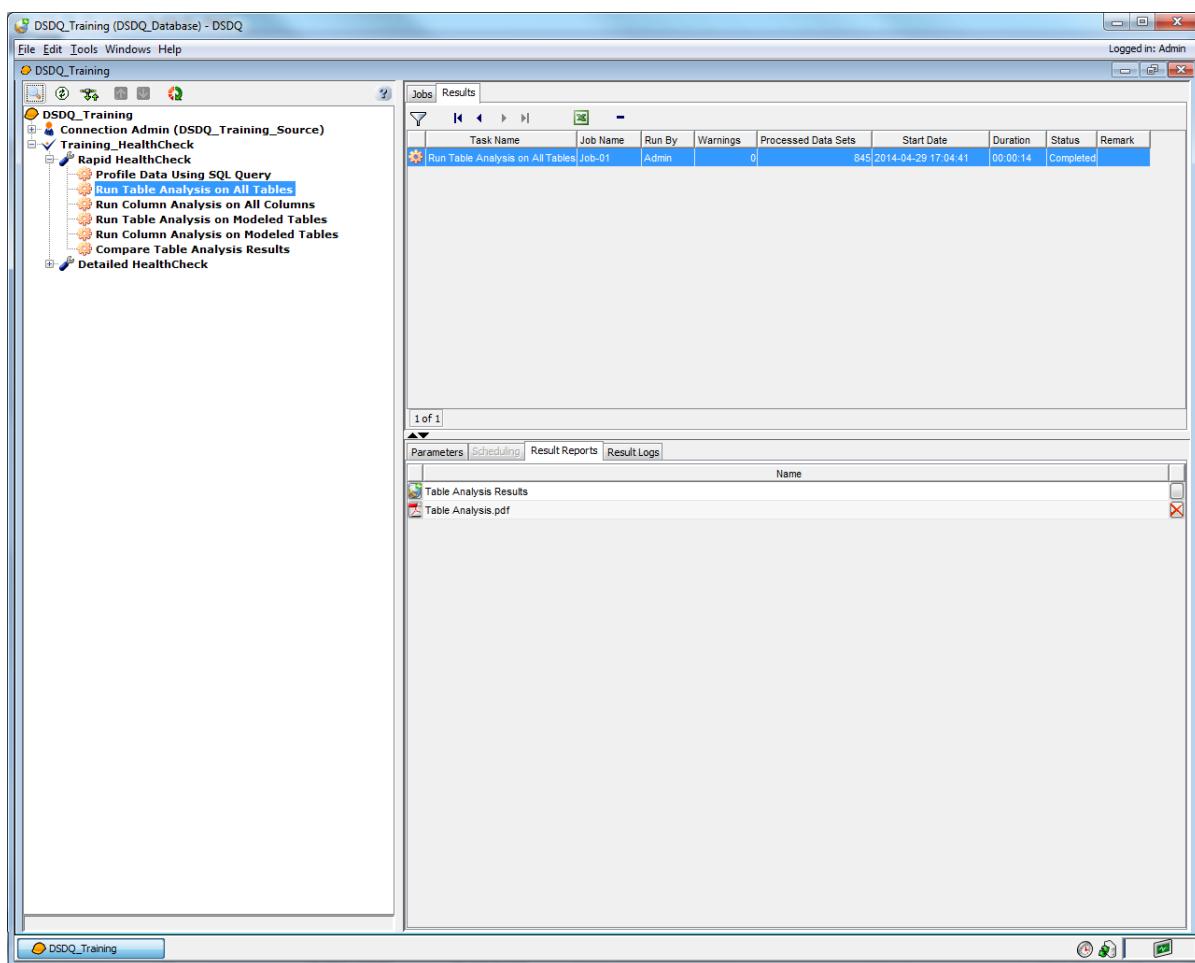


2. Enter **Job-01** in the **Job Name** field.
3. Enter **Table Analysis on All Tables** in the **Job Description** field.
4. Select the **Yes** option for **Generate Printable Reports?**

5. Select the **After** option for **Delete Results?** Leave the number of days as **7**.
6. Click  to save changes in the **Parameters** tab.
7. Click .

The **Run Table Analysis on All Tables** task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

8. 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**.



9. Click  on the **Result Reports** tab to display **Table Analysis on All Tables** results in PDF format.

Table Analysis

HALLIBURTON
Landmark Software
& Services

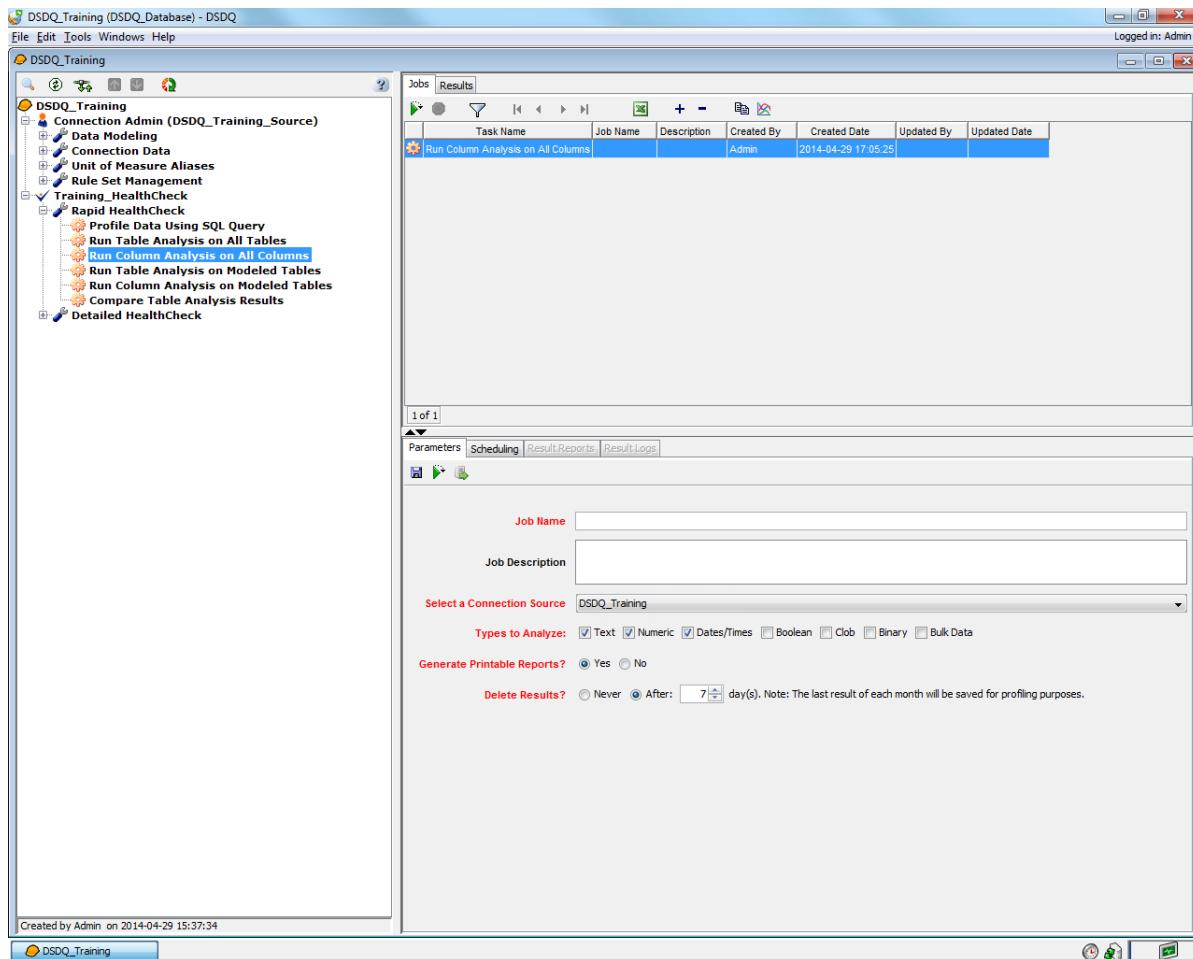
Table Name:	Row Count
WellSurvey	0
WellTemplate	30
WellTest	0
WellTestAnal	0
WellTestCont	0
WellTestCush	0
WellTestEquip	0
WellTestFlow	0
WellTestMud	0
WellTestPer	0
WellTestPress	0
WellTestRec	0
WellTestRecov	0
WellTestRemark	0
WellTestShut	0
WellTexasLoc	0
WellTreatment	0
WellTstCompAnal	0
WellTstFlwMeas	0
WellTstPresMeas	0
WellUwi	0
WellWorkOver	0
WellZConversion	117
XSecAnno	0

Exercise: Running Column Analysis on All Columns

To run column analysis on all the 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.

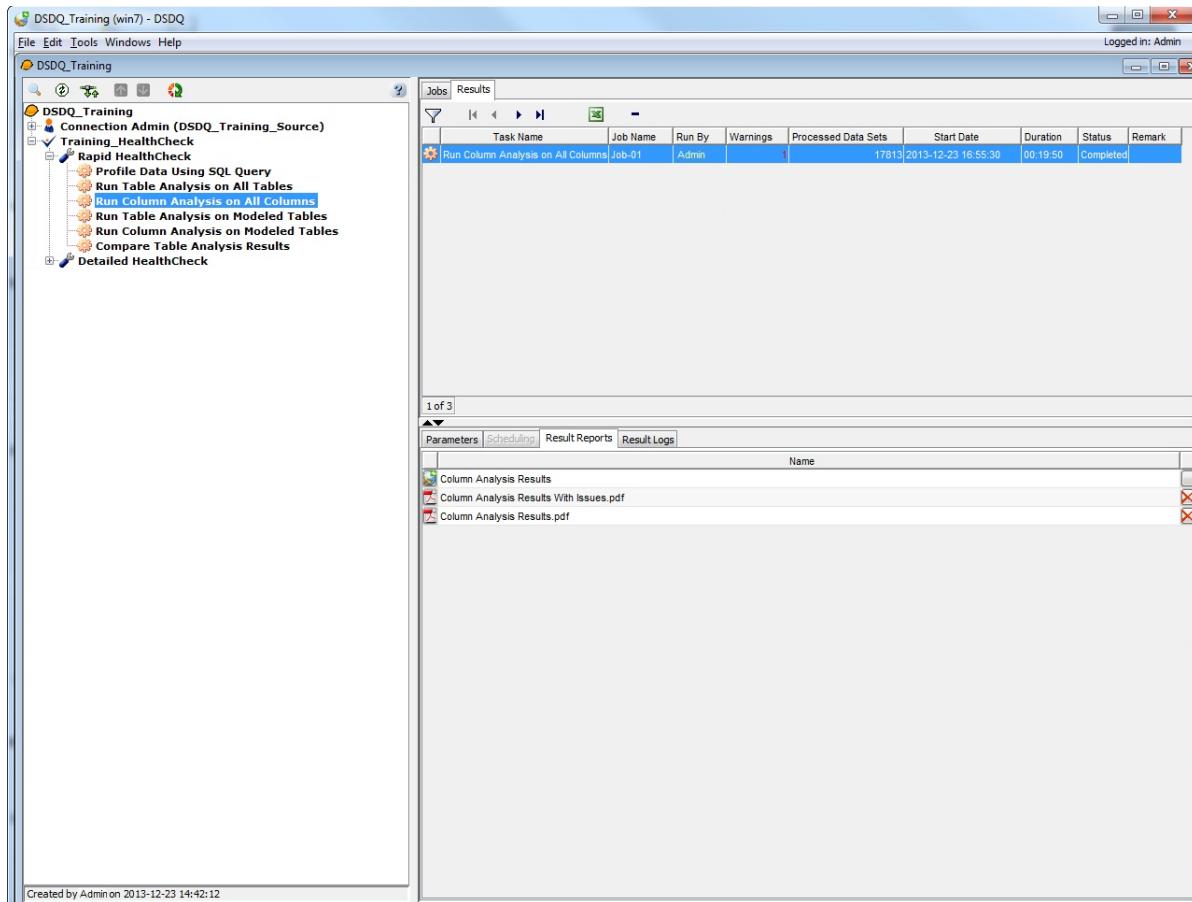


2. Enter **Job-01** in the **Job Name** field.
3. Enter **Column Analysis on All Columns** in the **Job Description** field.
4. Select **DSDQ_Training** from the **Select a Connection Source** drop-down list.
5. Select all options for **Types to Analyze**.
6. Select the **Yes** option for **Generate Printable Reports?**
7. Select the **After** option for **Delete Results?** Leave the number of days as **7**.
8. Click to save changes in the **Parameters** tab.

9. Click .

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



11. Click  on the **Result Reports** tab to display **Column Analysis Results with Issues** in PDF format.

Column Analysis With Issues										HALLIBURTON	
Project:	DSDQ_Training									Landmark Software & Services	
Phase:	Training_HealthCheck										
Task:	Run Column Analysis on All Columns										
Job:	Job-01										
Connection:	DSDQ_Training_Source										
Result Date:	Mon, Dec 23, 2013 16:55										
Column Name	Rows Analyzed by	Number Not Null	% Populated	Number Unique	Number of Mixed Case	Non Printable Characters	Preceding White Space	Trailing White Space	Double White Space	Minimum Value	Maximum Value
Table Name: AdbArchive											
archive_description	1	1	100	1	1	0	0	0	0	initial snapshot copy for head office	initial snapshot copy for head office
archive_stage_dir	1	1	100	1	1	0	0	0	0	\$ARCHIVE_STAGING_DIR/Test_project	\$ARCHIVE_STAGING_DIR/Test_project
archive_stub_dir	1	1	100	1	1	0	0	0	0	\$ARCHIVE_STUB_DIR/Test_project	\$ARCHIVE_STUB_DIR/Test_project
Table Name: AdbObject											
object_description	3	3	100	3	3	0	0	0	0	Grids used in simulation model	Pressure data from offset wells
Table Name: AdbProject											
project_description	1	1	100	1	1	0	0	0	0	This is a test project (installed by default)	This is a test project (installed by default)
project_name	1	1	100	1	1	0	0	0	0	Bonanza Field	Bonanza Field
project_source	1	1	100	1	1	0	0	0	0	Government data package	Government data package
Table Name: AdbProjectUser											
user_role	1	1	100	1	1	0	0	0	0	Geophysicist	Geophysicist
Table Name: AdbRAppDatatype											
application	32	32	100	12	30	0	0	0	0	EDMS	Z-MAPPplus R2003
datatype	32	32	100	21	32	0	0	0	0	2D Nav	Wells
Table Name: AdbRApplication											
application	13	13	100	13	11	0	0	0	0	EDMS	Z-MAPPplus R2003

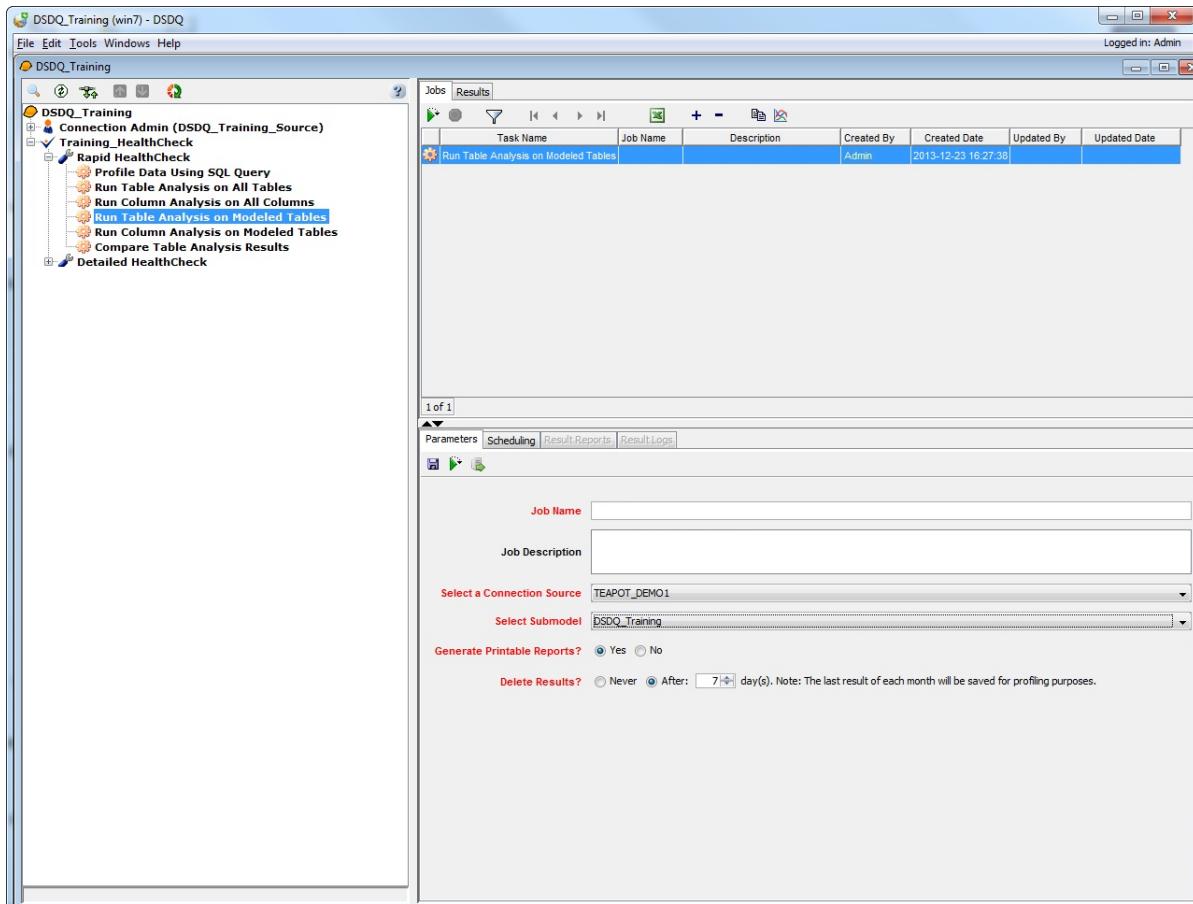
Exercise: Running Table Analysis on Modeled Tables

The **Run Table Analysis on Modeled Tables** Task runs only on tables that have been modeled in the **Perform Table Modeling** Tool.

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 displays on the **Job and Results**

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



2. Enter **Job-01** in the **Job Name** field.
3. Enter **Table Analysis on Modeled Tables** in the **Job Description** field.
4. Select **TEAPOT_DEMO1** from the **Select a Connection Source** drop-down list.

Note

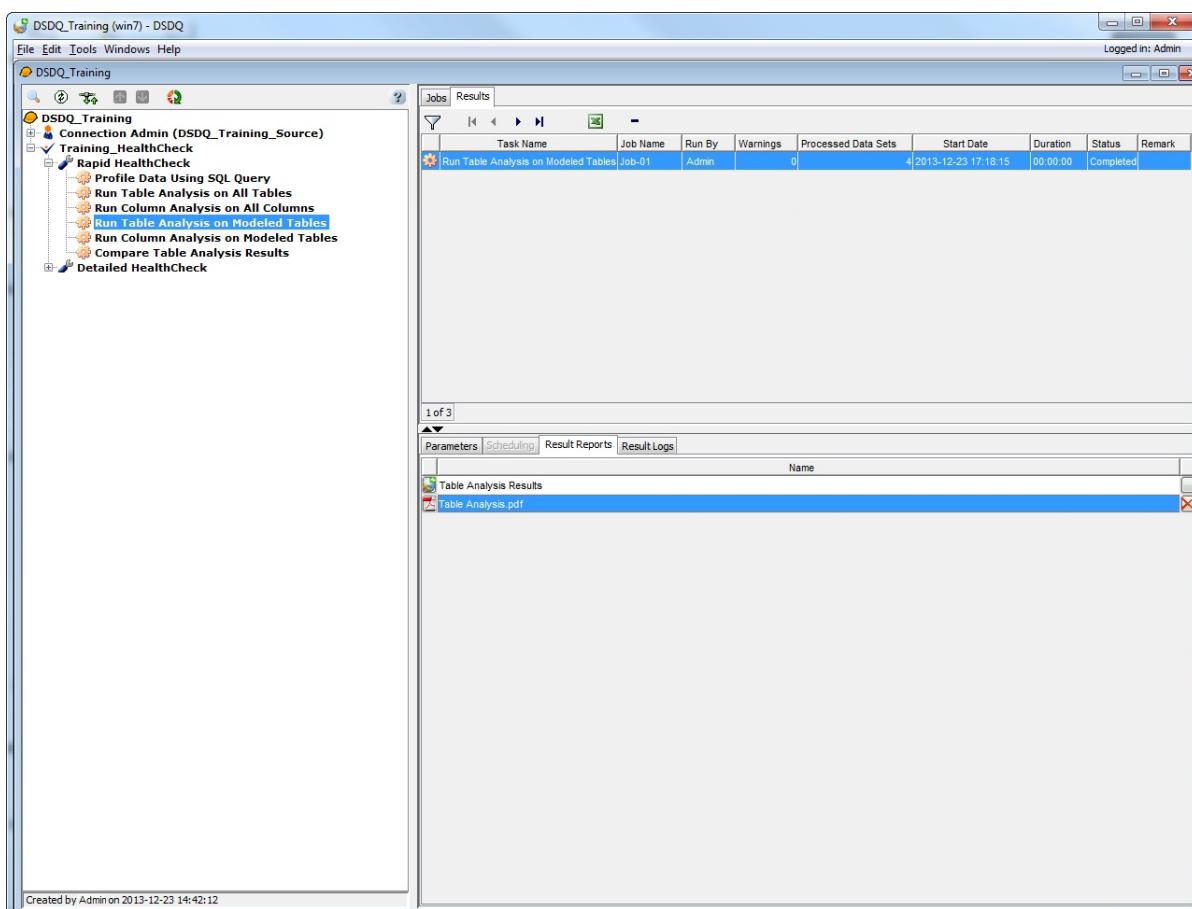
For more information on data owner connections, refer to Adding a New Data Owner Connection section in Chapter 2, Connecting DecisionSpace Data Quality with DecisionSpace Data Server.

5. Select **DSDQ_Training** 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 **7**.
8. Click  to save changes in the **Parameters** tab.
9. Click .

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



11. Click  on the **Result Reports** tab to display **Table Analysis on Modeled Tables** results in PDF format.

Table Analysis

Project:	DSDQ_Training	HALLIBURTON
Task:	Run Table Analysis on Modeled Tables	Landmark Software S & Services
Job:	Job-01	
Connection:	DSDQ_Training_Source	
Source:	TEAPOT_DEMO1	
Sub-Model:	DSDQ_Training	
Result Date:	Mon, Dec 23, 2013 17:18	

Table Name:	Row Count
Well	1395
WellEntire	1395
RCountry	158
RStateName	133

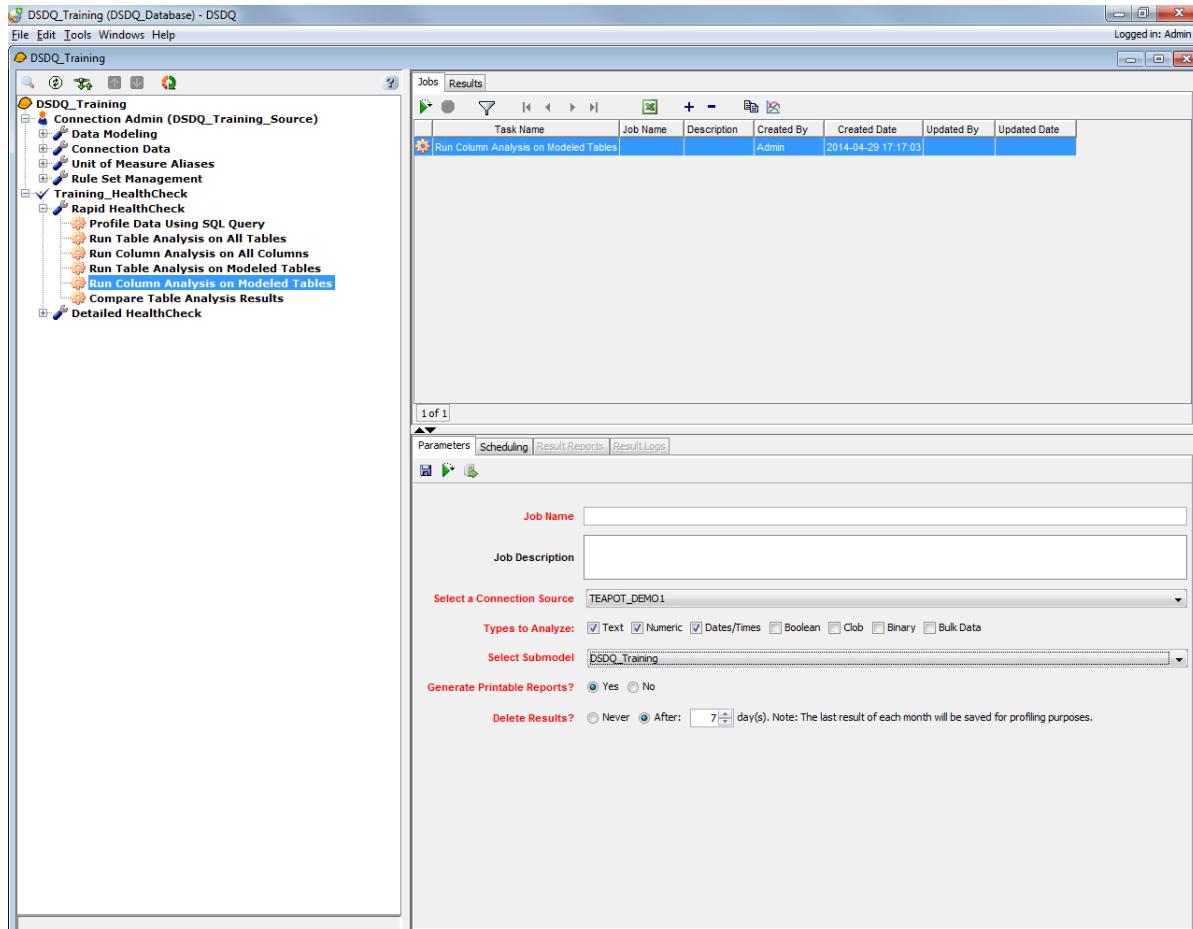
Exercise: Running Column Analysis on Modeled Tables

The **Run Column Analysis on Modeled Tables** Task runs only on tables that have been modeled in the **Perform Table Modeling** Tool.

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 displays on the **Job and Results**

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



2. Enter **Job-01** in the **Job Name** field.
3. Enter **Column Analysis on Modeled Tables** in the **Job Description** field.
4. Select **TEAPOT_DEMO1** from the **Select a Connection Source** drop-down list.

Note

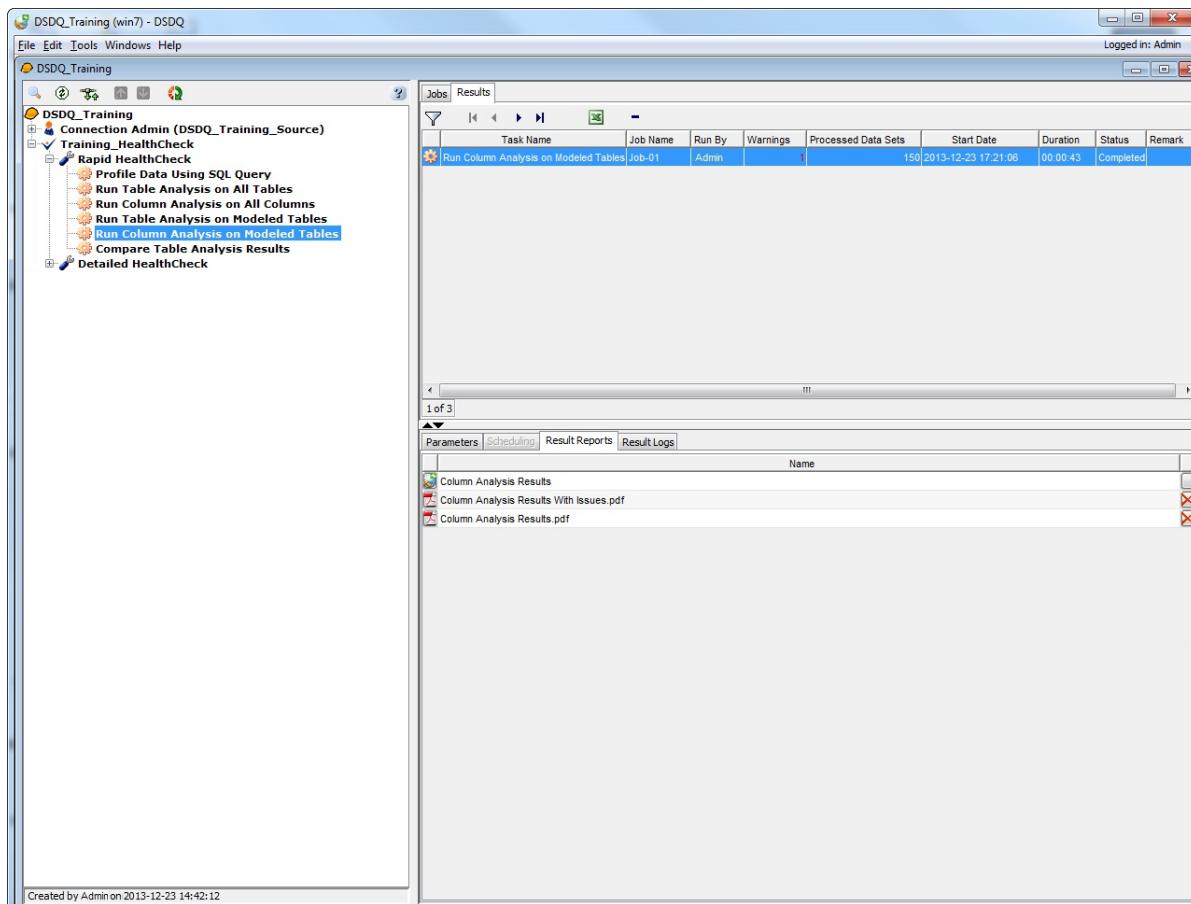
For more information on data owner connections, refer to Adding a New Data Owner Connection section in Chapter 2, Connecting DecisionSpace Data Quality with DecisionSpace Data Server.

5. Select all the options for **Types to Analyze**.
6. Select **DSDQ_Training** from the **Select Submodel** drop-down list.

7. Select the **Yes** option for **Generate Printable Reports?**
8. Select the **After** option for **Delete Results?** Leave the number of days as **7**.
9. Click  to save changes in the **Parameters** tab.
10. Click .

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

11. 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**.



12. Click  on the **Result Reports** tab to display **Column Analysis Results with Issues** in PDF format.

Column Analysis With Issues

HALLIBURTON
Landmark Software
& Services

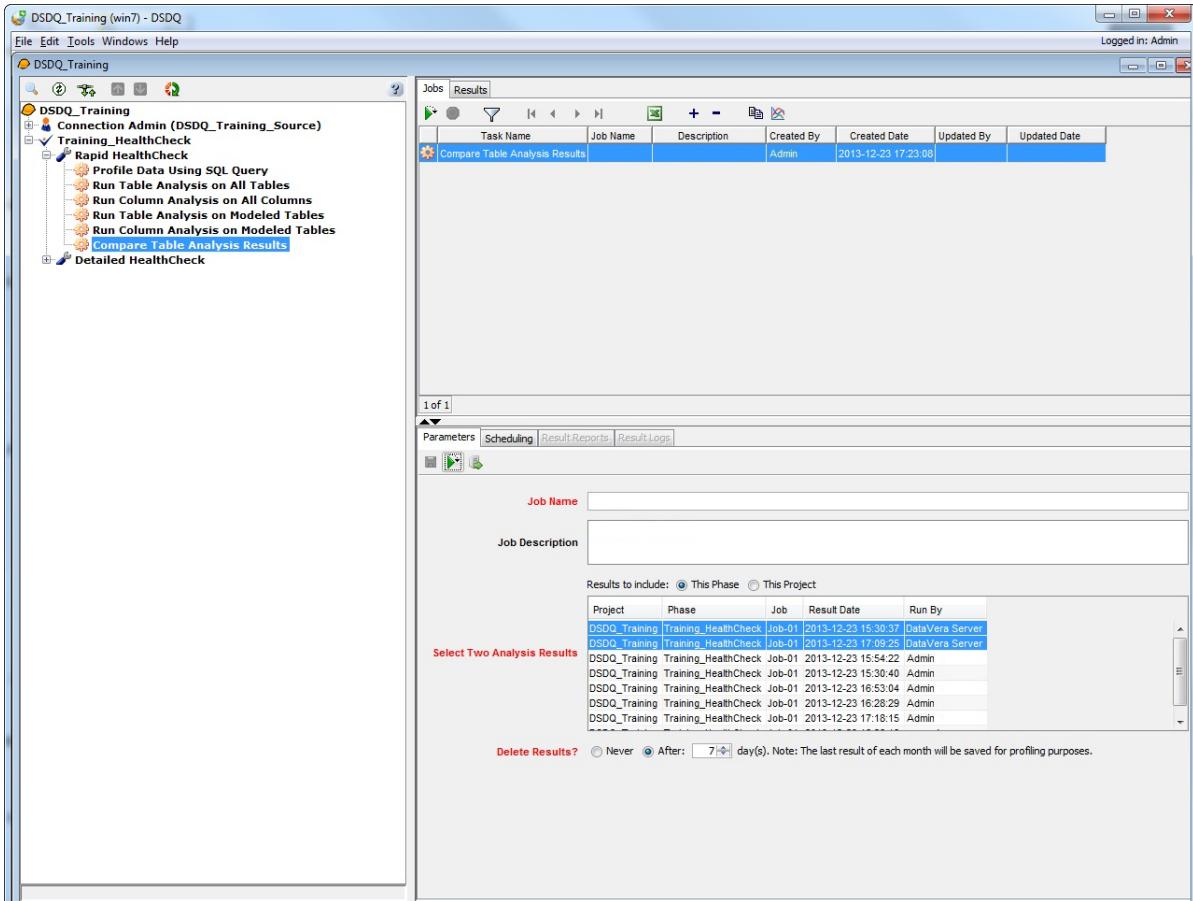
Column Name	Rows Analyzed by	Number Not Null	% Populated	Number Unique	Number of Mixed Case	Non Printable Characters	Preceding White Space	Trailing White Space	Double White Space	Minimum Value	Maximum Value
Table Name: RCountry											
country_name	158	158	100	158	2	0	0	0	0	AFGHANISTAN	ZAMBIA
remark	158	156	99	134	0	0	0	8	0	SHEIKDOM OCCUPYING A PENINSULA ON THE E. COAST OF ARABIA	
Table Name: RStateName											
state_name	133	133	100	133	3	0	0	0	0	ALABAMA	ZEELAND
Table Name: Well											
common_well_name	1395	1363	98	1305	20	0	0	0	0	1-10	Test Well 1
county	1395	1395	100	3	61	0	0	0	0	NATRONA	UNKNOWN
field	1395	1395	100	7	61	0	0	0	0	East Teapot	Wildcat
state	1395	1395	100	3	61	0	0	0	0	UNKNOWN	Wyoming
uwi	1395	1395	100	1395	4	0	0	0	0	490250625600	Tst490251031300
well_name	1395	1385	99	20	25	0	0	0	0	Beartooth Federa	ukn
well_number	1395	1384	99	1353	218	0	0	0	0	04211 31 A	State No.
well_operator	1395	1395	100	24	114	0	0	0	0	ADVENTURE	ukn
Table Name: WellEntire											
basin	1395	1395	100	3	61	0	0	0	0	POWDER RIVER	UNKNOWN

Exercise: Comparing Table Analysis Results

To compare the table analysis results:

1. Double-click the **Comparing Table Analysis Results** Task or right-click the **Comparing Table Analysis Results** 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.

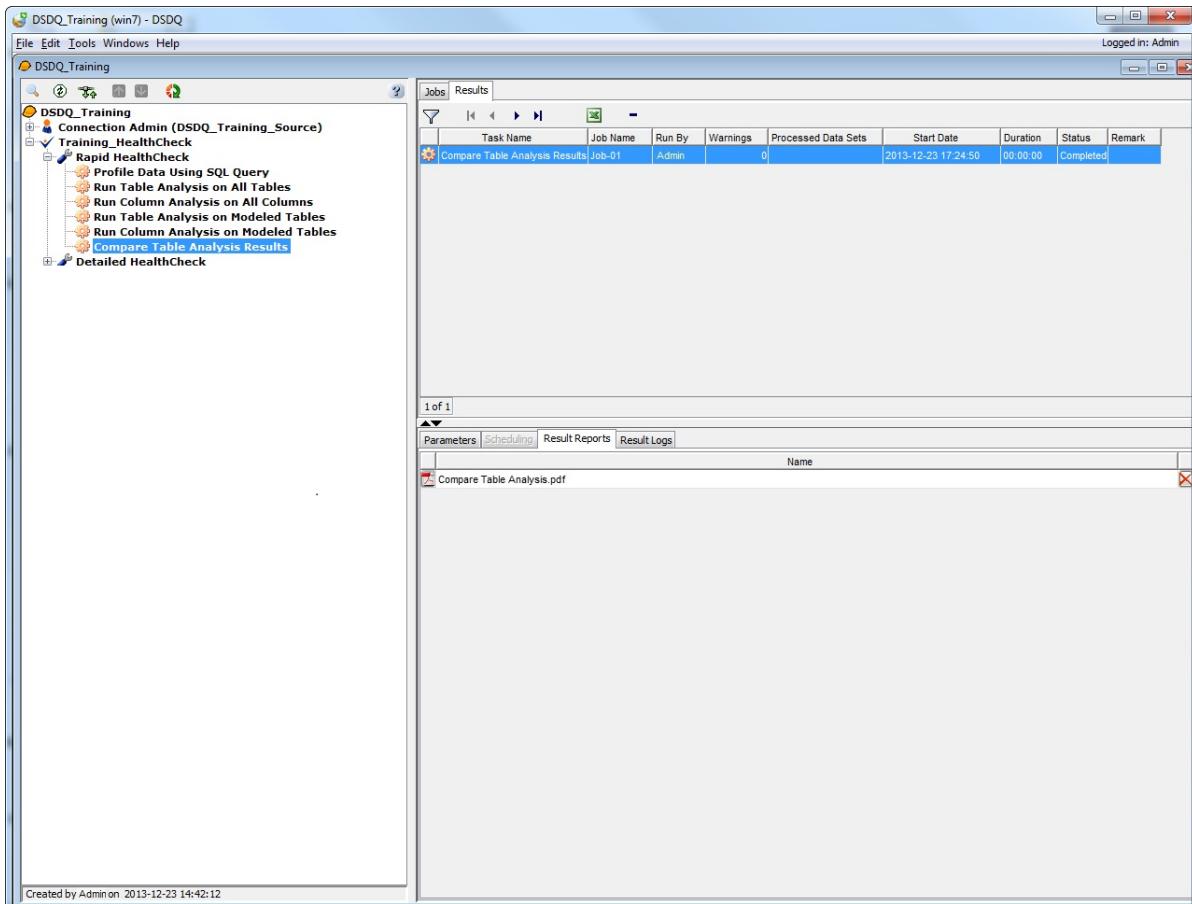


2. Enter **Job-01** in the **Job Name** field.
3. Enter **Compare Table Analysis** in the **Job Description** field.
4. Select the **This Phase** option for **Results to include**:
 - **This Phase:** The results can be compared within the same phase.
 - **This Project:** The results can be compared within multiple phases in the project.
5. Select two results that you want to compare from the **Select Two Analysis Results** list.
6. Select the **After** option for **Delete Results?** Leave the number of days as **7**.
7. Click to save changes in the **Parameters** tab.

8. Click .

The **Compare Table Analysis Results** task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

9. 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**.



10. Click  on the **Result Reports** tab to display **Compare Table Analysis** results in PDF format.

Compare Table Analysis

HALLIBURTON
Landmark Software
& Services

Result 1:		Result 2:	
Project:	DSDQ_Training	Project:	DSDQ_Training
Phase:	Training_HealthCheck	Phase:	Training_HealthCheck
Task:	Run Table Analysis on All Tables	Task:	Run Table Analysis on All Tables
Job:	Job-01	Job:	Job-01
Connection:	DSDQ_Training_Source	Connection:	DSDQ_Training_Source
Result Date:	Mon, Dec 23, 2013 17:09	Result Date:	Mon, Dec 23, 2013 15:30

Table Name	Result 1	Result 2	Difference
Activity	0	0	0
AdbArchive	1	1	0
AdbFileObject	0	0	0
AdbObject	3	3	0
AdbProject	1	1	0
AdbProjectBoundary	0	0	0
AdbProjectUser	1	1	0
AdbRAppDatatype	32	32	0
AdbRApplication	13	13	0
AdbRArchiveFmt	2	2	0
AdbRdatatype	27	27	0
AdbRMediaType	9	9	0
AdbRProjectClass	2	2	0
AdbRProjectStatus	3	3	0
AdbRProjectType	5	5	0
AdbSet	4	4	0
AdbVArchive	1	1	0
AdbVObject	3	3	0
AdbVProject	1	1	0
AdbVProjectBoundary	0	0	0
AdbVProjectUser	1	1	0
AdbVSet	4	4	0
AnalysisLogUse	0	0	0

Identifying Data Issues

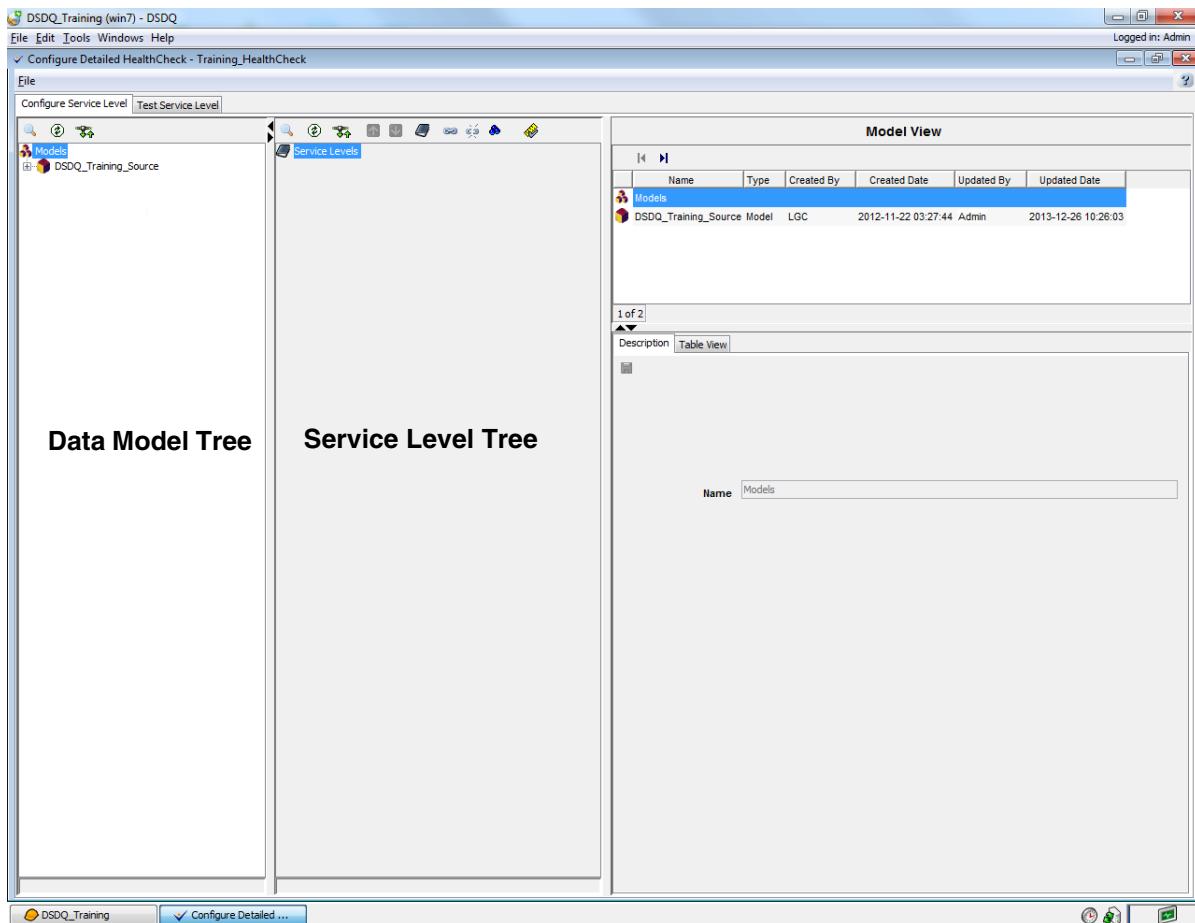
The **Detailed HealthCheck** Activity allows you to run business rules against the dataset to identify data problems. It produces a series of comprehensive reports that assist in rating the quality of the data. The tables must be modeled and the elements assigned prior to running the **Detailed HealthCheck** Activity. It also allows you to assign columns from selected submodels to HealthCheck requirements and test the service level to view results in the **Configure Detailed HealthCheck** Tool. The user can select which requirements are to be enabled/disabled in the service level. The user can also select a subset of the total data to be used when testing a service level.

Exercise: Configuring the 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**

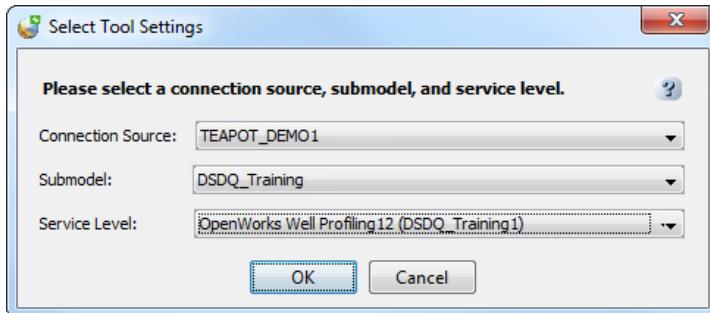
Tool from the pop-up menu.
The **Configure Detailed HealthCheck** window appears.



Note

The **Configure Detailed HealthCheck** Tool has two tabs: **Configure Service Level** and **Test Service Level** located at the top left. With the **Configure Service Level** tab selected, the Data Model Tree (left tree) displays tables used in the currently selected submodel. The Service Level Tree (right tree) shows the currently selected Service Level. The Table View Pane populates with table names or requirements, depending on what tree is selected. The **Description** tab displays the model name.

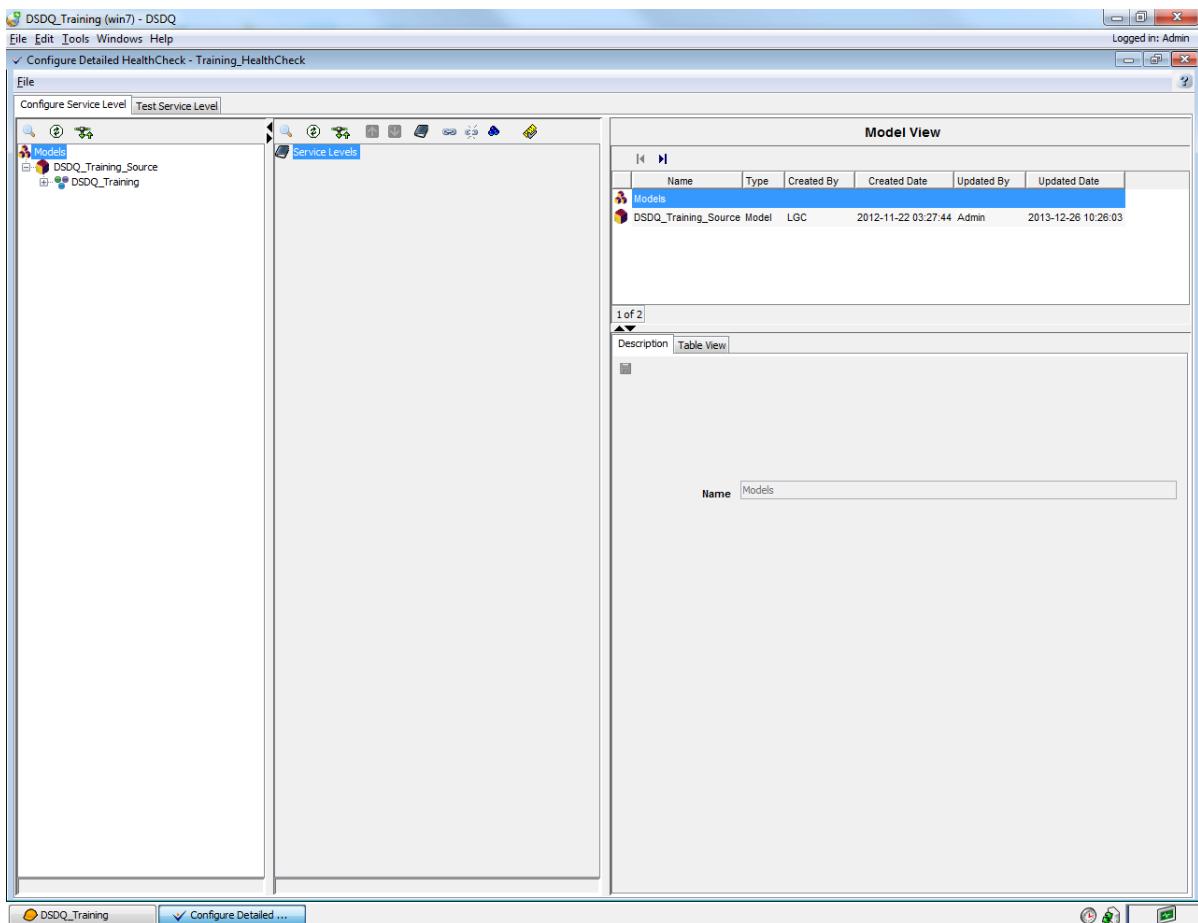
3. Select **File > Settings** from the File menu bar.
The **Select Tool Settings** dialog box appears.



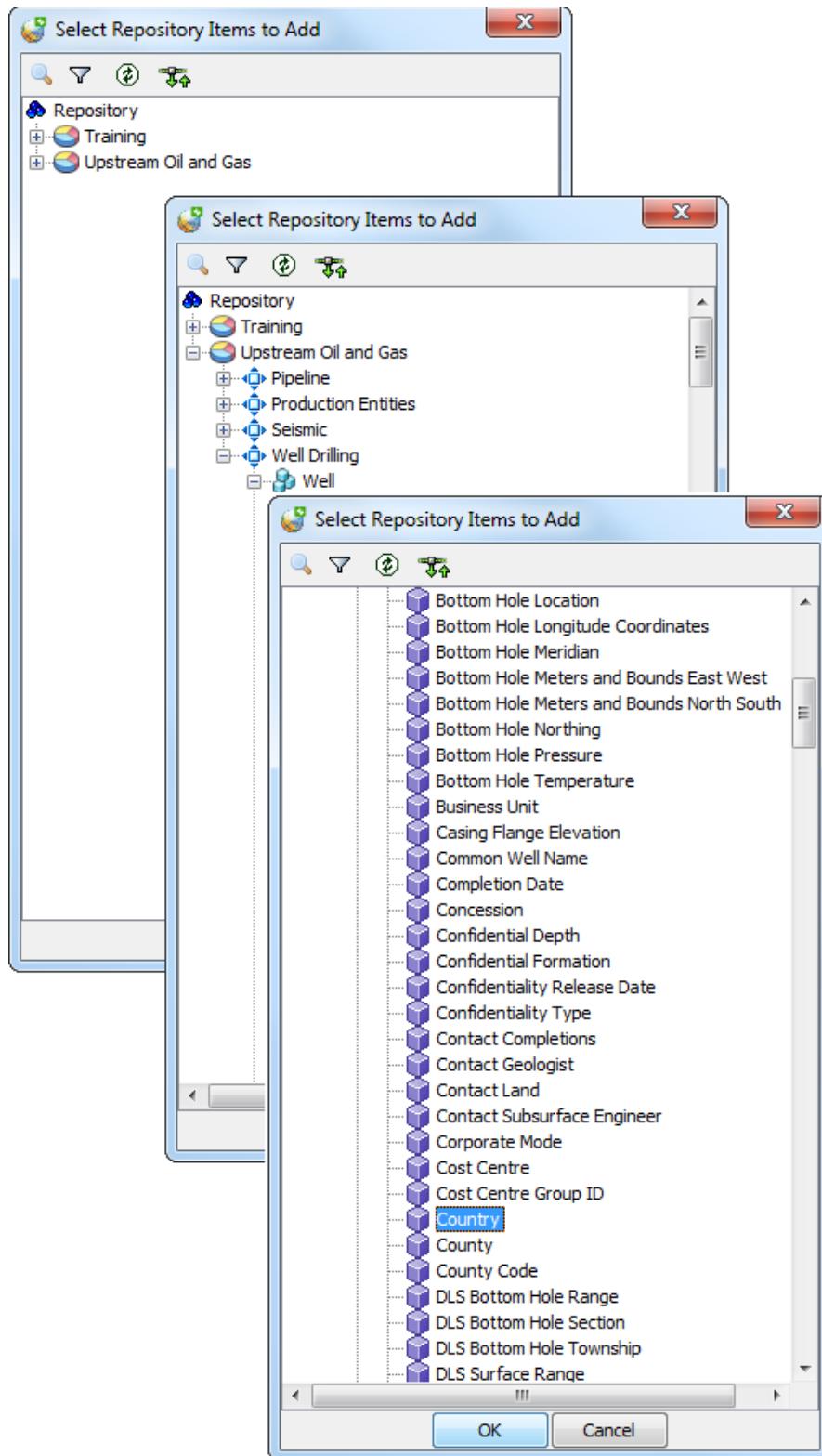
4. Select **TEAPOT_DEMO1** from the **Connection Source** drop-down list.
5. Select **DSDQ_Training** from the **Submodel** drop-down list.
6. Select **OpenWorks Well Profiling 12 (DSDQ_Training1)** from the **Service Level** drop-down list.

7. Click **OK**.

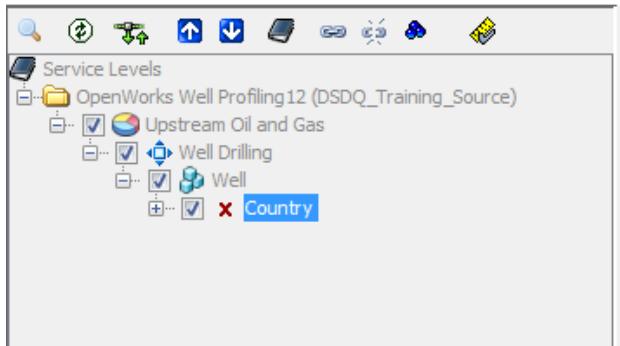
The **Configure Detailed HealthCheck** window appears with **DSDQ_Training** displaying in the Data Model Tree.



8. Click  on the Service Level Tree toolbar.
The Select Repository Items to Add dialog box opens.

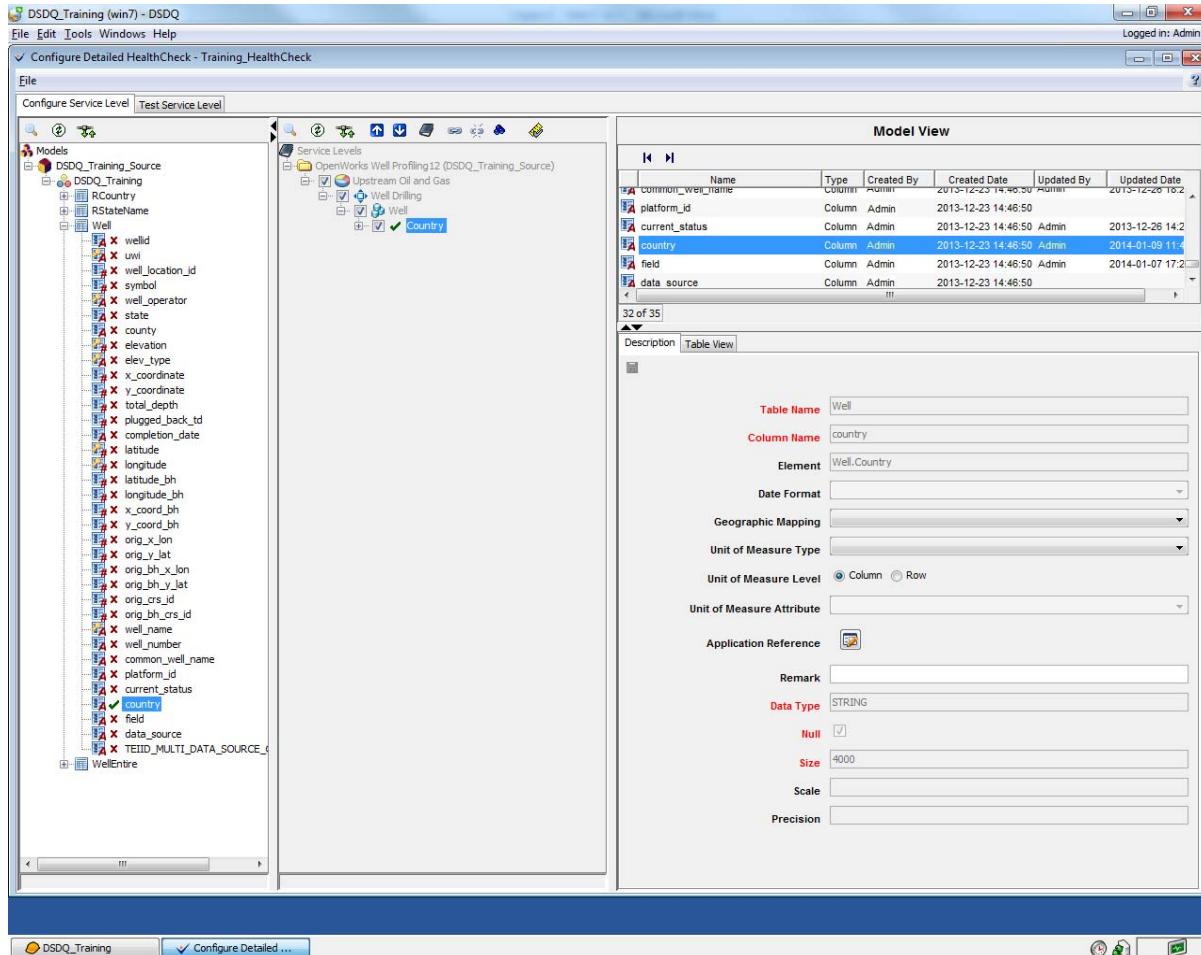


9. Click  to expand the **Upstream Oil & Gas** sector.
 10. Expand the **Well Drilling** area.
 11. Expand the **Well** element group.
 12. Select the **Country** element and click **OK**.
- The **Country** element is added to the Service Level Tree. A red cross will appear adjacent to the element in the Service Level Tree indicating that the element is not linked to any Column.



13. Click  on the Data Model Tree to expand **DSDQ_Training** table.
 14. Expand the **Well** table and select the **Country** column.
 15. Drag and drop the **Country** column onto the **Country** element in the Service Level Tree.
- A green check mark will appear 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 many columns to the same element if the columns come from different tables. Alternatively you can select the column & element

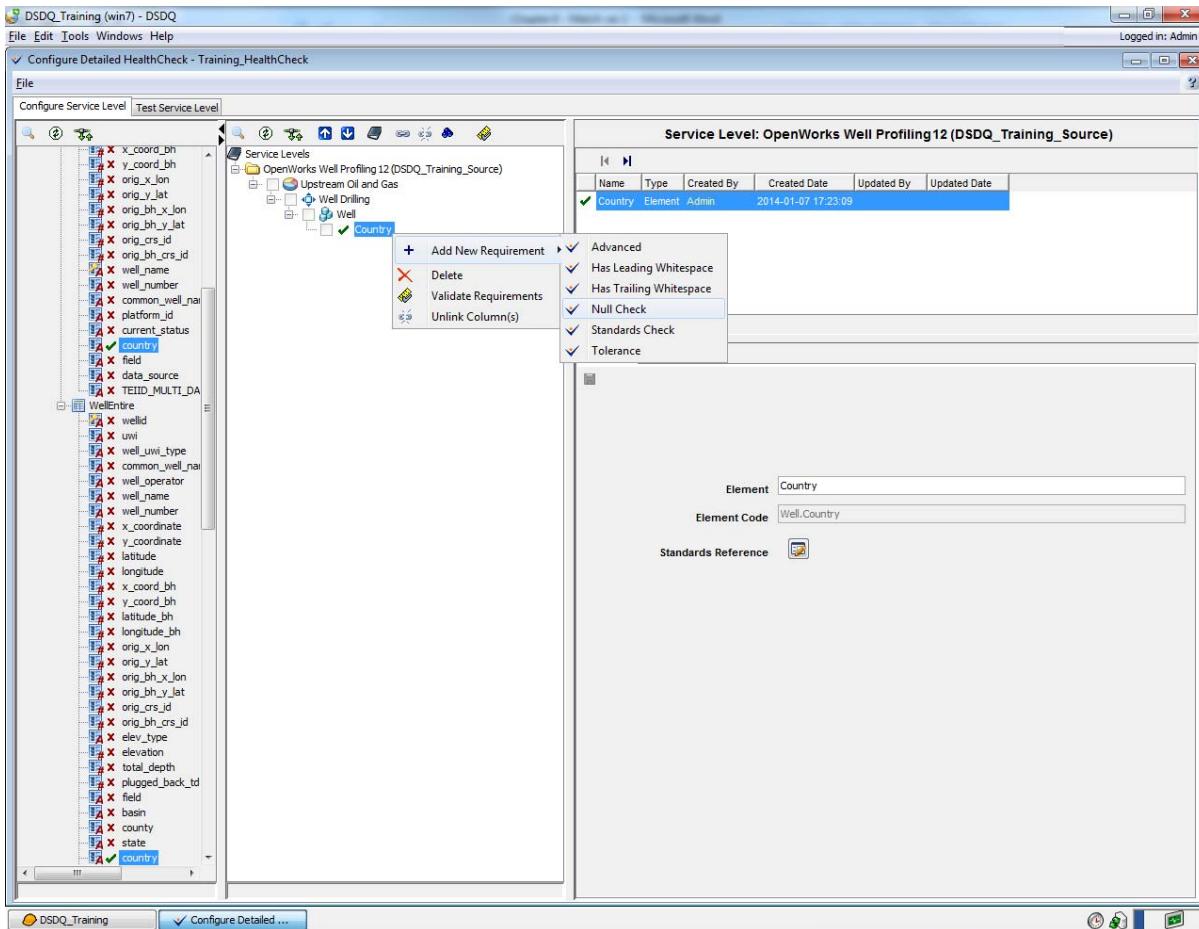
to link and click the **Link Column to Element** button on the Service Level Tree toolbar.



Note

You can unlink elements or columns. Select the element or column to unlink and click the **Unlink Columns from an Element** button on the Service Level Tree toolbar.

16. Right-click on the **Country** element in the Service Level Tree and select **Add New Requirement > Null Check** from the pop-up menu.



Note

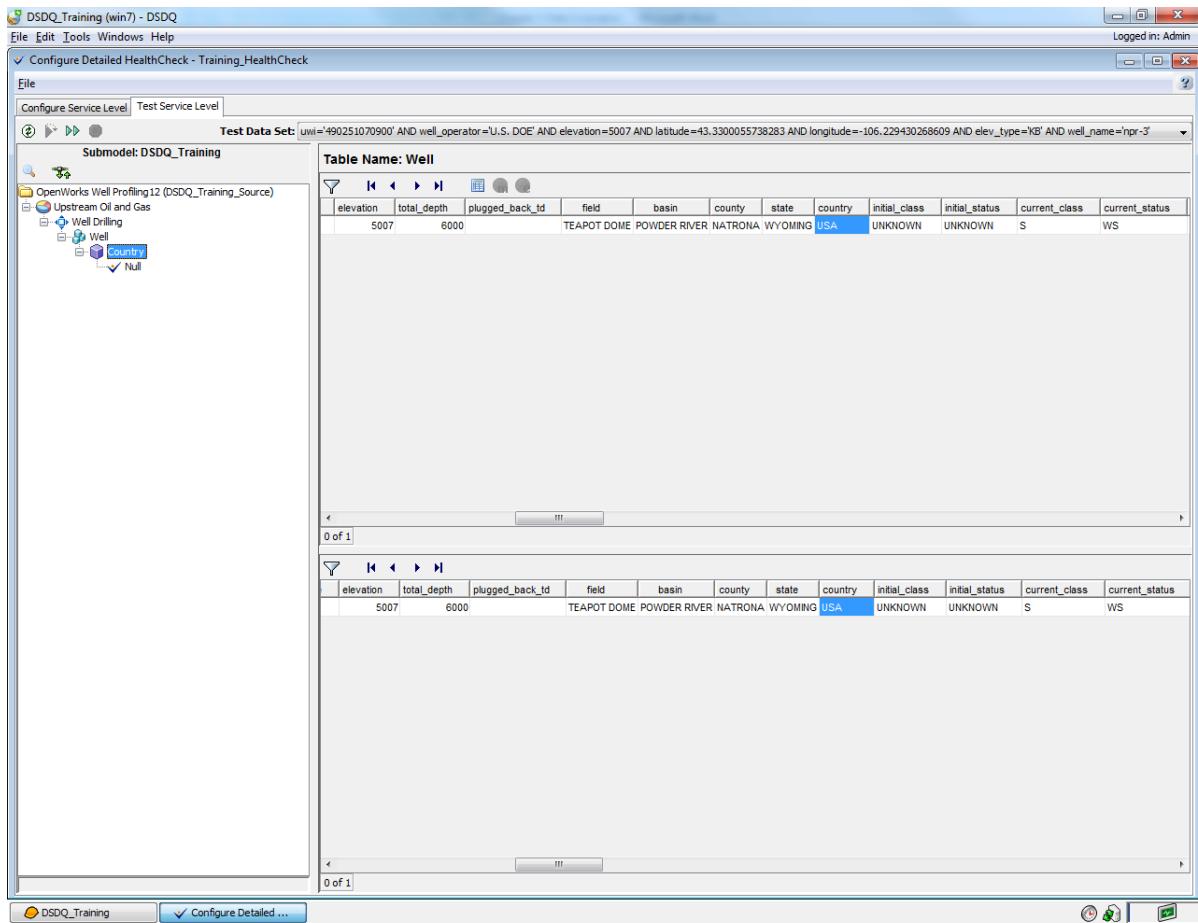
The Null Check Requirement determines the number of rows in the column are not populated.

The **Enter Name** dialog box appears.



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

18. Click **OK** to add the requirement to the selected element.
19. Optionally, repeat steps **8** to **18** to add all elements for HealthCheck.
20. Click the **Test Service Level** tab.
The test is automatically executed for the first record of the test data subset.



The corresponding column that the element has been assigned to is highlighted in both panes on the right. The original row data for the record is displayed in the top right pane. The bottom right pane shows the row of data after the service level rules and requirements have been run. Temporary and check columns defined in the rules and requirements are also shown (to see them you may have to scroll all the way to the right). By looking at the columns that have been changed and temporary columns, the user can verify that the behavior of the service level is correct prior to running the **Run Detailed HealthCheck Task**. The service level is tested one record

at a time.

Note

To view only the affected columns for the selected element in the table view, click the **View Affected Columns** button in the table view toolbar.

21. Click the **Next Data Set** button to test the next record.
22. Repeat step 21 to test all records.
23. Select **File > Exit** from the menu bar on the **Configure Detailed HealthCheck** window.

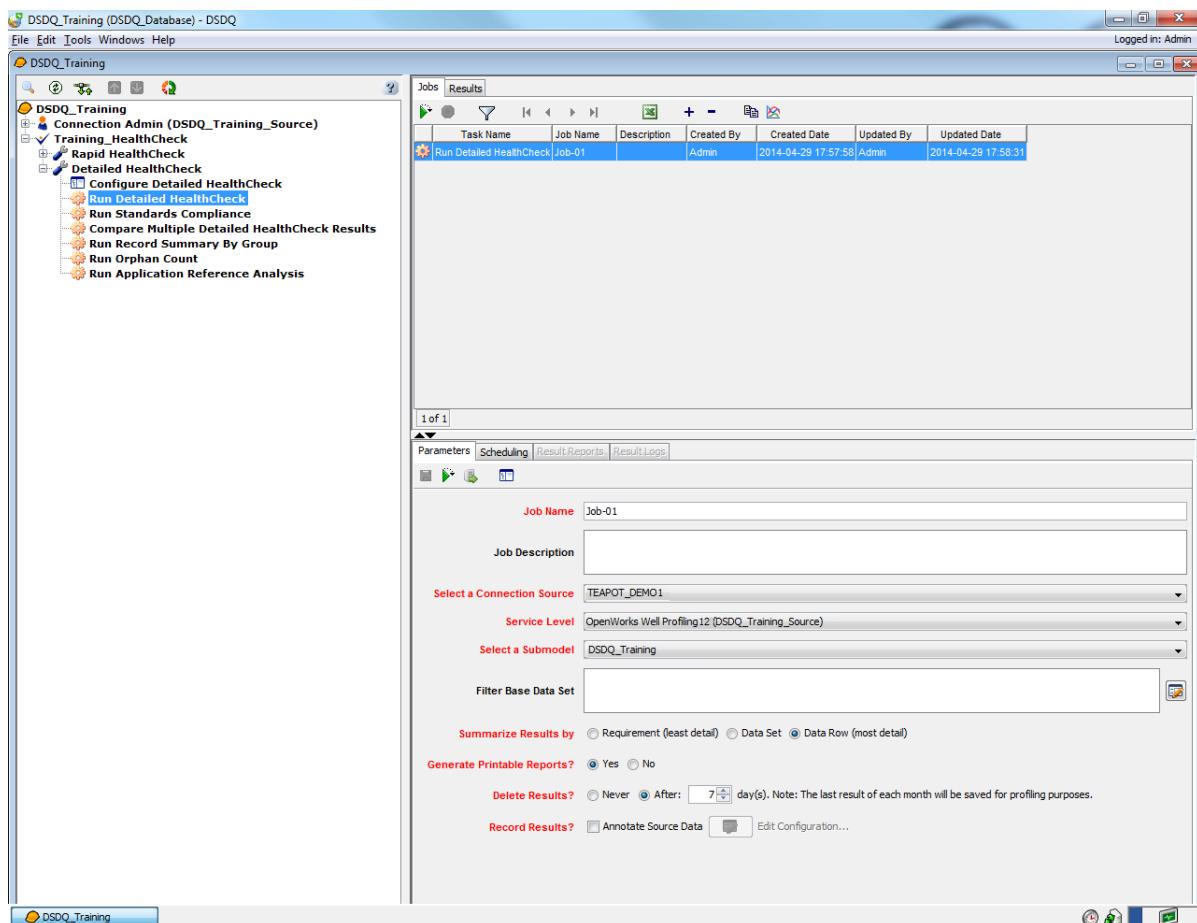
Exercise: Running Detailed HealthCheck

To run detailed HealthCheck:

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 displays on the **Job and Results**

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



2. Enter **Job-01** in the **Job Name** field.
3. Enter **Detailed HealthCheck** in the **Job Description** field.
4. Select **TEAPOT_DEMO1** from the **Select a Connection Source** drop-down list.

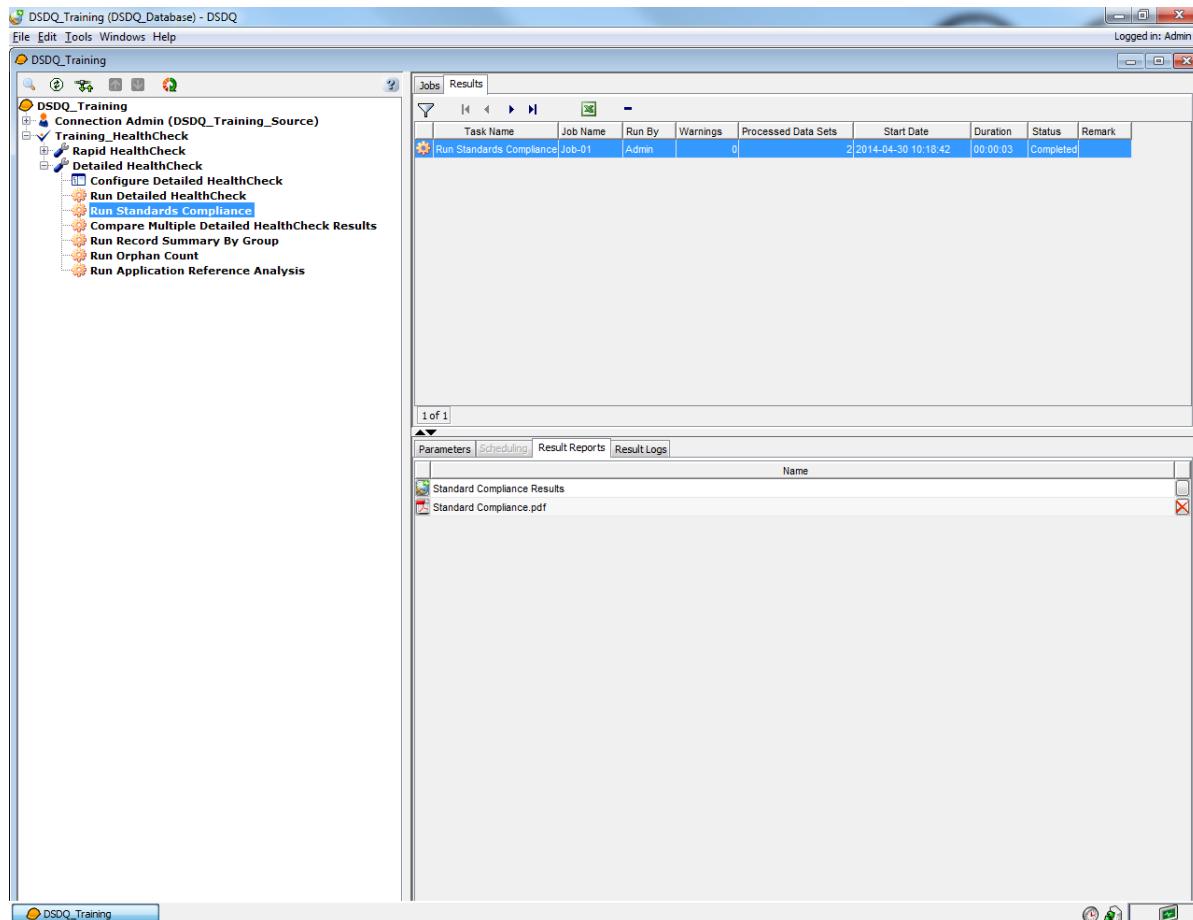
Note

For more information on data owner connections, refer to **Adding a New Data Owner Connection** section in Chapter 2, **Connecting DecisionSpace Data Quality with DecisionSpace Data Server**.

5. Select **OpenWorks Well Profiling 12 (DSDQ_Training_Source)** from the **Service Level** drop-down list.

6. Select **DSDQ_Training** from the **Select a Submodel** drop-down list.
 7. Optionally, you can select a filter for the dataset.
 8. Select **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 **7**.
 11. Do not select the check box for **Record Results?**
 12. Click  to save changes in the **Parameters** tab.
 13. Click .
- The **Detailed HealthCheck** task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

14. 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**.



15. Click  on the **Results Reports** tab to display **Detailed HealthCheck Grouped by Table** results in PDF format.

Detailed HealthCheck Grouped By Table



 Landmark Software & Services

Project:	DSDQ_Training		
Task:	Run Detailed HealthCheck		
Job:	Job-01		
Connection:	DSDQ_Training_Source		
Source:	TEAPOT_DEMO1		
Sub-Model:	DSDQ_Training		
Result Date:	Tue, Dec 24, 2013 12:50		
Table Name:	Well	Number Of Issues In This Table:	
Column Name	Rule Result	Result %	Rule Remark
common_well_name	32	2	Wells that do not have Common Well Name populated
completion_date	1395	100	Wells that do not have a Well Location ID populated.
county	15	1	Wells that have County populated as unknown.
elev_type	0	0	Wells where the Elevation Type is non-standard
elev_type	0	0	The wells elevation type is null
elevation	11	0	Wells where the Elevation is null
elevation	4	0	Wells where the Elevation is equal to zero
latitude	0	0	Wells with a Latitude of 0
latitude	9	0	Wells with a null Latitude
latitude	1	0	Wells with an invalid location Latitude coordinate
longitude	9	0	Wells with a null Longitude
longitude	0	0	Wells with a Longitude of 0
longitude	1	0	Wells with an invalid location Longitude coordinate
total_depth	18	1	Wells where the Total Depth is not populated
total_depth	0	0	Wells where Total Depth is 0
well_name	10	0	Wells that do not have a Well Name populated
well_number	11	0	Wells that do not have a Well Number populated
well_operator	10	0	Wells that have an operator set to unknown
well_operator	0	0	Wells that do not have an operator populated
Table Name:	WellEntire	Number Of Issues In This Table:	
Column Name	Rule Result	Result %	Rule Remark
country	12	0	Wells that have Country populated as unknown.
well_name	10	0	Wells that do not have a Well Name populated

Exercise: Running Standards Compliance

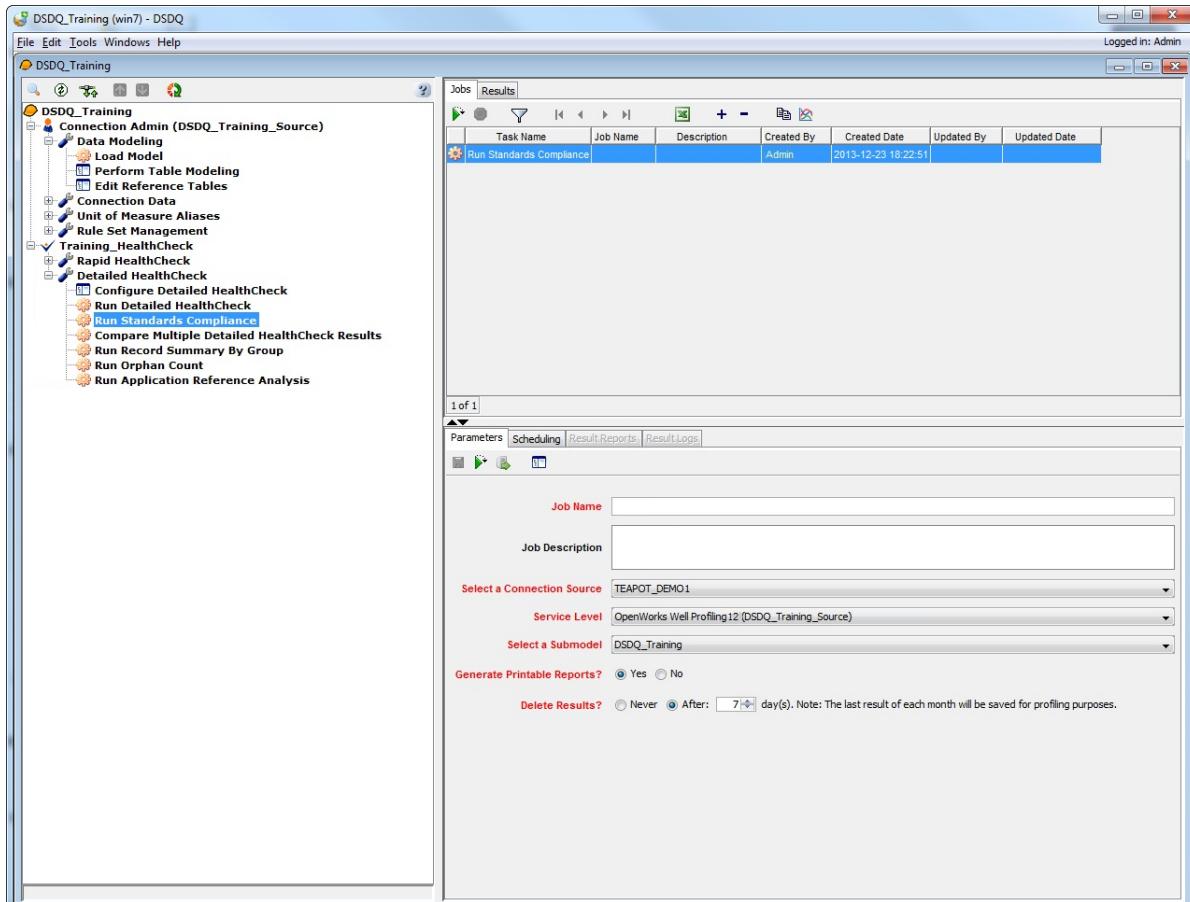
The **Run Standards Compliance** Task checks the number of values within a column. It can also be used to compare the values against the pre-configured values in either the Applications or Standards Reference tables.

To run standards compliance:

1. Double-click the **Run Standards Compliance** Task or right-click the **Run Standards Compliance** 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 **Job-01** in the **Job Name** field.
3. Enter **Standards Compliance** in the **Job Description** field.
4. Select **TEAPOT_DEMO1** from the **Select a Connection Source** drop-down list.

Note

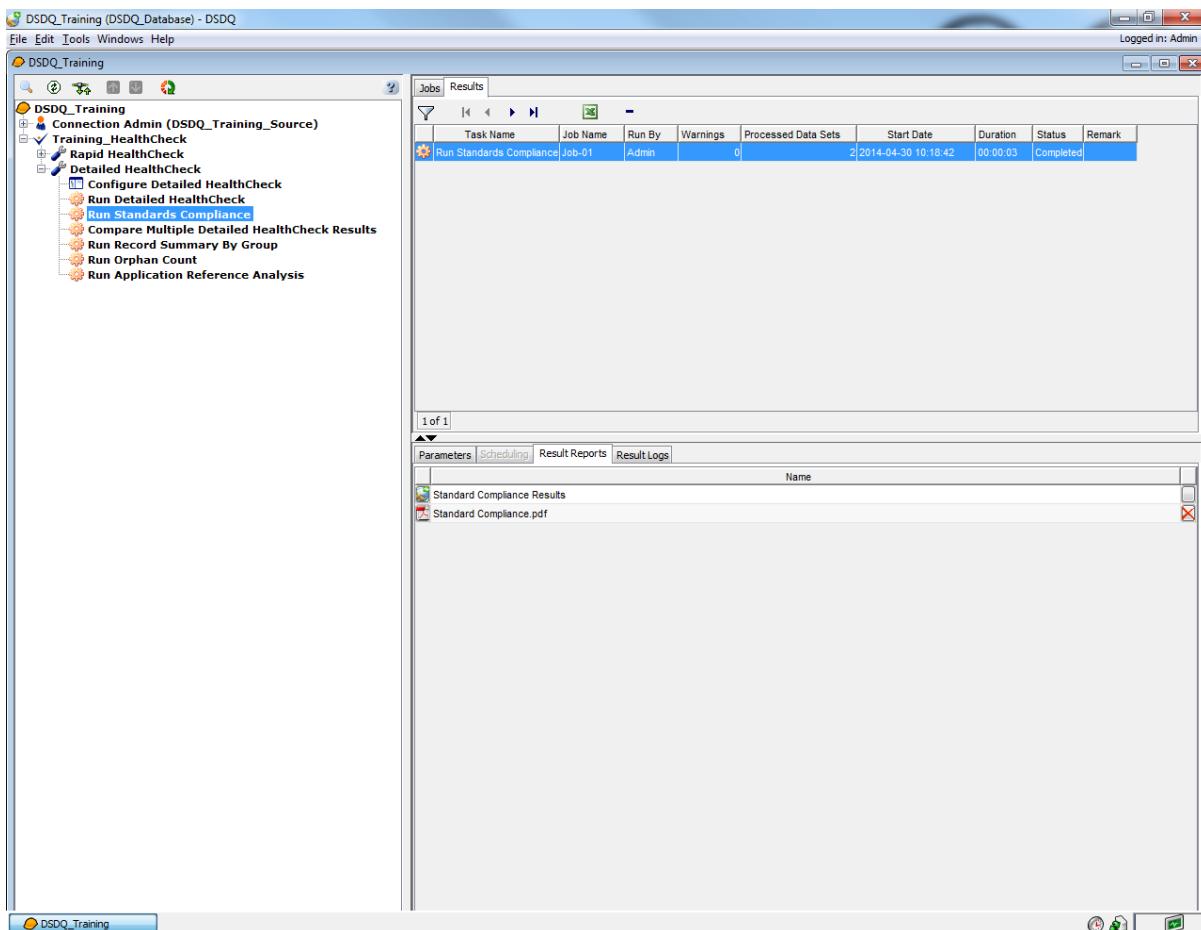
For more information on data owner connections, refer to Adding a New Data Owner Connection section in Chapter 2, Connecting DecisionSpace Data Quality with DecisionSpace Data Server.

5. Select **OpenWorks Well Profiling 12 (DSDQ_Training_Source)** from the **Service Level** drop-down list.

6. Select **DSDQ_Training** from the **Select a Submodel** drop-down list.
7. Select the **Yes** option for **Generate Printable Reports?**
8. Select the **After** option for **Delete Results?** Leave the number of days as **7**.
9. Click  to save changes in the **Parameters** tab.
10. Click .

The **Run Standards Compliance** task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

11. 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**.



12. Click  on the **Results Reports** tab to display **Standard Compliance** results in PDF format.

Standard Compliance			HALLIBURTON Landmark Software & Services	
Project:	DSDQ_Training			
Task:	Run Standards Compliance			
Job:	Job-01			
Connection:	DSDQ_Training_Source			
Source:	TEAPOT_DEMO1			
Result Date:	Thu, Dec 26, 2013 15:47			
Table Name: Well	Column Name: county			
Column Value	Application Reference Value	Standards Reference Value	Column Value Count	Column Value %
NATRONA			1319	95
Natrona			61	5
UNKNOWN			15	2
Table Name: Well	Column Name: well_operator			
Column Value	Application Reference Value	Standards Reference Value	Column Value Count	Column Value %
ADVENTURE			1	1
Anadarko Petroleum Corp.			1	1
Anschutz Corp.			1	1
Continental Industries, LC			1	1
Davis Oil Co.			2	1
F D SERV INC			1	1
Harpel John P Jr.			1	1
Howell Petroleum Corp.			7	1
MKM Exploration			1	1
Mammoth Production			83	6
Milestone Petroleum			1	1
Mutual Oil Co.			1	1
Pacific Enterprises Oil C			1	1
Pacific Enterprises Oil Com			6	1
Plains Petroleum			1	1
Quinisk C R			1	1
RMOTC			13	1
Snyder Oil Corp			2	1
TXO Production Cp.			1	1
Trigood Oil			2	1
U.S. DOE			1255	90
UNKNOWN			10	1
Wind River Resources			1	1
ukn			1	1

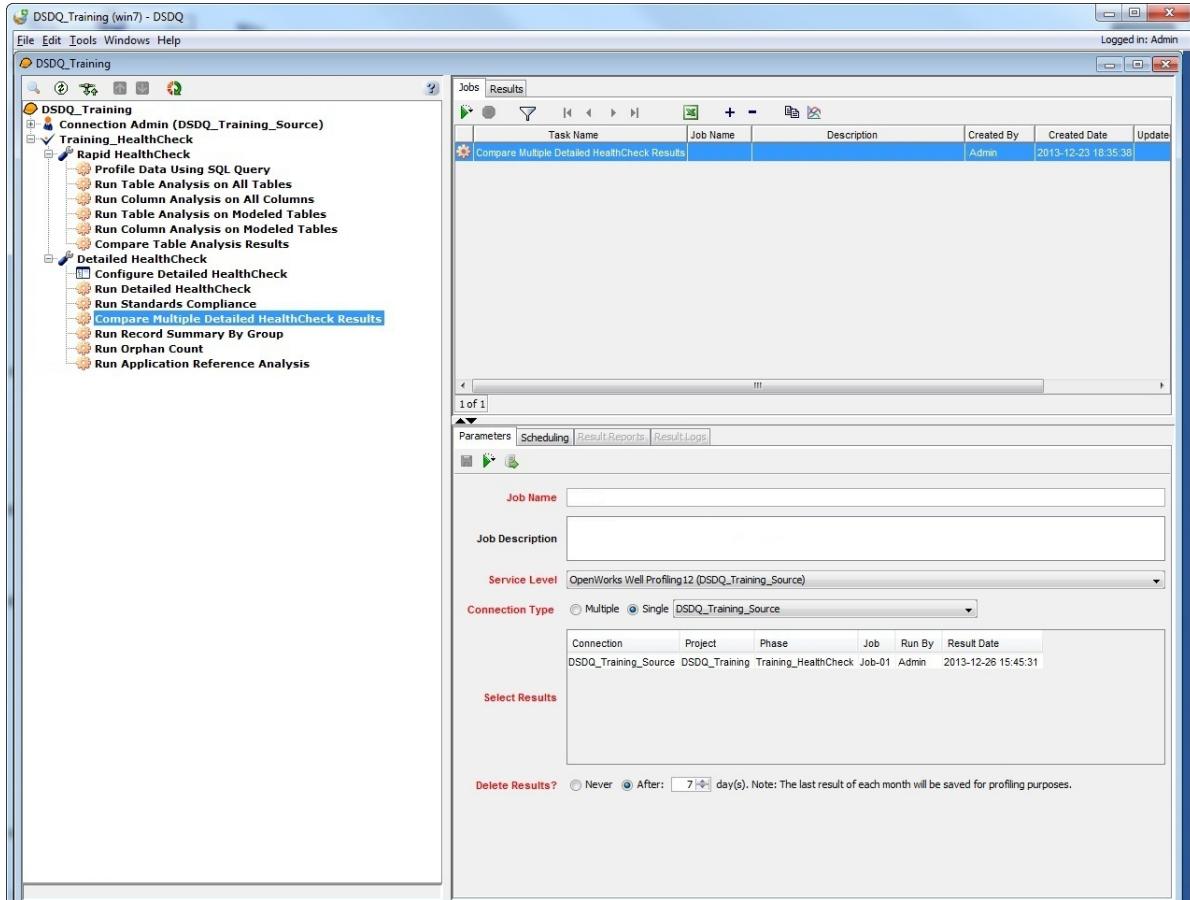
Exercise: Comparing Multiple Detailed HealthCheck Results

To compare multiple Detailed HealthCheck results:

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

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

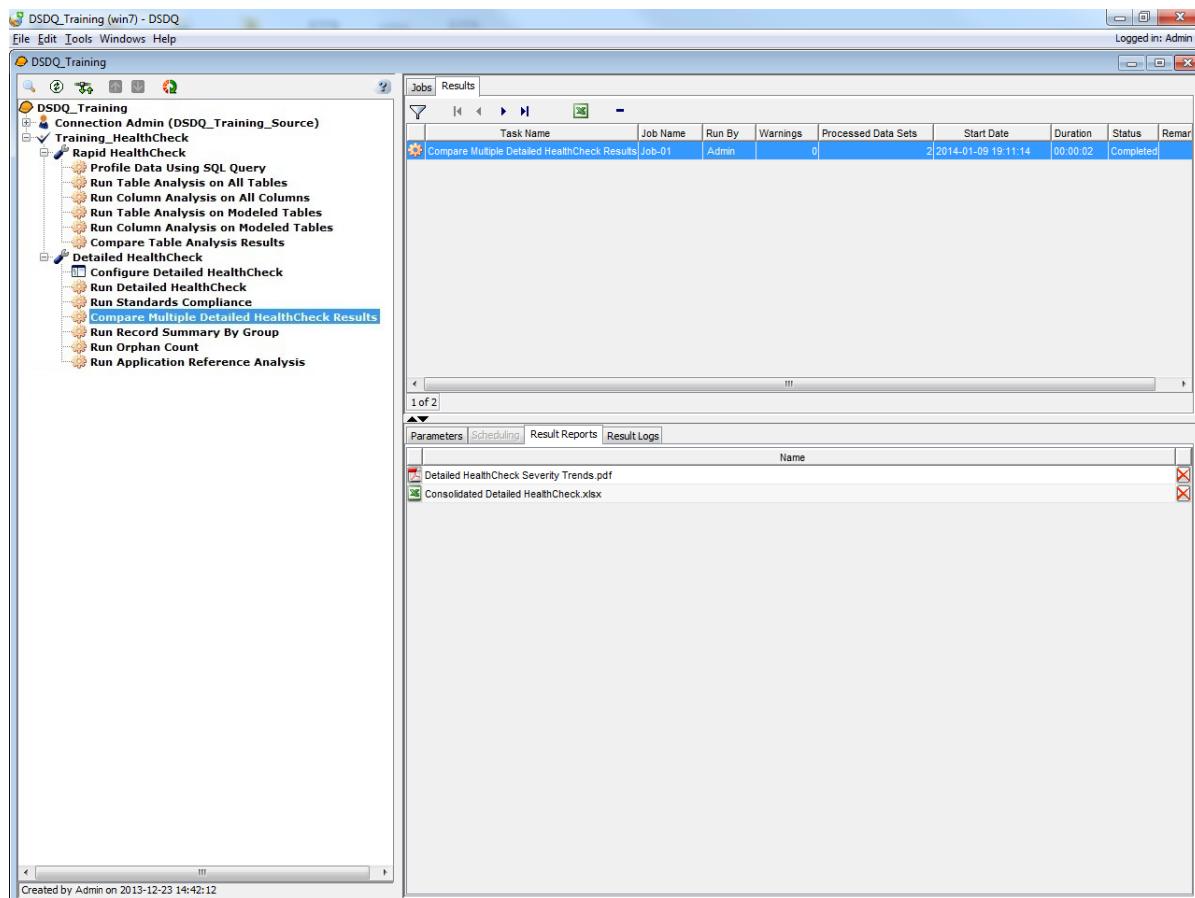
Information Pane on the right side of the DecisionSpace Data Quality Project Window.



2. Enter **Job-01** in the **Job Name** field.
3. Enter **Compare Multiple Detailed HealthCheck Results** in the **Job Description** field.
4. Select **OpenWorks Well Profiling 12 (DSDQ_Training_Source)** from the **Service Level** drop-down list.
5. Select the **Single** option for **Connection Type**.
 - **Multiple** connections automatically list the latest Detailed HealthCheck job run for a connection (i.e. PPDM, WellView, etc.) against the selected Service Level.
 - **Single** connections list only the jobs run under the specified connection for the selected Service Level.

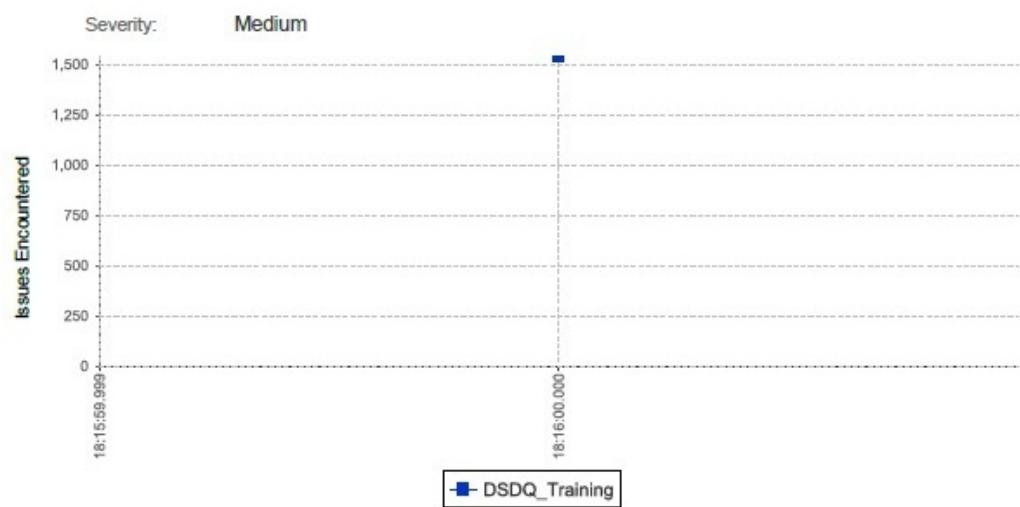
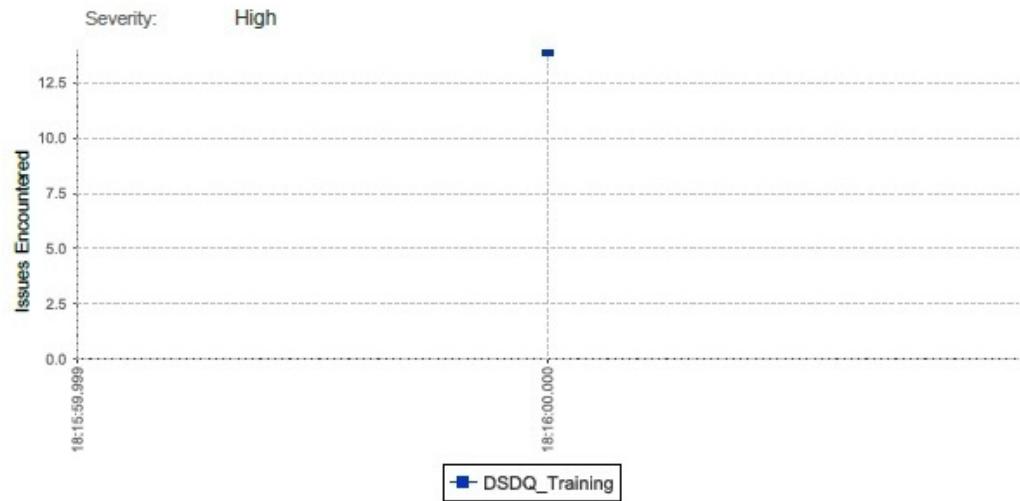
6. Select **DSDQ_Training_Source** from the **Connection Type** drop-down list.
7. Select **DSDQ_Training_Source** from the **Select Results** list.
8. Select the **After** option for **Delete Results?** Leave the number of days as **7**.
9. Click  to save changes in the Parameters tab.
10. Click .

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



11. 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**.

12. Click  on the **Result Reports** tab to display **Detailed HealthCheck Severity Trends** in PDF format.



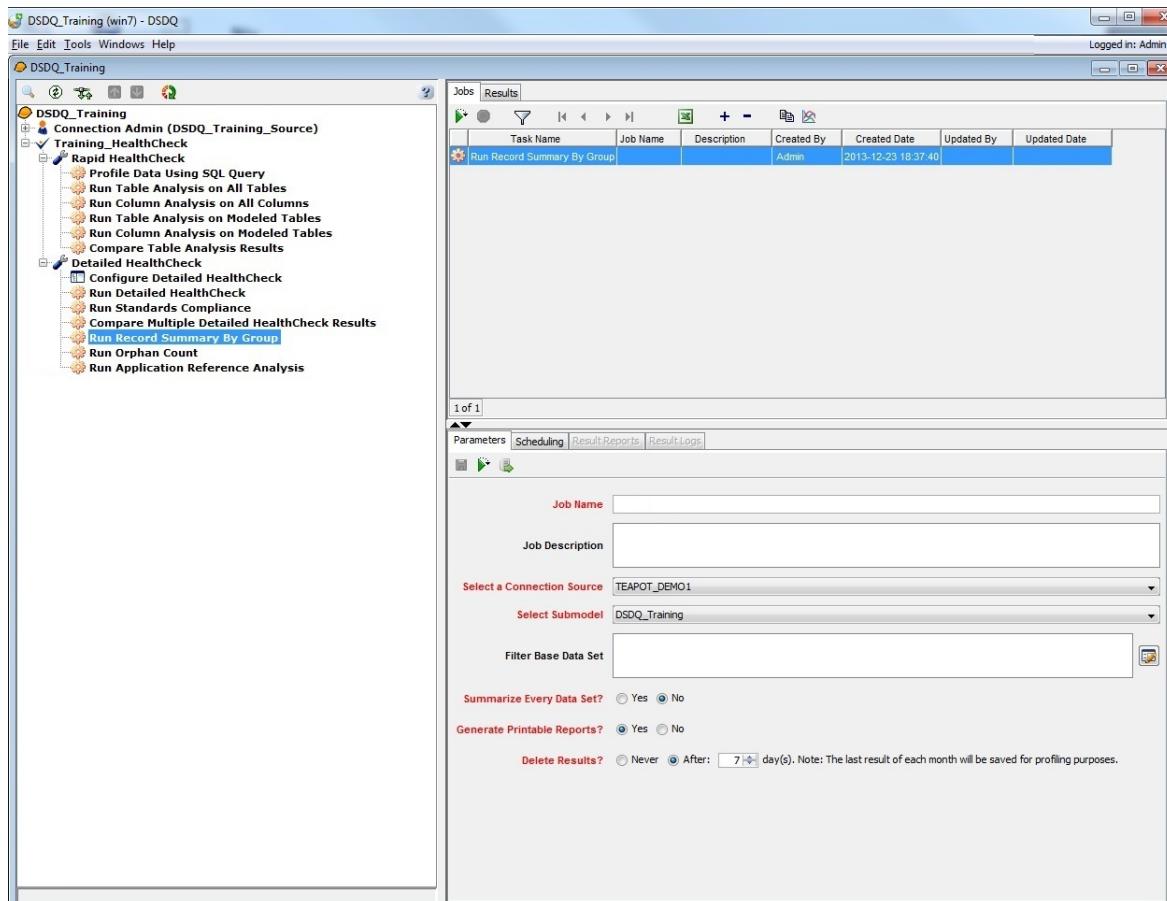
Exercise: Running Record Summary by Group

The **Run Record Summary by Group** Task provides a basic record count based on data groups. Depending on your choice, the task checks the basic row count by table or data group. For example, if the Well Header data group has 10 records and the Drilling Event data group has 50 records, the task counts the number of events that belong to each well.

To run record summary by group(s):

1. Double-click the **Run Record Summary By Group** Task or right-click the **Run Record Summary By Group** 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 **Job-01** in the **Job Name** field.
3. Enter **Summary by Group** in the **Job Description** field.

4. Select **TEAPOT_DEMO1** from the **Select a Connection Source** drop-down list.

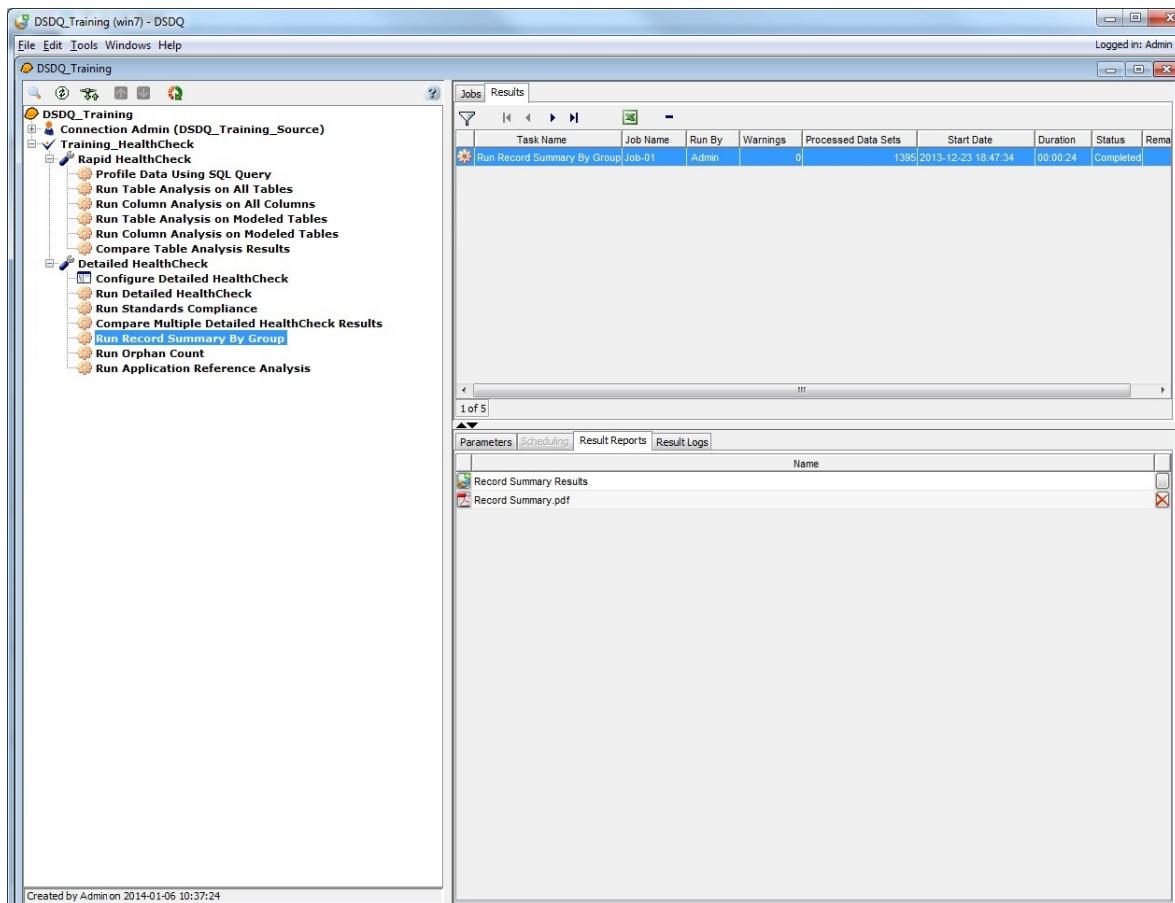
Note

For more information on data owner connections, refer to Adding a New Data Owner Connection section in Chapter 2, Connecting DecisionSpace Data Quality with DecisionSpace Data Server.

5. Select **DSDQ_Training** from the **Select a Submodel** drop-down list.
6. Optionally, set a filter on the data subset.
7. Select **No** from **Summarize Every Data Set**.
8. Select **Yes** from **Generate Printable Reports?**
9. Select the **After** option for **Delete Results?** Leave the number of days as **7**.
10. Click  to save changes in the **Parameters** tab.
11. Click .

The **Run Record Summary by Group** Task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

12. 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**.



13. Click  on the **Result Reports** tab to display **Record Summary By Group** in PDF format.

Record Summary By Group



Project:	DSDQ_Training
Task:	Run Record Summary By Group
Job:	Job-01
Connection:	DSDQ_Training_Source
Source:	TEAPOT_DEMO1
Sub-Model:	DSDQ_Training
Result Date:	Mon, Dec 23, 2013 18:47

Report Group: Complete Data Set	
Data Group	Group Row Count
Well	1395
WellEntire	1399
RCountry	1395
RStateName	1395

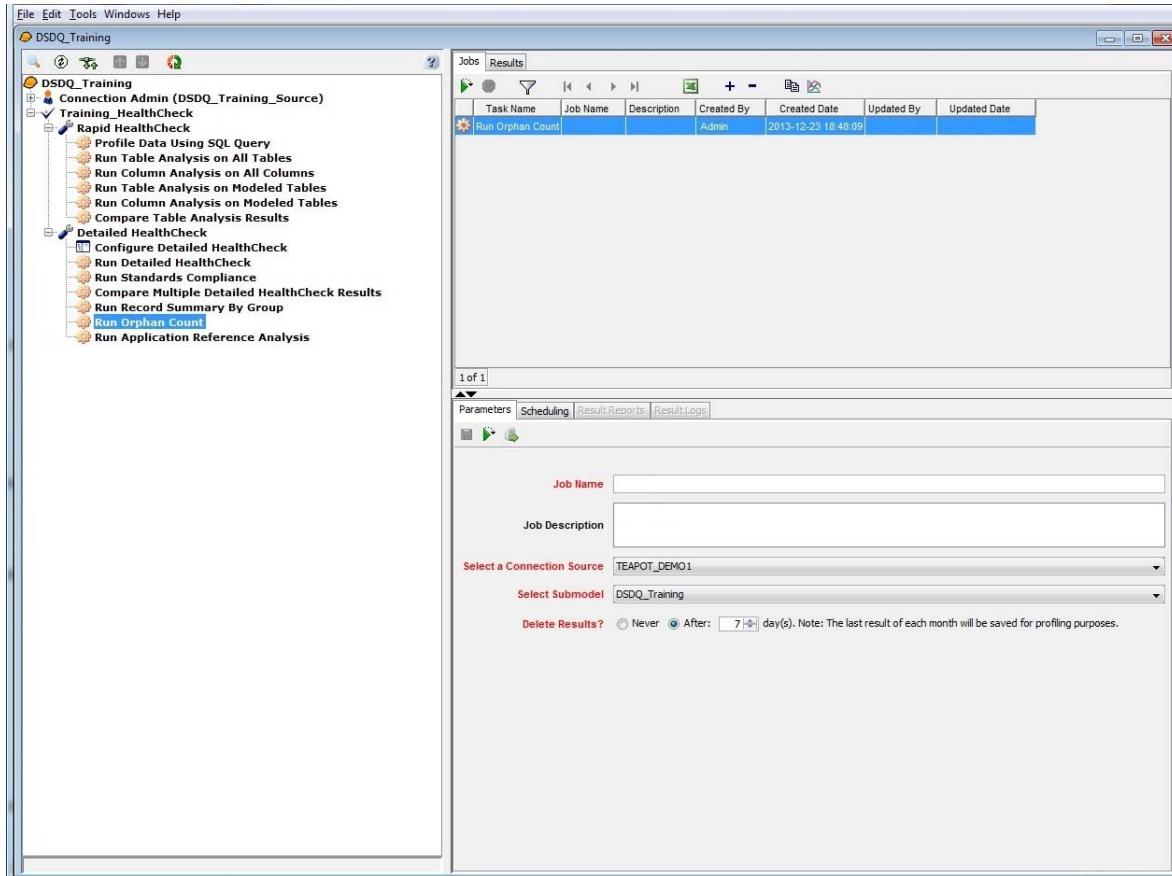
Task Name: Run Record Summary By Group Page 1 of 1
Report Date: Mon, Dec 23, 2013 18:47

Exercise: Running Orphan Count

To run orphan count:

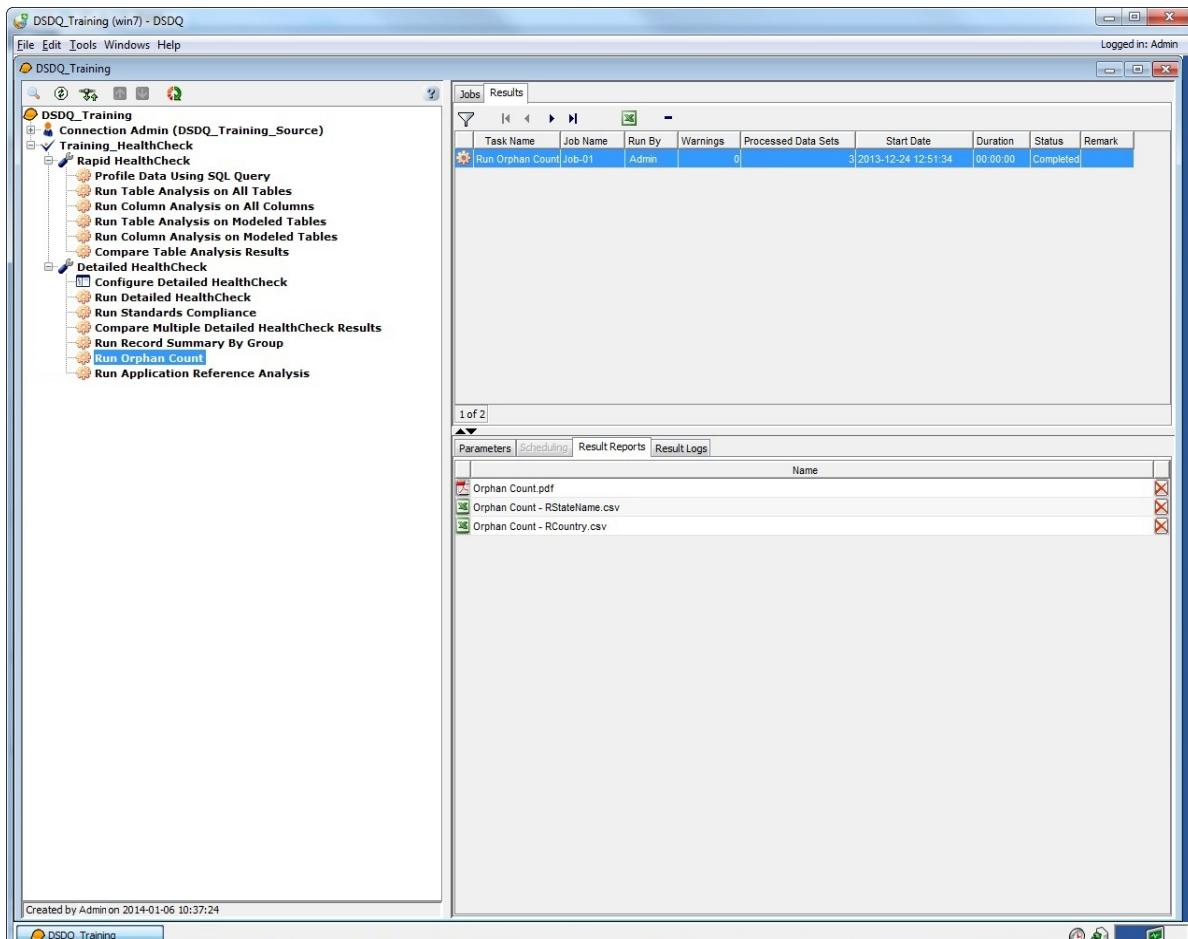
1. Double-click the **Run Orphan Count** Task or right-click the **Run Orphan Count** Task and select **Add Job** from the pop-up menu.
A new job is initiated and displays on the **Job and Results**

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

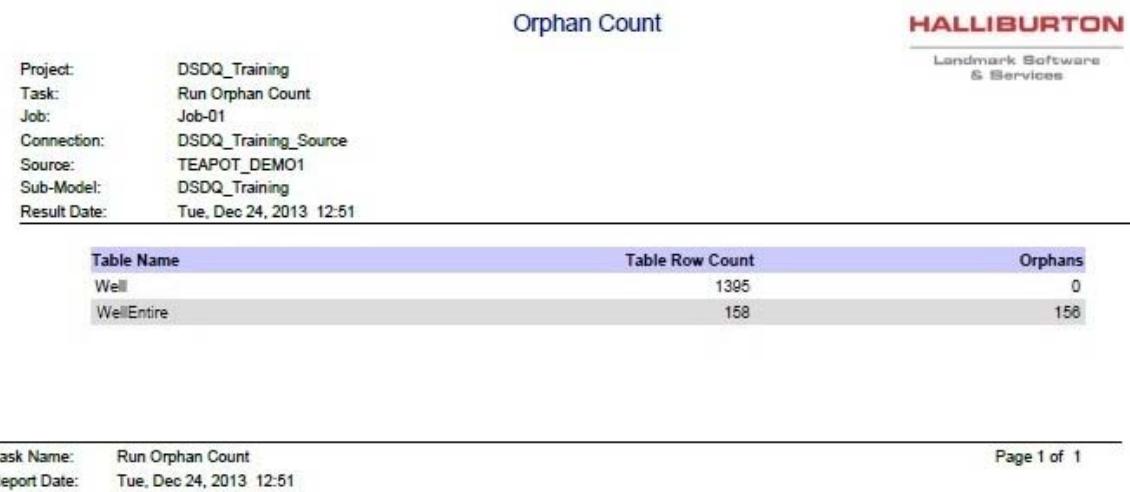


2. Enter **Job-01** in the **Job Name** field.
 3. Enter **Orphan Count** in the **Job Description** field.
 4. Select **TEAPOT_DEMO1** from the **Select a Connection Source** drop-down list.
 5. Select **DSDQ_Training** from the **Select a Submodel** drop-down list.
 6. Select the **After** option for **Delete Results?** Leave the number of days as **7**.
 7. Click to save changes in the **Parameters** tab.
 8. Click .
- The **Run Orphan Count** Task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

9. 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**.



10. Click  on the **Result Reports** tab to display **Orphan Count** in the PDF format.



Orphan Count		
Project:	DSDQ_Training	HALLIBURTON
Task:	Run Orphan Count	Landmark Software & Services
Job:	Job-01	
Connection:	DSDQ_Training_Source	
Source:	TEAPOT_DEMO1	
Sub-Model:	DSDQ_Training	
Result Date:	Tue, Dec 24, 2013 12:51	
Table Name	Table Row Count	Orphans
Well	1395	0
WellEntire	158	158

Task Name: Run Orphan Count
Report Date: Tue, Dec 24, 2013 12:51

Page 1 of 1

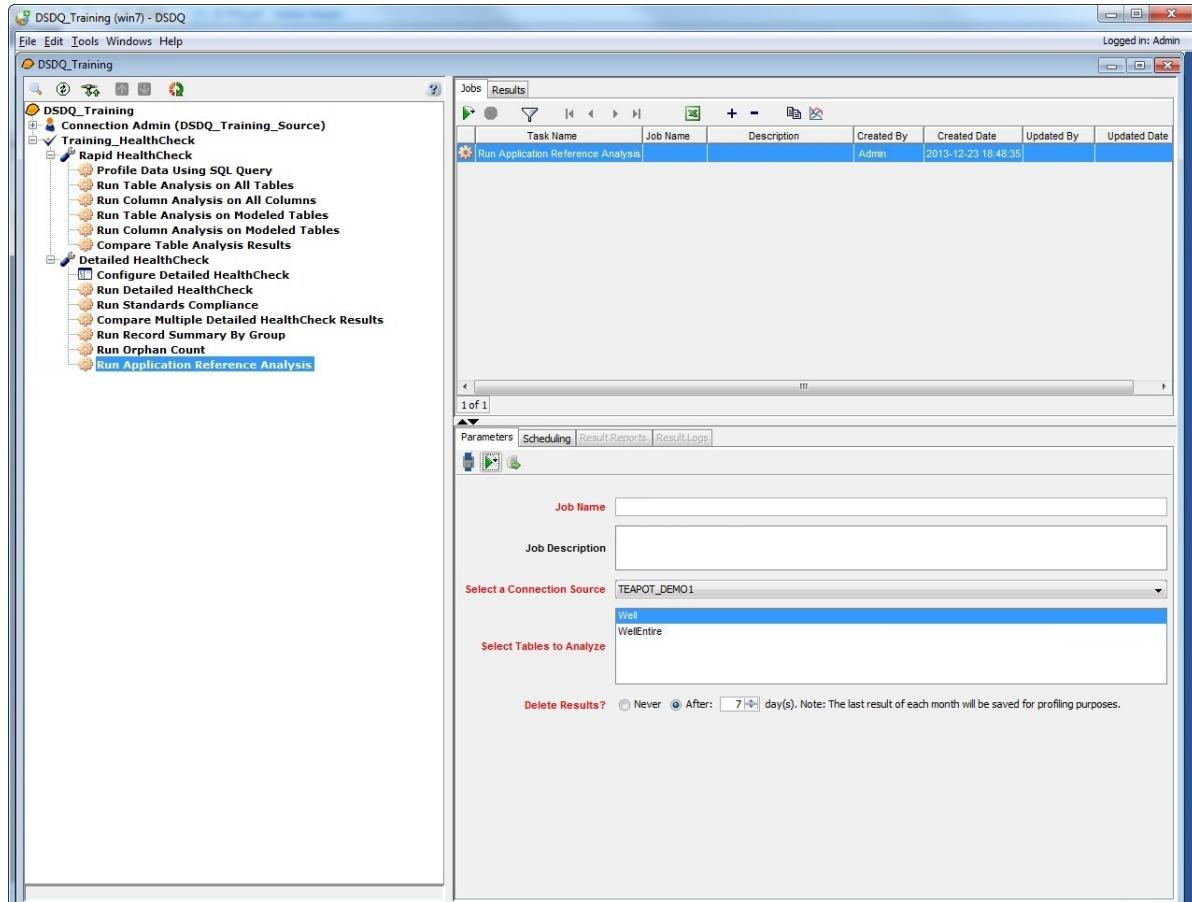
Exercise: Running Application Reference Analysis

The **Run Application Reference Analysis** Task is used to check which columns are associated with a particular Application Reference table, and how many values in each column match a specific Reference table value.

To run application reference analysis:

1. Double-click the **Run Application Reference Analysis** Task or right-click the **Run Application Reference Analysis** 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 **Job-01** in the **Job Name** field.
3. Enter **Application Reference Analysis** in the **Job Description** field.
4. Select **TEAPOT_DEMO1** from the **Select a Connection Source** drop-down list.

Note

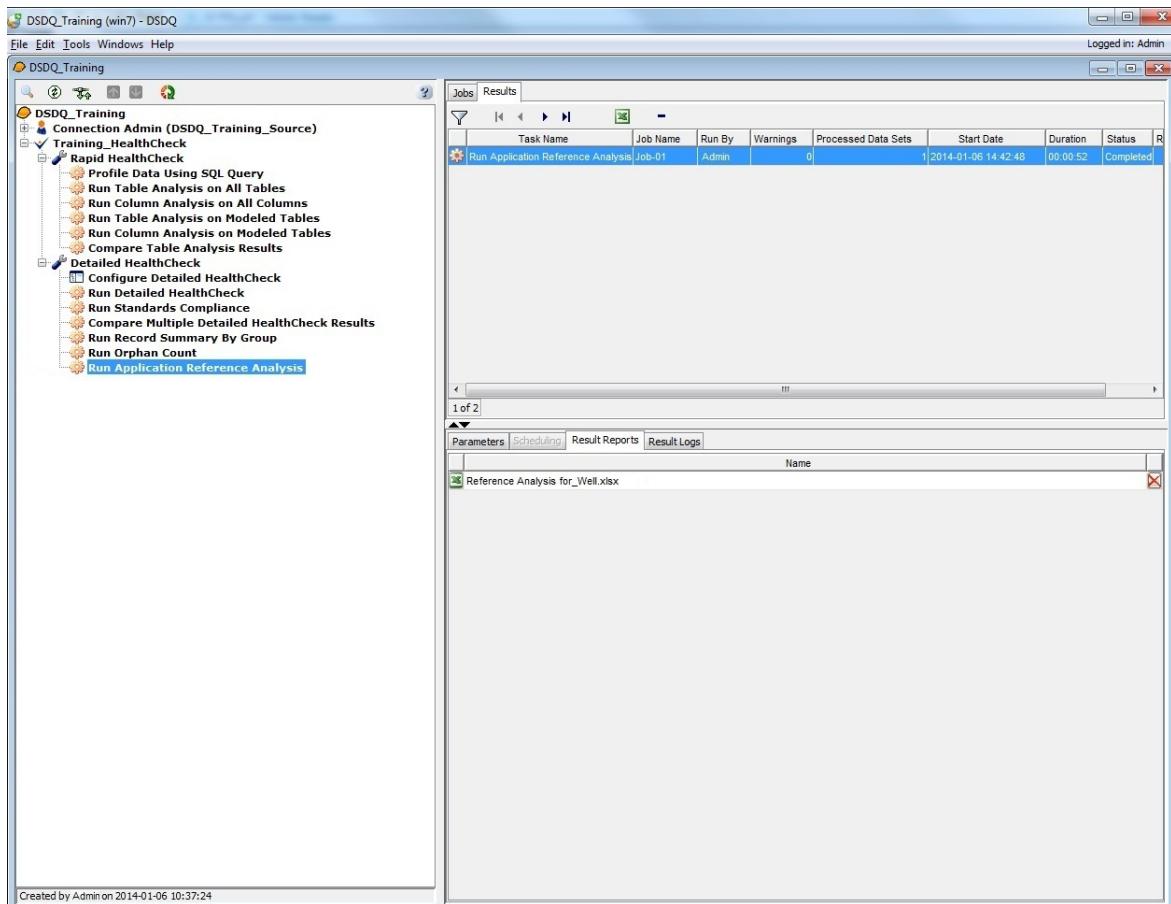
For more information on data owner connections, refer to Adding a New Data Owner Connection section in Chapter 2, **Connecting DecisionSpace Data Quality with DecisionSpace Data Server**.

5. Select **Well** from the **Select Tables to Analyze** list.

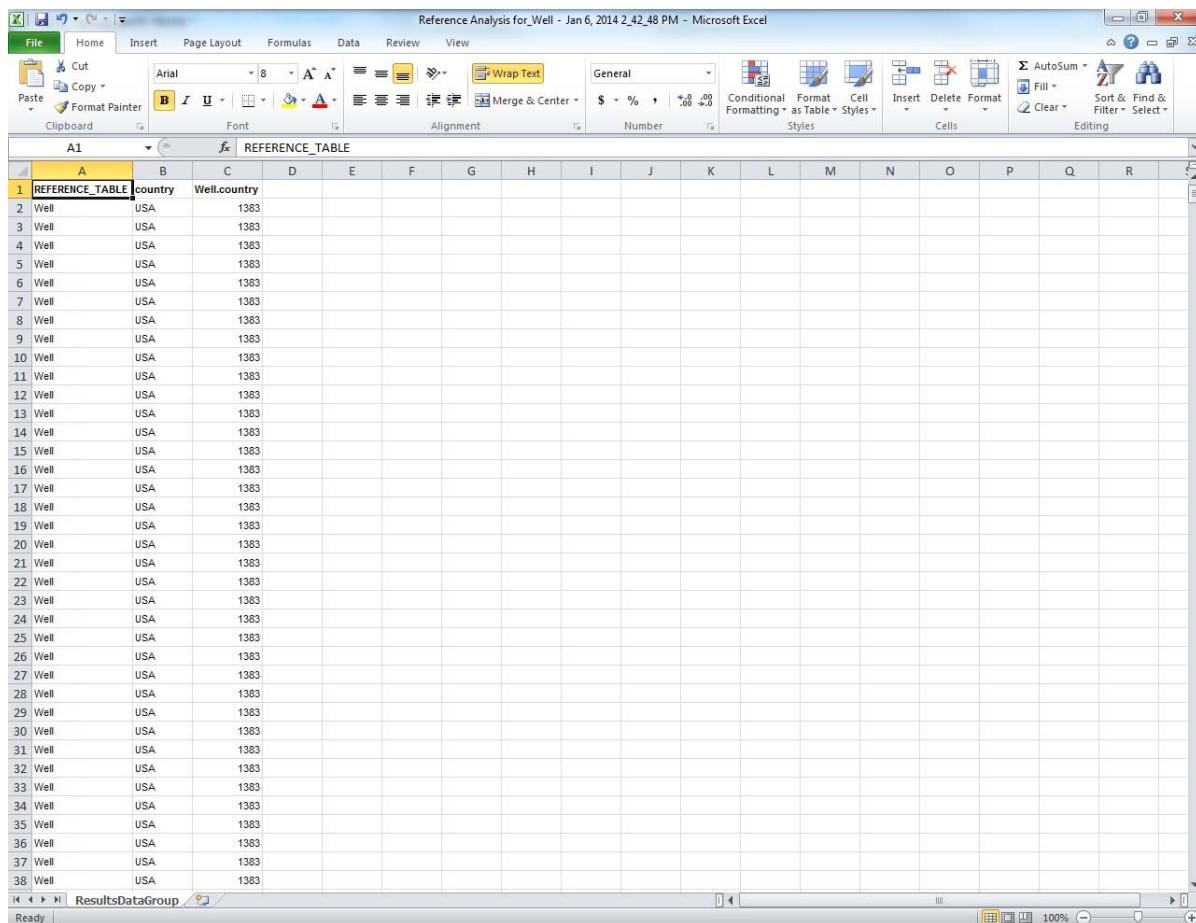
6. Select the **After** option for **Delete Results?** Leave the number of days as **7**.
7. Click  to save changes in the **Parameter** tab.
8. Click .

The **Run Application Reference Analysis Task** is executed and displays results in the **Result Reports** tab.

9. 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**.



10. Double-click  on the **Result Reports** tab to display the **Reference Analysis for_Well** spreadsheet in Microsoft Excel.
 The **Reference Analysis for_Well** spreadsheet appears.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	country	Well.country																
1	REFERENCE_TABLE	USA	1383															
2	Well	USA	1383															
3	Well	USA	1383															
4	Well	USA	1383															
5	Well	USA	1383															
6	Well	USA	1383															
7	Well	USA	1383															
8	Well	USA	1383															
9	Well	USA	1383															
10	Well	USA	1383															
11	Well	USA	1383															
12	Well	USA	1383															
13	Well	USA	1383															
14	Well	USA	1383															
15	Well	USA	1383															
16	Well	USA	1383															
17	Well	USA	1383															
18	Well	USA	1383															
19	Well	USA	1383															
20	Well	USA	1383															
21	Well	USA	1383															
22	Well	USA	1383															
23	Well	USA	1383															
24	Well	USA	1383															
25	Well	USA	1383															
26	Well	USA	1383															
27	Well	USA	1383															
28	Well	USA	1383															
29	Well	USA	1383															
30	Well	USA	1383															
31	Well	USA	1383															
32	Well	USA	1383															
33	Well	USA	1383															
34	Well	USA	1383															
35	Well	USA	1383															
36	Well	USA	1383															
37	Well	USA	1383															
38	Well	USA	1383															

Chapter 5

Data Cleansing and Standardization

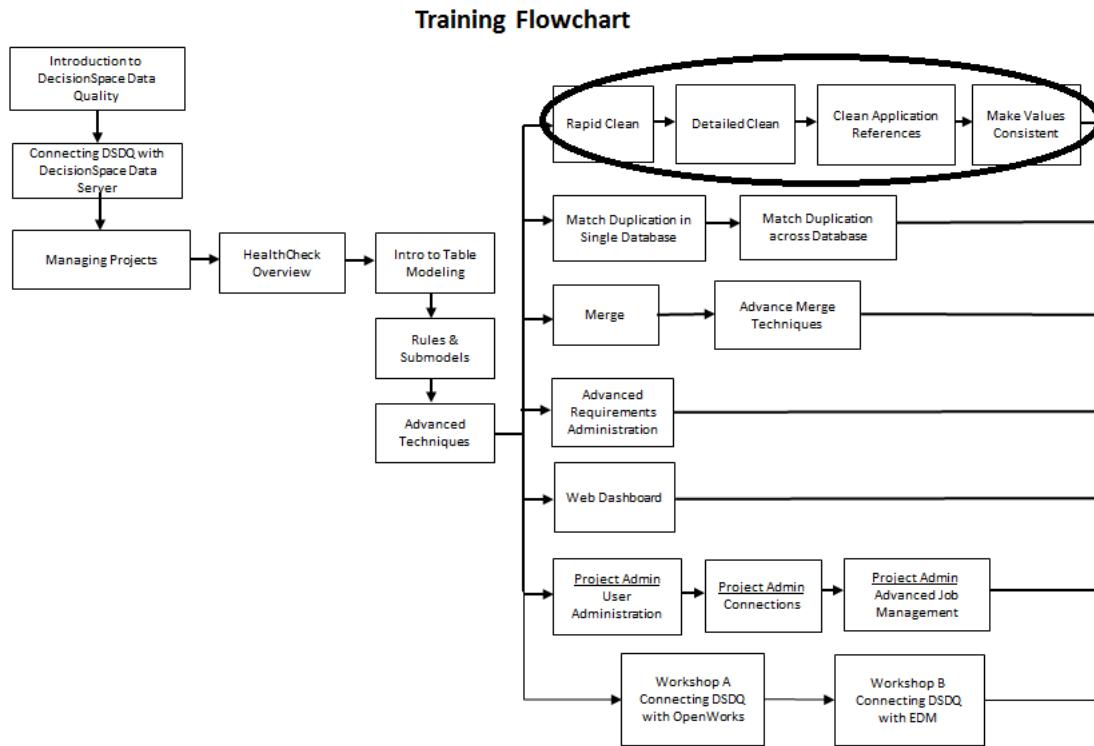
The Clean Phase enables you to address data quality issues across disparate data stores by means of quality control queries and a repeatable cleansing methodology. This Phase includes two Activities i.e. **Rapid Clean** and **Detailed Clean** and it is designed to work through data issues found during the HealthCheck phase.

Chapter Overview

In this chapter, you will learn about:

- Using the Rapid Clean Activity
- Using the Detailed Clean Activity
- Cleaning Application References
- Using the Making Values Consistent Tool

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:



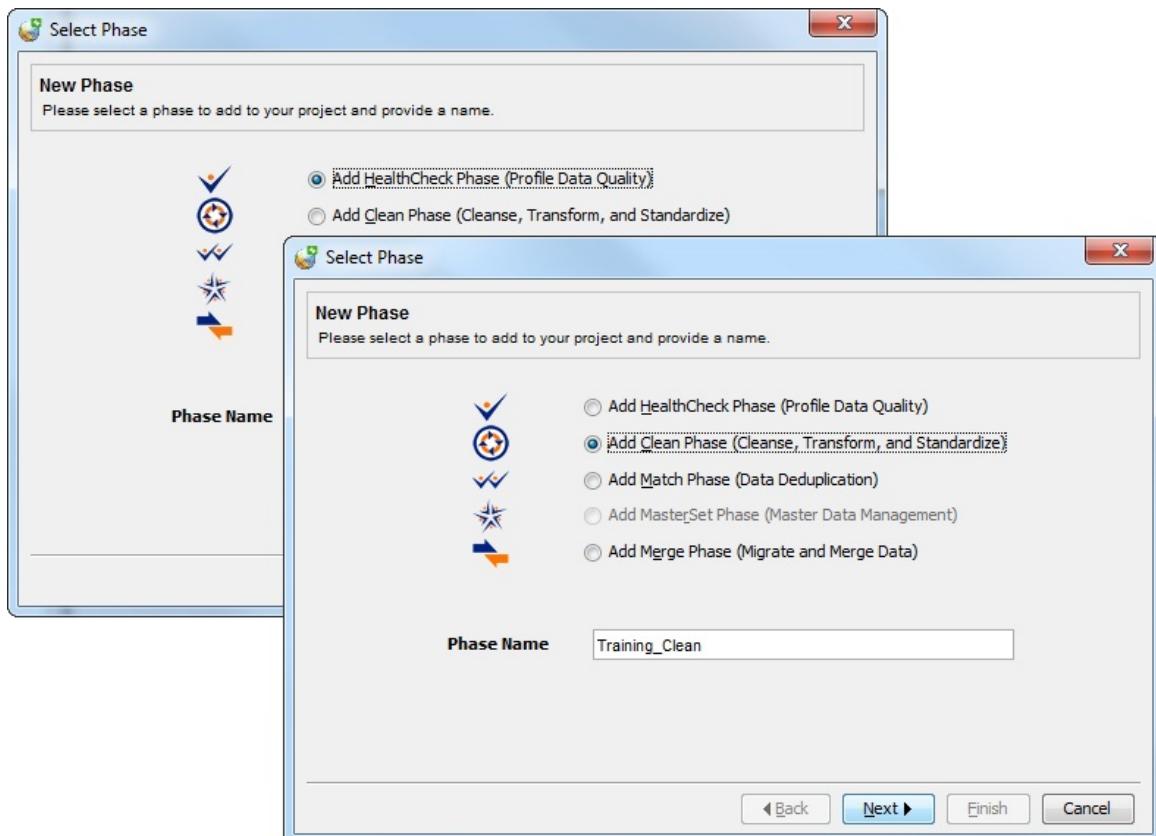
Resolving Data Issues

Issues found in the data during the HealthCheck Phase are cleaned by using the Clean Phase. The Clean Phase comprises of Rapid Clean and Detailed Clean Activities. By using this Data Quality feature, you can apply individual policies to transform your data while ensuring that the original database is updated with the correct policies.

Exercise: Adding a Clean Phase

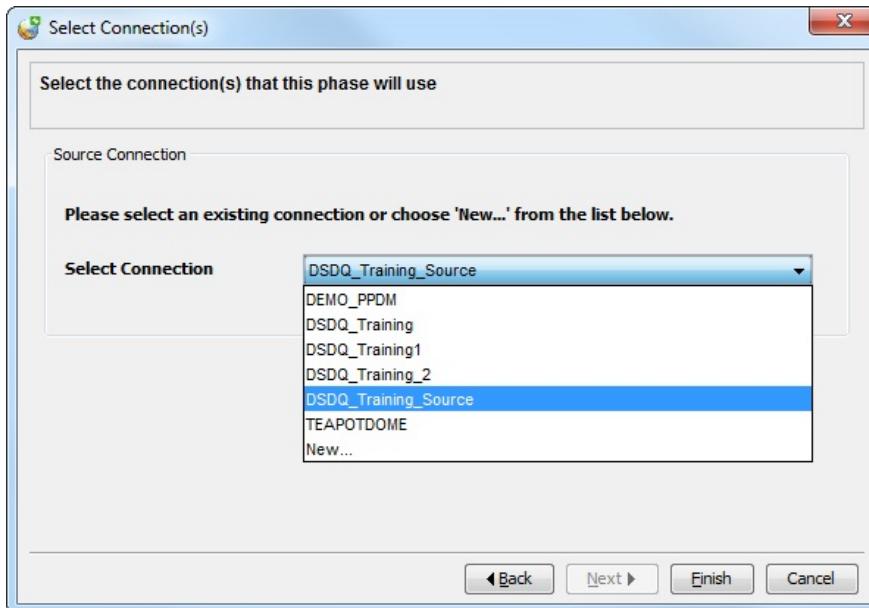
To add a Clean Phase:

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



2. Select the **Add Clean Phase (Cleanse, Transform, and Standardize)** option.

3. Enter **Training_Clean** in the **Phase Name** field.
4. Click **Next** to continue.
The **Select Connection(s)** window appears.



5. Select **DSDQ_Training_Source** 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 issues. Issues that can be cleaned out or corrected with this tool are:

- Mixed Case
- Non Printable Characters
- Preceding White Space
- Trailing White Space
- Double While Space

Note

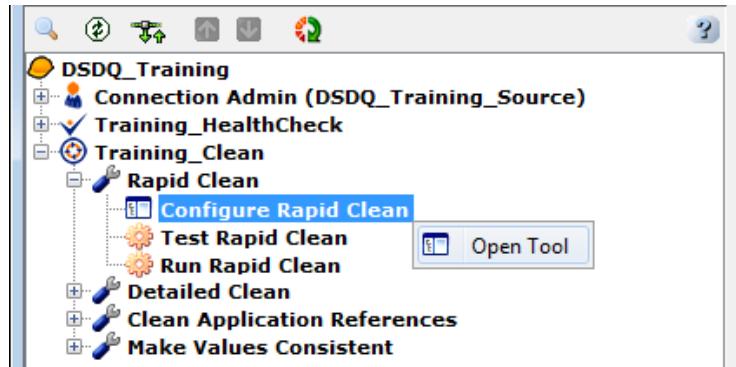
You must have already run a **Column Analysis** on Modeled Tables to run Rapid Clean. The Analysis information is required to display the issues found.

Exercise: Configuring the Rapid Clean Tool

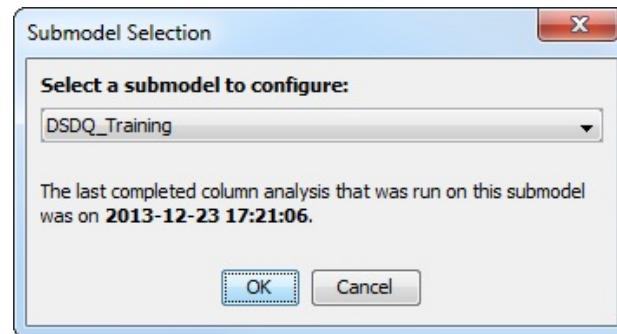
To configure the Rapid Clean Tool:

1. Click  on the DecisionSpace Data Quality Tree to expand the **Training_Clean** Phase.
2. Click  to expand the **Rapid Clean** Activity.
3. 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.



Note

The purpose of the date and time display is to let the user know that any clean changes made after this stated date and time will not be shown when the tool is opened. To see the updated changes after the date and time displayed here requires re-running the **Run Column Analysis on Modeled Tables** Task.

4. Select **DSDQ_Training** from the **Select a submodel to configure** drop-down list
5. Click **OK**.

Note

The **Submodel Selection** dialog box only displays the submodels on which **Run Column Analysis on Modeled Tables** has been run.

The **Configure Rapid Clean** window appears. On the left side of the window is the Column Issues Tree, a tree that displays issues found during column analysis, and the columns and tables in which they occur. On the right side is the **Column Analysis Details** Pane,

which displays column analysis information for the currently selected node in the tree. You can expand a table or an issue by clicking the plus + sign next to it. When a table is expanded, all issues for that table are displayed. When an issue is expanded, all column(s) in the table with the issue are displayed.

The screenshot shows the 'Configure Rapid Clean - Training_Clean' application window. The main area is titled 'Column Analysis and Details Pane'. On the left, there is a tree view of tables and issues. Annotations with arrows point to specific elements:

- Table:** Points to the 'WellEntre' table node in the tree view.
- To LowerCase:** Points to the 'common_well_name' column under the 'Well' table node. It indicates that this column converts characters to lowercase.
- Issue:** Points to the 'Mixed Case' issue node under the 'WellEntre' table node. It indicates that this issue converts characters to uppercase.
- ToUpperCase:** Points to the 'common_well_name' column under the 'Well' table node. It indicates that this column converts characters to uppercase.

On the right, there is a grid table with the following columns and data:

Table Name	Column Name	Rows	# Not Null	% Populated	# Unique	# Mixed Case	# NPC	# PWS	# TWS	# DWS
WellEntre	basin	1395	1395	100	3	61	0	0	0	0
WellEntre	common_well_name	1395	1383	98	1305	20	0	0	0	0
WellEntre	county	1395	1395	100	3	61	0	0	0	0
WellEntre	drilling_operator	1395	1395	100	29	114	0	0	0	0
WellEntre	field	1395	1395	100	7	61	0	0	0	0
WellEntre	remark	1395	1310	94	39	0	0	0	0	0
WellEntre	state	1395	1395	100	3	61	0	0	0	0
WellEntre	uwi	1395	1395	100	1395	4	0	0	0	0
WellEntre	well_location_uwi	1395	1395	100	1393	2	0	0	0	0
WellEntre	well_name	1395	1385	99	20	25	0	0	0	0
WellEntre	well_number	1395	1384	99	1353	218	0	0	0	0
WellEntre	well_operator	1395	1395	100	24	114	0	0	0	0

6. Select the check box(s) next to table(s) with issues.

Issue(s) in the column for the selected table will be highlighted in the **Column Analysis Details Pane**. The selected issues will be corrected when the Rapid Clean task is run. For columns containing Mixed Case characters, an extra icon appears to the left of the column name. Clicking this icon will toggle between converting mixed case characters to uppercase or lowercase. The default is set to uppercase. Selecting will convert all characters in the column to uppercase and selecting will convert all characters in the column to lowercase when the Rapid Clean task is run.

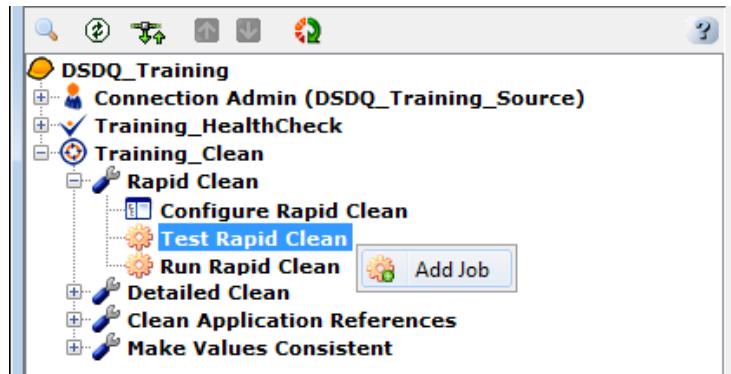
7. Select **File > Exit** from the menu bar on the **Configure Rapid Clean** window to close it.

Exercise: Running the Test Rapid Clean Task

After cleaning the issues using the **Configure Rapid Clean** Tool, the **Test Rapid Clean Task** is run to make sure that the expected results are

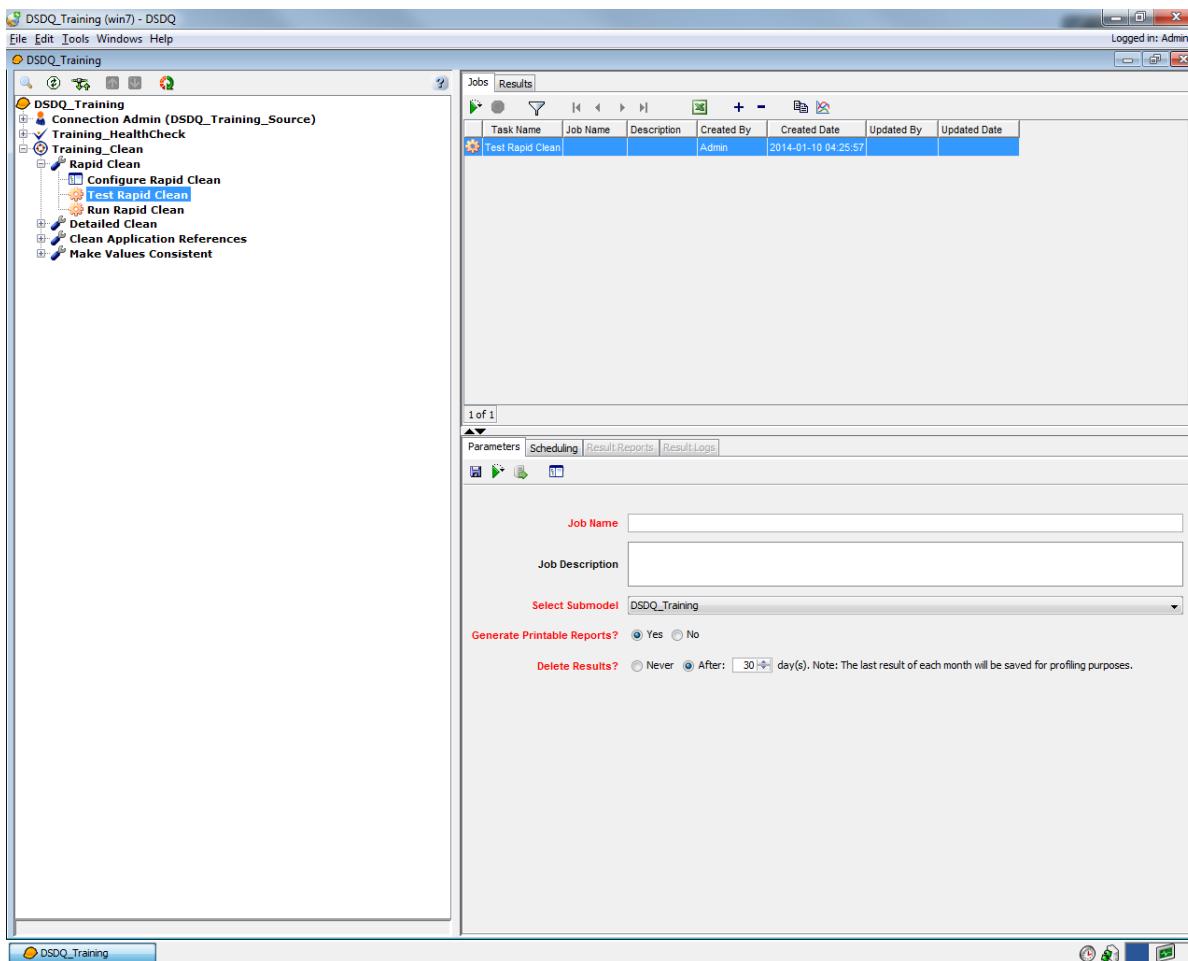
seen before running the **Run Rapid Clean** Task to fix the entire dataset in the submodel. To run the **Test Rapid Clean** Task:

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.



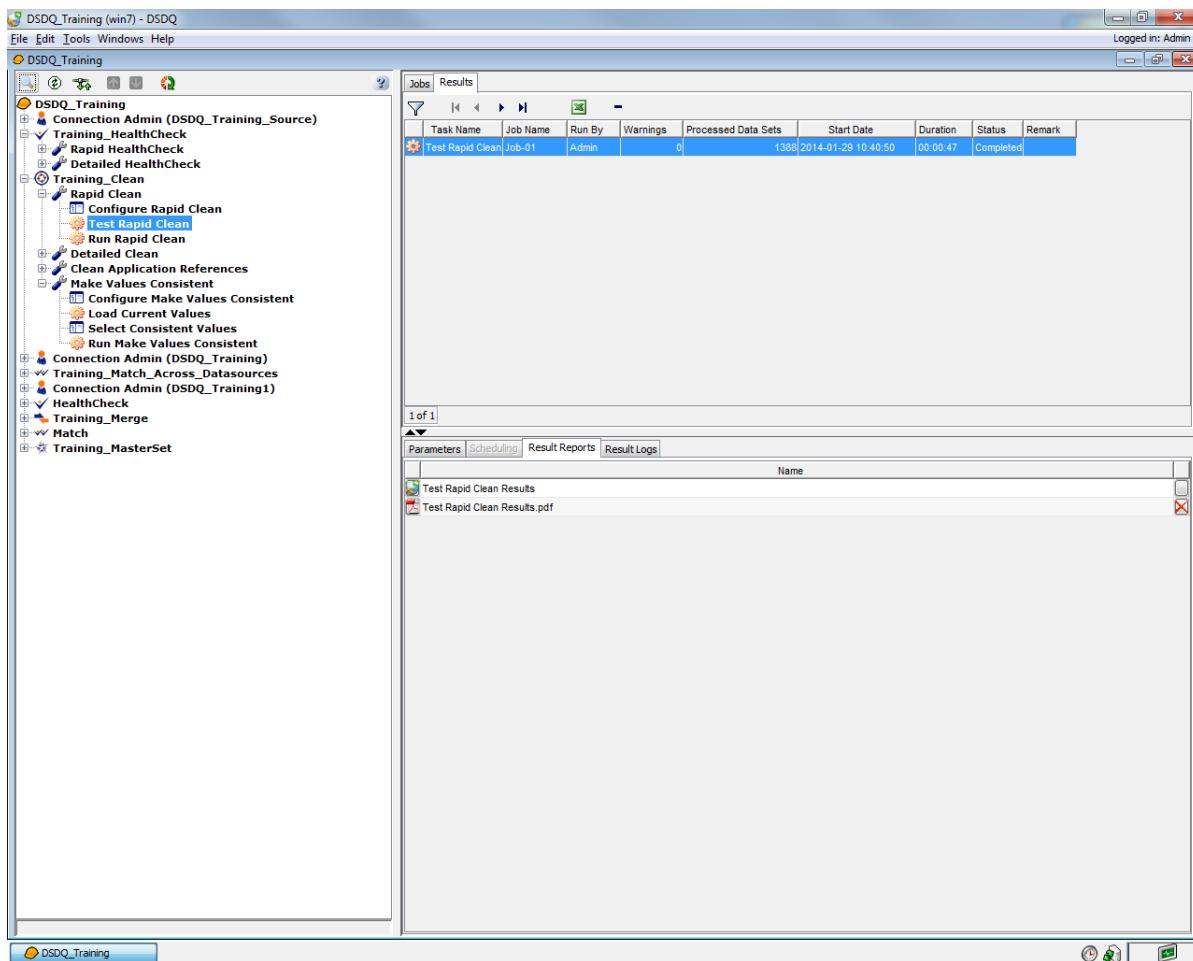
A new job is initiated and it displays on the **Job and Results**

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

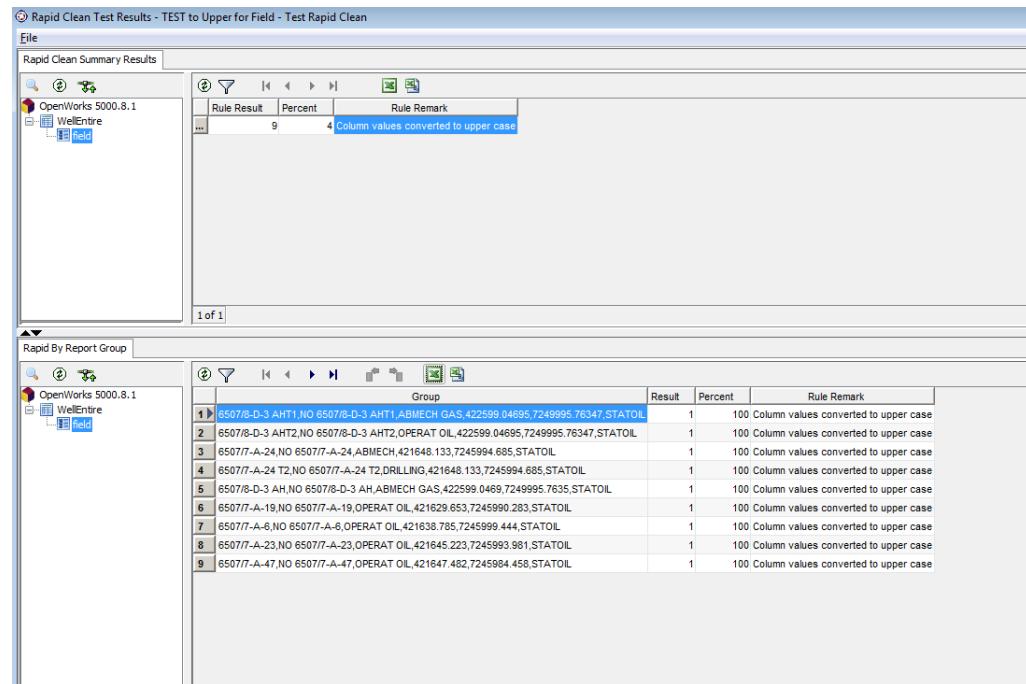


2. Enter **Job-01** in the **Job Name** field.
 3. Enter **Rapid Clean Test** in the **Job Description** field.
 4. Select **DSDQ_Training** from the **Select Submodel** drop-down list.
 5. Select the **Yes** option for **Generate Printable Reports?**
 6. Select the **After** option for **Delete Results?** Leave the number of days as **7**.
 7. Click to save changes in the **Parameters** tab.
 8. Click .
- The **Test Rapid Clean** task runs and displays results in the **Result Reports** tab of the **Job and Results Information** Pane.

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



10. Click  on the **Result Reports** tab of the Job and Results Information Pane to display the Test Rapid Clean Results.



Group	Result	Percent	Rule Remark
1 6507/B-D-3 AHT1,NO 6507/B-D-3 AHT1,ABMECH GAS,422599.04695,7249995.76347,STATOIL	1	100	Column values converted to upper case
2 6507/B-D-3 AHT2,NO 6507/B-D-3 AHT2,OPERAT OIL,422599.04695,7249995.76347,STATOIL	1	100	Column values converted to upper case
3 6507/T-A-24,NO 6507/T-A-24,ABMECH,421648.133,7245994.685,STATOIL	1	100	Column values converted to upper case
4 6507/T-A-24 T2,NO 6507/T-A-24 T2,DRILLING,421648.133,7245994.685,STATOIL	1	100	Column values converted to upper case
5 6507/B-D-3 AH,NO 6507/B-D-3 AH,ABMECH GAS,422599.04695,7249995.7635,STATOIL	1	100	Column values converted to upper case
6 6507/T-A-19,NO 6507/T-A-19,OPERAT OIL,421629.653,7245990.283,STATOIL	1	100	Column values converted to upper case
7 6507/T-A-6,NO 6507/T-A-6,OPERAT OIL,421638.785,7245999.444,STATOIL	1	100	Column values converted to upper case
8 6507/T-A-23,NO 6507/T-A-23,OPERAT OIL,421645.223,7245993.981,STATOIL	1	100	Column values converted to upper case
9 6507/T-A-47,NO 6507/T-A-47,OPERAT OIL,421647.482,7245984.458,STATOIL	1	100	Column values converted to upper case

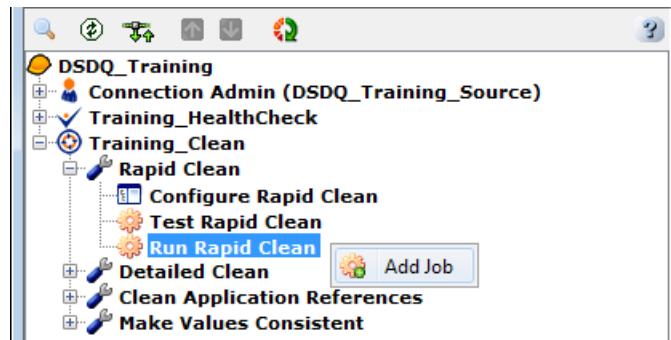
11. Click  on the **Result Reports** tab on the **Job and Results Information** Pane to display the **Test Rapid Clean Results** in PDF format.

Test Rapid Clean Results				HALLIBURTON Landmark Software & Services
Table Name: RCountry				Number of Changes in This Table: 0
Column Name	Rule Result	Result %	Result Description	
remark	0	0	Trailing white space removed	
Table Name: Well				Number of Changes in This Table: 457
Column Name	Rule Result	Result %	Result Description	
common_well_name	20	1	Column values converted to upper case	
county	61	4	Column values converted to upper case	
field	61	4	Column values converted to upper case	
state	61	4	Column values converted to upper case	
well_name	36	2	Column values converted to upper case	
well_number	218	15	Column values converted to upper case	
Table Name: WellEntire				Number of Changes in This Table: 1234
Column Name	Rule Result	Result %	Result Description	
basin	61	4	Column values converted to upper case	
common_well_name	20	1	Column values converted to upper case	
county	61	4	Column values converted to upper case	
drilling_operator	114	8	Column values converted to upper case	
field	61	4	Column values converted to upper case	
remark	479	34	Double white space removed	
state	61	4	Column values converted to upper case	
uwi	6	0	Column values converted to upper case	
well_location_uwi	2	0	Column values converted to upper case	
well_name	36	2	Column values converted to upper case	
well_number	218	15	Column values converted to upper case	
well_operator	115	8	Column values converted to upper case	

Exercise: Running the Rapid Clean Task

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

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 DSDQ interface with the 'Job and Results' pane open. In the 'Jobs' tab, there is one job entry:

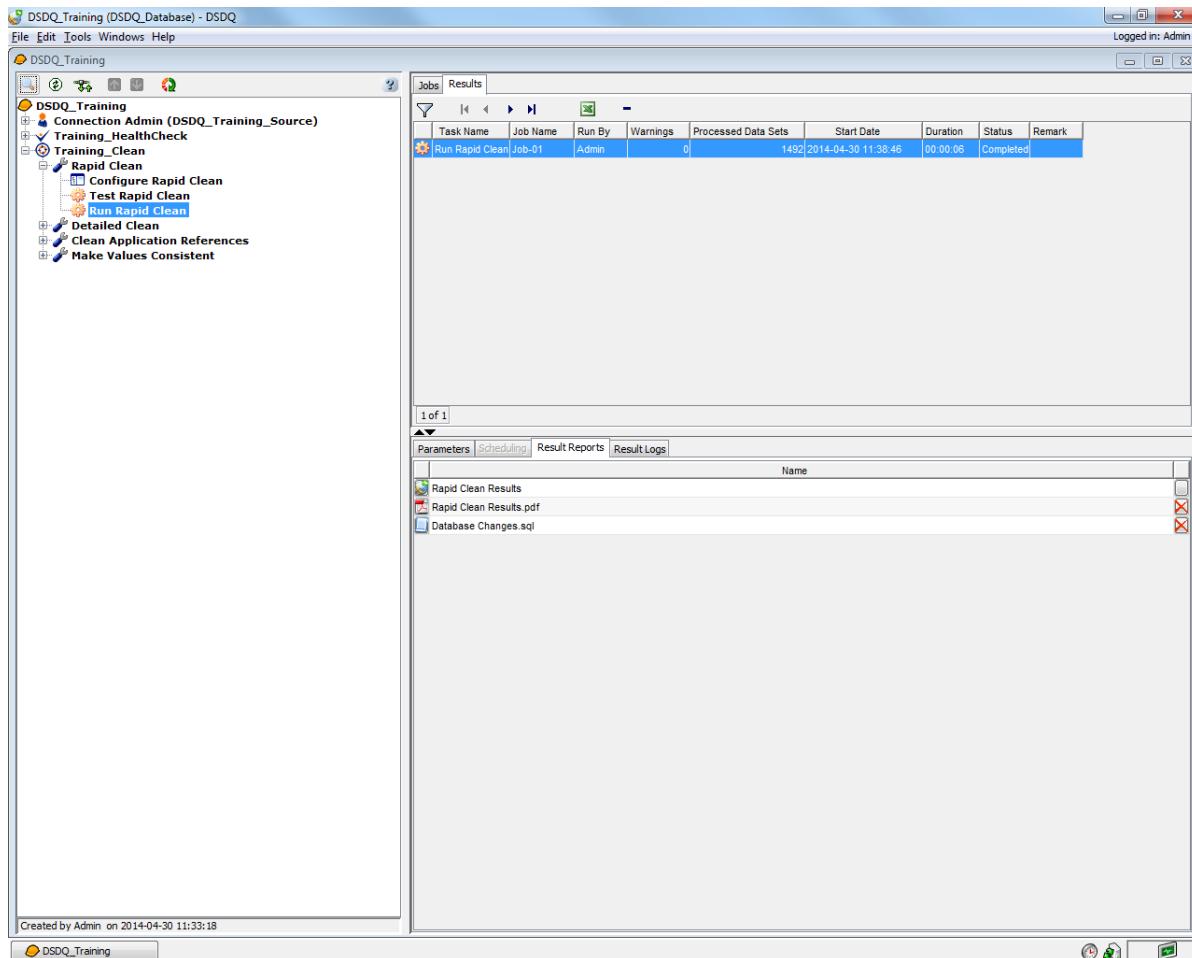
Task Name	Job Name	Description	Created By	Created Date	Updated By	Updated Date
Run Rapid Clean			Admin	2013-12-24 10:30:15		

Below the table, the 'Parameters' tab is active, showing fields for 'Job Name' (empty), 'Job Description' (empty), 'Select Submodel' (set to 'DSDQ_Training'), 'Generate Printable Reports?' (radio button set to 'Yes'), 'Delete Results?' (radio button set to 'After: 30 day(s)'), 'Data Change Action' (radio button set to 'Only Write the Changes to an SQL Script'), and 'Enable Audit Logging' (radio button set to 'No').

2. Enter **Job-01** in the **Job Name** field.
3. Enter **Rapid Clean** in the **Job Description** field.

4. Select **DSDQ_Training** 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 **7**.
 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 **Parameter** tab.
 10. Click .
- The **Rapid Clean** task runs and displays results in the **Result Reports** tab of the **Job and Results Information Pane**.

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



12. Click  on the **Result Reports** tab on the Job and Results Information Pane to display **Rapid Clean Results** in PDF format.

Rapid Clean Results				HALLIBURTON
				Landmark Software & Services
Project:	DSDQ_Training			
Phase:	Training_Clean			
Task:	Run Rapid Clean			
Job:	Job-01			
Connection:	OpenWorks 5000.8.3			
Sub-Model:	OpenWorks Well			
Result Date:	Wed, Apr 30, 2014 11:38			

Table Name:	Well	Number of Changes in This Table:	3259
Column Name	Rule Result	Result %	Result Description
current_status	1492	100	Column values converted to upper case
field	30	2	Column values converted to upper case
state	1492	100	Column values converted to upper case
well_number	131	8	Column values converted to upper case
well_operator	114	7	Column values converted to upper case

Detailed Clean Activity

The **Detailed Clean** Activity gives you the ability to assign columns from selected submodels to clean requirements, test a service level with selected test data and view test results in the **Configure Detailed Clean** Tool.

Exercise: Configuring the Detailed Clean Tool

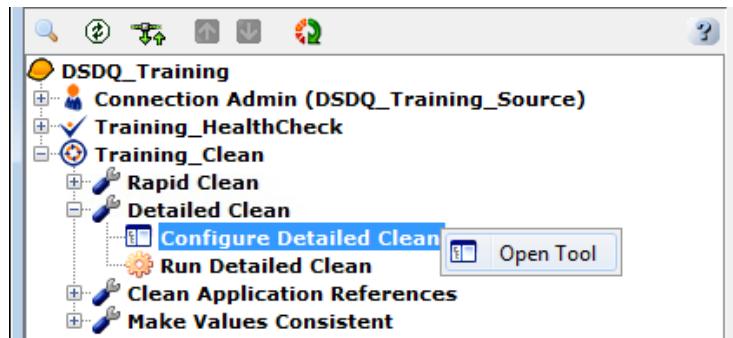
The **Configure Detailed Clean** Tool is used to configure service levels for testing prior to running the **Run Detailed Clean** Task. During this process, you can select which requirements in the service level to enable/disable and when testing a service level, what subset of the total data to use. A service level containing clean requirements must exist prior to opening the **Configure Detailed Clean** Tool. This process requires that you execute the following steps:

- Select a submodel
- Select a service level
- Assign elements to columns
- Add a new service level requirement
- Validate the service level requirement
- Test a service level

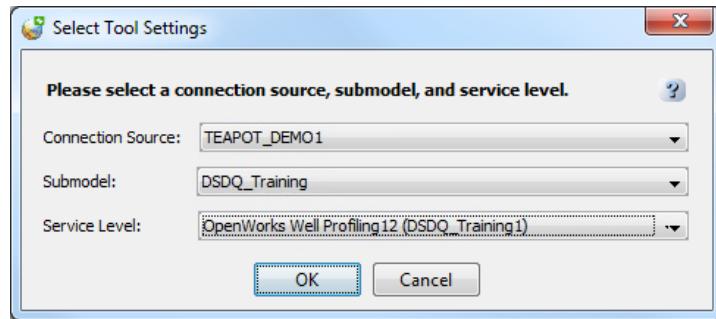
Information about each step is provided in the following section:

1. Click  on the DecisionSpace Data Quality Tree to expand the **Detailed Clean** Activity.
2. Double-click the **Configure Detailed Clean** Tool or right-click the **Configure Detailed Clean** Tool and select **Open Tool** from the

pop-up menu.



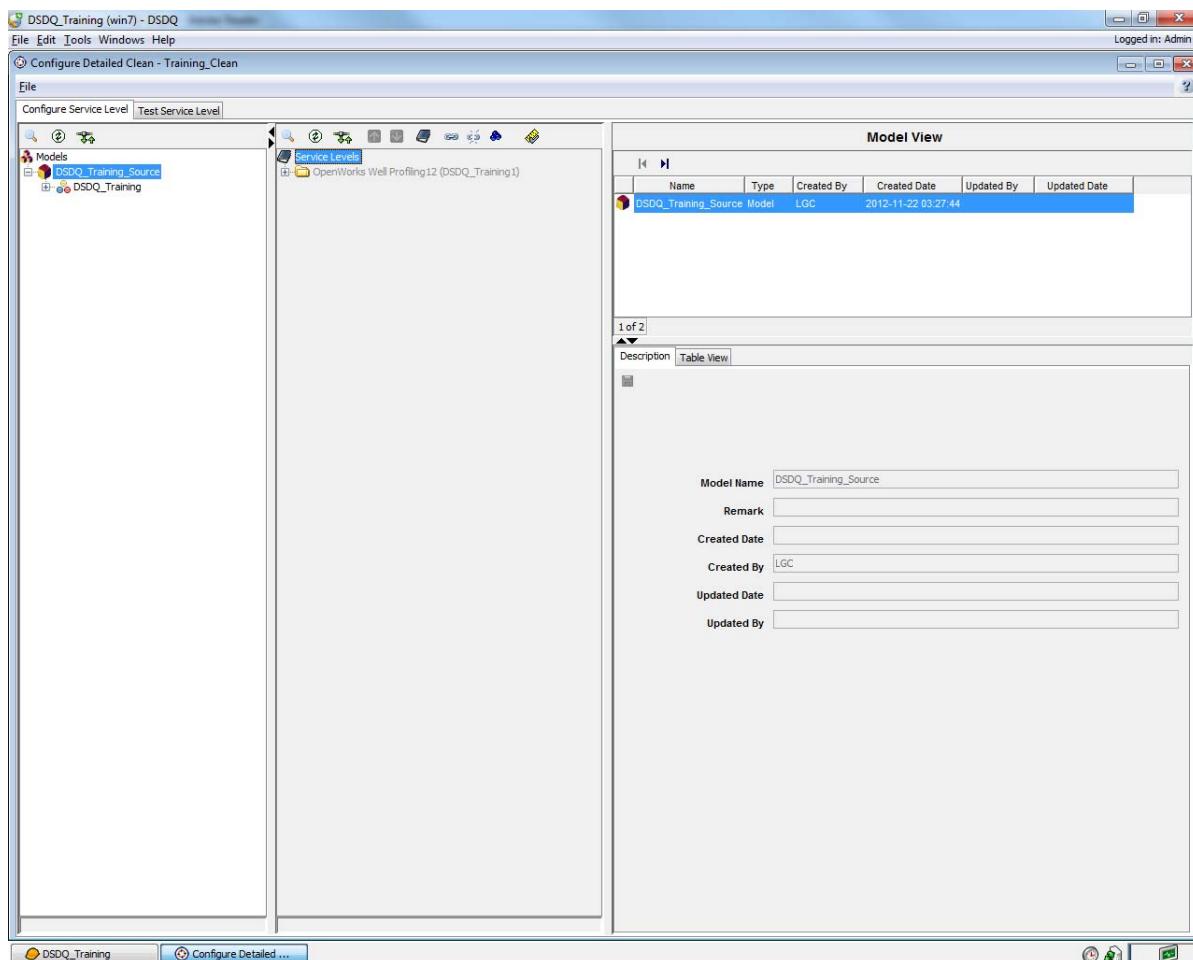
The Select Tool Settings window appears:



3. Select **TEAPOT_DEMO1** from the **Connection Source** drop-down list.
4. Select **DSDQ_Training** from the **Submodel** drop-down list.
5. Select **OpenWorks Well Profiling 12 (DSDQ_Training1)** from the **Service Level** drop-down list.

6. Click **OK**.

The **Configure Detailed Clean** window appears with the selected service level displaying in the Service Level Tree..



Note

The selected service level is remembered when the **Configure Detailed Clean** Tool is closed and is automatically displayed when the Tool is re-opened. Only one service level can be configured at any given time.

7. Click **+** on the Data Model Tree to expand the **DSDQ_Training** submodel.
8. Expand the **WellEntire** Table.
9. Click **+** on the Service Level Tree to expand the **OpenWorks Well Profiling 12 (DSDQ_Training1)** service level.
10. Expand the **Upstream Oil & Gas** sector.

11. Expand the **Well Drilling** area.

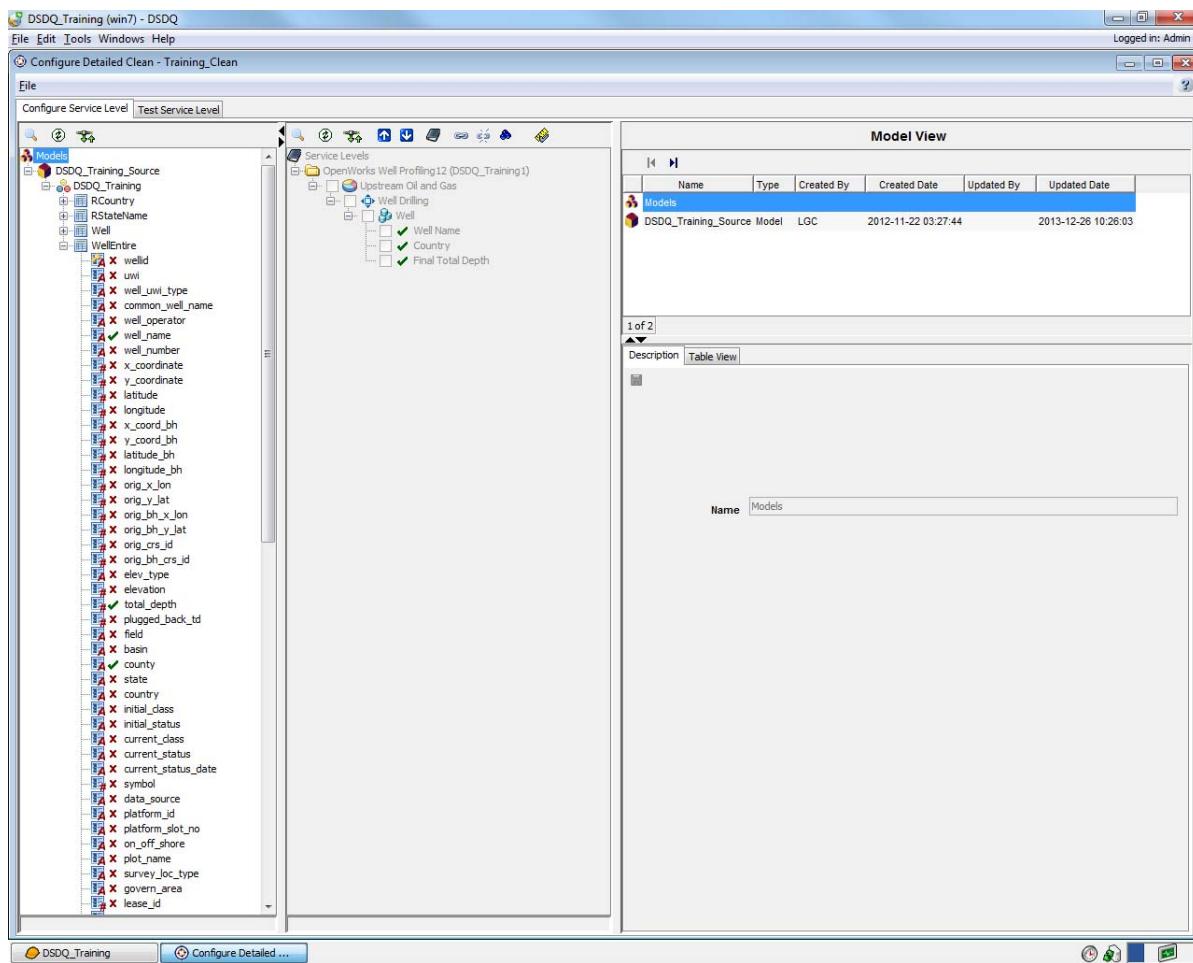
12. Expand the **Well** element group.

13. Drag and drop the **well_name** column from the Data Model Tree to the **Well Name** Element in the Service Level Tree.

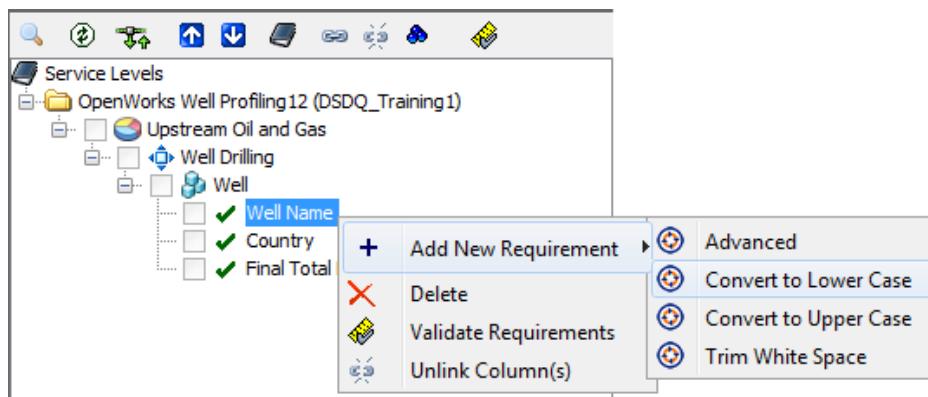
A green check mark will appear adjacent to the column and element that have just been associated. Alternatively you can select the column & element to link and click the **Link Column to Element** button on the Service Level Tree toolbar. Only one column from the same table can be linked to the same element. However, it is possible to link many columns to the same element if the columns come from different tables.

14. Drag and drop the **country** column from the Data Model Tree to the **Country** element in the Service Level Tree.

15. Drag and drop the **total_depth** column from the Data Model Tree to the **Final Total Depth** element in the Service Level Tree.



16. Right-click the **Well Name** element in the Service Level Tree and select **Add New Requirement > Convert to Lower Case** from the pop-up menu.

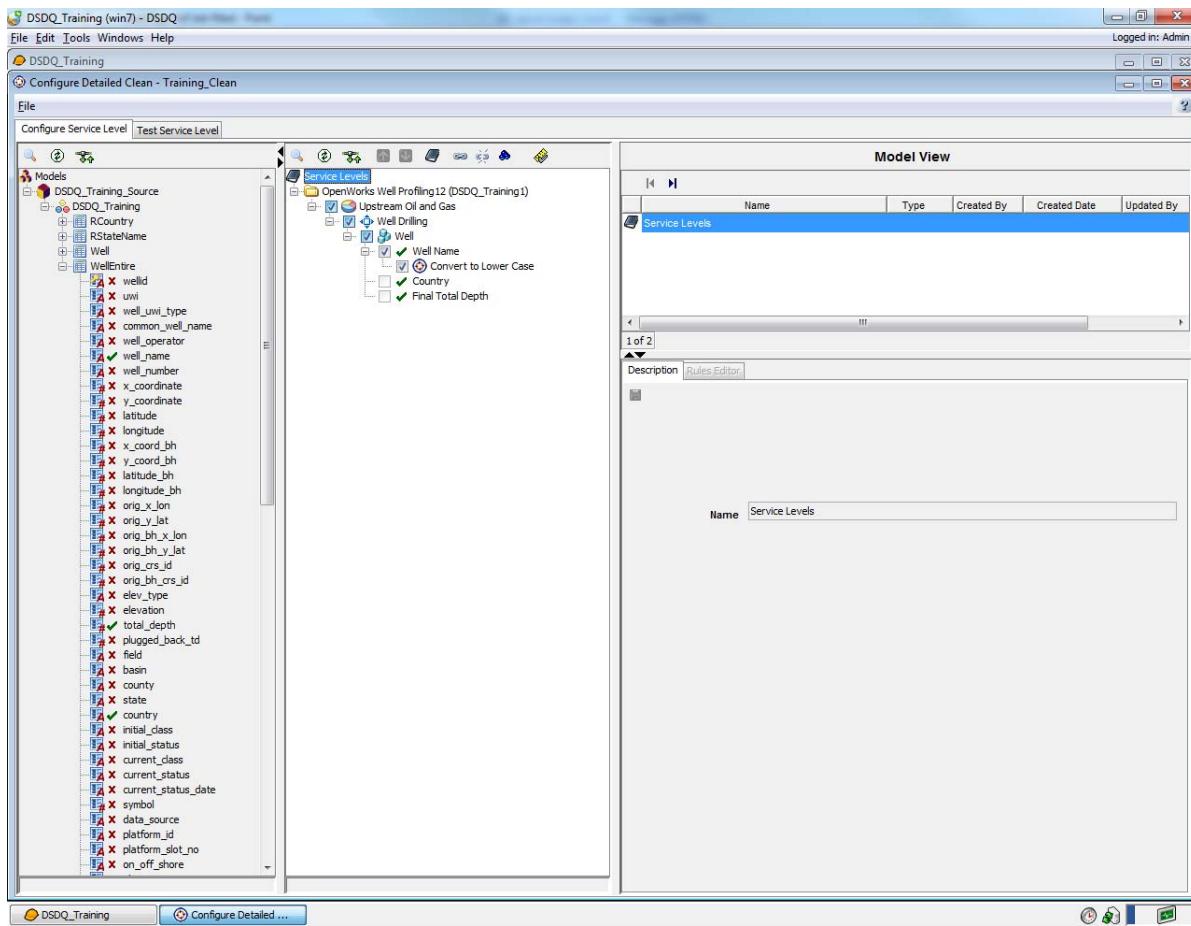


The **Enter Name** dialog box appears.

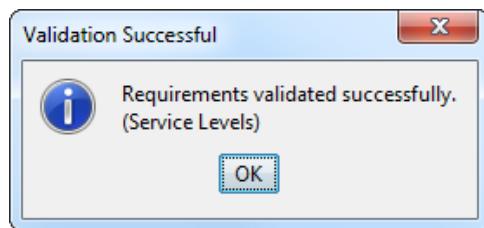


17. Optionally, specify a user-defined name for the requirement.

18. Click **OK** to add the requirement to the selected element.

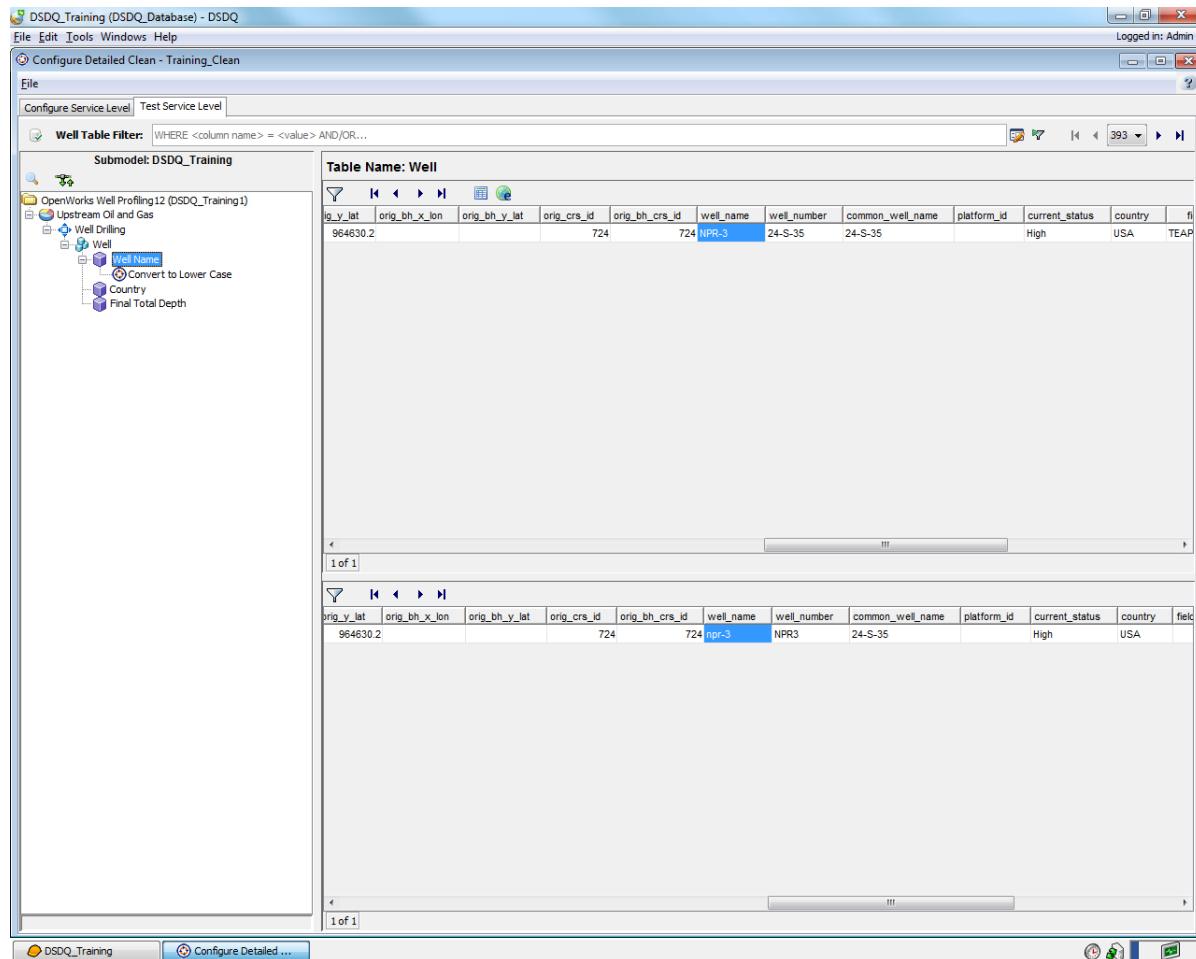


19. Select the **Convert to Lower Case** requirement and click on the Service Level Tree toolbar to validate the requirement. A **Validation Successful** message is displayed once the selected requirement has been validated.



20. You can test a service level by selecting the **Test Service Level** tab. The service level test is automatically executed. Results are

displayed on the top and bottom panes on the right side of the window.

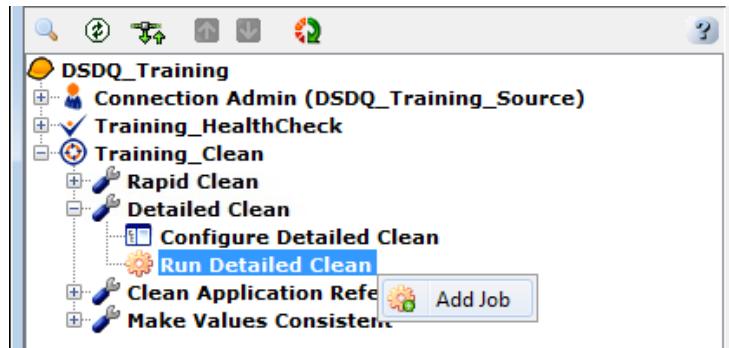


21. Select **File > Exit** from the menu bar on the **Configure Detailed Clean** window.

Exercise: Running the Detailed Clean Task

To run the Detailed Clean Task:

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



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

The screenshot shows the DSDQ interface with the 'DSDQ_Training' project selected. The 'Jobs' tab is active in the central pane, displaying a table with one row:

Task Name	Job Name	Description	Created By	Created Date	Updated By	Updated Date
Run Detailed Clean			Admin	2014-04-30 12:58:23		

Below the table, there is a configuration panel for the job:

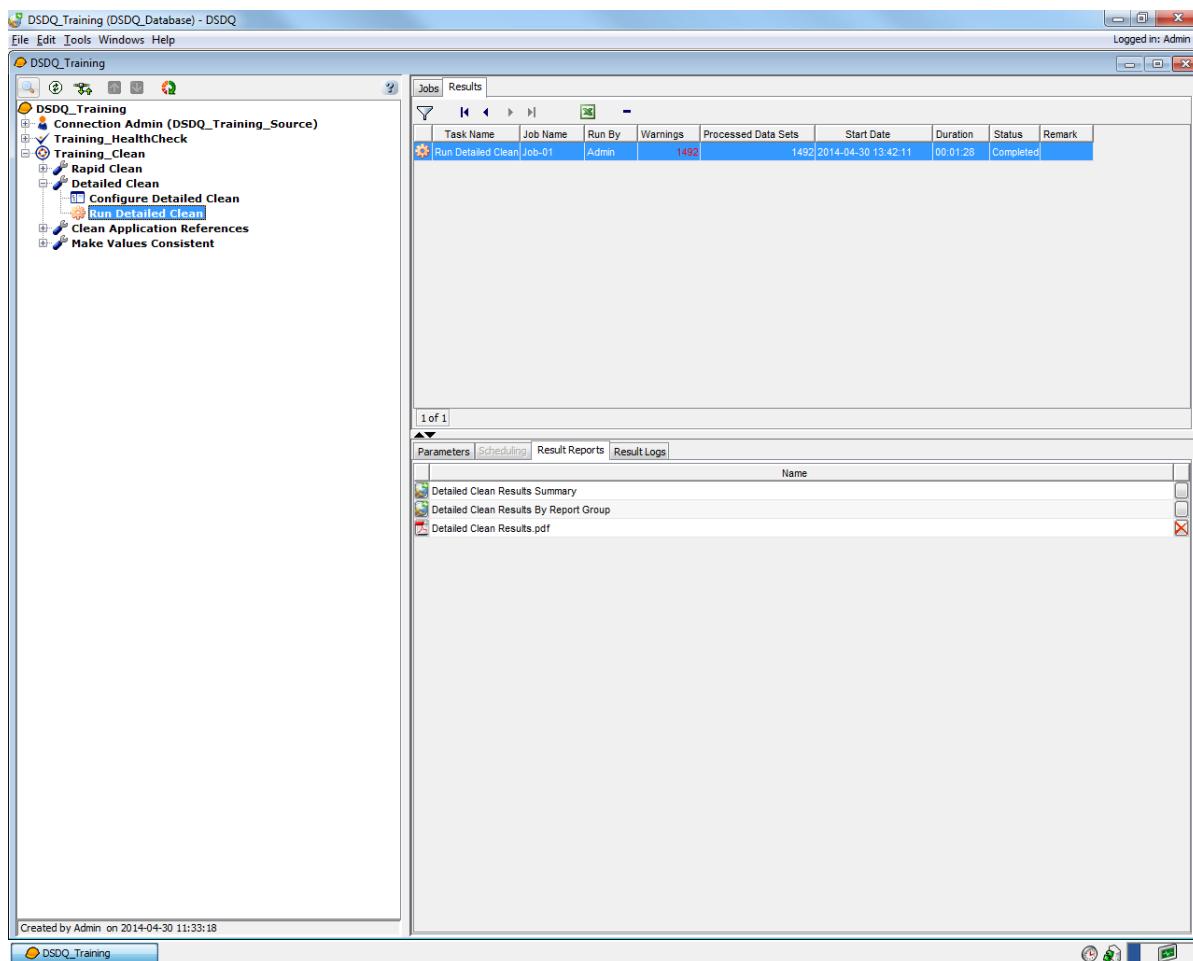
- Job Name:**
- Job Description:**
- Service Level:** OpenWorks Well Profiling 12 (DSDQ_Training)
- Select a Submodel:** DSDQ_Training
- Filter Base Data Set:**
- Summarize Results by:** Requirement (least detail) Data Set Data Row (most detail)
- Generate Printable Reports?** Yes No
- Delete Results?** Never After: 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

At the bottom left of the interface, a status bar shows: Created by Admin on 2014-04-29 15:37:34

2. Enter **Job-01** in the **Job Name** field.

3. Enter **Detailed Clean** in the **Job Description** field.
 4. Select **OpenWorks Well Profiling 12 (DSDQ_Training1)** from the **Service Level** drop-down list.
 5. Select **DSDQ_Training** from the **Select Submodel** drop-down list.
 6. Optionally, set a filter on the data subset.
 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 **7**.
 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 .
- The **Detailed Clean** task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information** Pane.

14. Select the **Results** tab on the **Jobs and Results Listing** Pane to view the values in the **Results Report** tab on the **Job and Results Information** Pane.



15. Click on the **Results Reports** tab to display **Detailed Clean Results** in PDF format.

Detailed Clean Results				HALLIBURTON Landmark Software & Services
Element Group:	Element	Rule Result	Result %	Number of Changes in This Group:
Well	Well Name	1385	99	
	Well Name	26	1	
	Well Name	26	1	
	Country	1399	100	
	Well Name	0	Well name values that had quotes removed	
	Well Name	0	Well name values that had quotes removed	
	Well Name	1389	99	
	Well Common Name	1346	96	

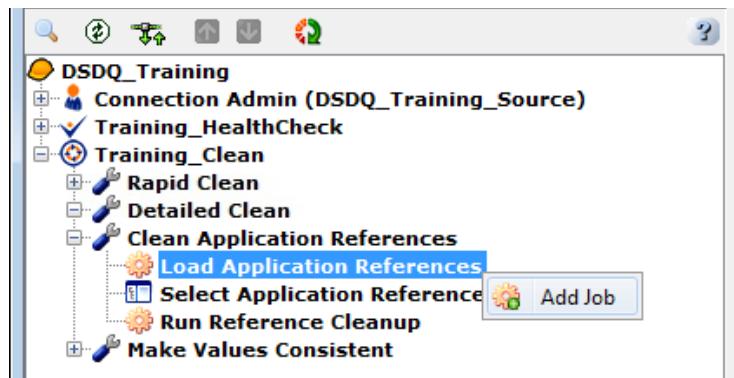
Cleaning Application References

The Clean Application References Activity enables you to clean up data references to the application reference tables. In order to do so, you will load the current values in the application reference tables followed by configuring values that are correct for your data and application reference tables.

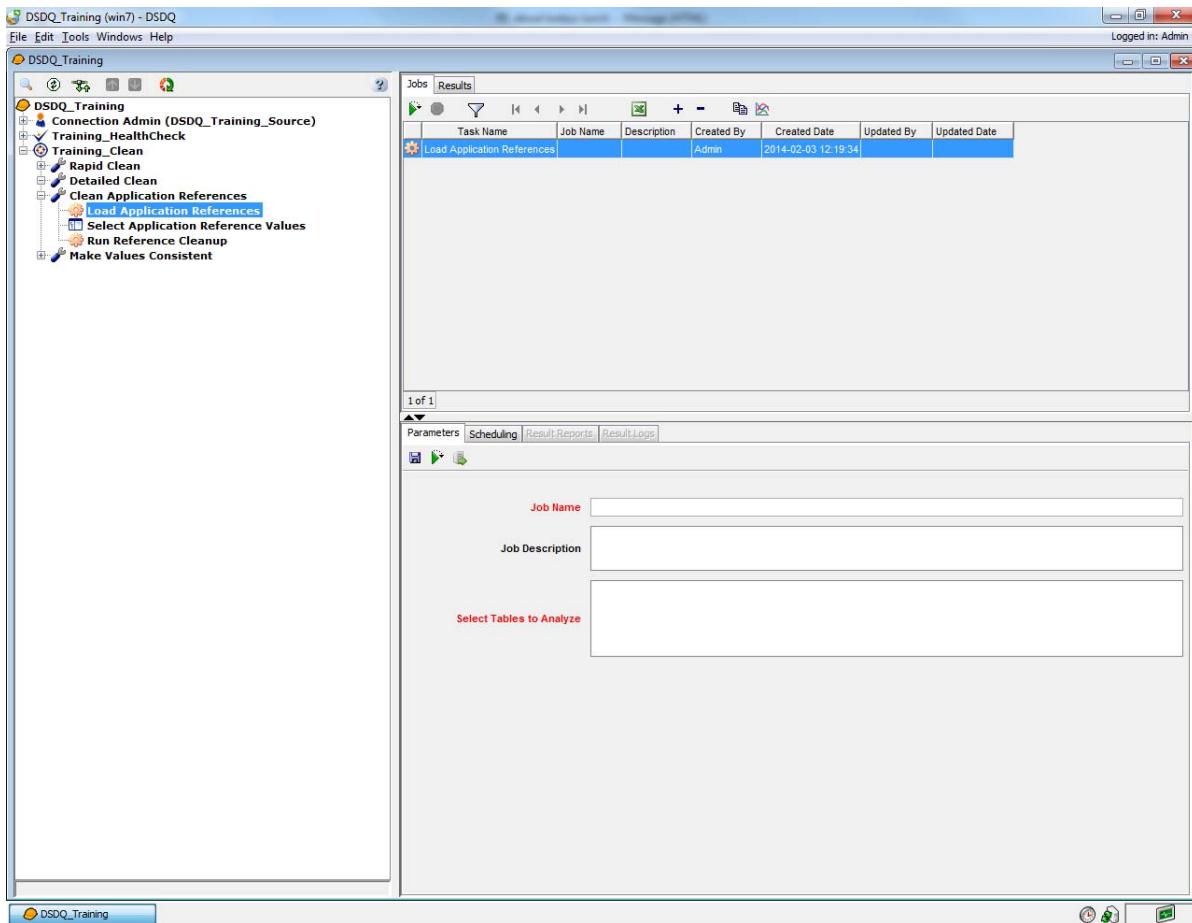
Exercise: Loading Application References

The **Load Application Reference** task loads the current values in the application reference tables that have been configured. To Load Application References:

1. Click  on the DecisionSpace Data Quality Tree to expand the **Clean Application References** Activity.
2. Double-click the **Load Application References** Task or right-click the **Load Application References** Task and select **Add Job** from the pop-up menu.

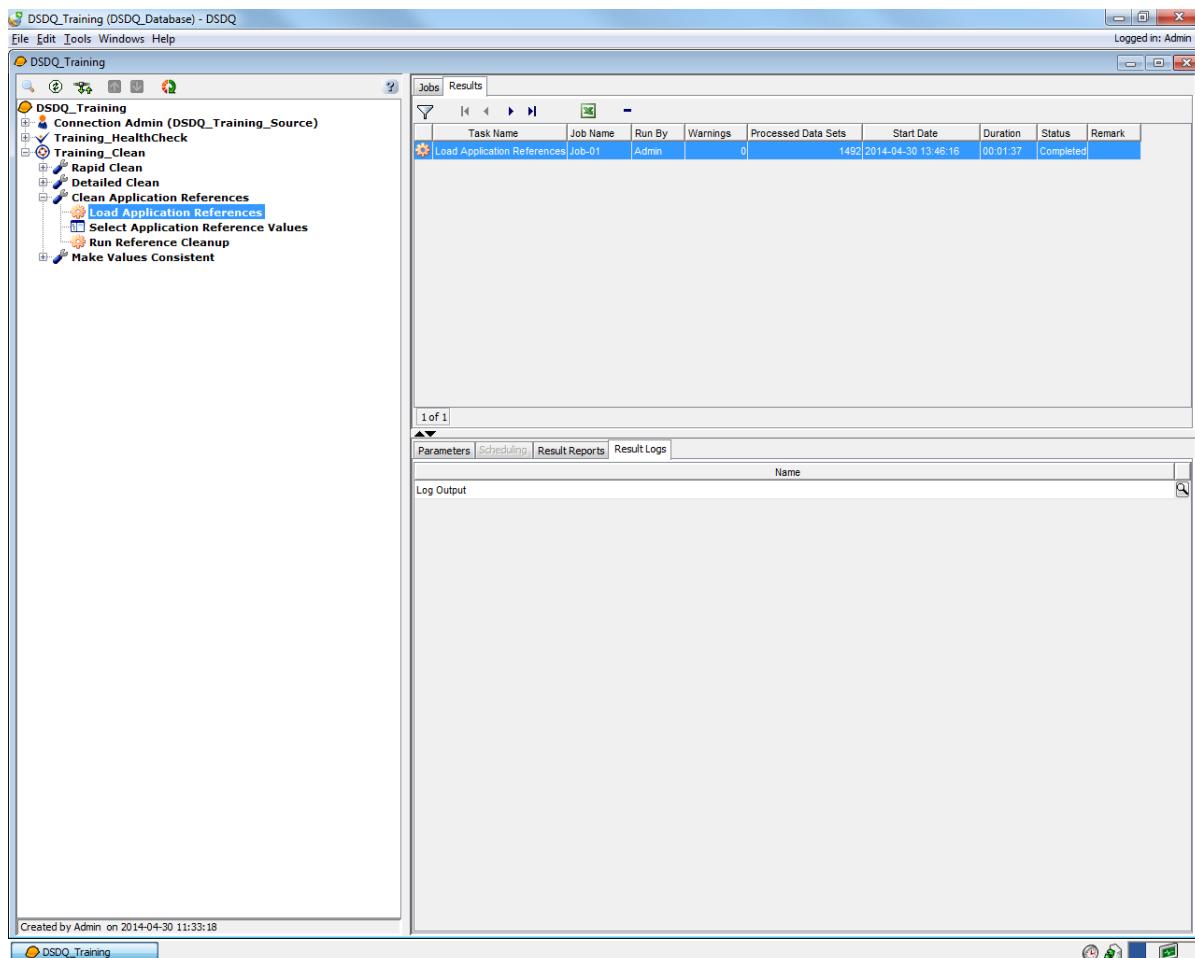


A new job is initiated and displays on the **Jobs and Results Listing Pane**.



3. Enter **Job-01** in the **Job Name** field.
 4. Enter **Application References** in the **Job Description** field.
 5. Select **RCountry** from the **Select Tables to analyze** list box.
 6. Click to save changes in the **Parameter** tab.
 7. Click .
- The **Load Application References** Task is executed and displays in the **Result Reports** tab on the Jobs and **Results Information** Pane.

8. Select the **Results** tab on the **Job and Results Listing** Pane and then the **Result Logs** tab on the **Jobs and Results Information Pane**.



- Double-click **Log Output** on the **Result Logs** tab to view the loaded references.

```

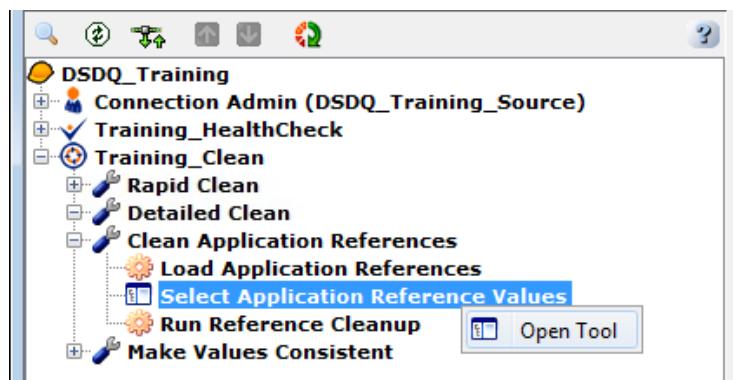
Log Output - Notepad
File Edit Format View Help
2013-12-26 15:49:55,919 INFO Processed data set #1
2013-12-26 15:49:55,929 INFO Processed data set #2
2013-12-26 15:49:55,939 INFO Processed data set #3
2013-12-26 15:49:55,948 INFO Processed data set #4
2013-12-26 15:49:55,957 INFO Processed data set #5
2013-12-26 15:49:55,968 INFO Processed data set #6
2013-12-26 15:49:55,978 INFO Processed data set #7
2013-12-26 15:49:55,987 INFO Processed data set #8
2013-12-26 15:49:55,997 INFO Processed data set #9
2013-12-26 15:49:56,006 INFO Processed data set #10
2013-12-26 15:49:56,015 INFO Processed data set #11
2013-12-26 15:49:56,024 INFO Processed data set #12
2013-12-26 15:49:56,032 INFO Processed data set #13
2013-12-26 15:49:56,039 INFO Processed data set #14
2013-12-26 15:49:56,046 INFO Processed data set #15
2013-12-26 15:49:56,058 INFO Processed data set #16
2013-12-26 15:49:56,066 INFO Processed data set #17
2013-12-26 15:49:56,075 INFO Processed data set #18
2013-12-26 15:49:56,083 INFO Processed data set #19
2013-12-26 15:49:56,092 INFO Processed data set #20
2013-12-26 15:49:56,101 INFO Processed data set #21
2013-12-26 15:49:56,110 INFO Processed data set #22

```

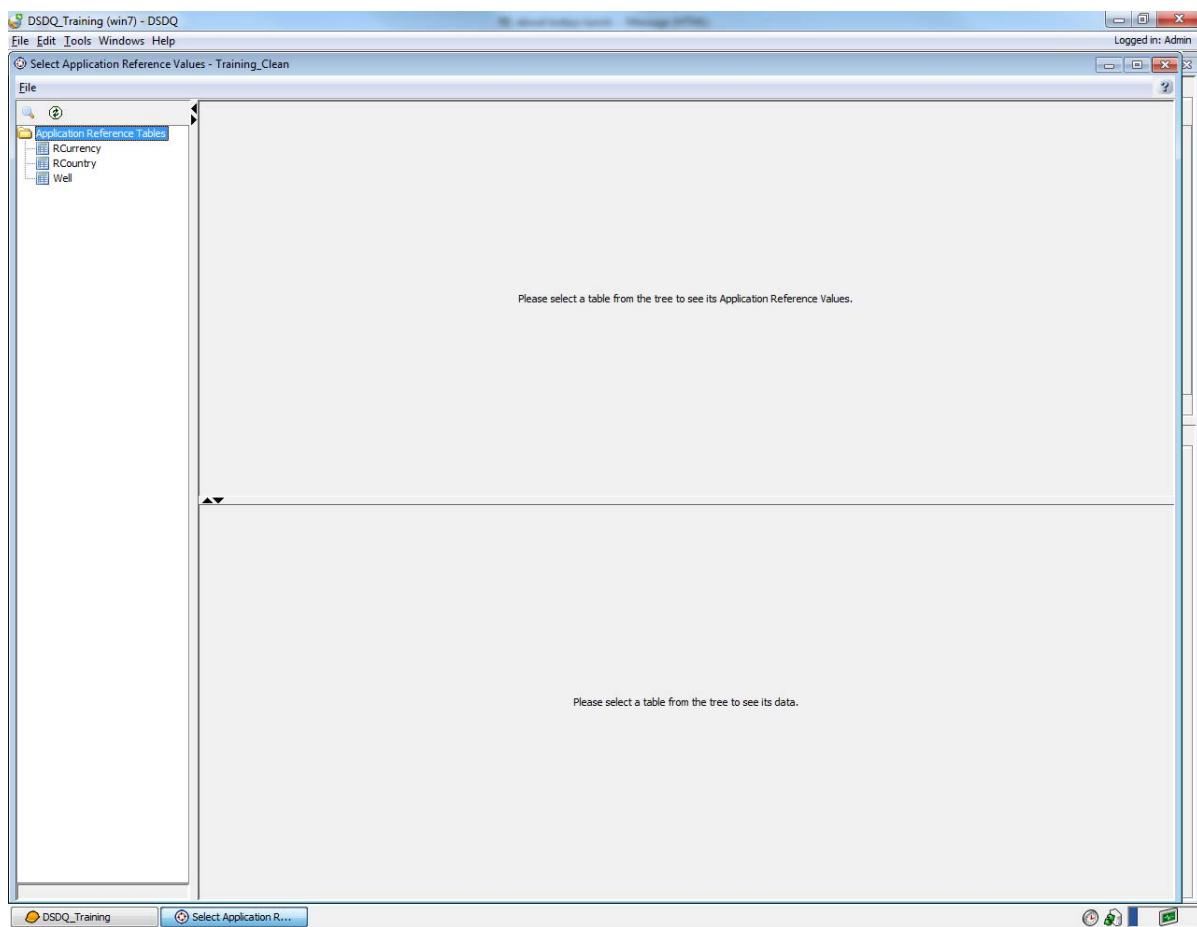
Exercise: Selecting Application Reference Values

The **Select Application Reference Values** Tool allows you to select and configure the values that are correct for your data and application reference tables. To select application reference values:

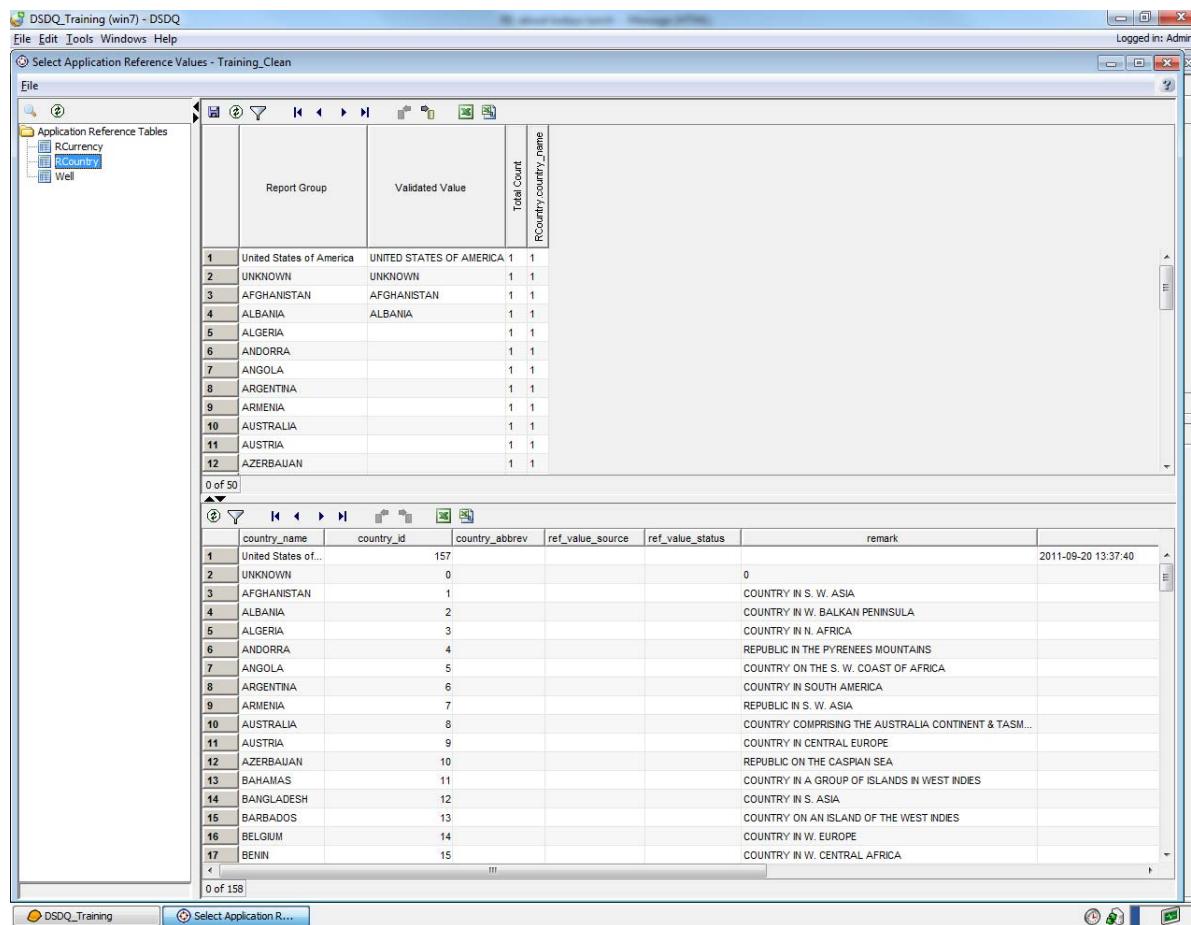
- Double-click the **Select Application Reference Values** Tool or right-click the **Select Application Reference Values** Tool and select **Open Tool** from the pop-up menu.



The **Select Application Reference Values** window appears.



2. Select **RCountry** table from the **Application Reference Table** to view the results of the table in the top right table.



Note

Table and column headers specify the application reference that was configured and the Total Count of matching values found. Selecting a result in the top right table displays the application reference in the bottom table. The bottom table displays the application reference values for the selected table.

3. To change a value, enter the correct value for application references in the **Validated Value** column. Only those values that exist in the application reference table can be entered here.
4. Click to save changes made to the reference table.

5. Select **File > Exit** from the menu bar on the **Select Application Reference Values** window.

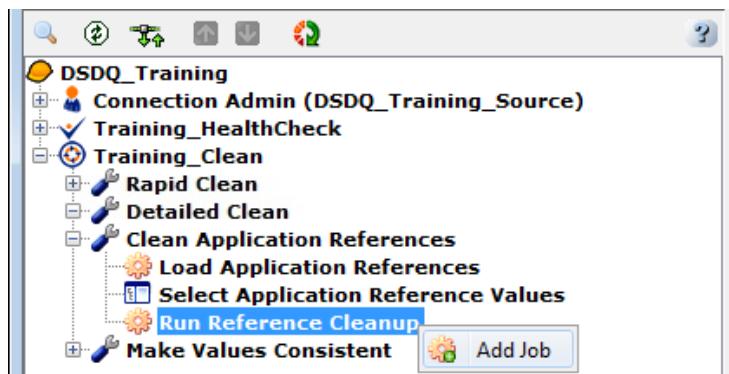
Note

The application reference tables have to be configured with at least one **Primary Key**. To view data from a specific column, configure it as a **Report Column** in the **Perform Table Modeling Tool**.

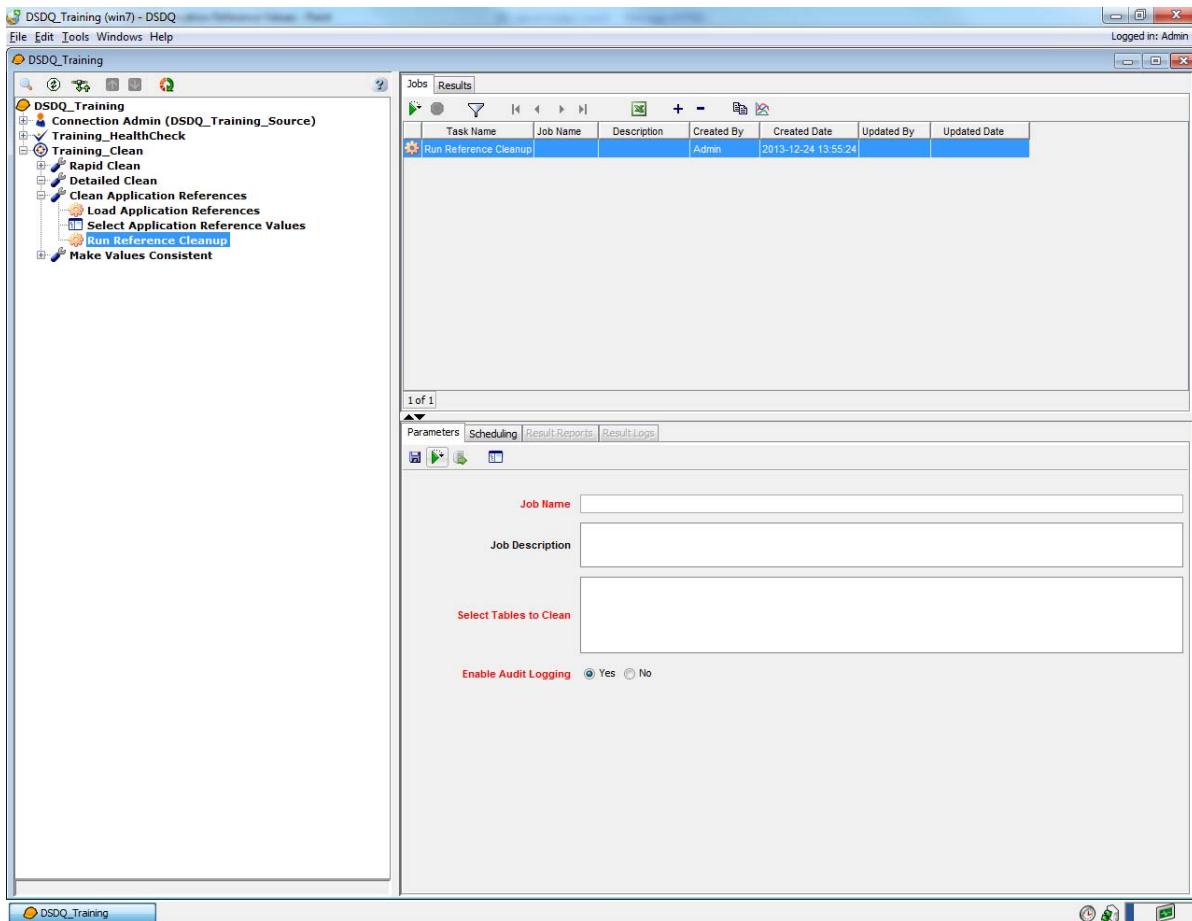
Exercise: Running Reference Cleanup

The **Run Reference Cleanup** task modifies the values as per the configuration in the **Select Application Reference Values** Tool.

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



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



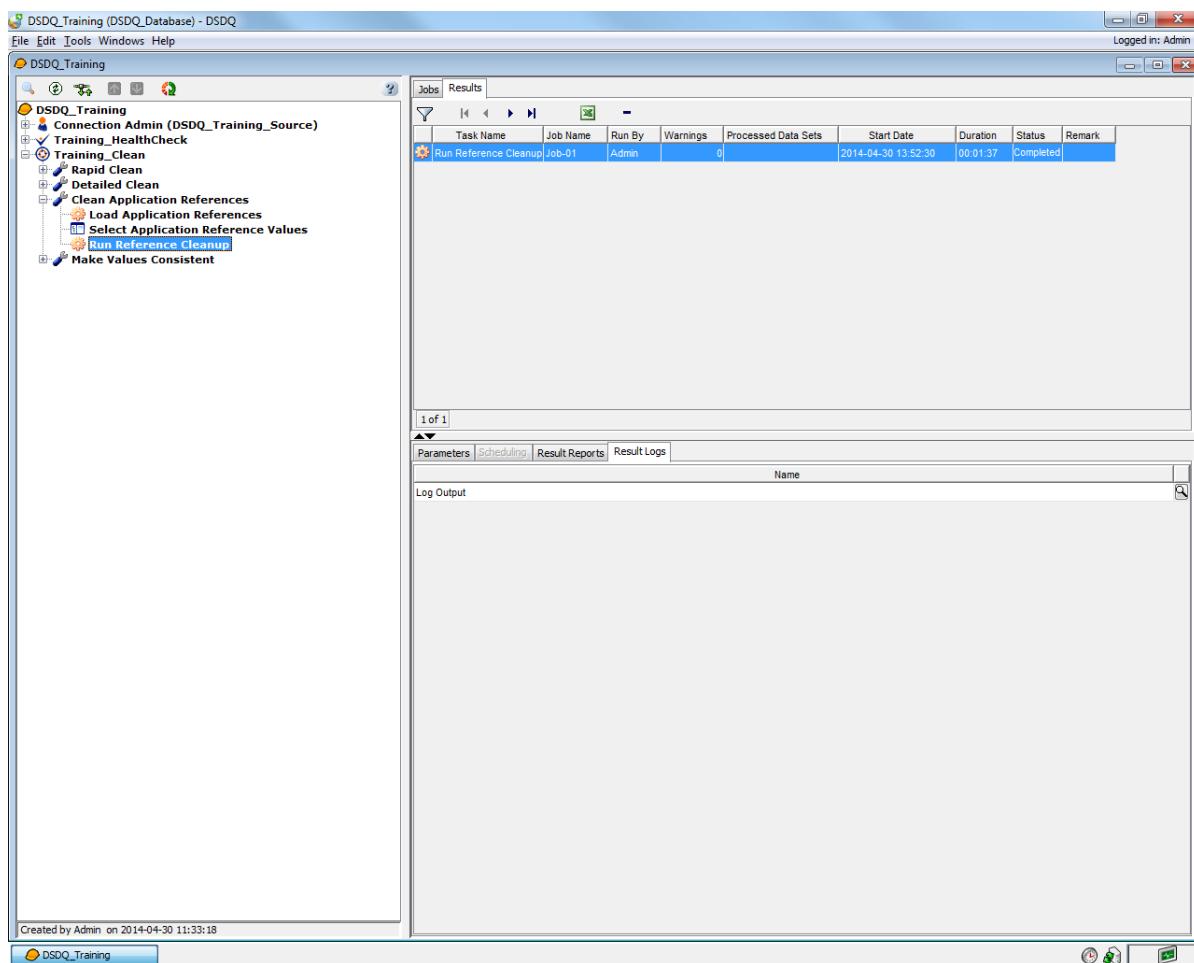
Note

The job creation is initiated only the first time the task is run. To add an additional job, right-click the **Run Reference Cleanup** Task and select the **Add Job** option from the pop-up menu, or click the **Add New Job**  button on the Jobs toolbar.

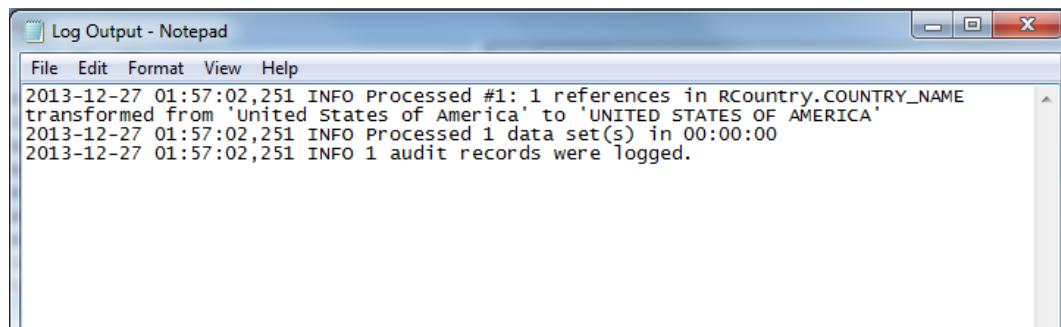
2. Enter **Job-01** in the **Job Name** field.
3. Enter **Reference Cleanup** in the **Job Description** field.
4. Select **RCountry** in the **Select Tables to Clean** field.
5. Select the **Yes** option for **Enable Audit Logging**.
6. Click  to save changes in the **Parameter** tab.

7. Click .

The **Run Reference Cleanup** task is executed and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

8. Select the **Results** tab on the **Job and Results Listing Pane** and then the **Result Logs** tab on the **Jobs and Results Information Pane**.

9. Double-click **Log Output** on the **Result Logs** tab to view results of the Run Reference Cleanup task.



Making Values Consistent

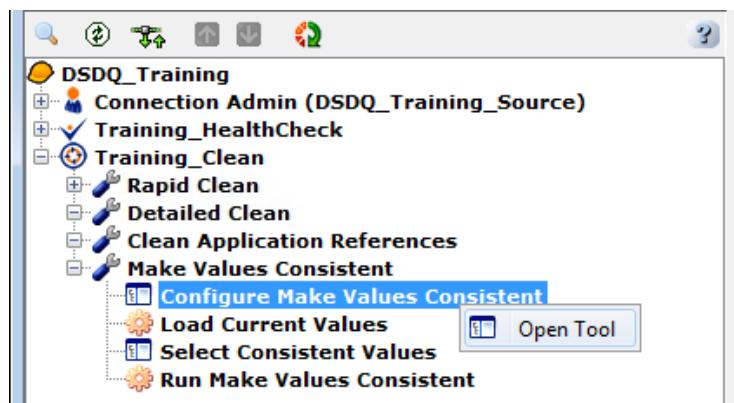
The Make Values Consistent tool is used to assign elements that have Make Values Consistent requirements to associated columns. In order to do so, you will load current values for this requirement, select consistent values (manually or by selecting these from reference tables) and finally update the data with the new values.

Exercise: Configuring Make Values Consistent

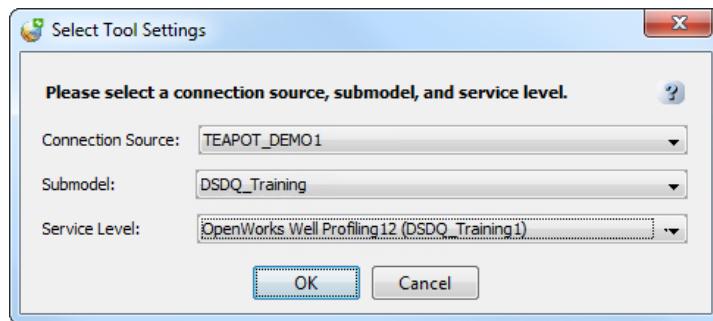
The **Configure Make Values Consistent** Tool is used to configure service levels with make values consistent requirements for testing prior to running the actual **Run Make Values Consistent** Task. The user can select which requirements in the service level to enable/disable when testing. A service level containing make values consistent requirements must exist prior to opening the **Configure Make Values Consistent** Tool. This process requires that you execute the following steps:

- Select a submodel
 - Select a service level
 - Assign elements to columns
 - Add a new service level requirement
1. Click  on the DecisionSpace Data Quality Tree to expand the **Make Values Consistent** Activity.
 2. Double-click the **Configure Make Values Consistent** Task or right-click the **Configure Make Values Consistent** Task and select

Open Tool from the pop-up menu.



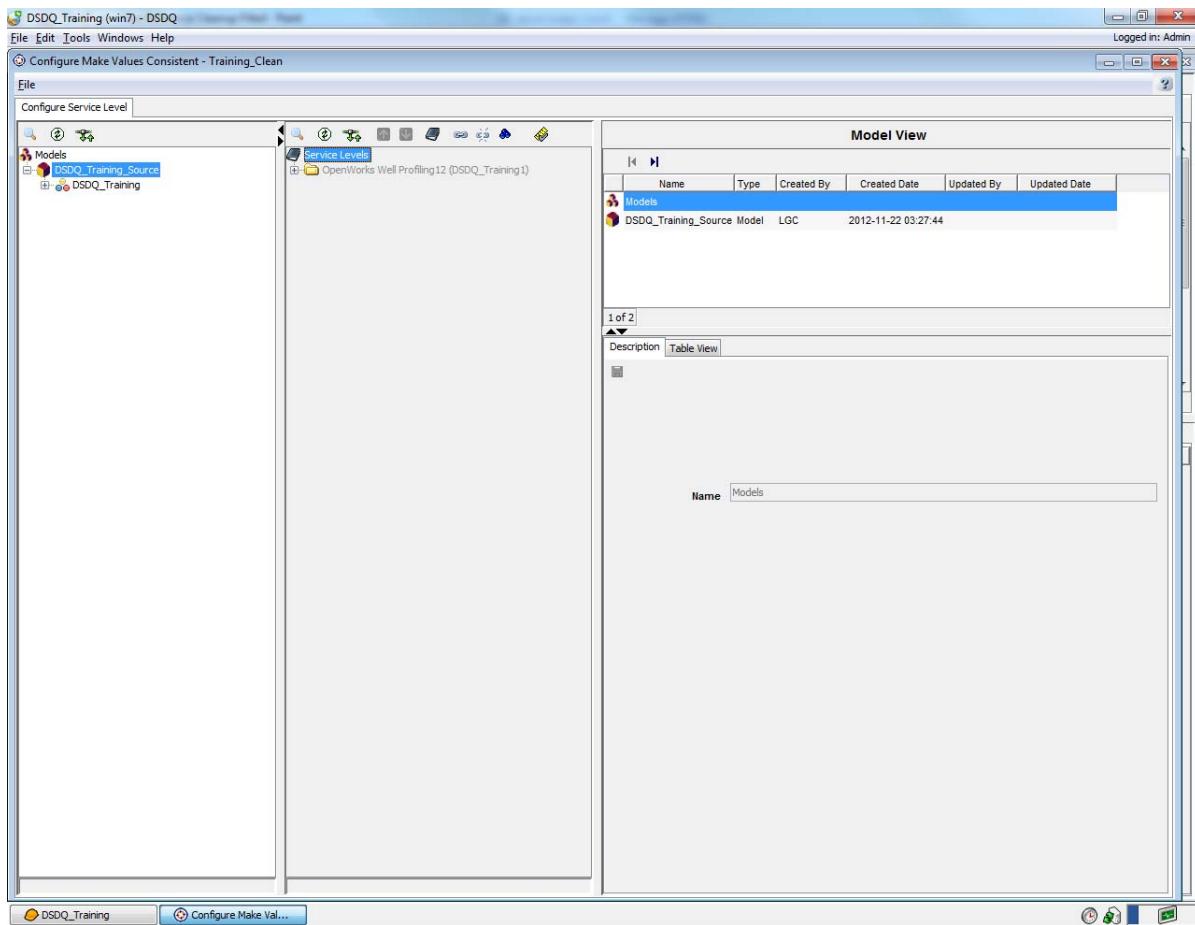
The Select Tool Settings dialog box appears:



3. Select **TEPOAT_DEMO1** from the **Connection Source** drop-down list.
4. Select **DSDQ_Training** from the **Submodel** drop-down list.
5. Select **OpenWorks Well Profiling 12 (DSDQ_Training 1)** from the **Service Level** drop-down list.

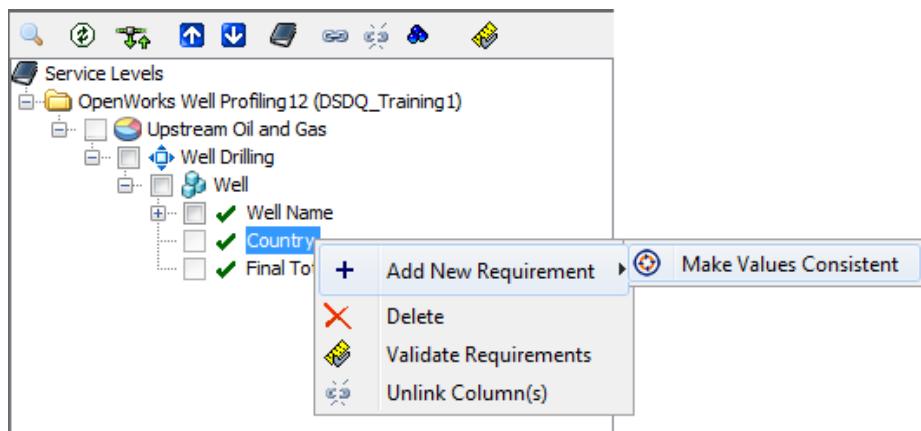
6. Click **OK**.

The **Configure Make Values Consistent** window appears with the selected service level displaying in the Service Level Tree.

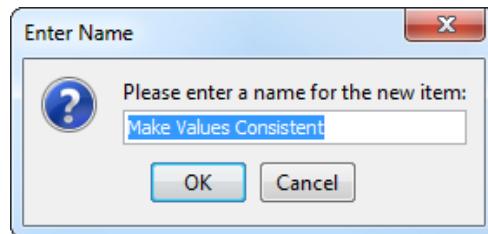


7. Click **+** on the Data Model Tree to expand the **DSDQ_Training** Submodel.
8. Expand the **RCountry** table.
9. Click **+** on the Service Level Tree to expand the **OpenWorks Well Profiling 12 (DSDQ_Training1)** service level.
10. Expand the **Upstream Oil & Gas** sector.
11. Expand the **Well Drilling** area.
12. Expand the **Well** element group.
13. Right-click the **Country** element on the Service Level Tree and select **Add New Requirement > Make Values Consistent** from

the pop-up menu.



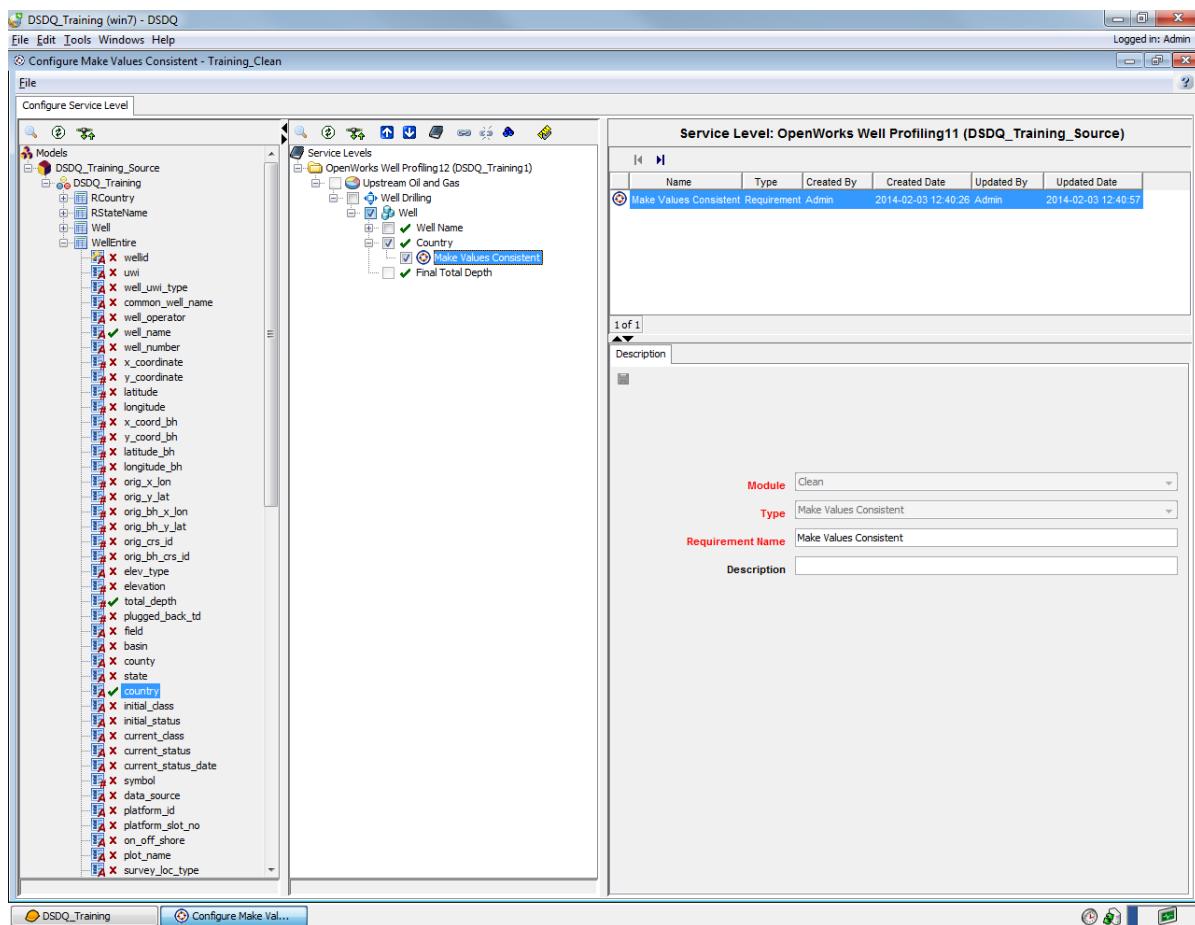
The **Enter Name** dialog box appears.



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

15. Click OK.

The requirement is added and displays in the Service Level Tree.

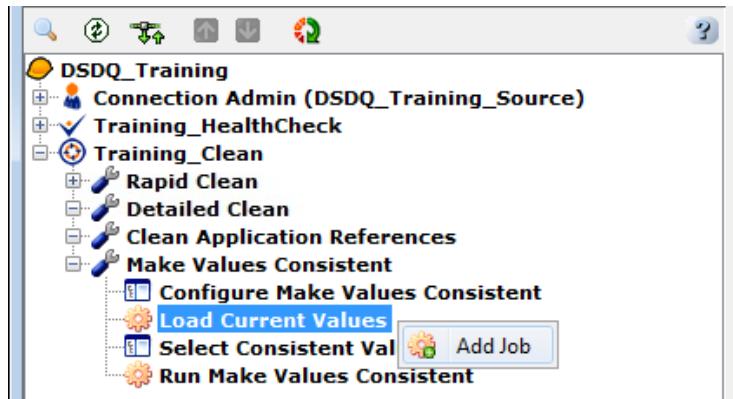


16. Select File > Exit from the menu bar on the Configure Make Values Consistent window.

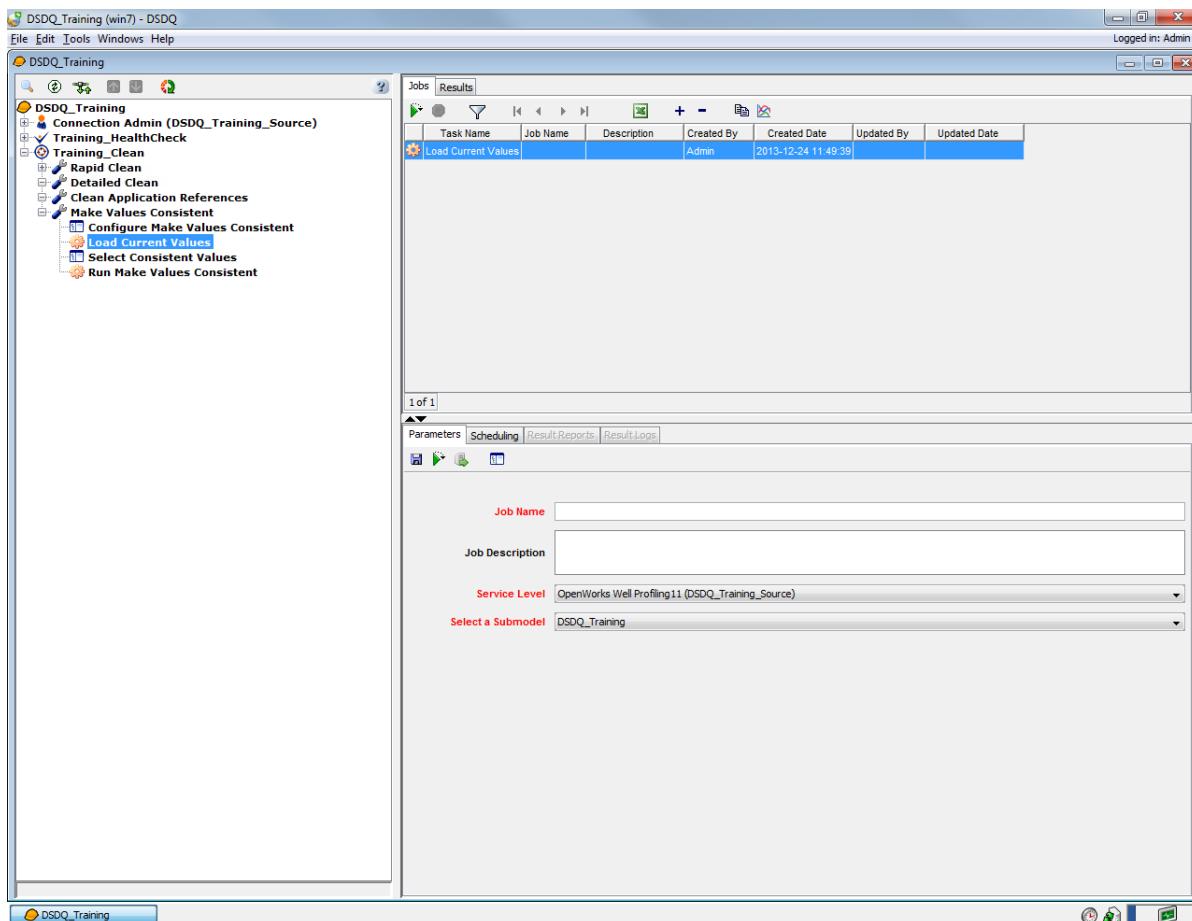
Exercise: Loading Current Values

To load current values in columns assigned to make values consistent requirements:

1. Double-click the **Load Current Values** Task or right-click the **Load Current Values** Task and select **Add Job** from the pop-up menu.

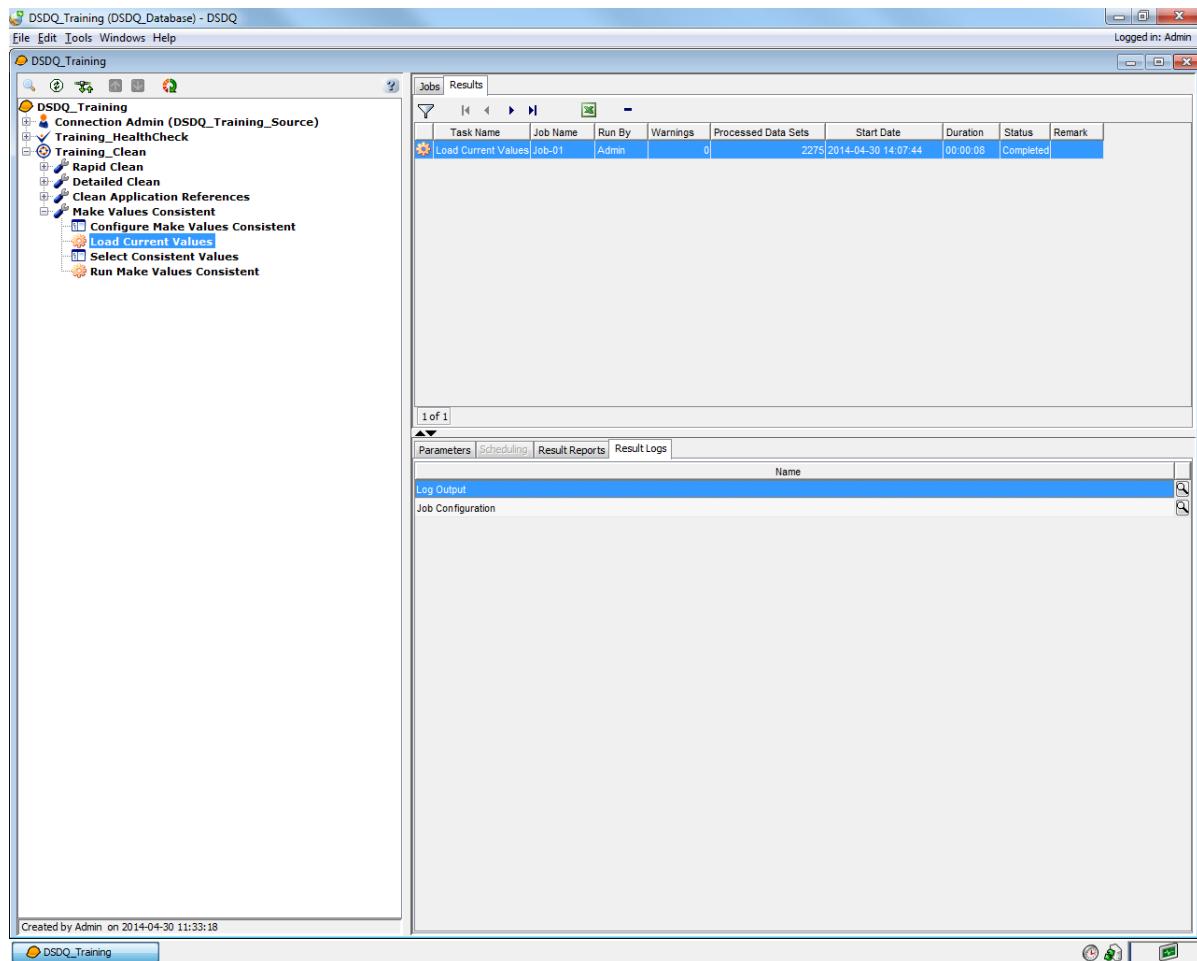


A new job is initiated and displays on the **Jobs and Results Listing Pane**.

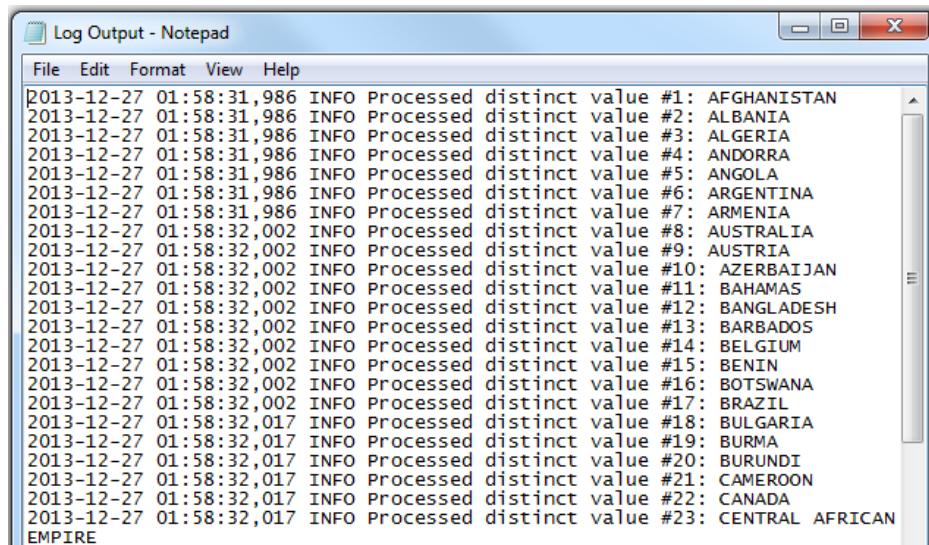


2. Enter **Job-01** in the **Job Name** field.
 3. Enter **Load Current Values** in the **Job Description** field.
 4. Select **OpenWorks Well Profiling12 (DSDQ_Training1)** from the **Service Level** drop-down list.
 5. Select **DSDQ_Training** from the **Select a Submodel** drop-down list.
 6. Click to save changes in the **Parameter** tab.
 7. Click .
- The **Load Current Values** task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

8. 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.



9. Double-click **Log Output** on the **Result Logs** tab to view results of the Run Reference Cleanup task.



```

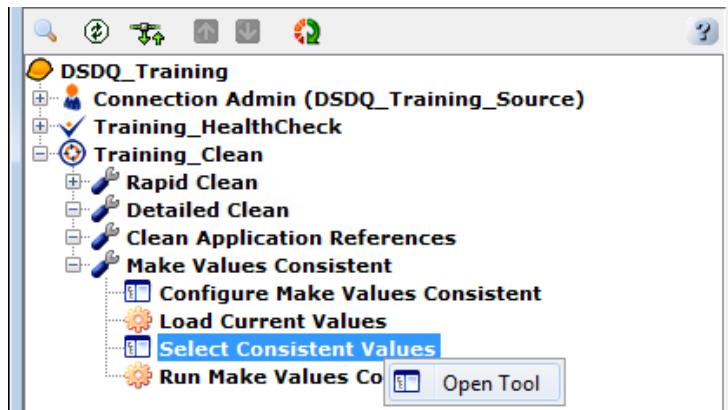
2013-12-27 01:58:31,986 INFO Processed distinct value #1: AFGHANISTAN
2013-12-27 01:58:31,986 INFO Processed distinct value #2: ALBANIA
2013-12-27 01:58:31,986 INFO Processed distinct value #3: ALGERIA
2013-12-27 01:58:31,986 INFO Processed distinct value #4: ANDORRA
2013-12-27 01:58:31,986 INFO Processed distinct value #5: ANGOLA
2013-12-27 01:58:31,986 INFO Processed distinct value #6: ARGENTINA
2013-12-27 01:58:31,986 INFO Processed distinct value #7: ARMENIA
2013-12-27 01:58:32,002 INFO Processed distinct value #8: AUSTRALIA
2013-12-27 01:58:32,002 INFO Processed distinct value #9: AUSTRIA
2013-12-27 01:58:32,002 INFO Processed distinct value #10: AZERBAIJAN
2013-12-27 01:58:32,002 INFO Processed distinct value #11: BAHAMAS
2013-12-27 01:58:32,002 INFO Processed distinct value #12: BANGLADESH
2013-12-27 01:58:32,002 INFO Processed distinct value #13: BARBADOS
2013-12-27 01:58:32,002 INFO Processed distinct value #14: BELGIUM
2013-12-27 01:58:32,002 INFO Processed distinct value #15: BENIN
2013-12-27 01:58:32,002 INFO Processed distinct value #16: BOTSWANA
2013-12-27 01:58:32,002 INFO Processed distinct value #17: BRAZIL
2013-12-27 01:58:32,017 INFO Processed distinct value #18: BULGARIA
2013-12-27 01:58:32,017 INFO Processed distinct value #19: BURMA
2013-12-27 01:58:32,017 INFO Processed distinct value #20: BURUNDI
2013-12-27 01:58:32,017 INFO Processed distinct value #21: CAMEROON
2013-12-27 01:58:32,017 INFO Processed distinct value #22: CANADA
2013-12-27 01:58:32,017 INFO Processed distinct value #23: CENTRAL AFRICAN EMPIRE

```

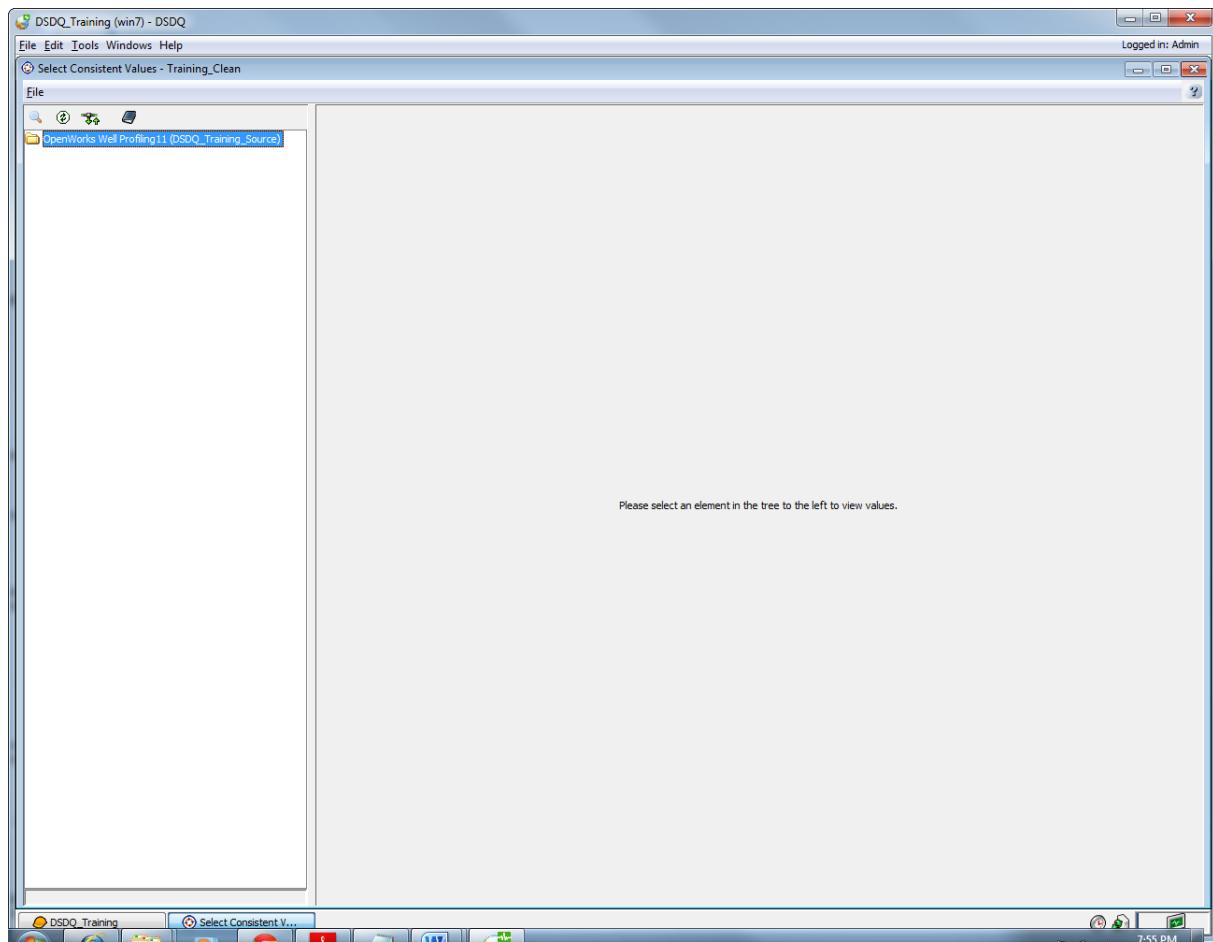
Exercise: Selecting Consistent Values

The **Select Consistent Values** Tool allows you to enter consistent values for columns that have been assigned elements with make values consistent requirements. To select consistent values:

1. Double-click the **Select Consistent Values** Tool or right-click the **Select Consistent Values** Tool and select **Open Tool** from the pop-up menu.

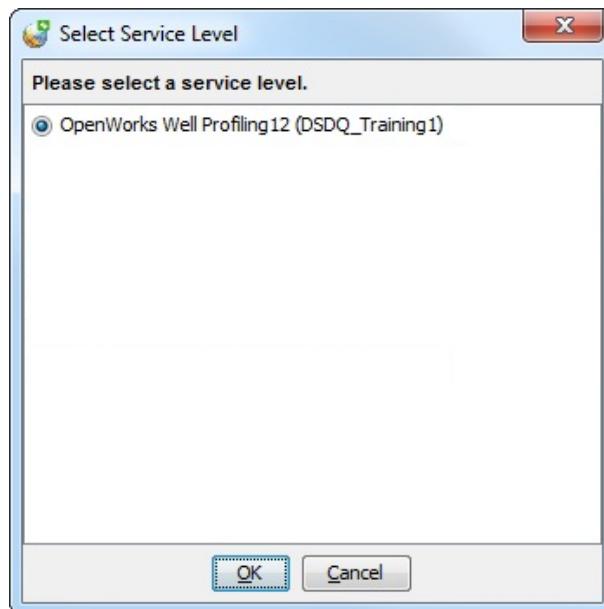


The **Select Consistent Values** window appears.



2. Click on the Service Level Tree toolbar.

The **Select Service level** window appears.



3. Select **OpenWorks Well Profiling12 (DSDQ_Training1)**

4. Click **OK**.

All service levels that have make values consistent requirements are displayed.

The screenshot shows the DSDQ Training software interface. The main window title is "DSDQ_Training (win7) - DSDQ". The menu bar includes File, Edit, Tools, Windows, Help, and a user status "Logged in: Admin". A toolbar with various icons is at the top. The left sidebar shows a project tree under "OpenWorks Well Profiling12 (DSDQ_Training1)", with "Upstream Oil and Gas" expanded, showing "Well Drilling" and "Well", with "Country" selected. The main area contains a data grid titled "Select Consistent Values - Training_Clean". The grid has columns: Value, #, %, Validated Value, Status, Validated, Action, and Suggested Value. Rows 1 through 20 list countries: AFGHANISTAN, ALBANIA, ALGERIA, ANDORRA, ANGOLA, ARGENTINA, ARMENIA, AUSTRALIA, AUSTRIA, AZERBAIJAN, BAHAMAS, BANGLADESH, BARBADOS, BELGIUM, BENIN, BOTSWANA, BRAZIL, BULGARIA, BURMA, and BURUNDI. Most rows have a yellow warning icon in the "Action" column. Row 1 has a checkmark in the "Validated" column. A note at the bottom of the grid says "1 of 102 haveed 2013-12-28 02:11:15.0" and "Indicates value was not found in the application reference." Below the grid is a "Rules Editor" panel with tabs for Rules Editor and Reference Data. The Rules Editor tab is active, showing a "Rules" palette with "Convert to Title Case" selected. The Reference Data tab is also visible. To the right is a "Rules Palette" and "Rule Configuration" panel, which is currently empty. The bottom of the screen shows the taskbar with the DSDQ_Training icon and a "Select Consistent V..." button.

5. Select an element in the Service Level Tree for which you want to select consistent values. The top right table displays the values for the column that is associated to the element.

The screenshot shows the 'DSDQ_Training (win7) - DSDQ' application window. The title bar indicates 'Logged in: Admin'. The main window is titled 'Select Consistent Values - Training_Clean'. On the left, a tree view shows 'Upstream Oil and Gas' and 'Well Drilling' with a 'Well' node expanded, showing a 'country' node. The main area contains two tables:

	Value	#	%	Validated Value	Status	Validated	Action	Suggested Value	Standards Reference
1	AFGHANISTAN	1	0	Afghanistan	?	<input checked="" type="checkbox"/>		Afghanistan	
2	ALBANIA	1	0	Albania	?	<input type="checkbox"/>		Albania	ALBANIA
3	ALGERIA	1	0	ALBANIA	?	<input type="checkbox"/>		Algeria	ALGERIA
4	ANDORRA	1	0		?	<input type="checkbox"/>			ANDORRA
5	ANGOLA	1	0	angola	?	<input checked="" type="checkbox"/>			ANGOLA
6	ARGENTINA	1	0	ARGENTINA	?	<input type="checkbox"/>			ARGENTINA
7	ARMENIA	1	0		?	<input type="checkbox"/>			ARMENIA
8	AUSTRALIA	1	0		?	<input type="checkbox"/>			AUSTRALIA
9	AUSTRIA	1	0		?	<input type="checkbox"/>			AUSTRIA
10	AZERBAIJAN	1	0		?	<input type="checkbox"/>			AZERBAIJAN
11	BAHAMAS	1	0		?	<input type="checkbox"/>			BAHAMAS
12	BANGLADESH	1	0		?	<input type="checkbox"/>			BANGLADESH
13	BARBADOS	1	0		?	<input type="checkbox"/>			BARBADOS
14	BELGIUM	1	0		?	<input type="checkbox"/>			BELGIUM
15	BENIN	1	0		?	<input type="checkbox"/>			BENIN
16	BOTSWANA	1	0		?	<input type="checkbox"/>			BOTSWANA
17	BRAZIL	1	0		?	<input type="checkbox"/>			BRAZIL
18	BULGARIA	1	0		?	<input type="checkbox"/>			BULGARIA
19	BURMA	1	0		?	<input type="checkbox"/>			
20	BURUNDI	1	0		?	<input type="checkbox"/>			BURUNDI

Below the table, a message says: 'Indicates value was not found in the application reference.'

Below the main table is a 'Rules Editor' section with a 'Reference Data' table:

	COUNTRY	ABBREVIATION	ACTIVE_IND	EFFI
1	Afghanistan	AF	Y	
2	ALBANIA	AL	Y	
3	ALGERIA	DZ	Y	
4	AMERICAN SAMOA	AS	Y	
5	ANDORRA	AD	Y	
6	ANGOLA	AO	Y	
7	ANGUILLA	AI	Y	
8	ANTARCTICA	AQ		
9	ANTIGUA AND BARBUDA	AG		
10	ARGENTINA	AR		
11	ARMENIA	AM		
12	ARUBA	AW		
13	AUSTRALIA	AU		

Below this table, a note says: 'No application reference data is associated with the row selected in the table above.'

At the bottom of the dialog, it says 'Standards Reference: DV_REFERENCE.R_COUNTRY'.

6. Select a value in the top table. If the value does not exist in the application reference table, its **Validated Value** will appear in yellow.

	Value	#	%	Validated Value	Status	Validated	Action	Suggested Value	Standards Reference
1	AFGHANISTAN	1	0	Afghanistan	0	<input checked="" type="checkbox"/>		Afghanistan	
2	ALBANIA	1	0	Albania	0	<input type="checkbox"/>		Albania	ALBANIA
3	ALGERIA	1	0	ALBANIA	0	<input type="checkbox"/>		Algeria	ALGERIA
4	ANDORRA	1	0		0	<input type="checkbox"/>			ANDORRA
5	ANGOLA	1	0	angola	0	<input checked="" type="checkbox"/>			ANGOLA
6	ARGENTINA	1	0	ARGENTINA	0	<input type="checkbox"/>			ARGENTINA
7	ARMENIA	1	0		0	<input type="checkbox"/>			ARMENIA
8	AUSTRALIA	1	0		0	<input type="checkbox"/>			AUSTRALIA
9	AUSTRIA	1	0		0	<input type="checkbox"/>			AUSTRIA
10	AZERBAIJAN	1	0		0	<input type="checkbox"/>			AZERBAIJAN
11	BAHAMAS	1	0		0	<input type="checkbox"/>			BAHAMAS
12	BANGLADESH	1	0		0	<input type="checkbox"/>			BANGLADESH
13	BARBADOS	1	0		0	<input type="checkbox"/>			BARBADOS
14	BELGIUM	1	0		0	<input type="checkbox"/>			BELGIUM
15	BENIN	1	0		0	<input type="checkbox"/>			BENIN
16	BOTSWANA	1	0		0	<input type="checkbox"/>			BOTSWANA
17	BRAZIL	1	0		0	<input type="checkbox"/>			BRAZIL
18	BULGARIA	1	0		0	<input type="checkbox"/>			BULGARIA
19	BURMA	1	0		0	<input type="checkbox"/>			
20	BURUNDI	1	0		0	<input type="checkbox"/>			BURUNDI

1 of 162 | Admin | 2014-02-03 14:42:00.938 | Indicates value was not found in the application reference.

7. Enter the final value of the element in the **Validated Value** column. This value will be used when the **Run Make Values Consistent** job is processed.
8. Select the check box in the **Validated** column for the element for which you want to **Select Consistent Values**.



Note

The **Validated** column can only be checked if a Validated Value has been entered. Once the **Validated** column has been checked, all values in that row will no longer be editable. Once you unchecks the **Validated** column, the row becomes editable.

The **Action** column has the following three buttons:

- **Clear Validated Value** - clear the Validated Value for that row.
- **Flag Value** - set the value to be looked at in future.
- **Set Value to Null** - remove the current value and set it to Null.

On clicking the **Validate All Rows** button in the toolbar, the Status column will be set to **Updated by User** for all records and the Validated column will be checked, as long as a Validated Value exists for that row.

On clicking the **Clear All Validated Rows** button in the toolbar, the Status and Validated columns values will be unchecked.

Once the elements and validated values have been correctly configured, the values can be modified by running the **Run Make Values Consistent** task. If the **Load Current Values** task is run again, then the processed records are stored and can be viewable by clicking on the **Previous Validated Values** button. This is a toggle button and clicking it again will allow you to view only current data.

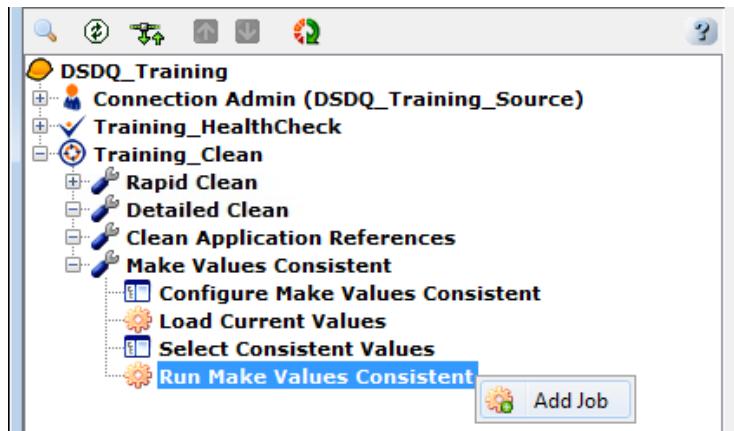
9. Click to save changes.
10. Select **File > Exit** from the menu bar on the **Select Consistent Value** window.

Exercise: Running Make Values Consistent

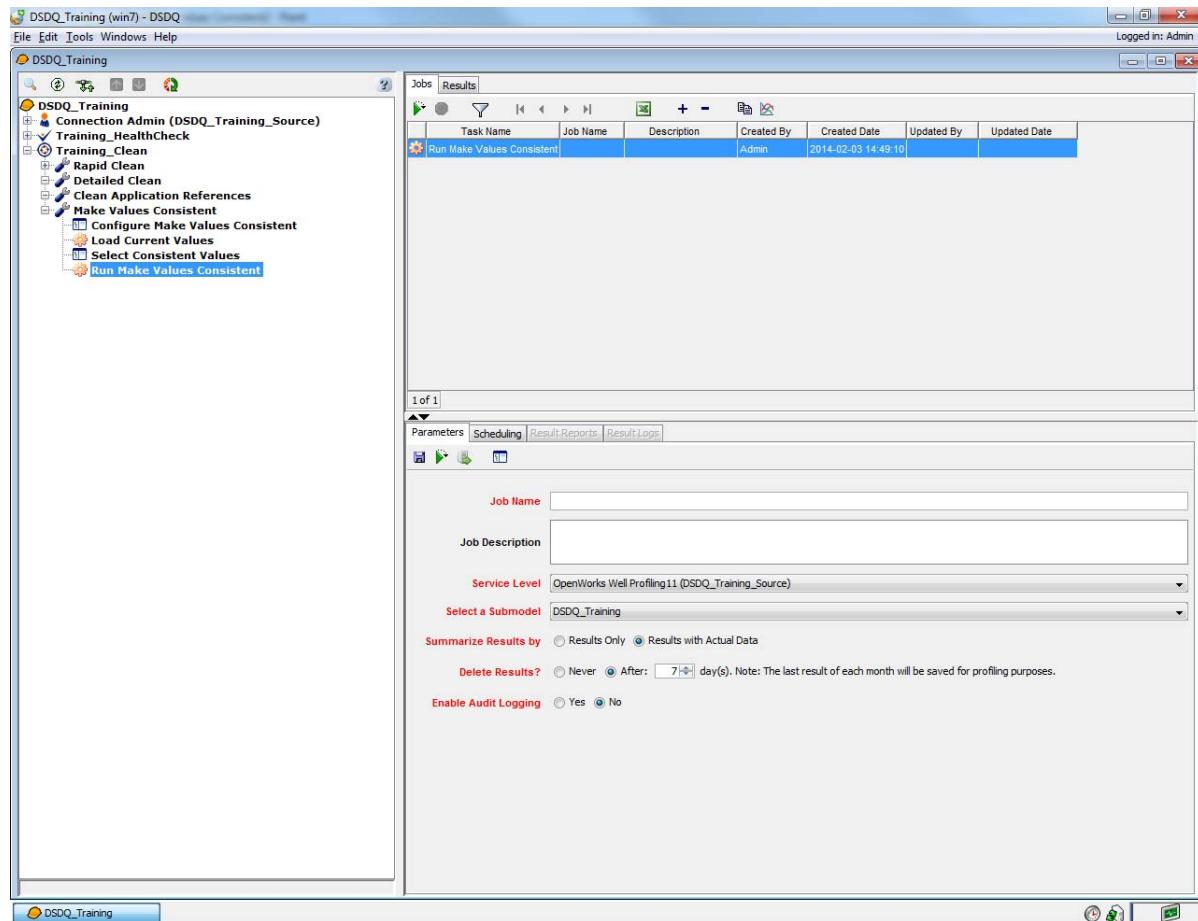
The Run Make Values Consistent Task modifies values with Validated Values for the data and updates the actual data. To run the Make Values Consistent task:

1. Double-click the **Run Make Values Consistent** Task or right-click the **Run Make Values Consistent** Task and select **Add Job** from the

pop-up menu.



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

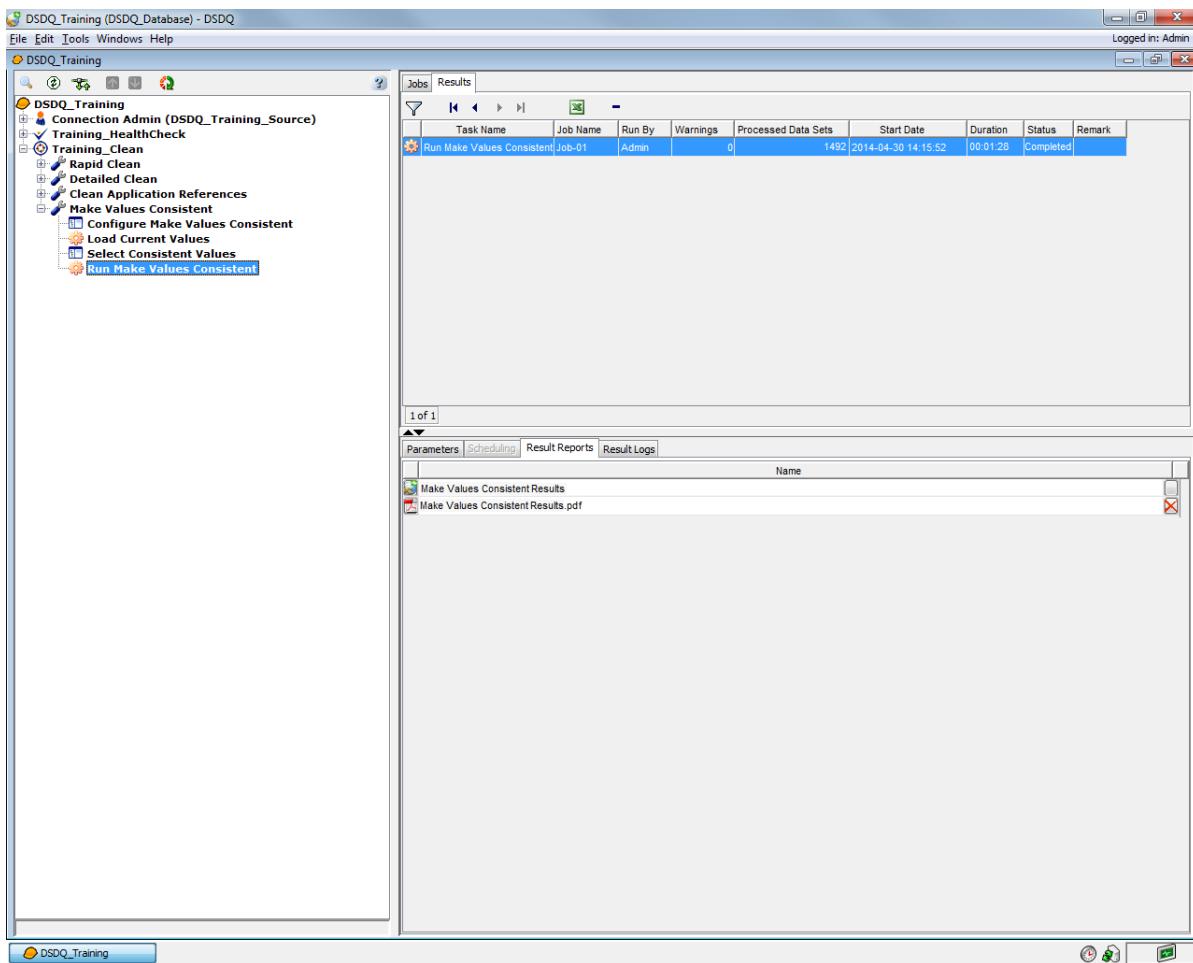


2. Enter **Job-01** in the **Job Name** field.
3. Enter **Make Values Consistent** in the **Job Description** field.

4. Select **OpenWorks Well Profiling12 (DSDQ_Training1)** from the **Service Level** drop-down list.
5. Select **DSDQ_Training** from the **Select Submodel** drop-down list.
6. Select the **Results with Actual Data** option for **Summarize Results by**.
7. Select the **After** option for **Delete Results?** Leave the number of days as **7**.
8. Select the **Yes** option for **Enable Audit Logging**.
9. Click  to save changes.
10. Click .

The **Run Make Values Consistent** task is executed and displays results in the **Result Reports** tab.

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



12. Click on the **Results Report** tab to display **Make Values Consistent Results** in PDF format.

Make Values Consistent Results		HALLIBURTON	
Project:	DSDQ_Training	Landmark Software & Services	
Phase:	Training_Clean		
Task:	Run Make Values Consistent		
Job:	job-01		
Connection:	DSDQ_Training_Source		
Result Date:	Sat, Dec 28, 2013 02:08		
Table Name:	RCountry	Column Name:	country_name
		Original	Accepted
		AFGHANISTAN	Afghanistan
		ALGERIA	Algeria
		ALBANIA	Albania
			Rows Affected
			Result %
			Remark

Chapter 6

Managing Data Duplication in DecisionSpace Data Quality

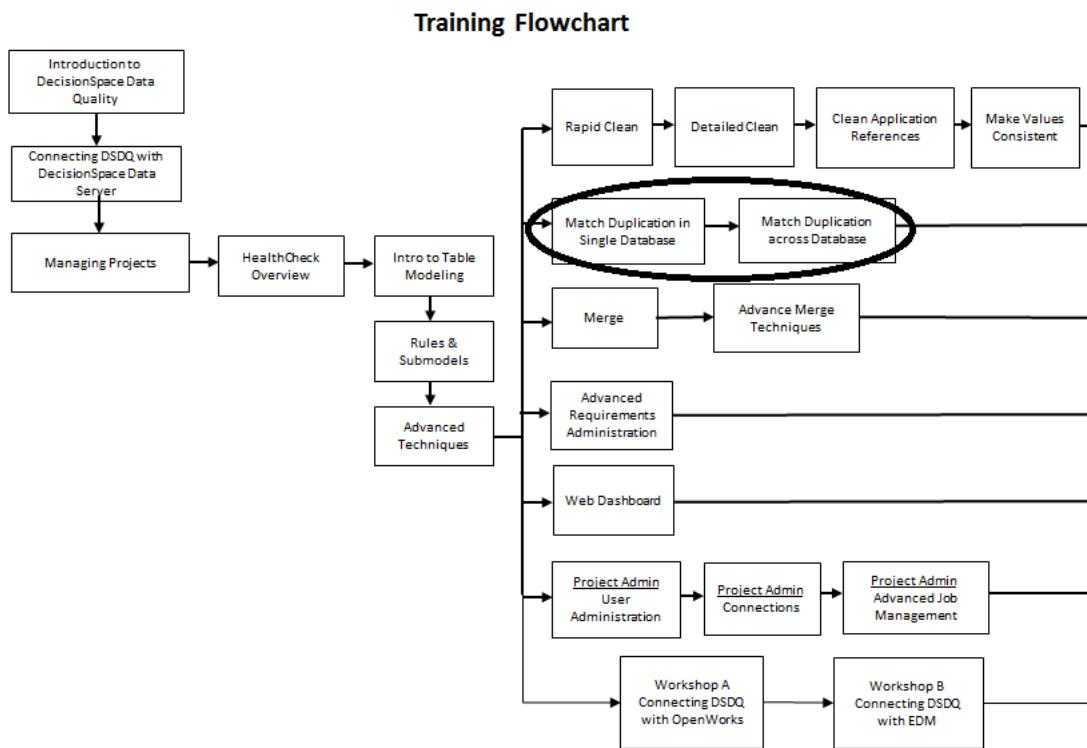
The Match Phase enables you to address data duplication issues found in data sources. Duplication can occur within a data source as well as across data sources. Once data duplication has been identified, you have the ability to clean or merge the duplicated data.

Chapter Overview

In this chapter, you will learn about:

- Data duplication
- Using the **Detailed Match** Activity for a single data source
- Managing duplication for a single data source
- Using the **Detailed Match** Activity across data sources
- Managing duplication across data sources

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:



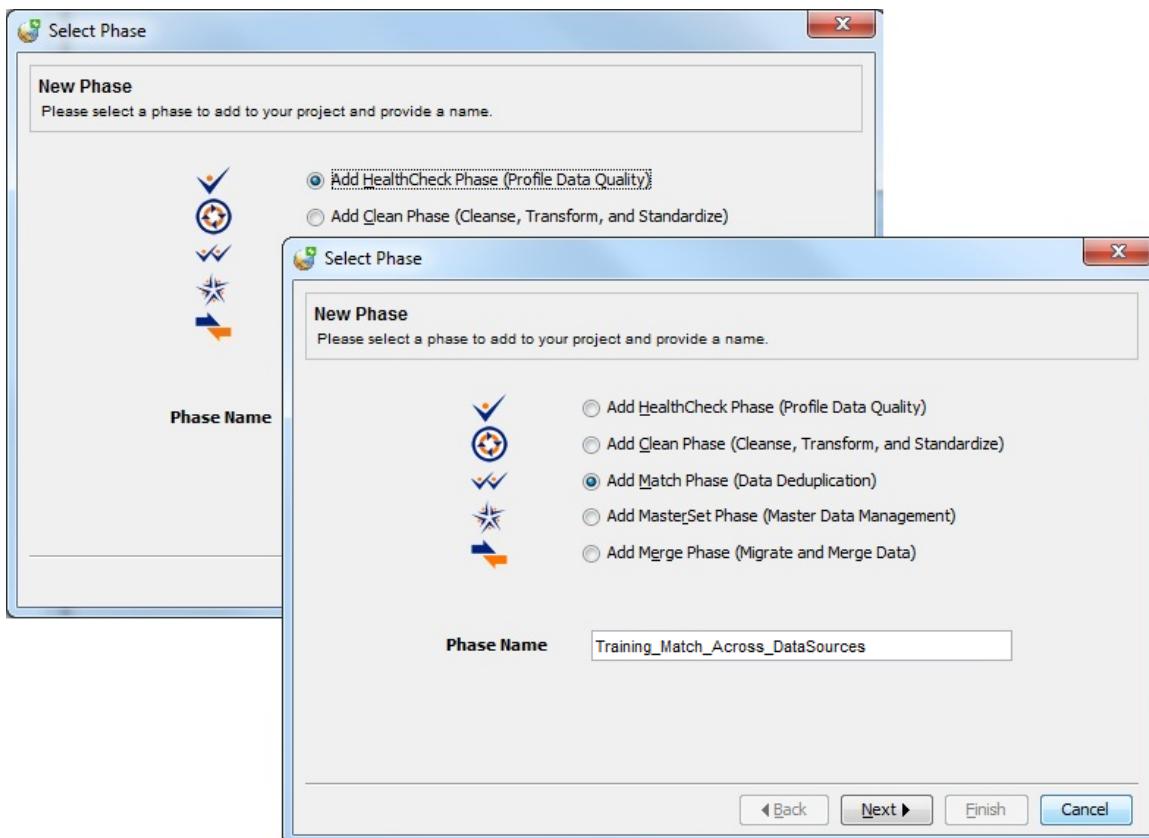
Data Duplication

Data Duplication occurs when multiple copies of the same data are created. This process is performed to create backups of specific data and/or can occur due to automated duplication of records. In situations where unwanted duplicate data copies are created, a procedure is implemented to remove or merge the duplicated copies into one clear source of accurate and updated information. The data quality application matches common records and removes or merges duplication. This not only helps in saving databases space but also assists in data organization.

Exercise: Adding a Match Phase for a Single Data Source

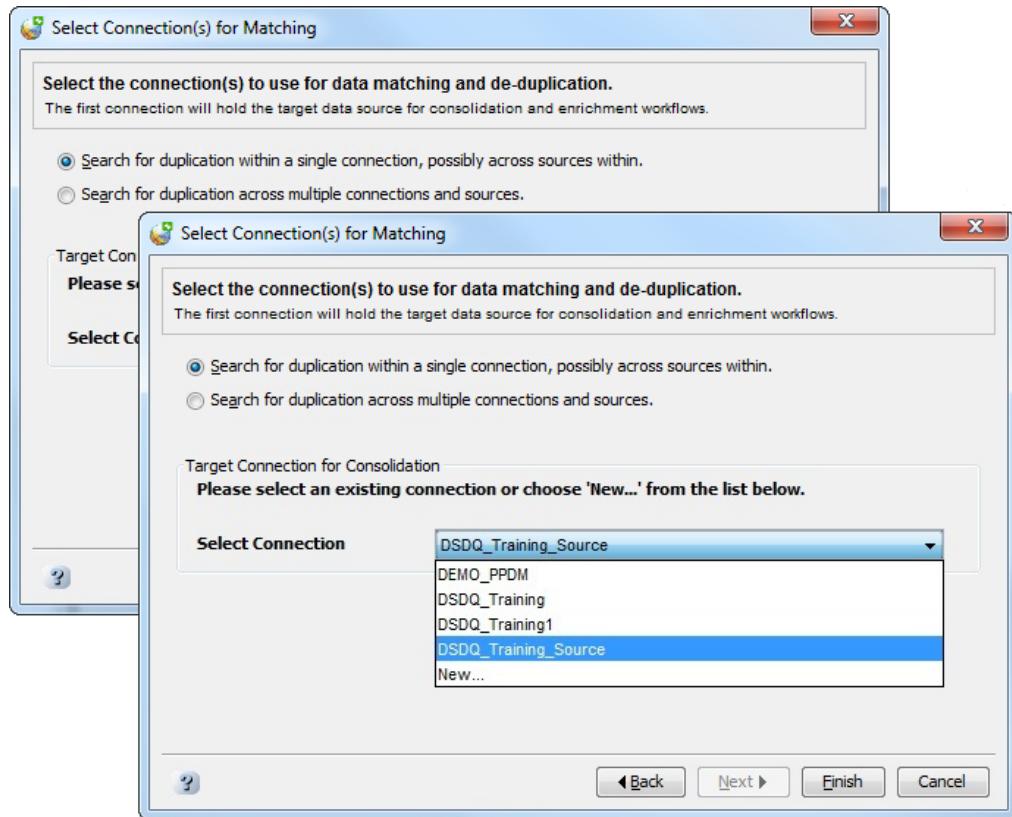
To add a Match Phase for a single data source:

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 Match Phase (Data Deduplication)** option.
3. Enter **Training_Match** in the **Phase Name** field.
4. Click **Next** to continue.
The **Select Connection(s) for Matching** window appears with the

Search for duplication within a single connection, possibly across sources within. option selected by default.



5. Select **DSDQ_Training_Source** from the **Select Connection** drop-down list and click **Finish**.
The Match Phase is created and displayed in the DecisionSpace Data Quality Project window.

Detailed Match for a Single Data Source

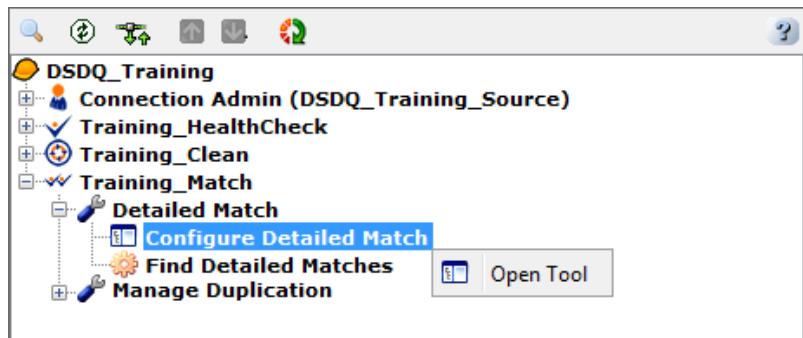
The **Detailed Match** Activity helps you in creating your own source data sets within match groups. Duplicated matches are found within the source data and you can setup specific match requirements for finding matches. Once the matches are found, they can be fixed. During the **Detailed Match for a Single Data Source** activity, only a specific data source is configured for duplication removal.

Exercise: Configuring the Detailed Match for a Single Data Source

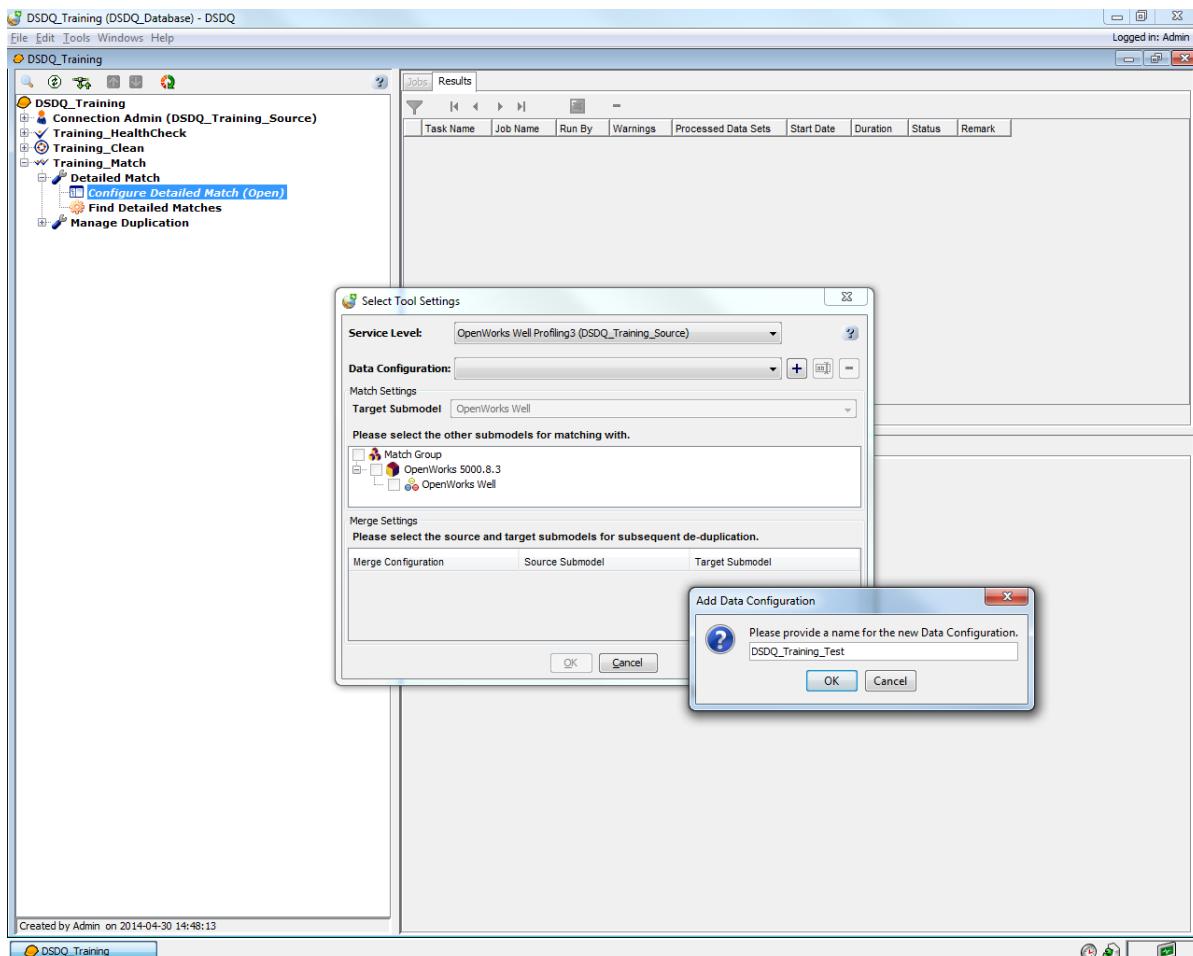
The **Configure Detailed Match** Tool is used to create match groups and configure service levels for testing prior to running the **Find Detailed Matches** Task. The match group consists of a single submodel. The user can select which requirements to enable/disable in the service level. You will also need to select a subset of data for testing.

To configure detailed match for a single data source:

1. Click on the DecisionSpace Data Quality Tree to expand the **Training_Match** Phase.
2. Click to expand the **Detailed Match** Activity.
3. Double-click the **Configure Detailed Match** Tool or right-click the **Configure Detailed Match** Tool and select **Open Tool** from the pop-up menu.



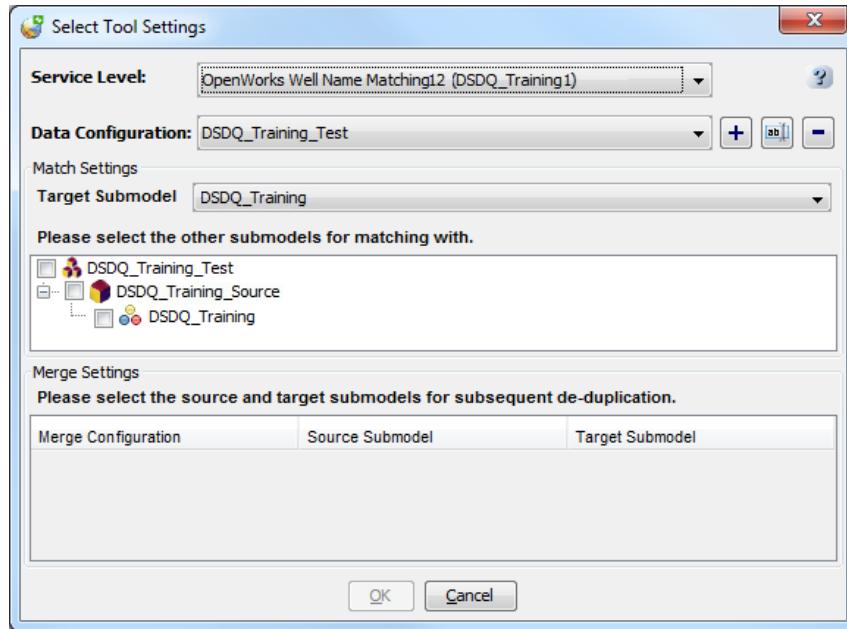
The **Add Data Configuration** dialog box appears by default the first time you run the **Configure Detailed Match** Tool.



4. Enter **DSDQ_Training_Test** in the **Add Data Configuration** dialog box.

5. Click **OK**.

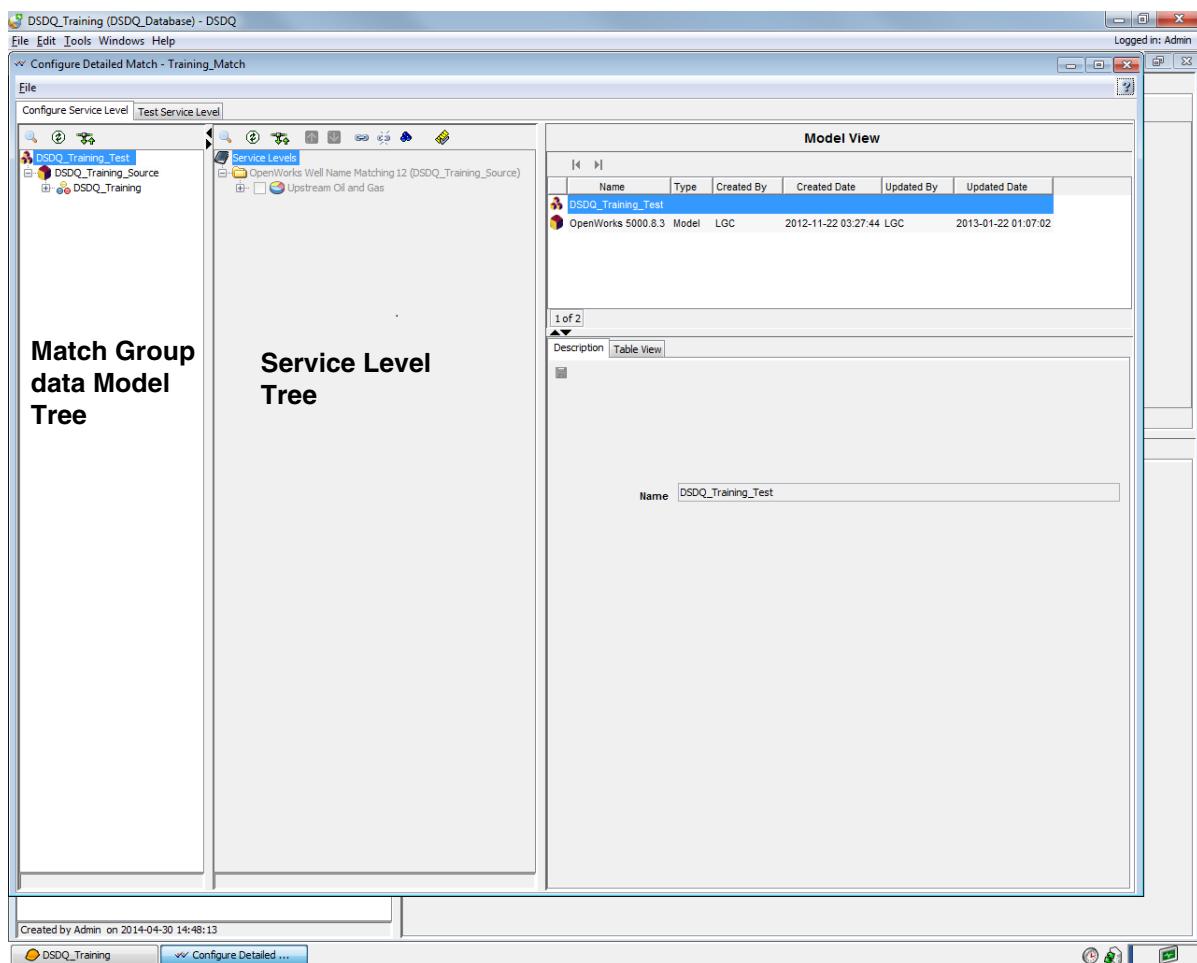
The **Select Tool Settings** window appears.



6. Select **OpenWorks Well Name Matching12 (DSDQ_Training1)** from the **Service Level** drop-down list
7. Select **DSDQ_Training_Test** from the **Data Configuration** drop-down list.
8. Select **DSDQ_Training** from the **Target Submodel** drop-down list.
9. Select the **DSDQ_Training_Test** check box.
10. Click **OK**.

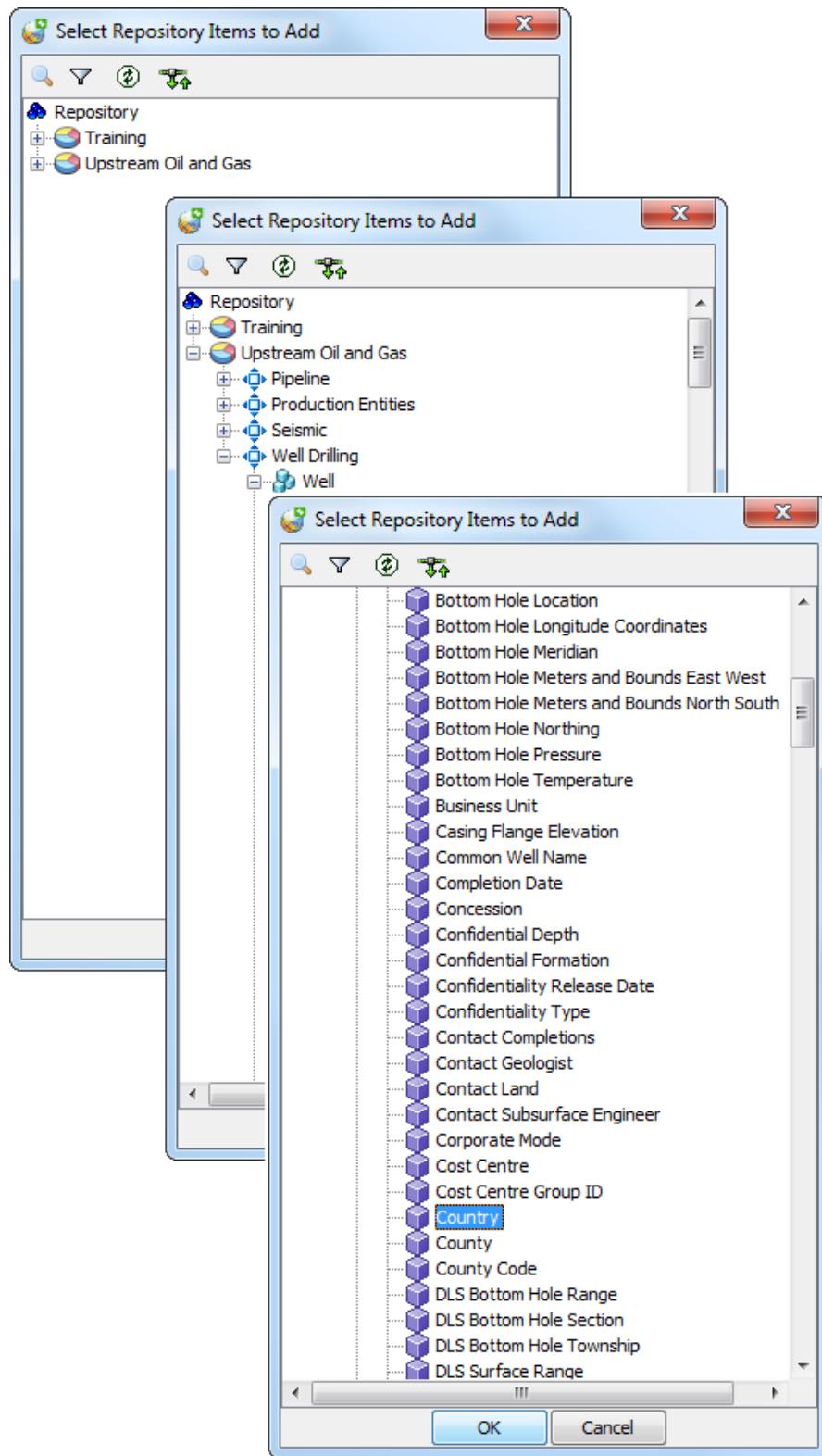
The **Configure Detailed Match** window appears with **DSDQ_Training_Test** and **OpenWorks Well Name Matching 12**

(DSDQ_Training_Source) displaying in the Match Group Data Model Tree and Service Level Tree respectively.



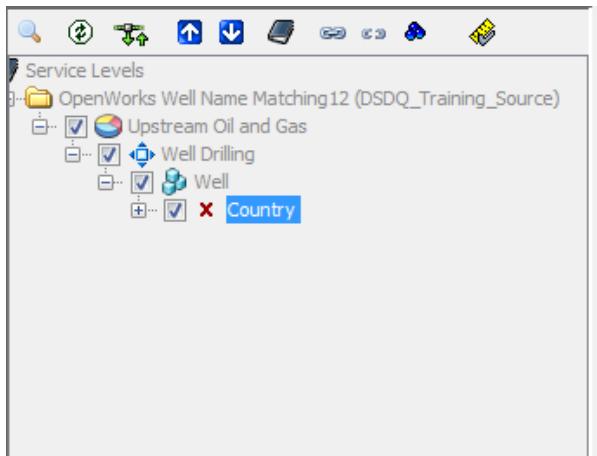
11. Click  on the Service Level Tree toolbar.

The Select Repository Items to Add dialog box opens.



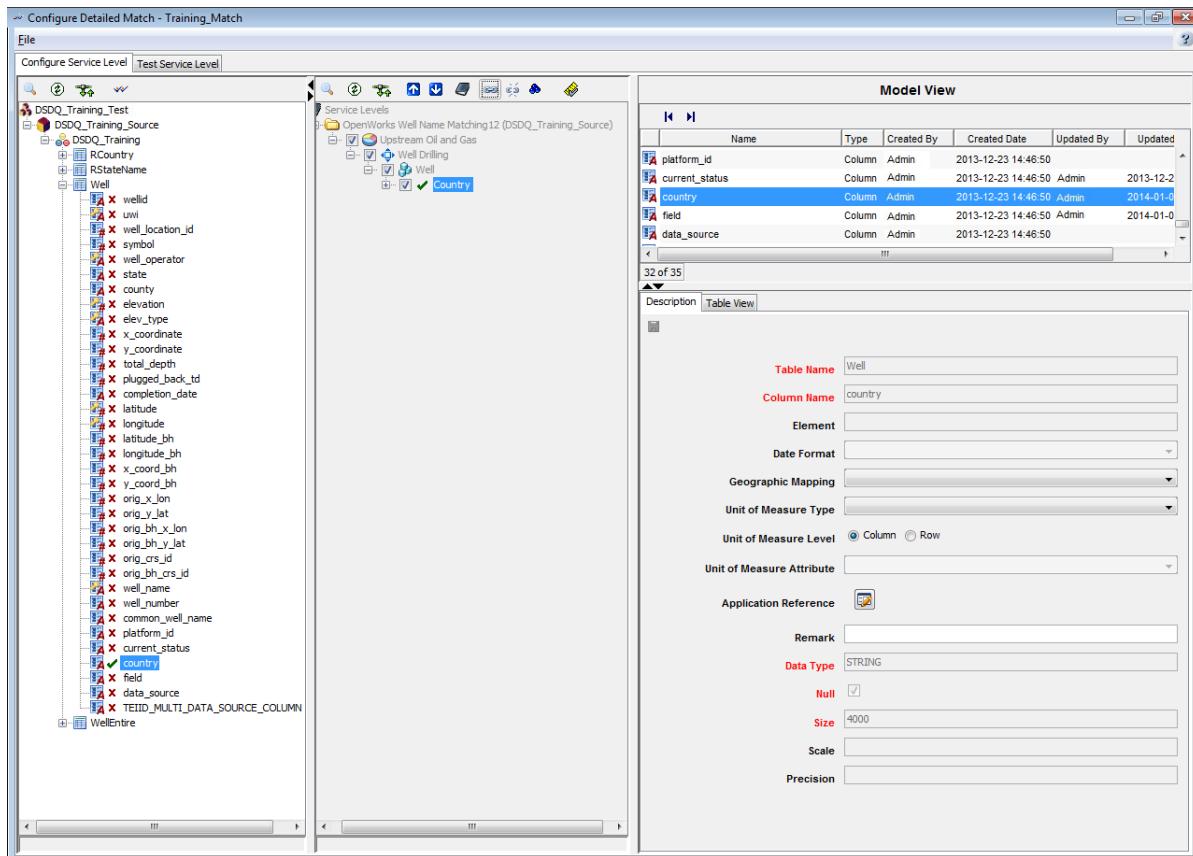
12. Click  to expand the **Upstream Oil & Gas** sector.
13. Expand the **Well Drilling** area.
14. Expand the **Well** element group, select the **Country** element
15. Click **OK**.

The **Country** element is added to the Service Level Tree. A red cross will appear adjacent to the element in the Service Level Tree indicating that the element is not linked to any column.



16. Click  on the Match Group Data Model Tree to expand **DSDQ_Training_Source** table.
17. Expand the **Well** table and select the **Country** column.
18. Drag and drop the **Country** column onto the **Country** element in the Service Level Tree.

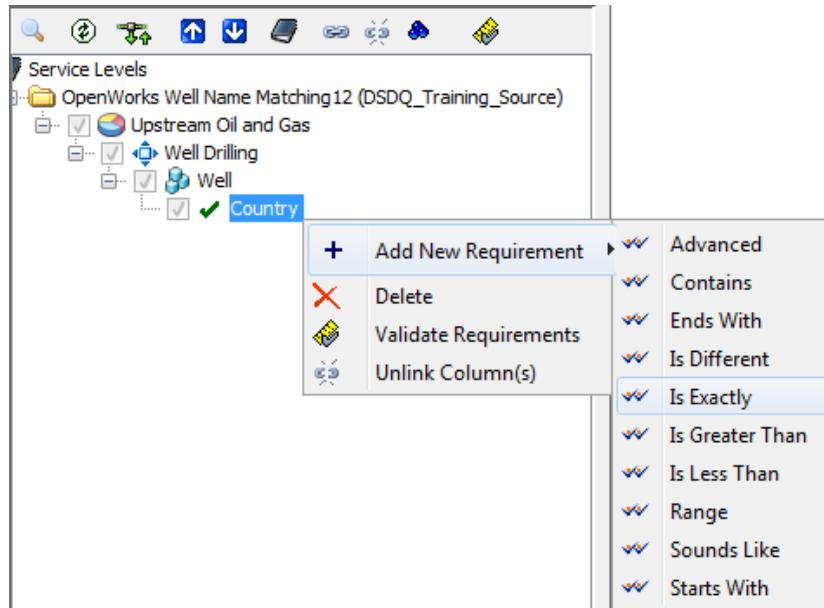
A green check mark will appear adjacent to the column and element that have just been associated. Only one column from the same table can be linked to the same element. However, it is possible to link many columns to the same element if the columns come from different tables. Alternatively you can select the column & element to link and click the **Link Column to Element** button on the Service Level Tree toolbar.



Note

You can unlink elements or columns. Select the element or column to unlink and click the **Unlink Columns from an Element** button on the Service Level Tree toolbar.

19. Right-click on the **Country** element in the Service Level Tree and select **Add New Requirement > Is Exactly** from the pop-up menu.



The **Enter Name** dialog box appears.



20. Optionally, specify a user-defined name for the requirement.
21. Click **OK** to add the requirement to the selected element.
22. Optionally, repeat steps **11** to **21** to add all elements for matching.

23. Click the Test Service Level tab.

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

Data Sets	Confidence	Well Country
393 (Well)	100%	USA
573 (Well)	80%	USA
578 (Well)	80%	USA
582 (Well)	80%	USA
586 (Well)	80%	USA
1487 (Well)	80%	USA
1499 (Well)	80%	USA
1507 (Well)	80%	USA
1512 (Well)	80%	USA
558 (Well)	80%	USA
514 (Well)	80%	USA
519 (Well)	80%	USA
523 (Well)	80%	USA
541 (Well)	80%	USA
545 (Well)	80%	USA
549 (Well)	80%	USA
553 (Well)	80%	USA
495 (Well)	80%	USA
491 (Well)	80%	USA
503 (Well)	80%	USA
499 (Well)	80%	USA
531 (Well)	80%	USA
527 (Well)	80%	USA
486 (Well)	80%	USA
536 (Well)	80%	USA
1164 (Well)	80%	USA
1160 (Well)	80%	USA
1173 (Well)	80%	USA
1169 (Well)	80%	USA
1147 (Well)	80%	USA
460 (Well)	80%	USA
1156 (Well)	80%	USA
1151 (Well)	80%	USA
1421 (Well)	80%	USA
651 (Well)	80%	USA
1411 (Well)	80%	USA
1416 (Well)	80%	USA
1402 (Well)	80%	USA
1407 (Well)	80%	USA
1177 (Well)	80%	USA
1394 (Well)	80%	USA
734 (Well)	80%	USA
739 (Well)	80%	USA
743 (Well)	80%	USA
753 (Well)	100%	USA

24. Verify all entries have correct matches; in this case the correct match is “USA”.

25. Click the Next Data Set button to test the next record.

26. Repeat steps **25 to test all records.**

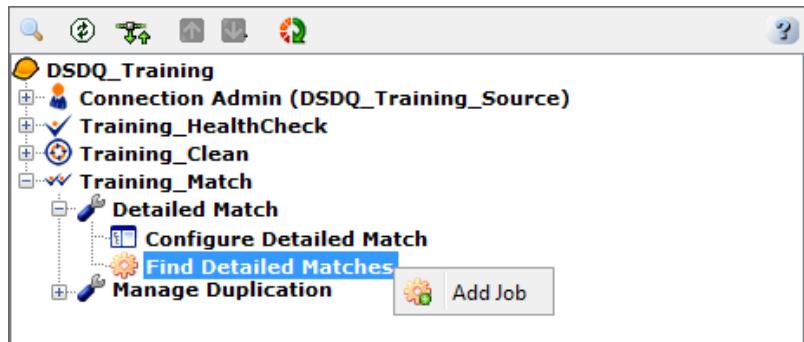
27. Select File > Exit from the menu bar on the Configure Detailed Match window.

Exercise: Finding Detailed Matches within a Single Data Source

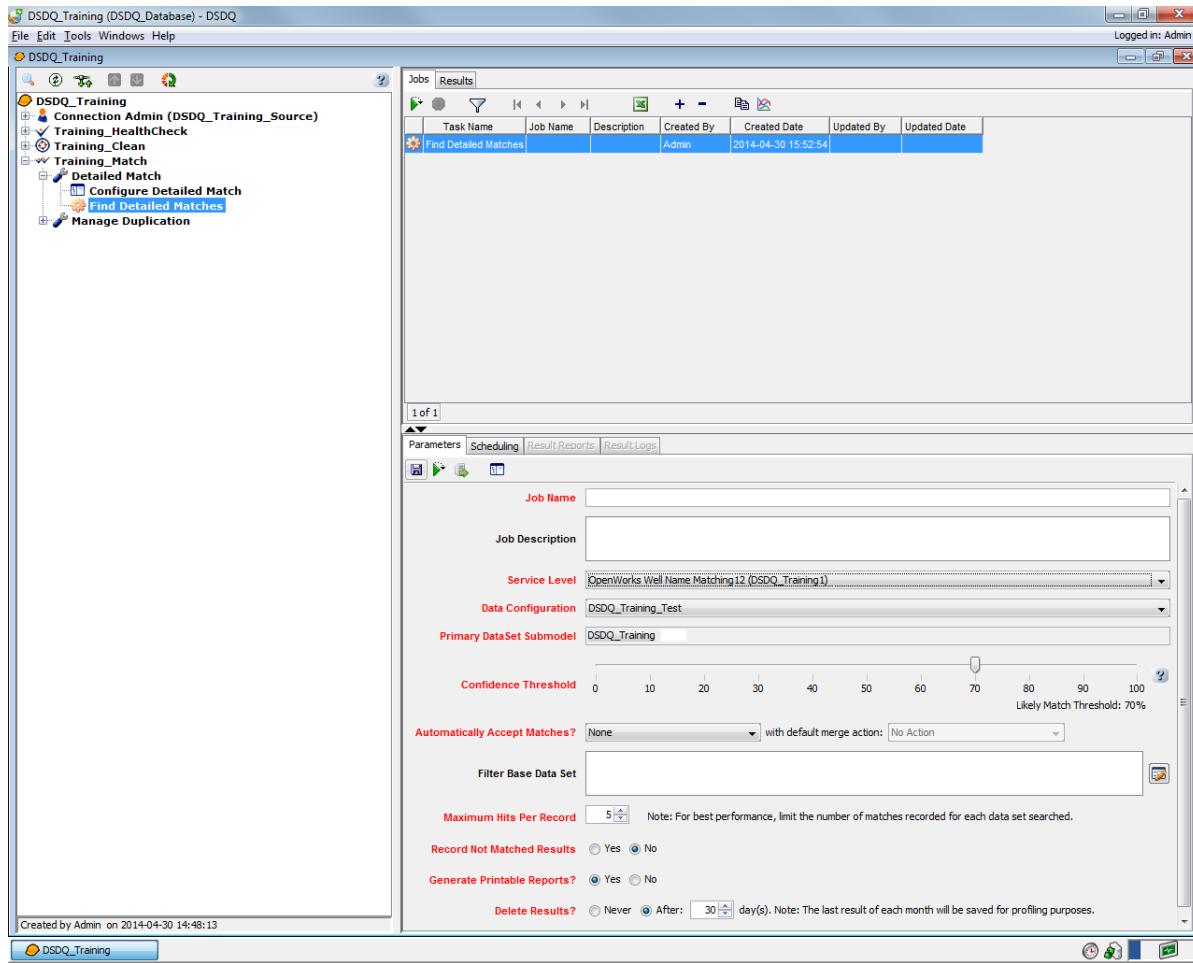
The **Find Detailed Match** Task allows you to check values that match, or are redundant within the database.

To find Detailed Matches within a Single Data Source:

1. Double-click the **Find Detailed Matches** Task or right-click the **Find Detailed Matches** task and select **Add Job** from the pop-up menu.



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



2. Enter **Job-01** in the **Job Name** field.
3. Enter **Find Detailed Matches** in the **Job Description** field.
4. Select **OpenWorks Well Name Matching 12 (DSDQ_Training_Source)** from the **Service Level** drop-down list.
5. Select **DSDQ_Training_Test** from the **Data Configuration** drop-down list.
The **Primary DataSet Submodel** field populates automatically.

6. Set the **Confidence Threshold** option as **70**.

Note

Confidence Thresholds allows you to set the minimum required match confidence for a selected match hit to be recorded. Setting a higher value here will be more selective in which possible matches appear in the Match results viewer. Setting a lower value will be more permissive, allowing for more matches to be presented, though they may possibly be incorrect upon expert review.

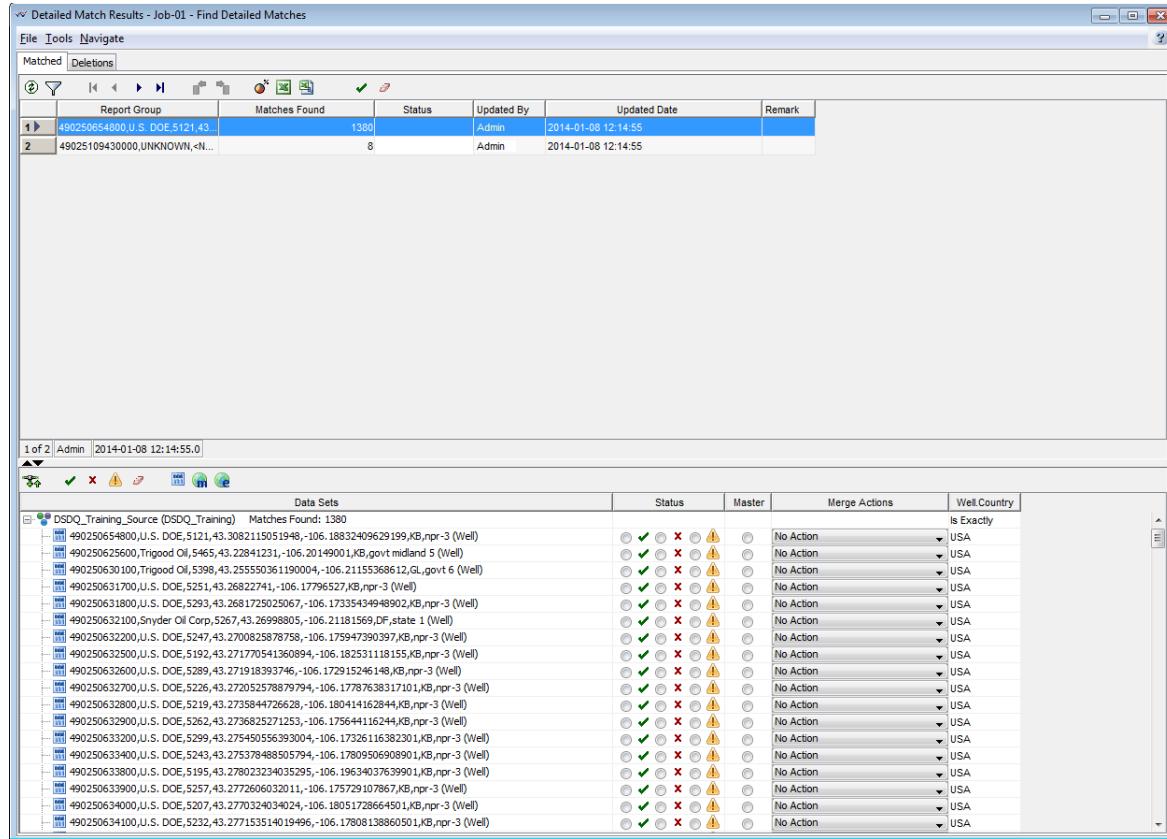
7. Select **None** from the **Automatically Accept Matches** drop-down list.

Note

Automatically Accept Matches allows the **Find Detailed Matches** Task to automatically match depending on the criteria selected: either **All** or **Having only one match per source**. It will then mark the merge action on each match according to the default merge action selected.

8. Optionally, set a filter on the data subset.
9. Set the **Maximum Hits Per Record** option as **5**.
10. Select the **No** option for **Record Not Matched Results**.
11. Select the **Yes** option for **Generate Printable Reports?**
12. Select the **No** option for **Enable Data Read Ahead?**
13. Select the **After** option for **Delete Results?** Leave the number of days as **30**.
14. Click  to save changes in the **Parameters** tab.
15. Click .
The **Find Detailed Matches** task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.
16. 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**.

17. Double-click **Detailed Match Results** on the **Result Reports** tab.
The **Detailed Match Results** window appears.



18. Select all the rows in the **Matches Details** Pane and click on the **Matches Found** toolbar.

Note

Selecting the **Yes Status** radio button implies that you want the result to match the source.

Selecting the **No Status** radio button implies that you do not want to match it to the source.

Selecting the **Review Status** radio button implies that the record still needs to be reviewed. You can set these statuses for all by clicking the respective buttons in the toolbar.

To clear the status for this result set, click the **Clear Status** button.

19. Optionally, select the data and click the **Show on Browser Map** button on the toolbar to display the location of the data set on a map in the **DecisionSpace Data Quality** Dashboard. You can also

display the data in a Pie chart by click the **Generate Status Chart** button.

20. Select **File > Exit** from the menu bar on the **Detailed Match Results** window.

Managing Duplication for a Single Data Source

The **Manage Duplication** Activity manages the process of duplication removal with the right set of policies. This ensures that duplication removal standards are met. A submodel is configured for adding matches, which are then removed or merged as desired.

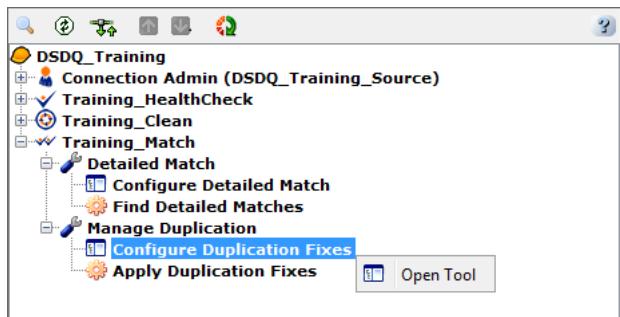
Exercise: Configuring Duplication Fixes for a Single Data Source

The **Configure Duplication Fixes** Tool enables you to setup table and column mappings for the automatically created merge group, generated from the Match Group that you created in the **Detailed Match** Activity.

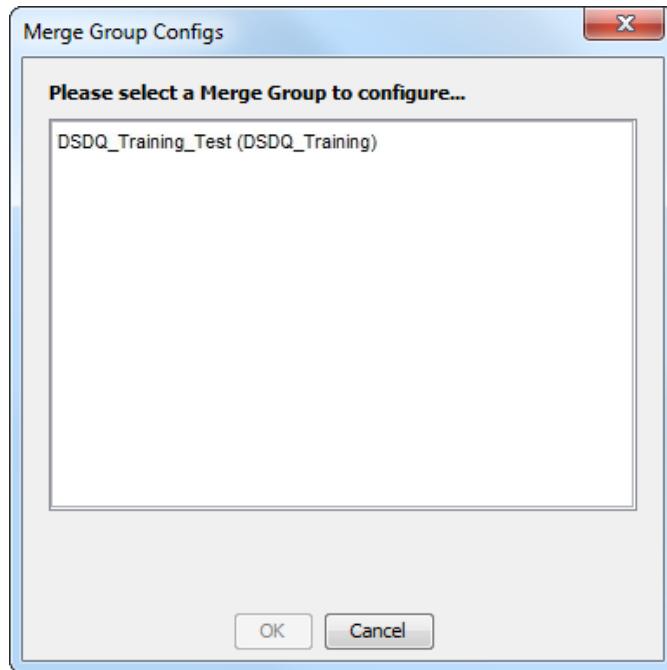
To Configure Duplication Fixes for a Single Data Source:

1. Click  to expand the **Manage Duplication** Activity on the DecisionSpace Data Quality Tree.
2. Double-click the **Configure Duplication Fixes** Tool or right-click the **Configure Duplication Fixes** Tool and select **Open Tool** from

the pop-up menu.



The **Merge Group Configs** window appears.

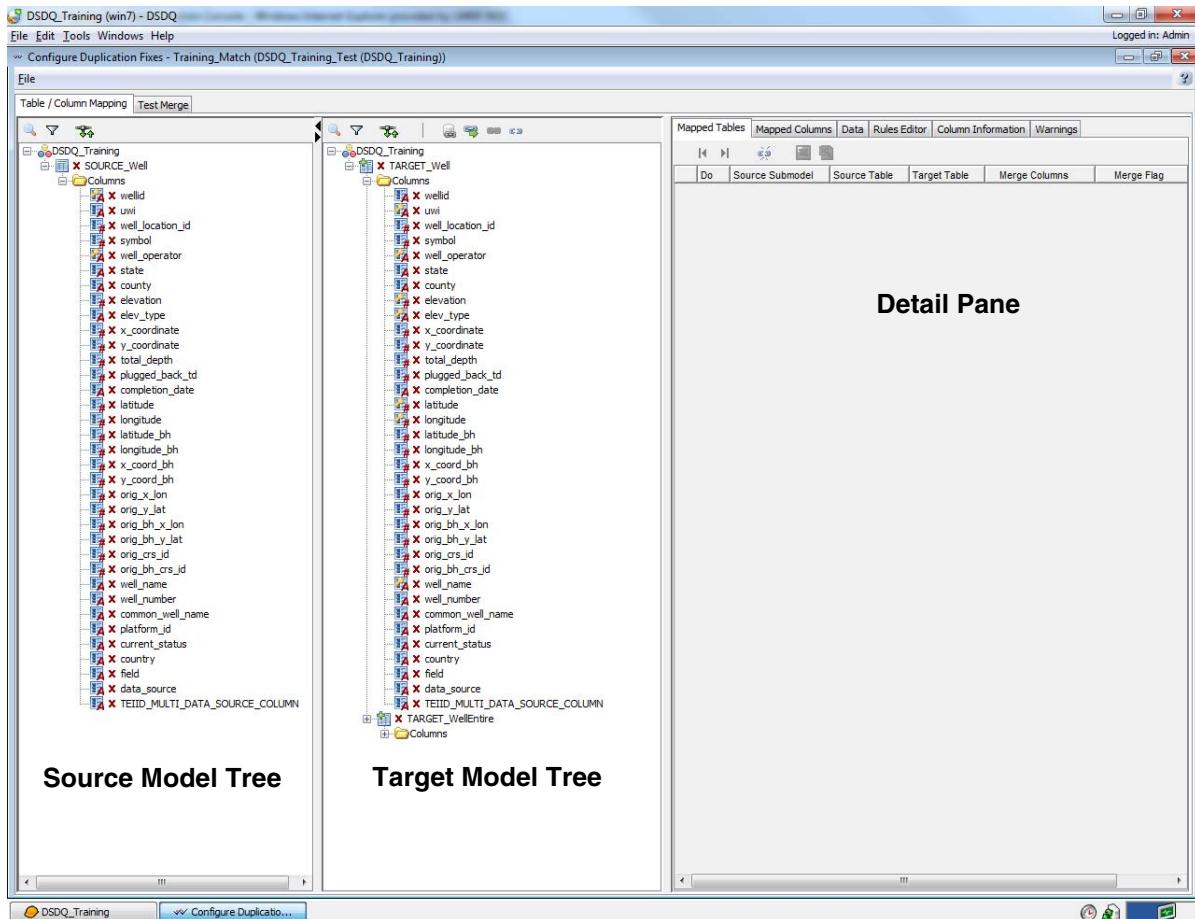


3. Select the **DSDQ_Training_Test (DSDQ_Training)** option.
 4. Click **OK**.
- The **Configure Duplication Fixes** window appears, displaying tables and columns for the Source Model Tree, Target Model Tree

and the Detail pane. The Detail Pane has six tabs:

Mapped Tables	This area displays information about the tables that have been mapped in the source and target trees.
Mapped Columns	This area displays the mapping between the target column and all its mapped sources.
Data	<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>The Source Data View and Target Data View toolbars can be used to Filter Data , to move to the First Record in Current View  or Last Record in Current View , or to move to the Next Data Set  or Previous Data Set  . The data can also be saved in Excel format by clicking on the corresponding button  or exported to a CSV file by selecting the Create a CSV Export File  button on the toolbar.</p> <p>Data in the Data Detail Pane can be sorted by clicking a column header. Column information on any column can be viewed by right-clicking a column header and selecting Column Info from the pop-up menu. Other columns information can be viewed by right-clicking a column header and selecting Columns Filter from the pop-menu.</p>
Rules Editor	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.
Column Information	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.

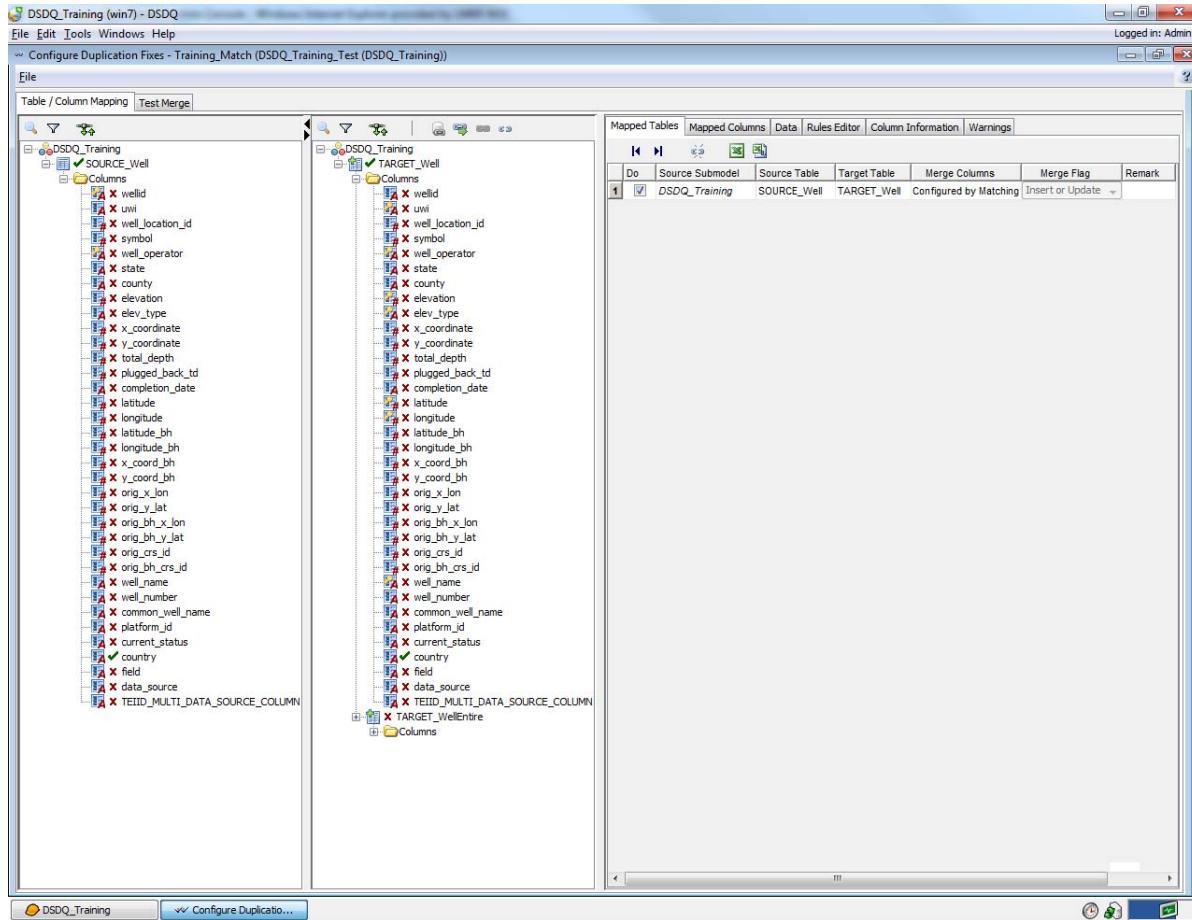
Warnings	This area will display any inconsistencies between the mapped columns, e.g. source column length is greater than the target column length. Initially this tab is blank. When the first warning is logged, the tab name turns red and a warning icon appears next to its name.
-----------------	---



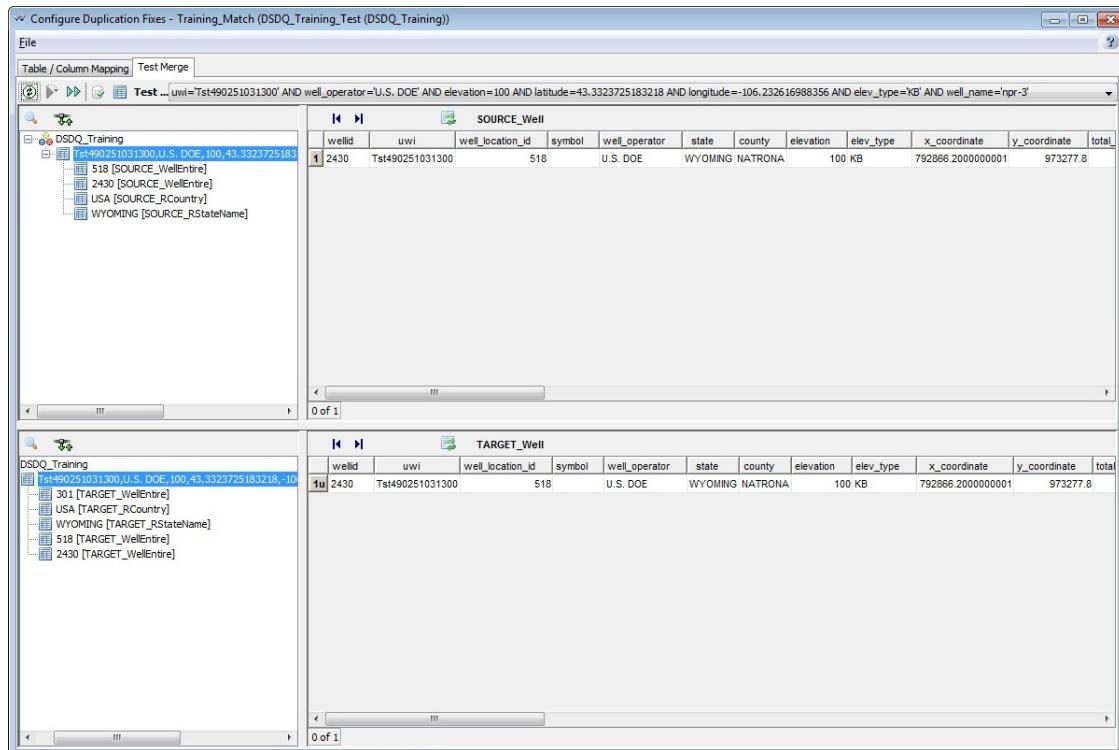
5. Select the **Source_Well** table from the Source Model Tree.
6. Select the **Target_Well** table from the Target Model Tree.
7. Click the **Auto Map selected Table/Columns** button on the toolbar.
A green check mark appears adjacent to the selected tables.
8. Select **Country** column from the **Source_Well** and **Target_Well** table.

9. Click the **Auto Map selected Table/Columns**  button on the toolbar.

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



10. Select the **Test Merge** tab to test all match result configurations.



11. Select **File > Exit** from the menu bar on the **Configure Duplication Fixes** window.

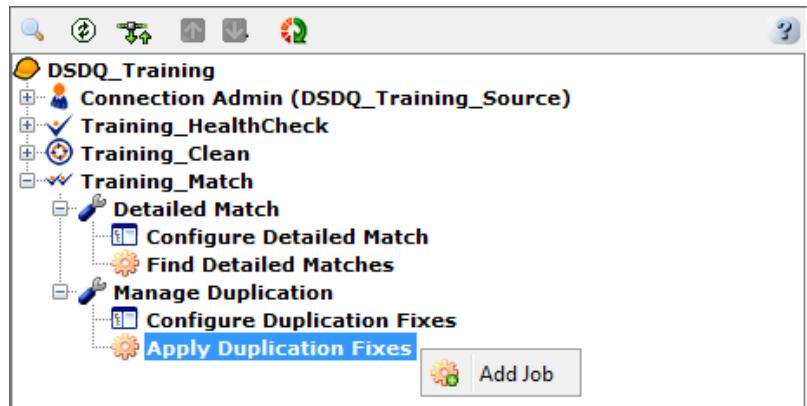
Exercise: Applying Duplication Fixes for a Single Data Source

After the duplications have been tested in the **Configure Duplication Fixes** Tool, the **Apply Duplication Fixes** Task can be used to move the complete data set over to the target database and remove the duplication.

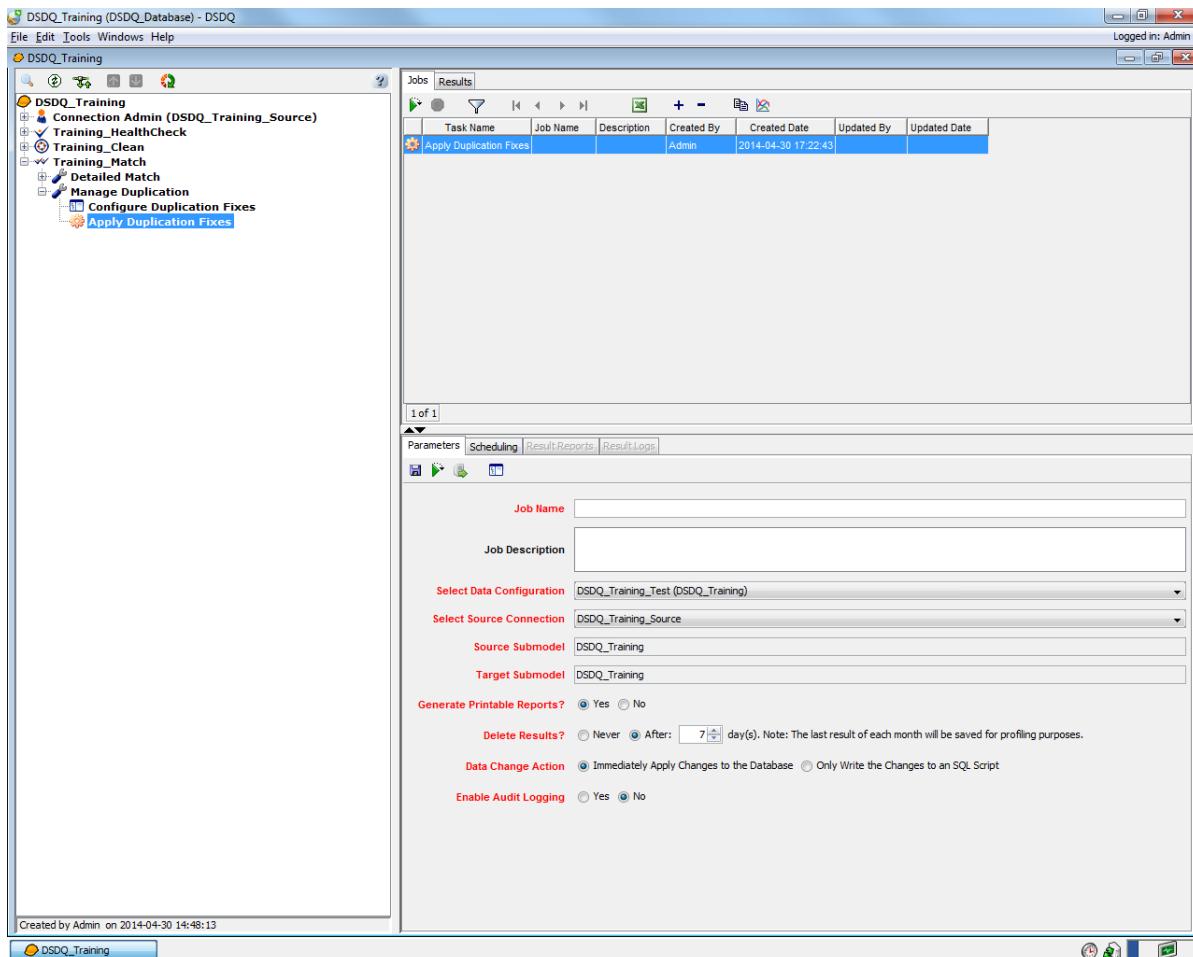
To Apply Duplication Fixes for a Single Data Source:

1. Double click the **Apply Duplication Fixes** task or right-click the **Apply Duplication Fixes** task and select **Add Job** from the pop-up

menu.



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

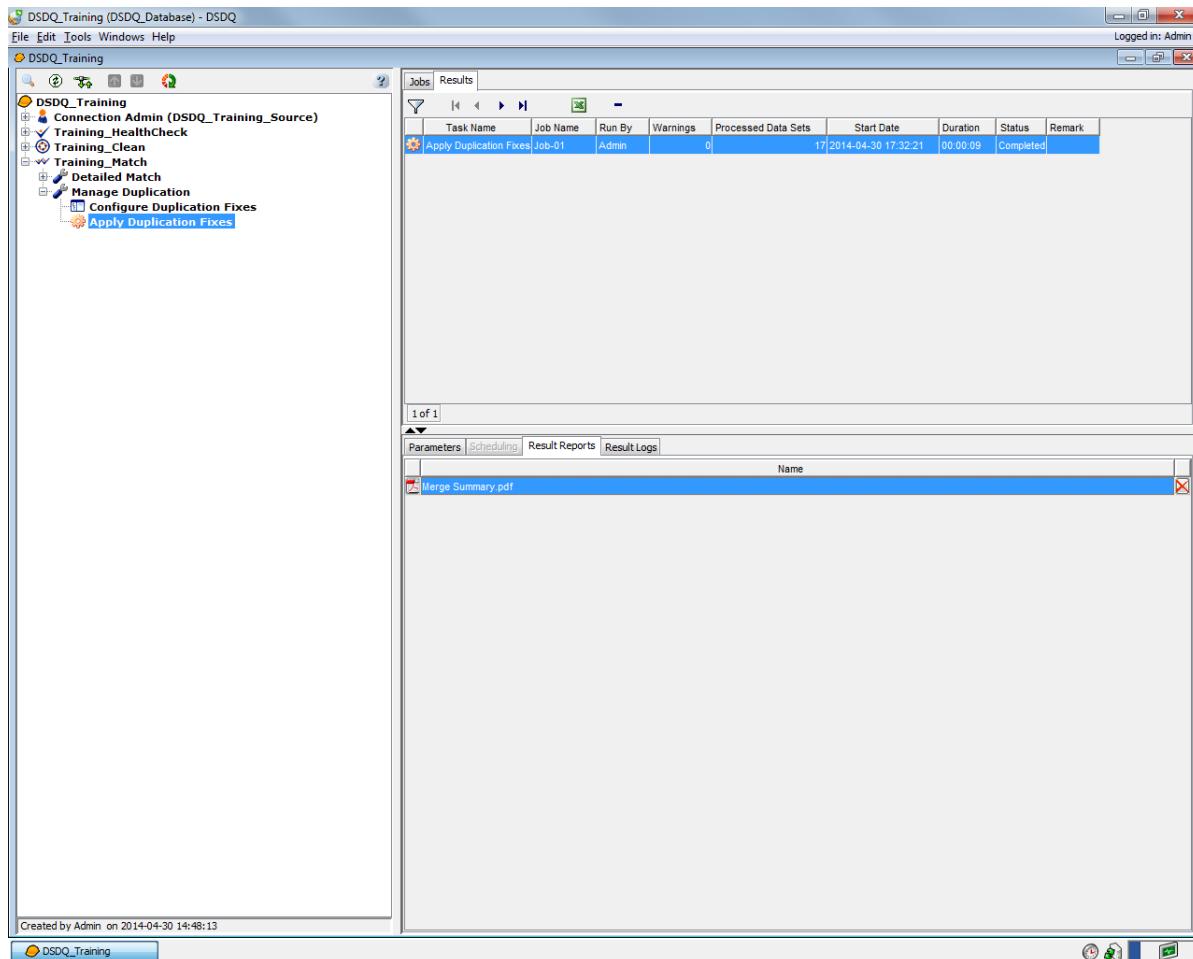


2. Enter **Job-01** in the **Job Name** field.
3. Enter **Apply Duplication Fixes** in the **Job Description** field.

4. Select **DSDQ_Training_Test (DSDQ_Training)** from the **Select Data Configuration** drop-down list.
5. Select **DSDQ_Training_Source** from the **Select Source Connection** drop-down list.
The **Source Submodel** and **Target Submodel** fields populates automatically.
6. Select the **Yes** option for **Generate Printable Reports?**
7. Select the **After** option for **Delete Results?** Set the number of days as **7**.
8. Select the **Immediately Apply Changes to the Database** option for **Data Change Action**.
9. Select the **No** option for **Enable Audit Logging**.
10. Click  to save changes in the **Parameter** tab.
11. Click .

The **Apply Duplication Fixes** Task is executed and displays results in the **Result Reports** tab.

12. 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**.



13. Click on the **Result Reports** tab to display **Duplication Fixes** results in PDF format.

Merge Summary						HALLIBURTON
Project: DSDQ_Training Phase: Training_Match Task: Manage Duplication Job: Job-01 Merge Group: DSDQ_Training_Test (DSDQ_Training) Result Date: 2014-01-07 16:10:03						Landmark Software & Services
Source Table	Target Table	Total Rows	Rows Processed	Rows Succeeded	Rows Failed	Target Table Remarks
Well	Well	1388	122	110	0	
WellEntire	WellEntire	1388	110	110	0	

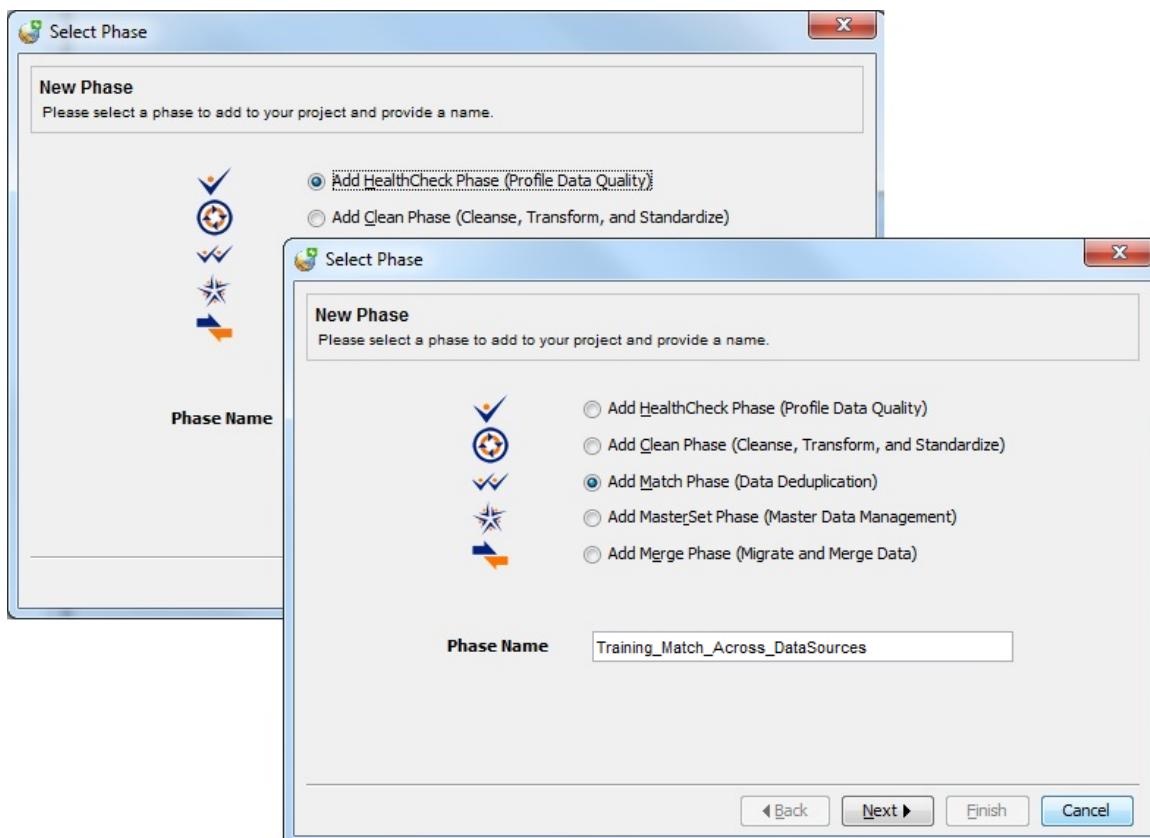
Detailed Match across Data Sources

The Detailed Match across Data Sources Activity helps in searching verifying and removing duplication across several data sources. The procedure is almost identical to the one followed for clearing out a single data source with the exception of creating a primary submodel and searchable submodels. The primary submodel contains the source data set that matches will be found for. The searchable submodels contain the data sets that matches will be found from.

Exercise: Adding a Match Phase across Data Sources

To add a Match Phase across Data Sources:

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.

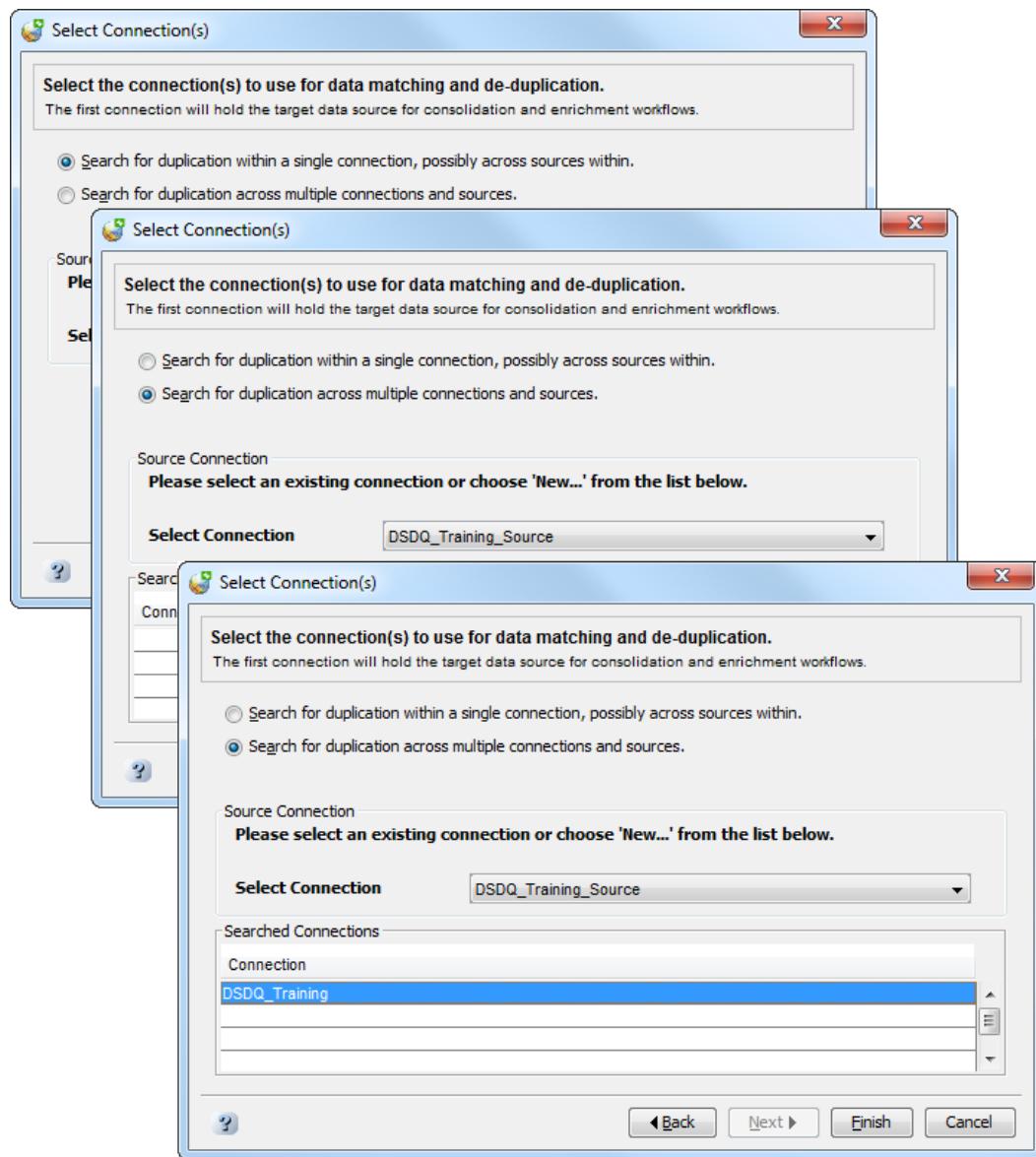


2. Select the **Add Match Phase (Data Deduplication)** option.

3. Enter **Training_Match_Across_Datasources** in the **Phase Name** field.

4. Click **Next** to continue.

The **Select Connection(s)** dialog box appears with the **Search for duplication within a single connection, possibly across sources within** option selected by default.



5. Select the **Search for duplication across multiple connections and sources** option.
6. Select **DSDQ_Training_Source** from the **Select Connection** drop-down list.

7. Click on the first empty row under the **Searched Connections** text box and select **DSDQ_Training** from the **Connections** drop-down list.
8. Click **Finish**.
The Match Phase is created and displayed in the DecisionSpace Data Quality Project Window.

Exercise: Configuring Detailed Match across Data Sources

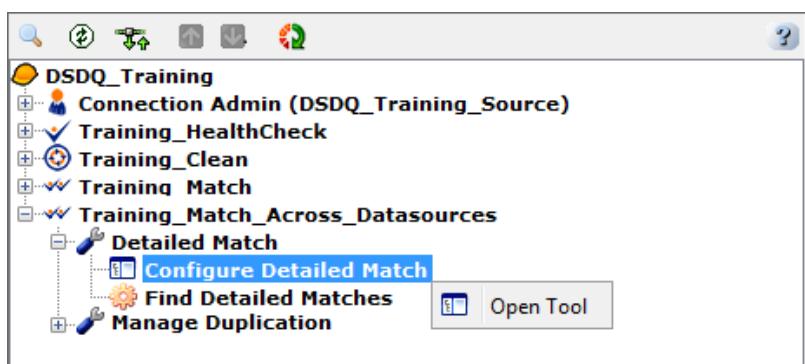
The **Configure Detailed Match** Tool is used to create Match Groups (a selection of primary/searchable submodels and their respective Merge Groups) and configure Service Levels for testing prior to running the **Find Detailed Matches** Task. The Match Groups consist of a primary submodel, searchable submodel(s), and their respective Merge Groups. The user 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 Match requirements must exist prior to opening the **Configure Detailed Match** Tool.

Note

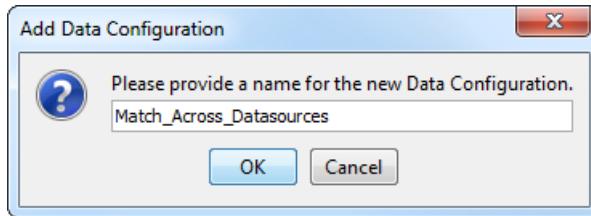
Only one Service Level can be configured at any given time.

To Configure Detailed Match across Data Sources

1. Click **+** on the DecisionSpace Data Quality Tree to expand the **Training_Match_Across_Datasources** Phase.
2. Click **+** to expand the Detailed Match Activity.
3. Double-click the **Configure Detailed Match** Tool or right-click the **Configure Detailed Match** Tool and select **Open Tool** from the pop-up menu.

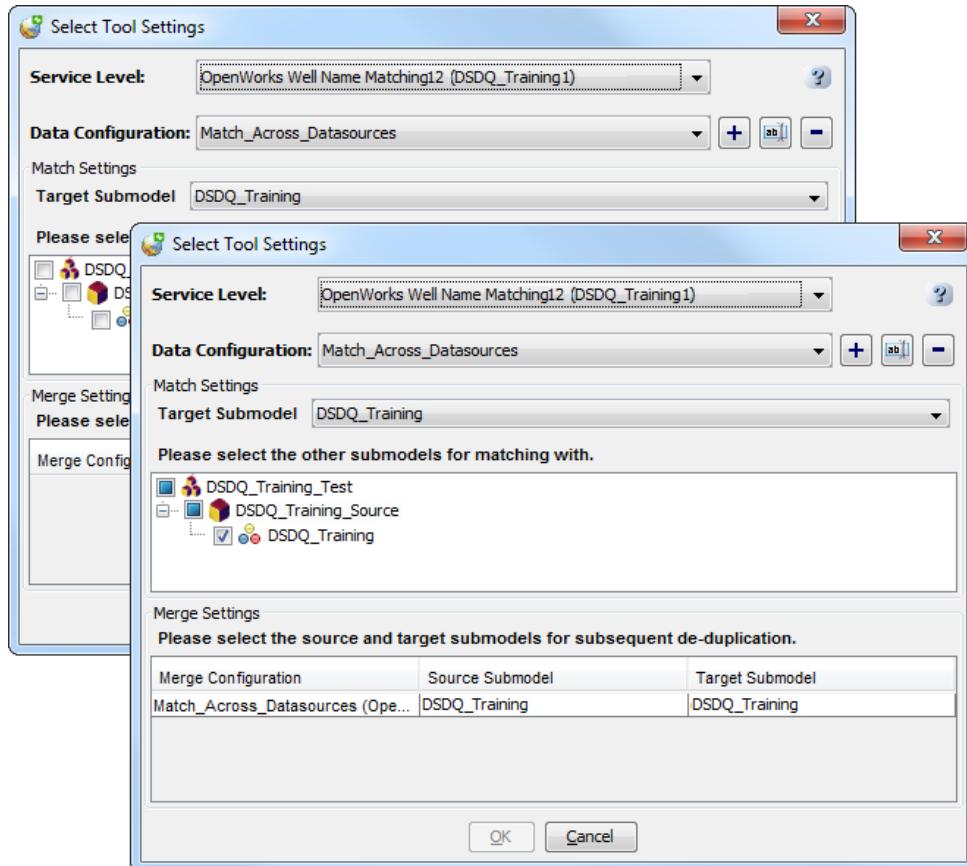


The **Configure Detailed Match** window appears. The **Add Data Configuration** dialog box appears by default the first time you run the **Configure Detailed Match** tool.



4. Enter **Match_Across_Datasources** in the **Please provide a name for the new Data Configuration** field.
5. Click **OK**.

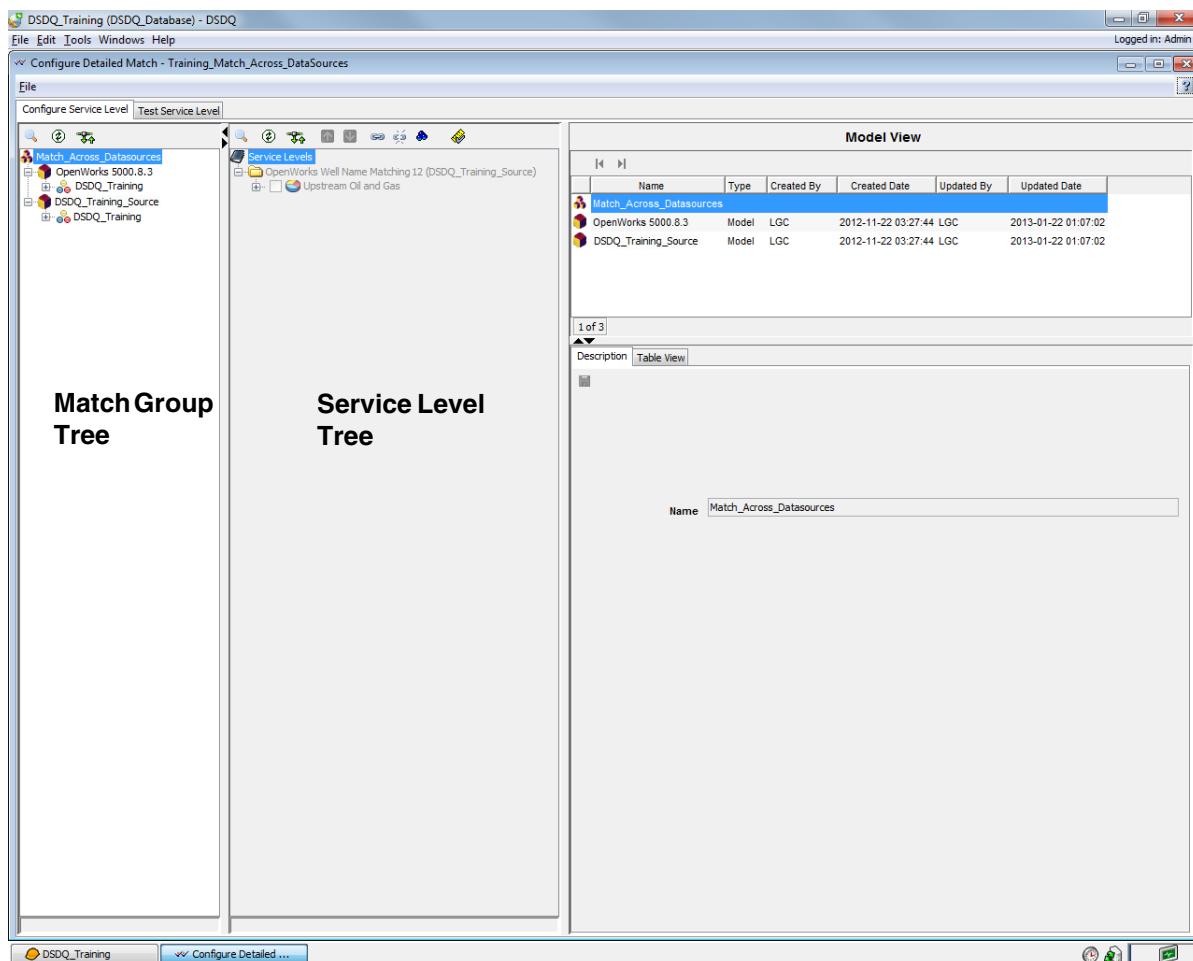
The **Select Tool Settings** window appears with **DSDQ_Training_Test** added to the **Searchable Submodel Tree**.



6. Select **OpenWorks Well Name Matching12(DSDQ_Training1)** from the **Service Level** drop-down list.

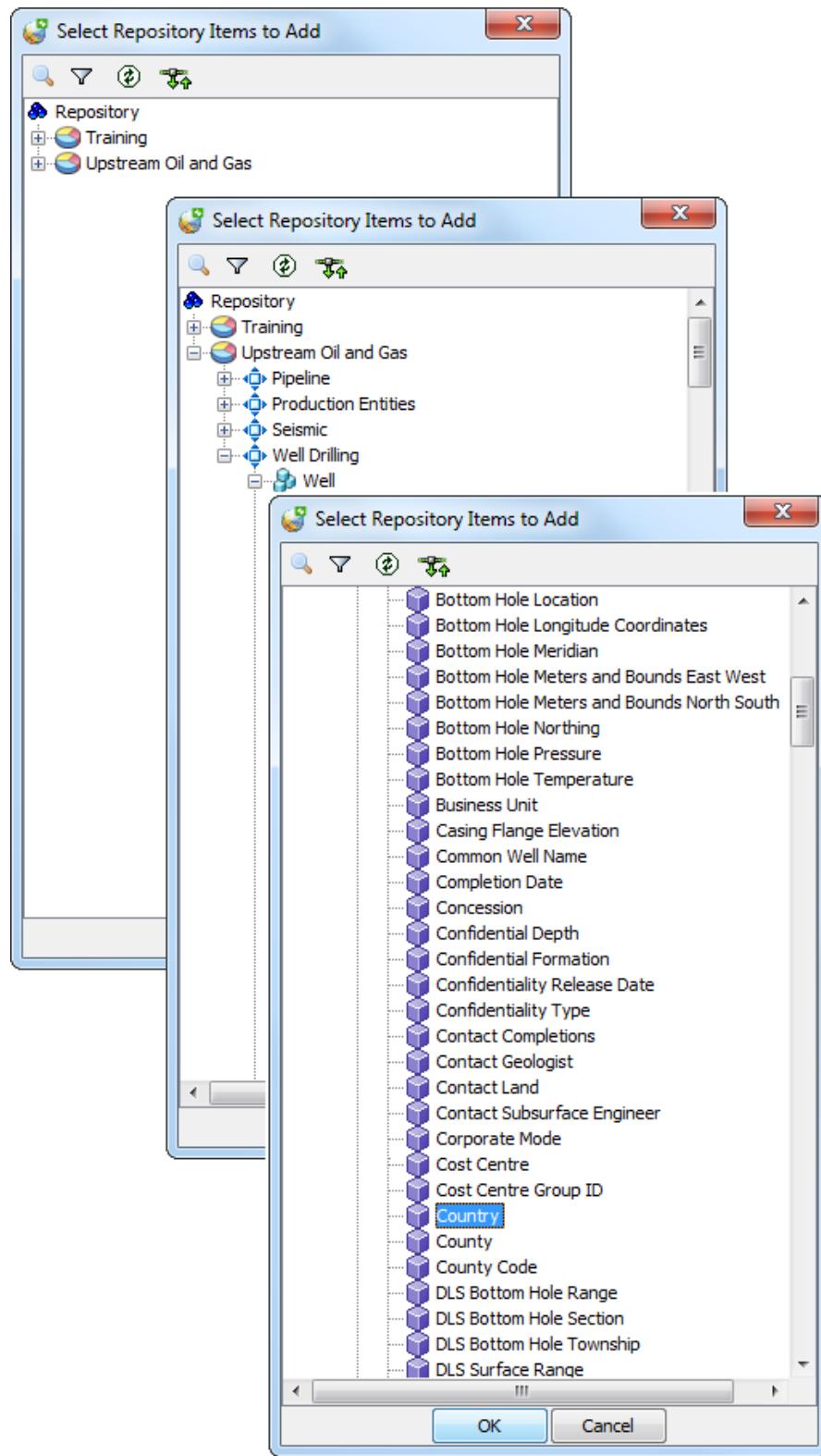
7. Select **Match_Across_Datasources** from the **Data Configuration** drop-down list.
8. Select **DSDQ_Training** from the **Target Submodel** drop-down list.
9. Select the **Match_Across_Datasources** check box from the **Searchable Submodel Tree**.
10. Click **OK**.

The **Configure Detailed Match** window appears with **DSDQ_Training** added to the **Match Group Tree**.

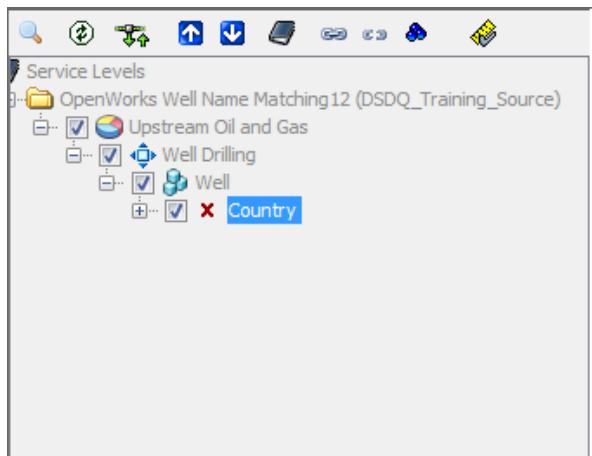


11. Click  on the Service Level Tree toolbar.

The **Select Repository Items to Add** dialog box opens.

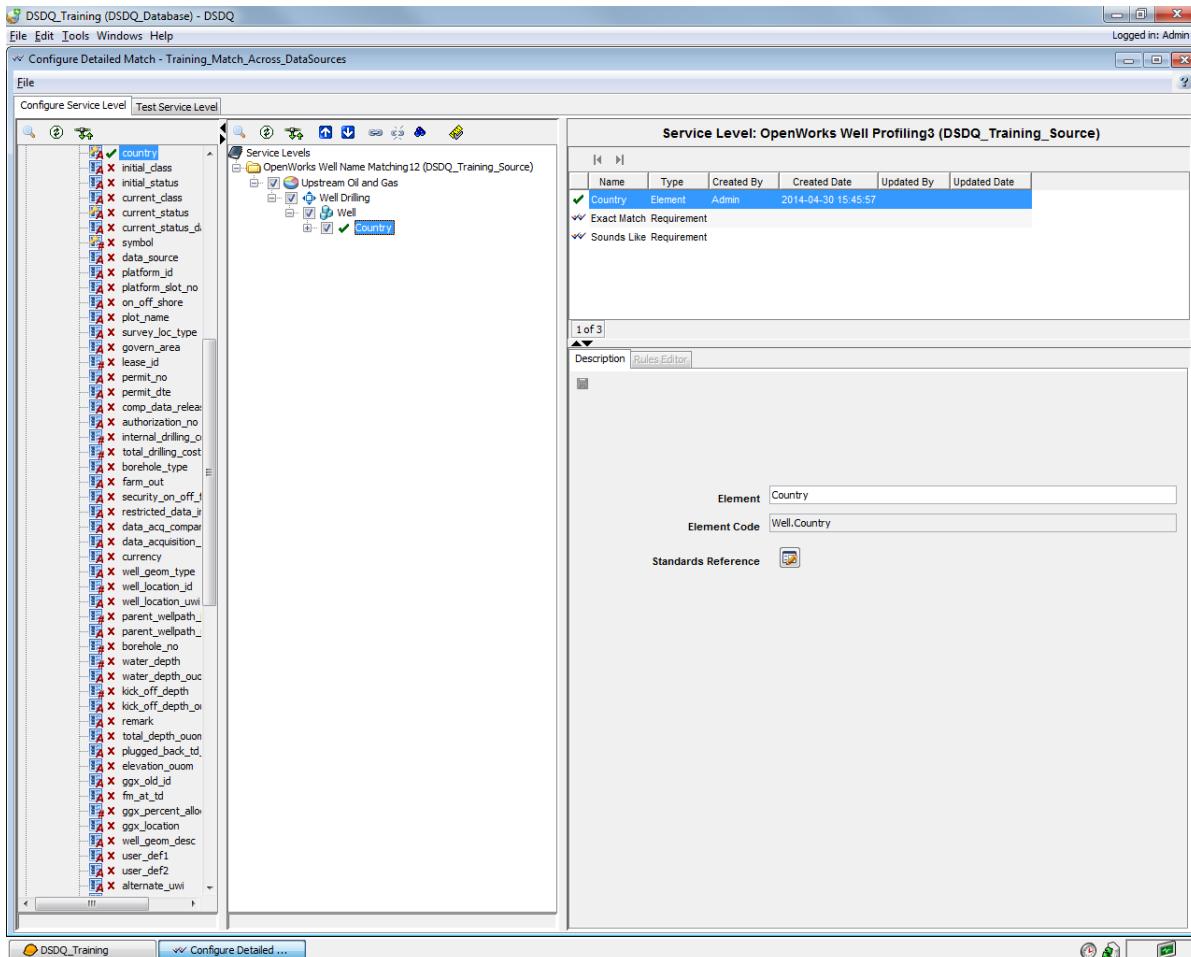


12. Click **[+]** to expand the **Upstream Oil & Gas** sector.
13. Expand the **Well Drilling** area.
14. Expand the **Well** element group, select the **Country** element and click **OK**.
The **Country** element is added to the Service Level Tree. A red cross will appear adjacent to the element in the Service Level Tree indicating that the element is not linked to any column.



15. Click **[+]** next to the **DSDQ_Training** submodel underneath the **OpenWorks 5000.8.3** model on the Match Group Data Model Tree to expand the source submodel.
16. Expand the **Well** table and select the **Country** column.
17. Drag and drop the **Country** column onto the **Country** element in the Service Level Tree.
An amber check mark will appear adjacent to the **Country** element.
18. Click **[+]** next to the **DSDQ_Training** submodel underneath the **DSDQ_Training_Source** model on the Match Group Data Model Tree to expand the searchable submodel.
19. Expand the **Well** table and select the **Country** column.
20. Drag and drop the **Country** column onto the **Country** element in the Service Level tree.
A green check mark will appear adjacent to the column and element that have just been associated. Alternatively you can select the

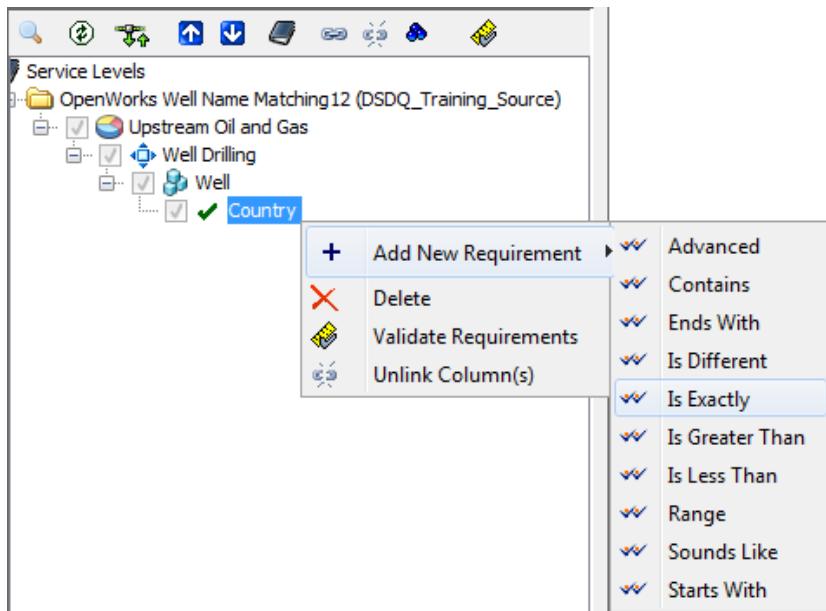
column & element to link and click the **Link Column to Element** button on the Service Level Tree toolbar.



Note

You can unlink elements or columns. Select the element or column to unlink and click the **Unlink Columns from an Element** button on the Service Level Tree toolbar.

21. Right-click on the **Country** element in the Service Level Tree and select **Add New Requirement > Is Exactly** from the pop-up menu.



The **Enter Name** dialog box appears.



22. Optionally, specify a user-defined name for the requirement.
23. Click **OK** to add the requirement to the selected element.
24. Optionally, repeat Steps **11** to **23** to add all elements for matching.

25. Click the Test Service Level tab.

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

Data Sets	Total Matches Found:	Well_Country
DSDQ_Training_Source (DSDQ_Training)	1378	Is Exactly
(a) 490250654800 (U.S. DOE,5121;43.3082115051946,-106.18832409629199 AND elev_type='KB' AND well_name='npr-3 (Well)		USA
DSDQ_Training1 (DSDQ_Training)	Matches Found: 1378	
U.S. DOE,1000 (Well)		USA
U.S. DOE,1001 (Well)		USA
U.S. DOE,1002 (Well)		USA
U.S. DOE,1003 (Well)		USA
U.S. DOE,1004 (Well)		USA
U.S. DOE,1005 (Well)		USA
U.S. DOE,1006 (Well)		USA
U.S. DOE,1007 (Well)		USA
U.S. DOE,1008 (Well)		USA
U.S. DOE,1009 (Well)		USA
U.S. DOE,1010 (Well)		USA
U.S. DOE,1011 (Well)		USA
U.S. DOE,1012 (Well)		USA
U.S. DOE,1013 (Well)		USA
U.S. DOE,1014 (Well)		USA
U.S. DOE,1015 (Well)		USA
U.S. DOE,1016 (Well)		USA
U.S. DOE,1017 (Well)		USA
U.S. DOE,1018 (Well)		USA
U.S. DOE,1019 (Well)		USA
U.S. DOE,1020 (Well)		USA
U.S. DOE,1021 (Well)		USA
U.S. DOE,1022 (Well)		USA
U.S. DOE,1023 (Well)		USA
U.S. DOE,1024 (Well)		USA
U.S. DOE,1025 (Well)		USA
U.S. DOE,1026 (Well)		USA
U.S. DOE,1027 (Well)		USA
U.S. DOE,1028 (Well)		USA
U.S. DOE,1029 (Well)		USA
U.S. DOE,1030 (Well)		USA
U.S. DOE,1031 (Well)		USA
U.S. DOE,1032 (Well)		USA
U.S. DOE,1033 (Well)		USA
U.S. DOE,1034 (Well)		USA
U.S. DOE,1035 (Well)		USA
U.S. DOE,1036 (Well)		USA
U.S. DOE,1037 (Well)		USA
U.S. DOE,1038 (Well)		USA
U.S. DOE,1039 (Well)		USA
U.S. DOE,1040 (Well)		USA

26. Verify all entries have correct matches; in this case the correct match is “USA”.

27. Click the Next Data Set button to test the next record.

28. Repeat steps **27** to test all records.

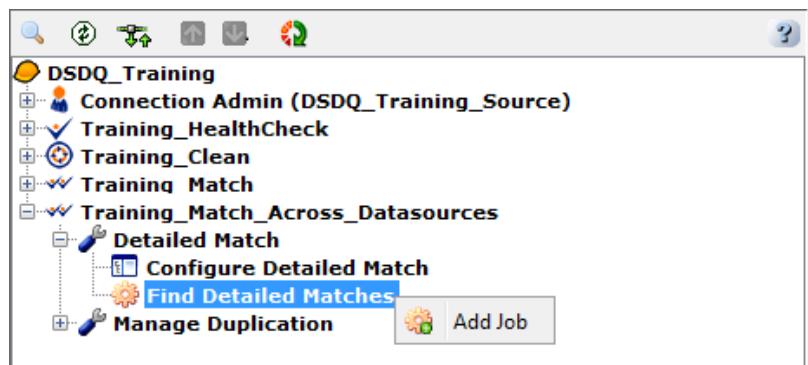
29. Select **File > Exit** from the menu bar on the **Configure Detailed** window.

Exercise: Finding Detailed Matches across Data Sources

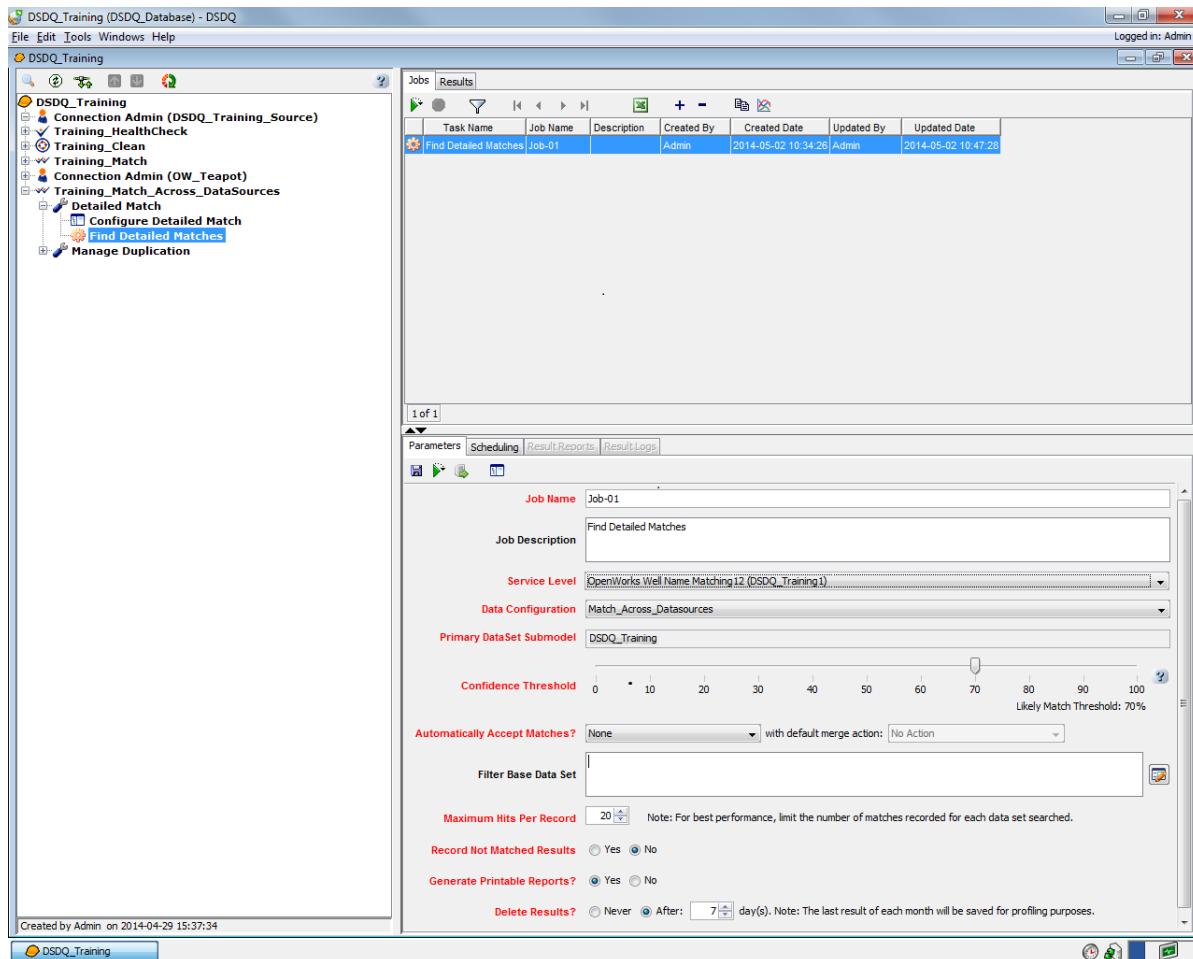
The **Find Detailed Matches** task looks for matching values within the specified columns using the associated match requirements.

To find Detailed Matches across Data Sources:

1. Double-click the **Find Detailed Matches** Task or right-click the **Find Detailed Matches** task and select **Add Job** from the pop-up menu.



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

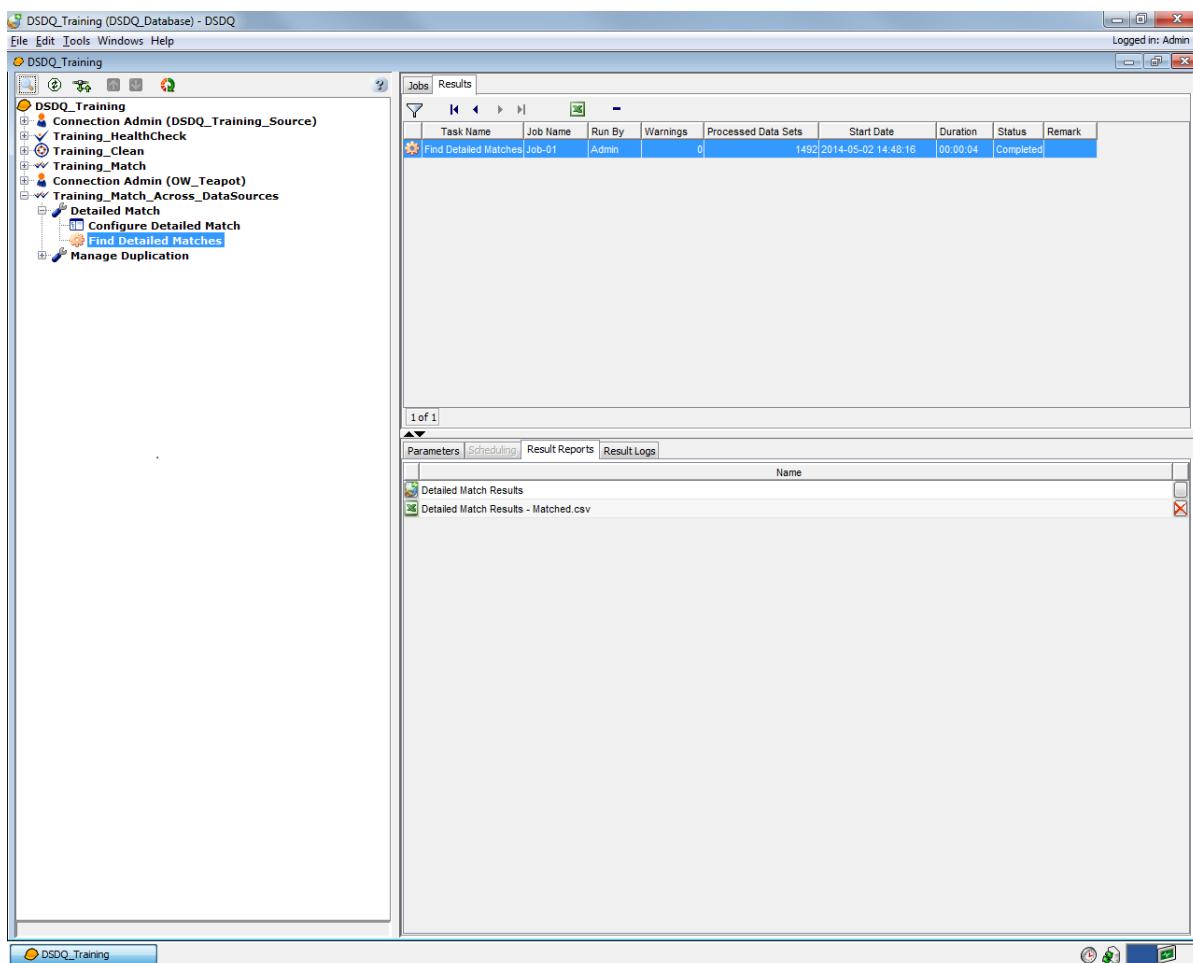


2. Enter **Job-01** in the **Job Name** field.
3. Enter **Find Detailed Matches** in the **Job Description** field.
4. Select **OpenWorks Well Name Matching12 (DSDQ_Training1)** from the **Service Level** drop-down list.
5. Select **Match_Across_Datasources** from the **Data Configuration** drop-down list.
The **Primary DataSet Submodel** field populates automatically.
6. Set the **Confidence Threshold** option as **70**.
7. Select **None** from the **Automatically Accept Matches** drop-down list.
8. Optionally, set a filter on the data subset.

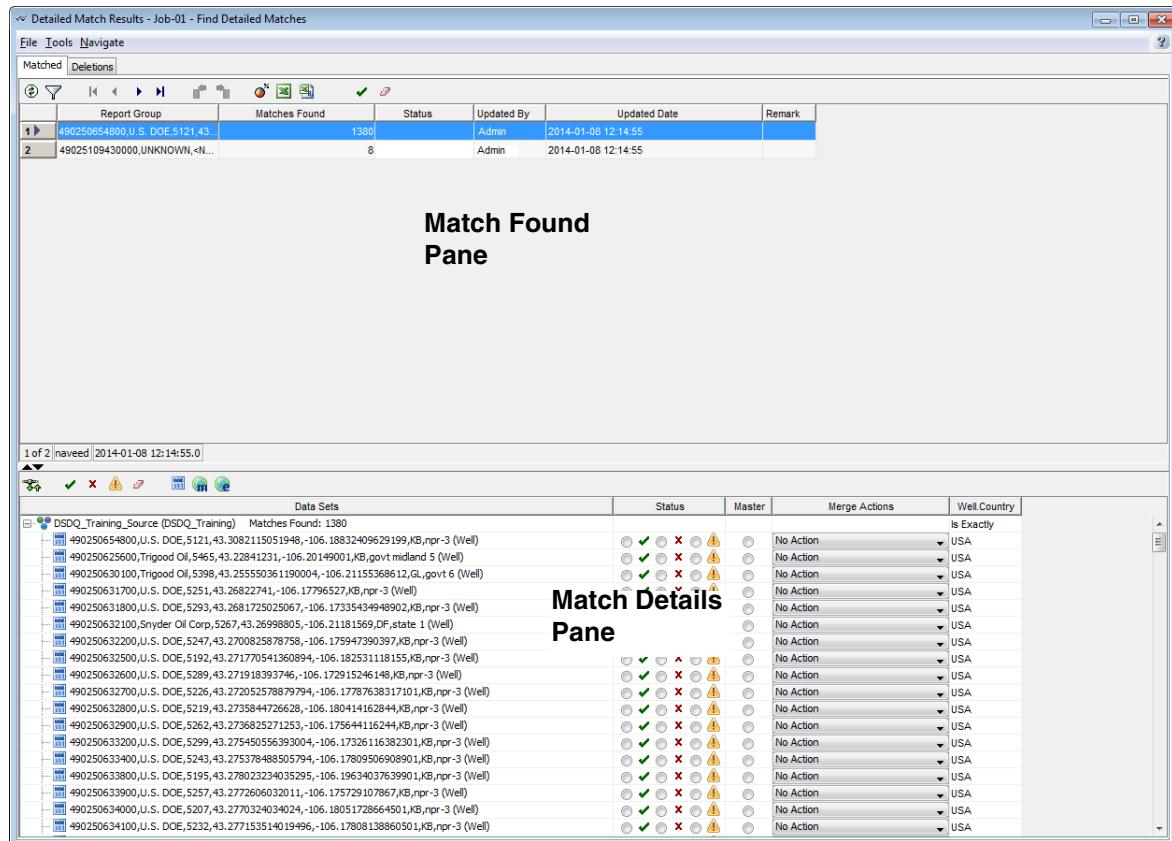
9. Set the **Maximum Hits Per Record** option as **20**.
10. Select the **No** option for **Record Not Matched Results**.
11. Select the **Yes** option for **Generate Printable Reports?**
12. Select the **After** option for **Delete Results?** Set the number of days as **7**.
13. Click  to save changes in the **Parameters** tab.
14. Click .

The **Find Detailed Matches** task runs and displays results in the **Result Reports** tab on the **Jobs 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. Double-click **Detailed Match Results** on the **Result Reports** tab.
The **Detailed Match Results** window appears.



17. Select all the rows in the **Matches Details** Pane and click on the **Matches Found** toolbar.
18. Optionally, select the data and click the **Show on Browser Map** button on the toolbar to display the location of the data set on a map in the **DecisionSpace Data Quality Dashboard**. You can also display the data in a Pie chart by click the **Generate Status Chart** button.
19. Select **File > Exit** from the menu bar on the **Detailed Match Results** window.

Manage Duplication across Data Sources

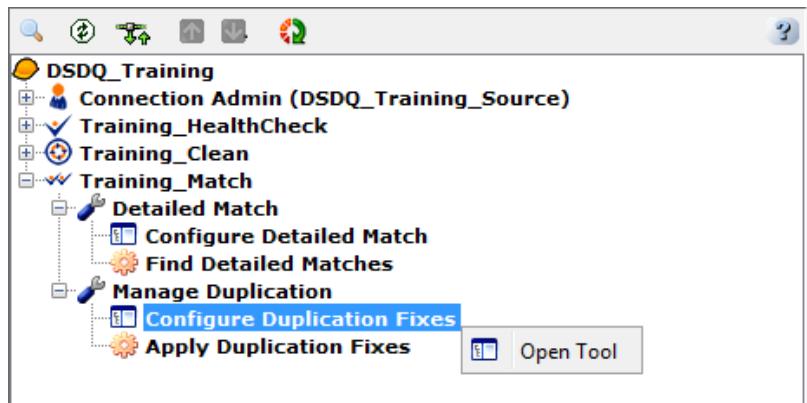
The Manage Duplication across Data Sources Activity helps in removing duplication from multiple data sources. Submodels are configured to make the process straightforward while meeting standards of duplication removal.

Exercise: Configuring Duplication Fixes across Data Sources

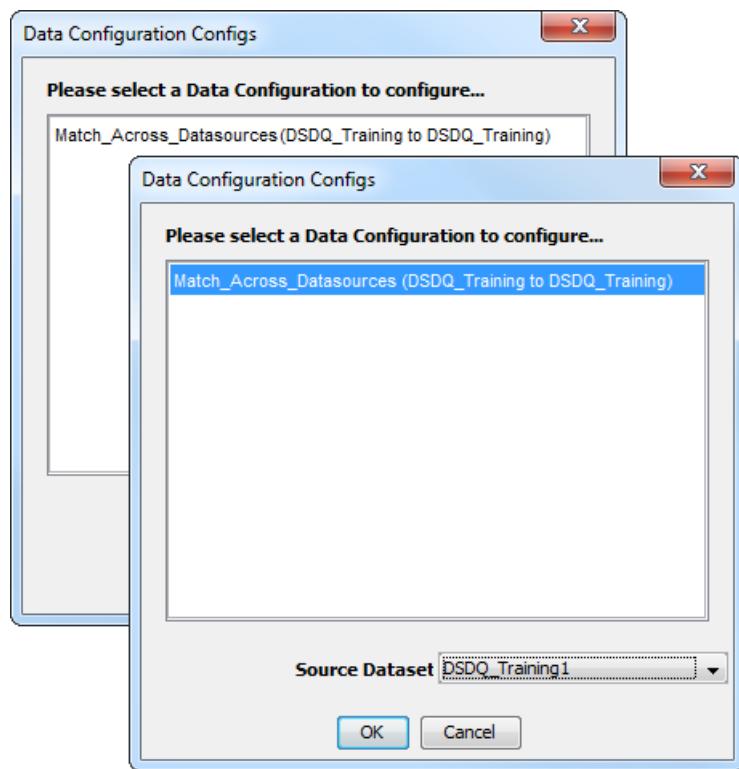
The **Configure Duplication Fixes** Tool enables you to set up table and column mappings for the selected **Merge Group**. The subset of data can then be selected for testing before running it on the actual data set. The interface also enables users to apply rules in order to modify data before moving it to the target database.

To Configure Duplication Fixes across Data Sources:

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



The **Data Configuration Configs** window appears.

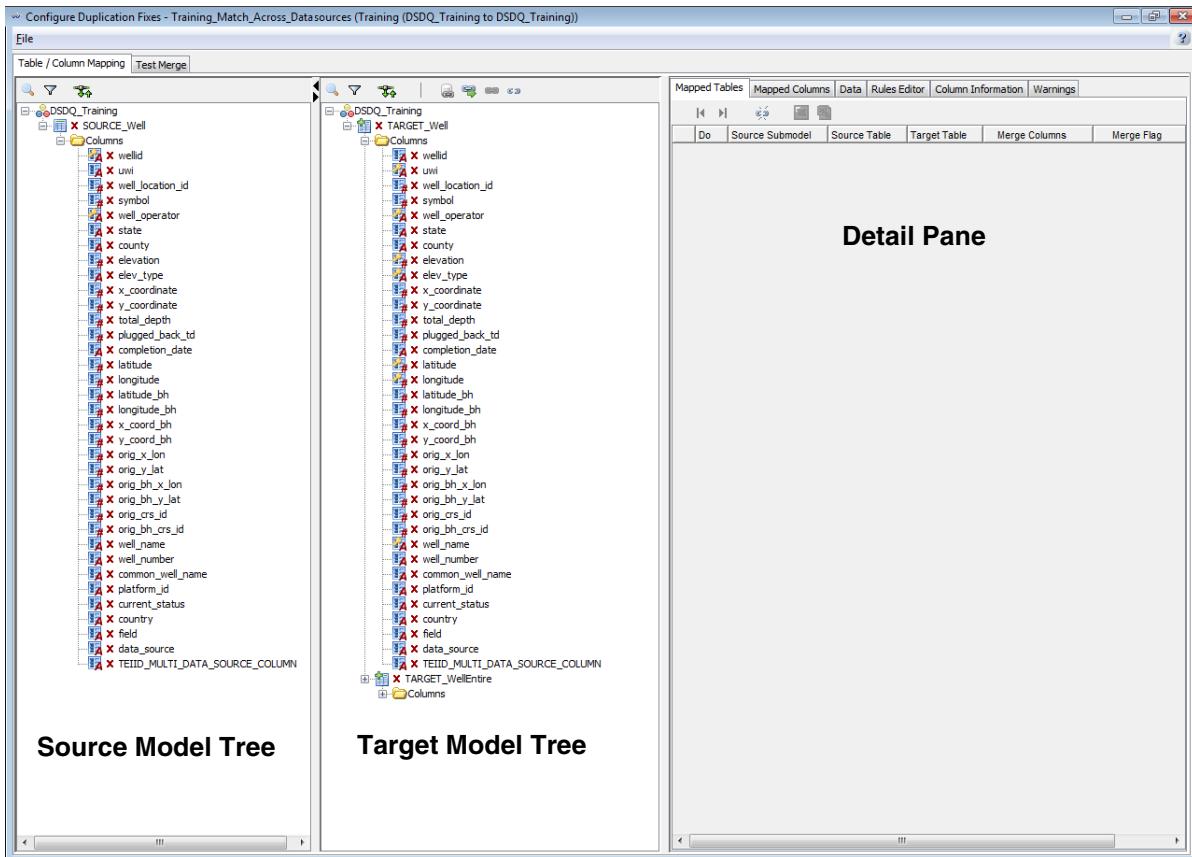


3. Select the **Match_Across_Datasources (DSDQ_Training to DSDQ_Training)** option.
4. Click **OK**.

The **Configure Duplication Fixes** window displays, displaying tables and columns for the Source Model Tree, Target Model Tree and the Detail pane. The Detail Pane has six tabs:

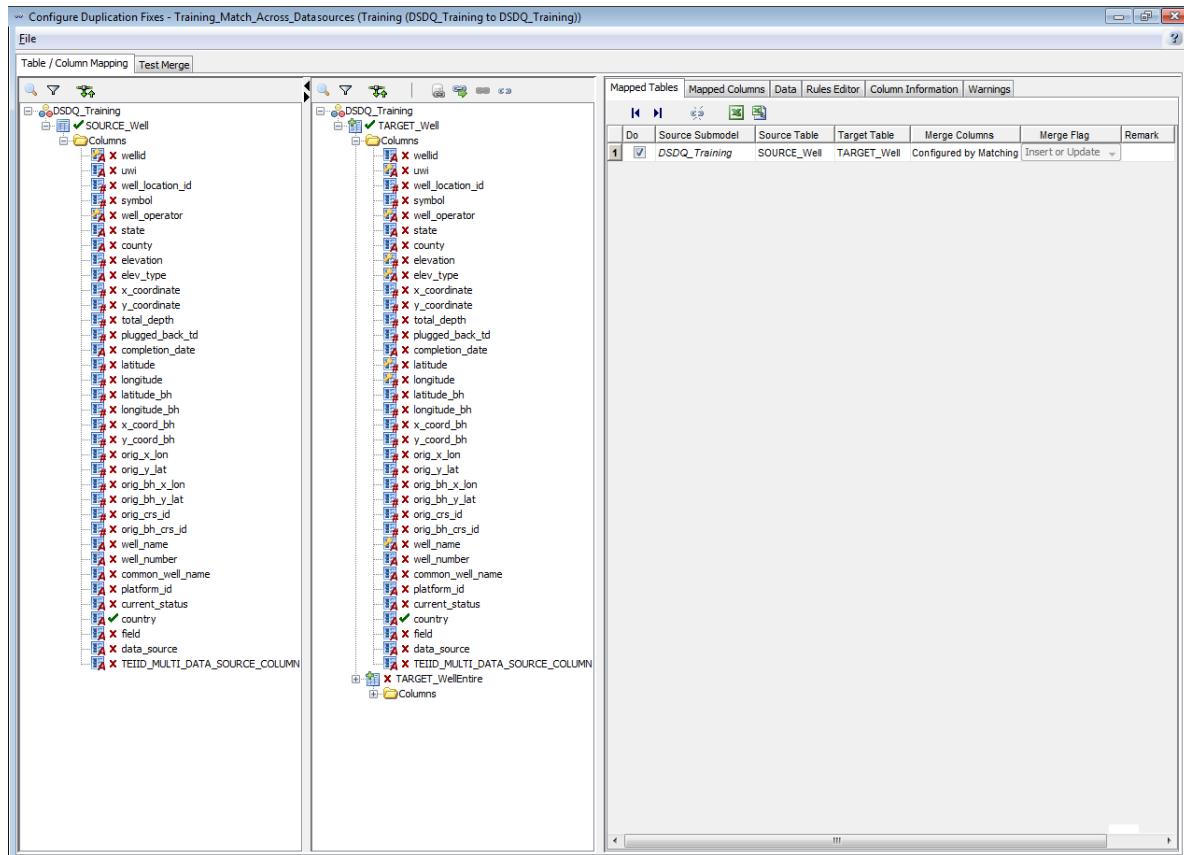
Mapped Tables	This area displays information about the tables that have been mapped in the source and target trees.
Mapped Columns	This area displays the mapping between the target column and all its mapped sources.
Data	<p>This area displays the source and target tables' data. Selecting a column in 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>The Source Data View and Target Data View toolbars can be used to Filter Data , to move to the First Record in Current View  or Last Record in Current View , or to move to the Next Data Set  or Previous Data Set  . The data can also be saved in Excel format by clicking on the corresponding button  or exported to a CSV file by selecting the Create a CSV Export File  button on the toolbar.</p> <p>Data in the Data Detail Pane can be sorted by clicking a column header. Column information on any column can be viewed by right-clicking a column header and selecting Column Info from the pop-up menu. Other columns information can be viewed by right-clicking a column header and selecting Columns Filter from the pop-menu.</p>
Rules Editor	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.
Column Information	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.

Warnings	This area will display any inconsistencies between the mapped columns, e.g. source column length is 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.
-----------------	---

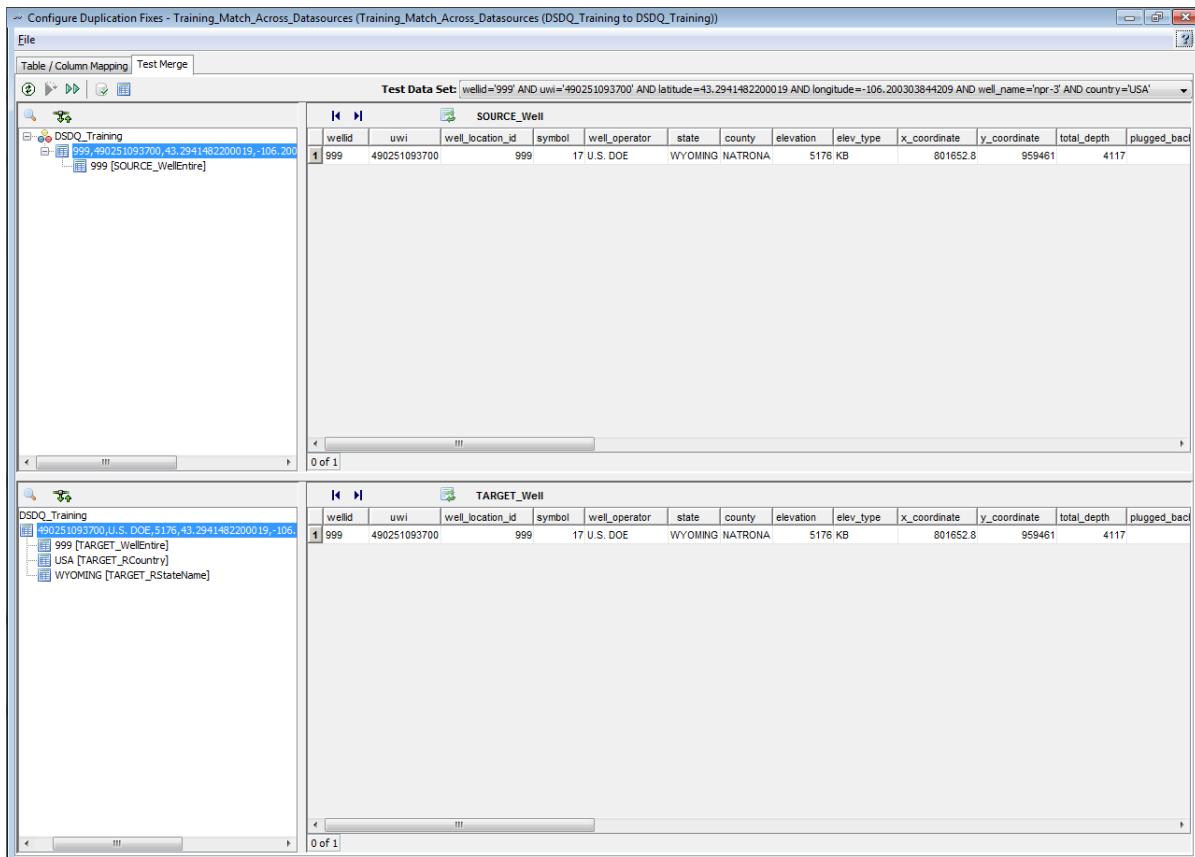


5. Select the **Source_Well** table from the Source Model Tree.
6. Select the **Target_Well** table from the Target Model Tree.
7. Click the **Auto Map selected Table/Columns** button on the toolbar.
A green check mark appears adjacent to the selected tables.
8. Select **Country** column from the Source_Well & Target_Well table.
9. Click the **Auto Map selected Table/Columns** button on the toolbar.

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



10. Select the **Test Merge** tab to test all match result configurations.



11. Select **File > Exit** from the menu bar on the **Configure Duplication Fixes** window.

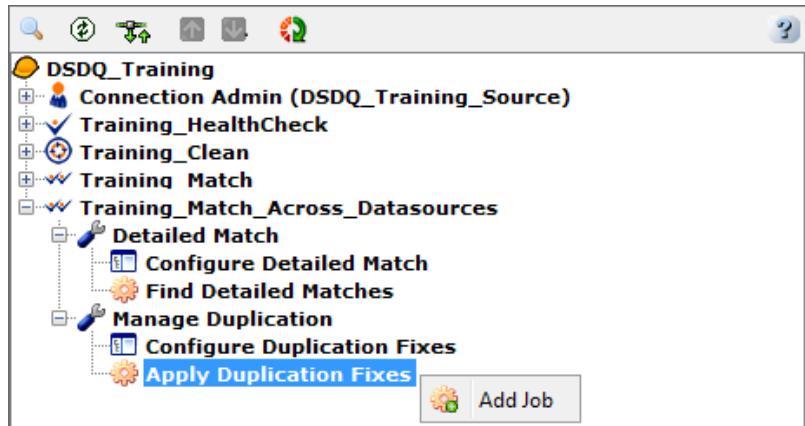
Exercise: Applying Duplication Fixes across Data Sources

After the duplications have been tested in the **Configure Duplication Fixes** Tool, the **Manage Duplication** Activity and **Apply Duplication Fixes** Task is used to move the complete data set over to the target database.

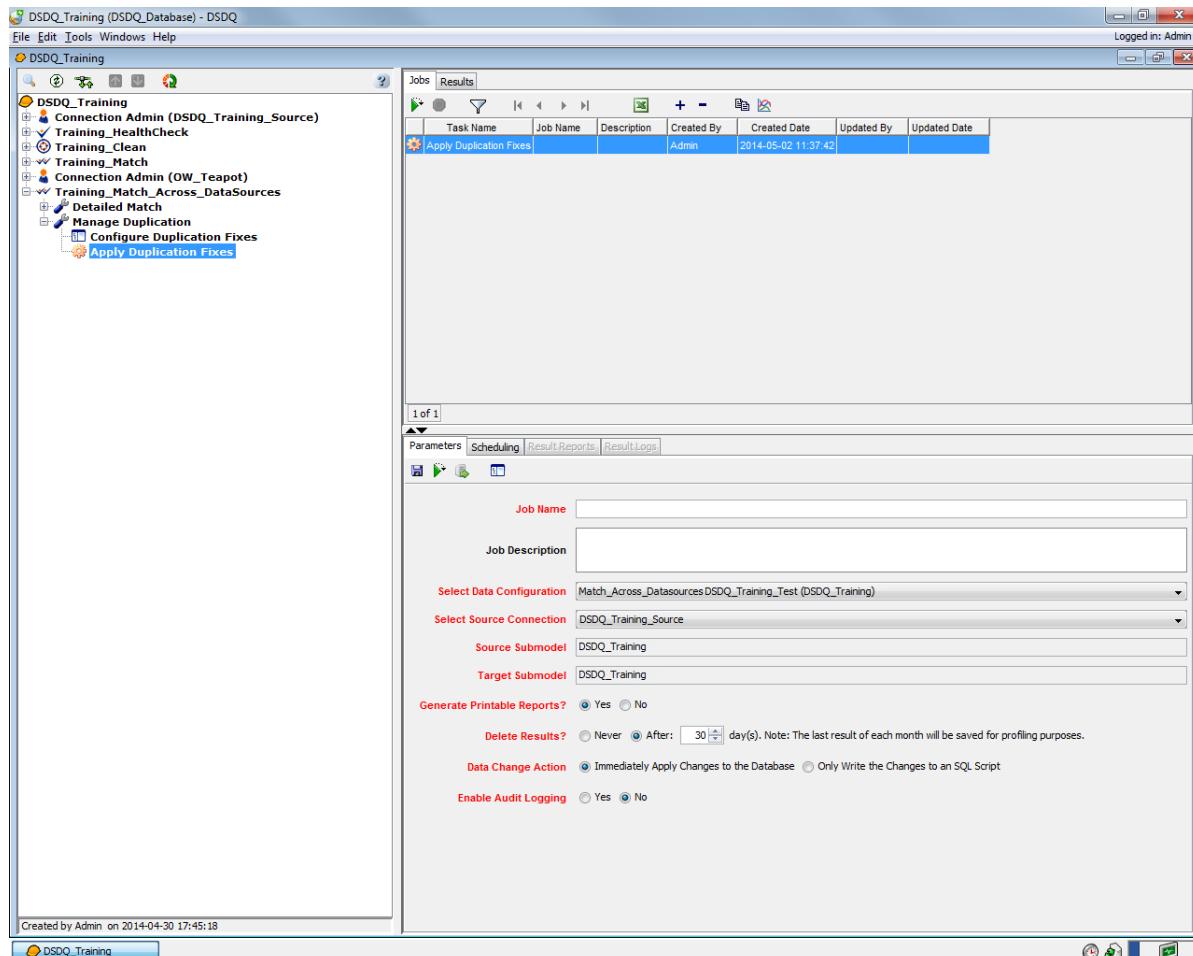
To Apply Duplication Fixes across Data Sources:

1. Double click the **Apply Duplication Fixes** Task or right-click the **Apply Duplication Fixes** task and select **Add Job** from the pop-up

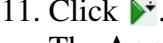
menu.



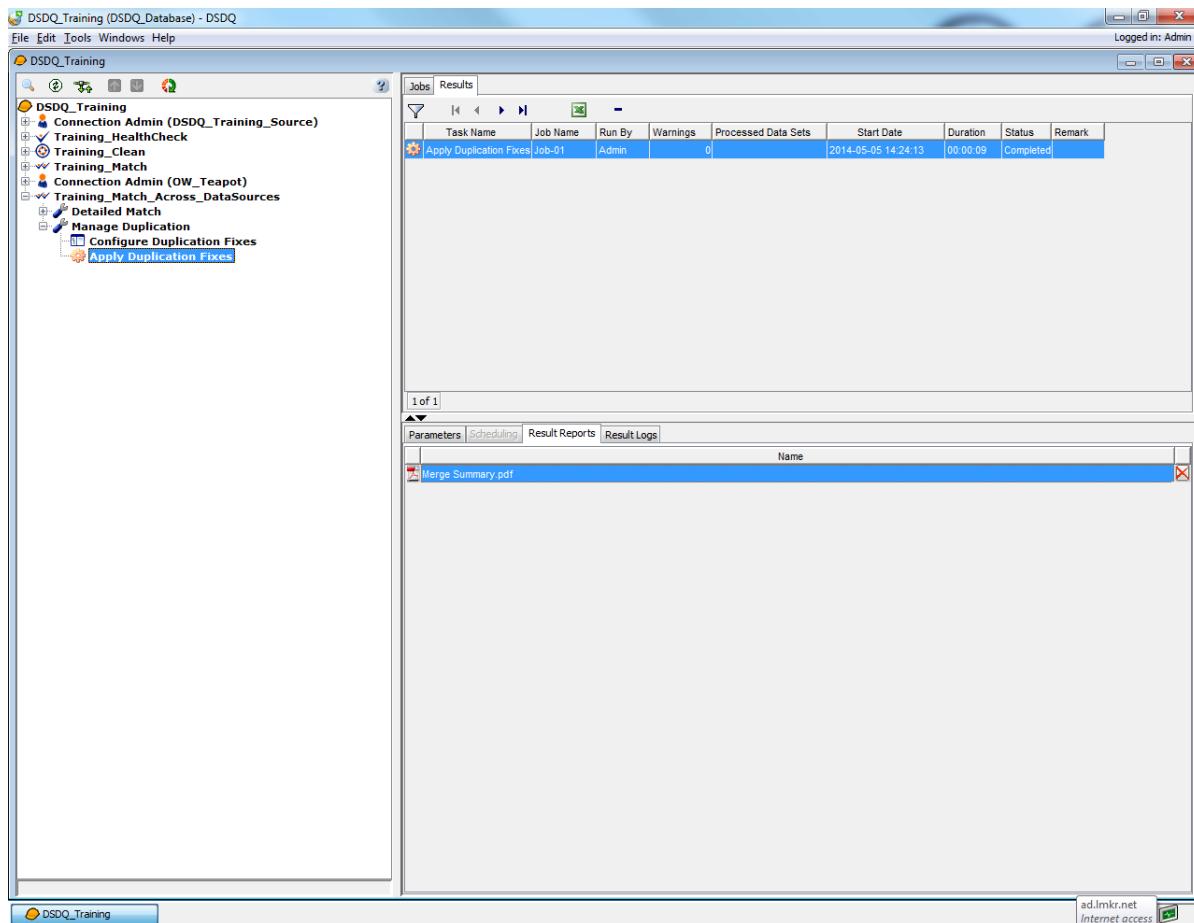
A new job is initiated and displays on the **Jobs and Results Listing Pane**.



2. Enter **Job-01** in the **Job Name** field.

3. Enter **Apply Duplication Fixes** in the **Job Description** field.
4. Select **Match_Across_Datasources DSDQ_Training_Test (DSDQ_Training)** from the **Select Data Configuration** drop-down list.
5. Select **DSDQ_Training_Source** from the **Select Source Connection** drop-down list.
The **Source Submodel** and **Target Submodel** fields populates automatically.
6. Select the **Yes** option for **Generate Printable Reports?**
7. Select the **After** option for **Delete Results?** Set the number of days as **7**.
8. Select the **Immediately Apply Changes to the Database** option for **Data Change Action**.
9. Select the **No** option for **Enable Audit Logging**.
10. Click  to save changes in the **Parameter** tab.
11. Click .
The **Apply Duplication Fixes** Task is executed and displays results in the **Result Reports** tab.

12. 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**.



13. Click on the **Results Reports** tab to display Duplication Fixes results in PDF format.

Merge Summary						HALLIBURTON
						Landmark Software & Services
Project:	DSDQ_Training					
Phase:	Training_Match_Across_Datasources					
Task:	Manage Duplication					
Job:	Job-01					
Merge Group:	Training (DSDQ_Training to DSDQ_Training)					
Result Date:	2014-01-09 09:48:00					
Source Table	Well	Target Table	Well	Total Rows	Rows Processed	Rows Succeeded
				1388	8	8
					0	0
						Target Table Remarks

Chapter 7

Data Consolidation in DecisionSpace

Data Quality

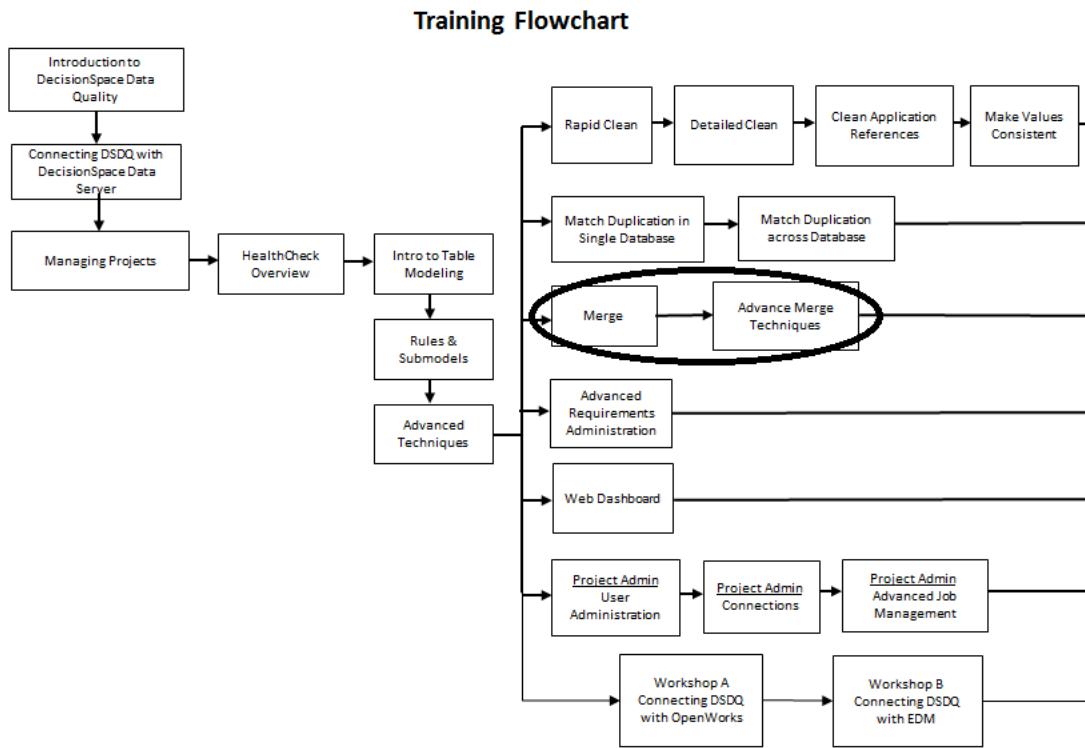
The Merge Phase enables you to consolidate data that is fit to be included in your dataset. Merge can apply individual policies to choose correct business values or transform data, and output the data to different databases or data formats.

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

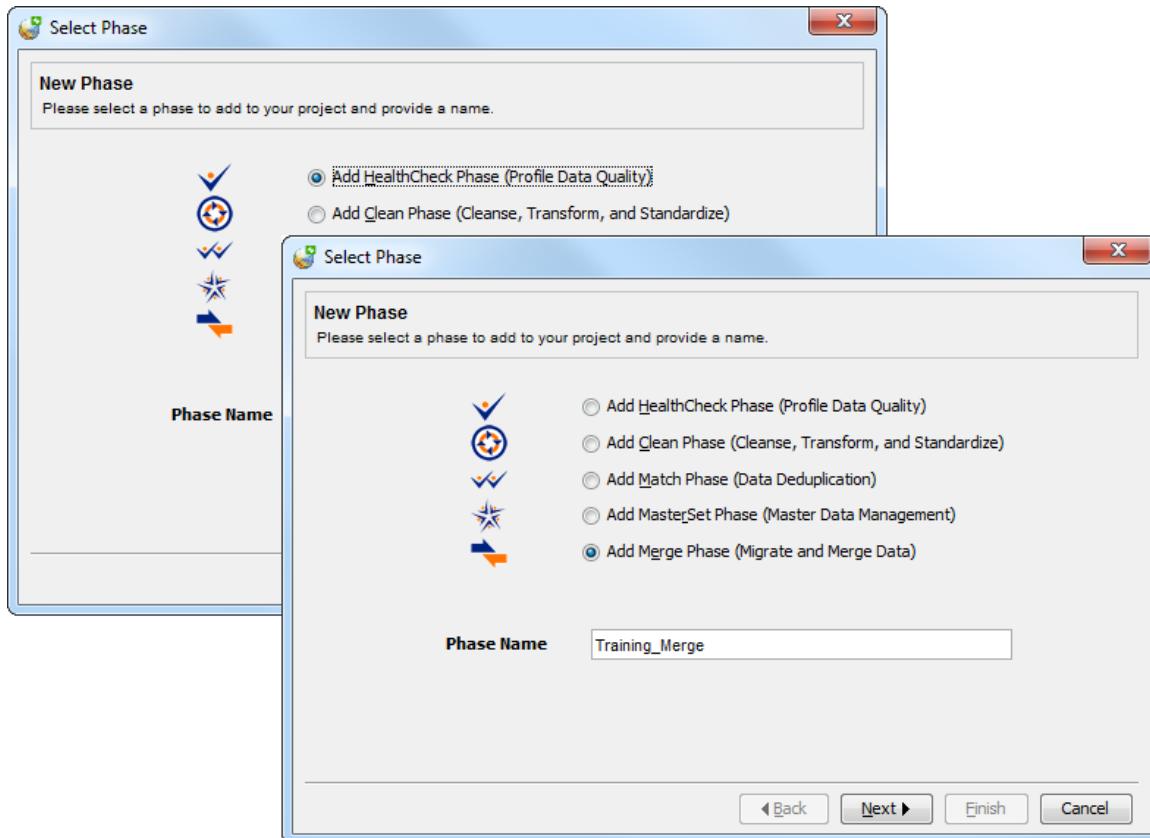
Once your data has passed through the various Phases of DSDQ and cleaned as per applicable policies, it needs to be consolidated in one database. The **Merge** module enables you to migrate/merge data in one data set. This also ensures that bad data is not loaded during the merge Phase.

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:

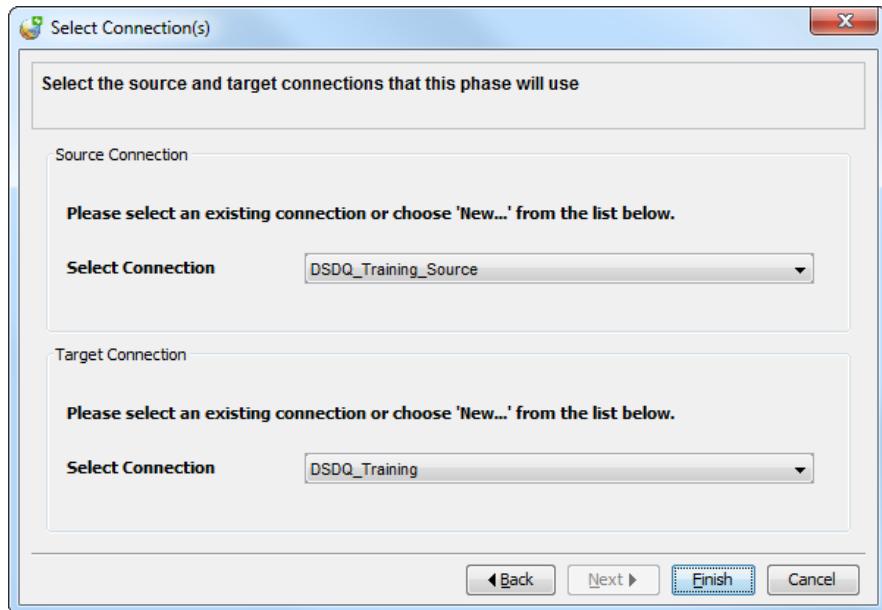
1. Click the **Add New Phase**  button on the project toolbar.
The **Select Phase** window appears with the **Add HealthCheck Phase (Profile Data Quality)** option selected by default.



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.

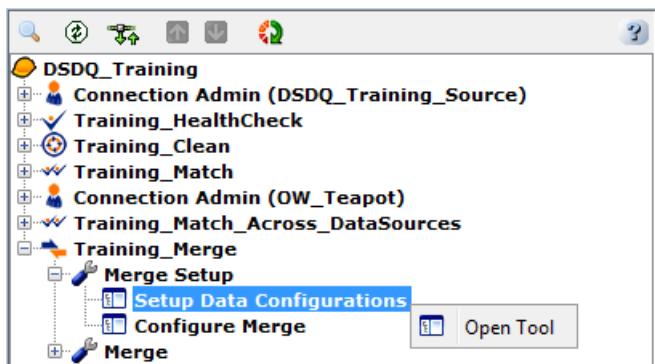
You can create, edit and delete Merge groups with the **Setup Data Configurations** Tool. The **Configure Merge** Tool allows you to map tables and columns.

Exercise: Adding a Merge Group

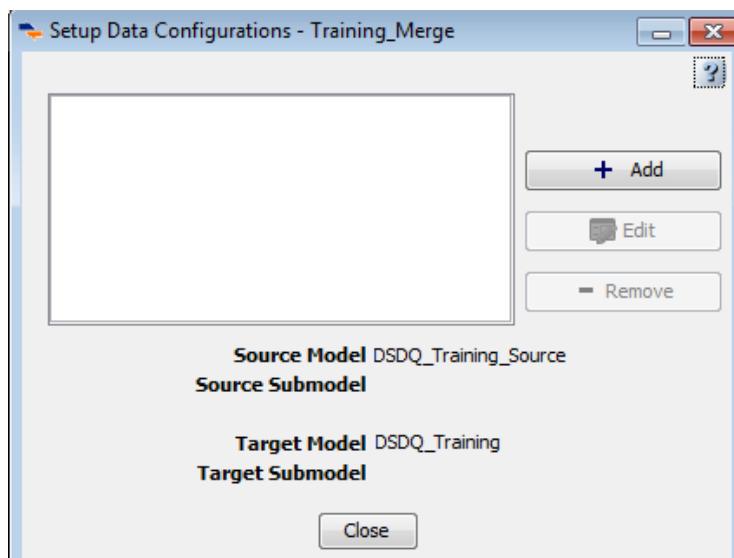
Merge groups allow you to effectively manage different merge configurations. Any settings that you want to apply to the source and target data are contained in Merge groups. These groups can be added from the **Setup Data Configurations** Tool. You will need to enter a name for the group, select a Source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

To add a Merge Group:

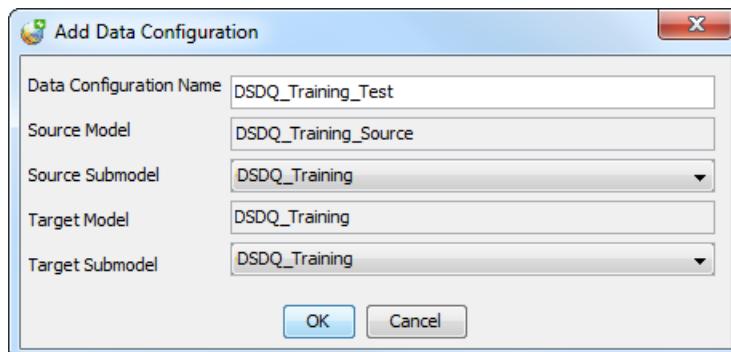
1. Click  on the DecisionSpace Data Quality Tree to expand the **Training_Merge** Phase.
2. Click  to expand the **Merge Setup** Activity.
3. Double-click the **Setup Data Configurations** Tool or right-click the **Setup Data Configurations** Tool and select **Open Tool** from the pop-up menu



The **Setup Data Configurations** window appears.



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



- 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.

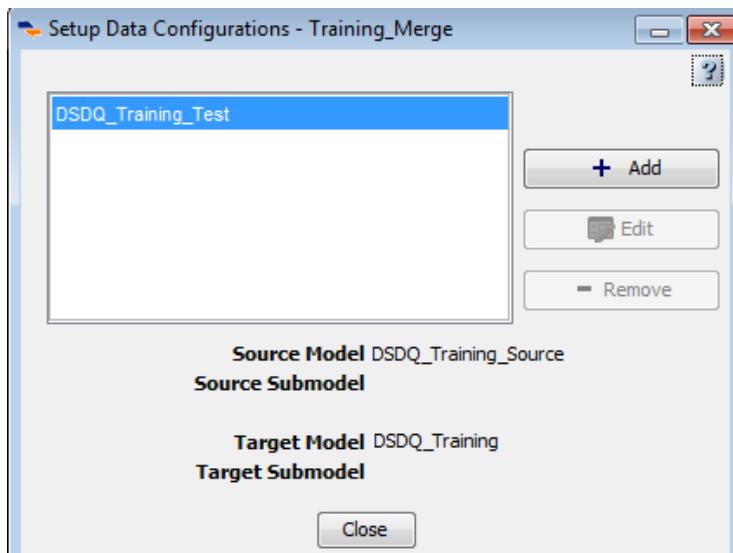
6. Select **DSDQ_Training** from the **Source Submodel** drop-down list.

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

7. Select **DSDQ_Training** from the **Target Submodel** drop-down list.

8. Click **OK**.

The **DSDQ_Training_Test** is added and displays in the **Setup Data Configurations** window.



9. 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.

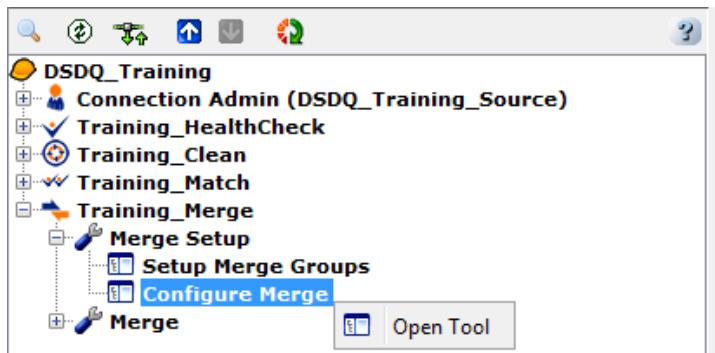
Exercise: Using the Configure Merge Tool

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.

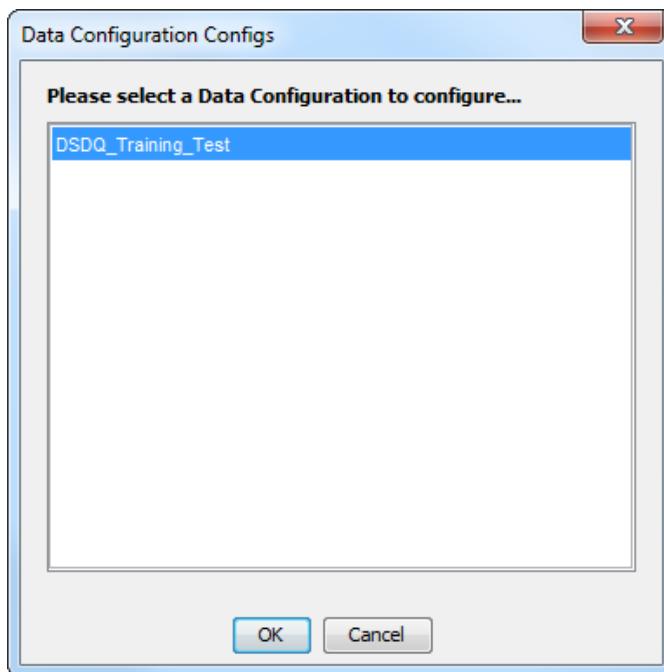
To use the Configure Merge Tool:

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.

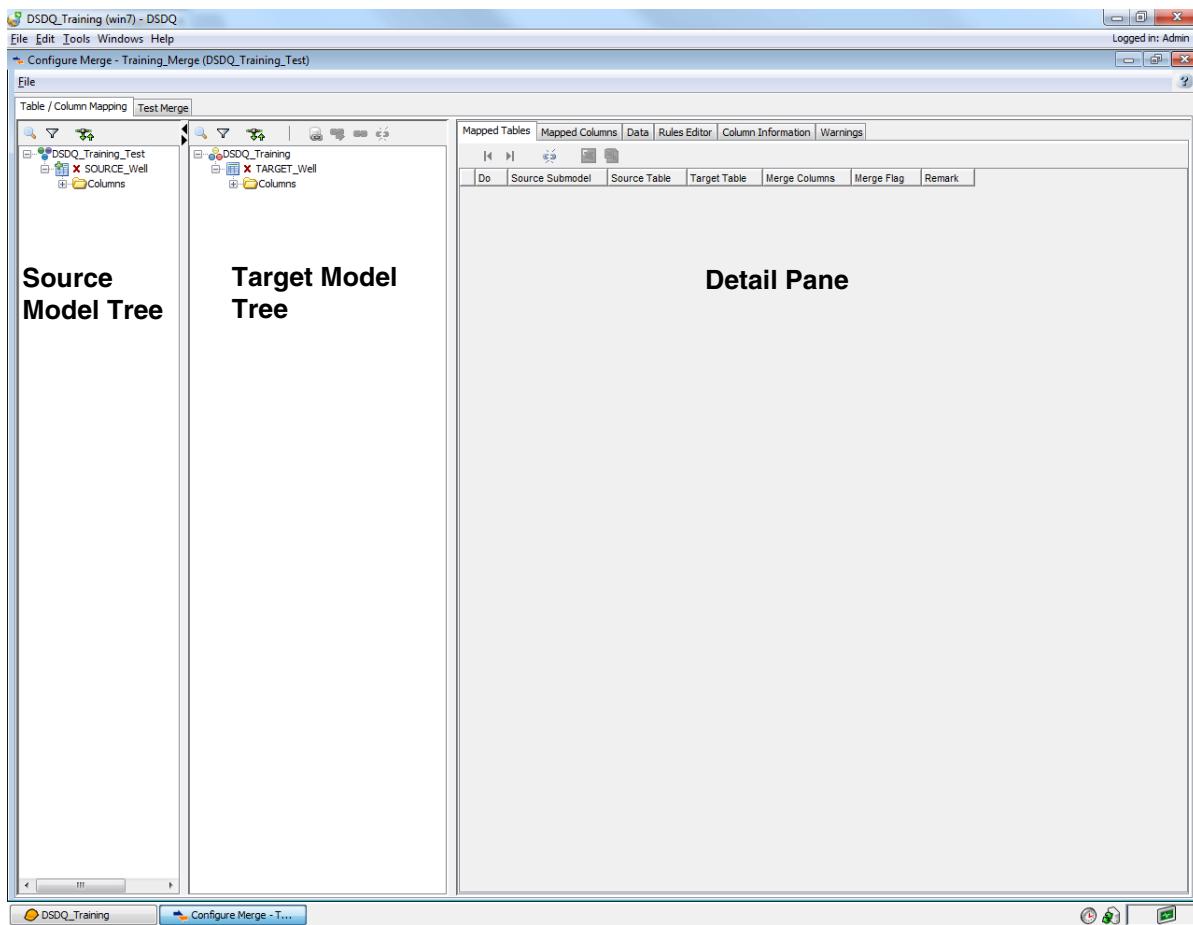


2. Select the **DSDQ_Training_Test** option.

3. Click **OK**.

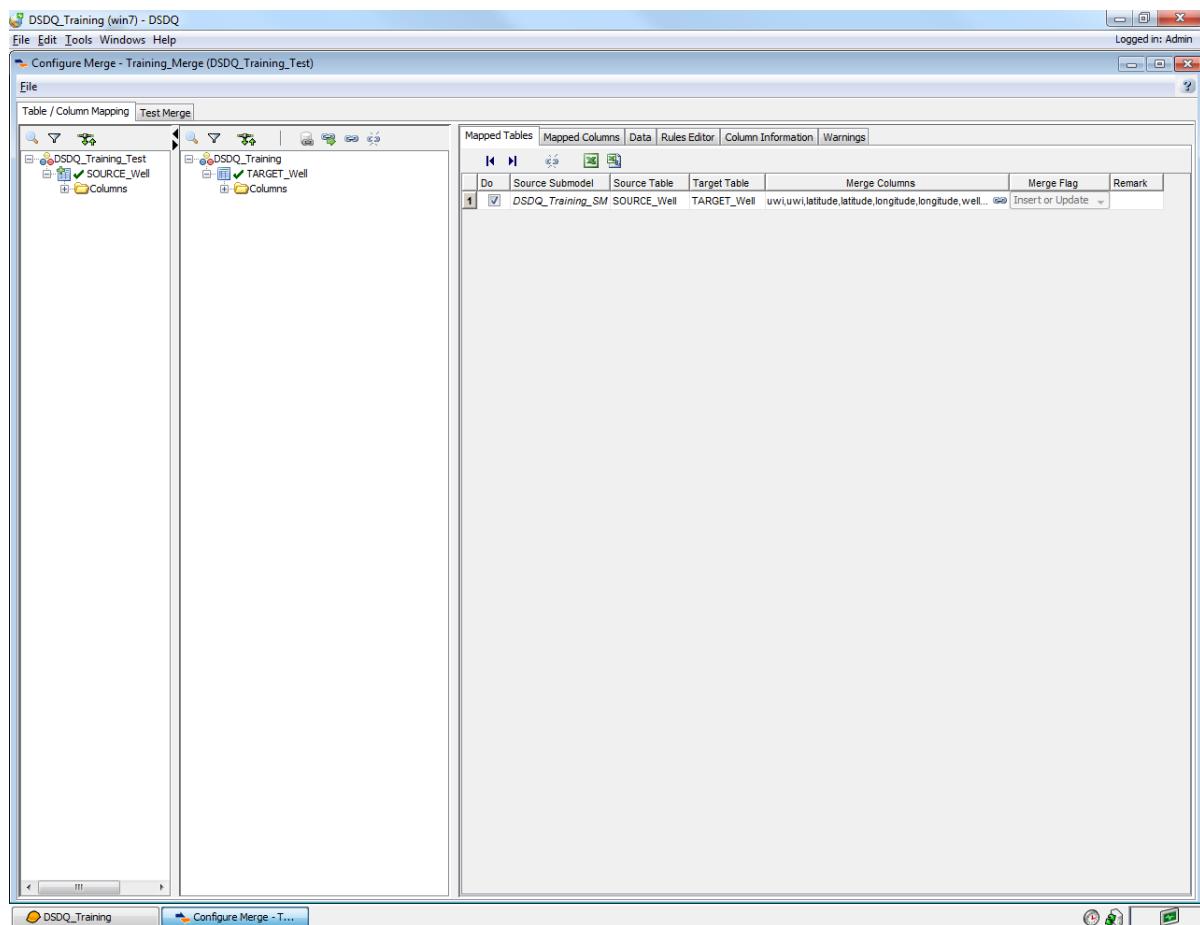
The **Configure Merge** window appears, displaying tables and columns for the Source Model Tree, Target Model Tree and the Detail pane. The Detail Panel has six tabs: **Mapped Tables**, **Mapped Columns**, **Data**, **Rules Editor**, **Column Information**, and **Warnings**:

Mapped Tables	This area displays information about the tables that have been mapped in the source and target trees. Clicking the Link  button under the "Merge Column" header causes the Record Association dialog to be launched and allows configuration of merge columns.
Mapped Columns	This area displays the mapping between the target column and all its mapped sources.
Data	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. 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.
Rules Editor	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.
Column Information	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.
Warnings	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.



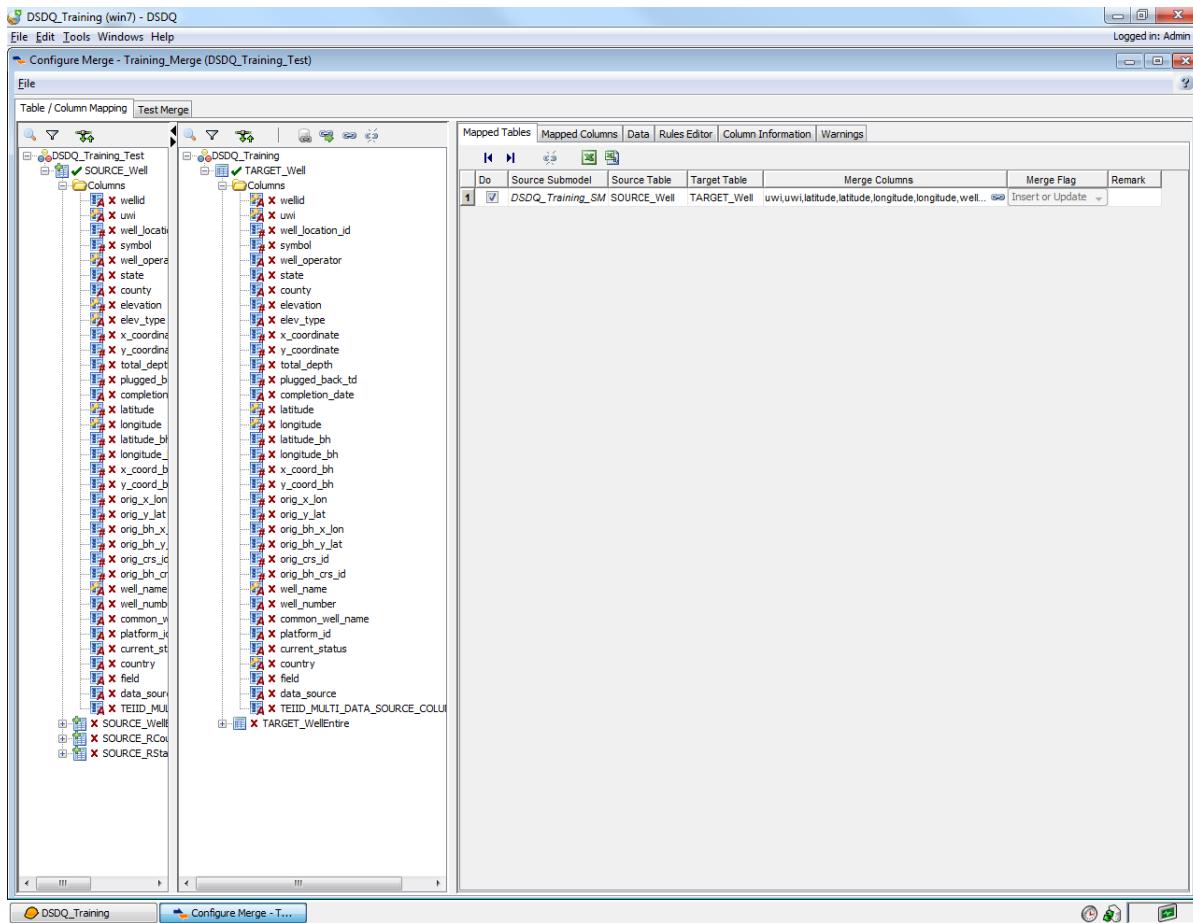
1. Select the **Source_Well** table from the Source Model Tree.
2. Select the **Target_Well** table from the Target Model Tree.

3. Click the **Auto Map Selected Tables/Columns**  button on the Target Model Tree toolbar
A green check mark appears adjacent to the selected tables.



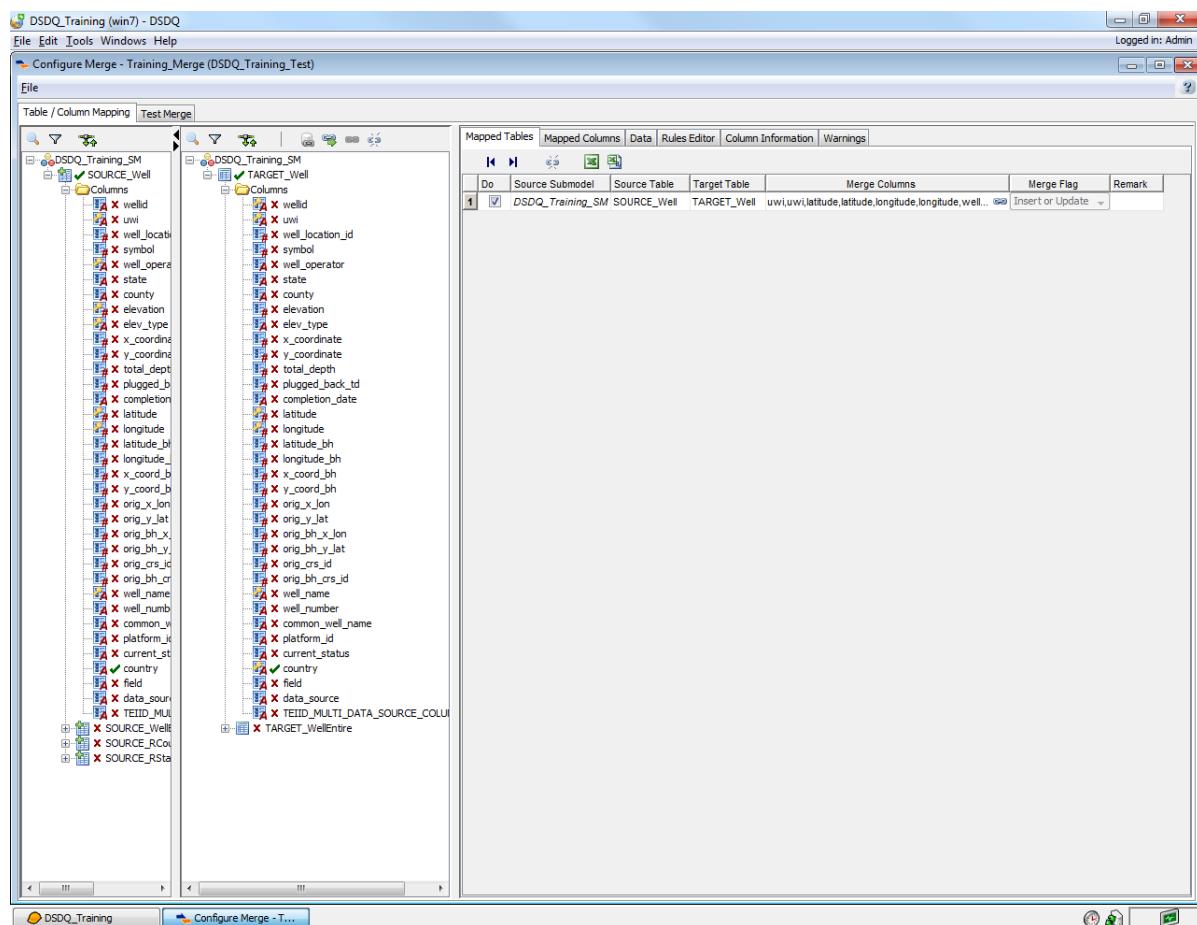
4. Click  to expand the columns on the **Source Model Tree**.

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



6. Select the **Country** element from the **Source_Well** and **Target_Well** table.
7. 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.

8. Optionally, repeat steps **6** and **7** 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

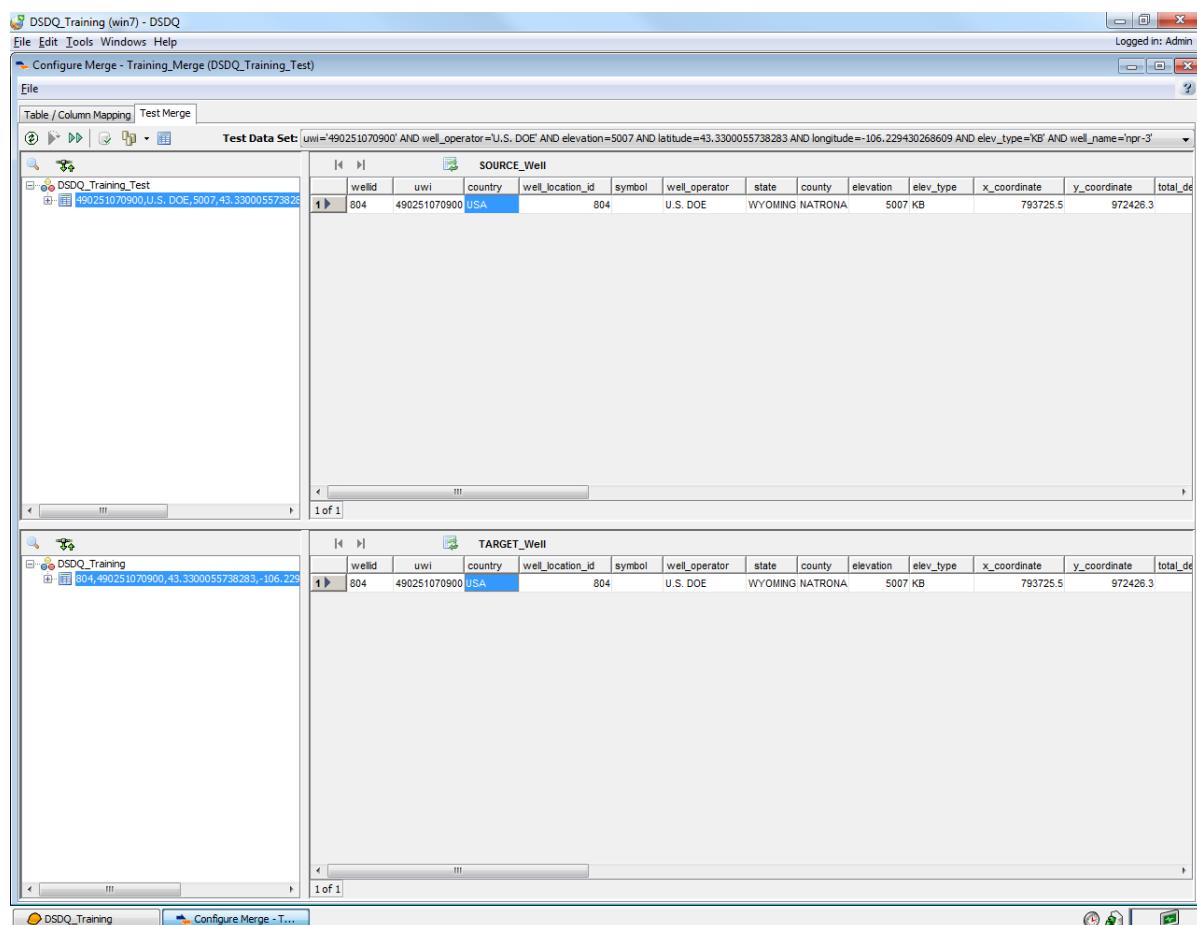
Merge flags allow configuring merge differently for every column. The **Update Always** merge flag is set as the default.

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.

9. 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.



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

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

12. 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.

13. To view just the affected mapped columns in the **Data** table view, click on the **View Affected Columns**  button in the **Test Merge** toolbar.
14. To **Test Merge** against database constraints, click the **Test Against Database**  button in the **Test Merge** toolbar.
15. 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.
16. Select **File > Exit** from the menu bar on the **Configure Merge** window.

Using Merge

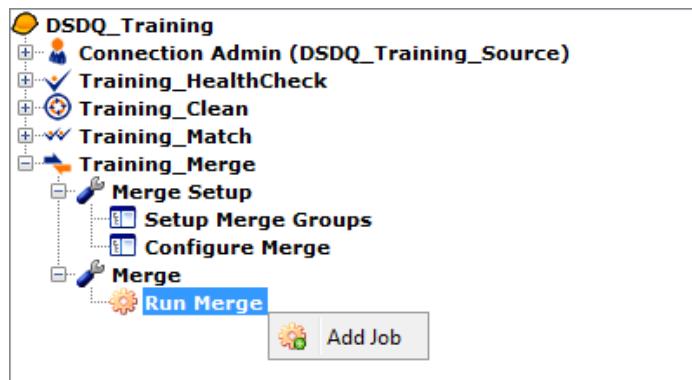
The Merge Activity is used after merge configurations have been tested on your data subset. This activity moves the data to the target database and generates a merge activity report.

Exercise: Running the Merge Tool

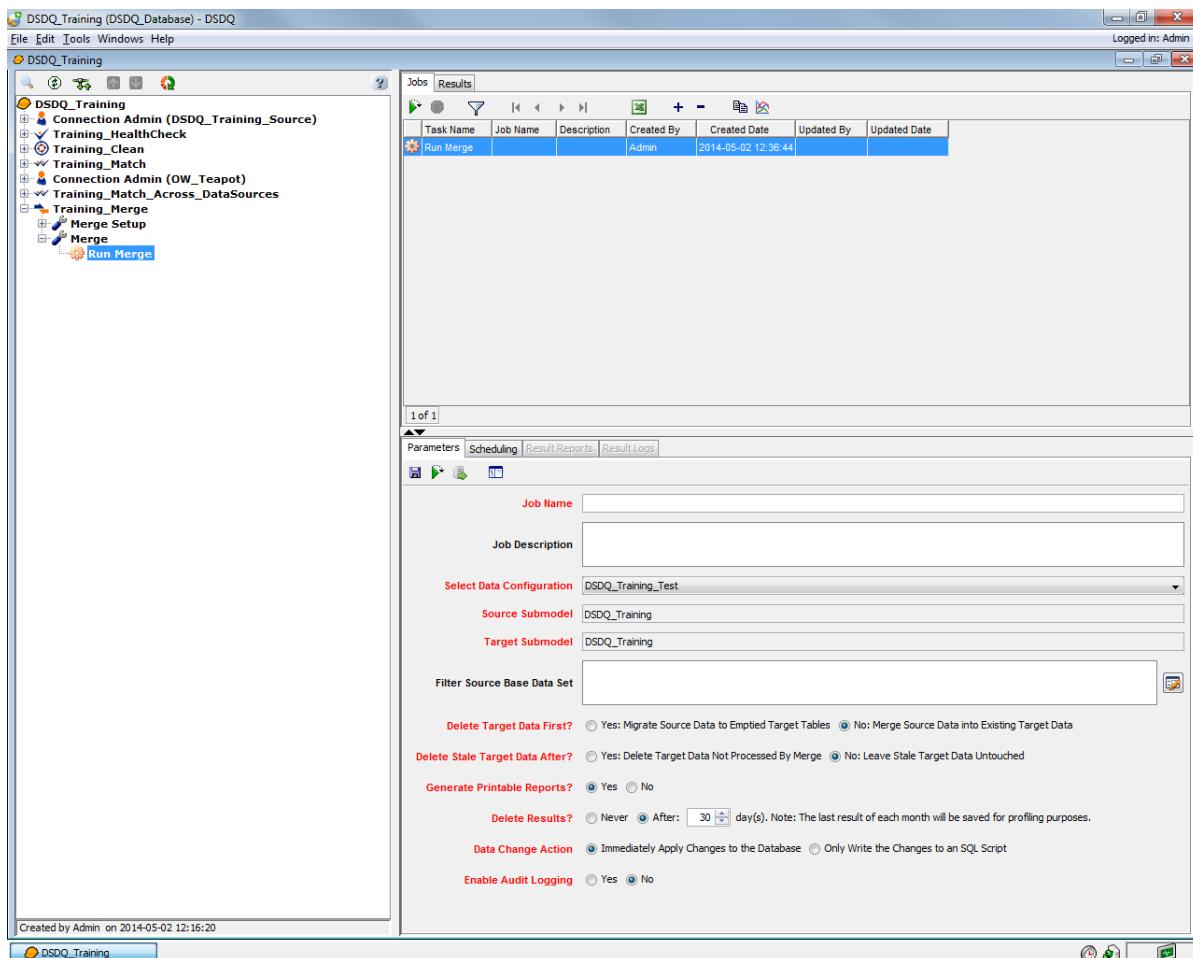
The **Run Merge** Tool is used to move the merge data to the target database. The parameters you would enter for the **Merge** Tool include: Job Name and Merge Group. After the Merge operation is complete, you can check the results on the **Result Reports** tab of the **Jobs and Results Information Pane**.

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

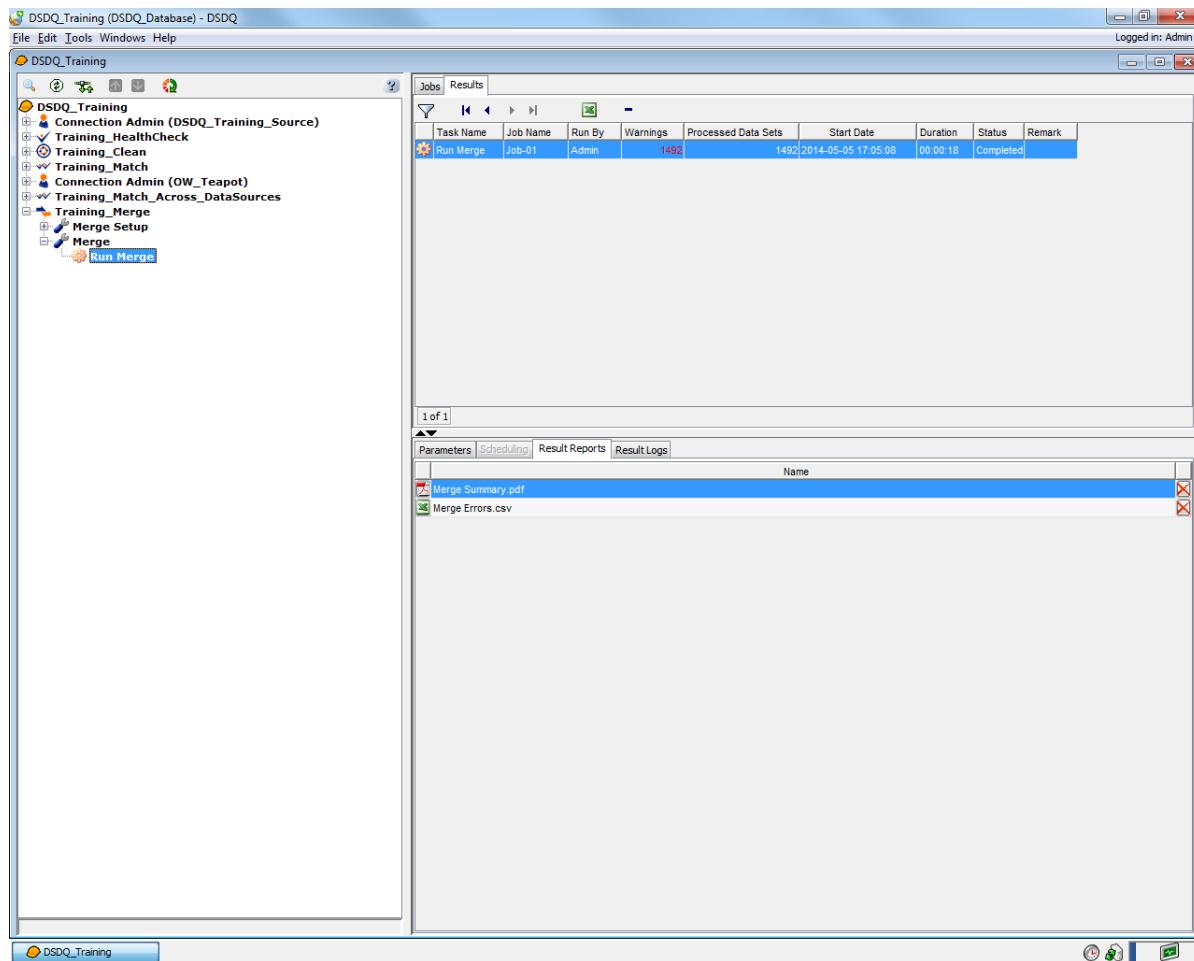
Selecting **Immediately Apply Changes to the Database** will apply the job changes directly to the configured data source. Or if you want to write a Sql script of the changes, select **Only Write the Changes to an SQL Script**. This creates a “Database Changes.sql” file to be produced and can be viewed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



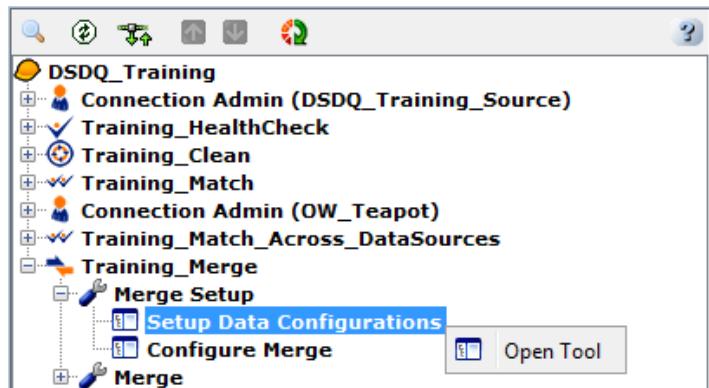
Merge Summary					
Source Table	Target Table	Total Rows	Rows Processed	Rows Succeeded	Rows Failed
Well	Well	1388	1388	1388	0

Target Table Remarks

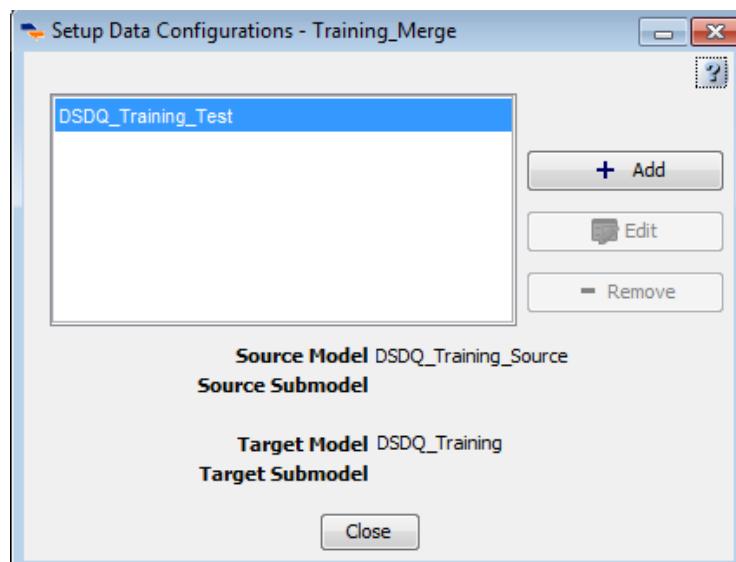
Exercise: Modifying Data while Merging

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.

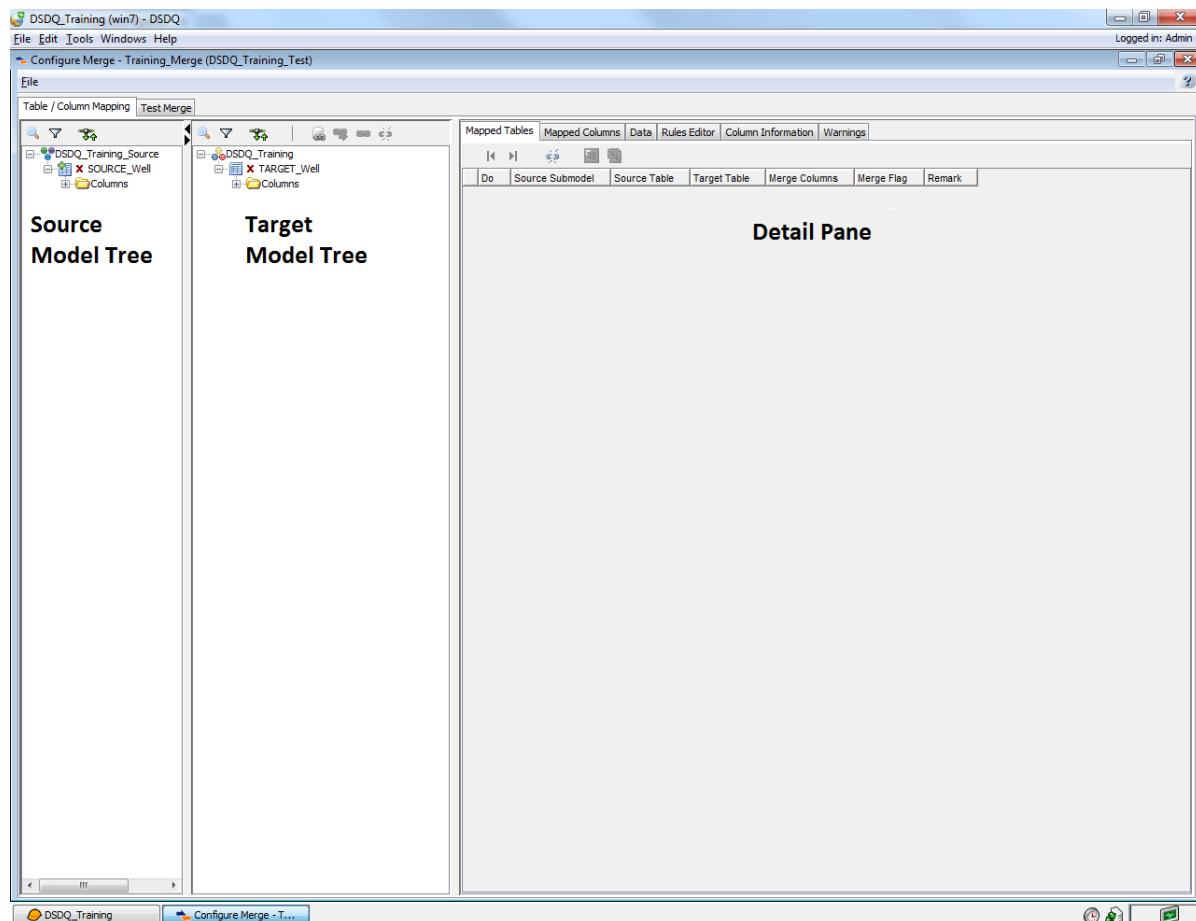


2. Select the **DSDQ_Training_Test** option.

3. Click **OK**.

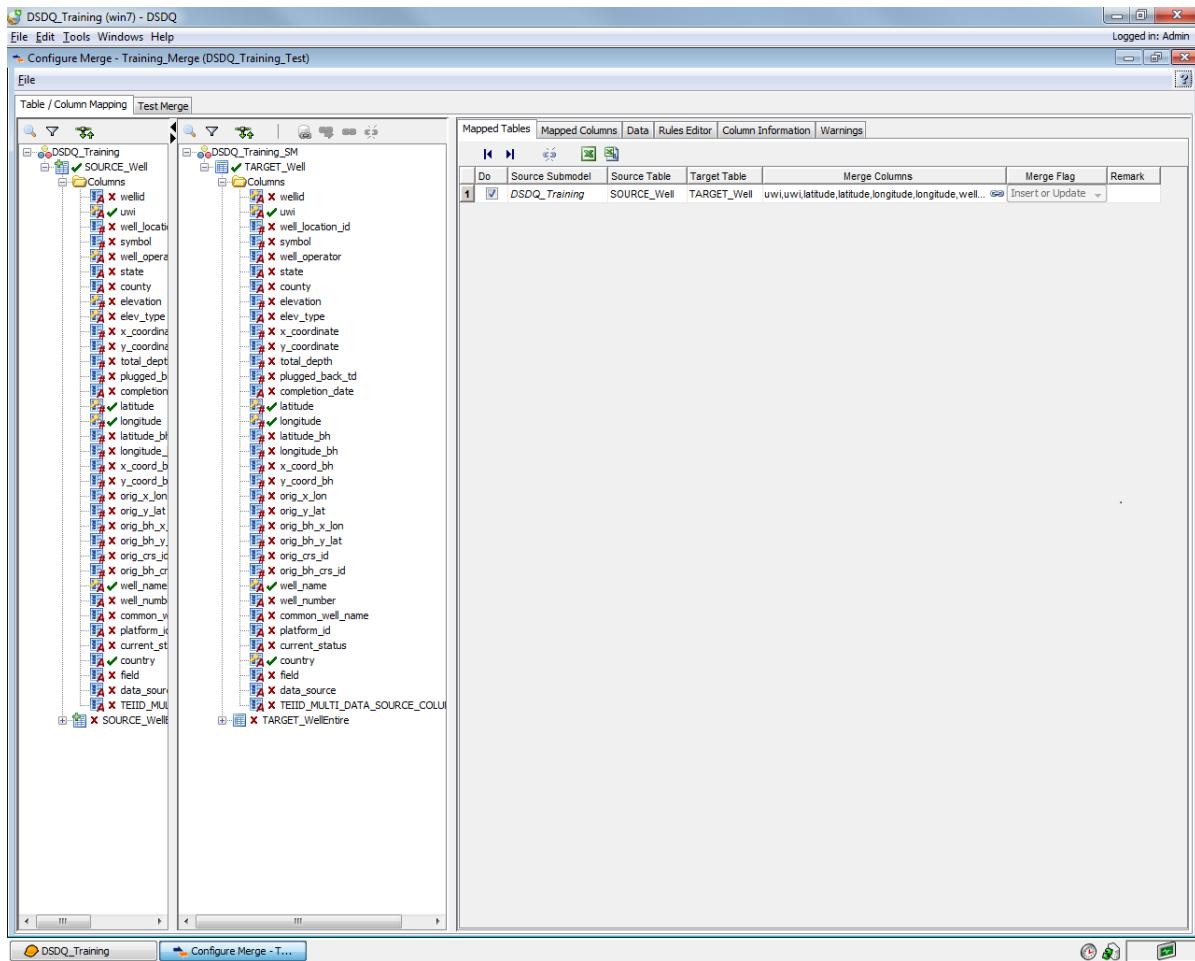
The **Configure Merge** window appears, displaying tables and

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**.

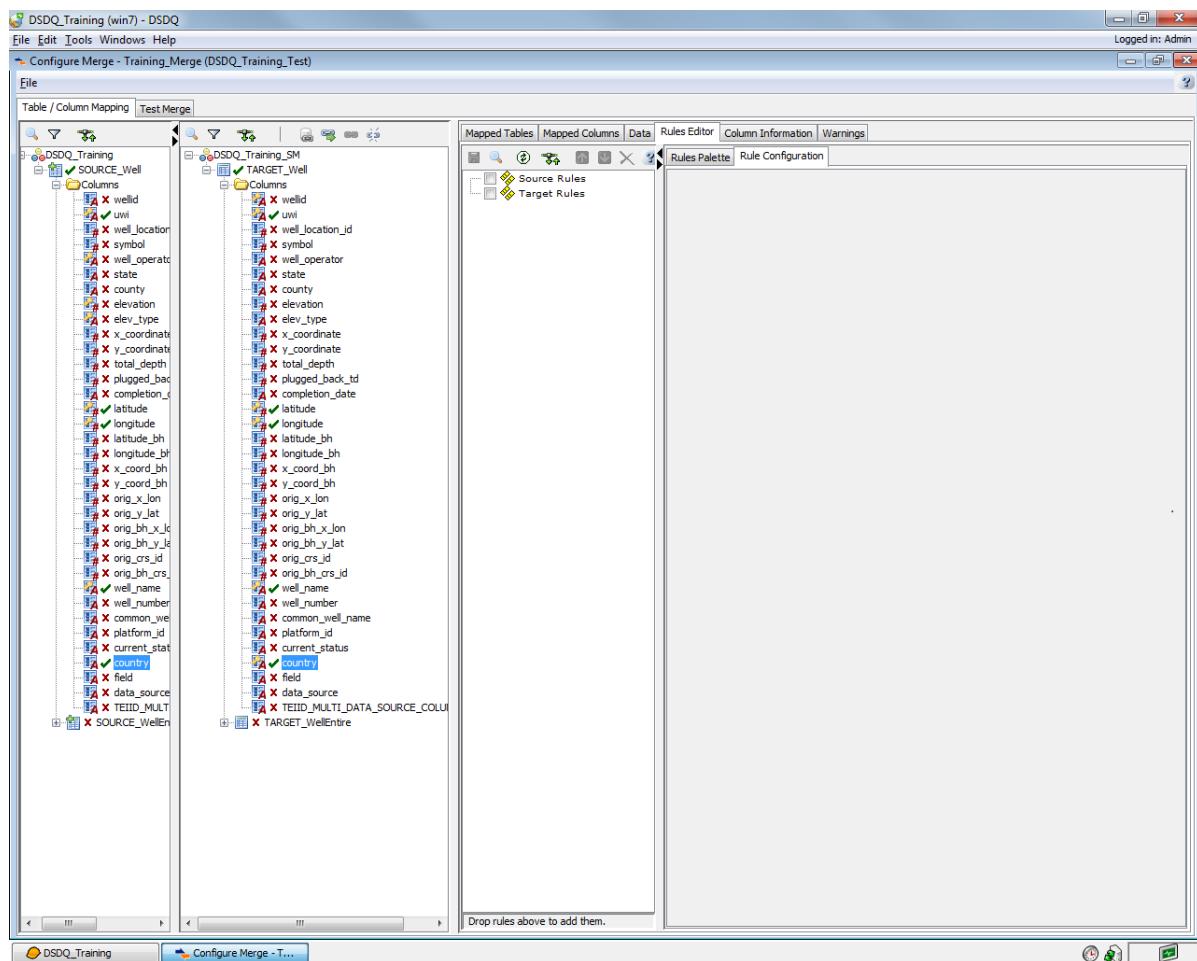
- Click **+** to expand the **Columns** folder on the **Target Model Tree**.



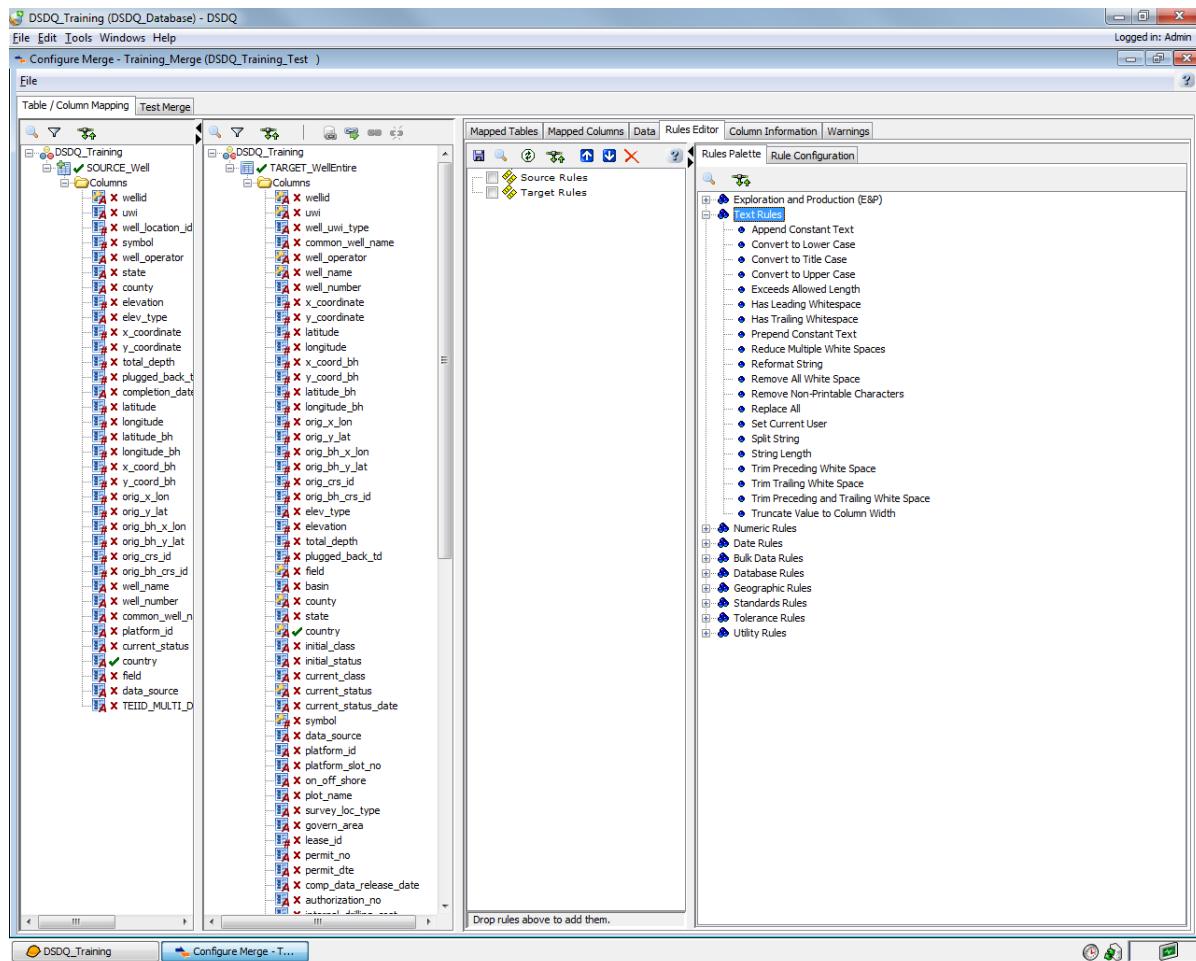
- Select the **Country** element from the Source_Well and Target_Well tables.

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

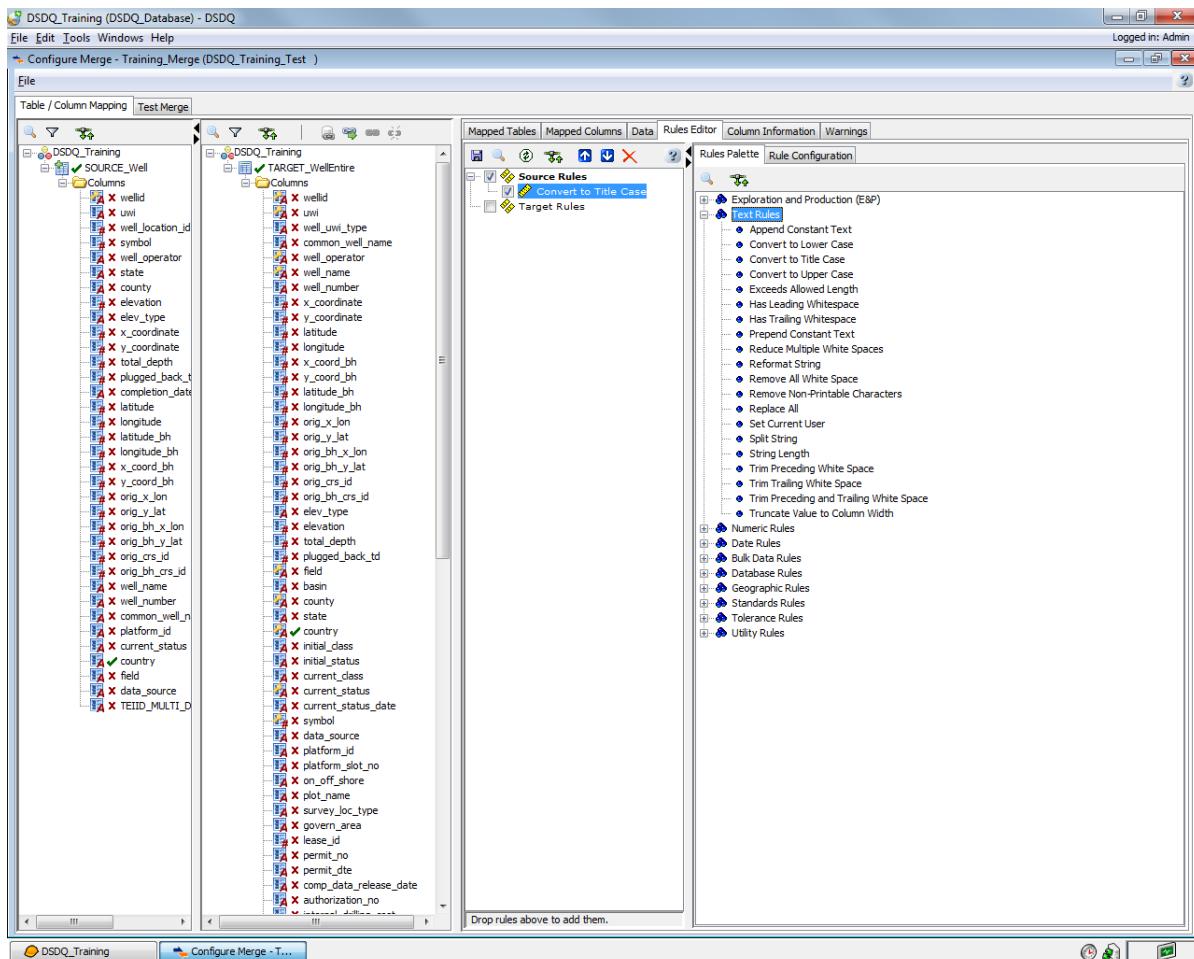
The **Rule Configuration** tab displays in the **Detail Pane** of the **Configure Merge** window.



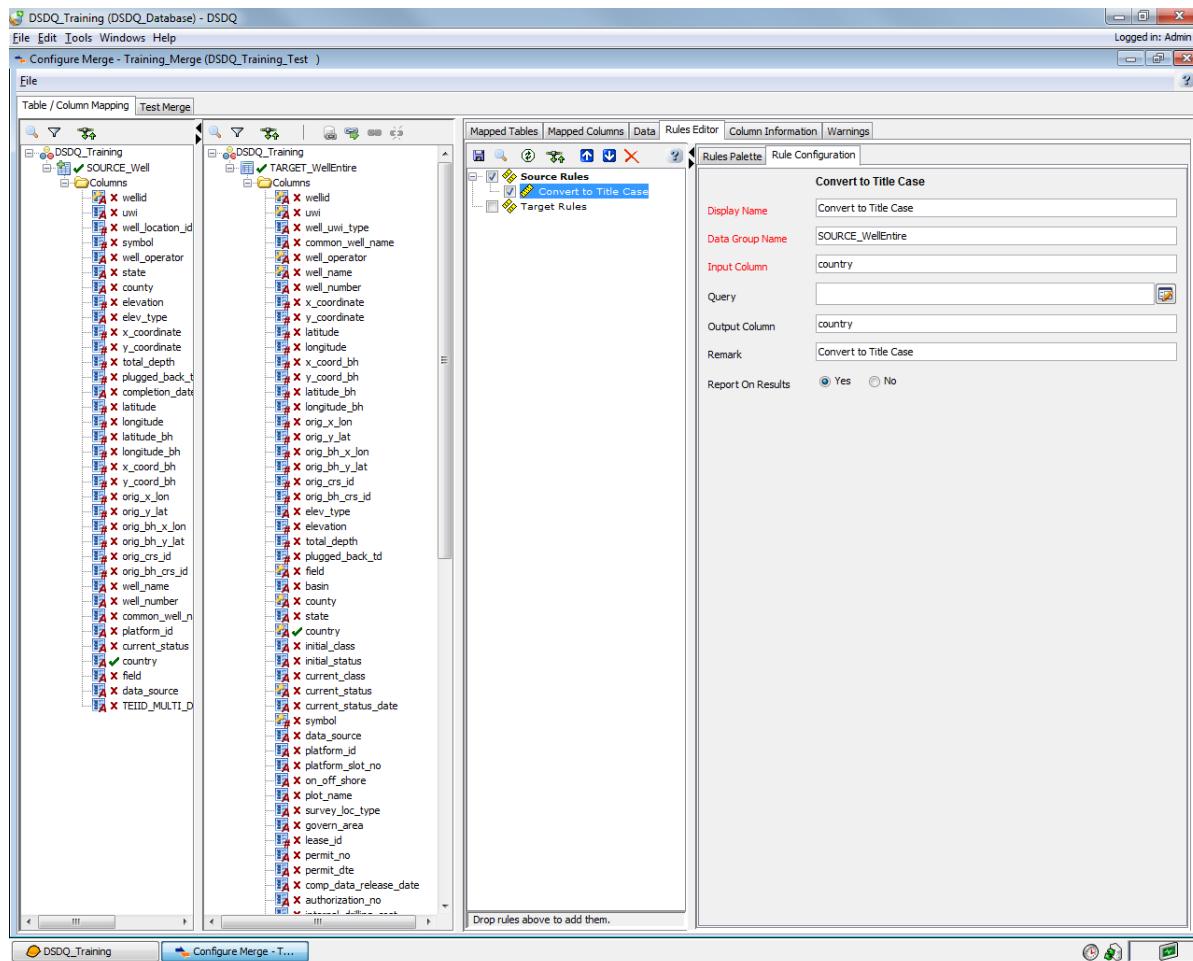
8. Select the **Rules Palette** tab and click to expand the **Text Rules**.



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



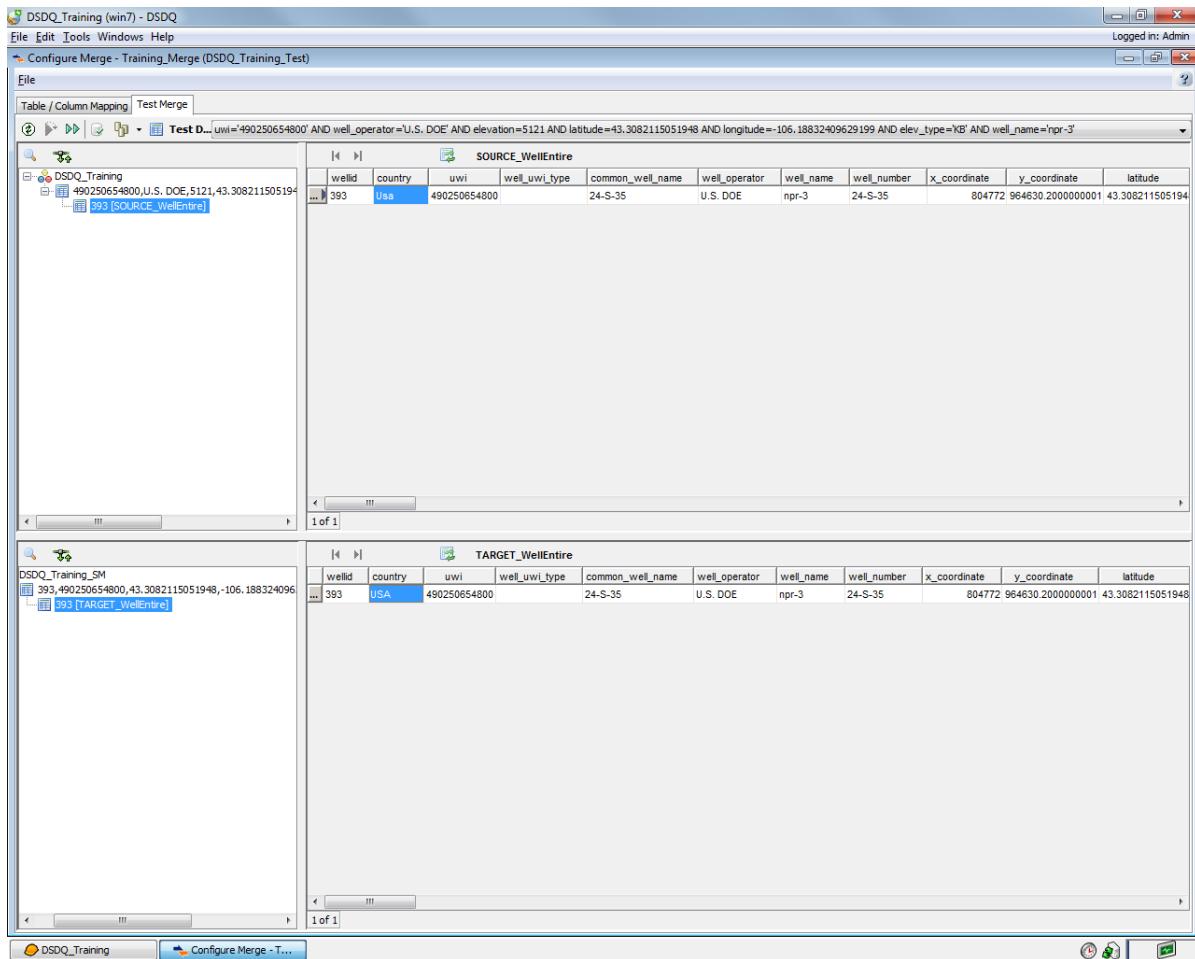
The **Rule Configuration** tab displays by default.



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

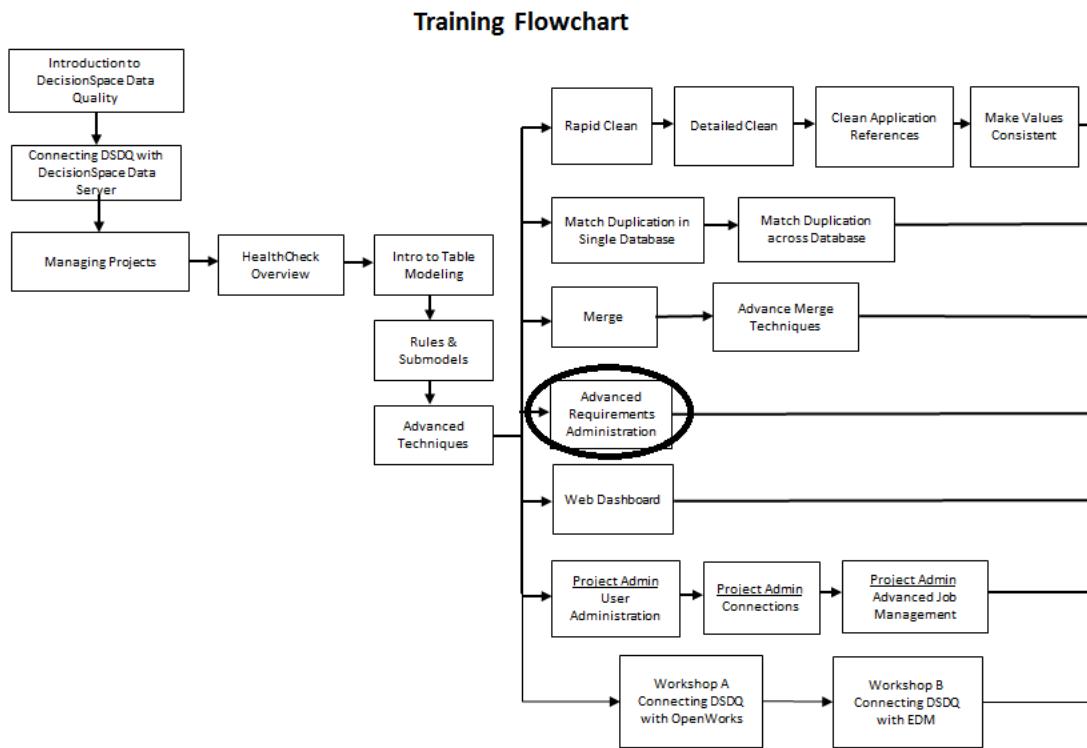
The Requirements Administrator tool manages the Data Quality Rules Repository. The tool is divided into two sections with respective folder trees: Repository and Service Level. The Repository Tree contains the original values in the repository and should only be updated if a new global requirement needs 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:



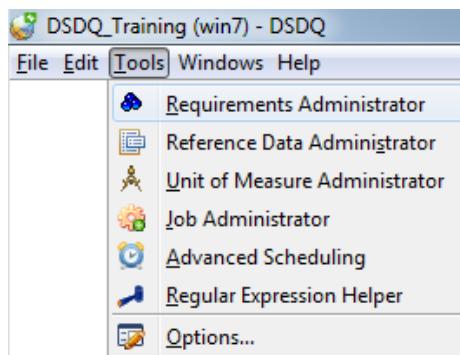
Repository

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.

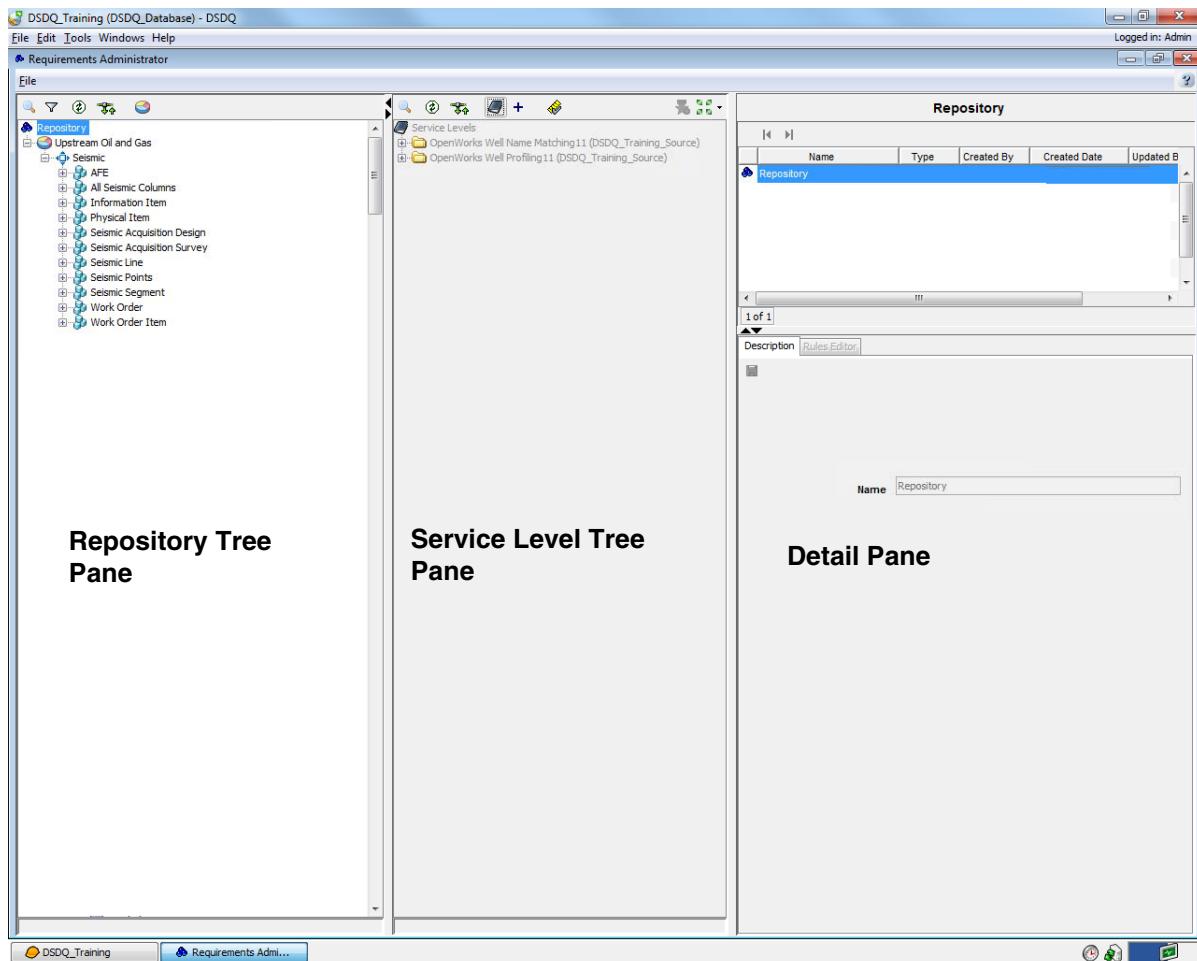
Exercise: Populating the Repository

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.

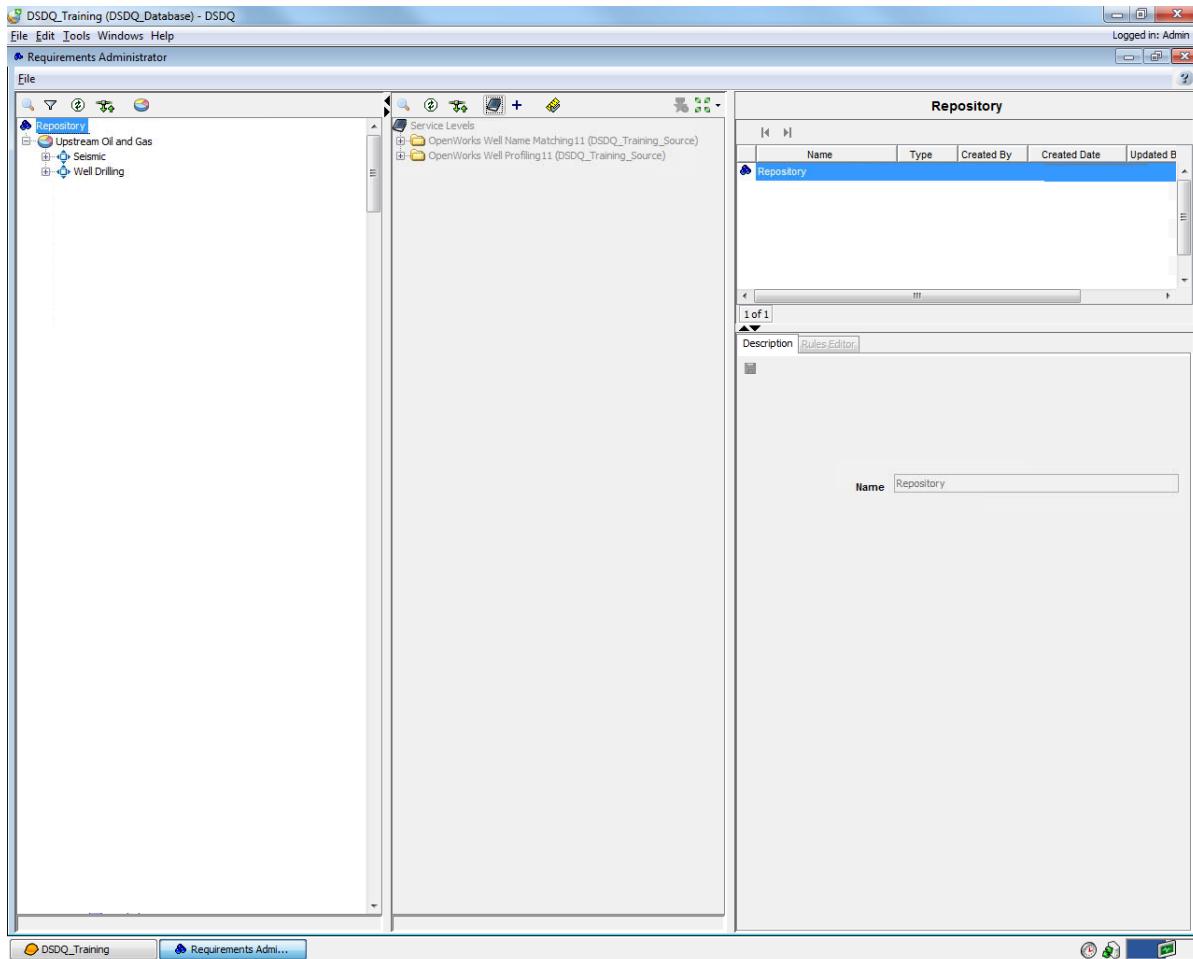


3. Select the **Well Drilling** area from the **Upstream Oil and Gas** sector.



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**.

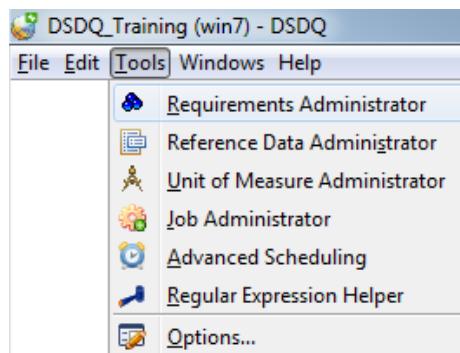
Service Levels

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.

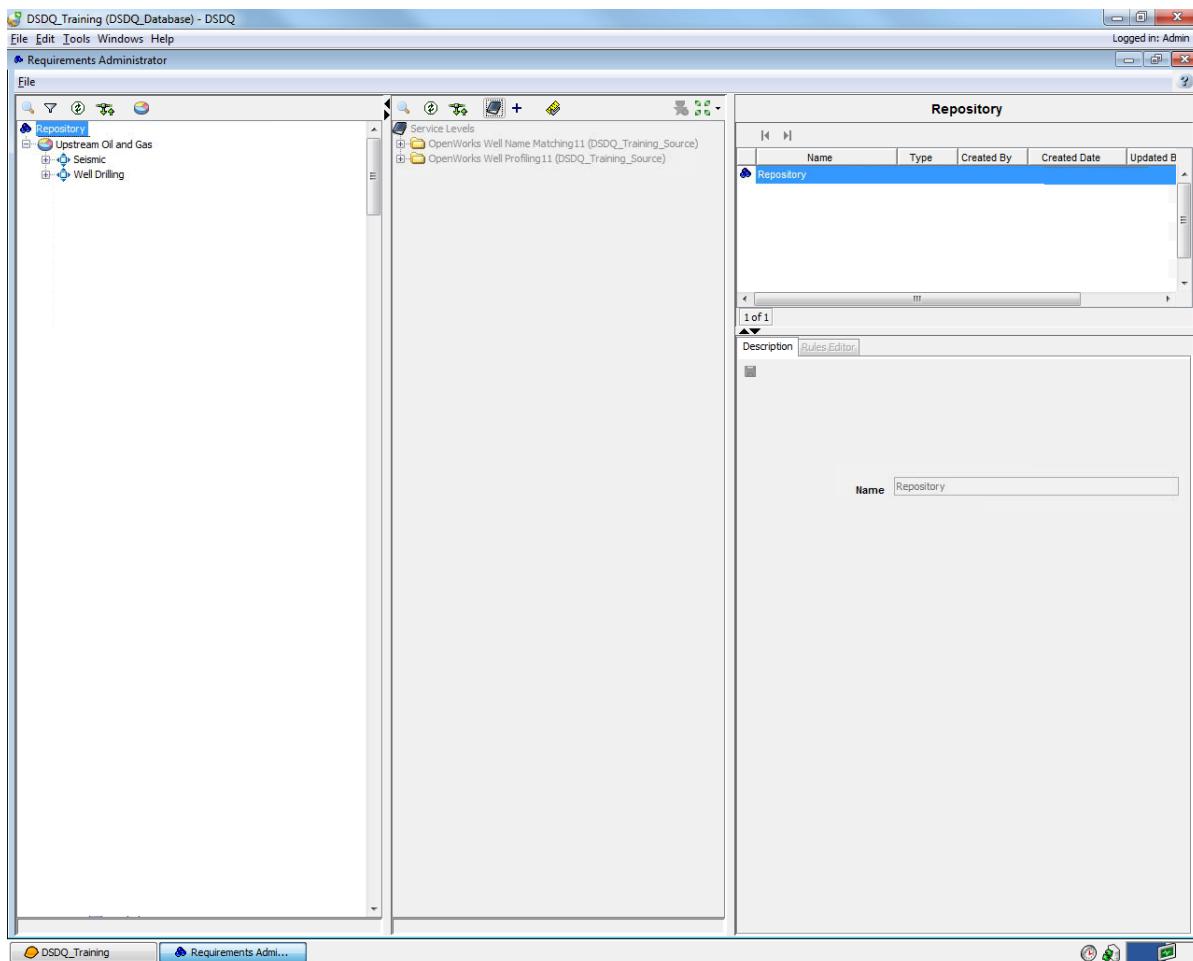
Exercise: Creating a Service Level

To create a service level:

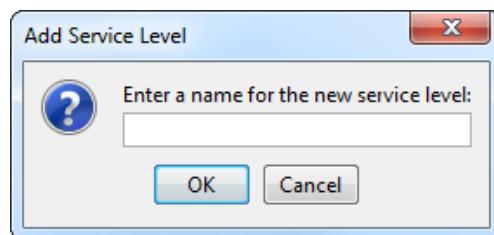
1. Select **Tools > Requirements Administrator** from the menu bar on the **DSDQ Project** window.



The Requirements Administrator window appears.



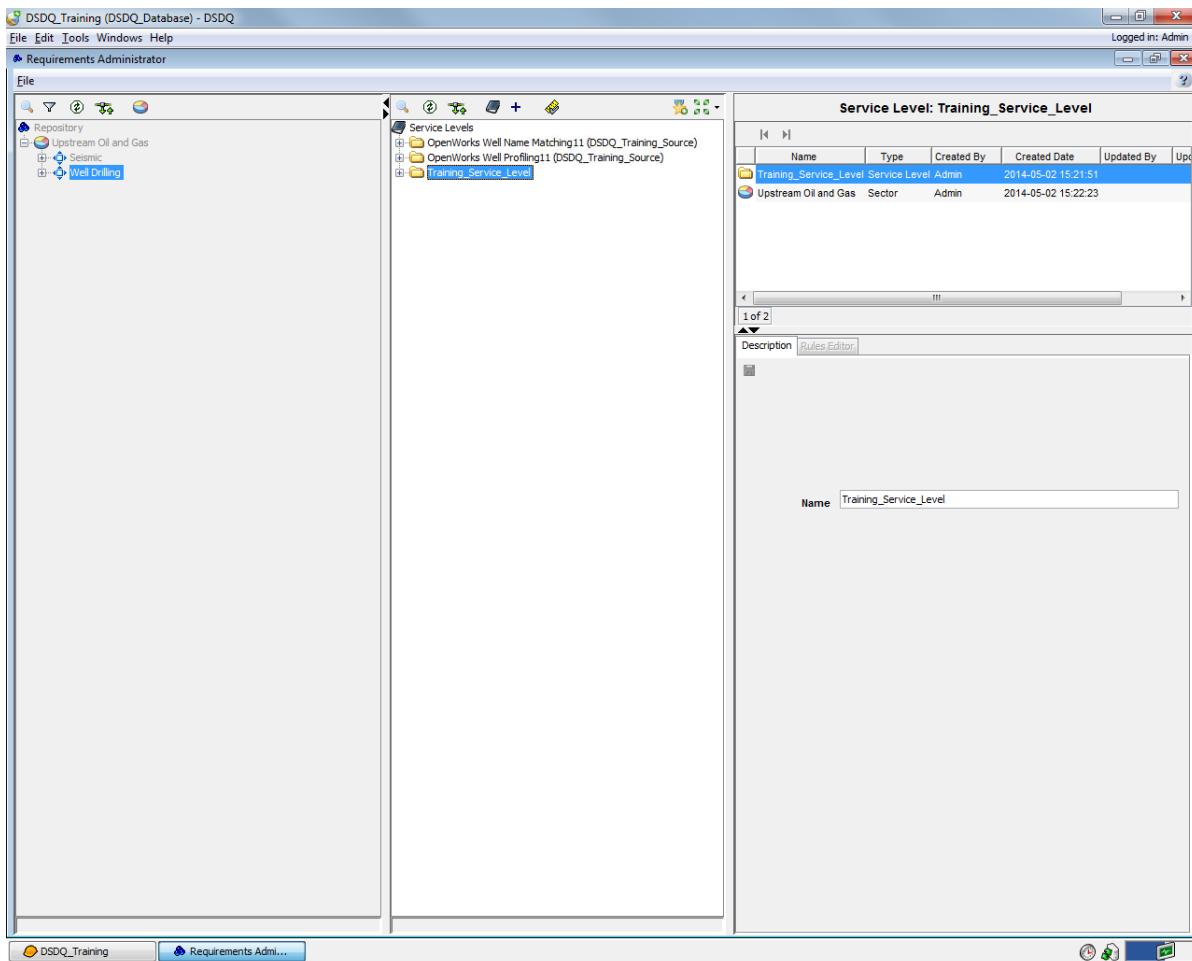
2. Click on the Service Level Tree Pane toolbar.
The Add Service Level dialog box appears.



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.



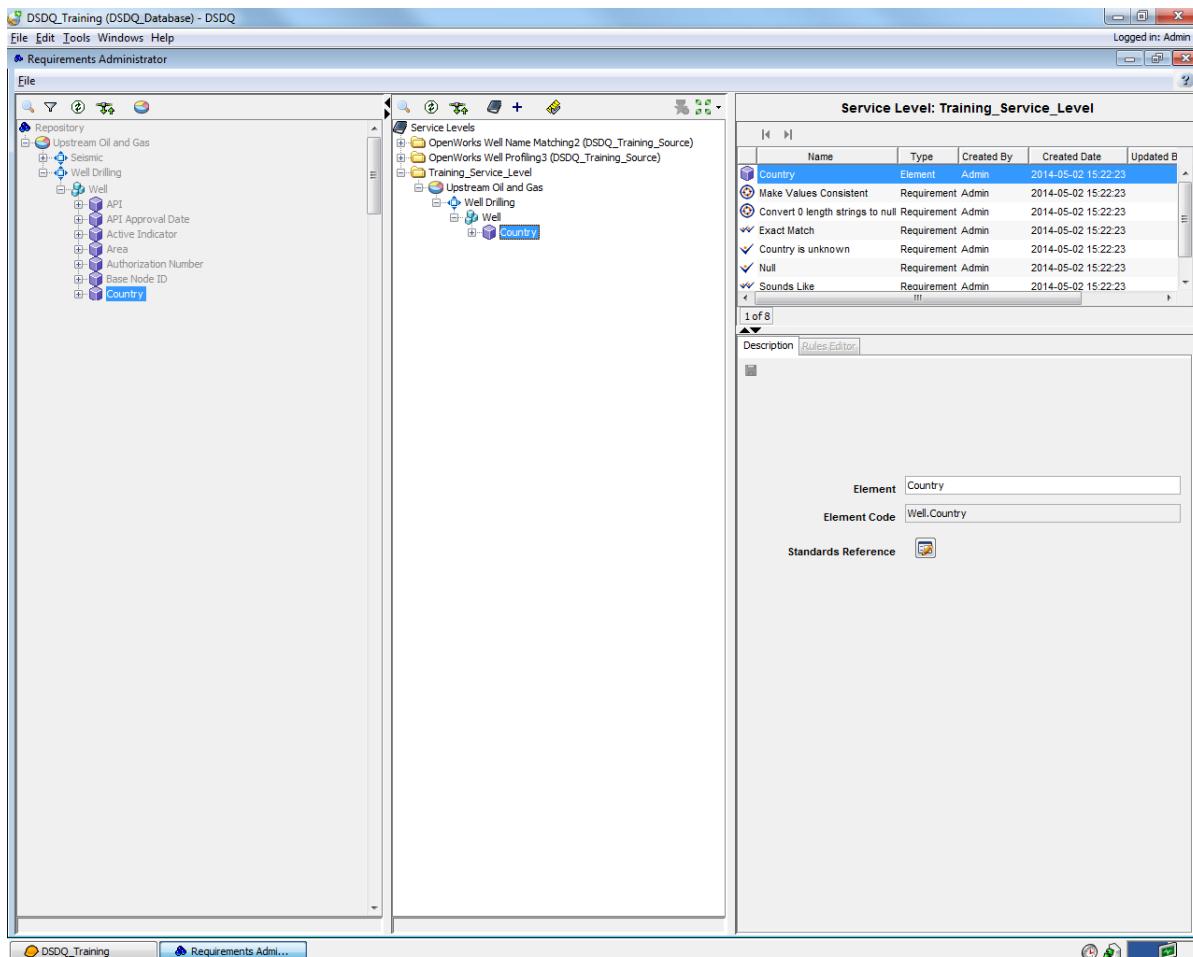
5. Optionally, add more service levels by repeating steps **2** and **3**.

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

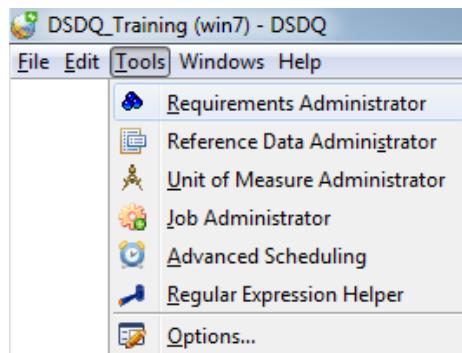
You can publish a service level to the repository. Select the desired Service Level from the Service Level Tree and click the **Publish Service Level to Repository** button from the Service Level Tree Pane toolbar.

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

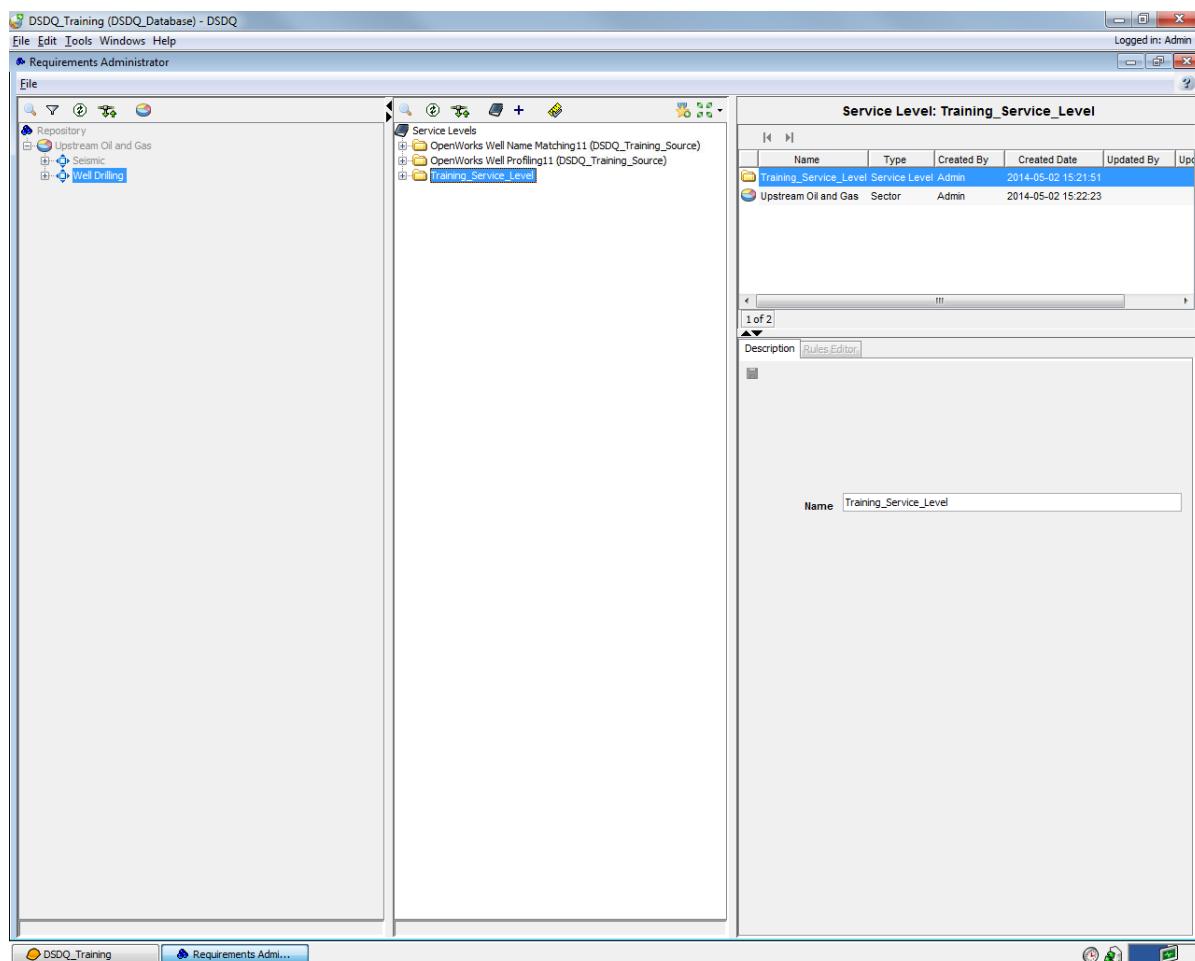
Exercise: Adding a Service Level

To add a service level:

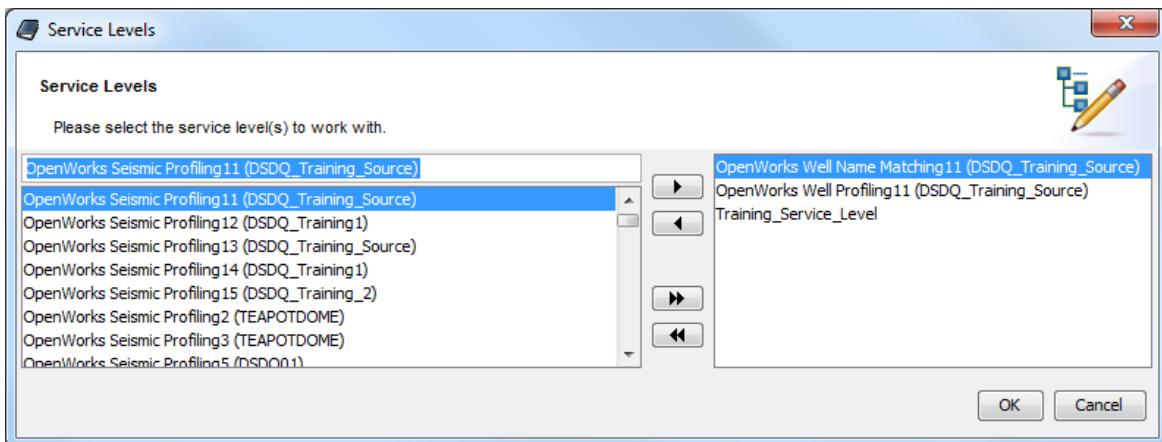
1. Select **Tools > Requirements Administrator** from the menu bar on the **DSDQ Project** window.



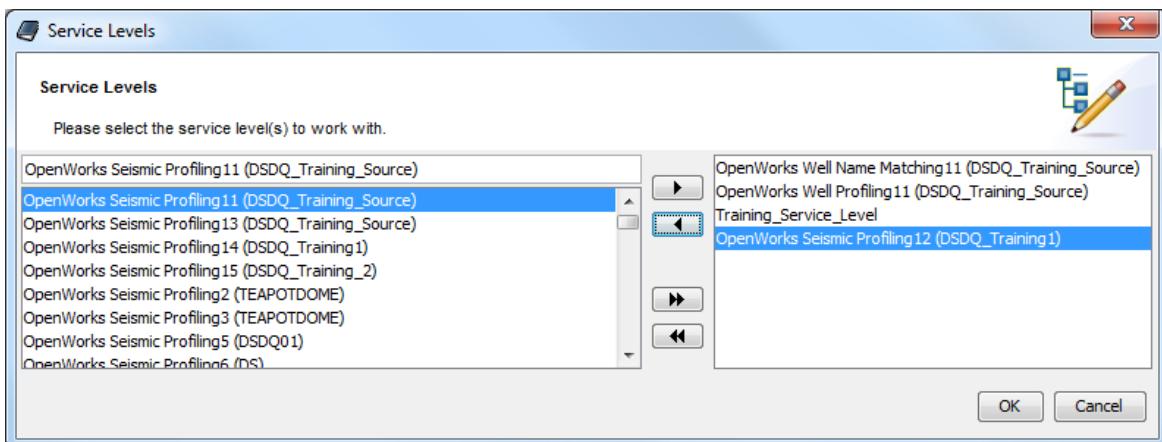
The **Requirements Administrator** window appears.



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

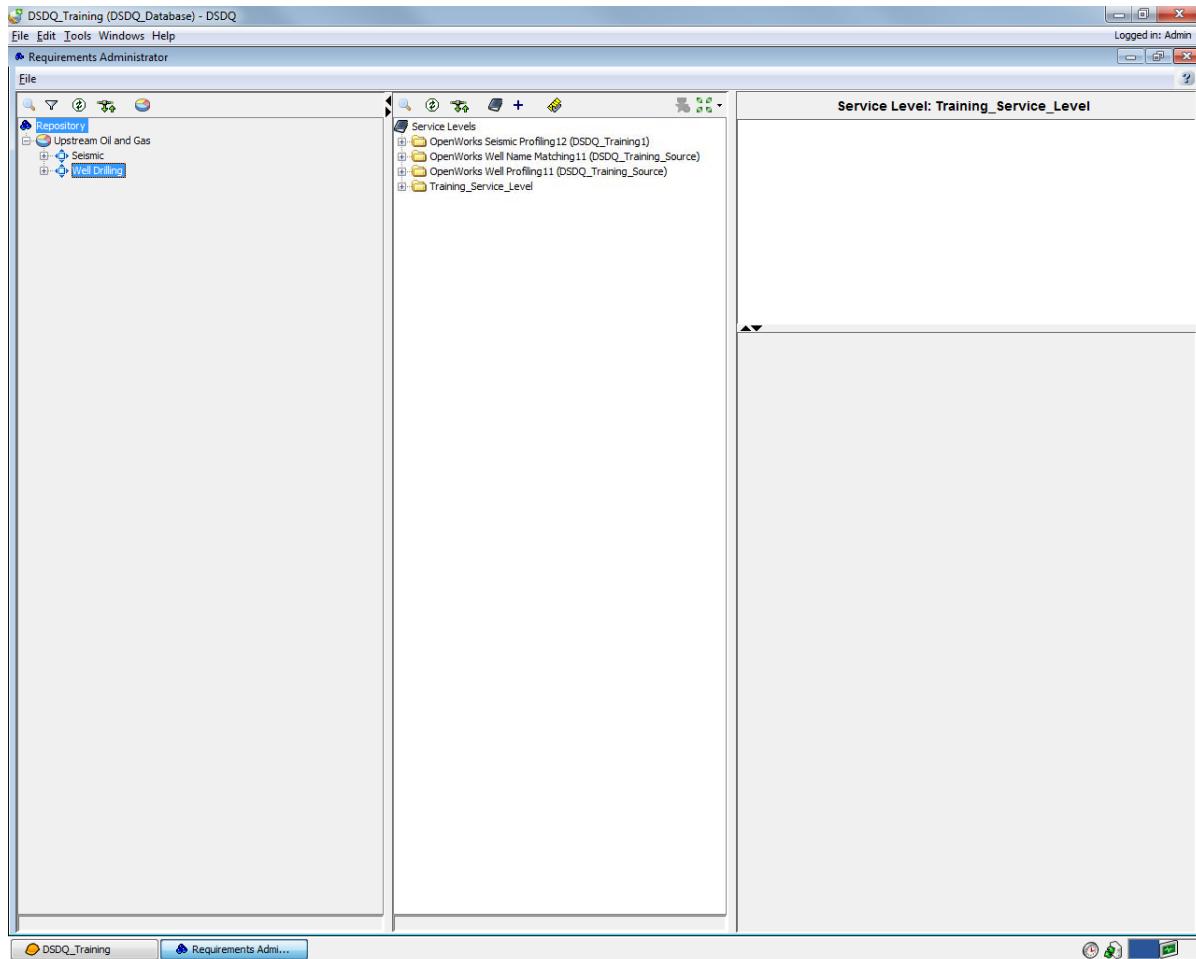


3. Select **OpenWorks Seismic Profiling 12(DSDQ_Training1)** from the left pane and click .
- The selected service level is moved to the right pane of the **Service Levels** dialog box.



4. Click **OK**.

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

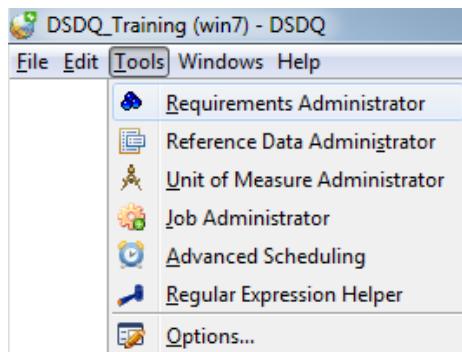


5. Optionally, you can add more than one service level by repeating steps **2** to **4**.
6. Select **File > Exit** from the menu bar on the **Requirements Administrator** window.

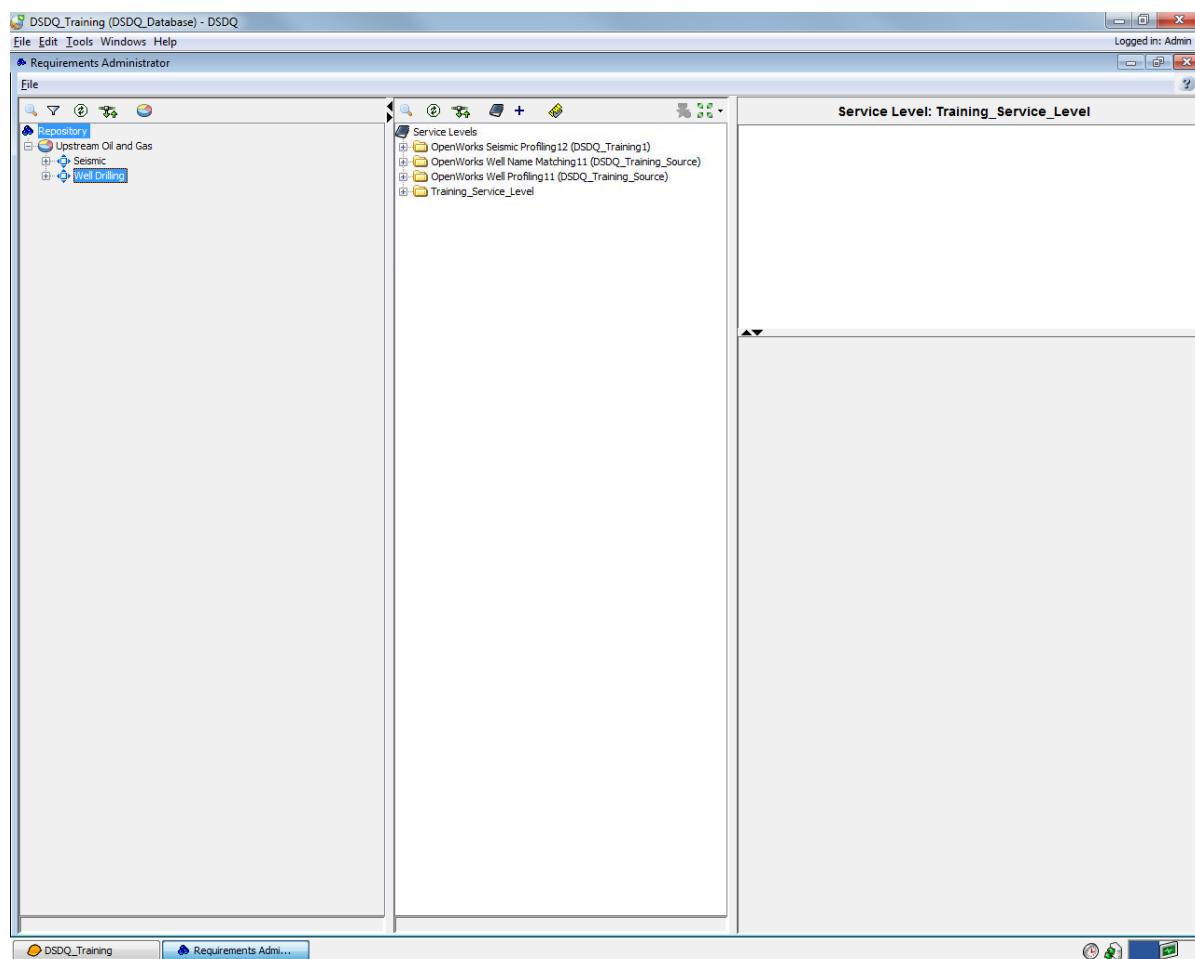
Exercise: Editing a Service Level

To edit a service level:

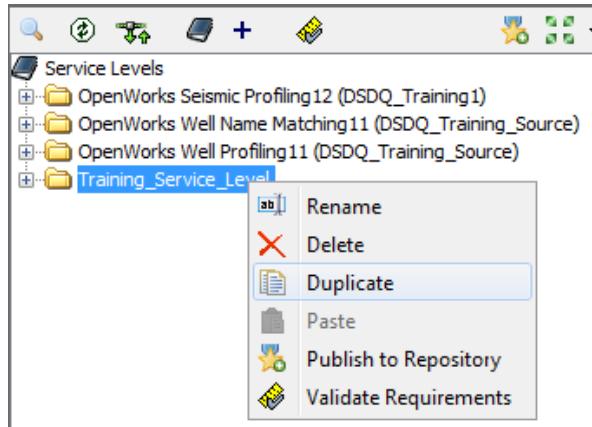
1. Select **Tools > Requirements Administrator** from the menu bar on the **DSDQ Project** window.



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.

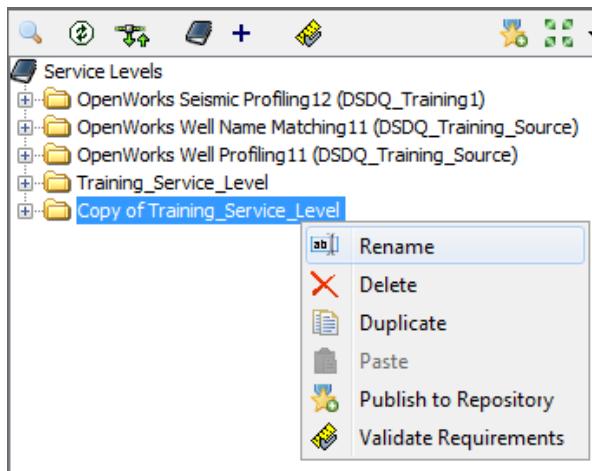


A duplicate service level is added in the **Service Level Tree**.

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.

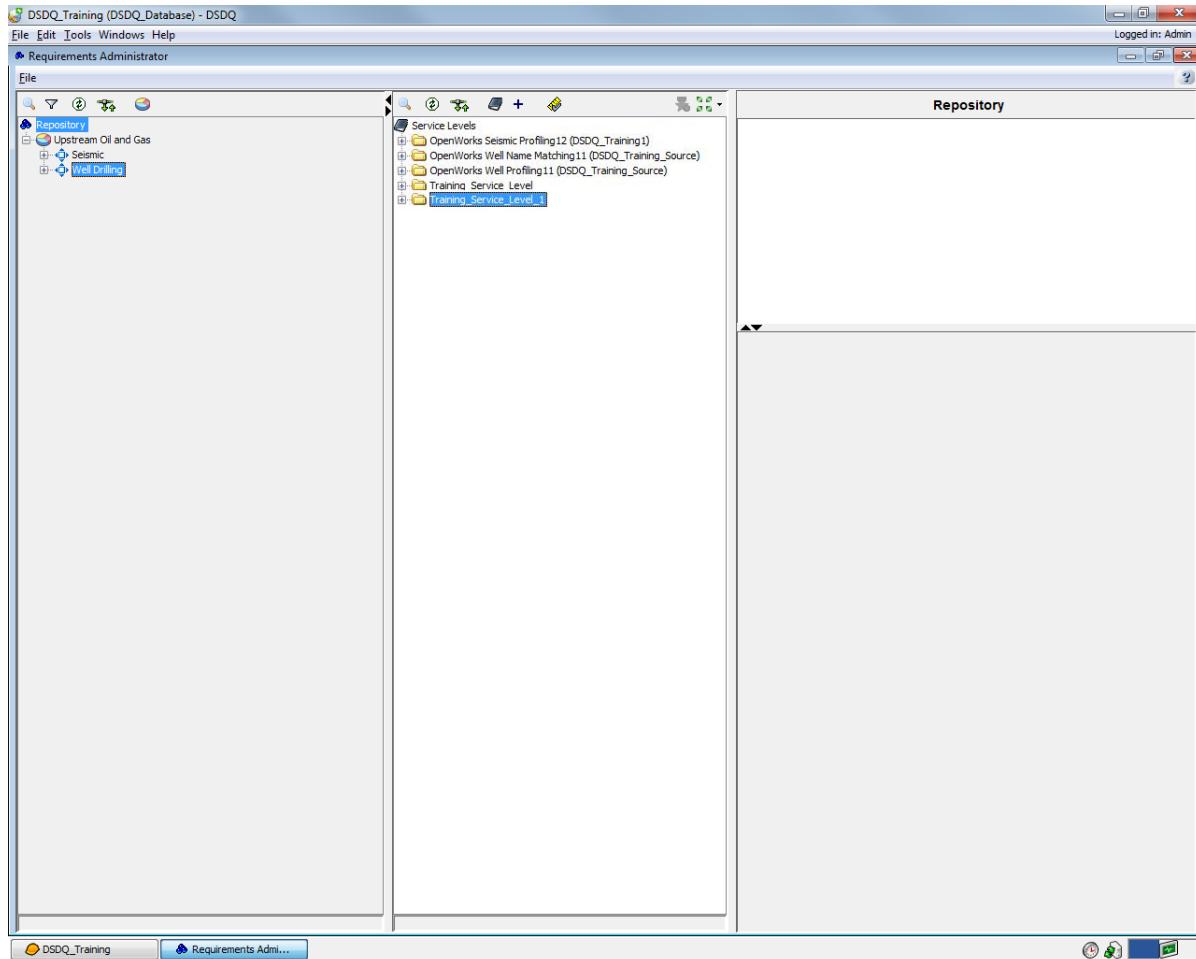


The **Rename Service Level Item** dialog box appears.



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**.

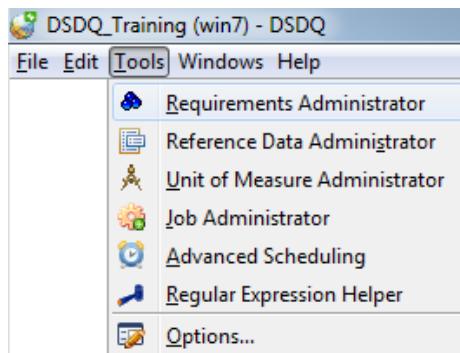


6. Select **File > Exit** from the menu bar on the **Requirements Administrator** window.

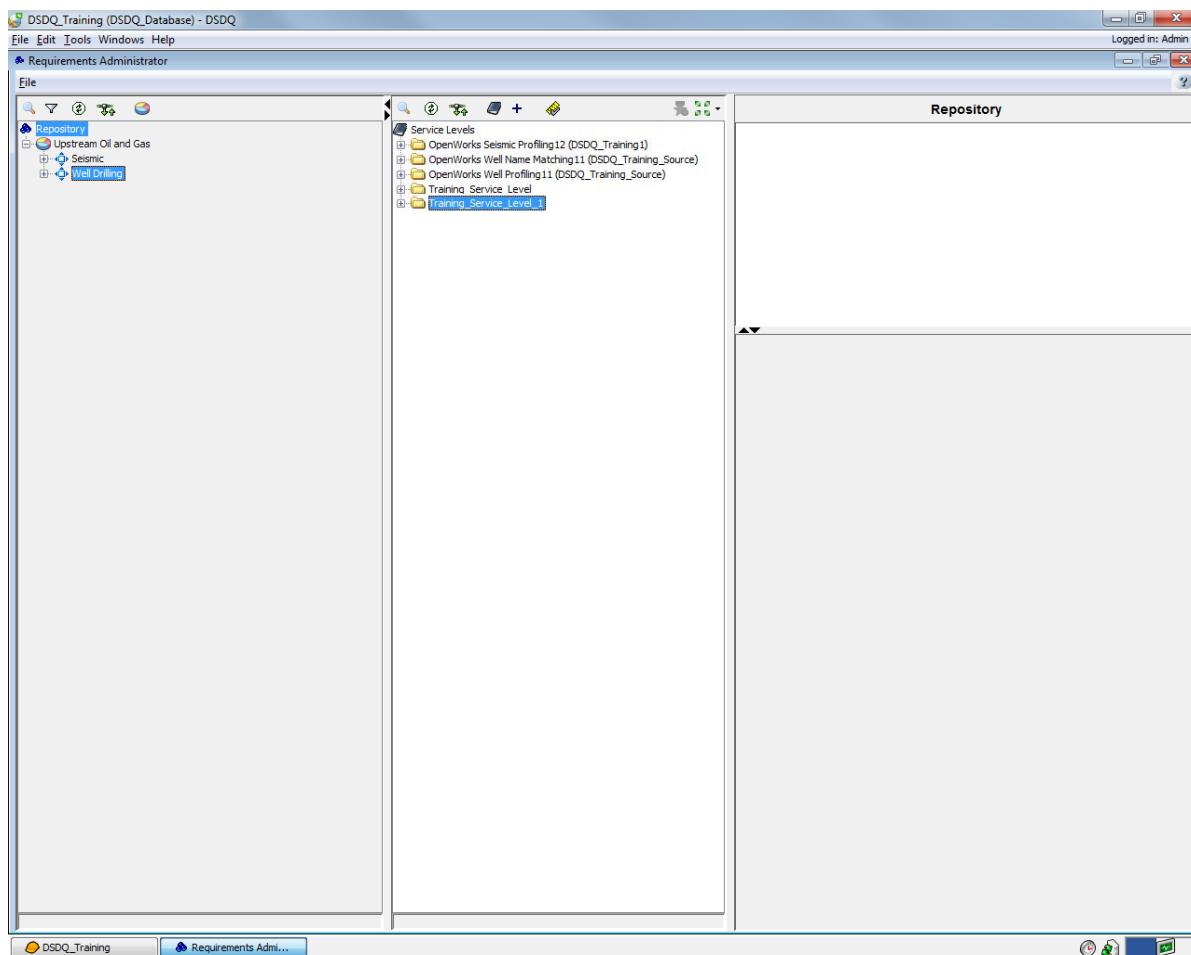
Exercise: Deleting a Service Level

To delete a service level:

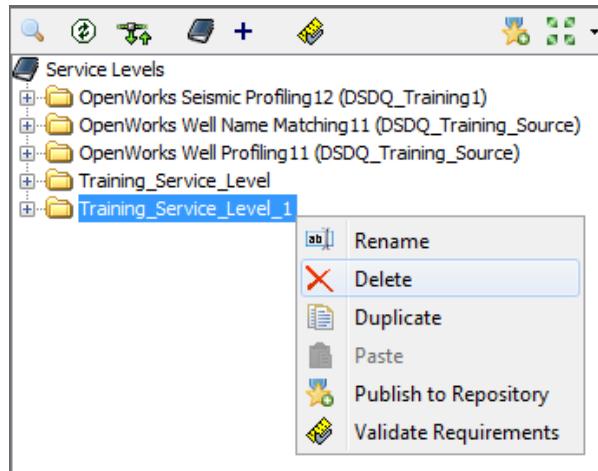
1. Select **Tools > Requirements Administrator** from the menu bar on the DSDQ Project window.



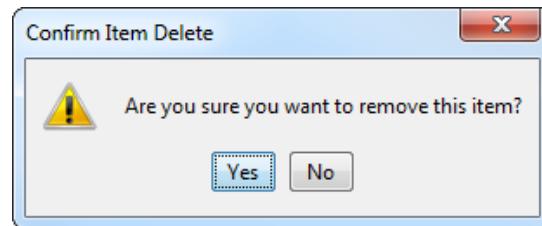
The Requirements Administrator window appears.



2. Right-click **Training_Service_Level_1** and select **Delete** from the pop-up menu.

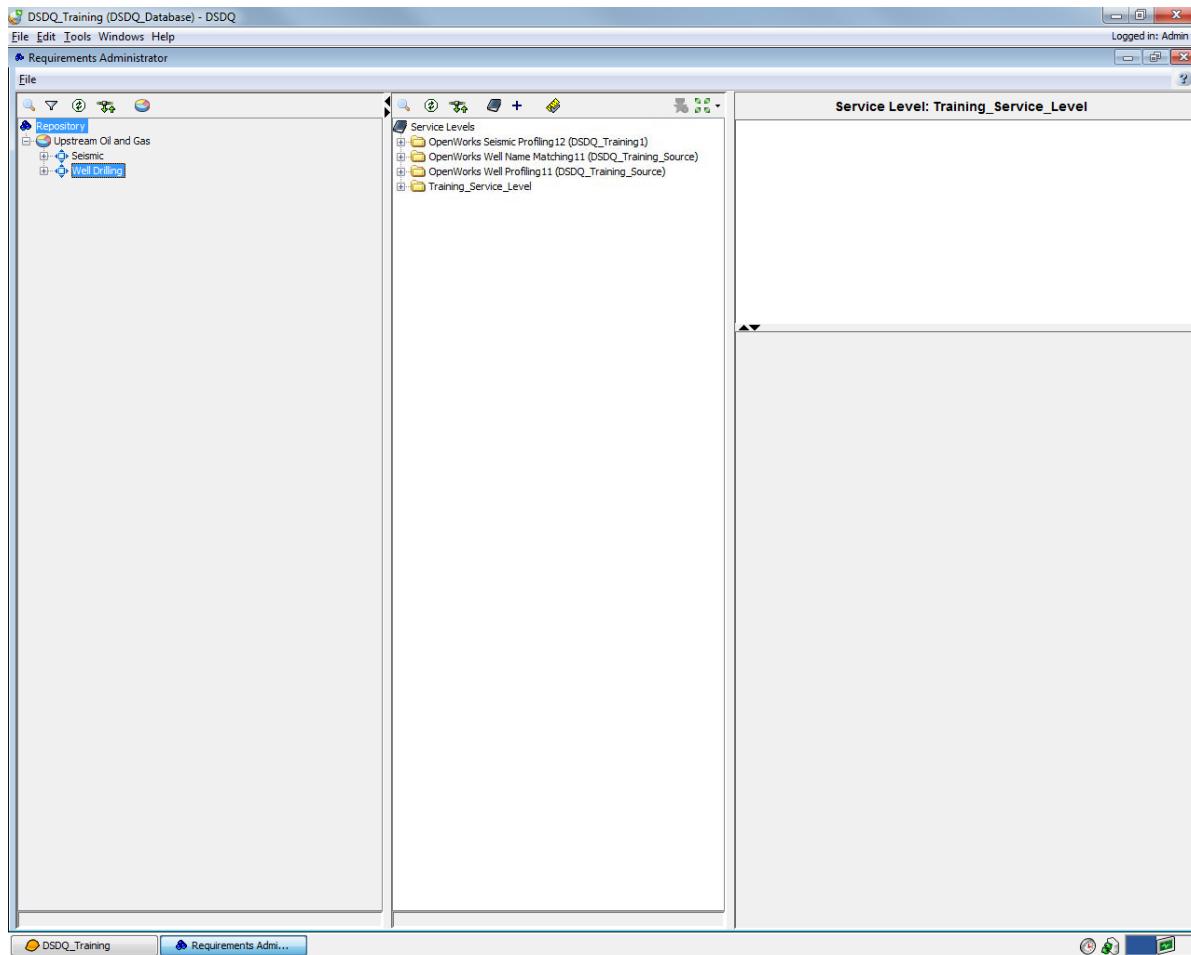


The **Confirm Item Delete** dialog box appears.



3. Click Yes.

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

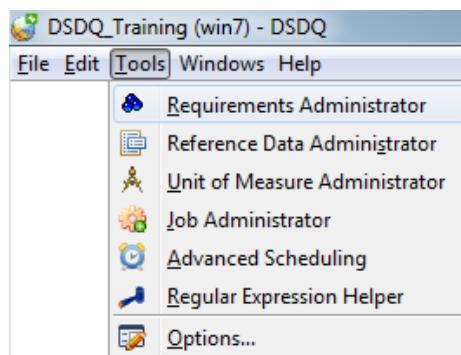


4. Select **File > Exit** from the menu bar on the **Requirements Administrator** window.

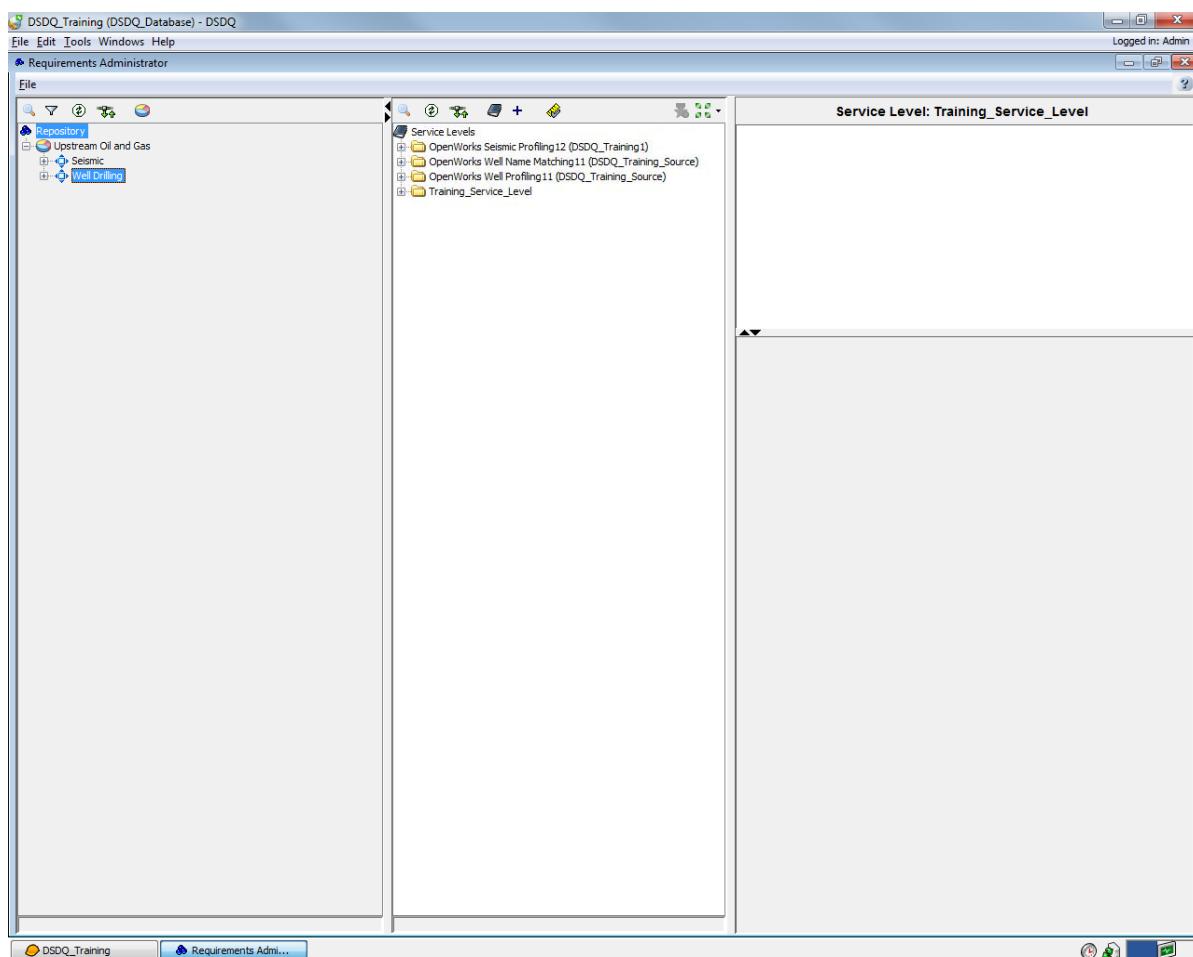
Exercise: Exporting a Service Level

To export a service level:

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

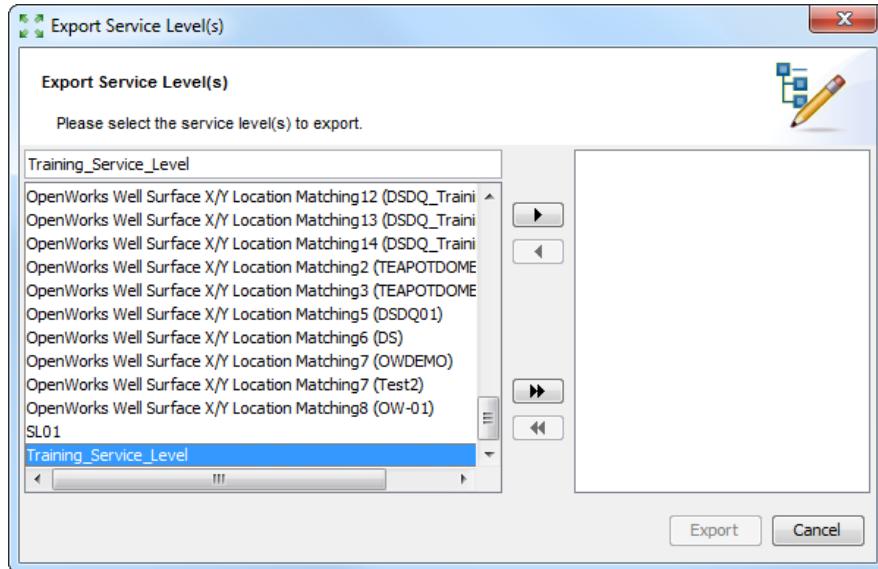


The **Requirements Administrator** window appears.

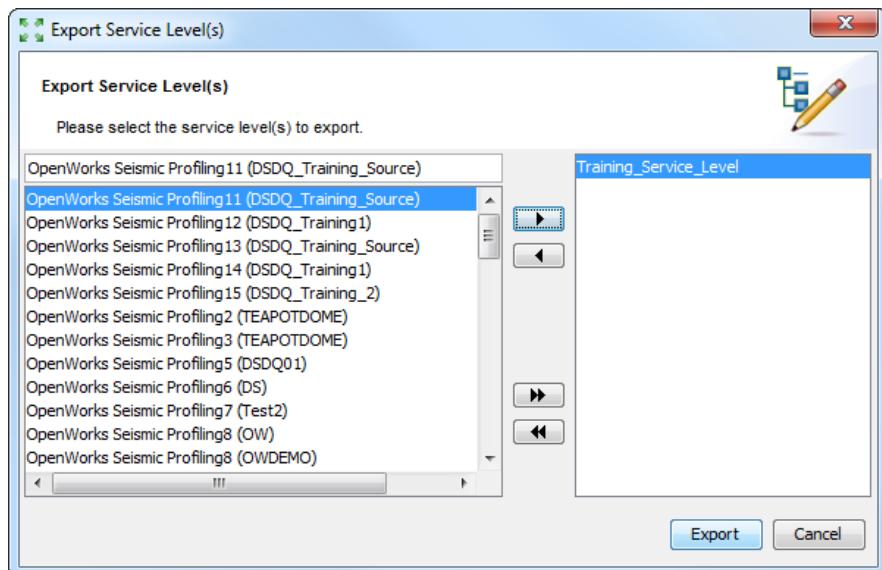


2. Click the arrow ▾ on the **Import/Export Service Levels** icon  and select the **Export Service Levels** option from the drop-down menu.

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

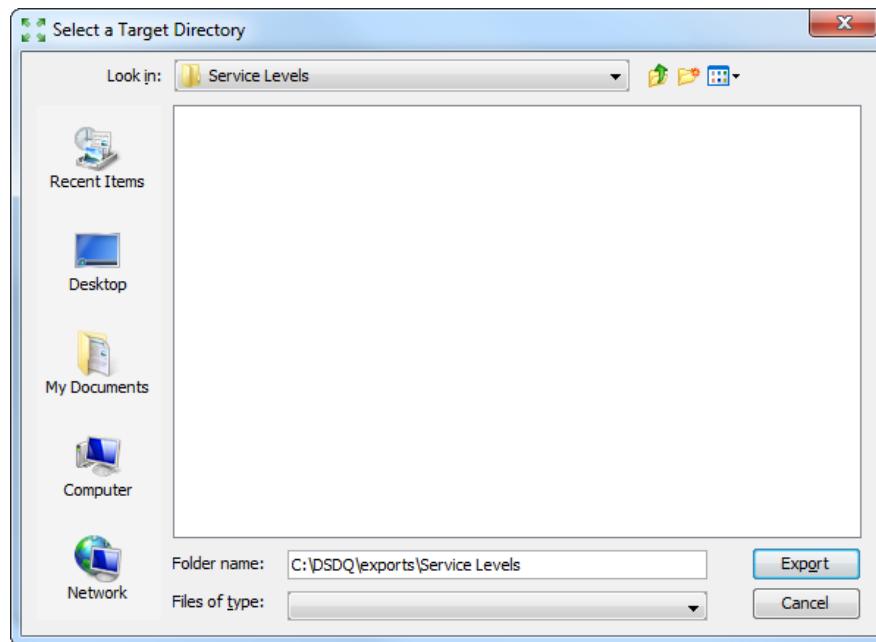


3. Select **Training_Service_Level**.
4. Click  to move the selected service level to the right pane of the **Export Service Level(s)** dialog box.



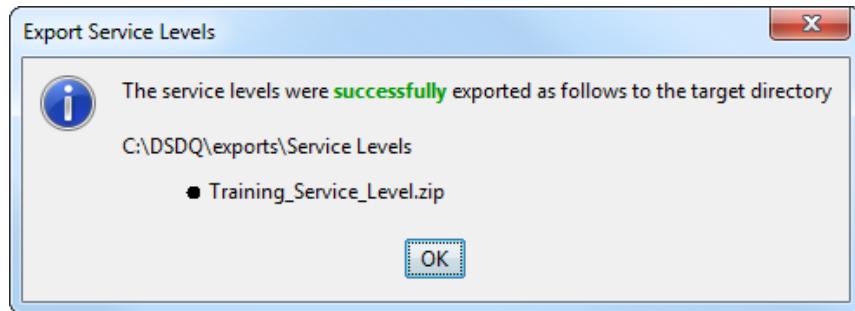
5. Click **Export**.

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.

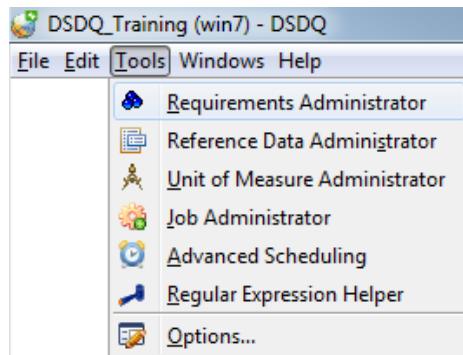
Note

This functionality is only accessible to users with Administrator rights.

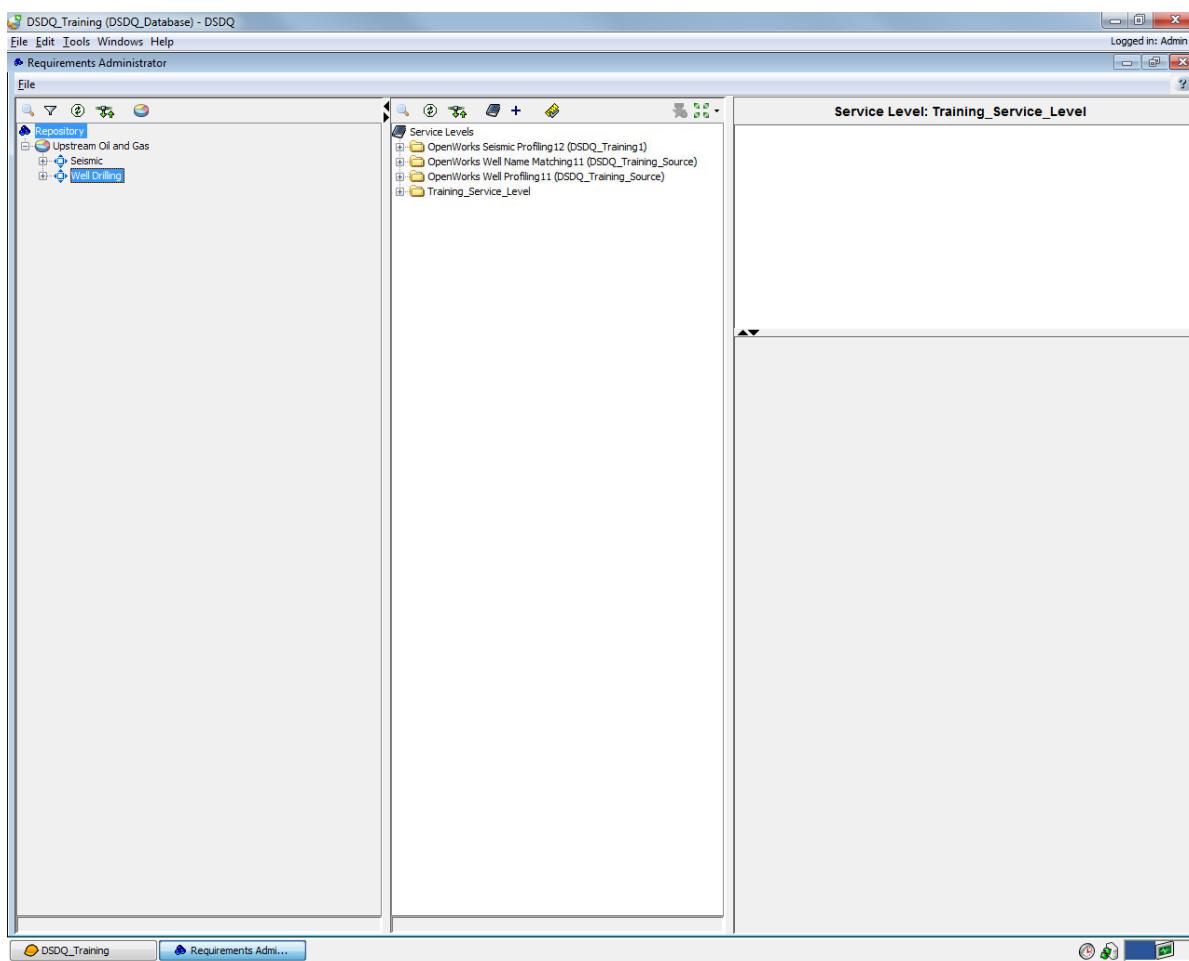
Exercise: Importing a Service Level

To import a service level:

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

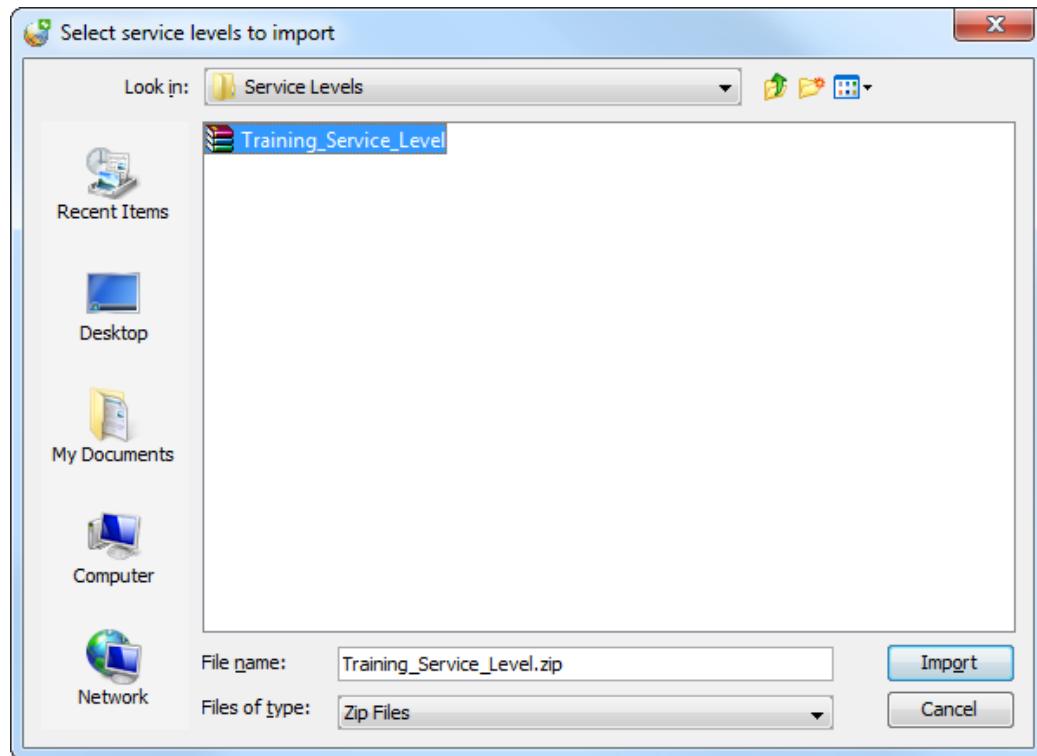


The **Requirements Administrator** window appears.

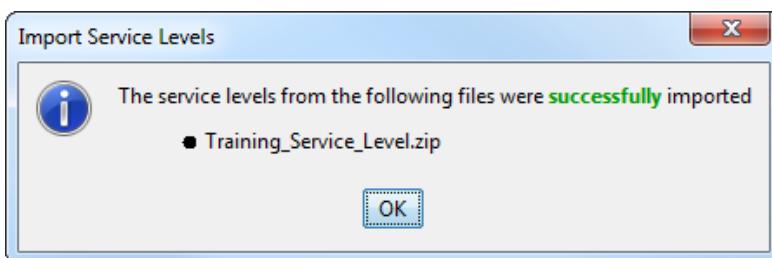


2. Click the arrow ▾ on the **Import/Export Service Levels** icon  and select the **Import Service Levels** option from the drop-down menu.

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:

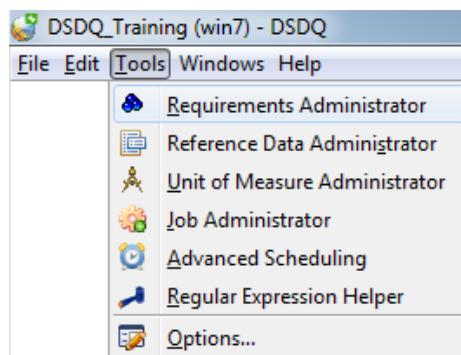
- **Sector**  – e.g.: Upstream Oil and Gas
A sector can be identified as an industry. For example, Finance or Utilities.
- **Area**  – e.g.: Well Drilling
An area is a smaller division within the sector. For example, a Financial Sector can have Financial Accounting and Production Accounting areas, as 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.

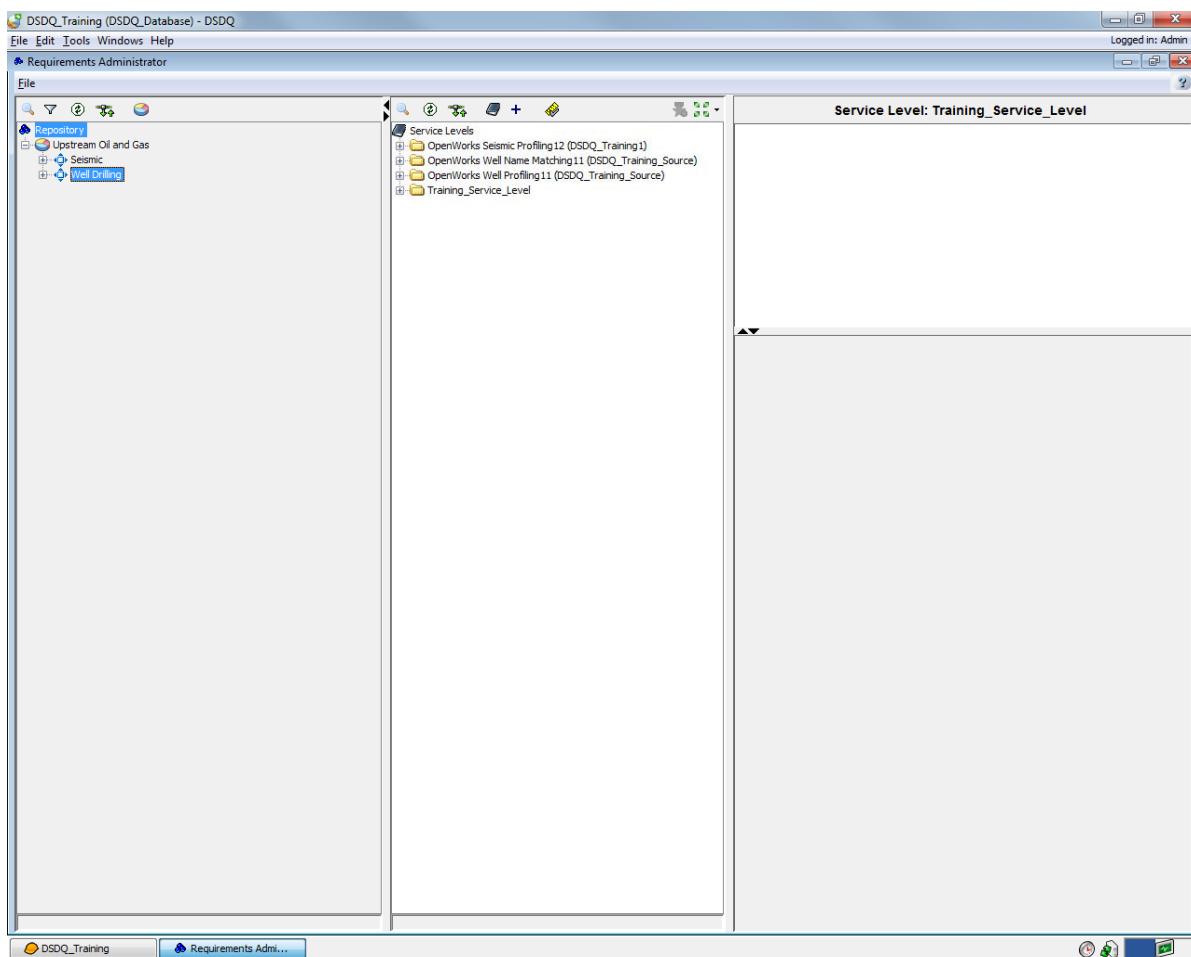
Exercise: Adding a Component

To add a component:

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



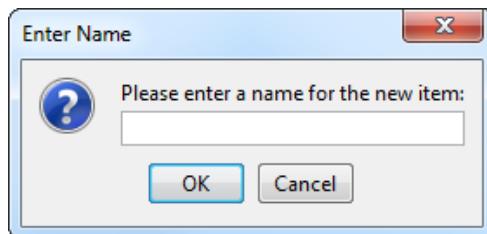
The Requirements Administrator window appears.



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



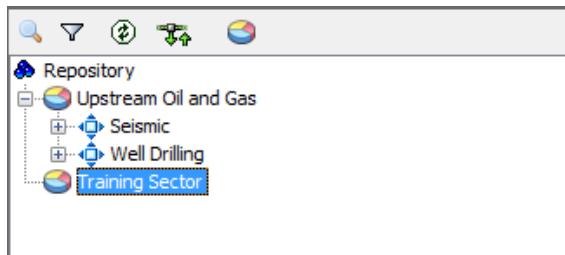
The **Enter Name** dialog box appears.



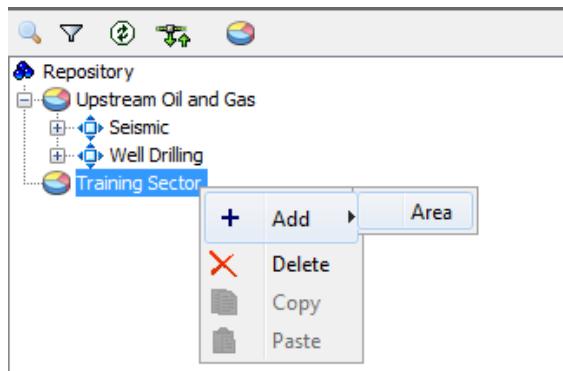
3. Enter **Training Sector** in the **Please enter a name for the new item** field.

4. Click **OK**.

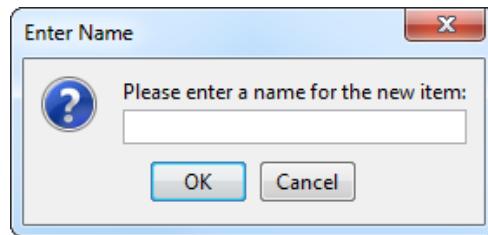
The sector is added and displays in the Repository Tree.



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

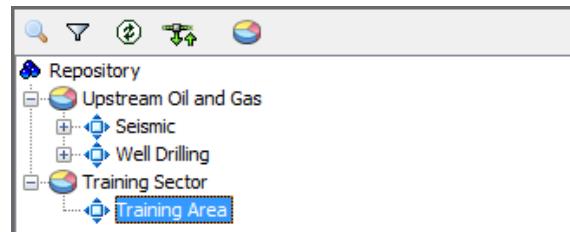


The **Enter Name** dialog box appears.

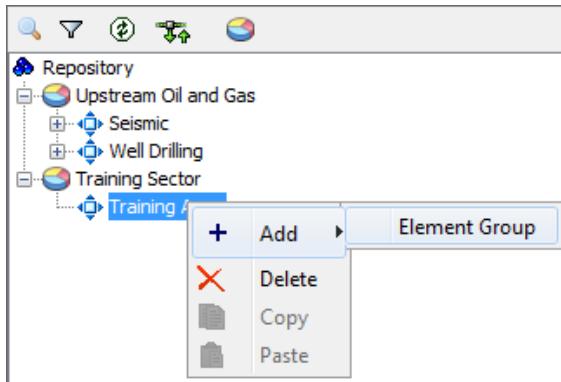


6. Enter **Training Area** in the **Please enter a name for the new item** field.
7. Click **OK**.

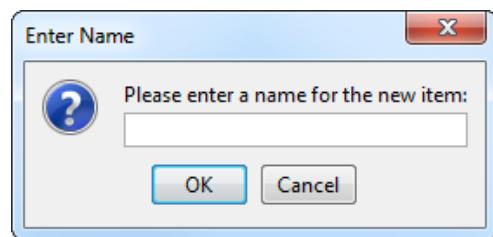
The area is added and displays in the Repository Tree.



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

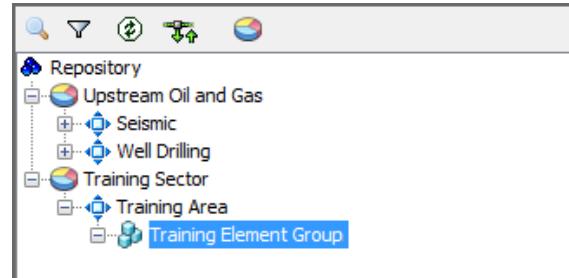


The **Enter Name** dialog box appears.

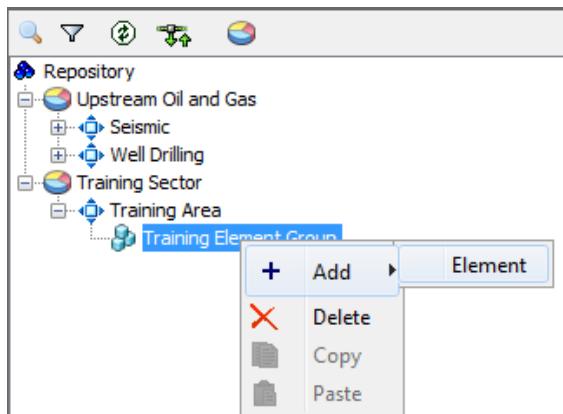


9. Enter **Training Element Group** in the **Please enter a name for the new item** field.
10. Click **OK**.

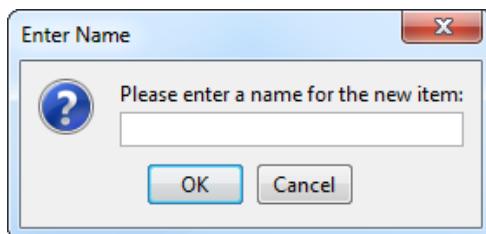
The element group is added and displays in the Repository Tree.



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



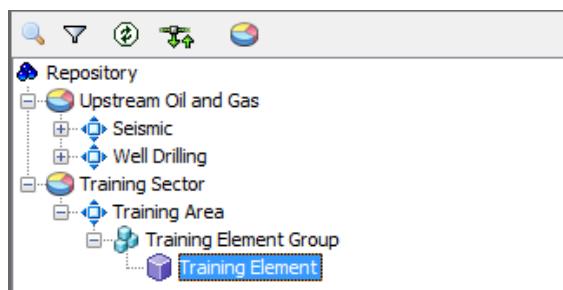
The **Enter Name** dialog box appears.



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.

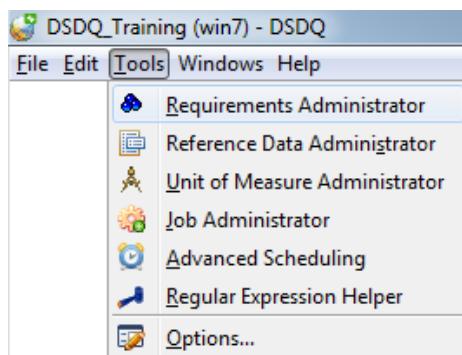


14. Select **File > Exit** from the menu bar on the **Requirements Administrator** window.

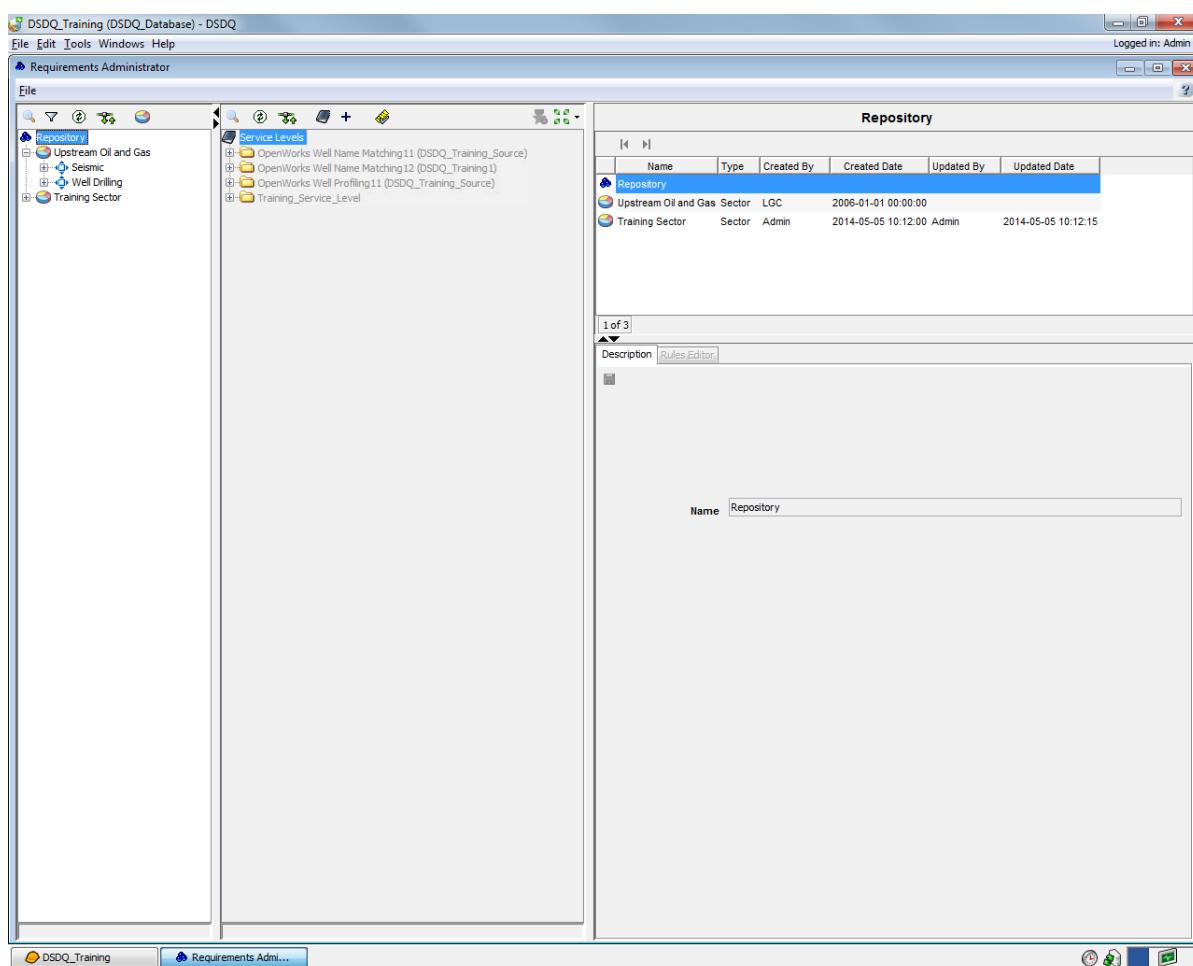
Exercise: Editing a Component

To edit a component:

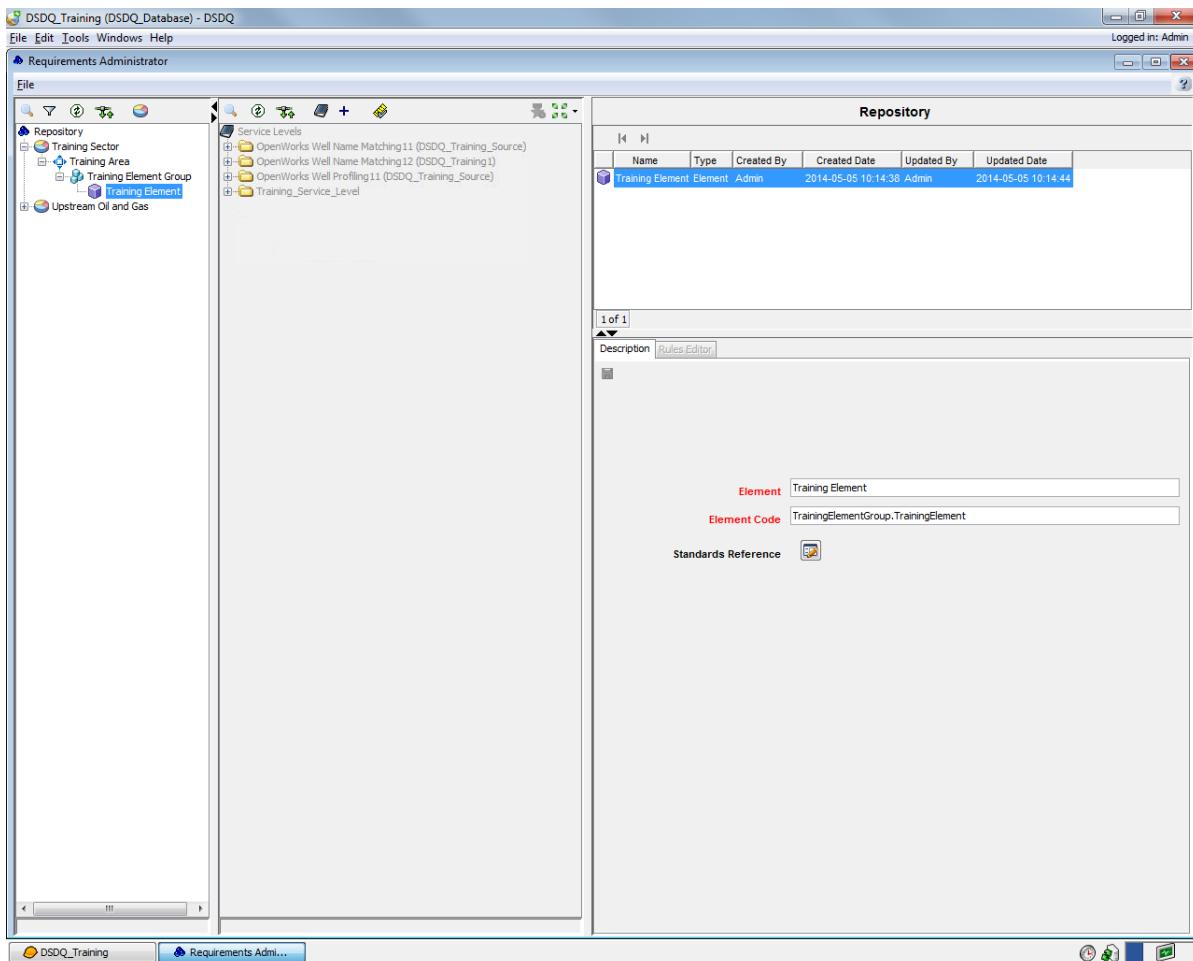
1. Select **Tools > Requirements Administrator** from the menu bar on the **DSDQ Project** window.



The **Requirements Administrator** window appears.



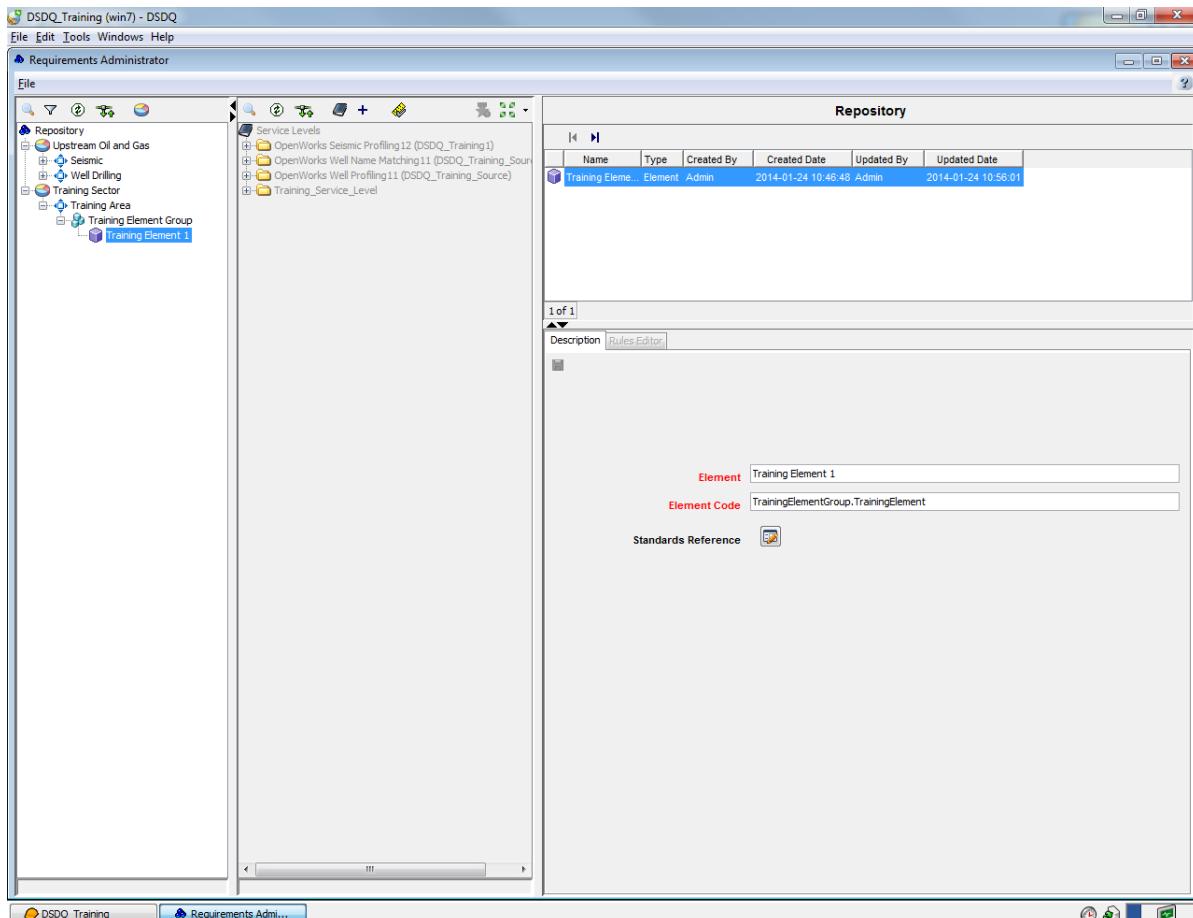
2. Click to expand **Training Sector** in the **Repository Tree**.
3. Click to expand **Training Area**.
4. Click to expand **Training Element Group**.
5. Select **Training Element** in the **Repository Tree**.
Information about **Training Element** appears in the **Description** tab of the Details Pane.



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 .

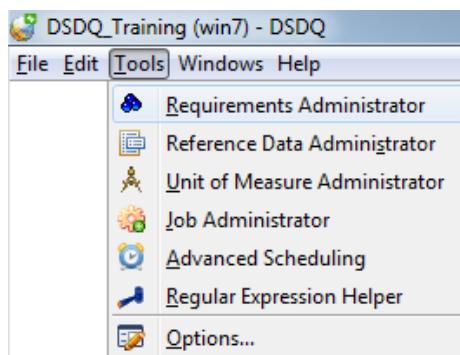
The updated element displays in the Repository Tree Pane.



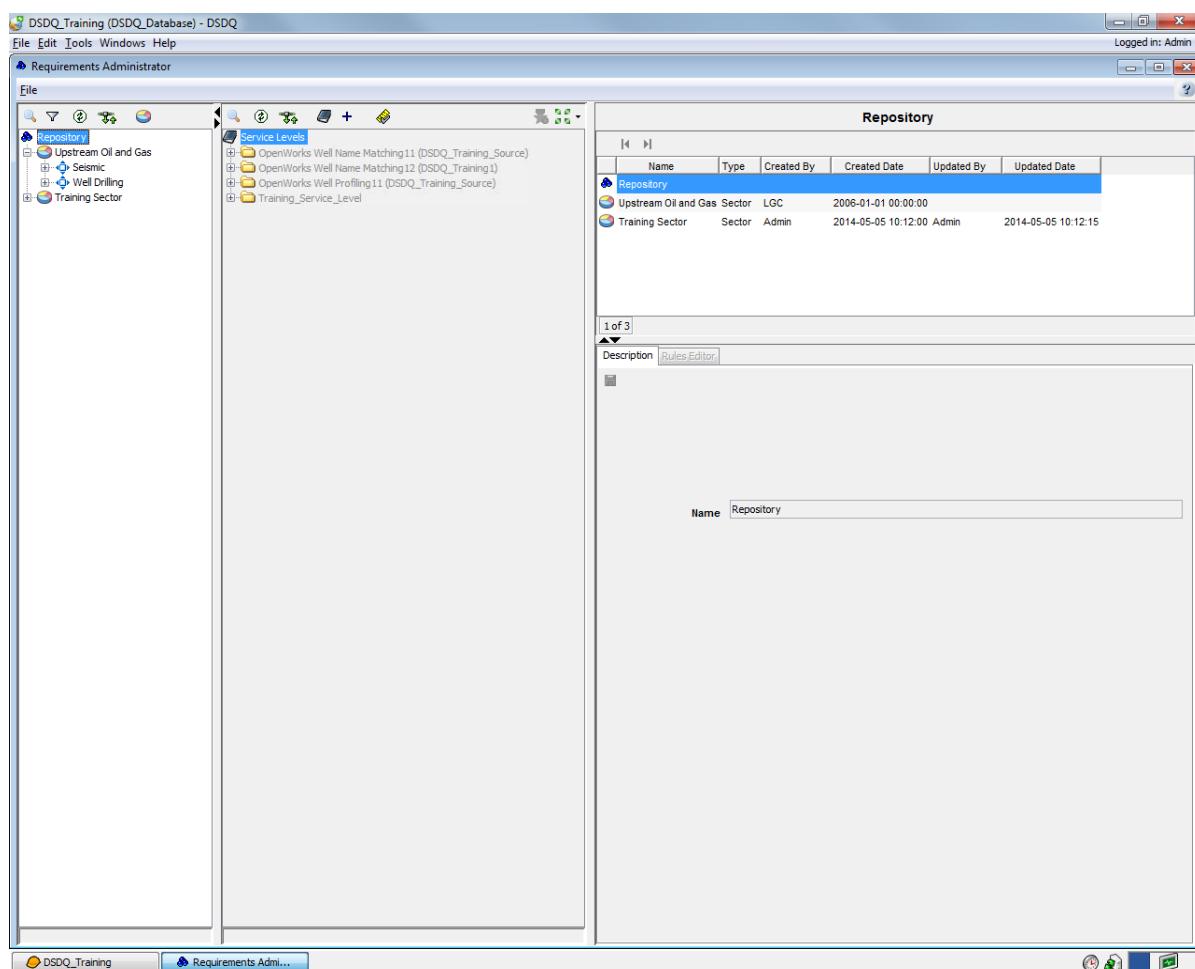
Exercise: Deleting a Component

To delete a component:

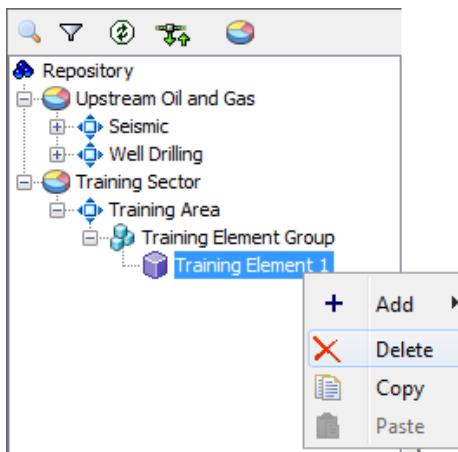
1. Select **Tools >Requirements Administrator** from the menu bar on the **DSDQ Project** window.



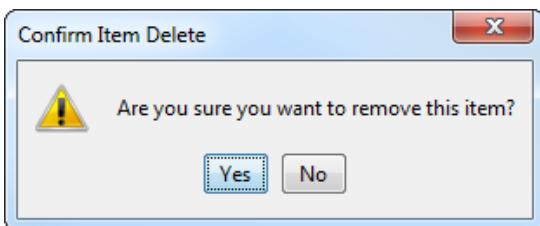
The **Requirements Administrator** window appears.



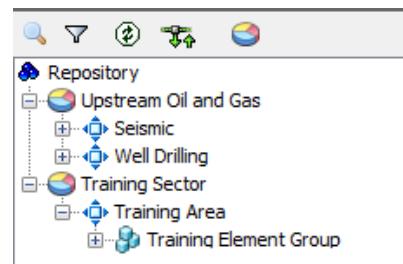
2. Click to expand **Training Sector** in the **Repository Tree**.
3. Click to expand **Training Area**.
4. Click to expand **Training Element Group**.
5. Right-click **Training Element** and select **Delete** from the pop-up menu.



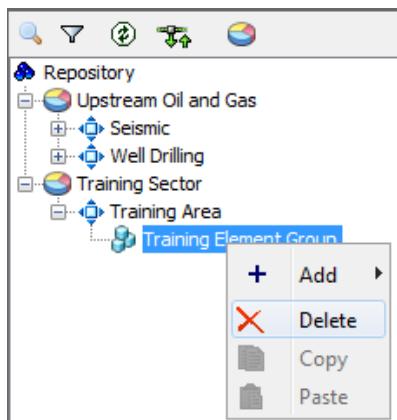
The **Confirm Item Delete** dialog box appears.



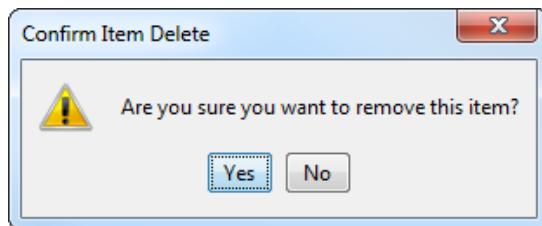
6. Click **Yes**.
The selected element is deleted.



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

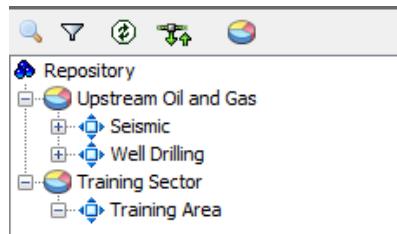


The **Confirm Item Delete** dialog box appears.

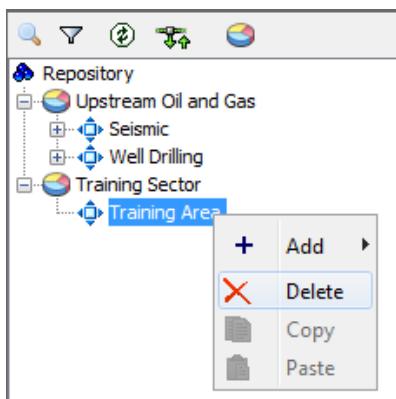


8. Click **Yes**.

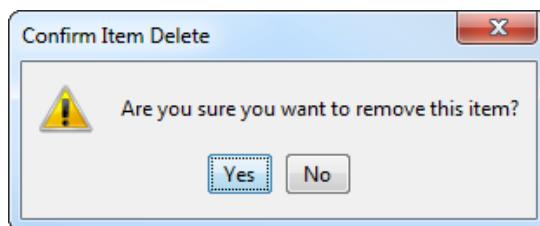
The selected element group is deleted.



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



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

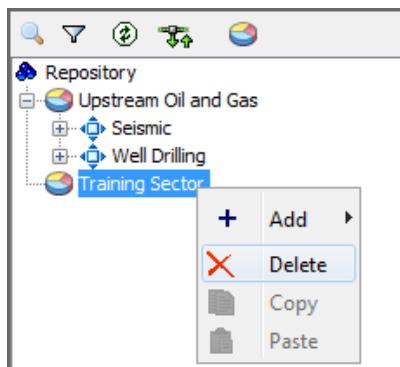


10. Click **Yes**.

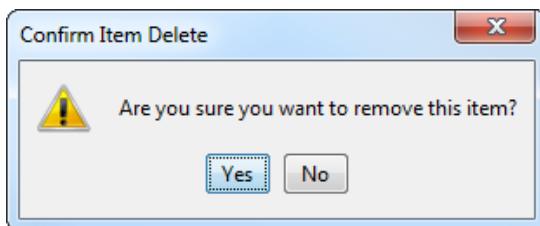
The selected area is deleted.



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.

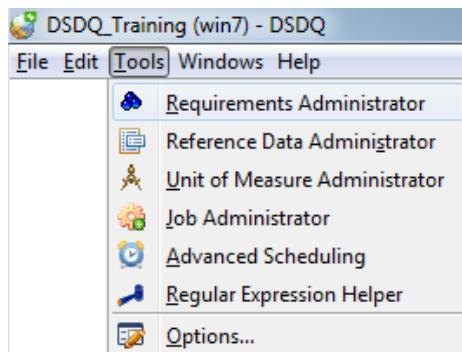


13. Select **File > Exit** from the menu bar on the **Requirements Administrator** window.

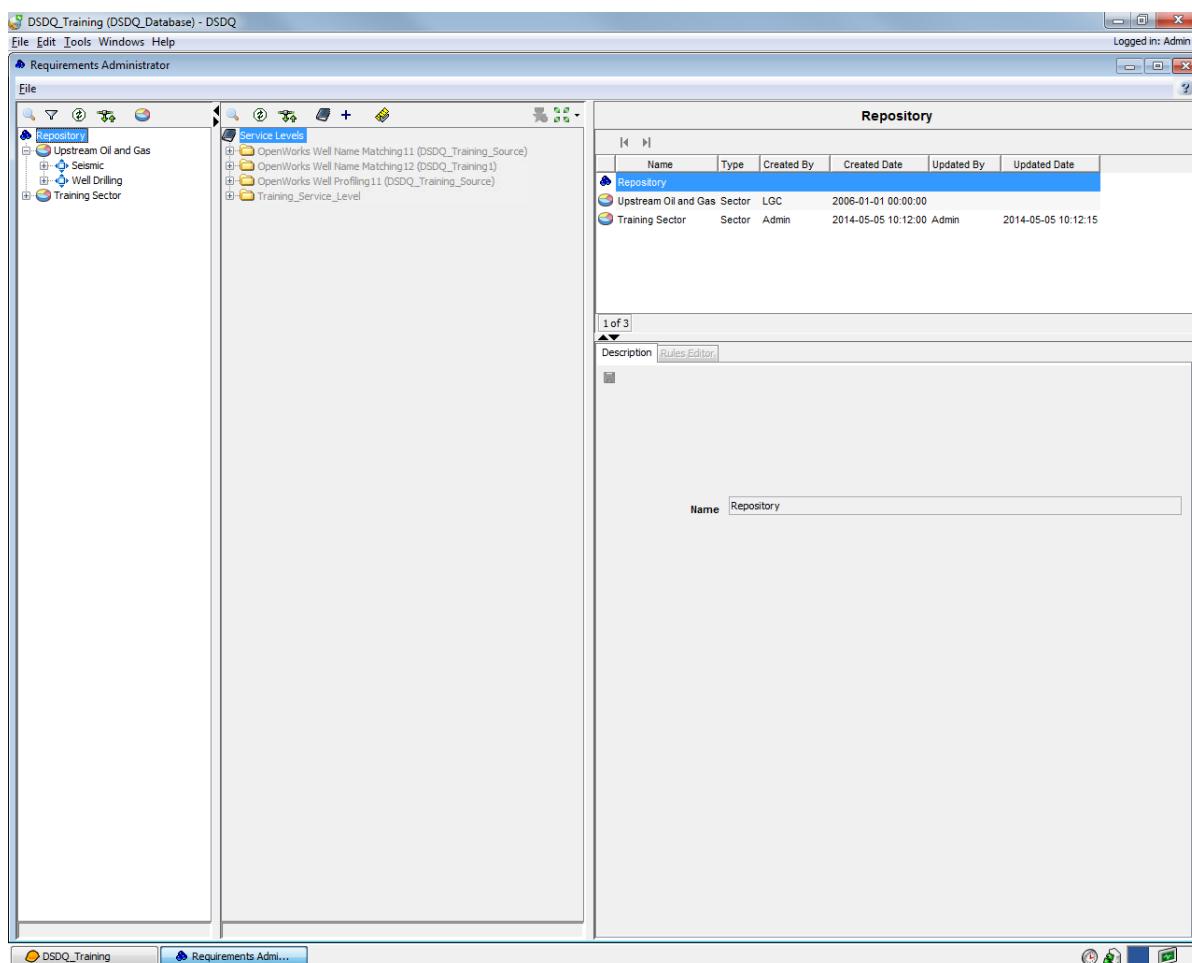
Exercise: Associating an Element with a Standards Reference Table

To associate an element with a Standards Reference Table:

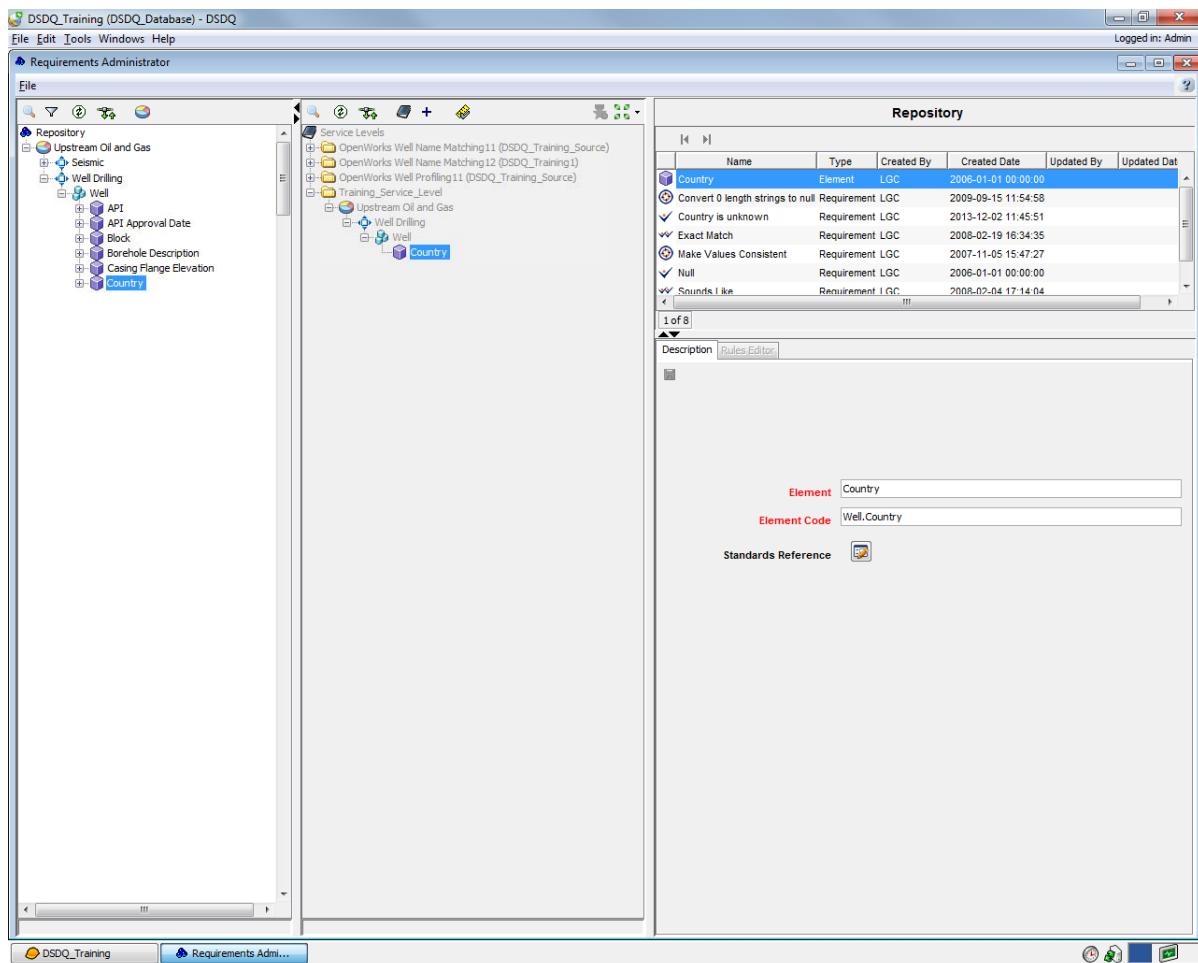
1. Select **Tools > Requirements Administrator** from the menu bar on the **DSDQ Project** window.



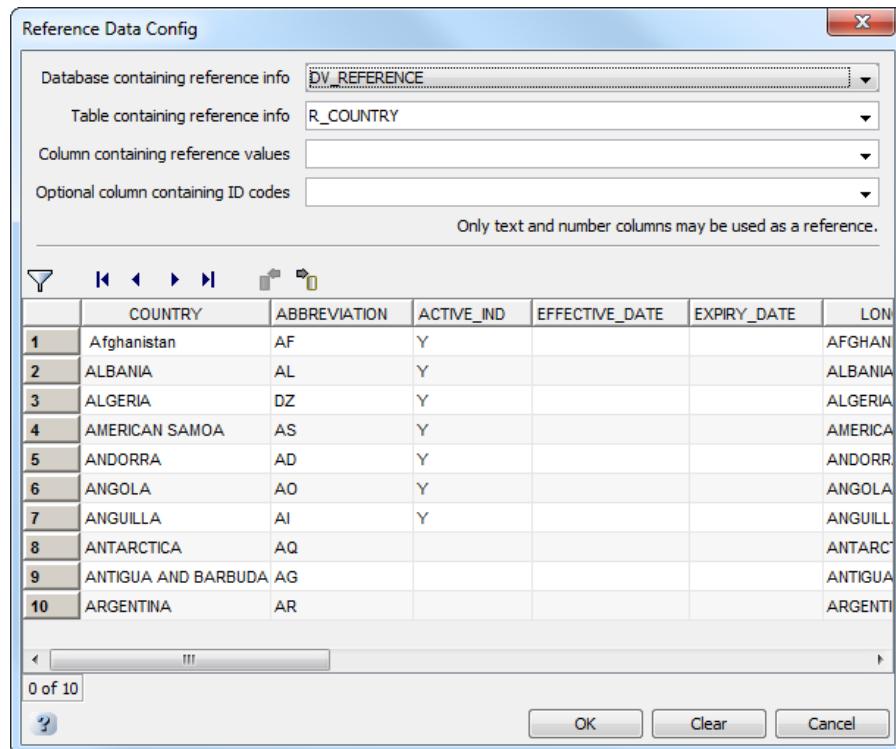
The **Requirements Administrator** window appears.



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.



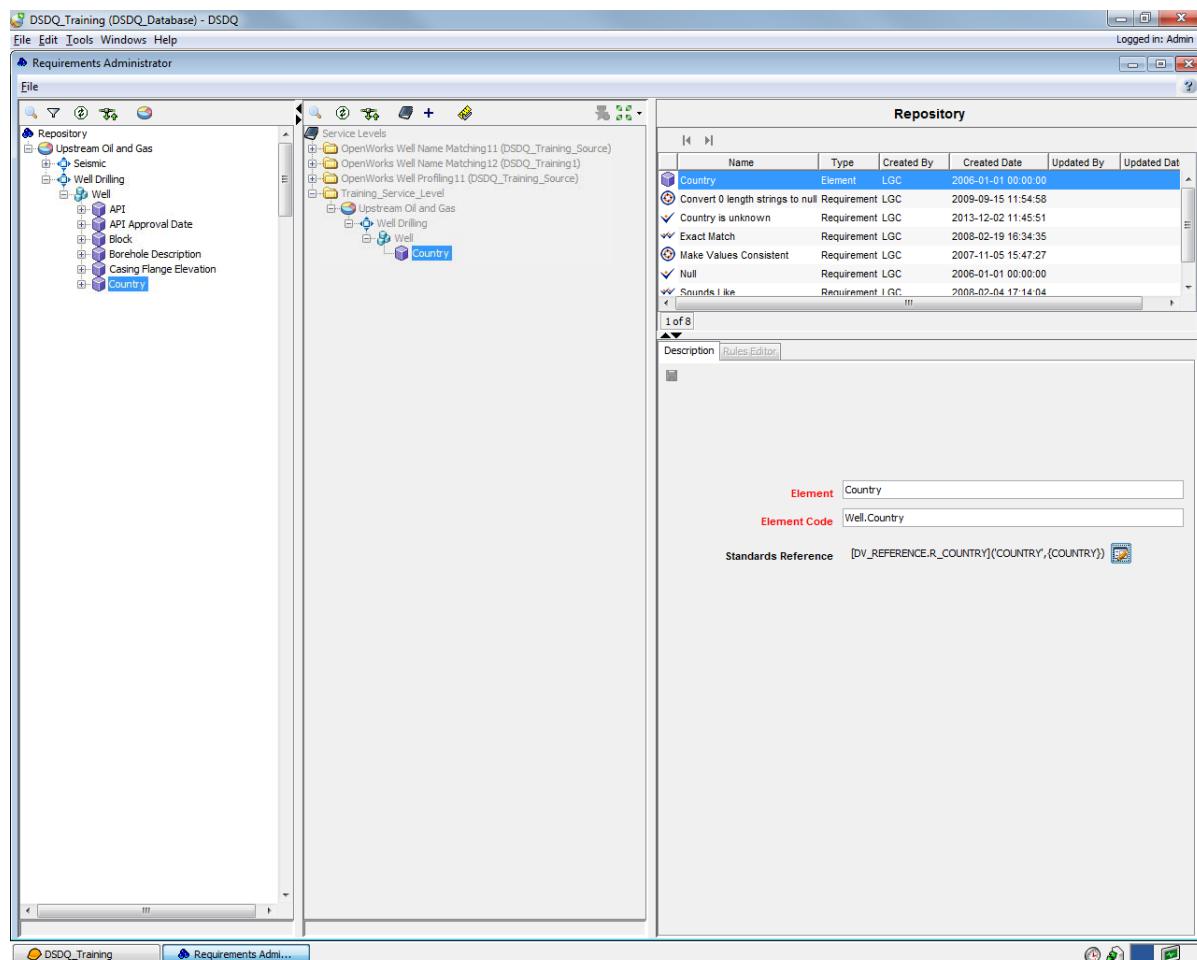
5. Click the **Standard Reference**  button in the **Description** tab.
The **Reference Data Config** window appears.



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.**

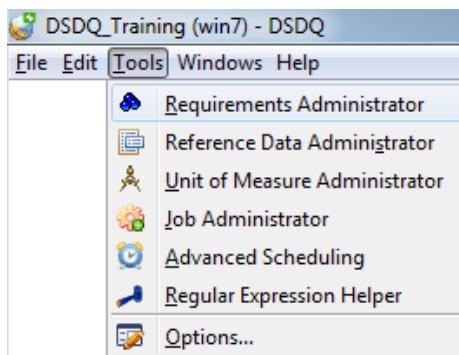
Requirements

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**.

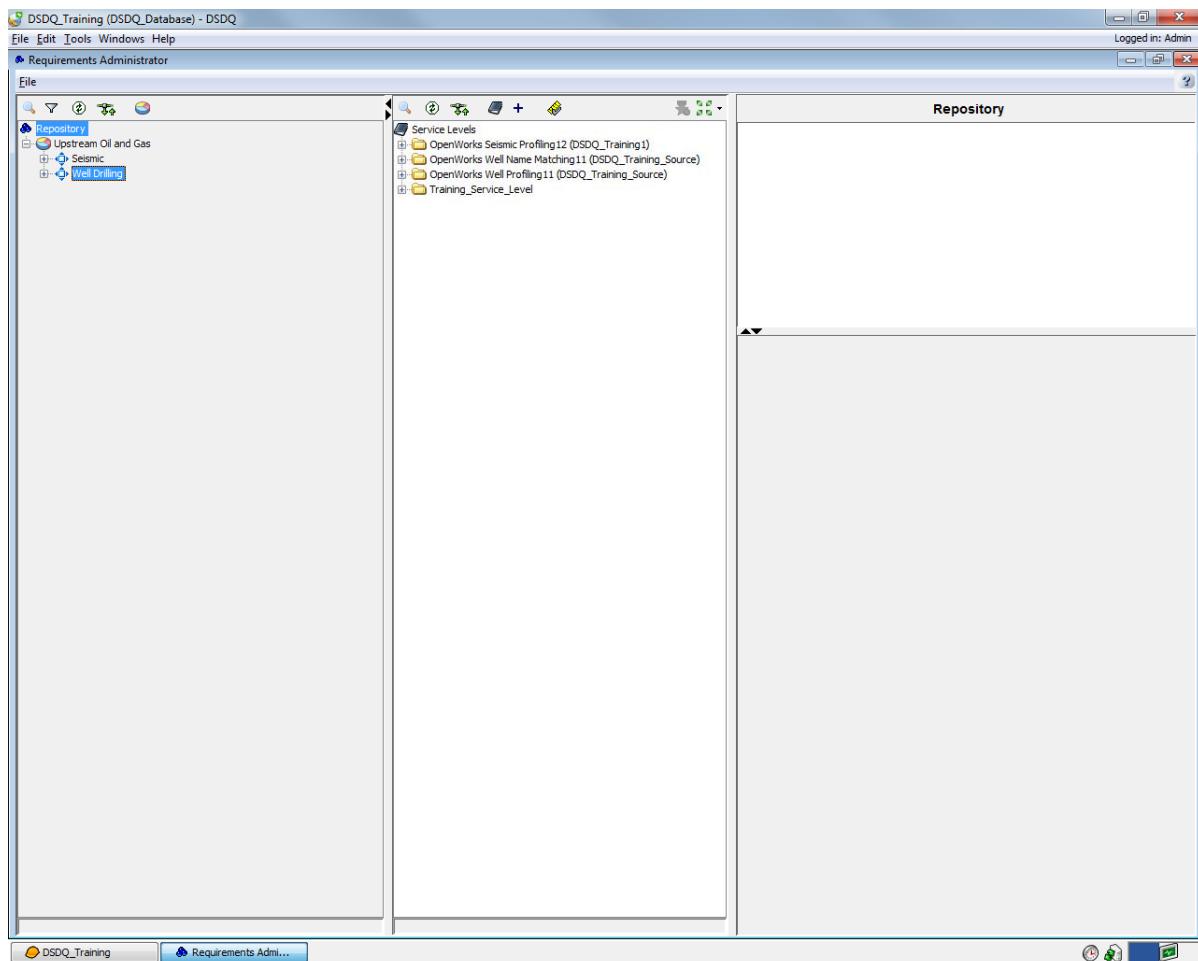
Exercise: Adding a Requirement

To add a requirement:

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

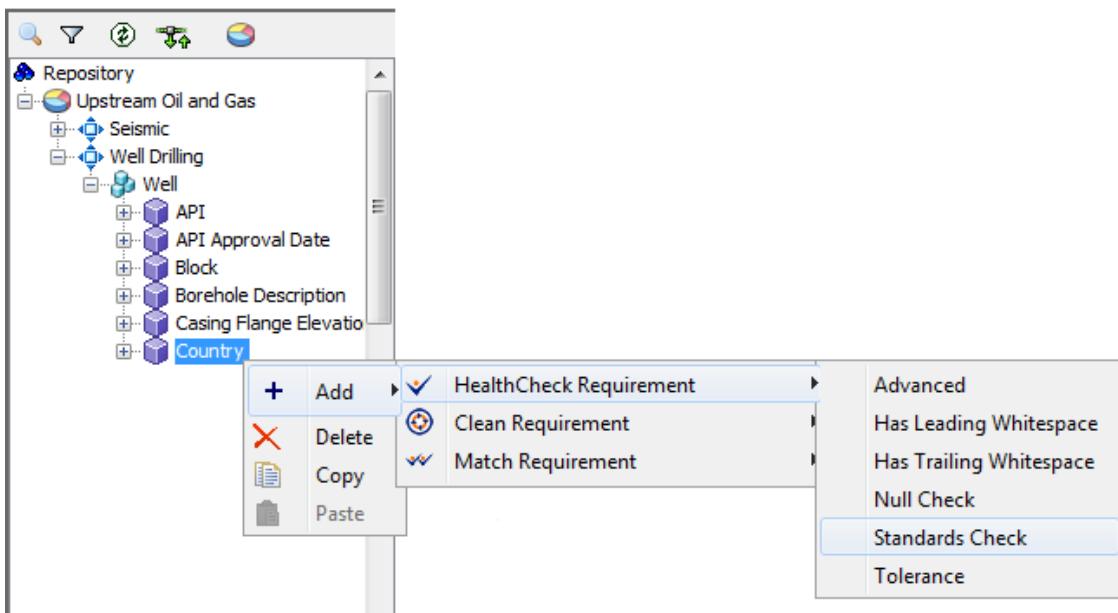


The Requirements Administrator window appears.

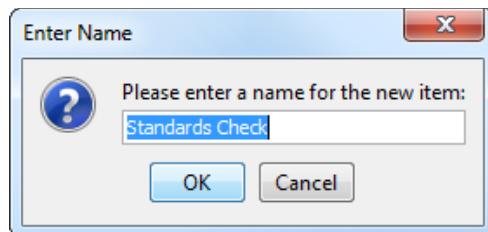


2. Click to expand the **Well Drilling** area in the **Repository Tree**.
3. Click to expand the **Well** element group.
4. Right-click the **Country** element in the **Repository Tree** and select **Add > HealthCheck Requirement > Standards Check** from the

pop-up menu.



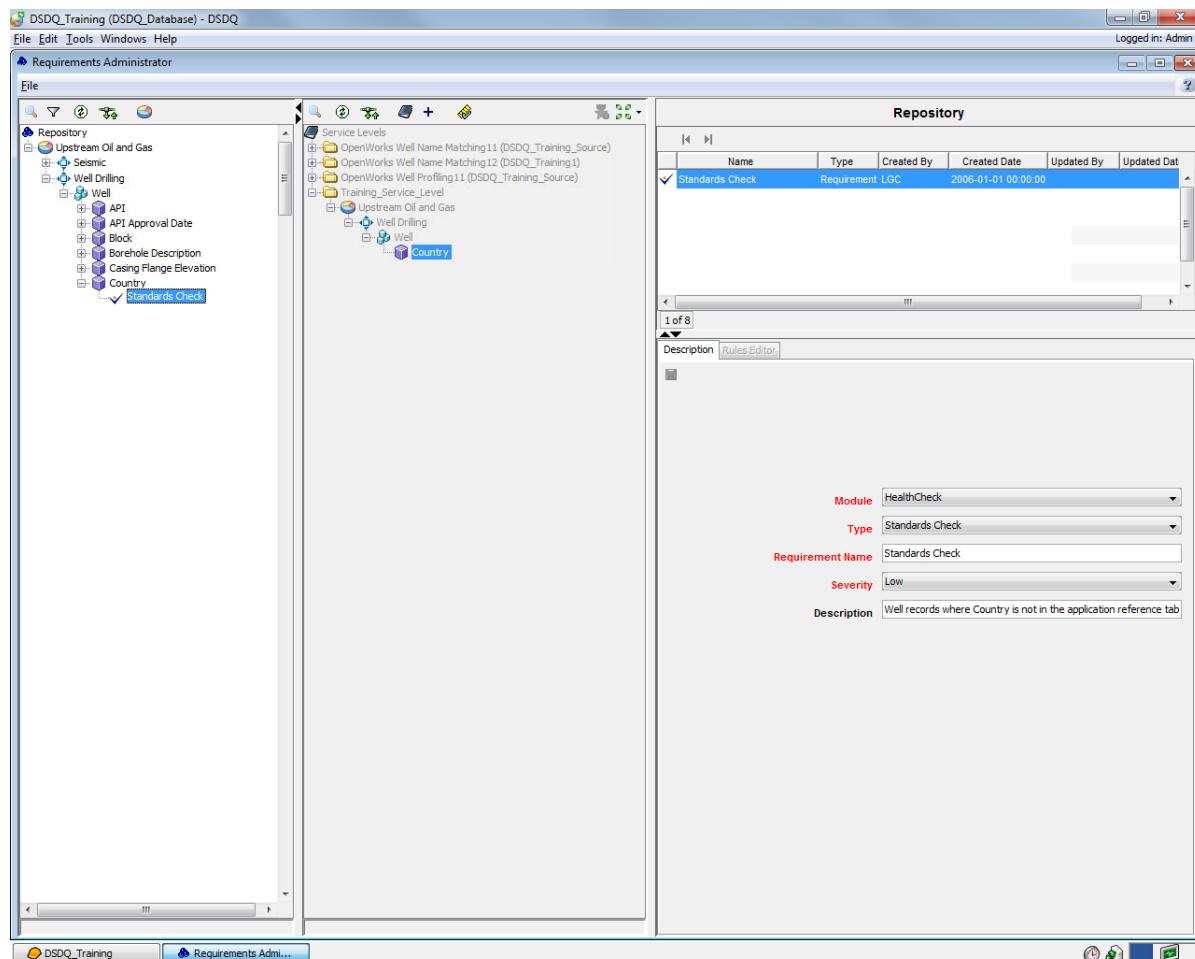
The Enter Name dialog box appears.



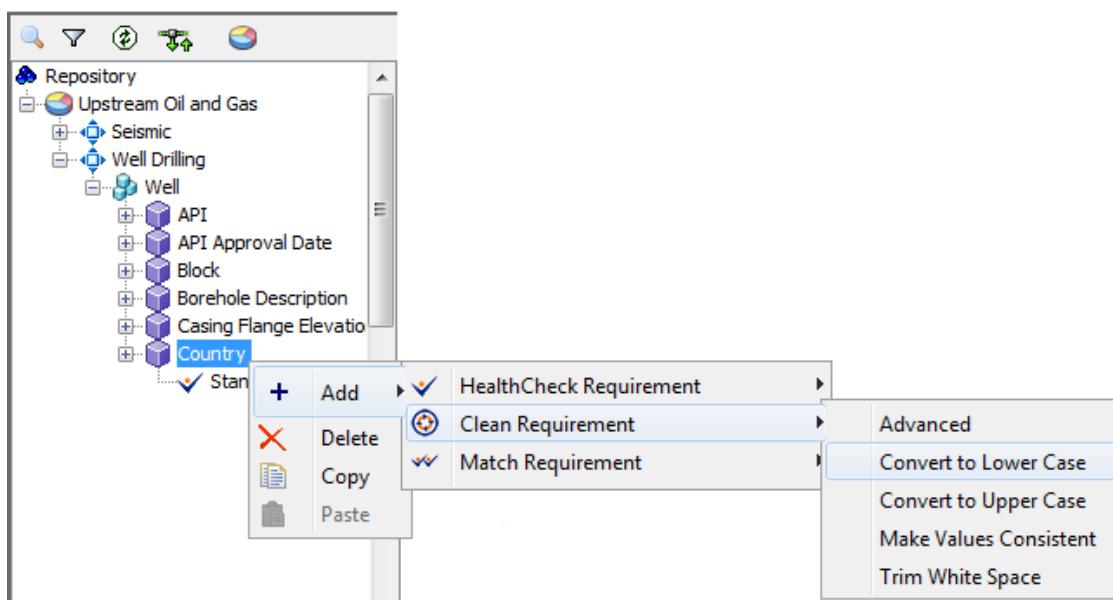
5. Optionally, specify a user-defined name for the requirement.

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.



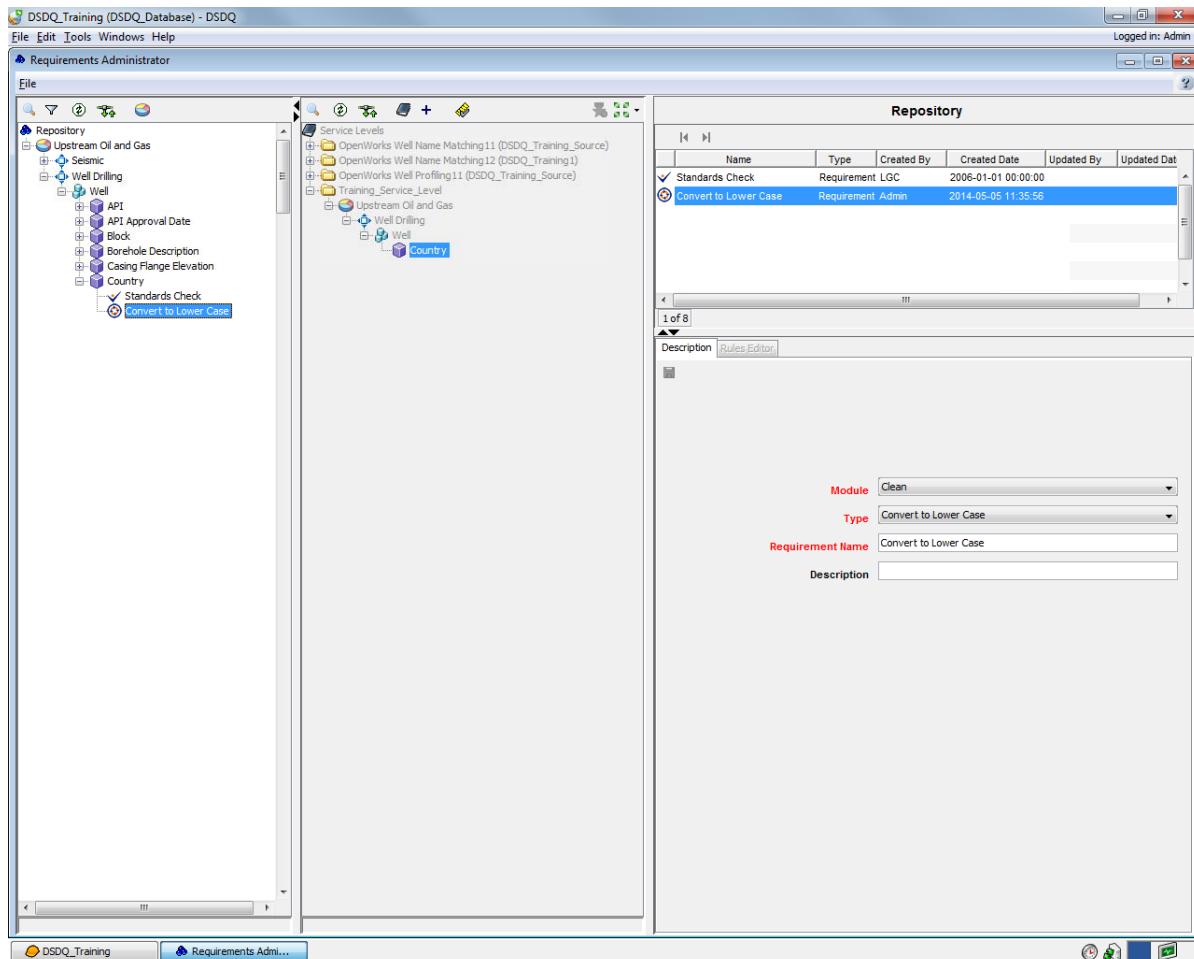
The **Enter Name** dialog box appears.



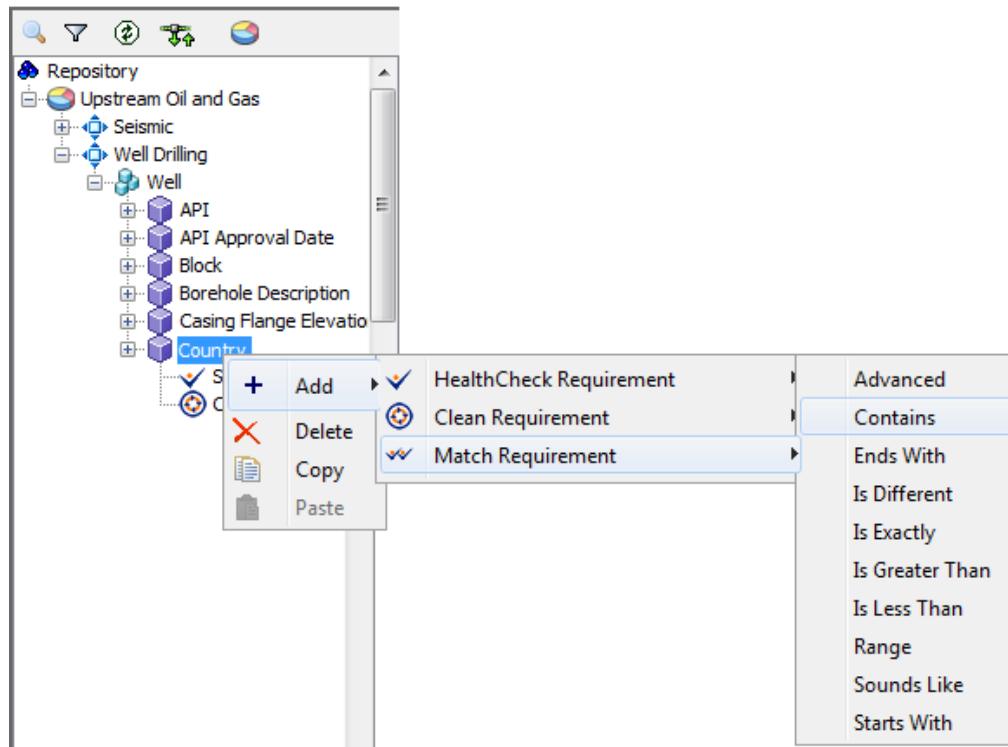
8. Optionally, specify a user-defined name for the requirement.

9. Click **OK**.

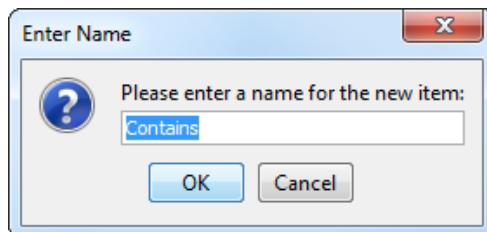
The selected requirement is added to the **Country** element and displays in the **Repository Tree**.



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



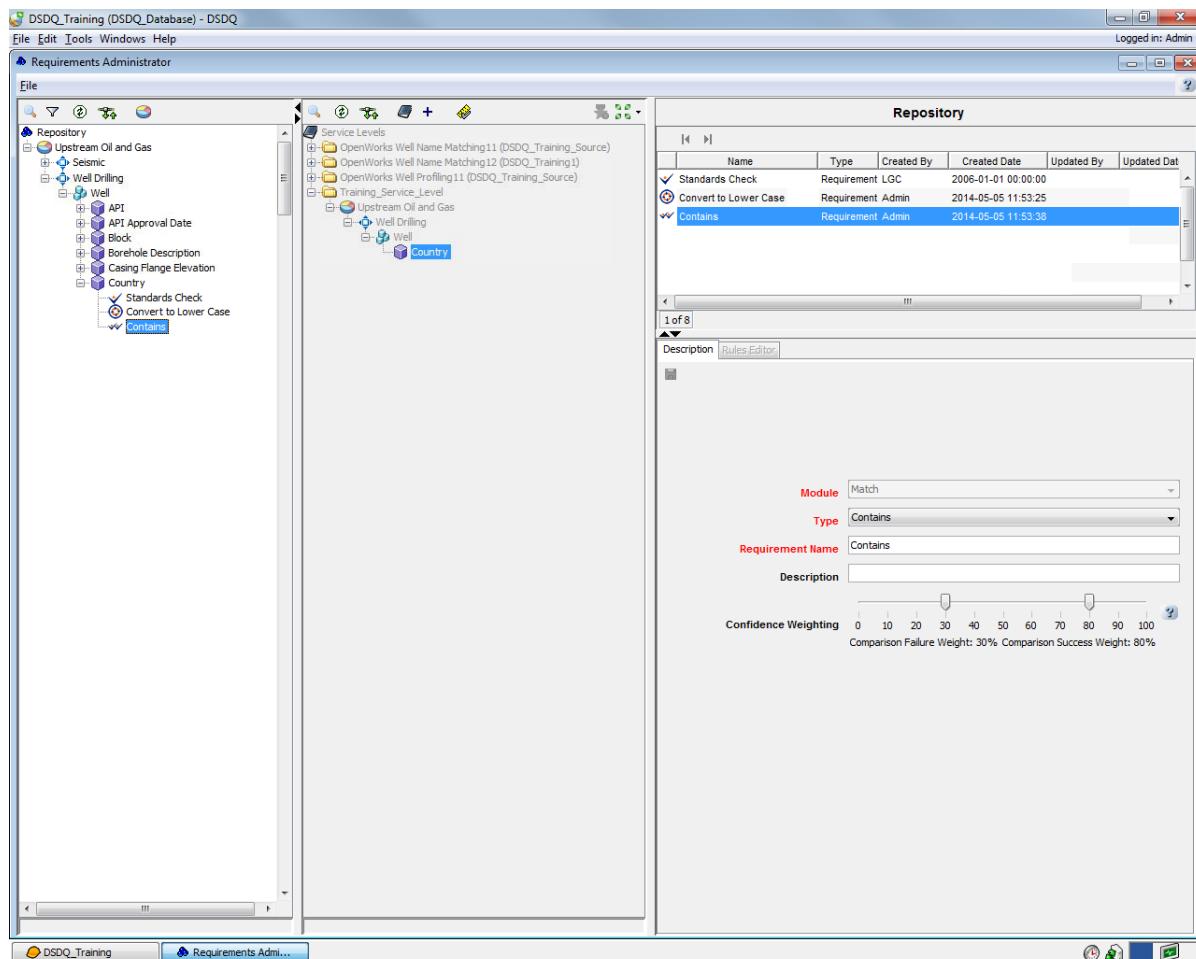
The **Enter Name** dialog box appears.



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**.



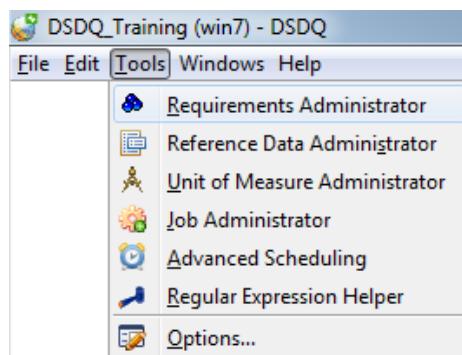
13. Optionally repeat steps 4 to 6 to add more requirements.

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

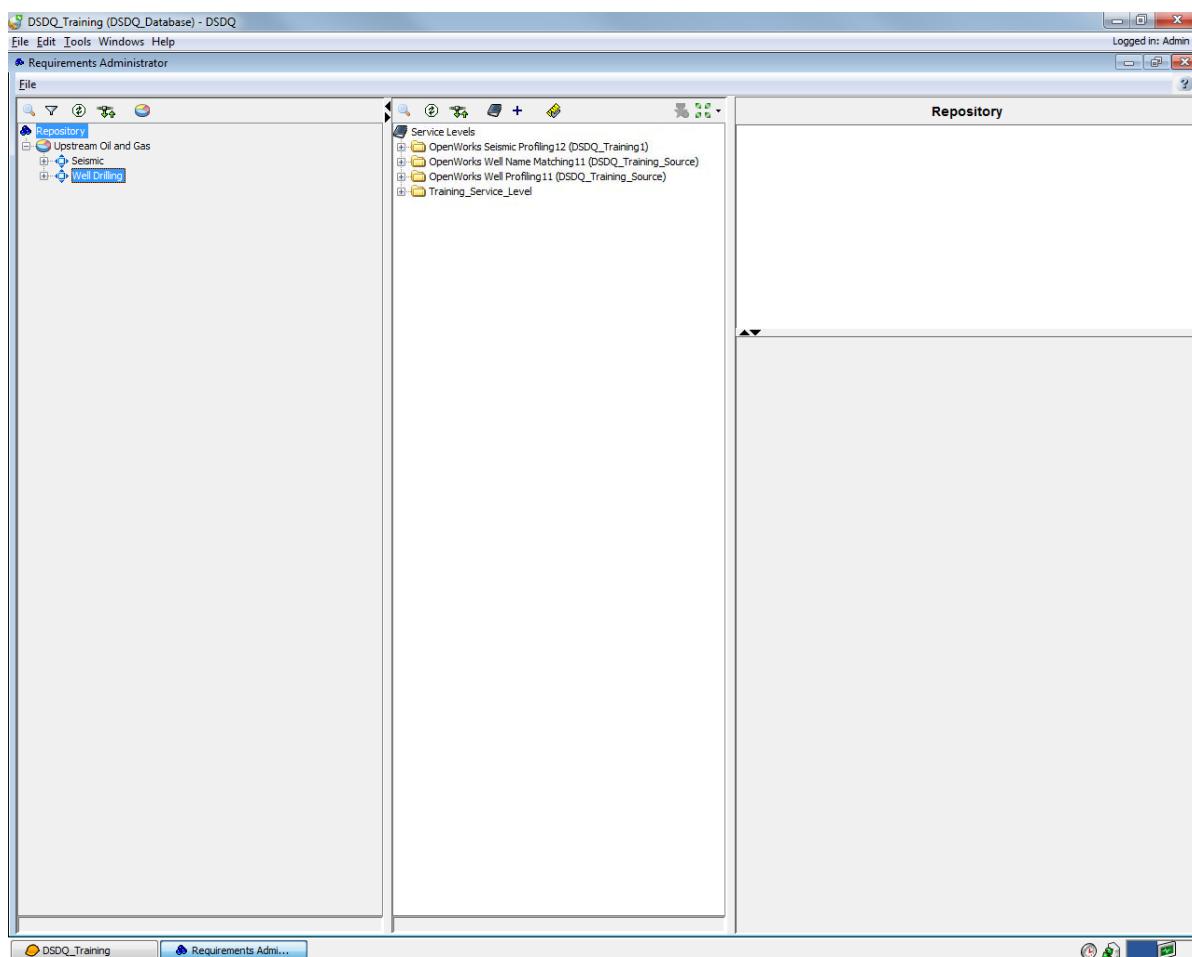
Exercise: Editing a Requirement

To edit a Requirement:

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

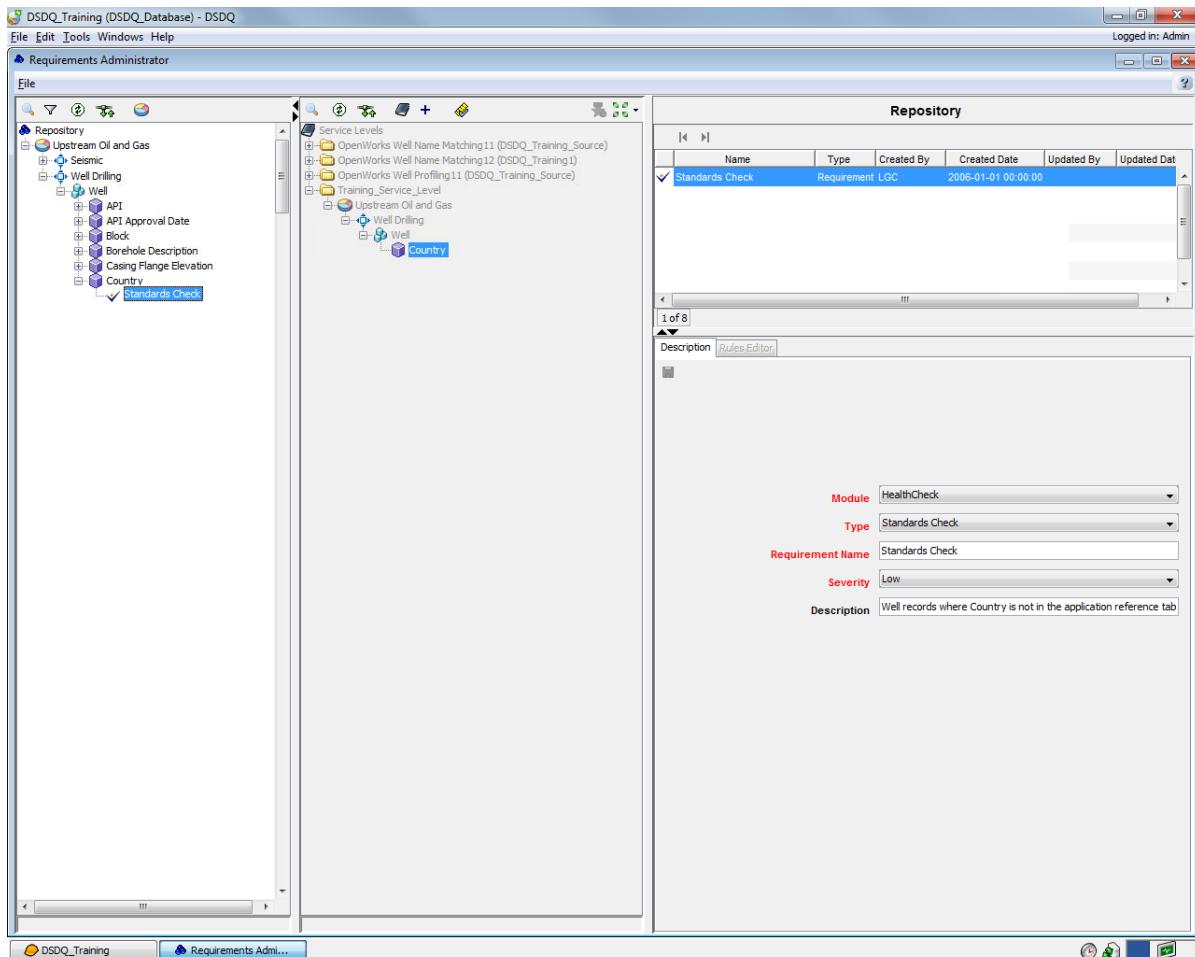


The **Requirements Administrator** window appears.



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**.



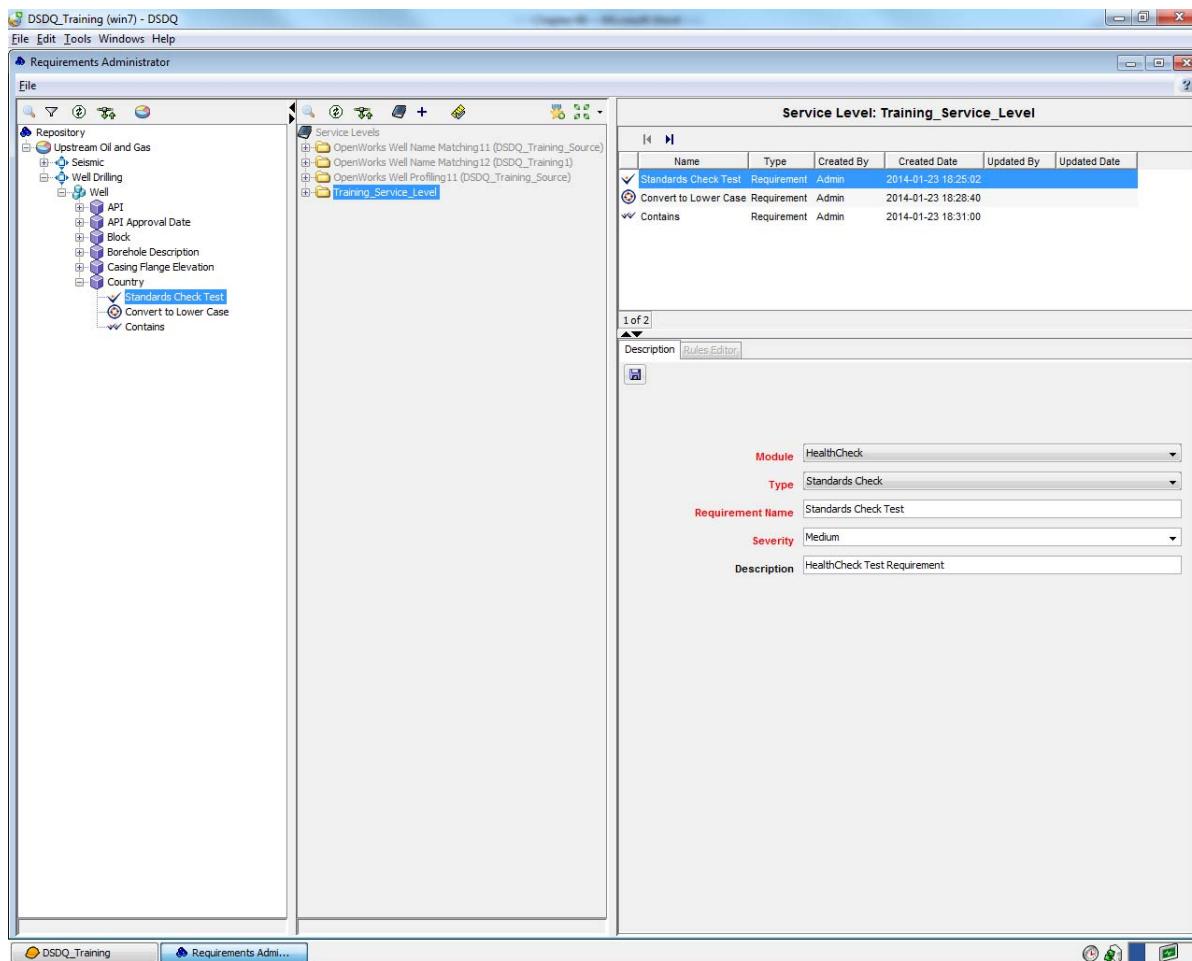
6. Enter **Standards Check** in the **Requirement Name** field of the **Description** tab.

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

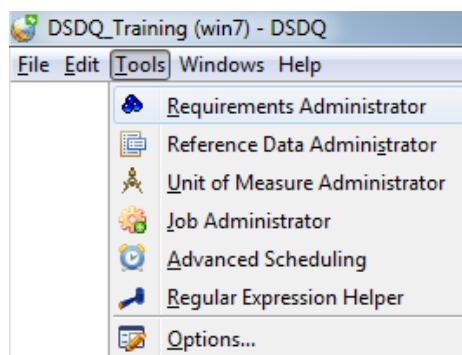
To duplicate a requirement, right-click on the desired requirement and select **Copy** from the pop-up menu. Right-click on the Element and select **Paste** from the pop-up menu to copy the requirement to the desired Element..

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

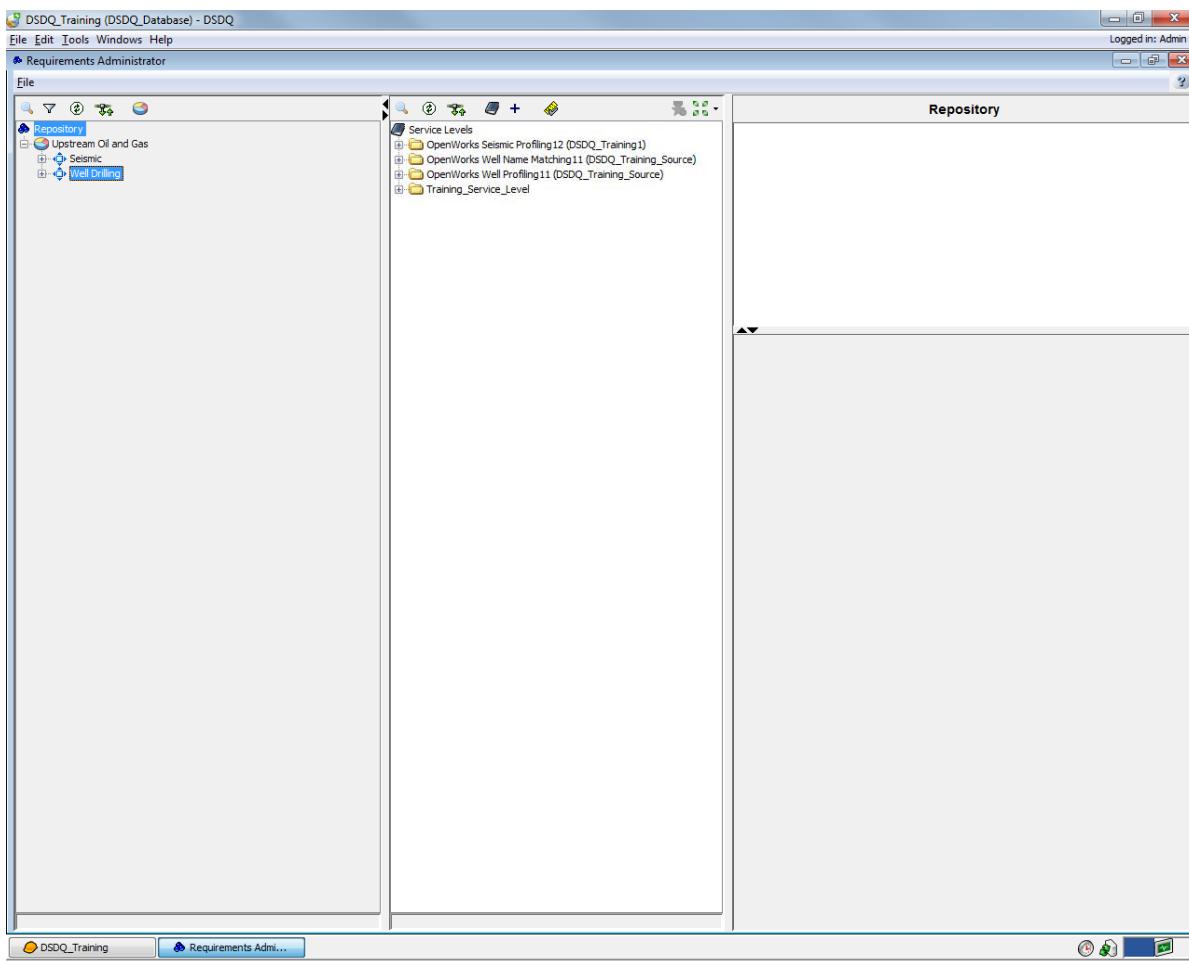
Exercise: Deleting a Requirement

To delete a Requirement:

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

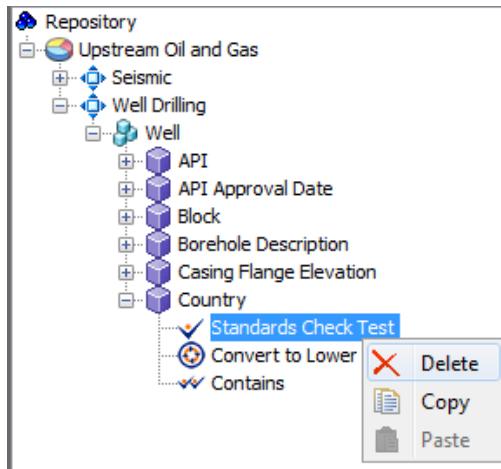


The Requirements Administrator window appears.

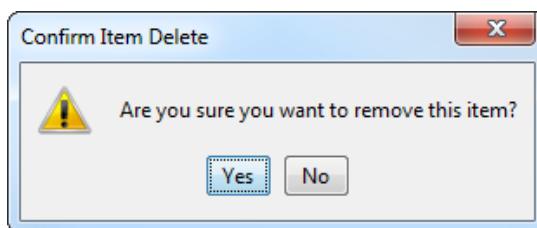


2. Click to expand the **Well Drilling** Area in the Repository Tree.
3. Click to expand the **Well** Element Group.
4. Click to expand the **Country** Element.

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

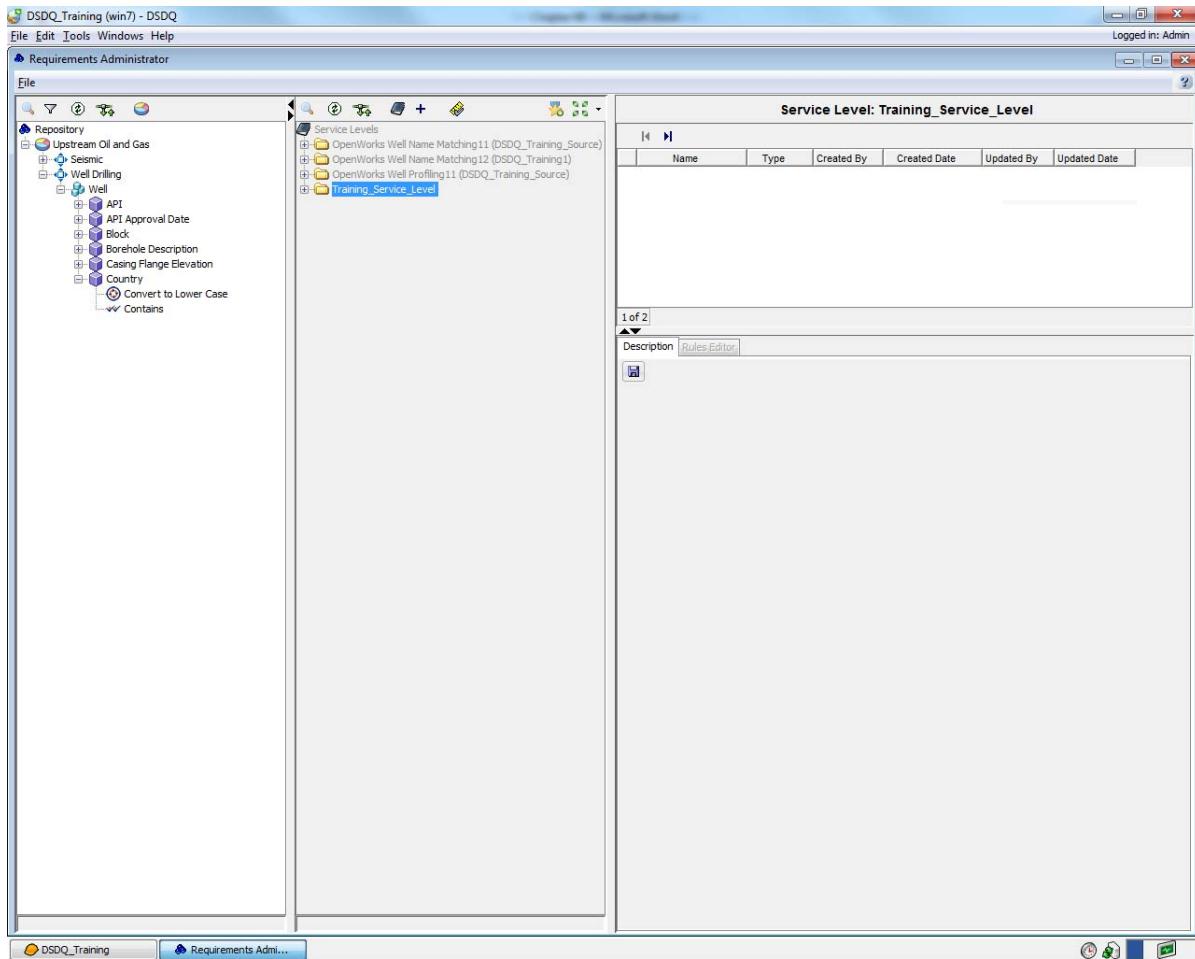


The **Confirm Item Delete** dialog box appears.



6. Click Yes.

The selected requirement is deleted.

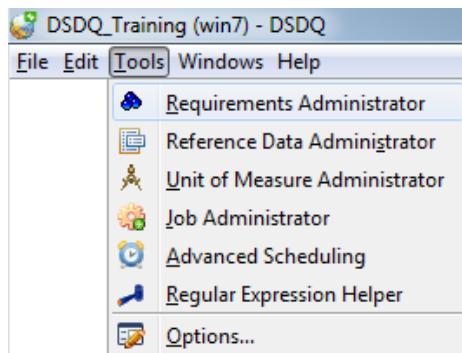


7. Select **File > Exit** from the menu bar on the **Requirements Administrator** window.

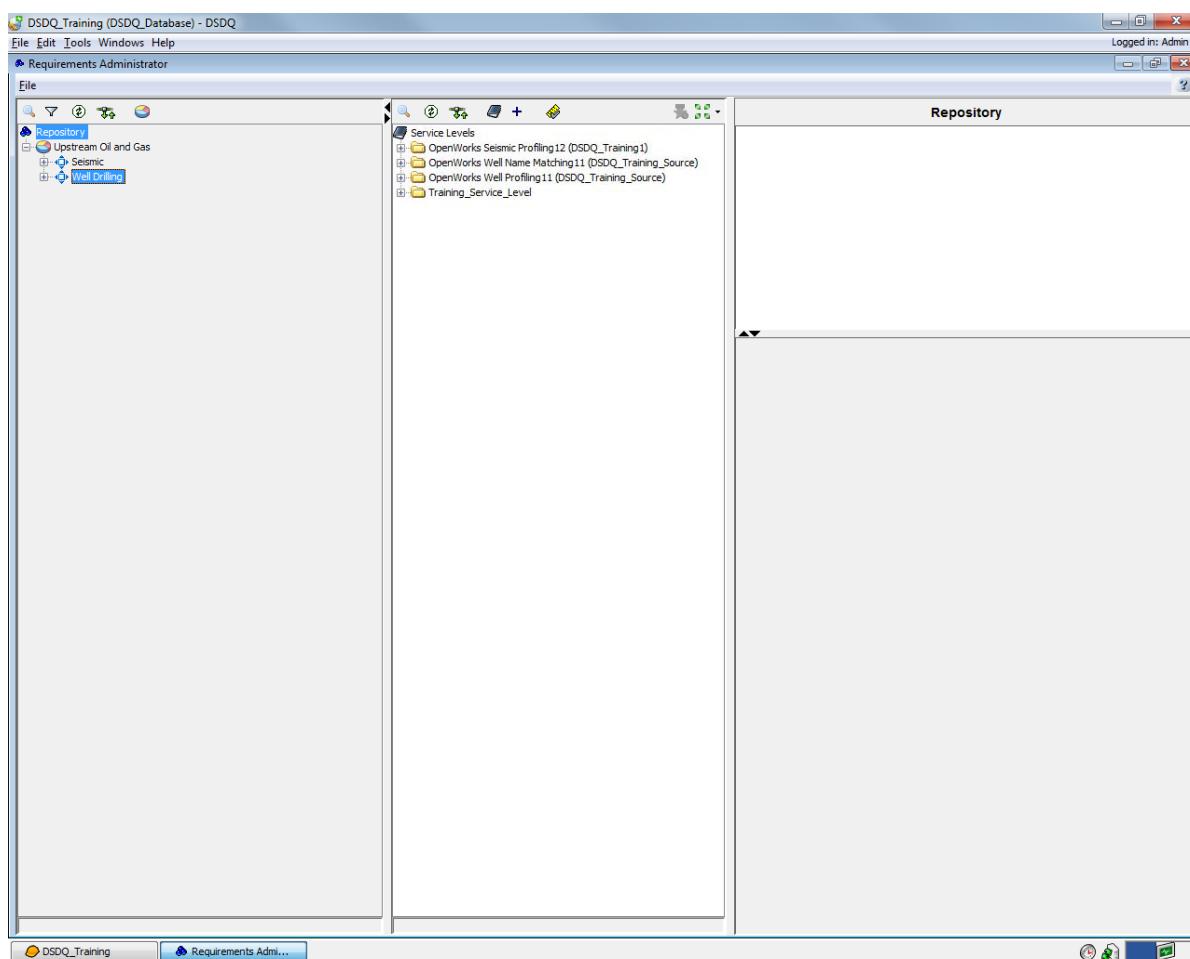
Exercise: Validating a Requirement

To validate a Requirement:

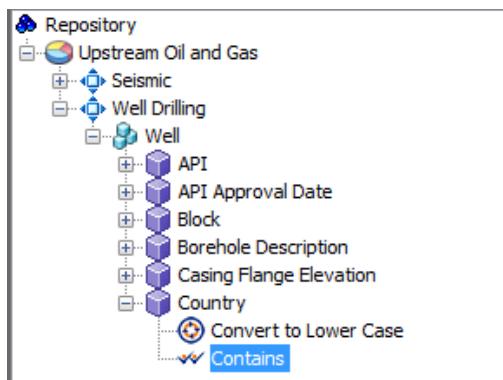
1. Select **Tools > Requirements Administrator** from the menu bar on the **DSDQ Project** window.



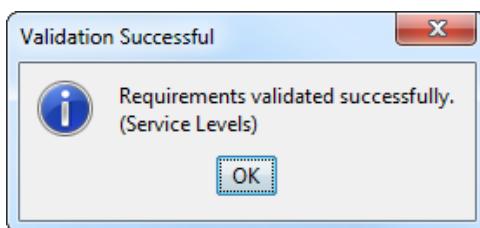
The **Requirements Administrator** window appears.



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.

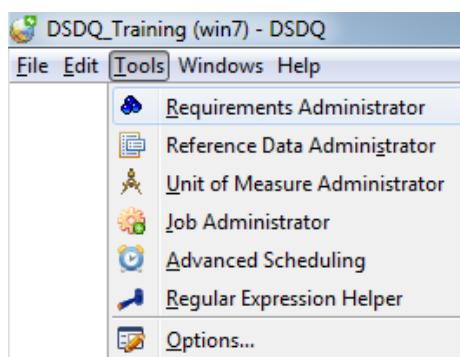
Note

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

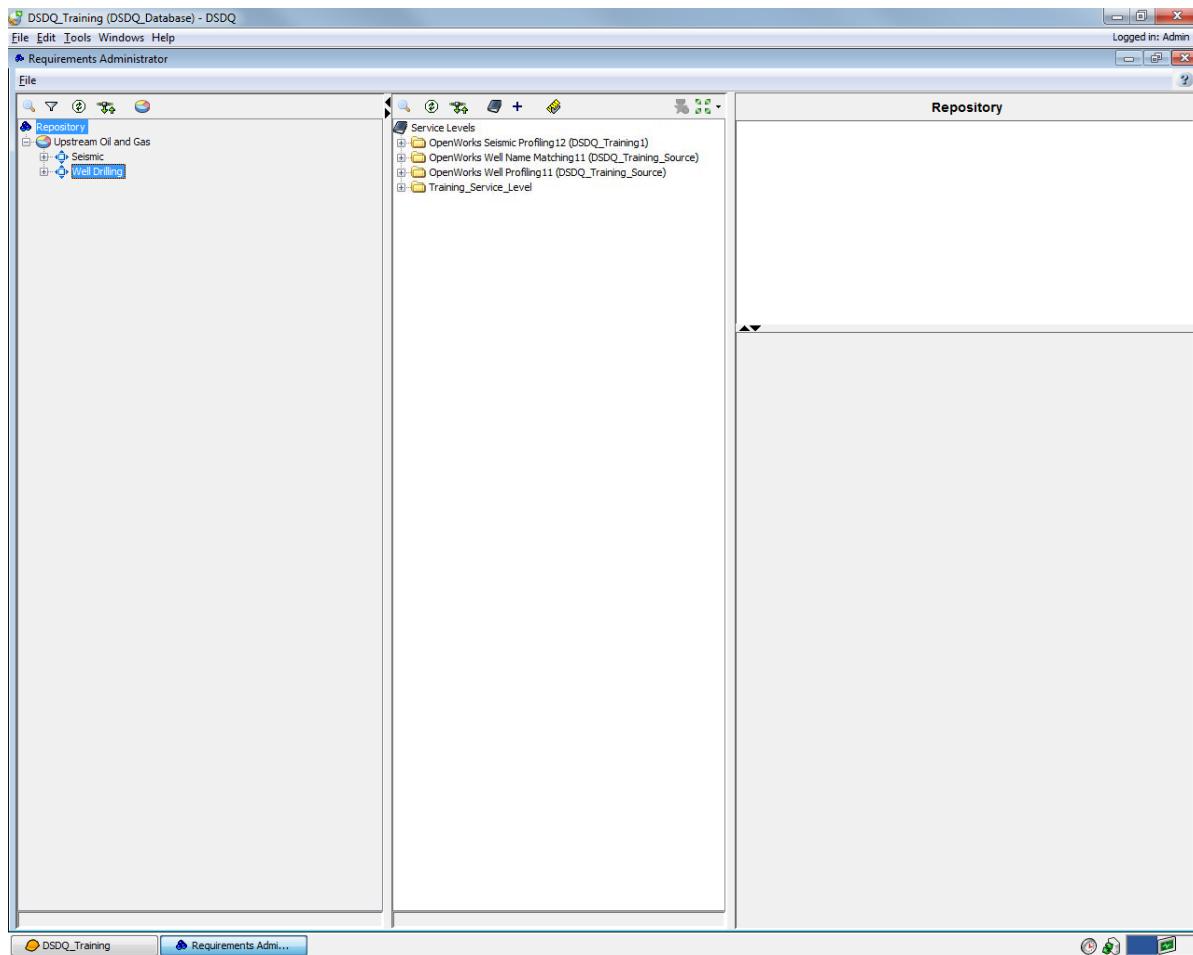
Exercise: Managing Advanced Requirements

To manage advanced requirements:

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

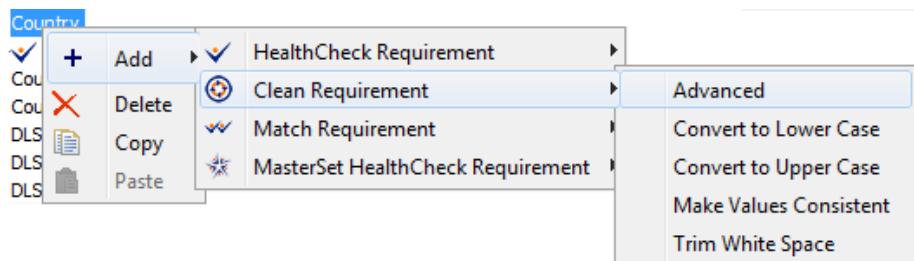


The Requirements Administrator window appears.

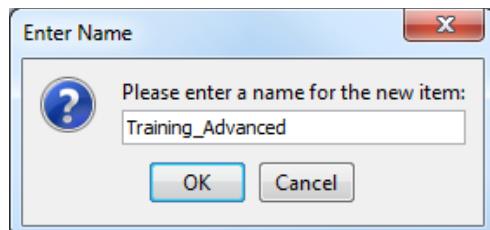


2. Click to expand the **Upstream Oil and Gas** sector in the Repository Pane.
3. Click to expand the **Well Drilling** area in the Repository Tree Pane.
4. Click to expand the **Well** element group in the Repository Tree Pane.

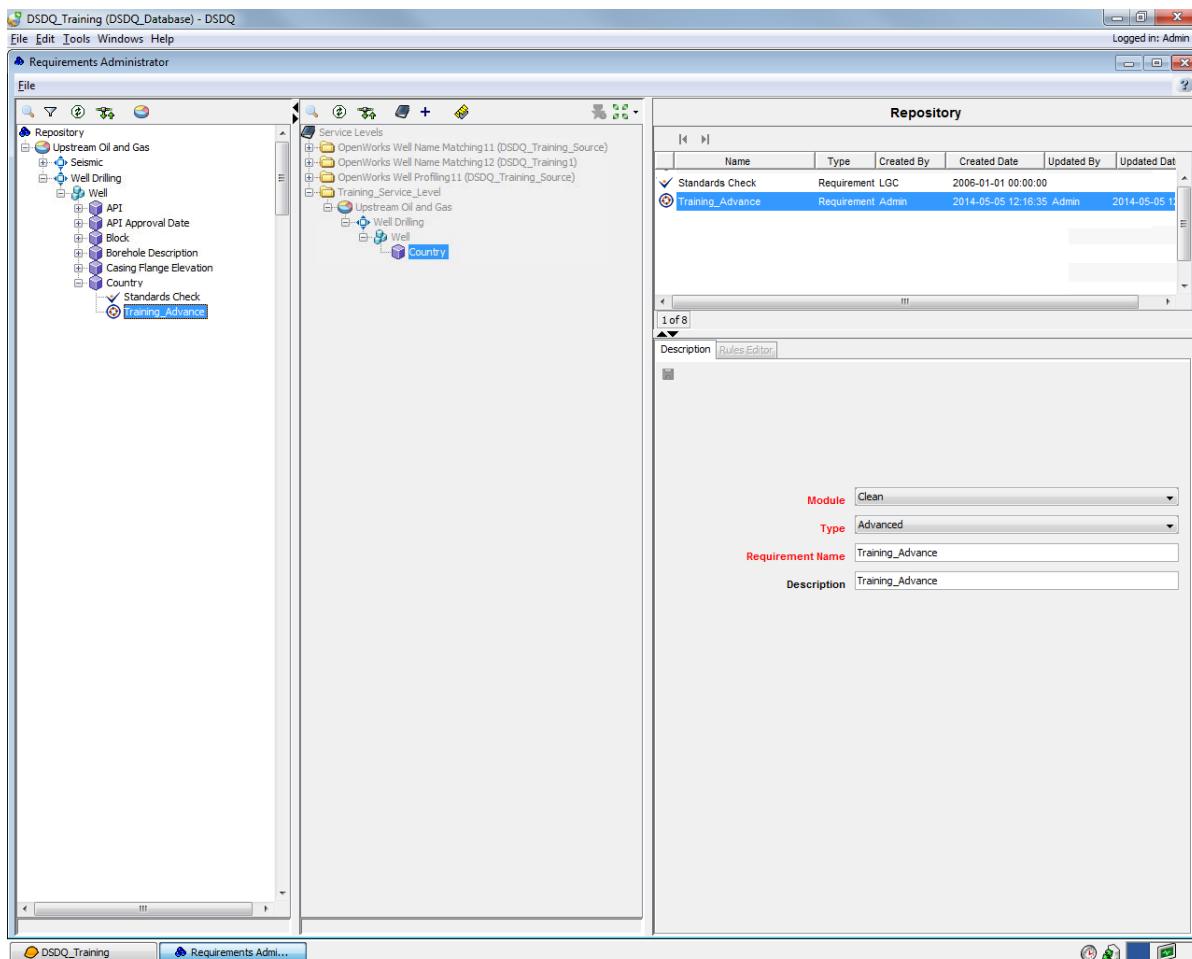
5. Right-click on the **Country** Element and select **Add > Clean Requirements > Advanced** from the pop-up menu.



The **Enter Name** dialog box appears.

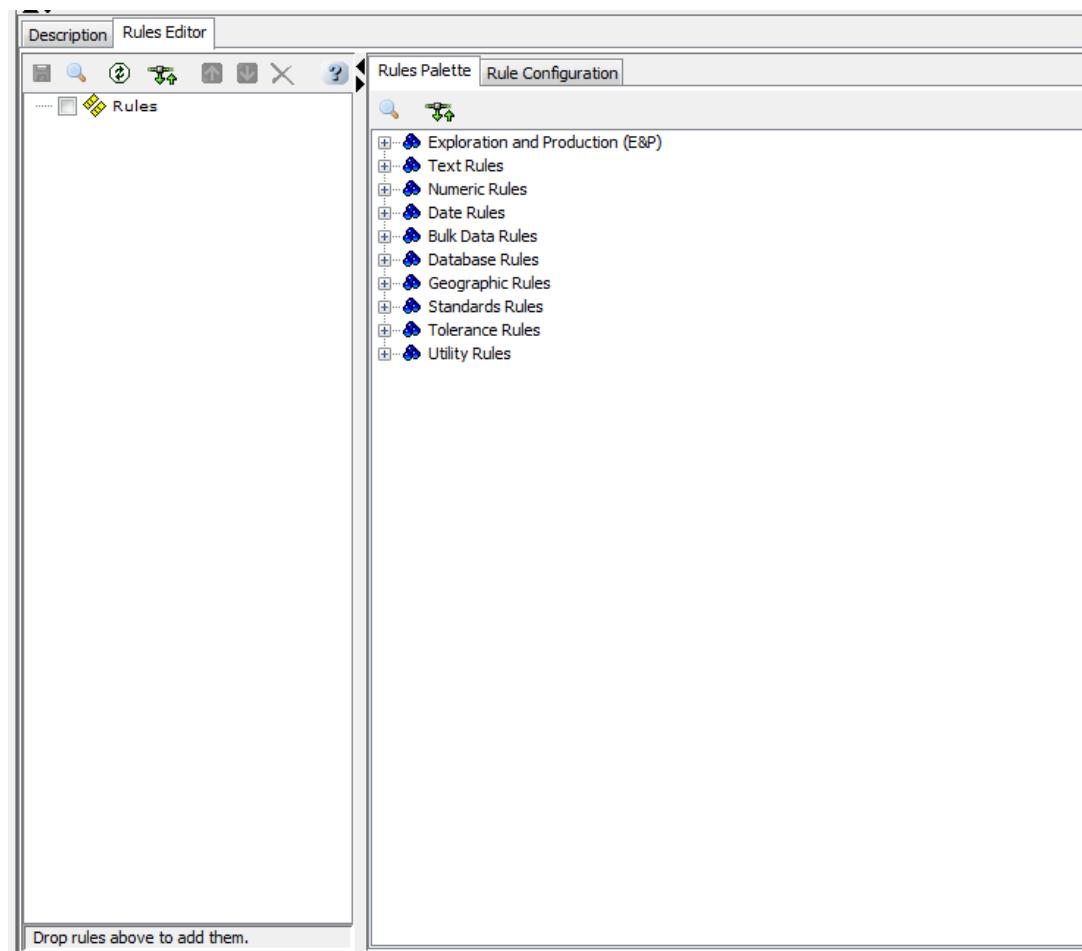


6. Optionally, enter a name and click **OK**.

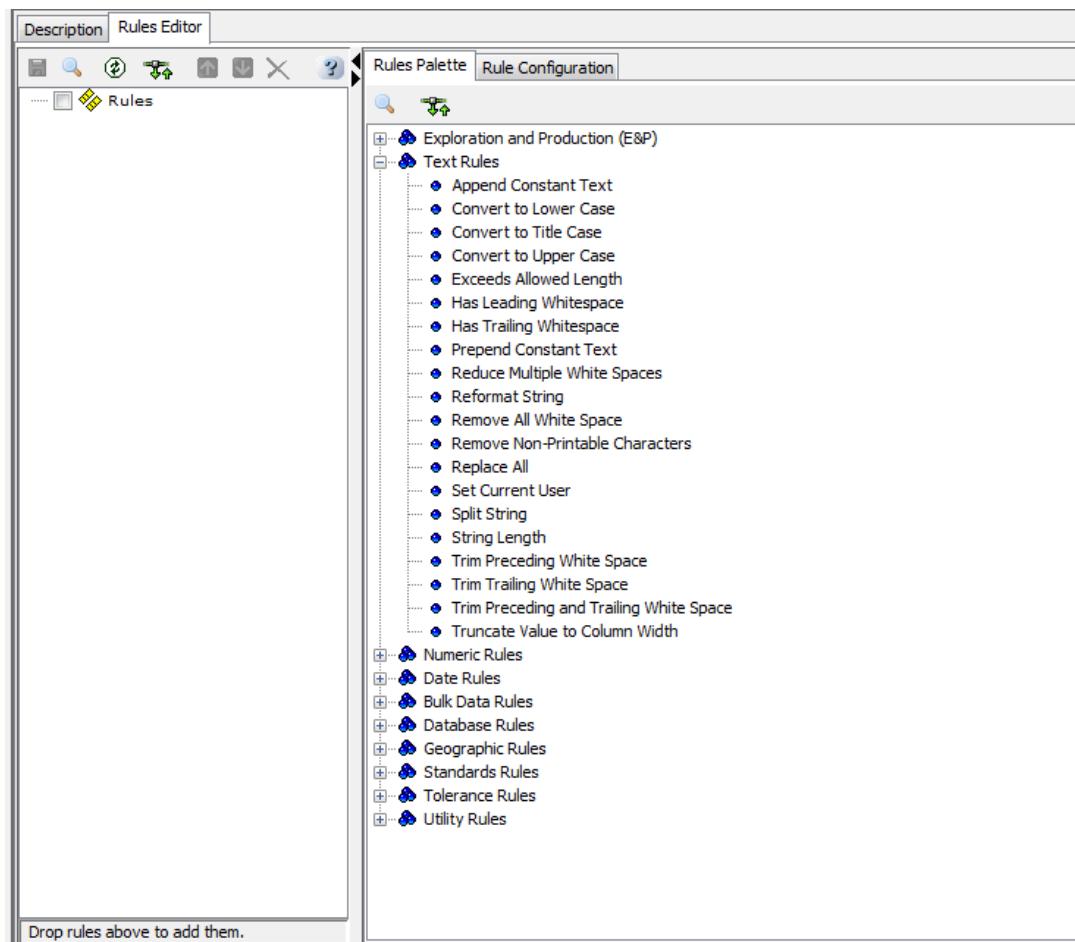


7. Select **Clean** from the **Module** drop-down list.
8. Select **Advanced** from the **Type** drop-down list.
9. Enter **Training_Advanced** in the **Requirement Name** field.
10. Enter **Training_Advanced** in the **Description** field.

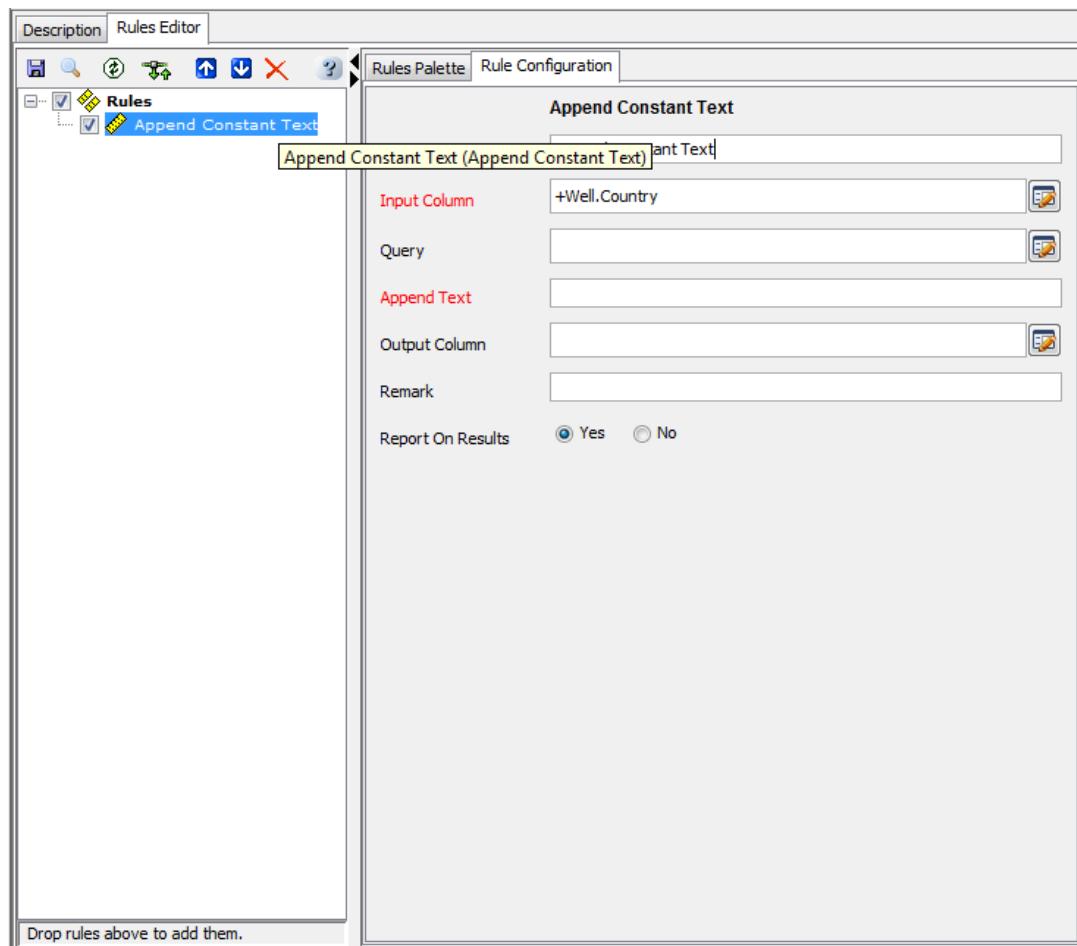
11. Select the **Rules Editor** tab.



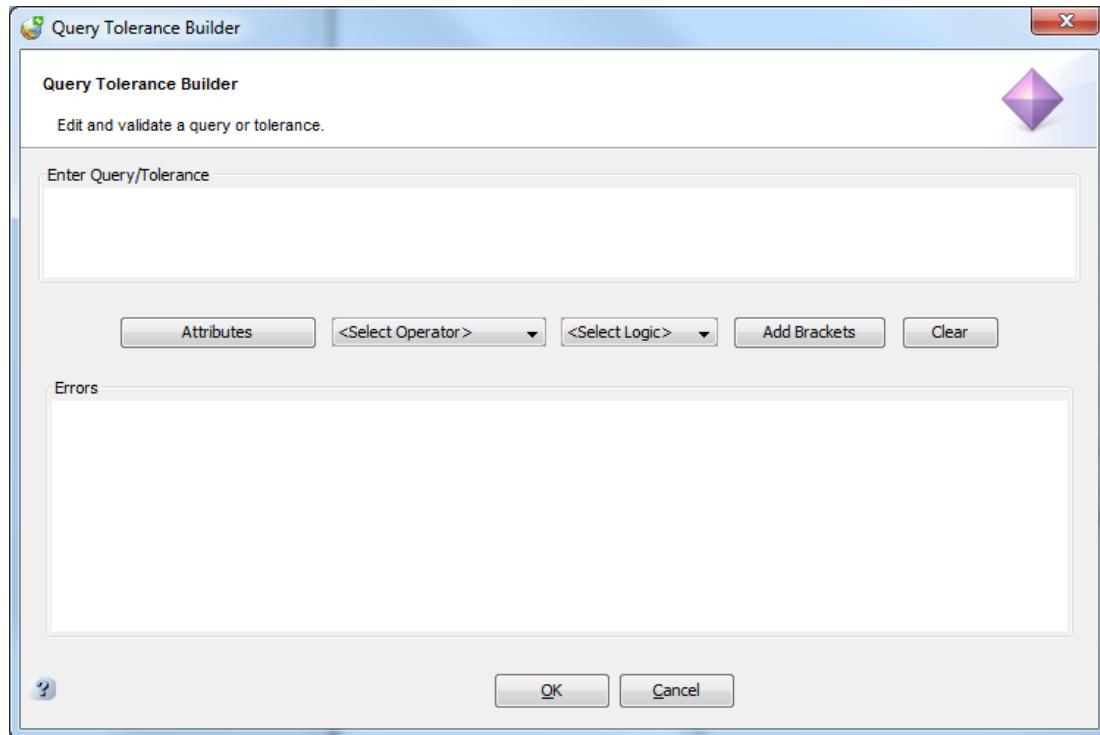
12. Click to expand **Text Rules** in the **Rules Palette** window.



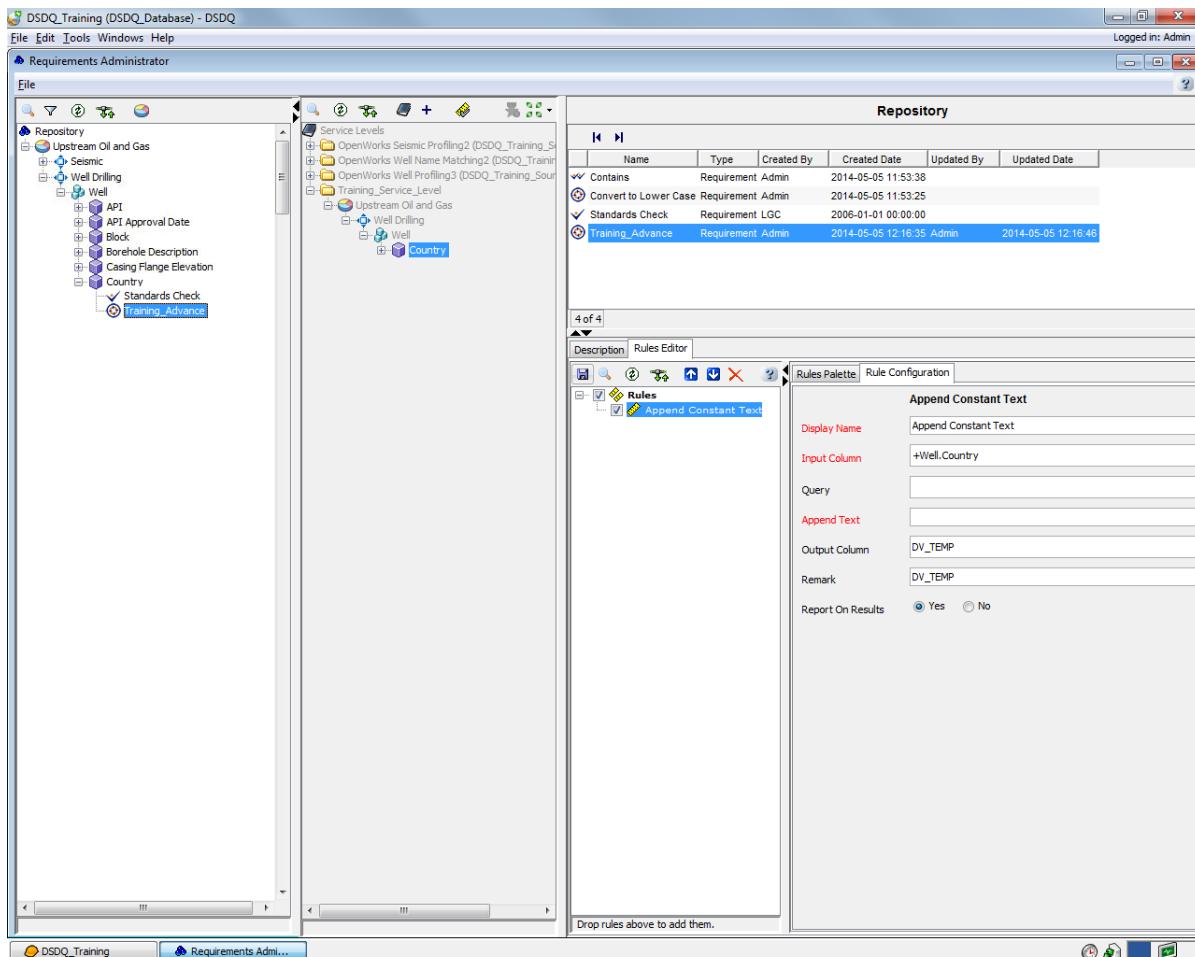
13. Drag **Convert to Lower Case** Rule from the **Rules Palette** to the **Rules Pane**.
The **Rule Configuration** tab opens automatically.



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.

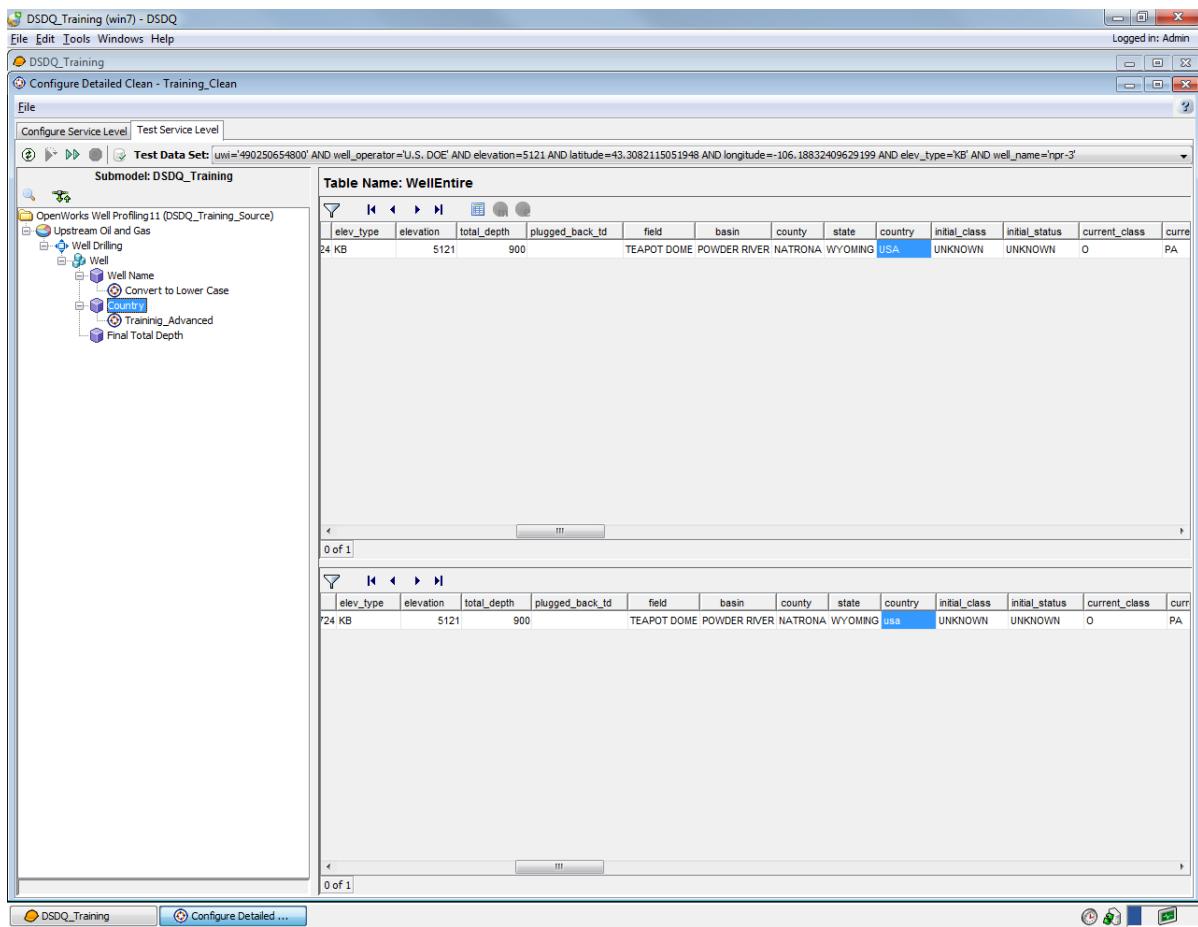
The advanced rule is created.

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.

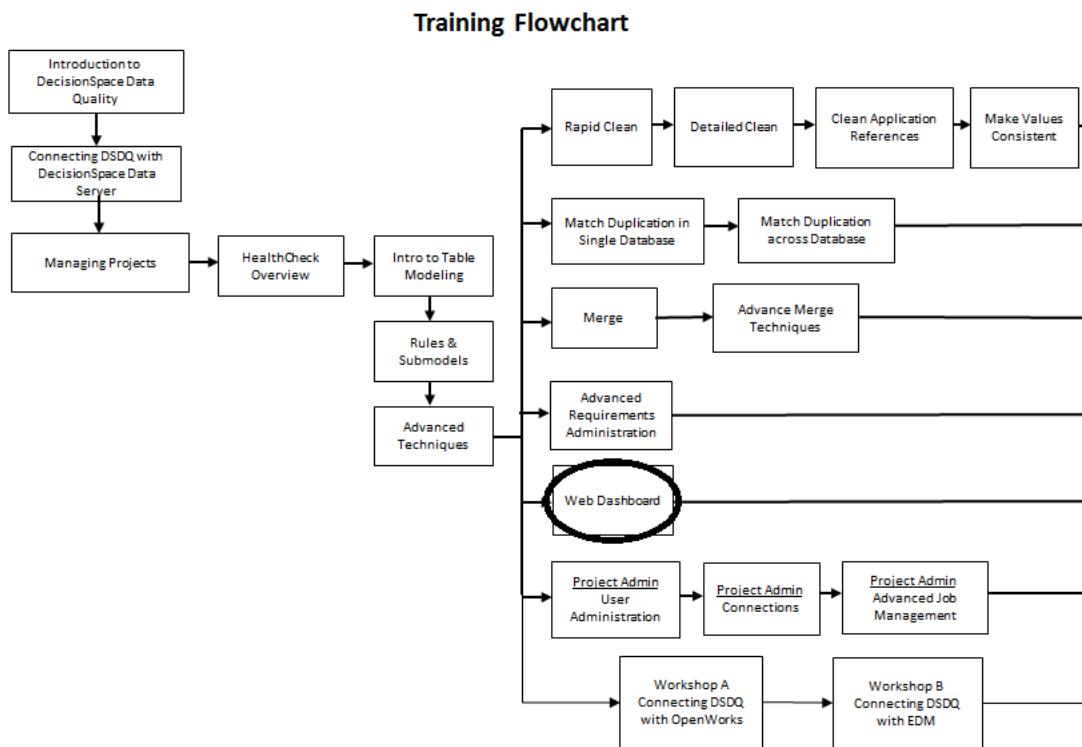


Chapter Overview

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

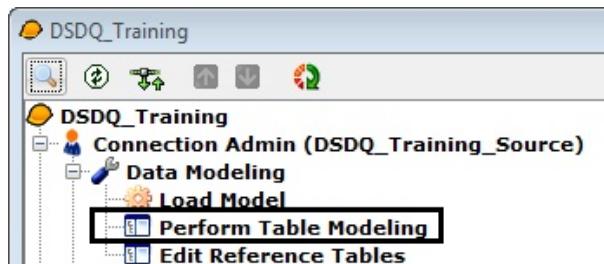
A submodel is a core data structure within the Data Quality application that is used by all Activities. It is created using the Perform Table Modeling task, a submodel is a grouping of data tables within a database and is integral to defining data configurations. Submodels can be used to represent one 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

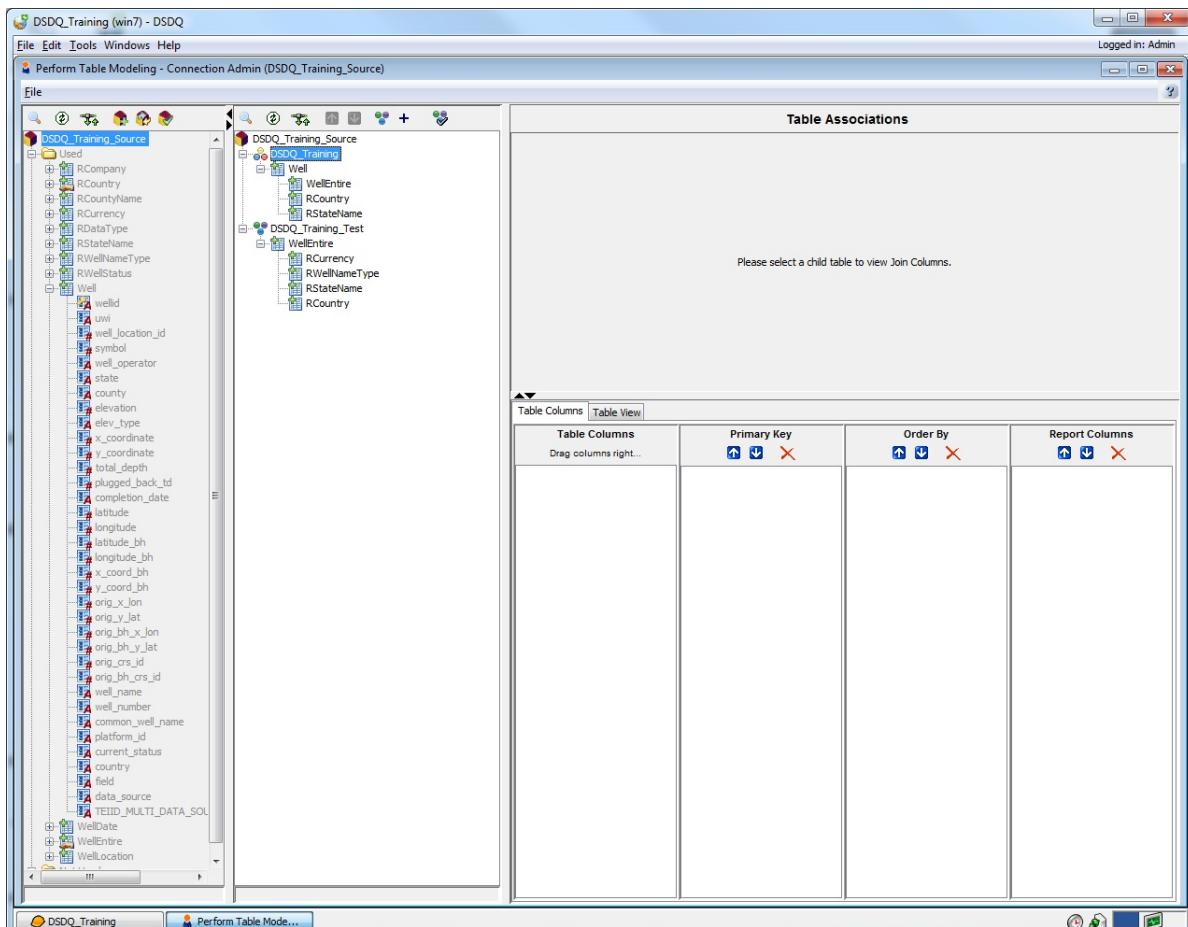
To view information about a DecisionSpace Data Quality submodel, you must publish it to the Web Dashboard. To publish a submodel to the Web Dashboard:

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.

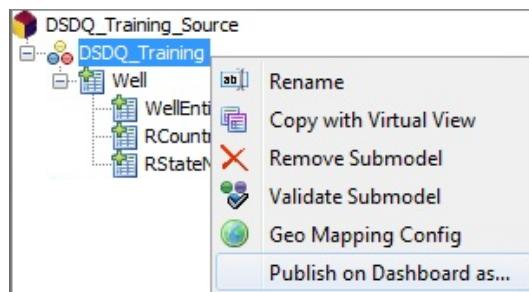


The **Perform Table Modeling - Connection Admin** window appears.



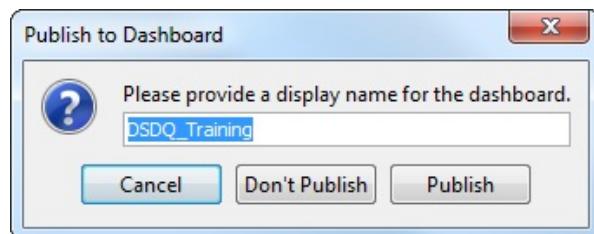
4. From the right tree, right-click **DSDQ_Training** from the Submodel Listing Tree and select **Publish on Dashboard as...**

from the pop-up menu.



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

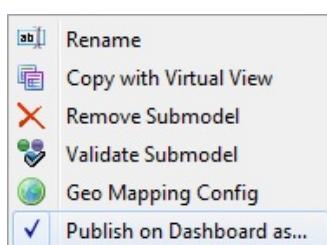
5. Enter **DSDQ_Training** in the **Publish to Dashboard** dialog box and click **Publish**.



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

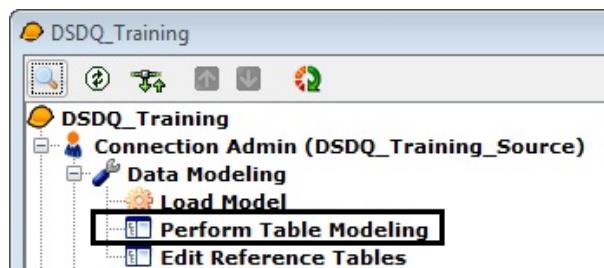
You group master data information into Column Display Groups for the purposes of displaying it on the Web Dashboard or using it as search criteria within the Web Dashboard.

Exercise: Configuring Columns Display Groups

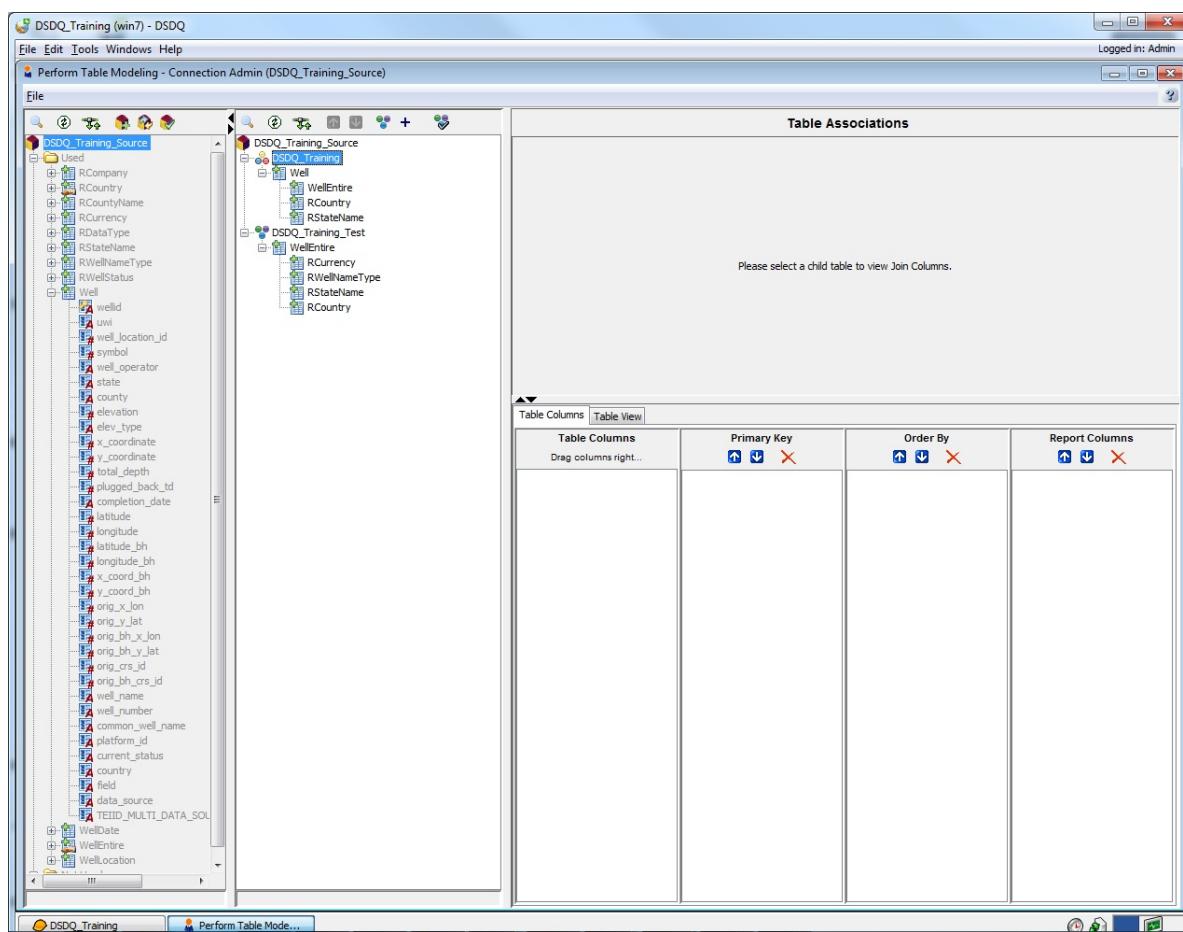
The web dashboard allows users to view master data information through the intranet, grouped by columns display groups.

To configure column display groups:

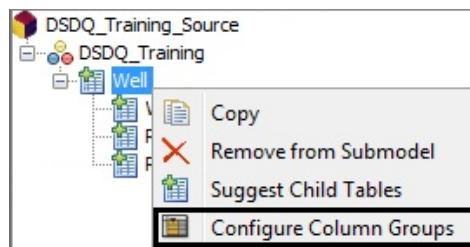
1. Ensure that **Connection Admin (DSDQ_Training_Source)** appears expanded on the DecisionSpace Data Quality Tree.
2. Ensure that the **Data Modeling** Activity appears expanded the DecisionSpace Data Quality Tree.
3. Double-click the **Perform Table Modeling** tool.



The **Perform Table Modeling** window appears.

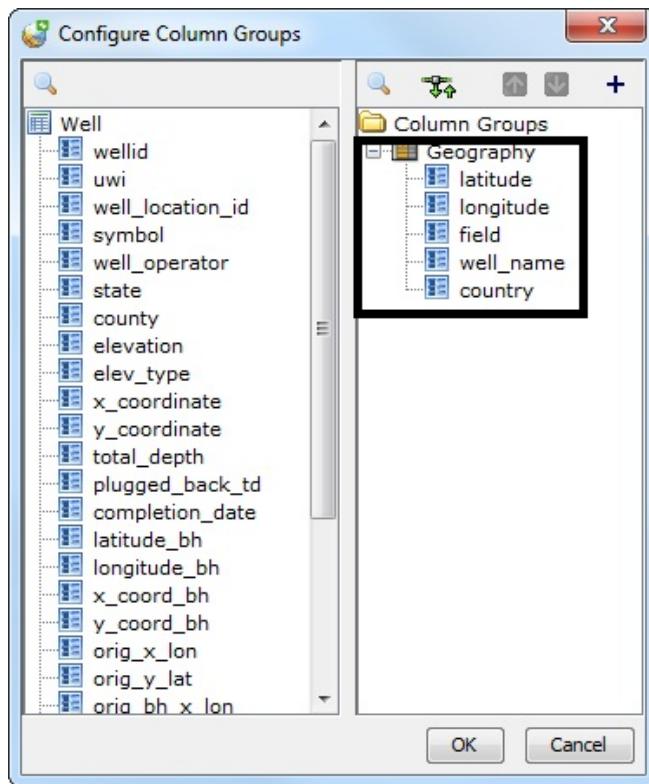


- From the right tree, right-click **Well** from the Submodel Listing Tree and select **Configure Column Groups** from the pop-up menu.



- 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

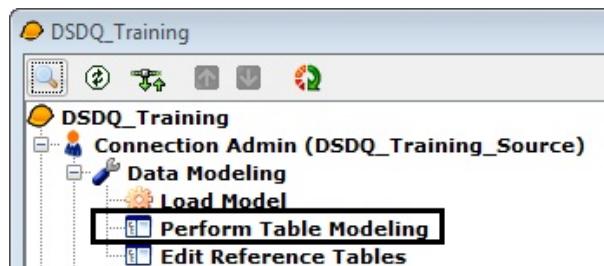
You can only drag and drop one column from the table tree to the **Column Groups** folder at any given time. Repeat step 6 as required for each column.

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.

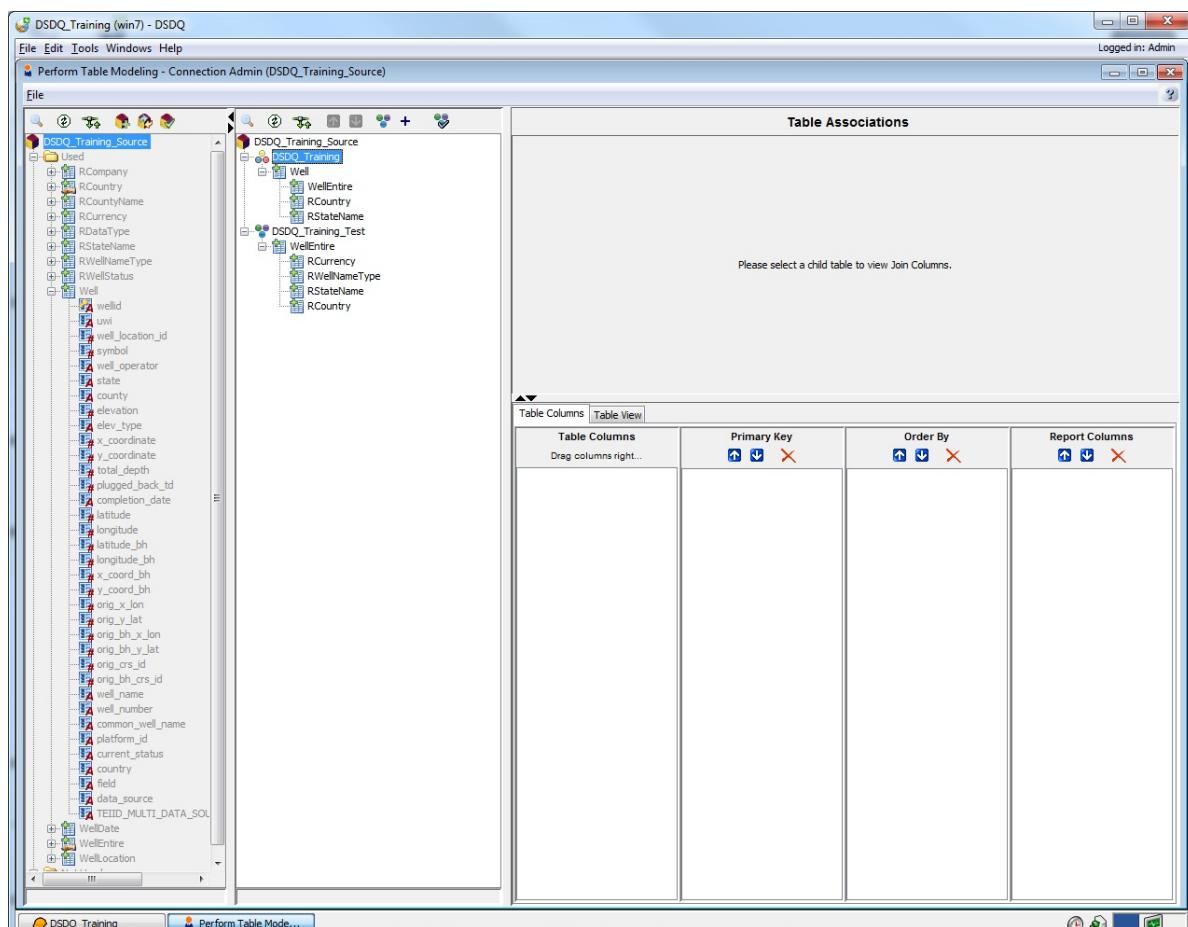
Exercise: Specifying a Column Dashboard Name

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



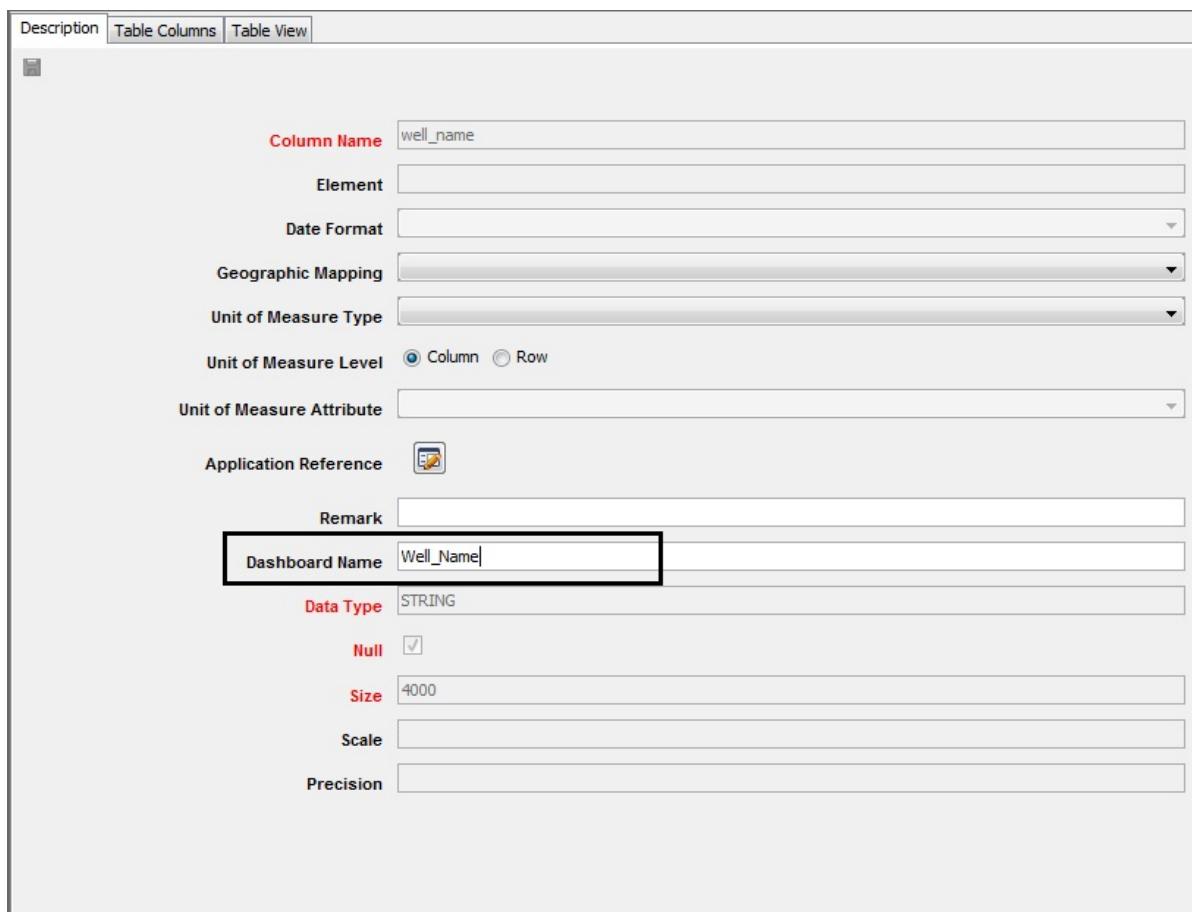
The **Perform Table Modeling** window appears.



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 interface includes tabs for 'Description', 'Table Columns', and 'Table View'. The 'Description' tab contains various configuration fields:

- Column Name:** well_name
- Element:** (empty)
- Date Format:** (empty)
- Geographic Mapping:** (empty)
- Unit of Measure Type:** (empty)
- Unit of Measure Level:** Column Row
- Unit of Measure Attribute:** (empty)
- Application Reference:** 
- Remark:** (empty)
- Dashboard Name:** Well_Name (highlighted with a black border)
- Data Type:** STRING
- Null:**
- Size:** 4000
- Scale:** (empty)
- Precision:** (empty)

7. Click  to save the specified column dashboard name.

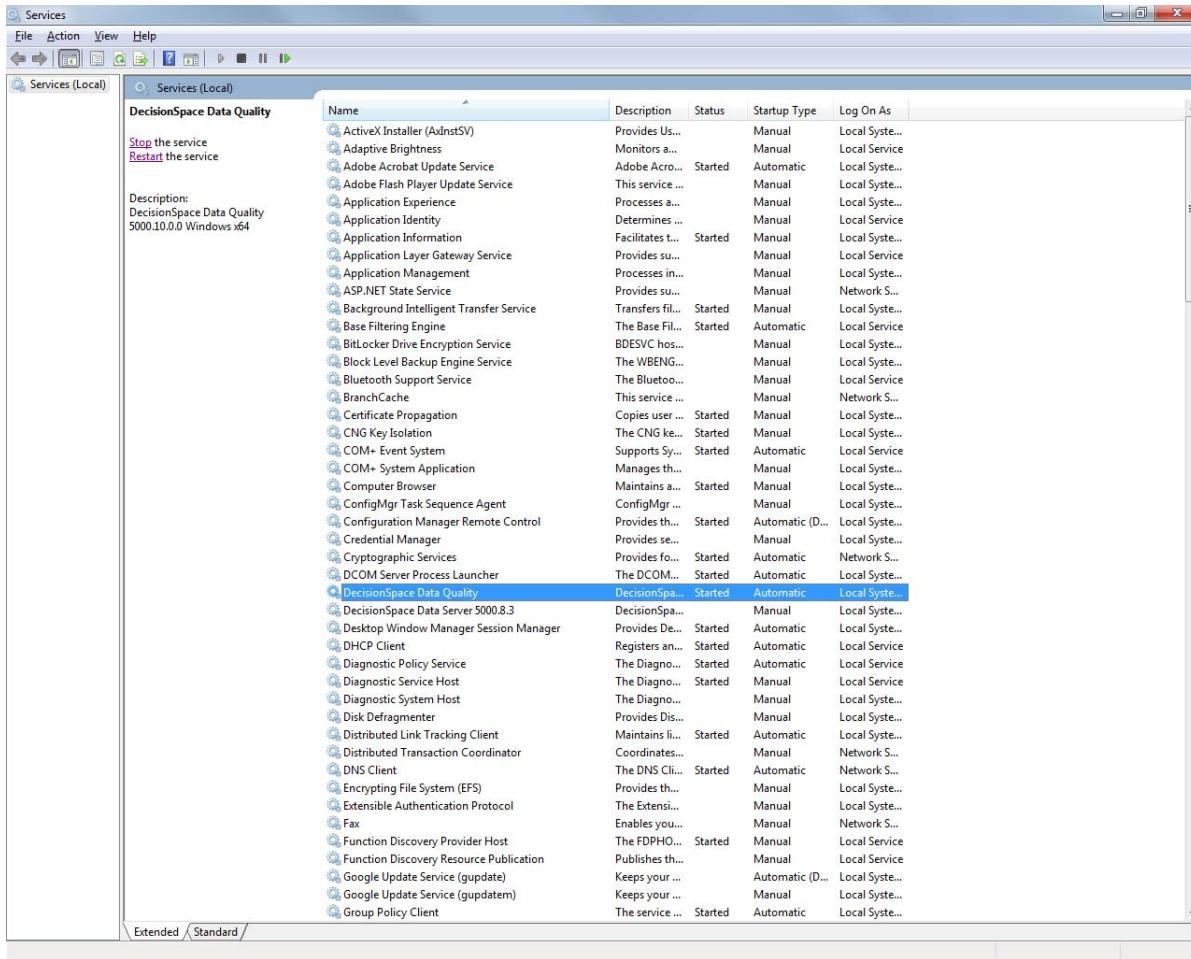
8. Select **File > Exit** from the menu bar to close the **Perform Table Modeling** window.

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

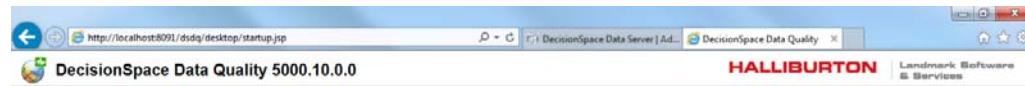
In earlier exercises of this chapter, you published a submodel, configured column display groups and specified a dashboard name for a column. This section includes information about viewing these changes on the Web Dashboard. To refresh data on the Web Dashboard:

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.

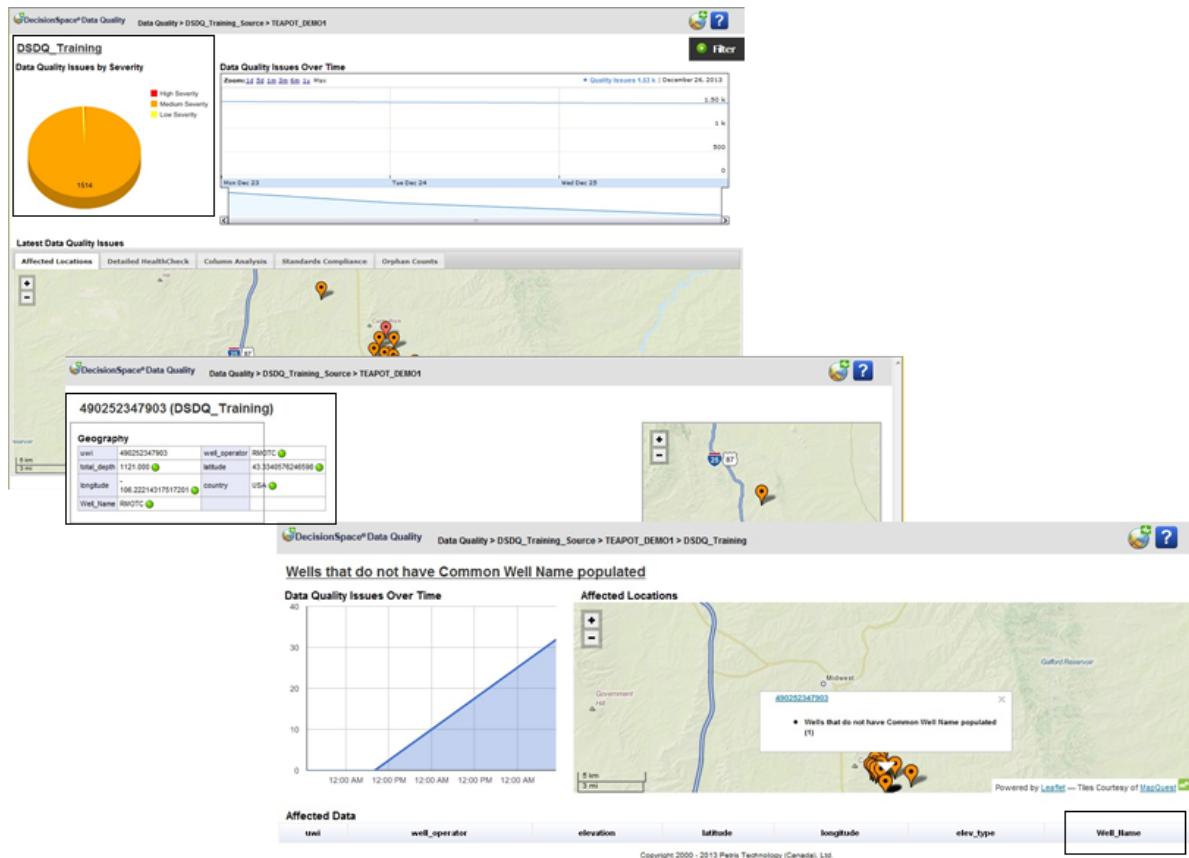


The **Web Dashboard** is launched in your web browser.



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.

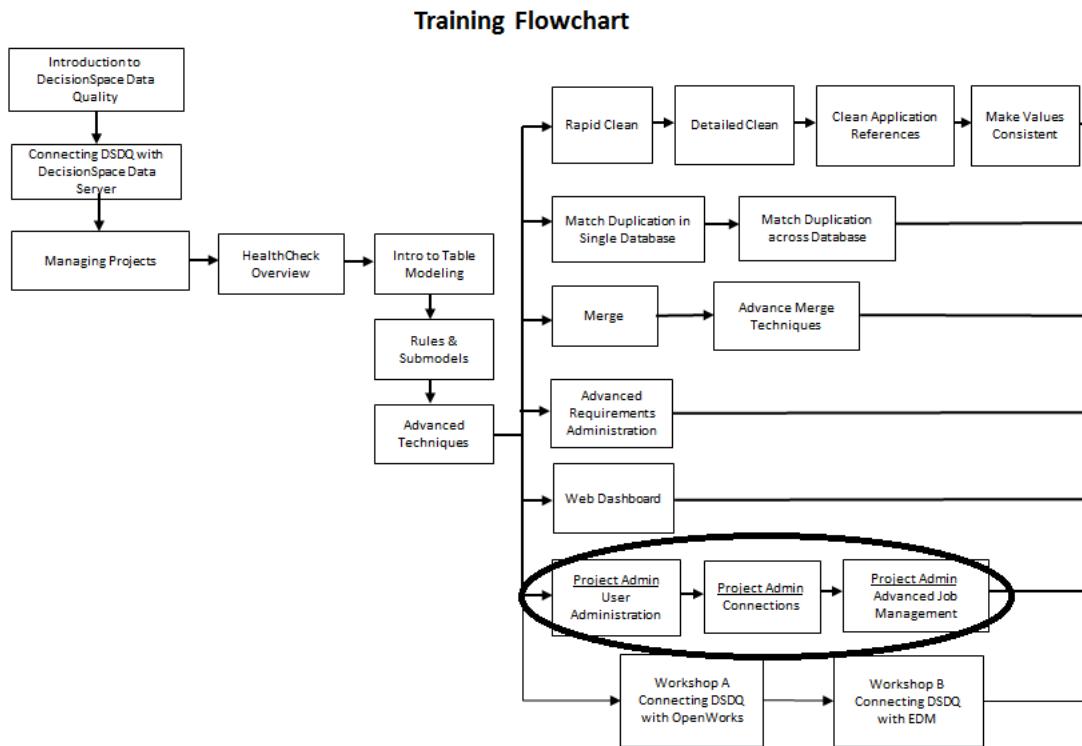
A DSDQ Administrator can also manage data owner and workspace connections, in order to provide access to the data sources.

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

Roles And Permission	Data Custodian	Data Steward
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

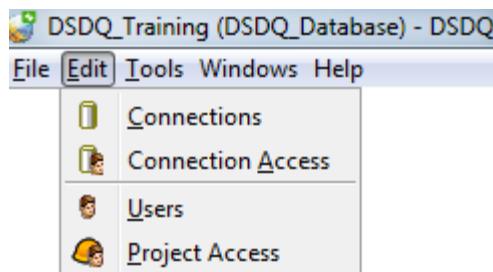
Managing Users

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

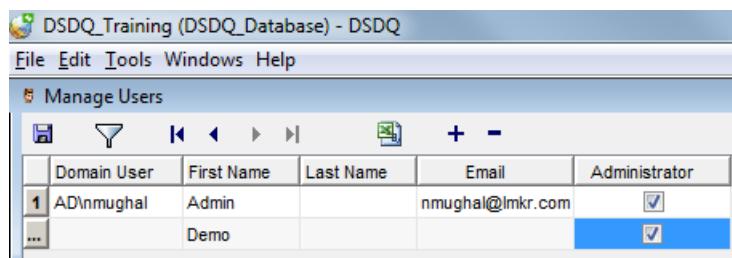
Exercise: Creating a User

To create a user:

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

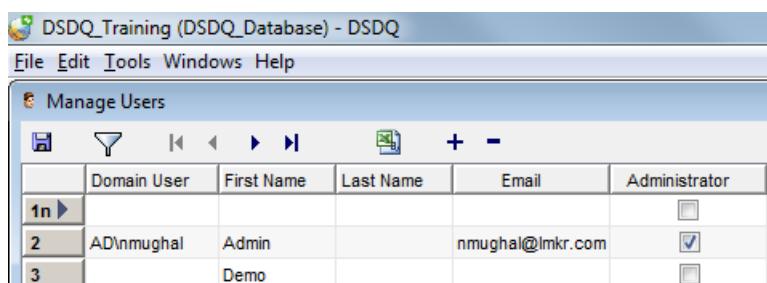


The **Manage Users** window appears.



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.

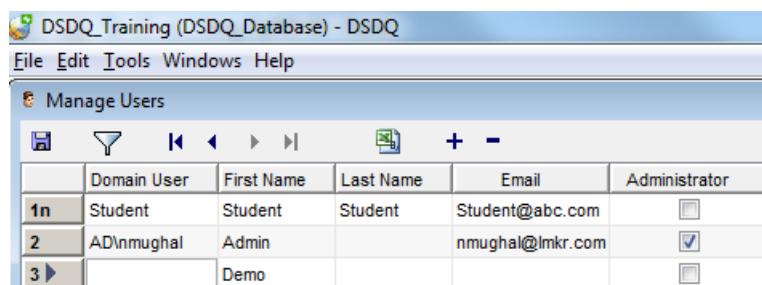
Note

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.

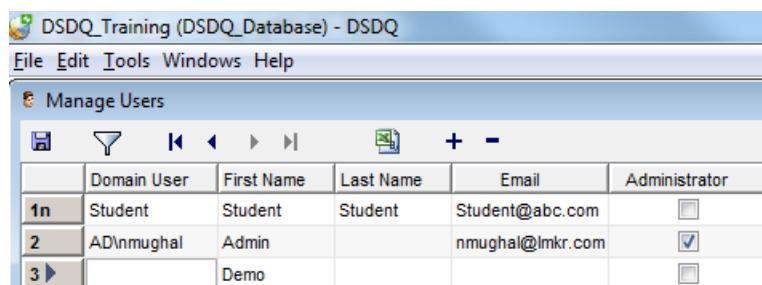
7. Click .



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.

- Click to update the record.

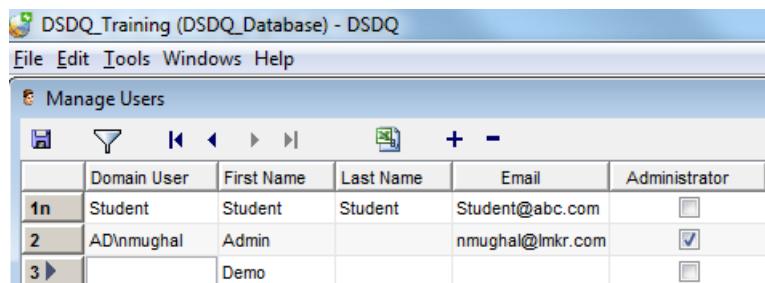
Note

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:

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



- 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.
- Click .
- The selected user is removed.
- 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.

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

Advanced Scheduling

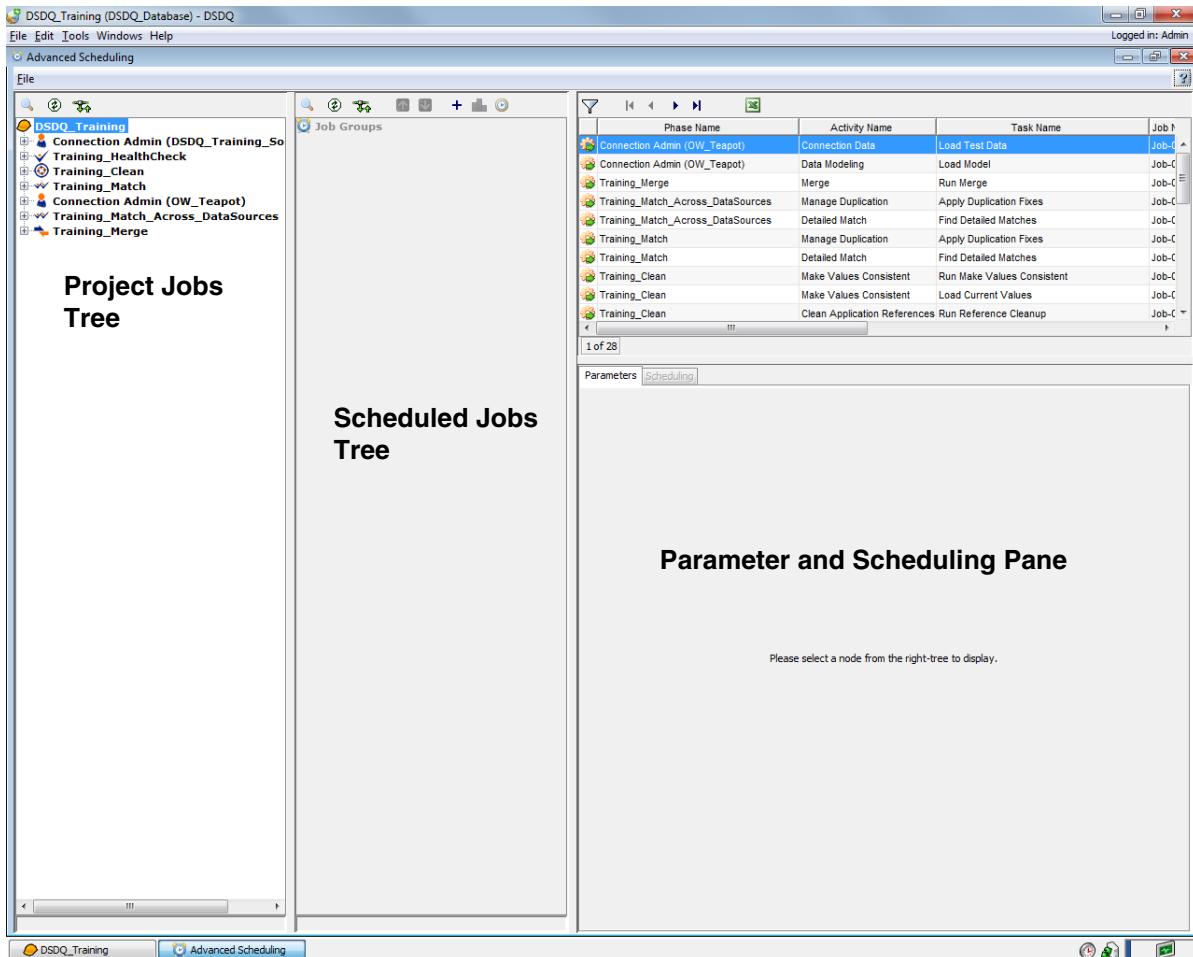
The **Advanced Scheduling** tool is used to set up and configure scheduled events (that consist of current jobs) to run in sequence. You can use this tool to schedule multiple events for multiple jobs.

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.

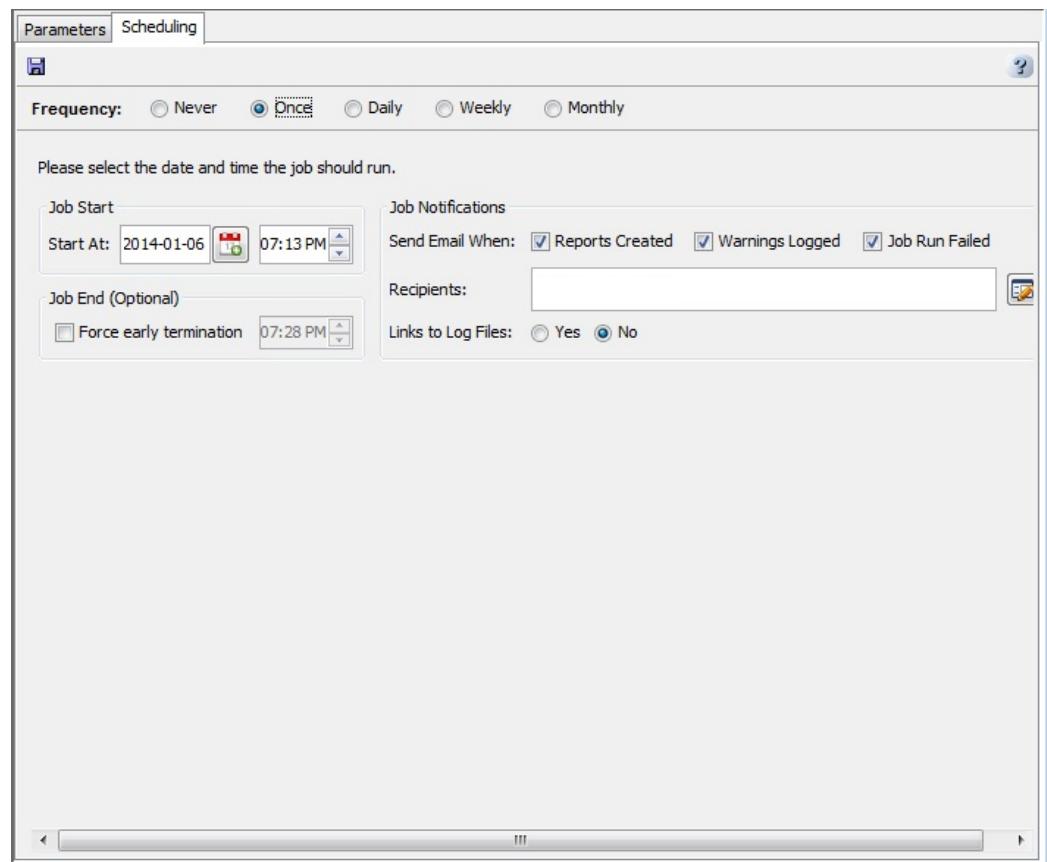
Note

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 .

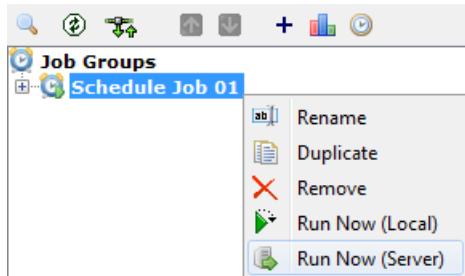
Note

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

Exercise: Running a Job Group on the Server

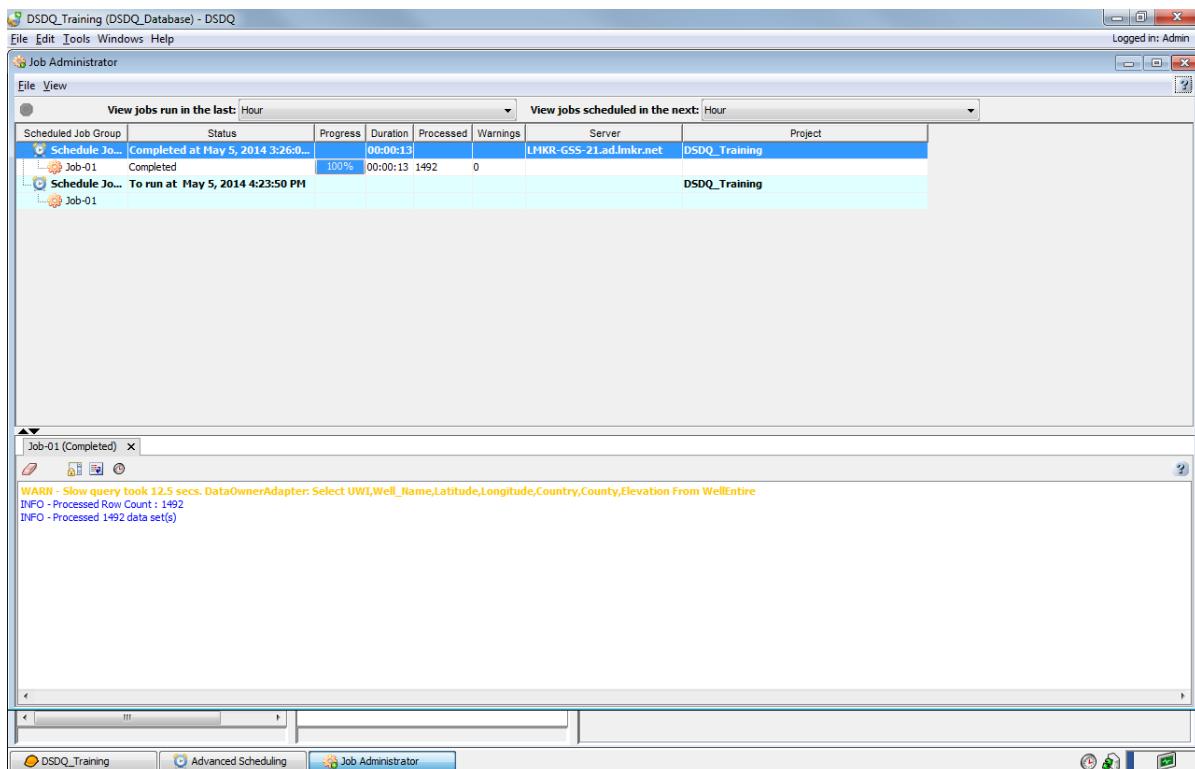
To run a job group on the server:

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.



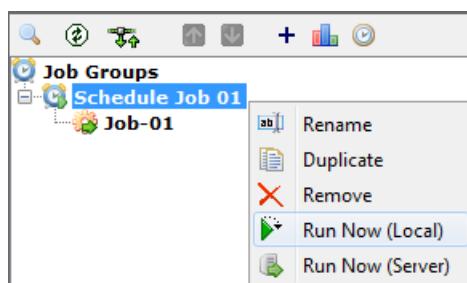
Note

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	LMKR-...	DSDQ_Training
Job-01	To run at Dec 28, 2013 10:00:00 AM						DSDQ_Training
Job-01	To run at Dec 28, 2013 12:27:33 PM						DSDQ_Training
Job-01	To run at Dec 29, 2013 10:00:00 PM						DSDQ_Training
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.

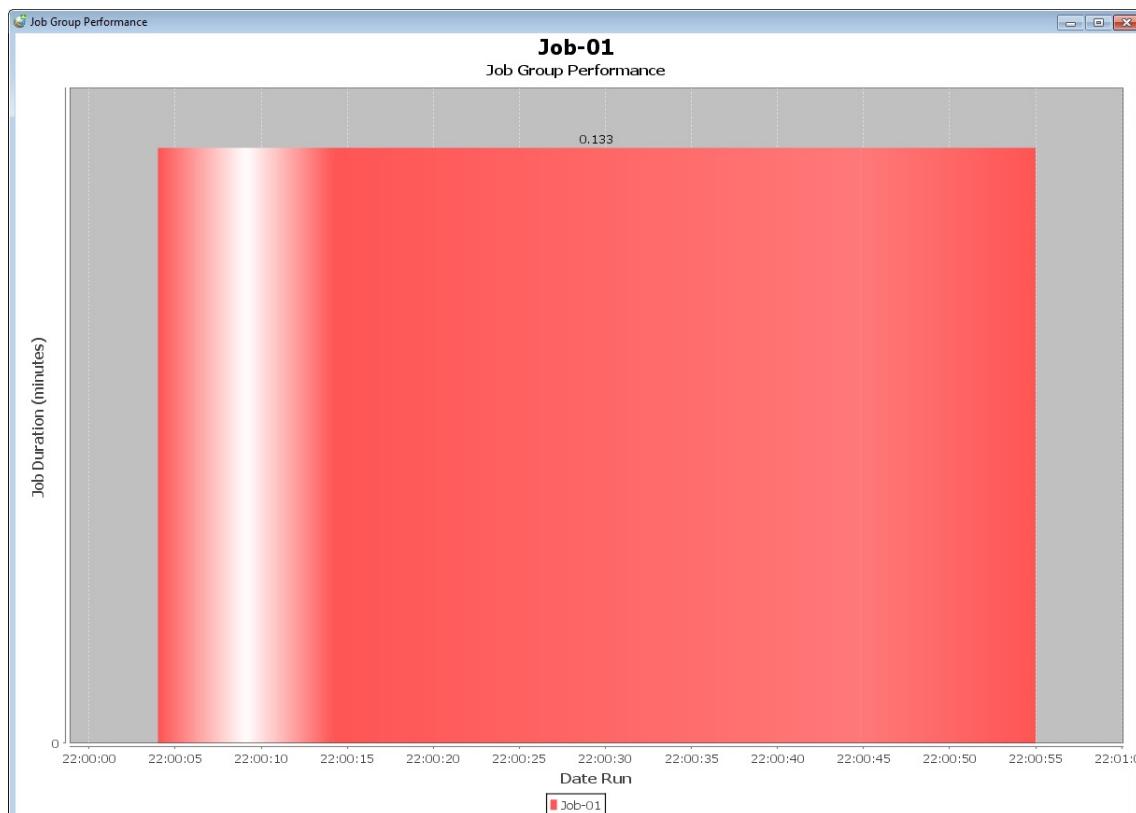
Viewing Job Group Charts

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.



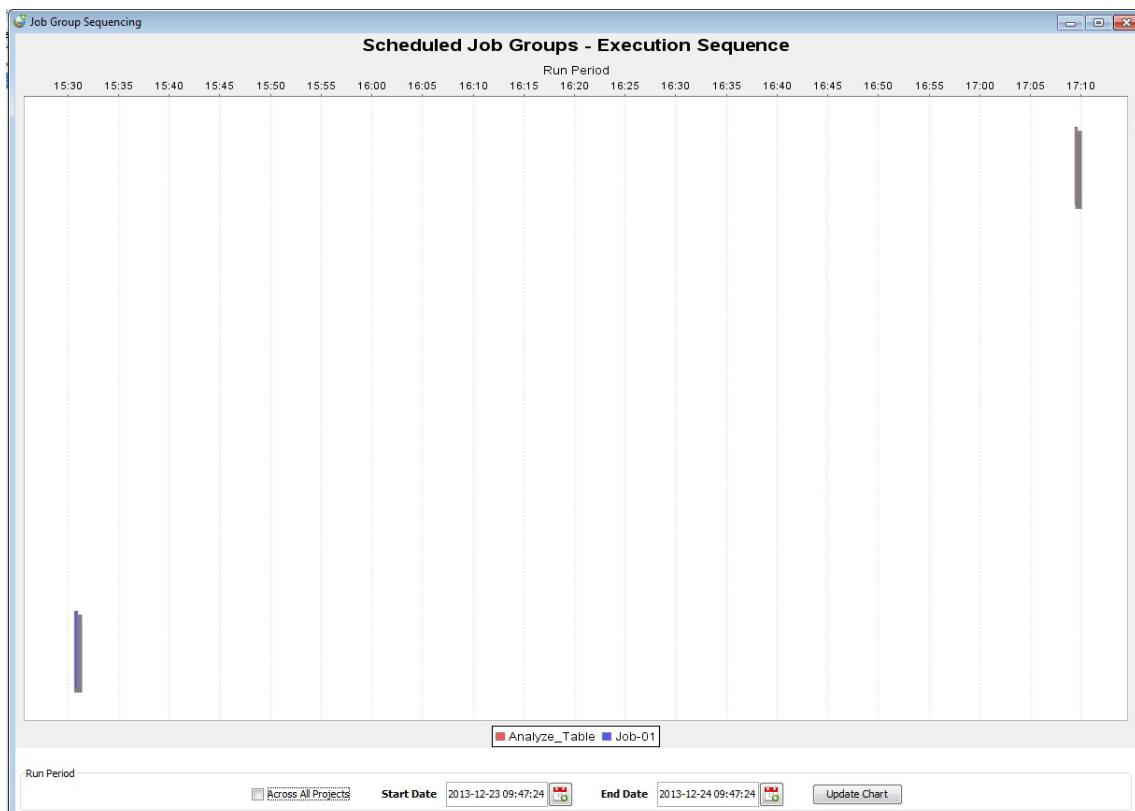
Note

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.

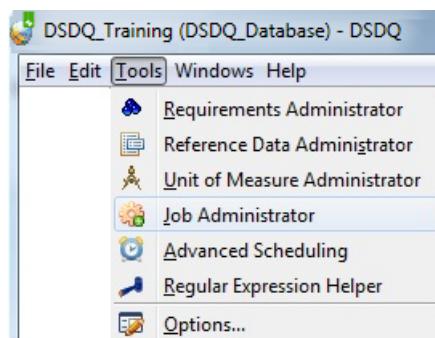
Jobs have the following three states:

- **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.

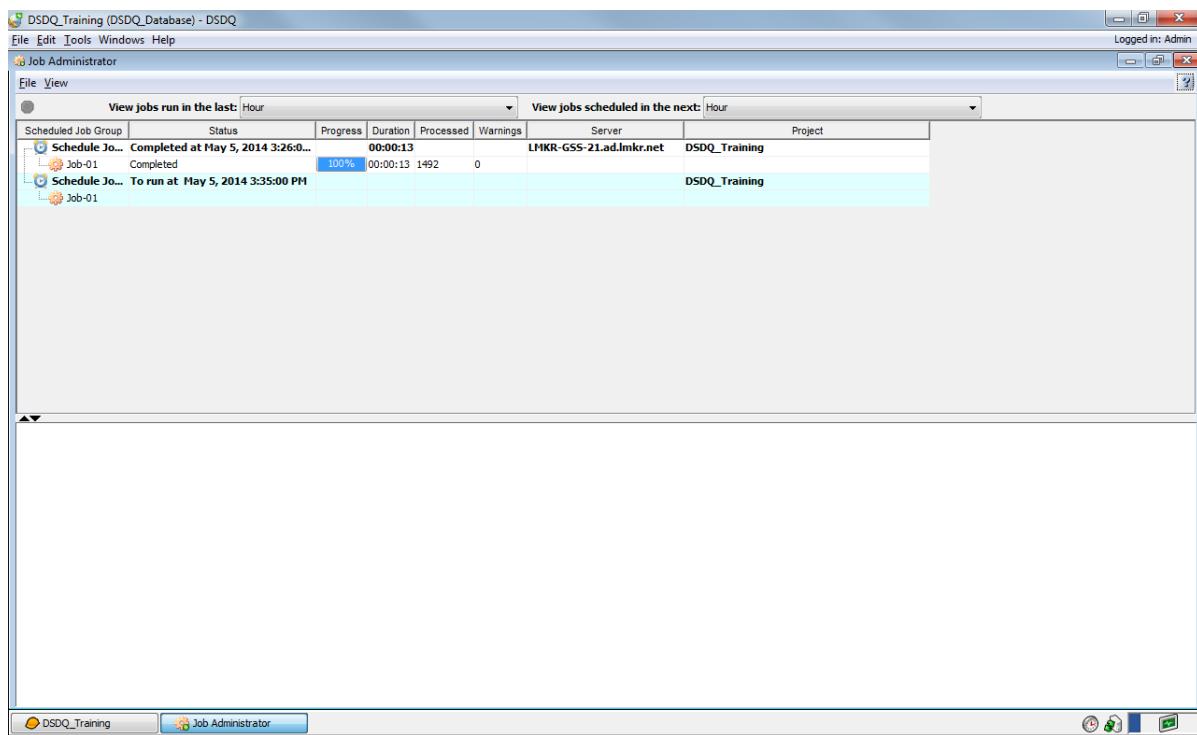
Exercise: Opening Job Administrator

To open Job Administrator:

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



The **Job Administrator** window appears.



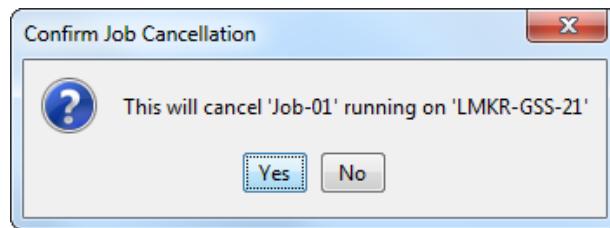
Note

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.

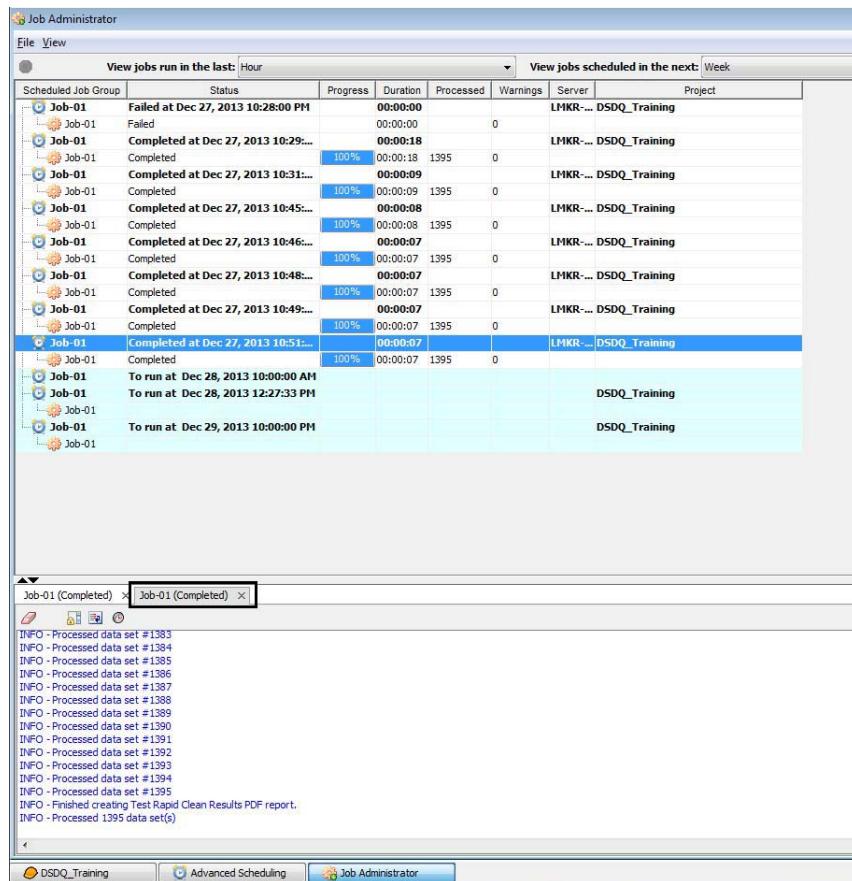
Note

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

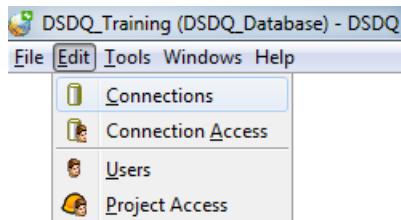
Every connection consists of three Source Dataset elements - Data Source, Workspace and Data Model. The application reads data from the data source. The Workspace is the schema that the application results are written to. The data model is the location to store detailed information about the data source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

Administrative users have rights to manage connections.

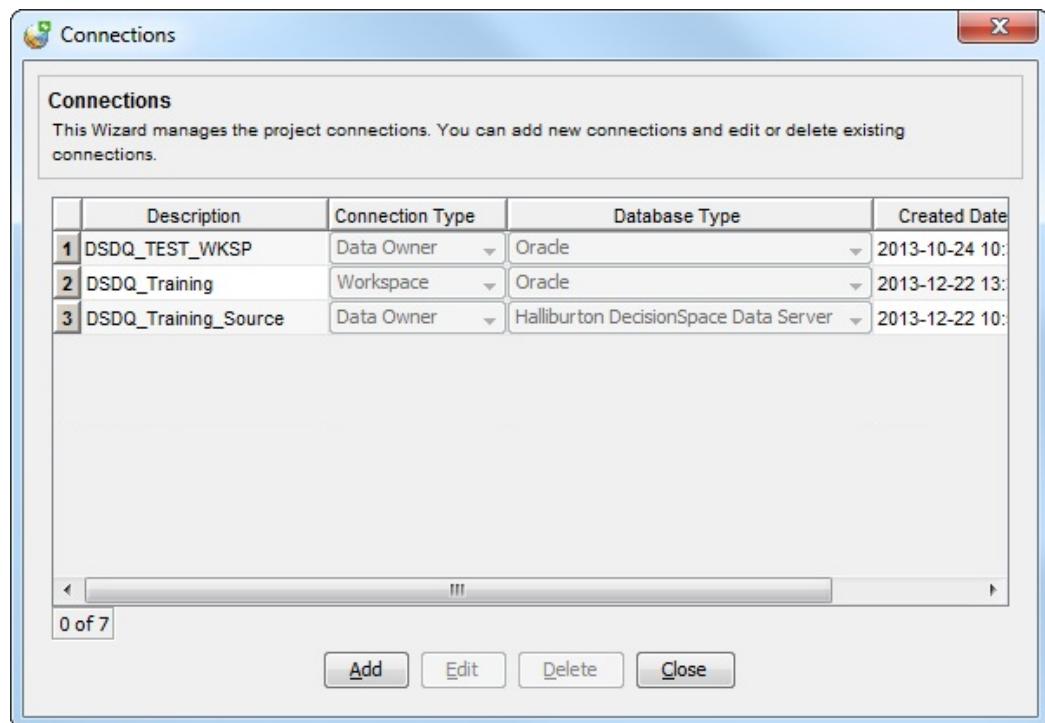
Exercise: Editing a Data Owner Connection

To edit a data owner connection:

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

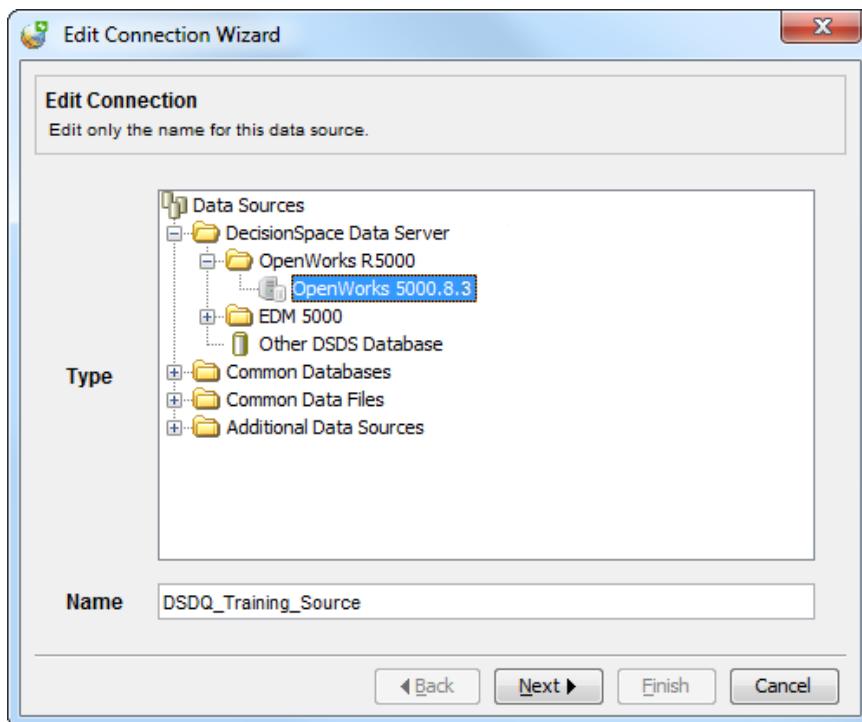


The **Connections** window appears.



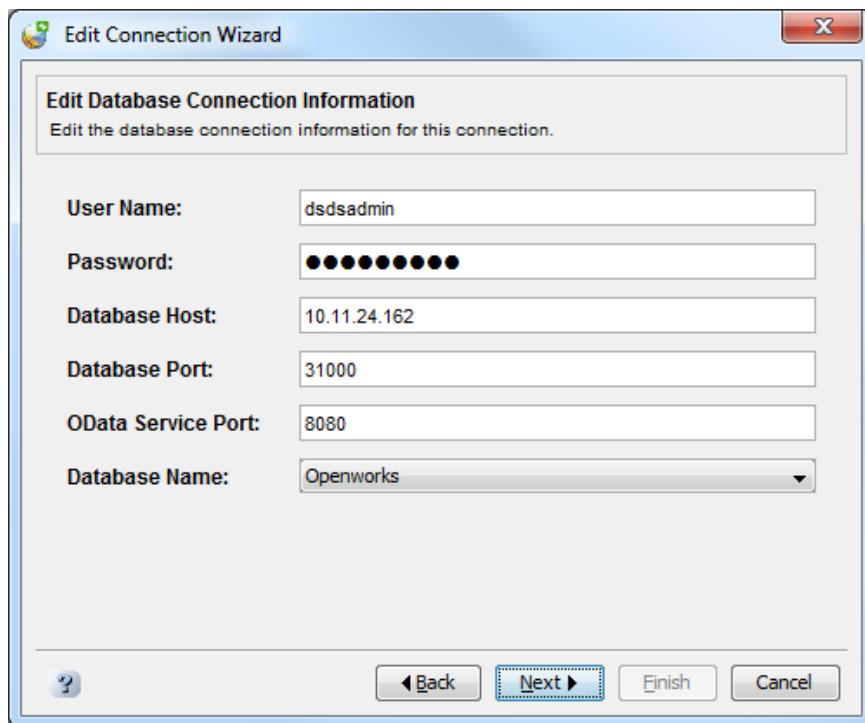
2. Select **DSDQ_Training_Source** and click **Edit**.

The **Edit Connection** window appears.



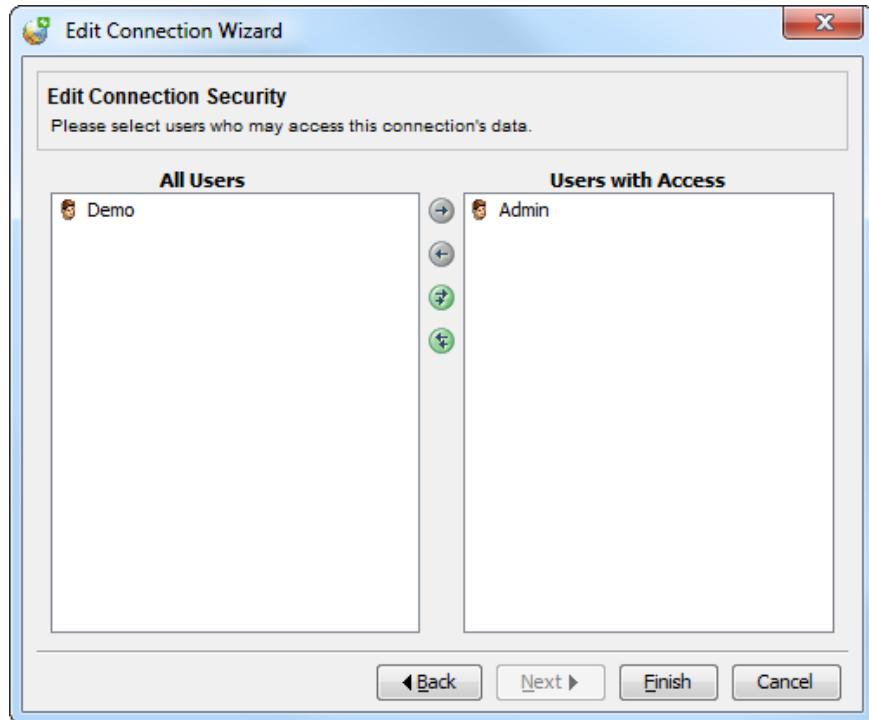
3. Enter DSDQ_Training_Source1 in the **Name** field and click **Next**.

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**.

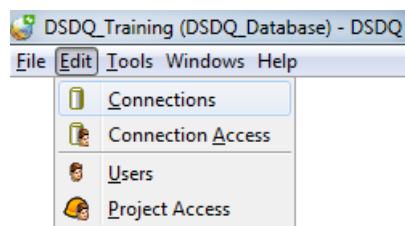
Note

System Administrators can edit all information for connections.

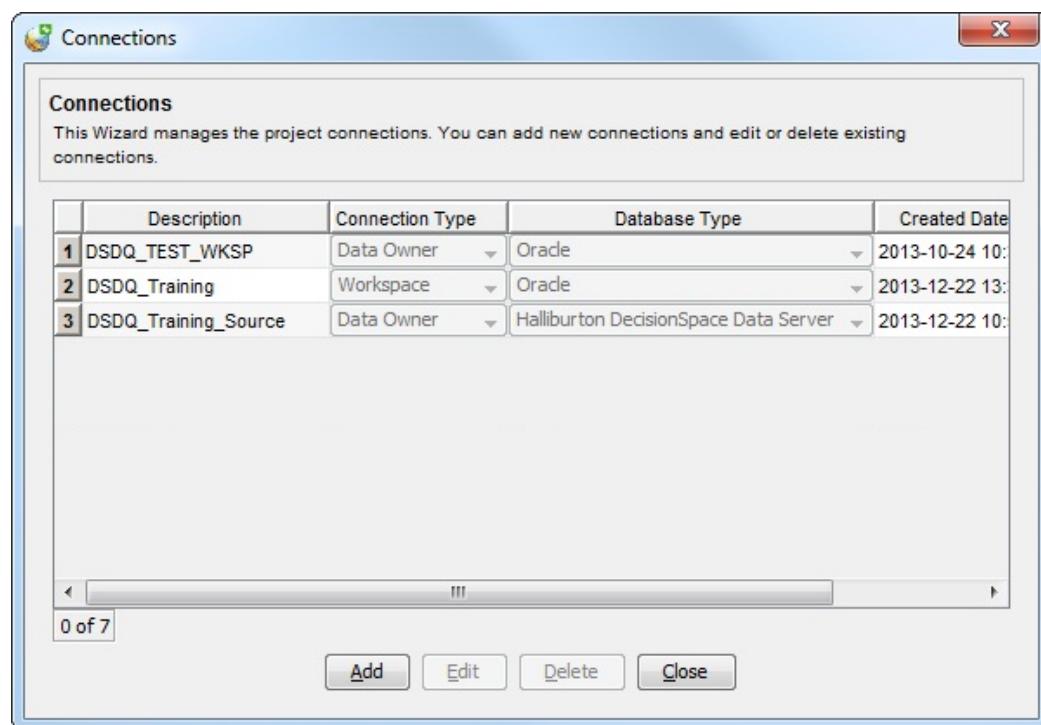
Exercise: Editing a Workspace Connection

To edit a data owner connection:

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

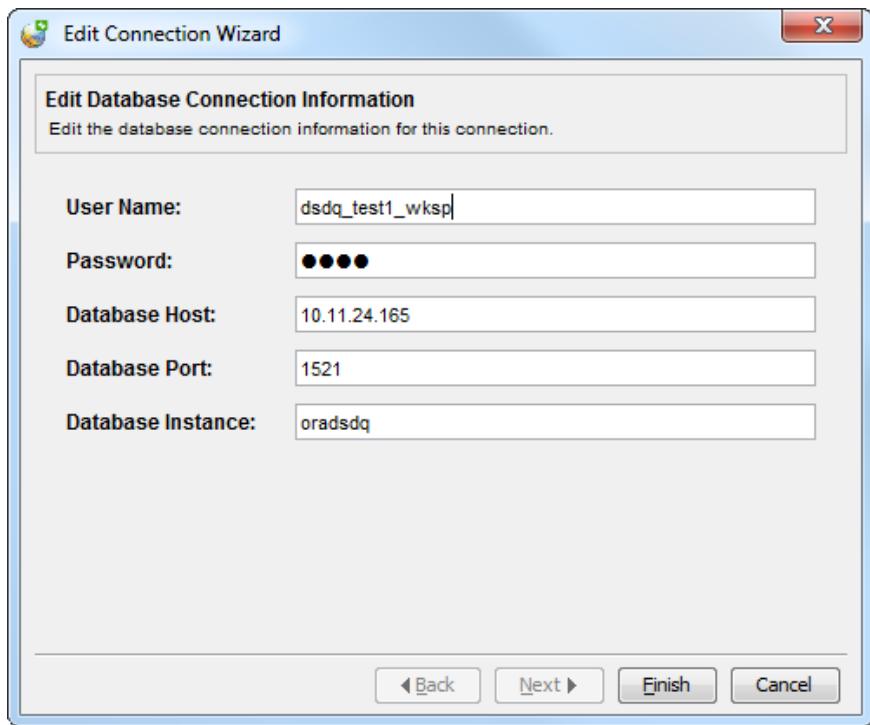


The **Connections** window appears.



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**.

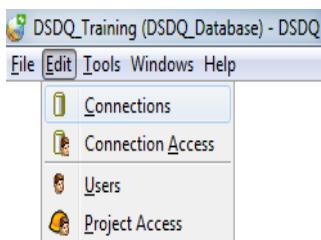
Note

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.

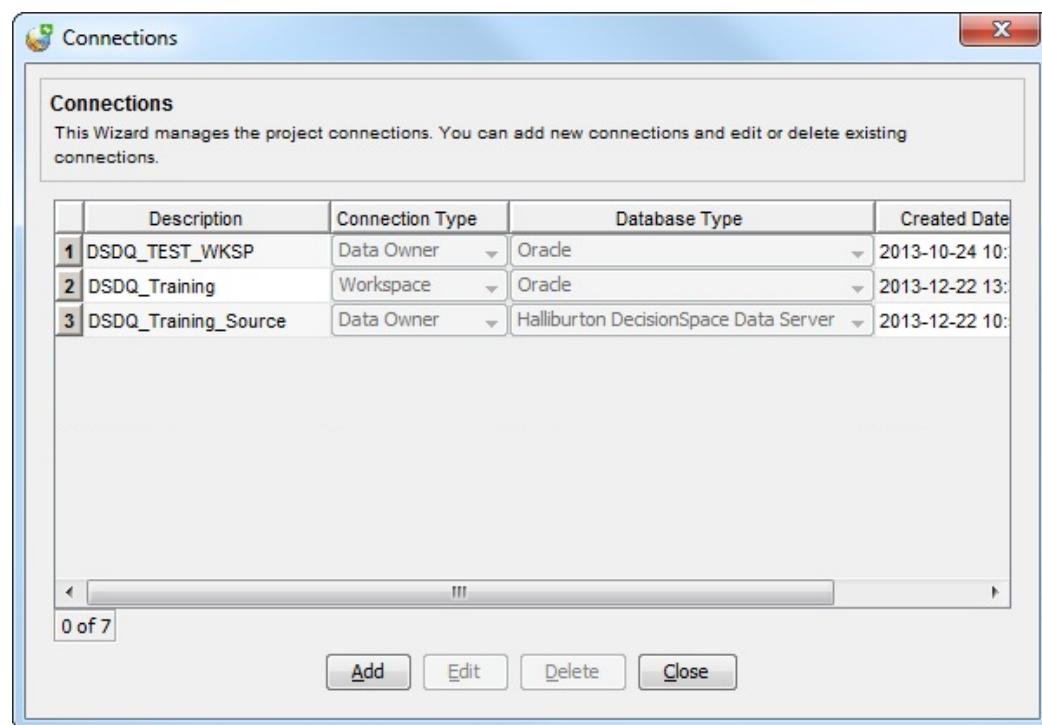
Exercise: Removing a Data Owner/Workspace Connection

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.

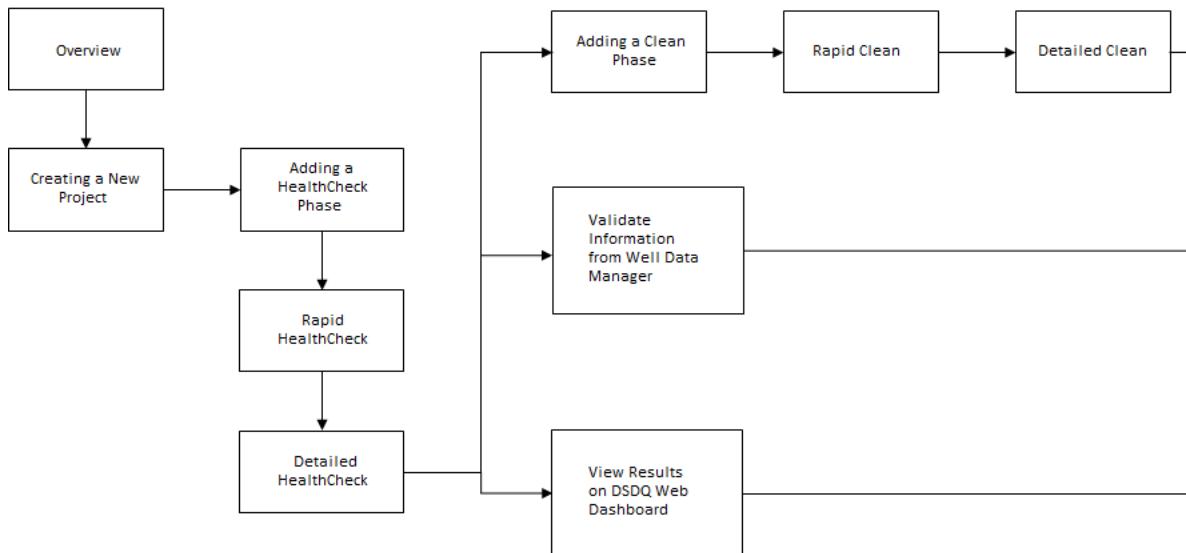
- Creating a New Project
- Evaluating data using the HealthCheck Phase
 - a) Rapid HealthCheck
 - Running Table Analysis on All Tables
 - Running Column Analysis on Columns
 - Running Table Analysis on Modeled Tables
 - Running Column Analysis on Modeled Tables
 - b) Detailed HealthCheck
 - Configuring 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



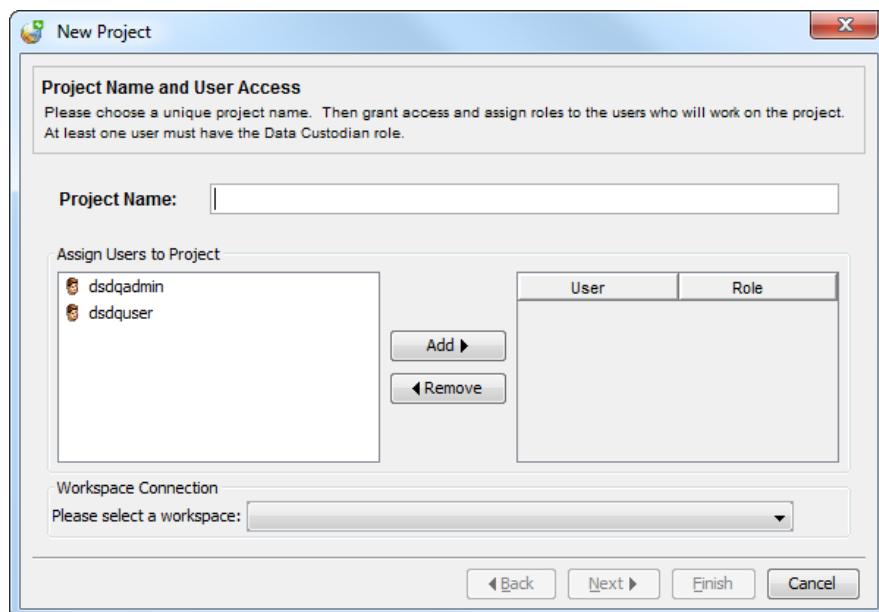
Creating a New Project

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).

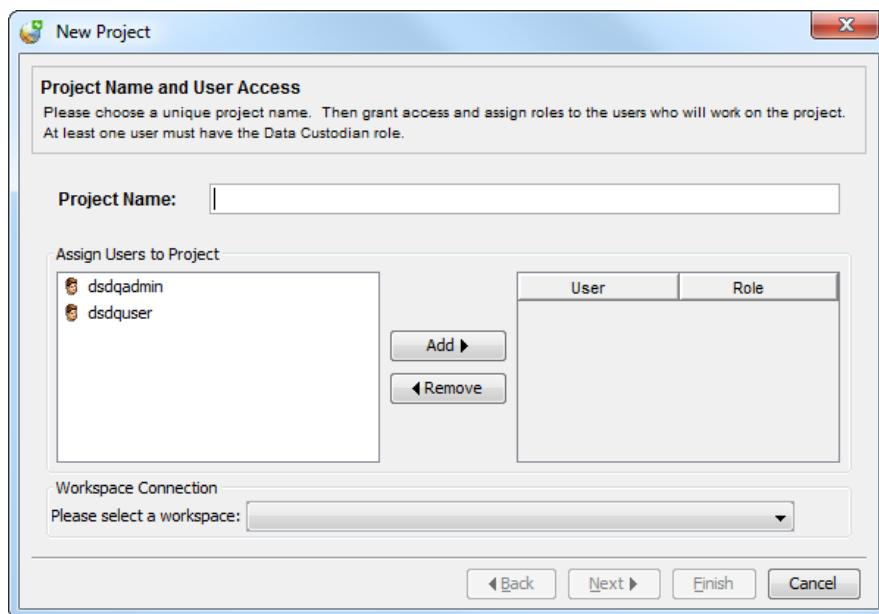
A new Project can be created:

- 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.



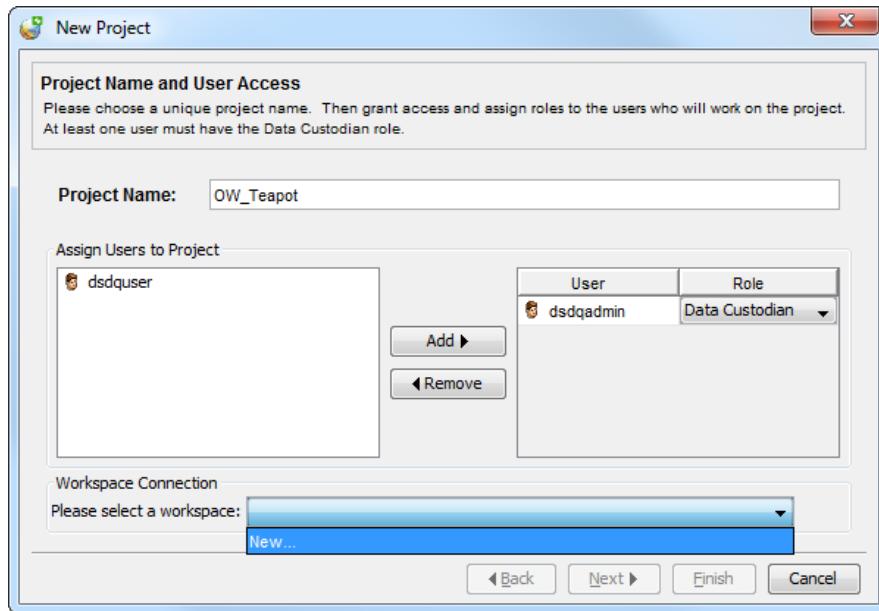
2. Enter **OW_Teapot** in the **Project Name** field.

Note

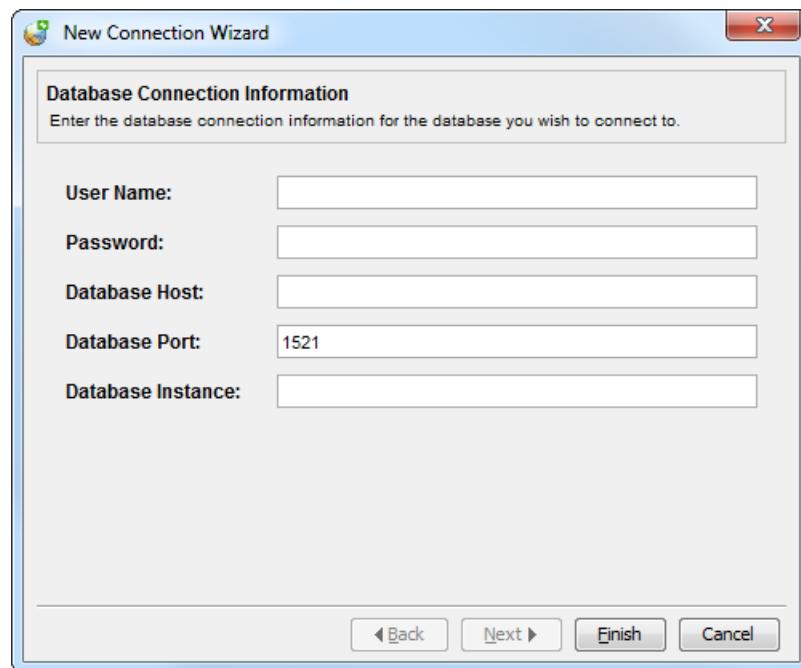
You will be using **OW_Teapot** throughout this workflow.

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.

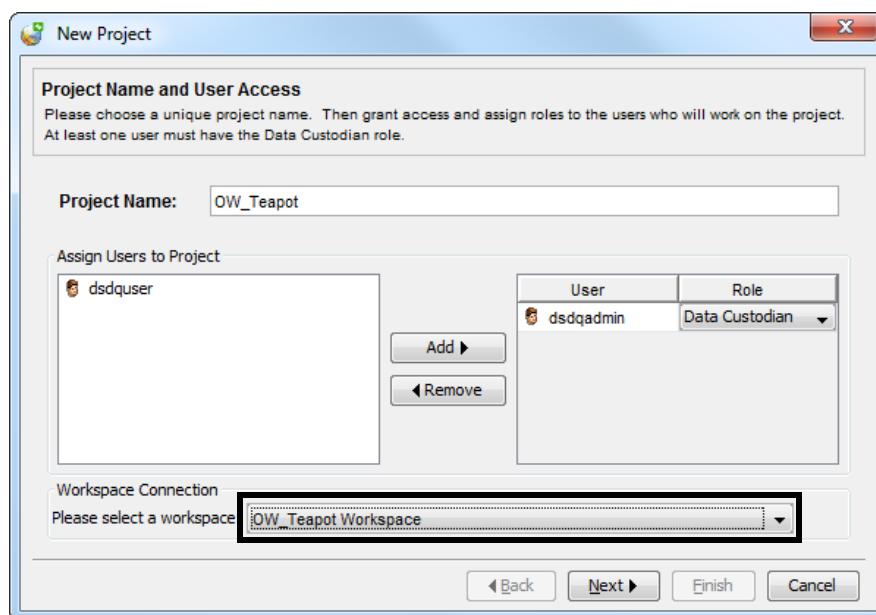


7. Enter **DSDQ_OW_Teapot_WKSP** in the **User Name** field.

Note

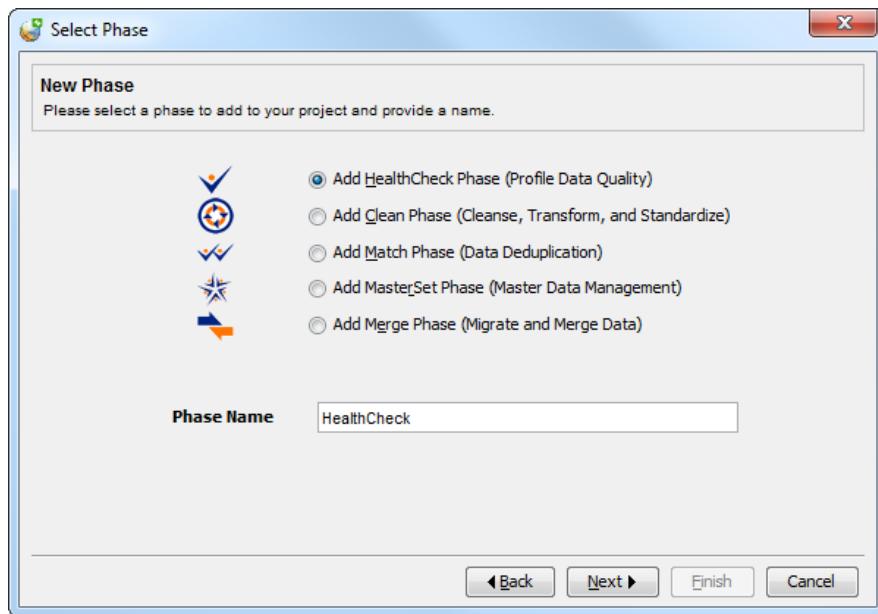
Make sure that the user and database connection have already been setup in Oracle (*reference: chapter 1, DSDQ Training Manual*).

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 **oradsdq10** 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. \

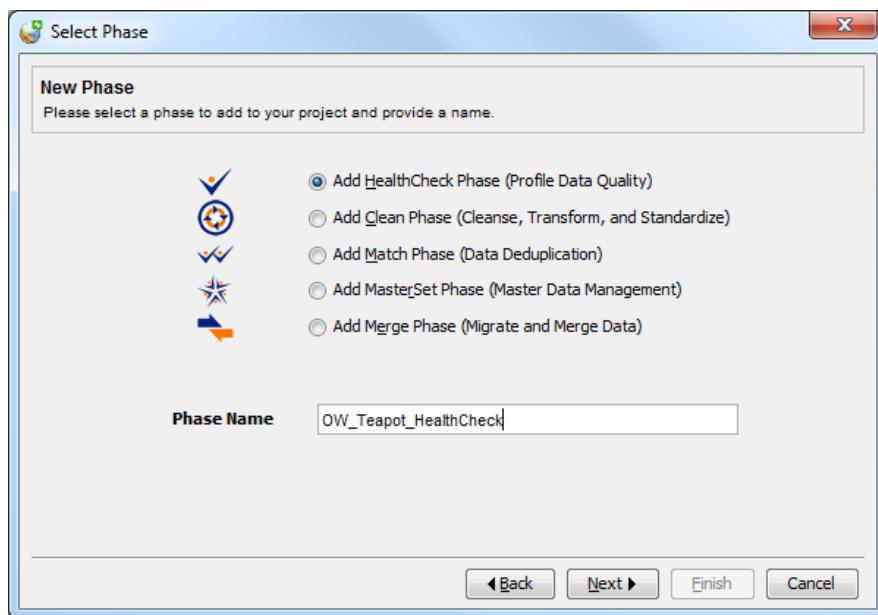


12. Click **Next** to continue.

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

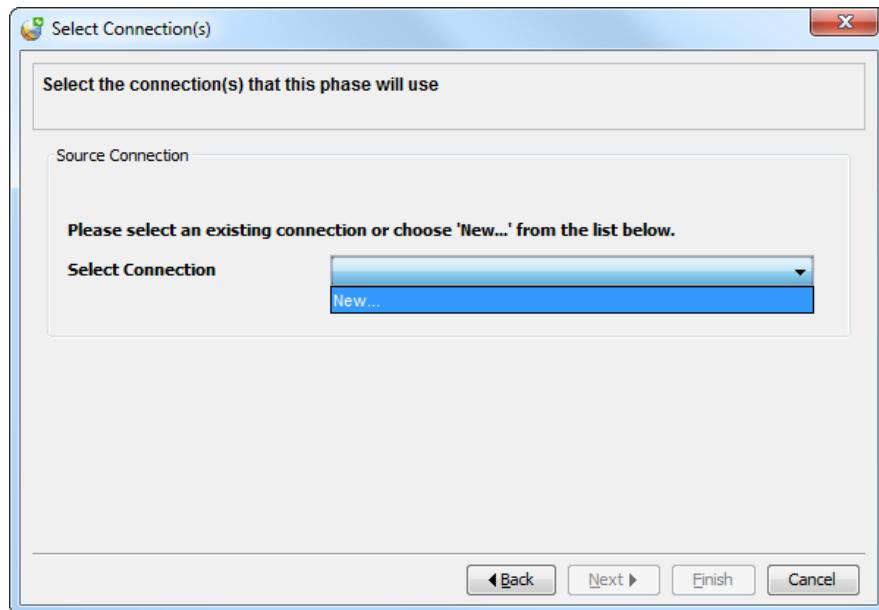


13. Enter **OW_Teapot_HealthCheck** in the **Phase Name** field.

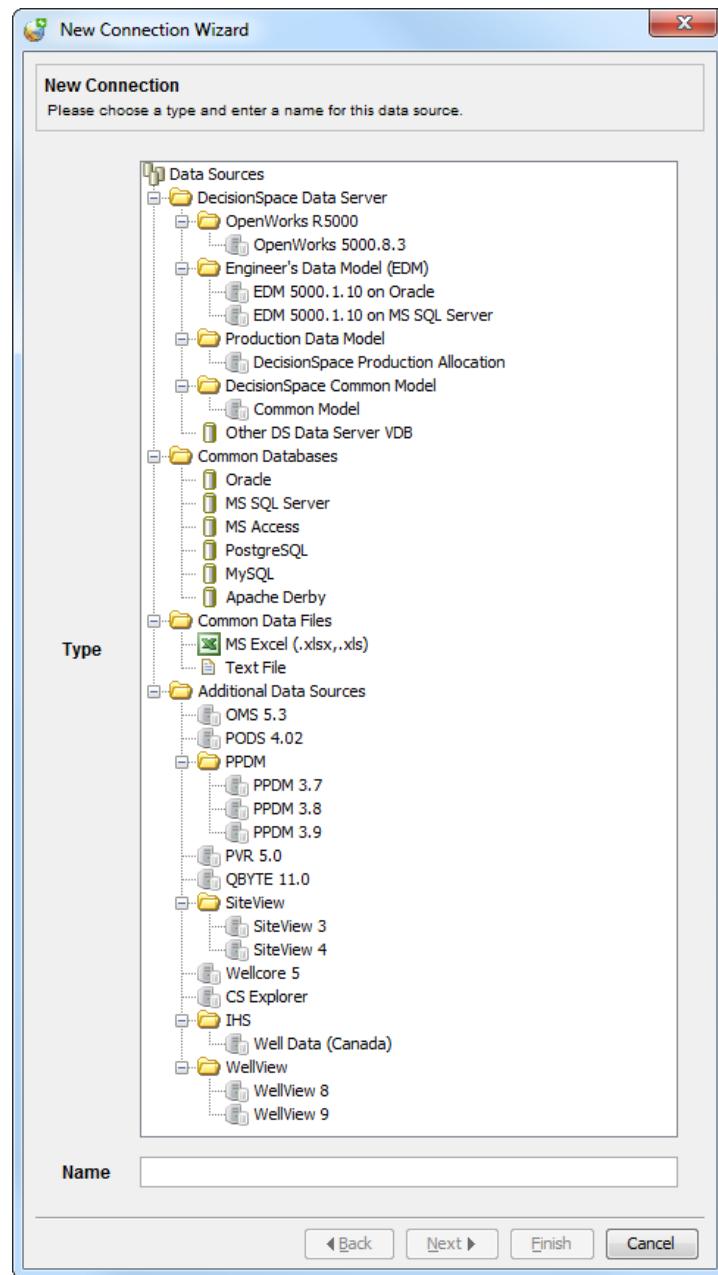


14. Click **Next** to continue.

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

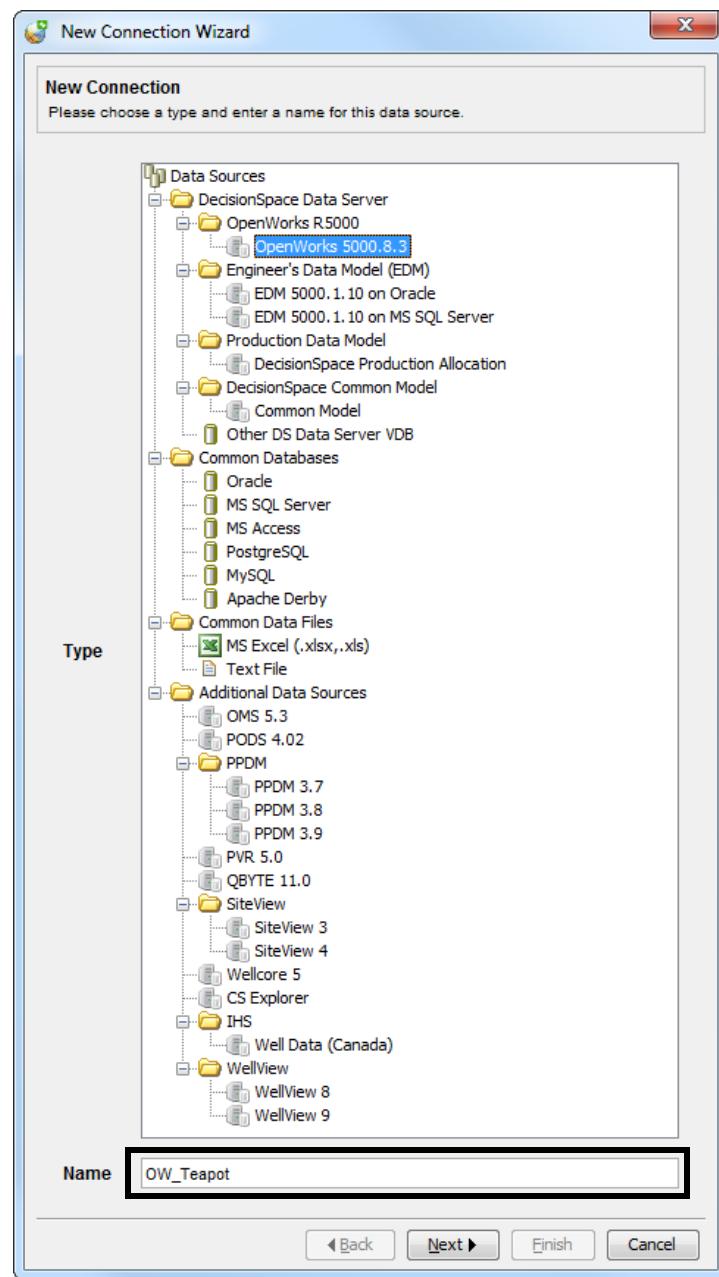


15. Select **New...** from the **Select Connection** drop-down list.
The **New Connection Wizard - New Connection** window appears.



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

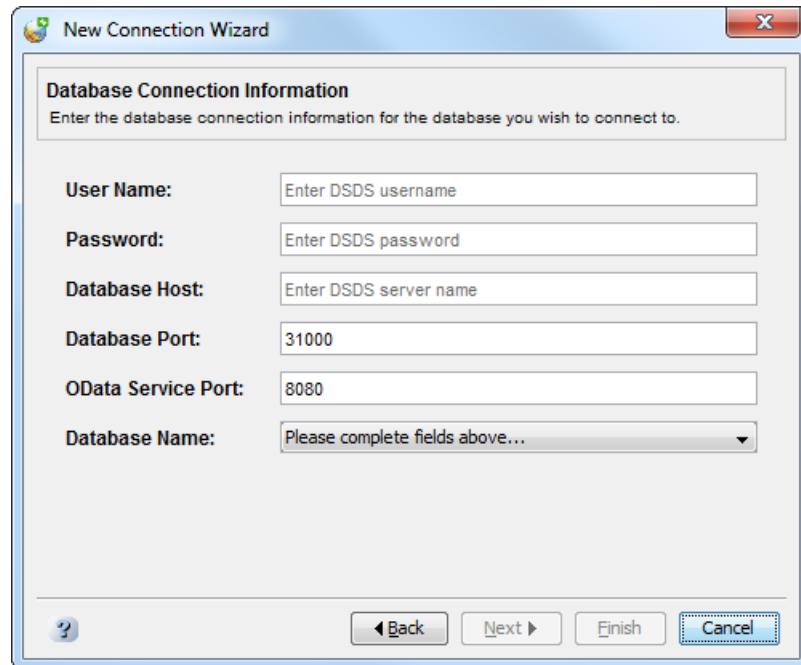
17. Enter **OW_Teapot** in the **Name** field.



18. Click **Next** to continue.

The New Connection Wizard -Database Connection

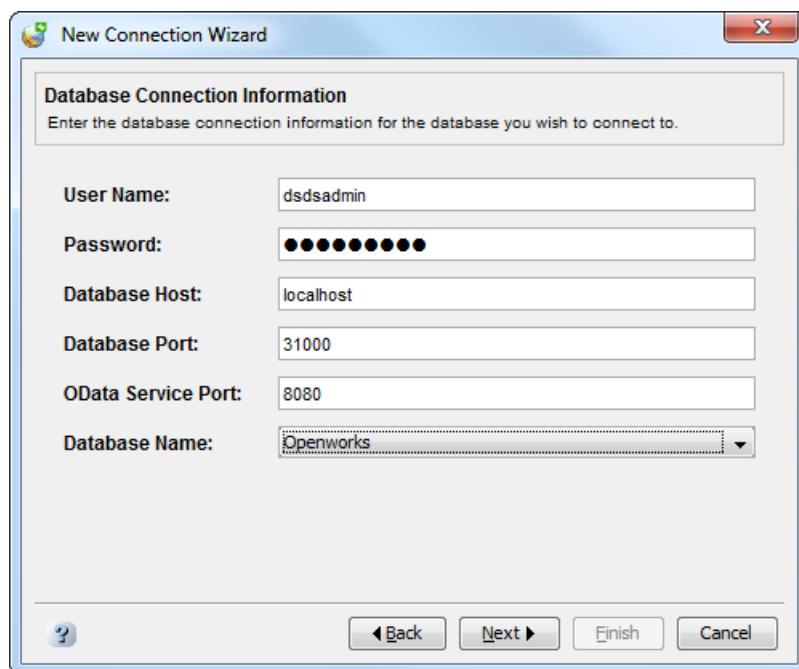
Information window appears displaying database connection information for DecisionSpace Data Server.



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**.

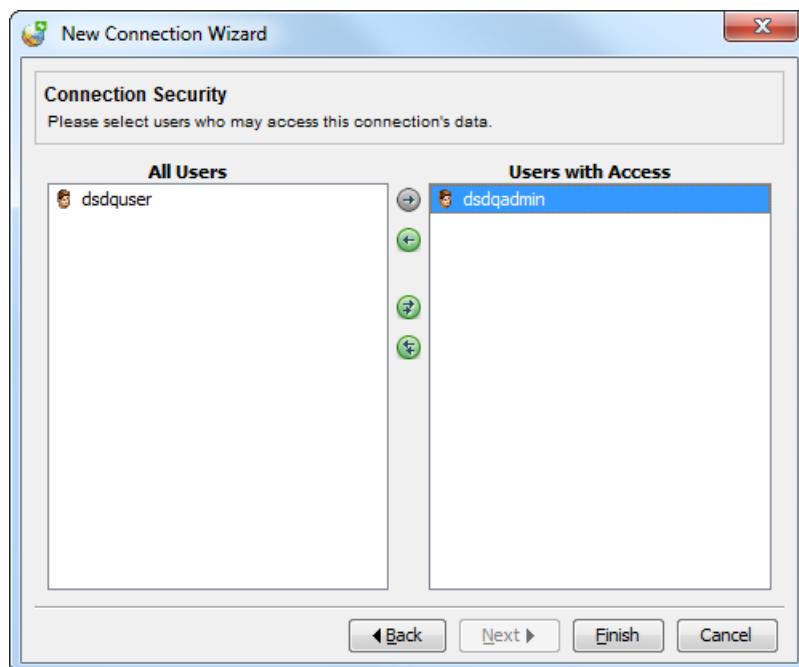
23. Accept the default **OData Service Port** value of **8080**.

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



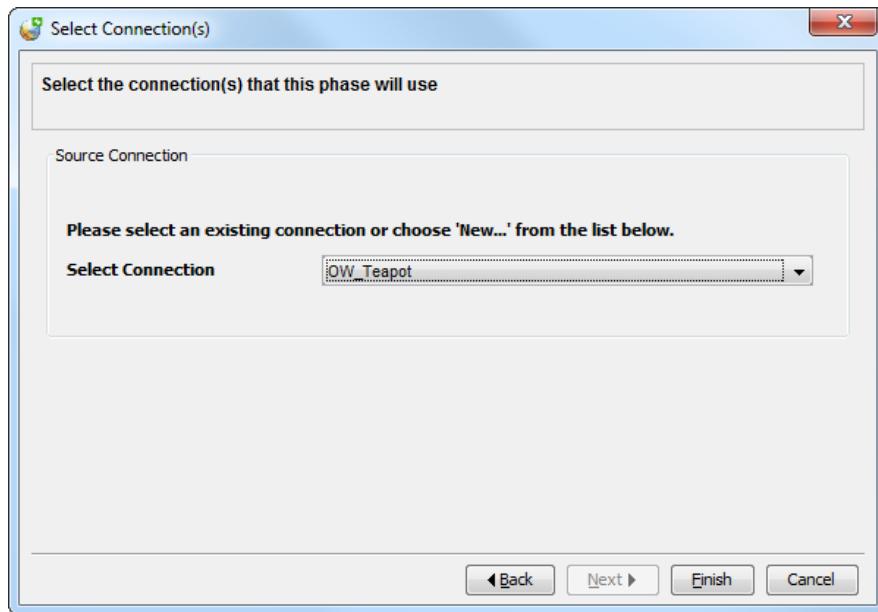
24. Click **Next** to continue.

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



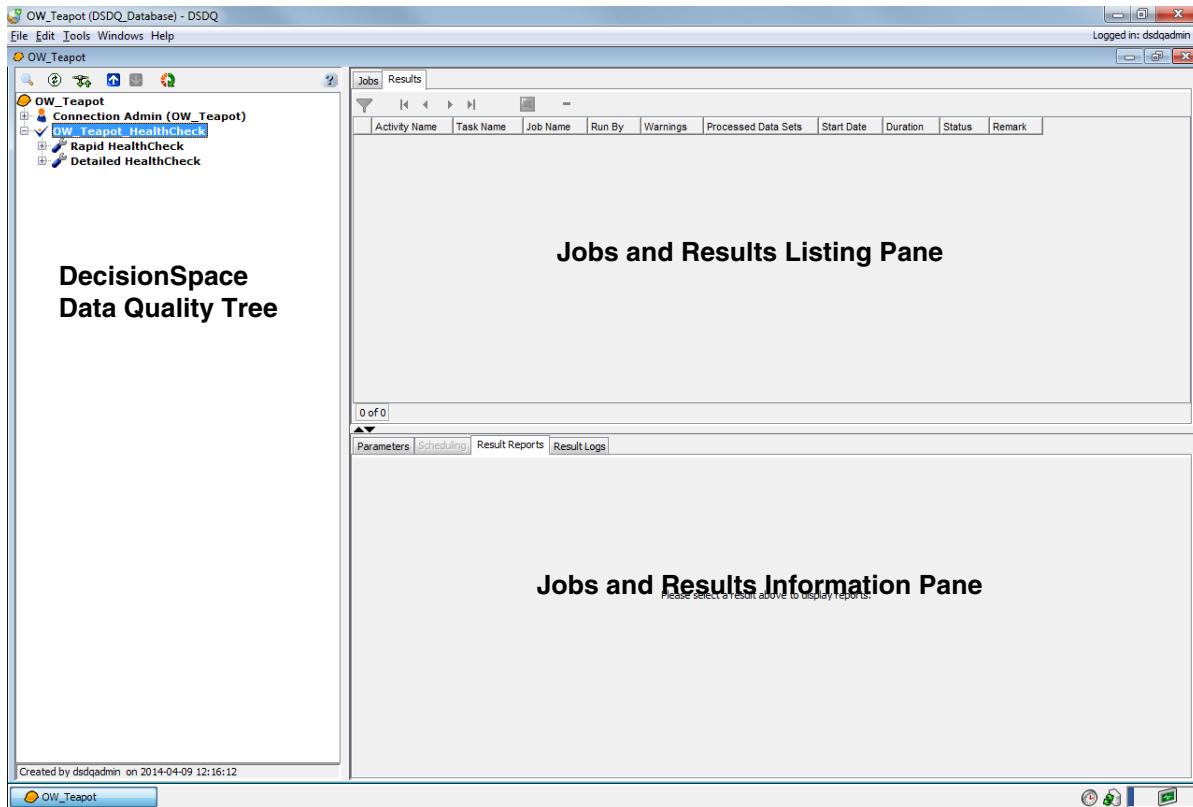
25. Click **Finish**.

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.

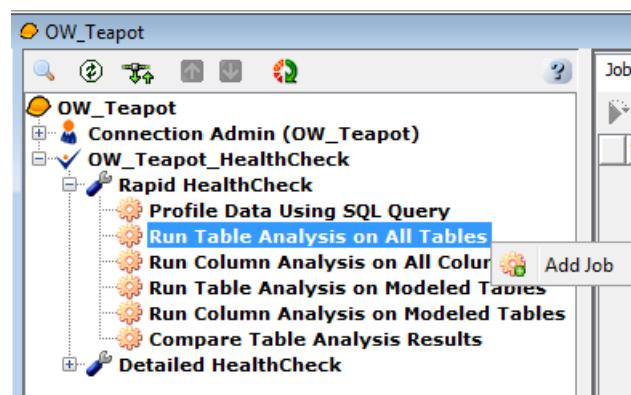
Rapid HealthCheck Activity

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

Exercise: Running Table Analysis on All Tables Task

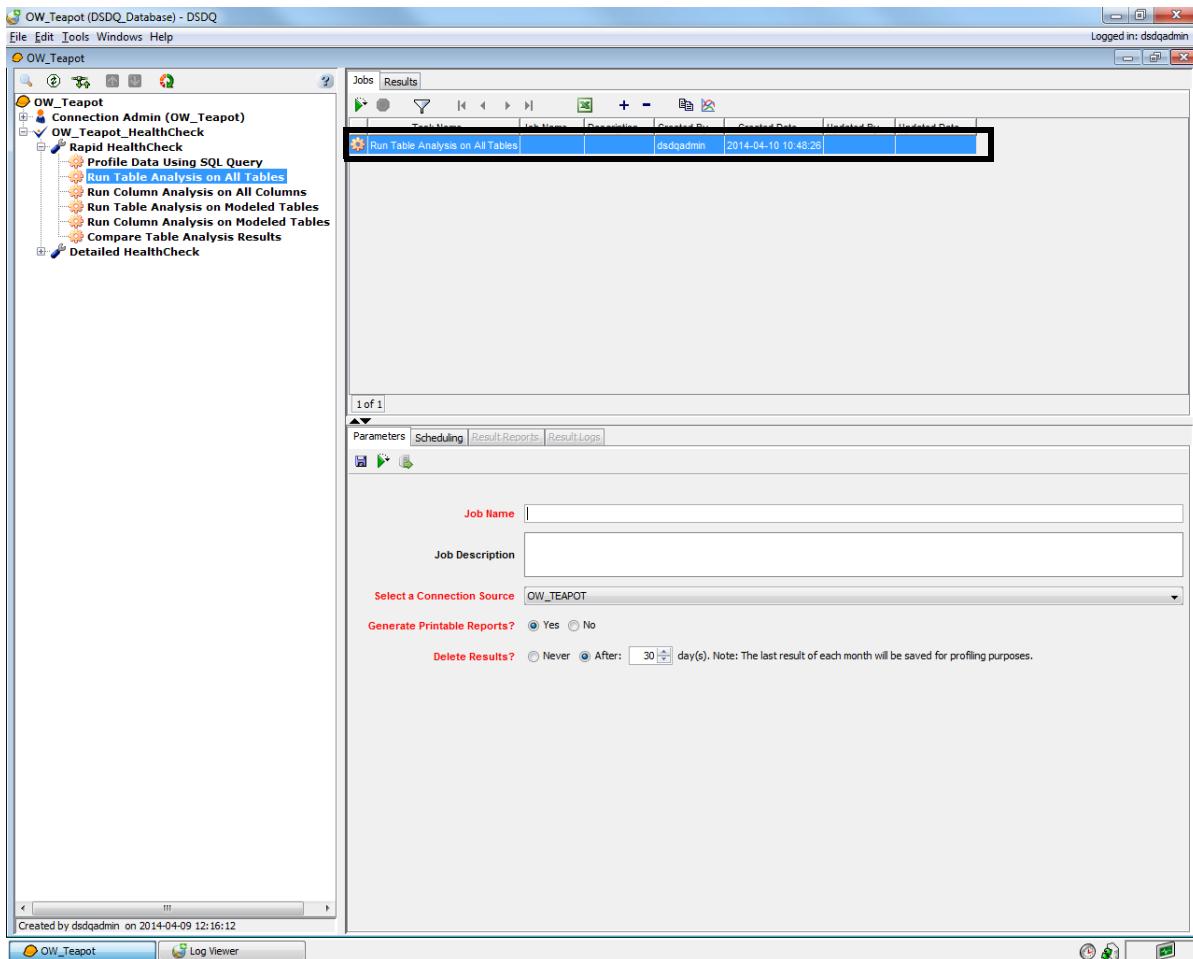
The **Run Table Analysis on All Tables** Task is used to analyze all the tables for issues and inconsistencies. In this particular exercise, we will analyze the tables and count the number of rows in them. Rows are counted when values are entered in them. To run Table Analysis on all tables:

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.

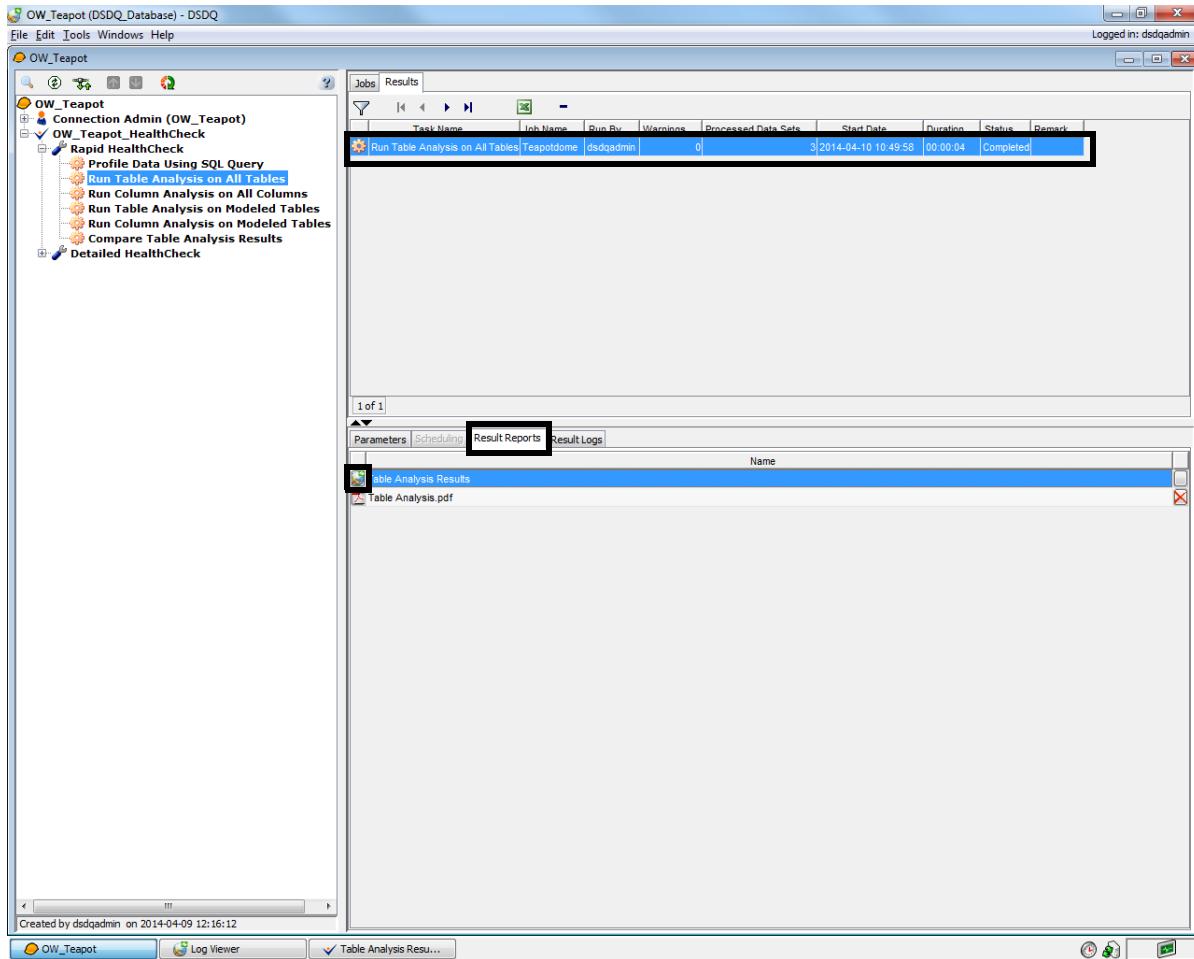


3. Enter **Teapotdome** in the **Job Name** field.
 4. Enter **Table Analysis on All Tables of OW_Teapot** in the **Job Description** field.
 5. Select **OW_Teapot** from the **Select a Connection Source** drop-down list.
 6. Select the **Yes** option for **Generate Printable Reports**.
 7. Select the **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

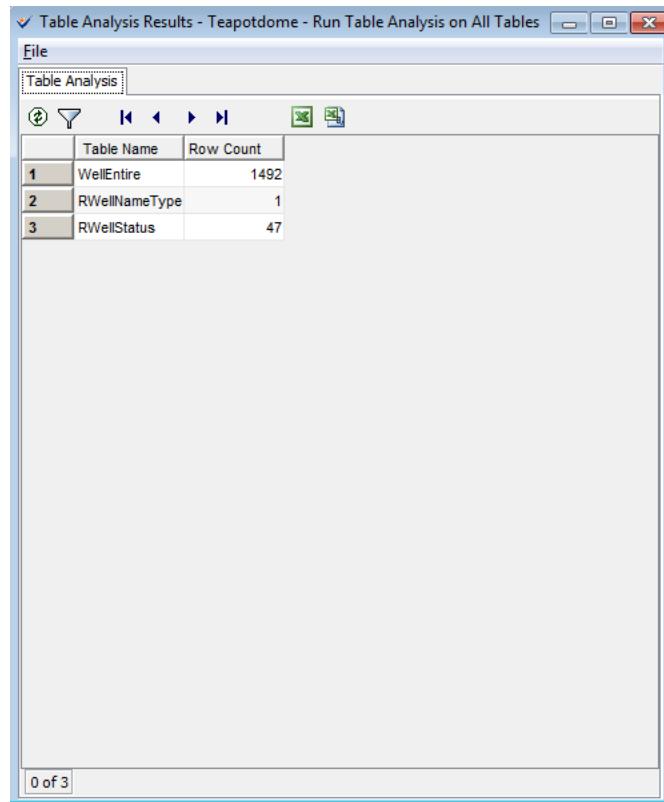
results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

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.



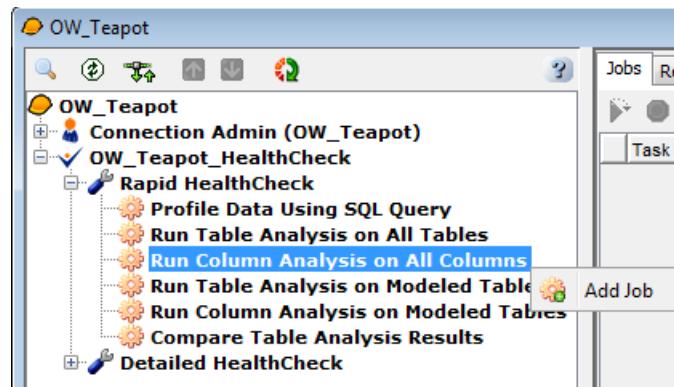
12. Select **File > Exit** to close the **Basic View Frame** window.

Exercise: Running Column Analysis on All Columns Task

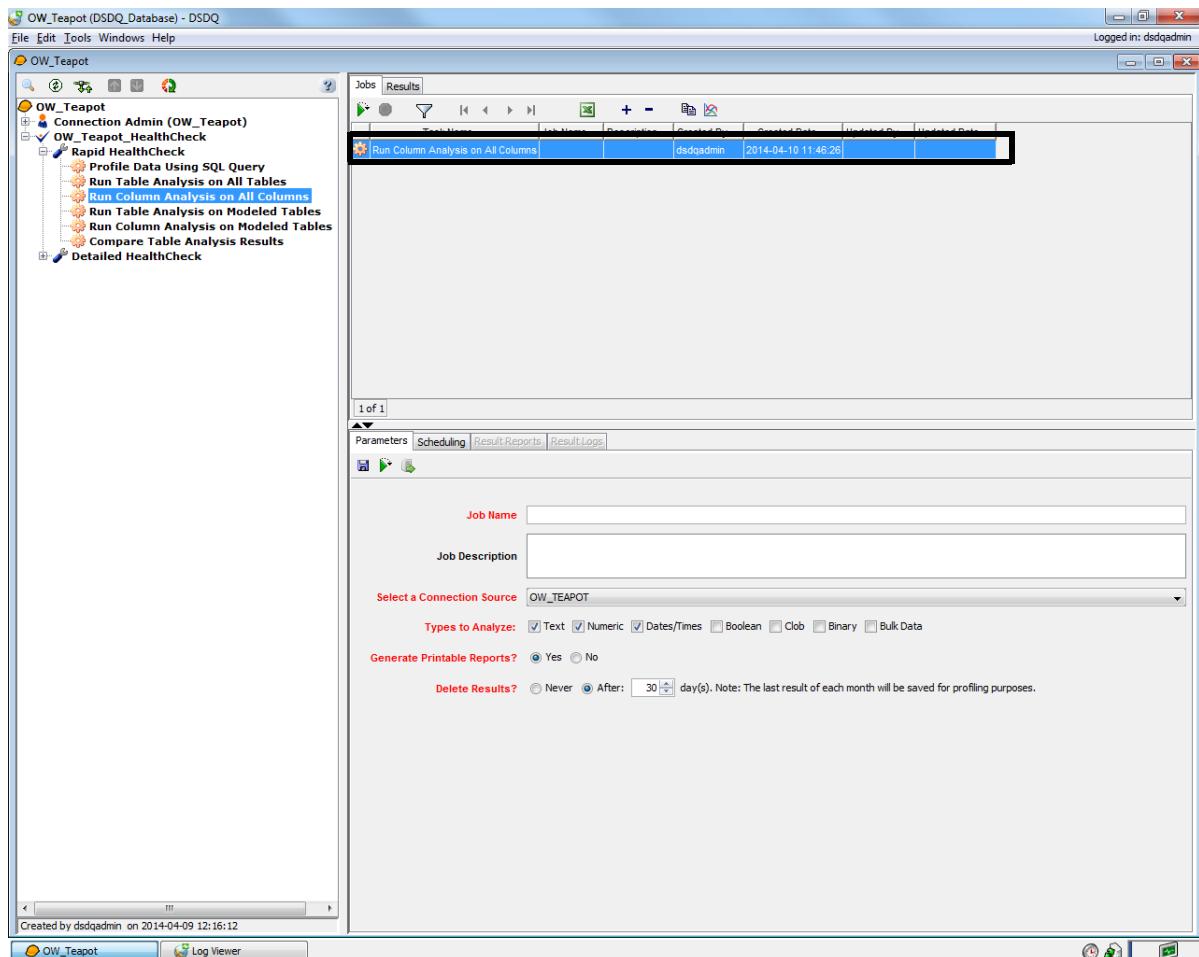
The **Run Column Analysis on All Columns** Task offers basic data profiling by checking the number of rows, not null values & unique values, percentage of row populated, minimum & maximum values, number of values with mixed cases, number of non-printable characters, and number of preceding, trailing & 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.

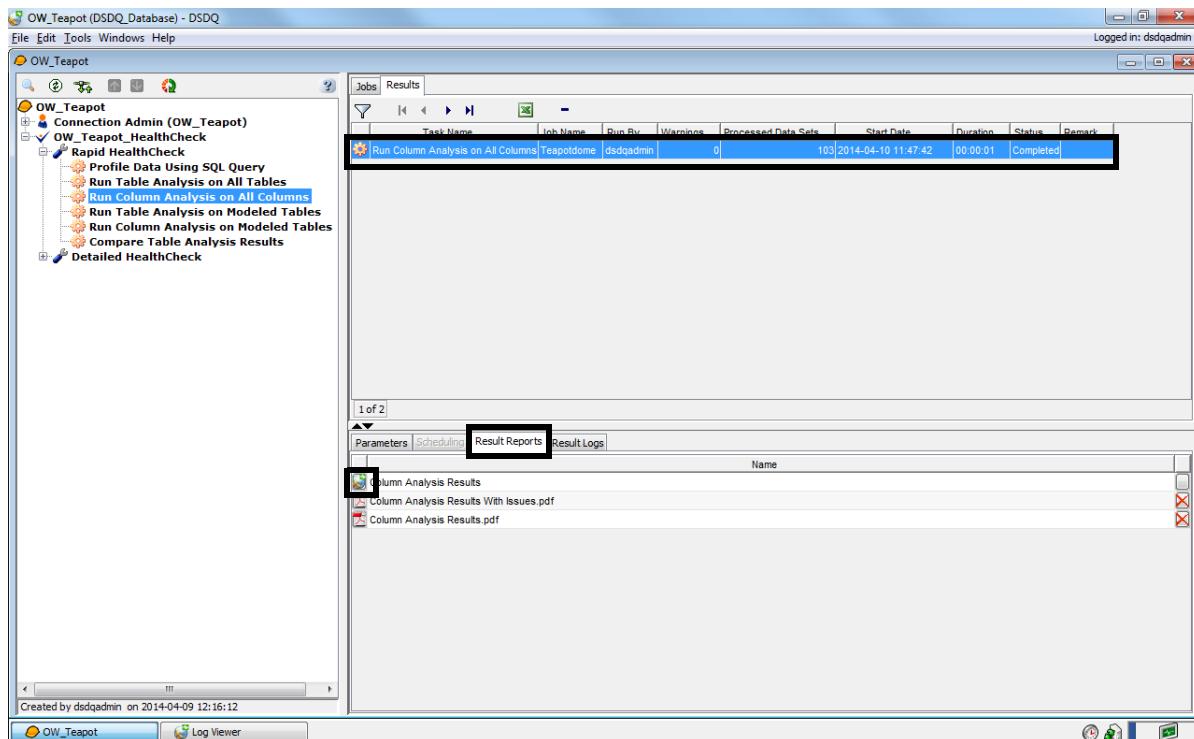


A new job is initiated and displays on the **Job and Results Information Pane** on the right side of the DecisionSpace Data Quality Project window.



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.

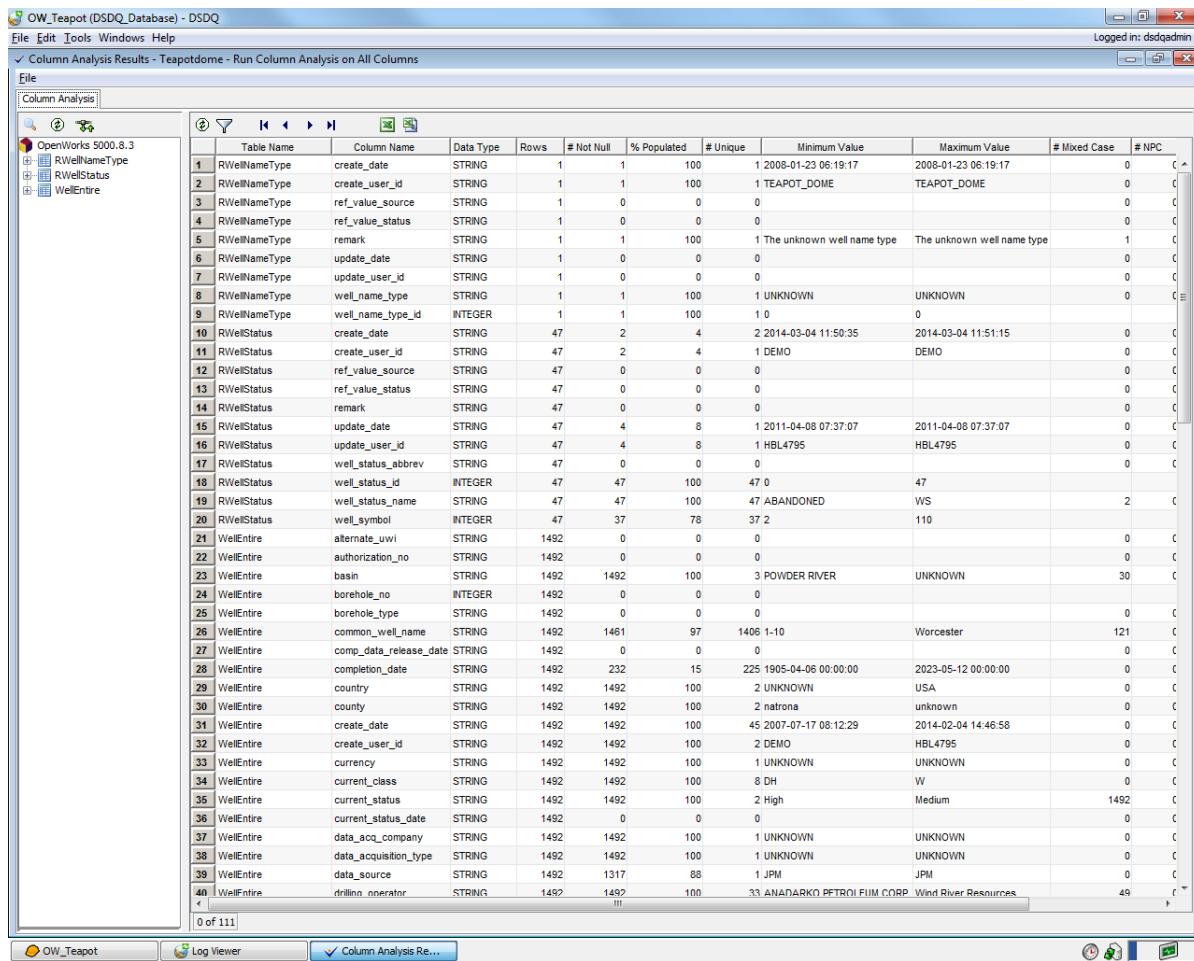


Table Name	Column Name	Data Type	Rows	# Not Null	% Populated	# Unique	Minimum Value	Maximum Value	# Mixed Case	# NPC
1 RWelNameType	create_date	STRING	1	1	100	1	2008-01-23 06:19:17	2008-01-23 06:19:17	0	C
2 RWelNameType	create_user_id	STRING	1	1	100	1	TEAPOT_DOME	TEAPOT_DOME	0	C
3 RWelNameType	ref_value_source	STRING	1	0	0	0			0	C
4 RWelNameType	ref_value_status	STRING	1	0	0	0			0	C
5 RWelNameType	remark	STRING	1	1	100	1	The unknown well name type	The unknown well name type	1	C
6 RWelNameType	update_date	STRING	1	0	0	0			0	C
7 RWelNameType	update_user_id	STRING	1	0	0	0			0	C
8 RWelNameType	wel_name_type	STRING	1	1	100	1	UNKNOWN	UNKNOWN	0	C
9 RWelNameType	wel_name_type_id	INTEGER	1	1	100	1	0	0	0	C
10 RWelStatus	create_date	STRING	47	2	4	2	2014-03-04 11:50:35	2014-03-04 11:51:15	0	C
11 RWelStatus	create_user_id	STRING	47	2	4	1	DEMO	DEMO	0	C
12 RWelStatus	ref_value_source	STRING	47	0	0	0			0	C
13 RWelStatus	ref_value_status	STRING	47	0	0	0			0	C
14 RWelStatus	remark	STRING	47	0	0	0			0	C
15 RWelStatus	update_date	STRING	47	4	8	1	2011-04-08 07:37:07	2011-04-08 07:37:07	0	C
16 RWelStatus	update_user_id	STRING	47	4	8	1	HBL4795	HBL4795	0	C
17 RWelStatus	wel_status_abbrv	STRING	47	0	0	0			0	C
18 RWelStatus	wel_status_id	INTEGER	47	47	100	47	0	47	0	C
19 RWelStatus	wel_status_name	STRING	47	47	100	47	ABANDONED	WS	2	C
20 RWelStatus	wel_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

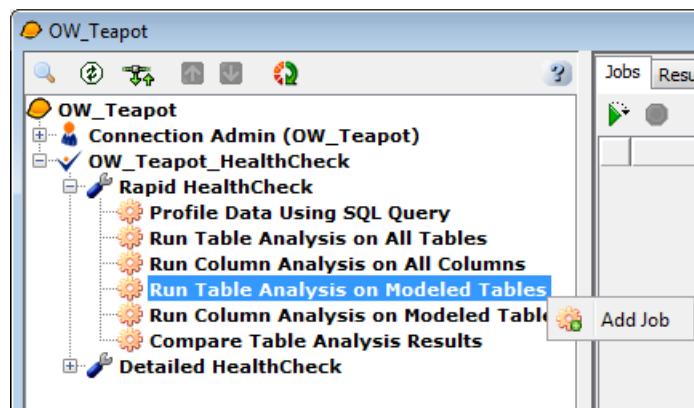
12. Select File > Exit to close the Basic View Frame window.

Exercise: Running Table Analysis on Modeled Tables Task

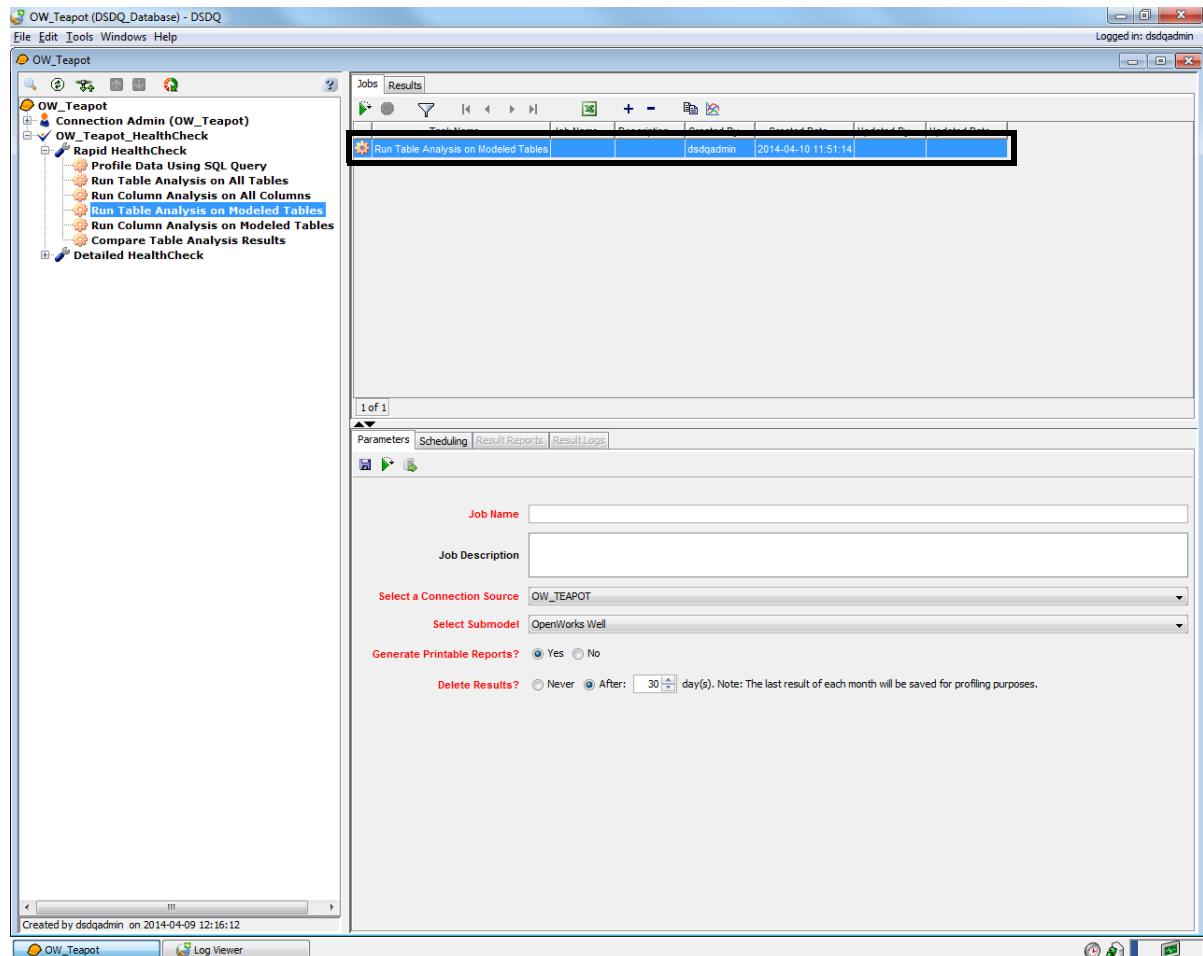
The **Run Table Analysis on Modeled Tables** Task runs only on tables that have been modeled in the **Perform Table Modeling Tool** (*reference: DecisionSpace Data Quality Training Manual, Chapter 4: Data Evaluation, Perform Table Modeling*). To run Table Analysis on all the modeled tables:

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.

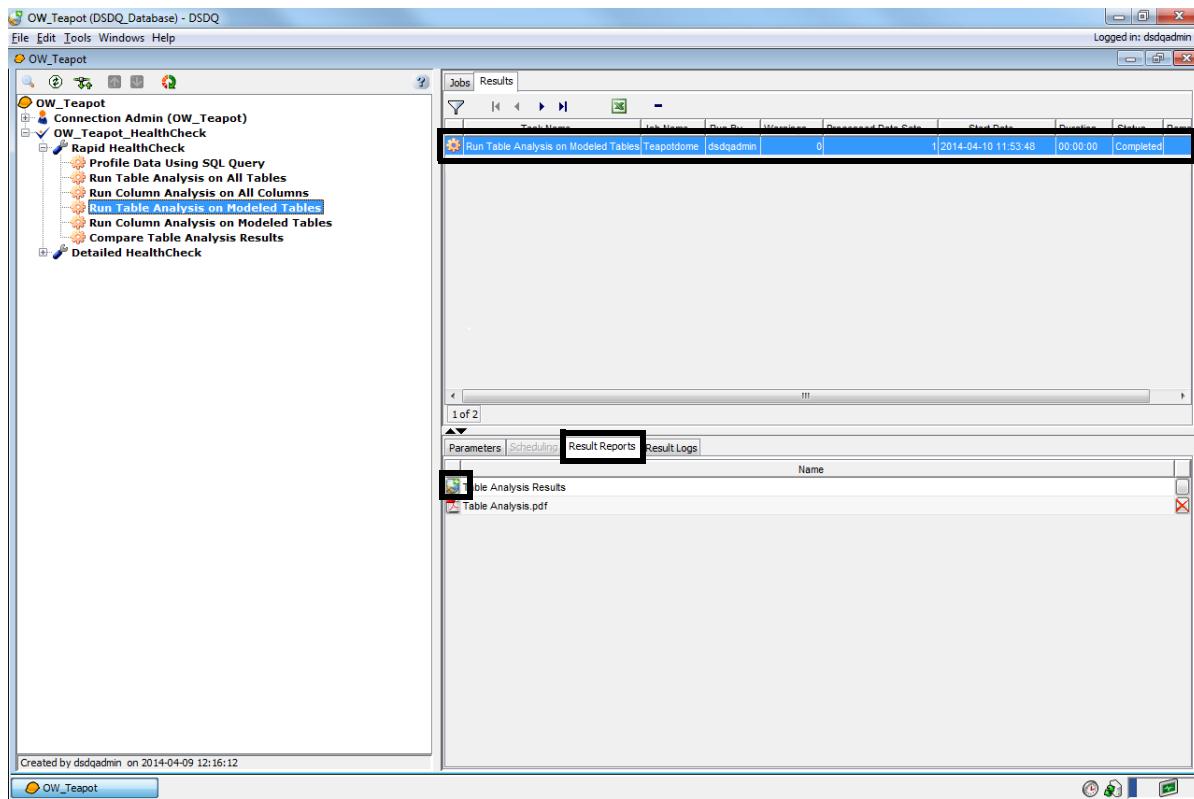


A new job is initiated and it displays on the **Job and Results Information Pane** on the right side of the DecisionSpace Data Quality Project window.

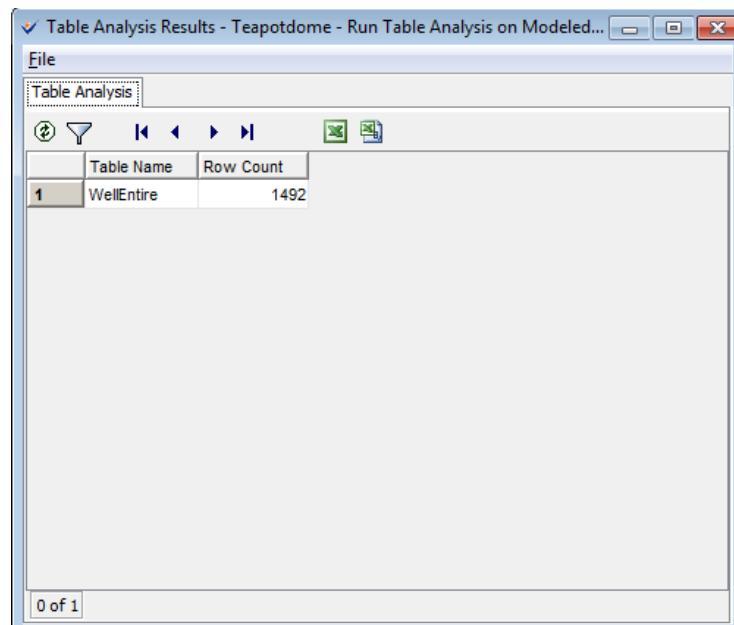


2. Enter **Teapotdome** in the **Job Name** field.

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.



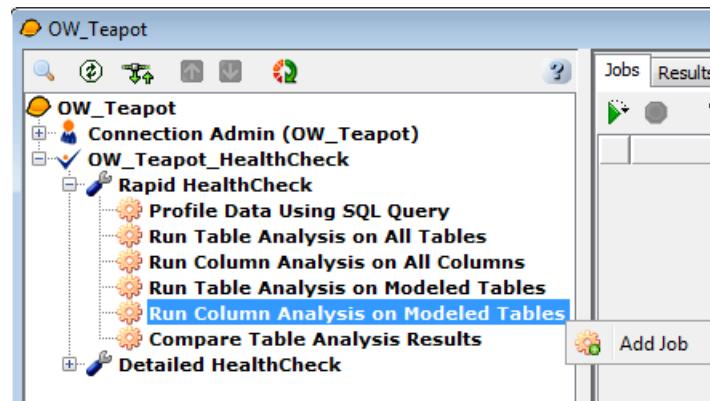
12. Select **File > Exit** to close the **Basic View Frame** window.

Exercise: Running Column Analysis on Modeled Tables Task

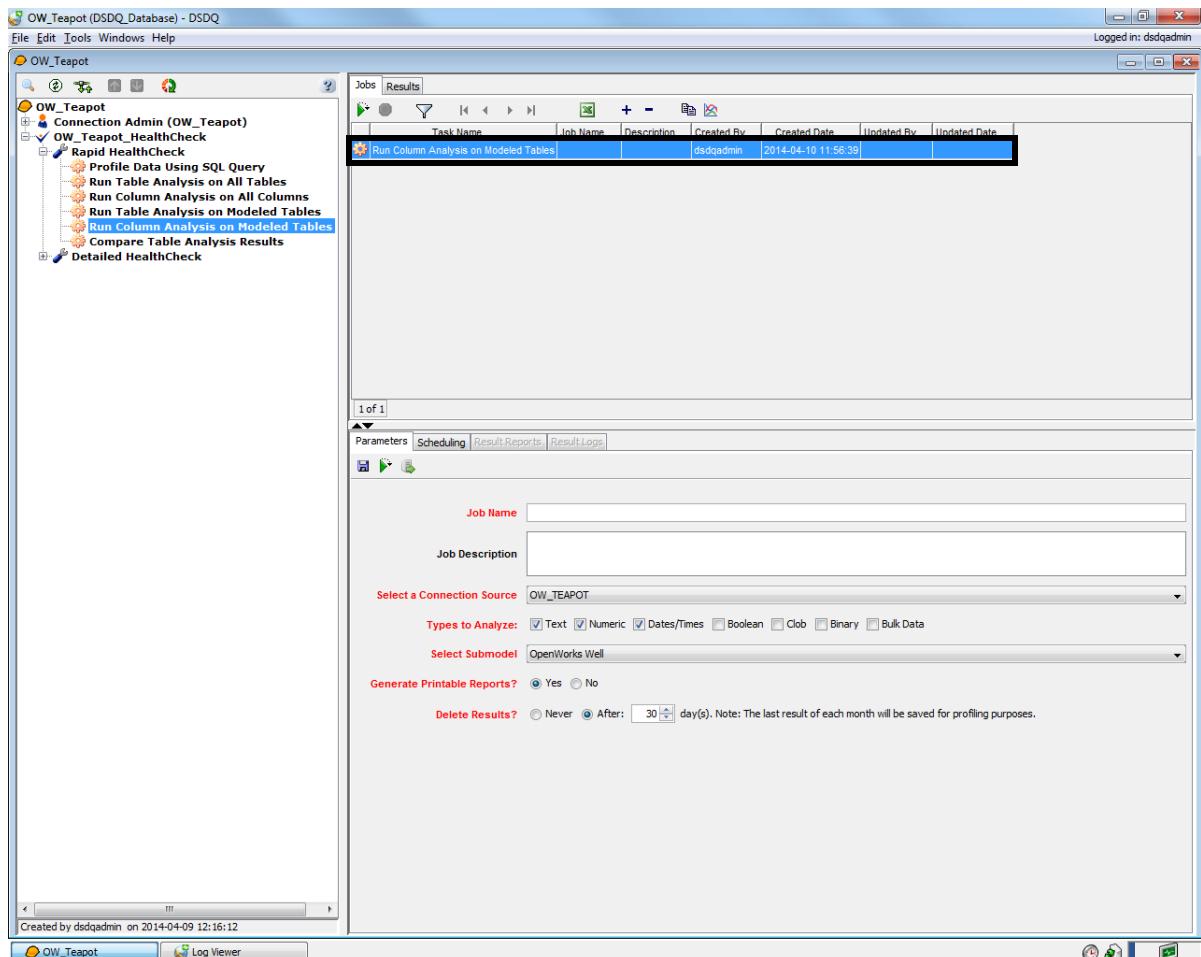
The **Run Column Analysis on Modeled Tables** Task runs only on tables that have been modeled in the **Perform Table Modeling** Tool (*reference: DecisionSpace Data Quality Training Manual, Chapter 4: Data Evaluation, Perform Table Modeling*). To run Column Analysis on all the modeled tables:

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.

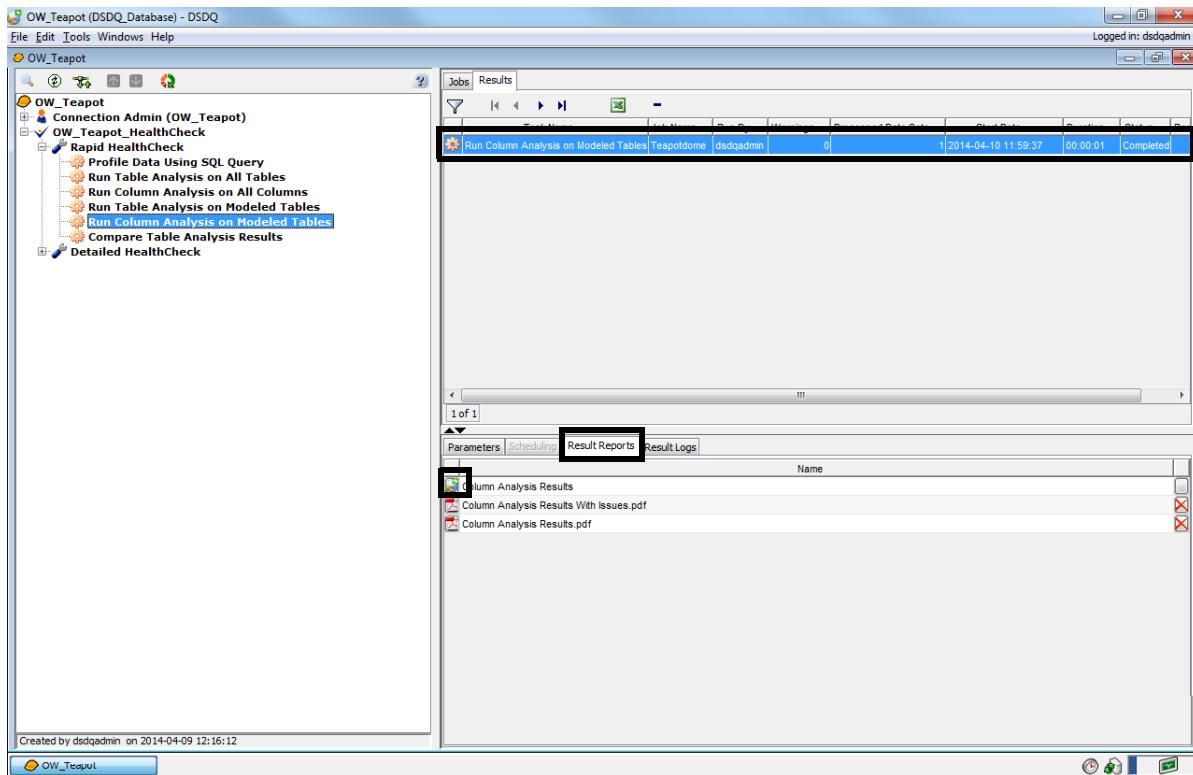


2. Enter **Teapotdome** in the **Job Name** field.

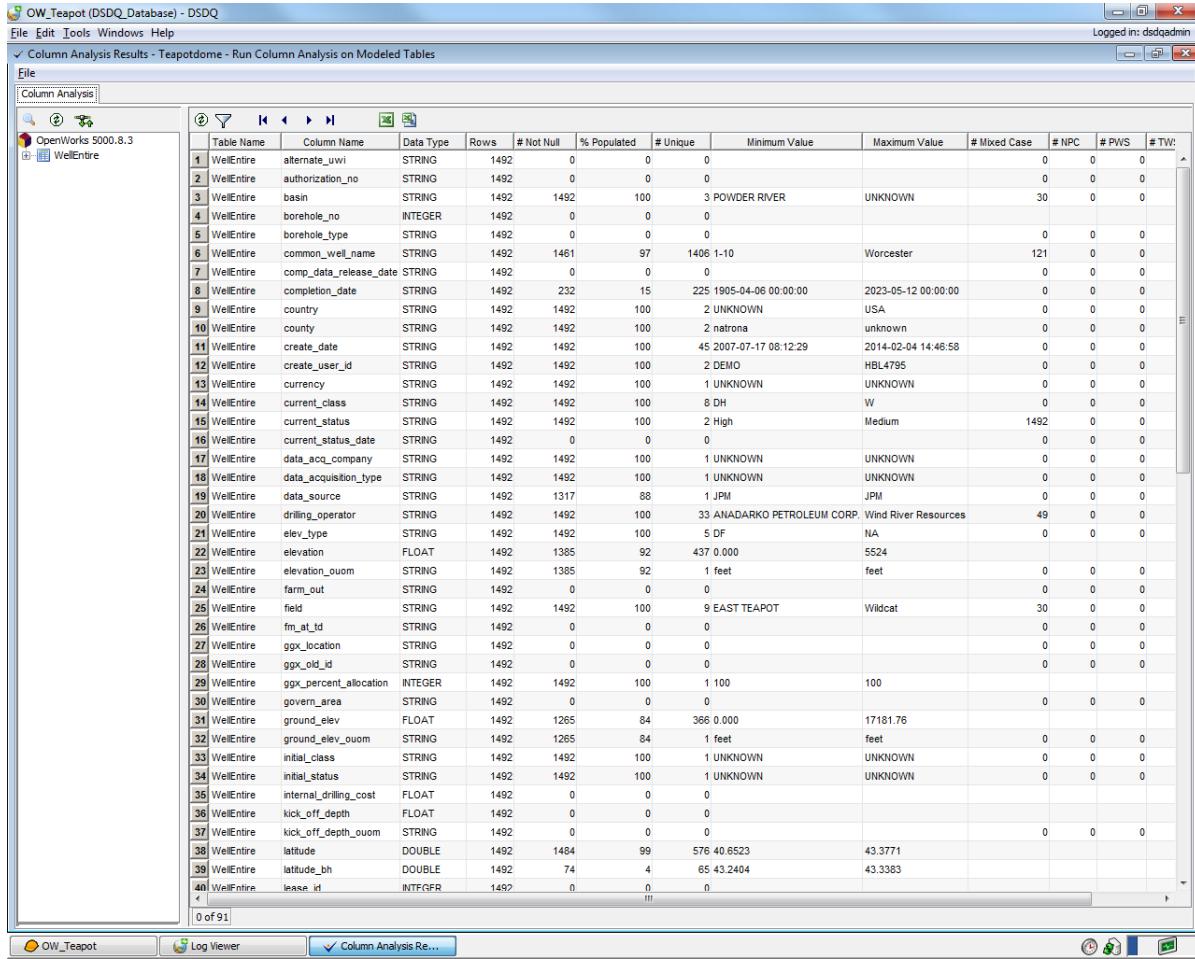
3. Enter **Column Analysis on Modeled Tables of OW_Teapot** in the **Job Description** field.
4. Select **OW_Teapot** from the **Select a Connection Source** drop-down list.
The **Text**, **Numeric** and **Dates/Times** options are selected by default for **Types to Analyze**.
5. Select **OpenWorks Well** 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.



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

Detailed HealthCheck Activity

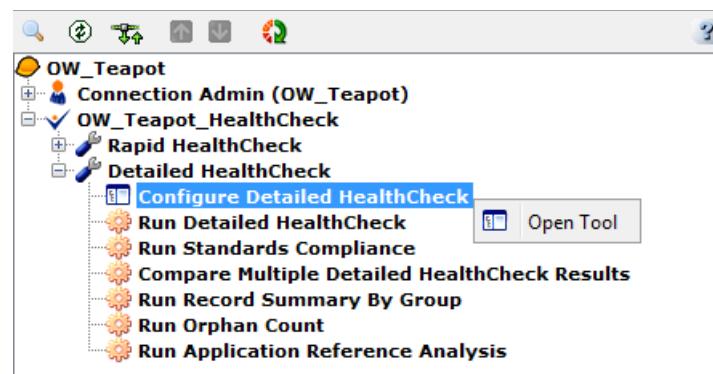
The **Detailed HealthCheck** Activity allows you to run business rules against the dataset to identify data problems.

Exercise: Configuring the Detailed HealthCheck Tool

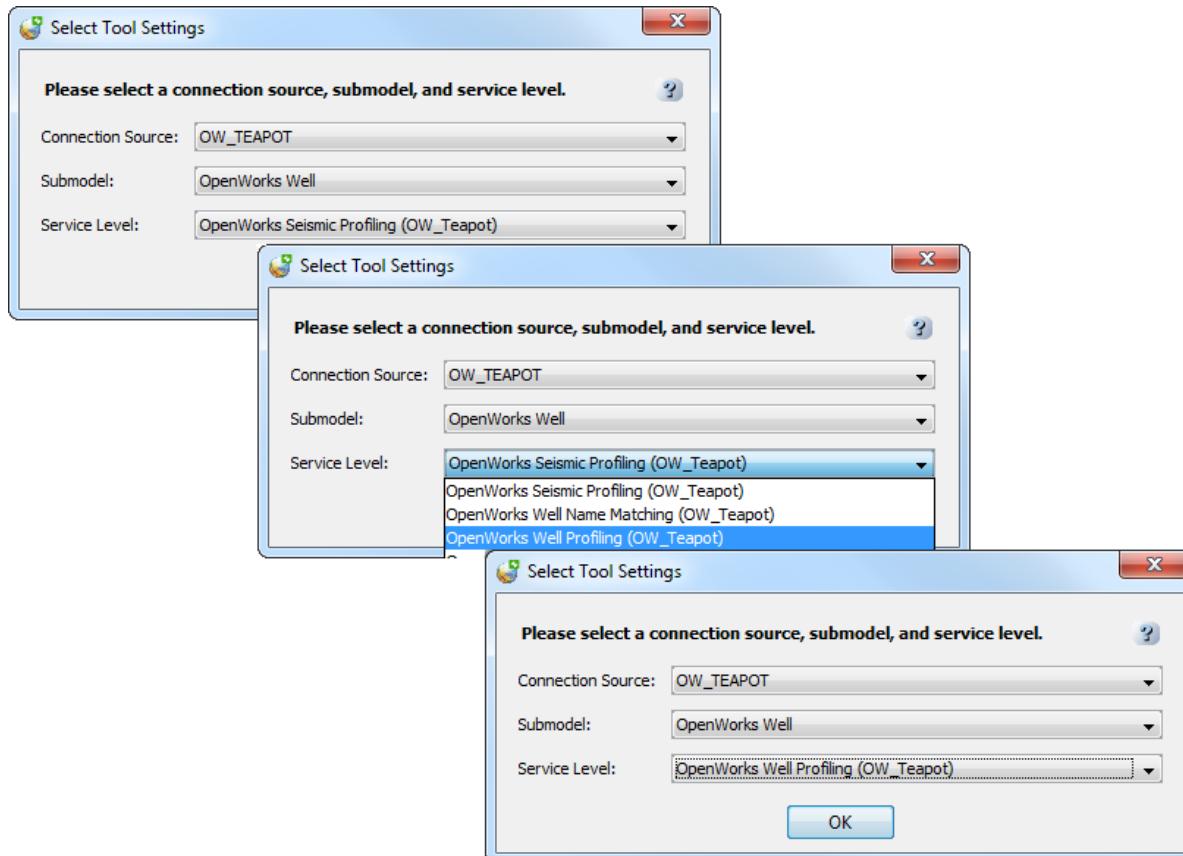
The **Configure Detailed HealthCheck** Tool configures service levels for testing prior to running the **Run Detailed HealthCheck** Task. You can select which requirements are to be enabled/disabled in the service level. You can also select subset of the total data to be used when testing a service level.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

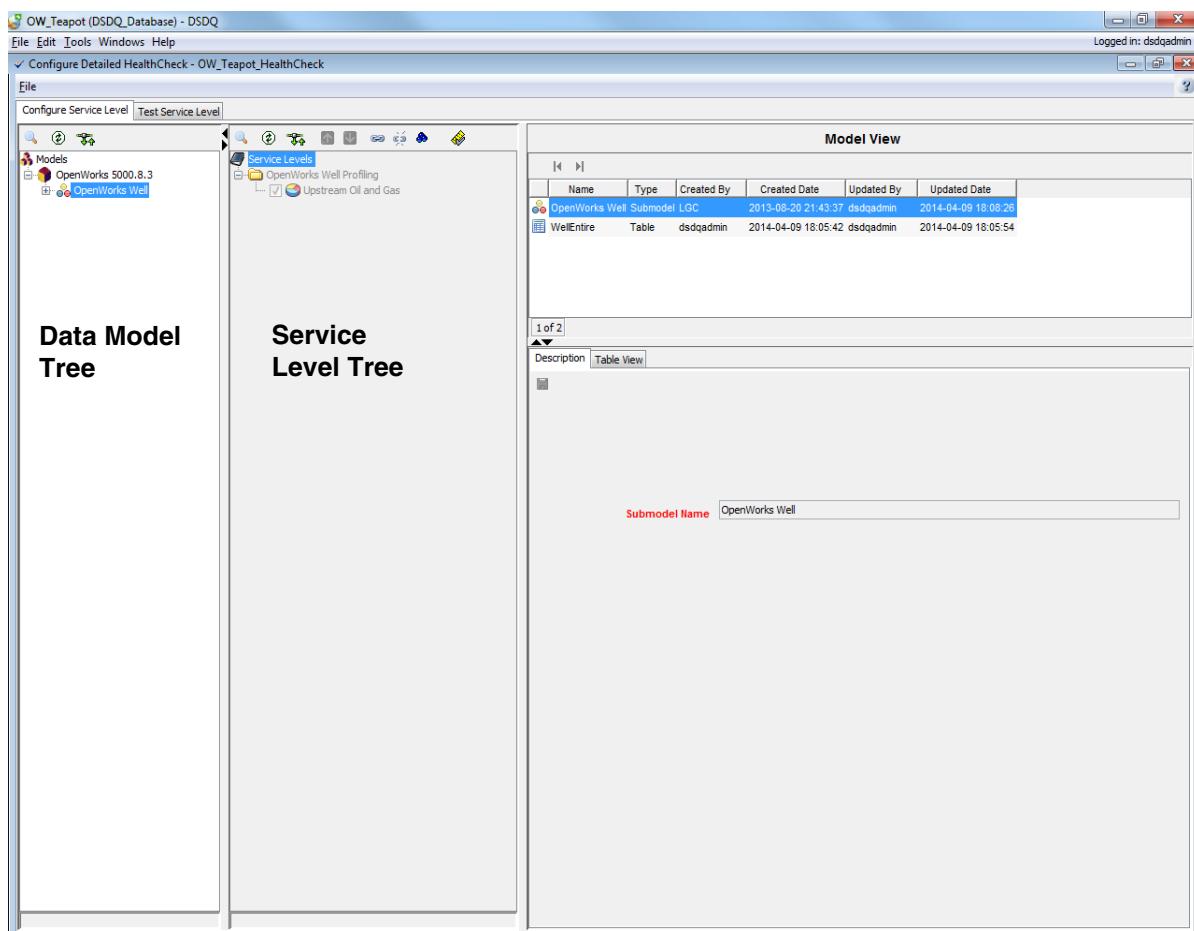


The **Select Tool Settings** window appears:



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**.

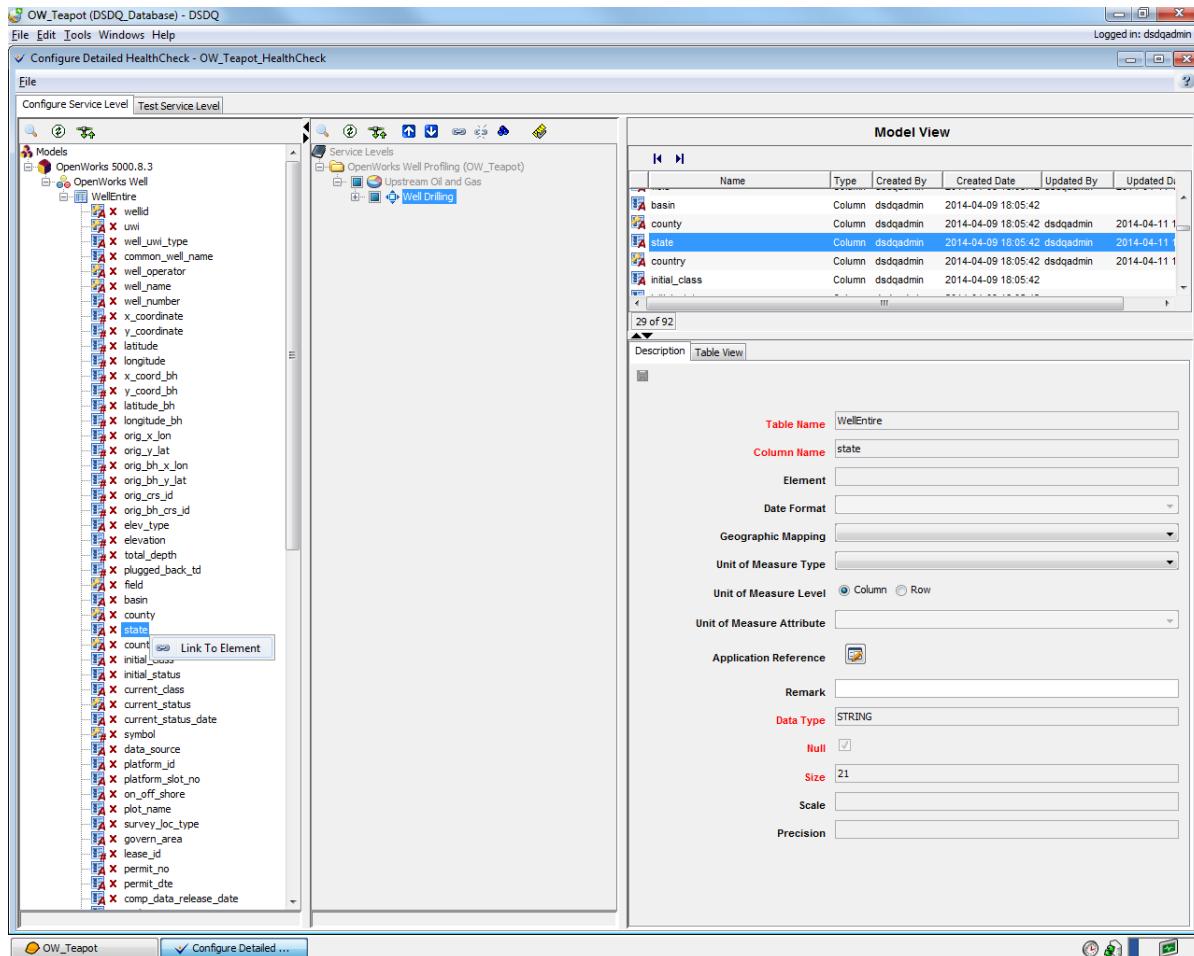
The **Configure Detailed HealthCheck** window appears.



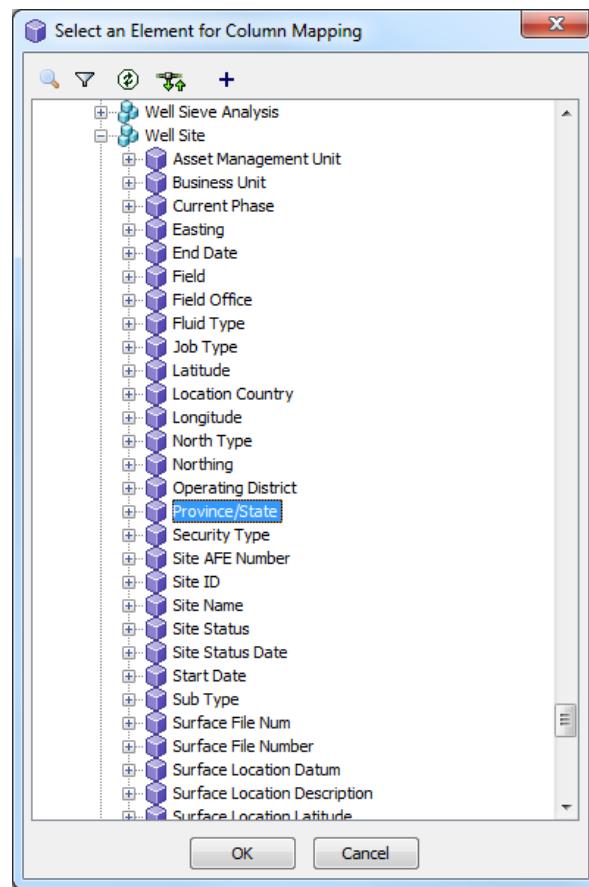
The Data Model Tree populates with the submodel selected in step 4. The Service Level Tree populates with the service level selected in step 5.

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.

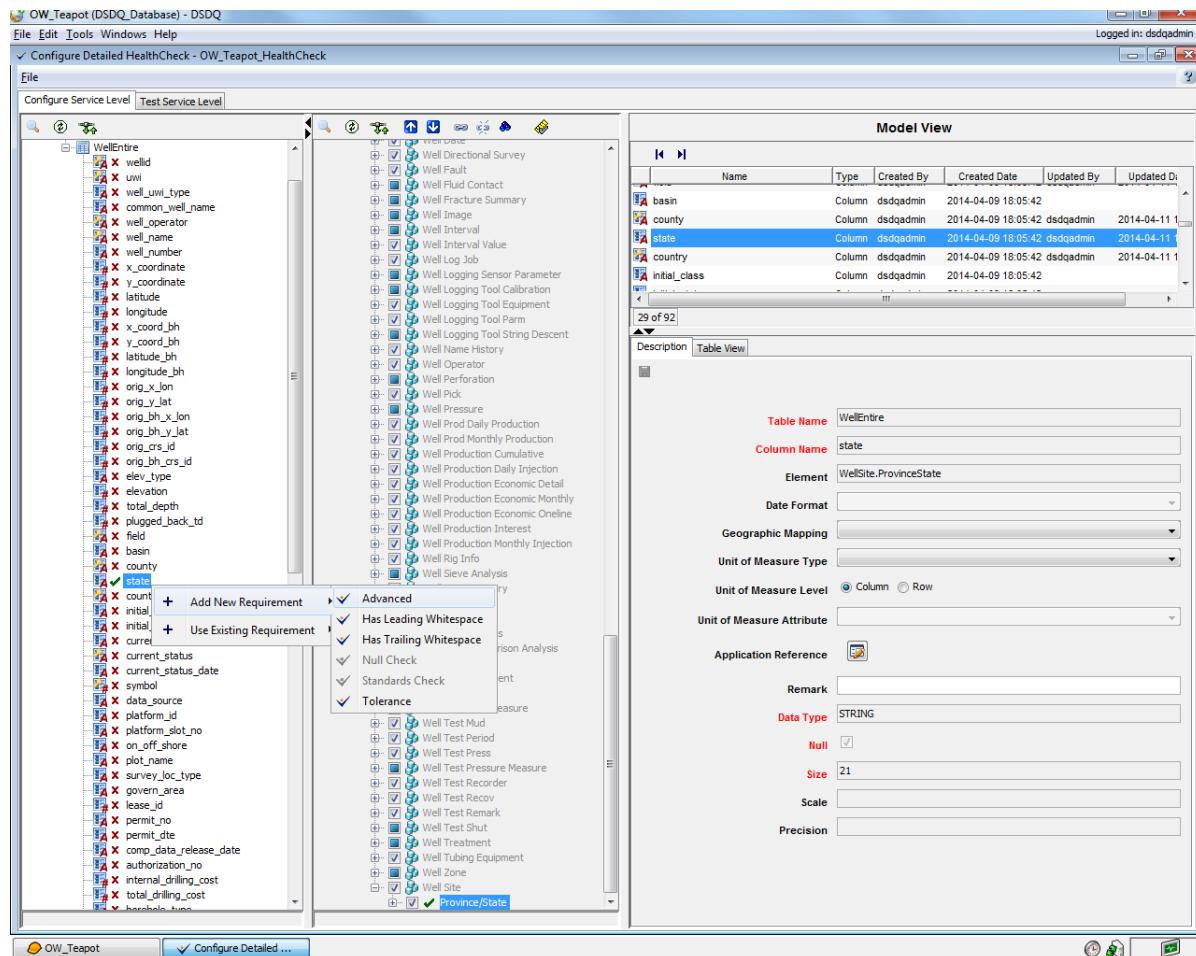
11. Click **OK**.

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.

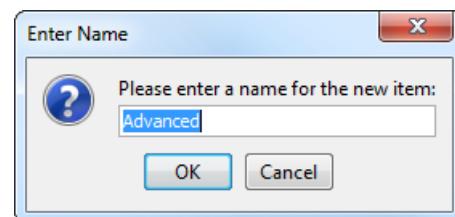
However, it is possible to link different tables' columns to the same element.

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

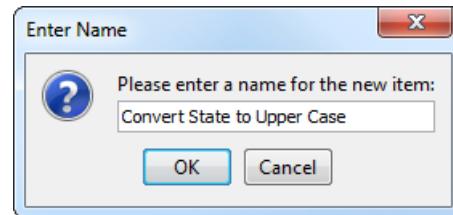
12. Right-click the **State** column in the Data Model Tree and select **Add New Requirement > Advanced** from the pop-up menu.



The **Enter Name** dialog box appears.



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



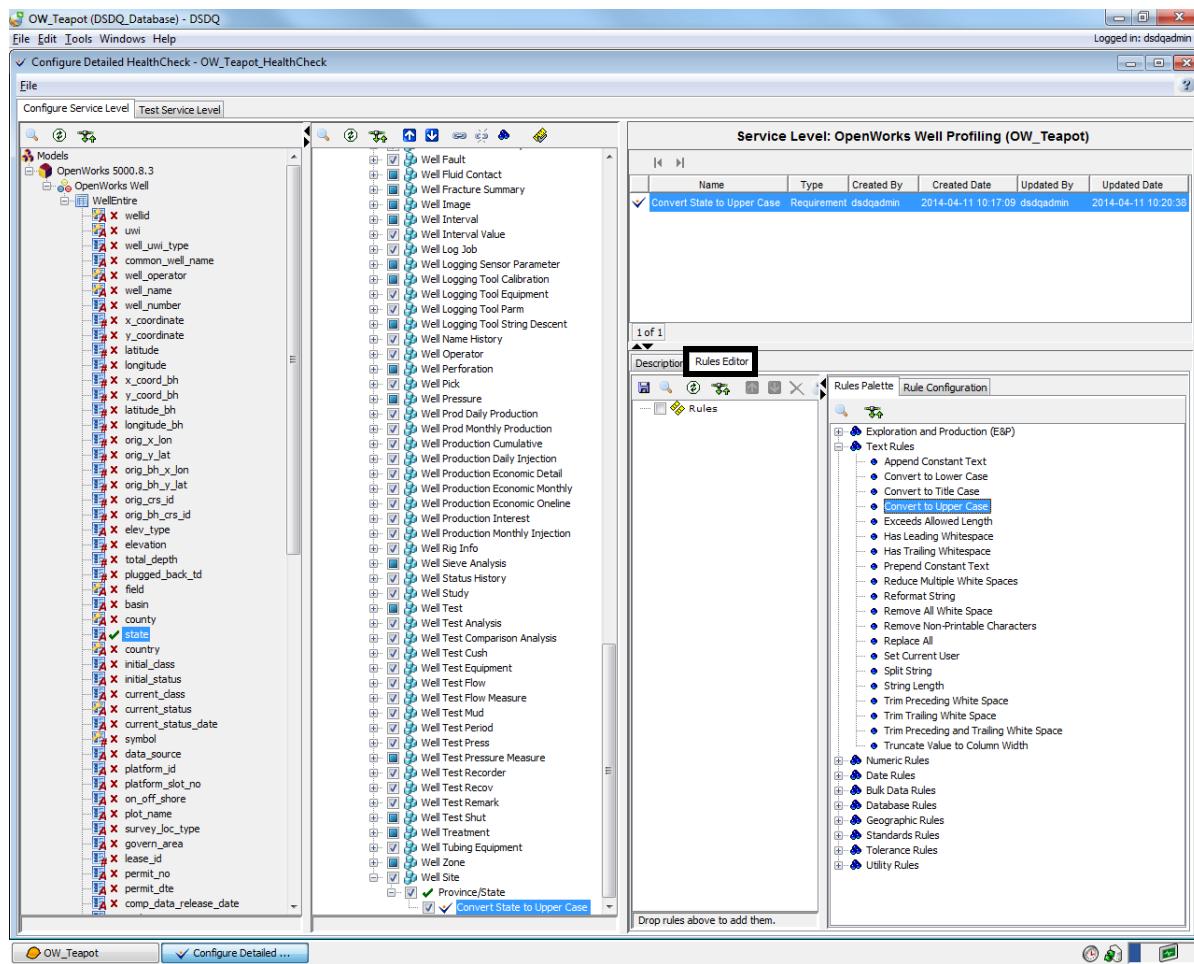
- 14. Click OK to add the requirement to the selected column. The Convert State to Upper Case requirement is added to the State element.**

Name	Type	Created By	Created Date	Updated By	Updated Date
Convert State to Upper Case	Requirement	dsdqadmin	2014-04-11 10:17:09	dsdqadmin	2014-04-11 10:18:53

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

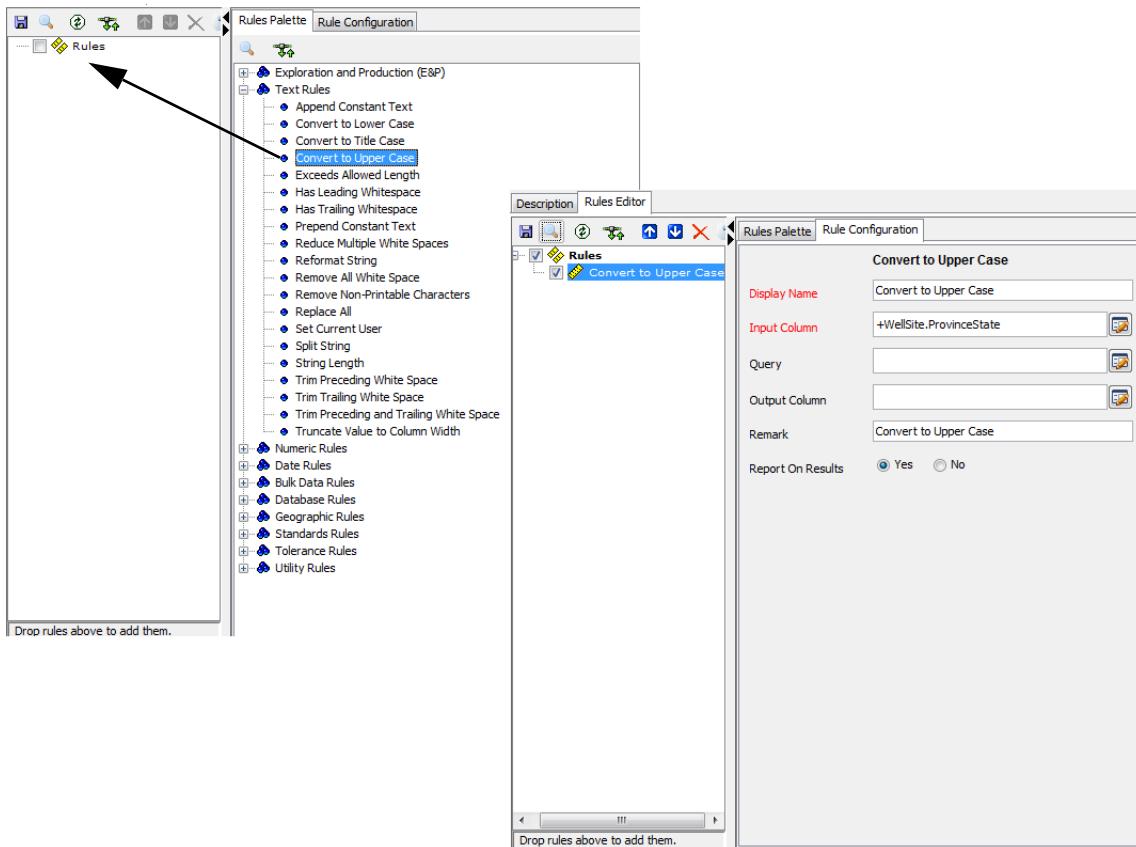
- 16. The Requirement name field populates with the name assigned in step 13.**

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.

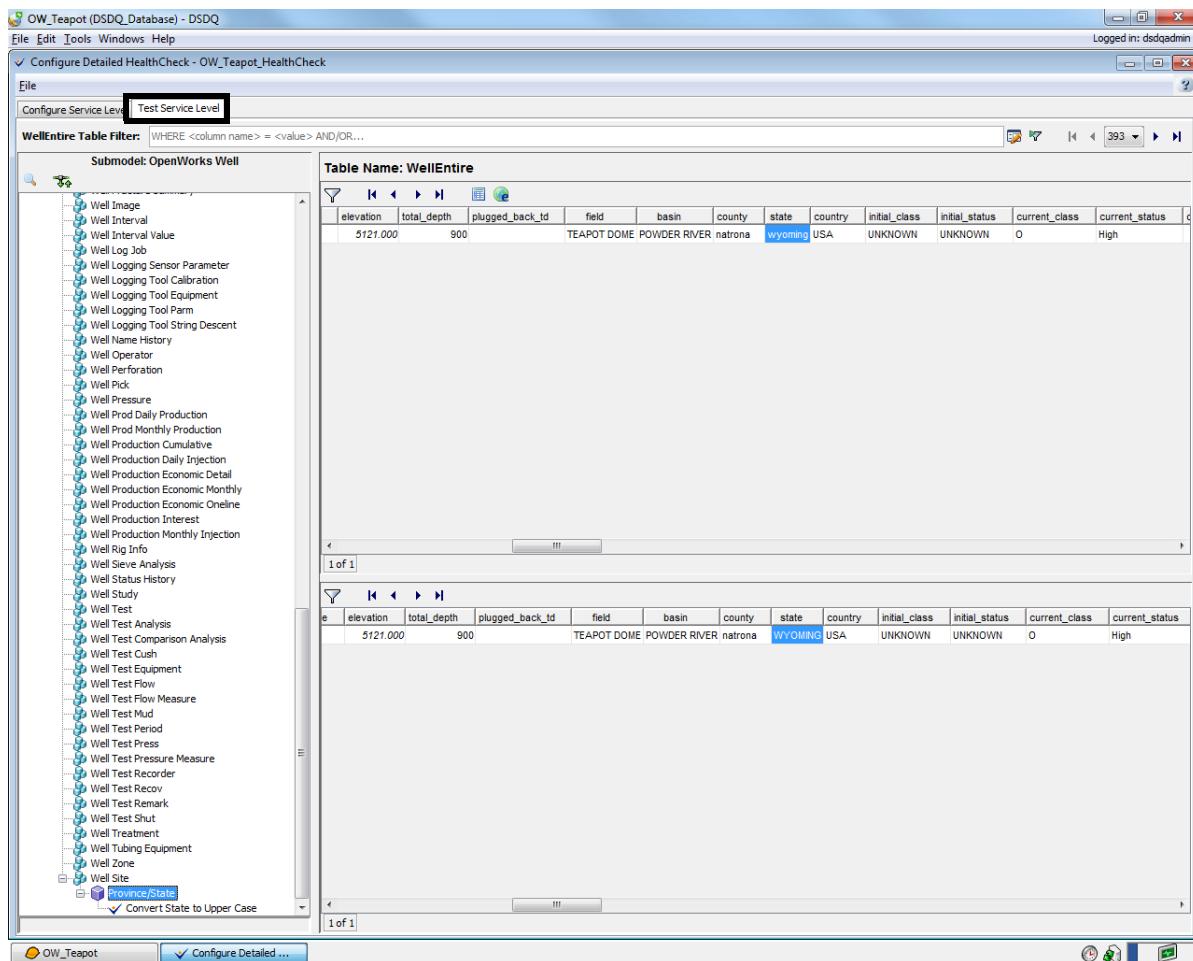
22. Drag and drop the **Convert to Upper Case** rule onto the **Rule** area.



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**.

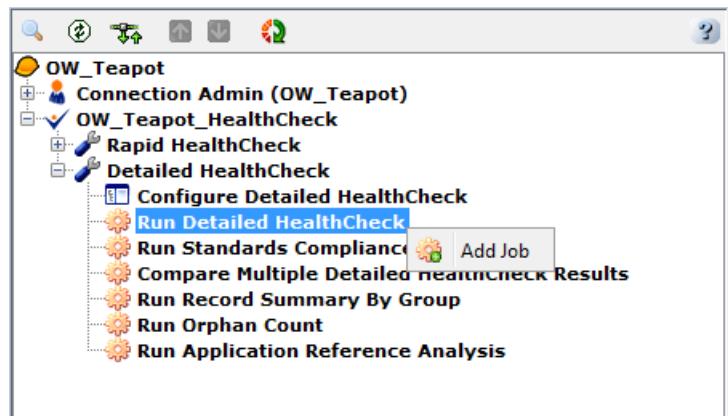
25. Click the **Next Data Set**  button to test the next record.
26. Repeat step 25 to test all records.
27. Select **File > Exit** to close the **Configure Detailed HealthCheck** window.

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

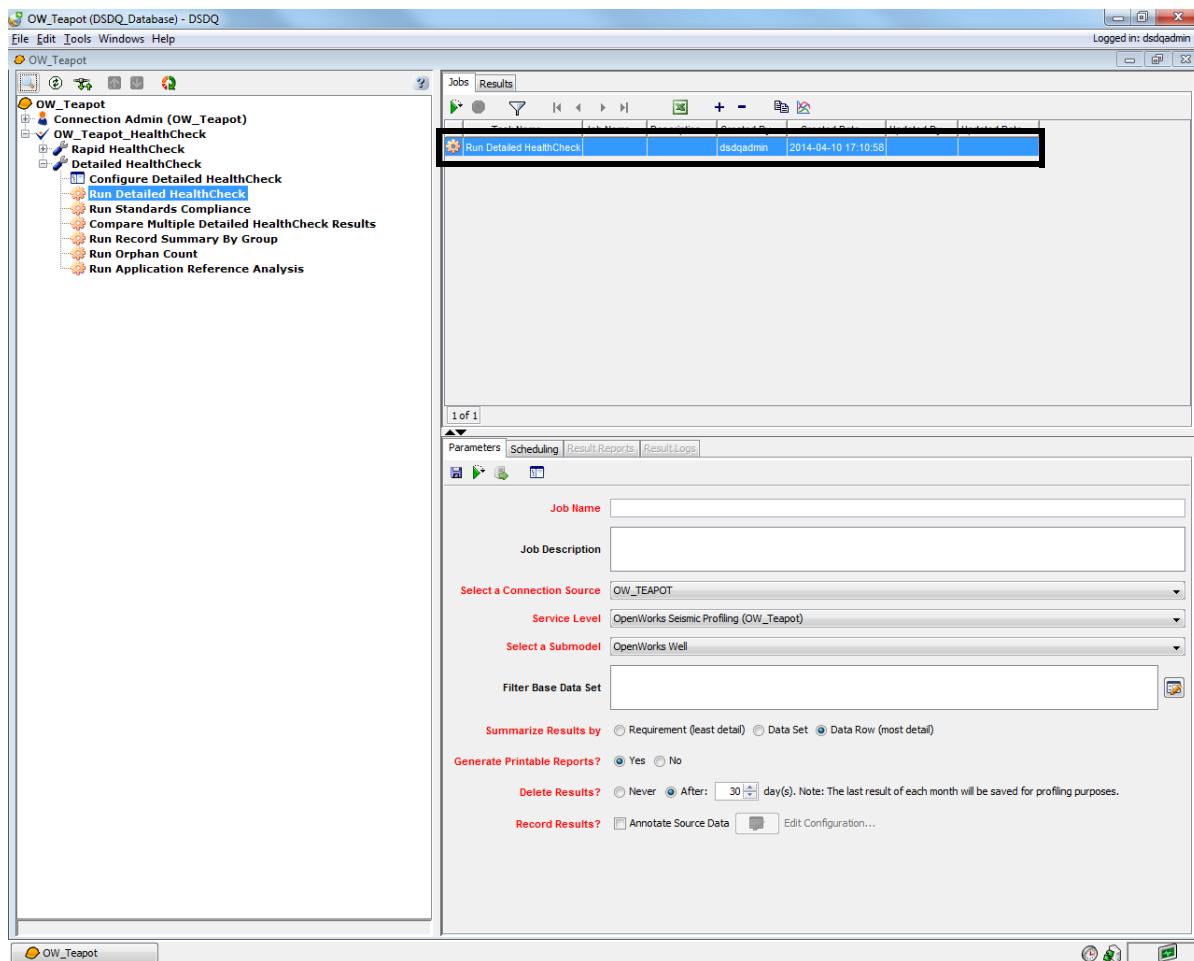
submodel table have been assigned elements from the desired service level. To run the Detailed HealthCheck task:

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.



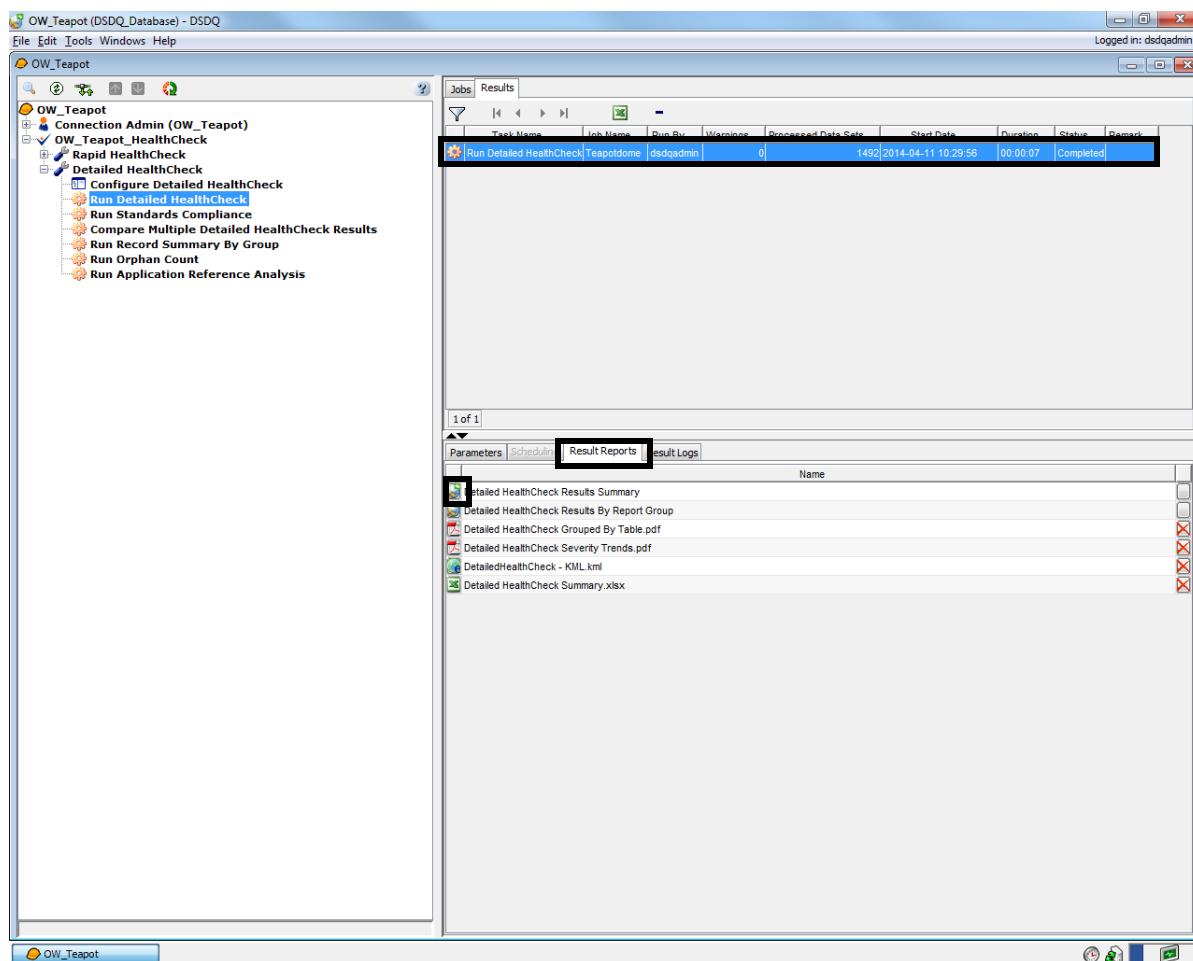
A new job is initiated and it displays on the **Job and Results**

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

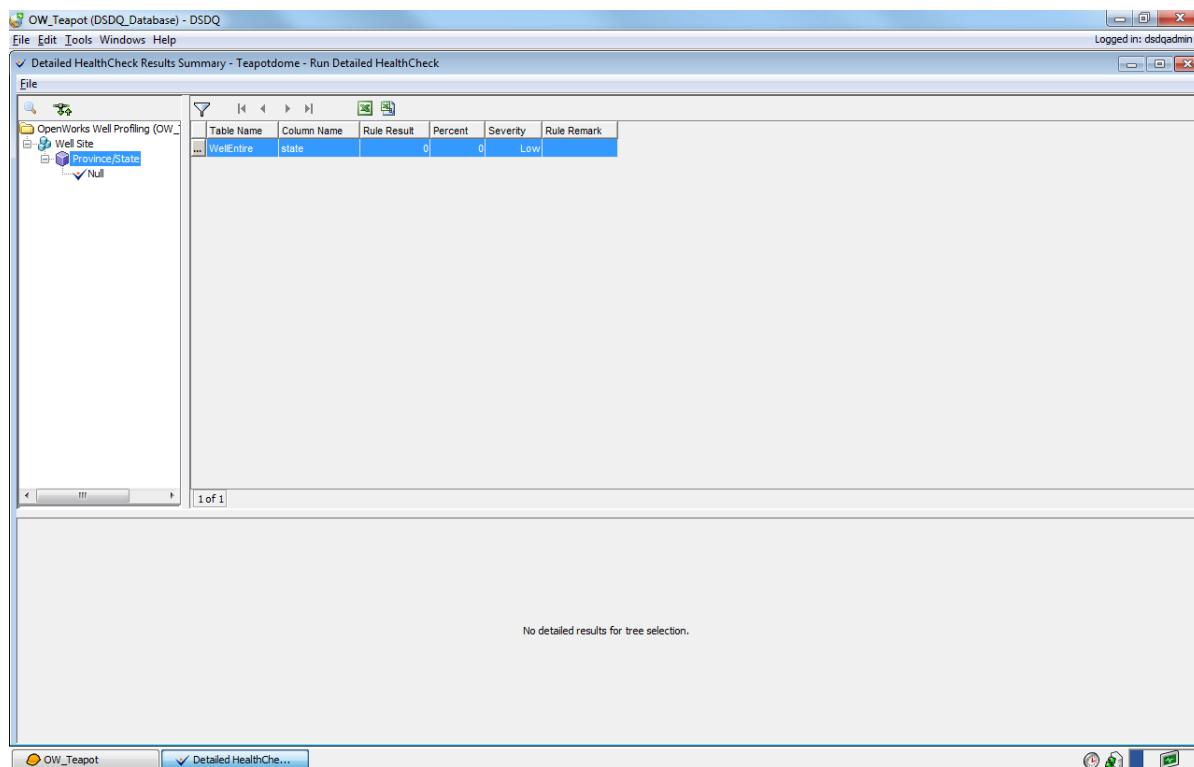


2. Enter **Teapotdome** in the **Job Name** field.
3. Enter **Detailed HealthCheck** for **OW_Teapot** in the **Job Description** field.
4. Select **OW_TEAPOT** from the **Select a Connection Source** drop-down list.
5. Select **OpenWorks Well Profiling (OW_Teapot)** from the **Service Level** drop-down 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.



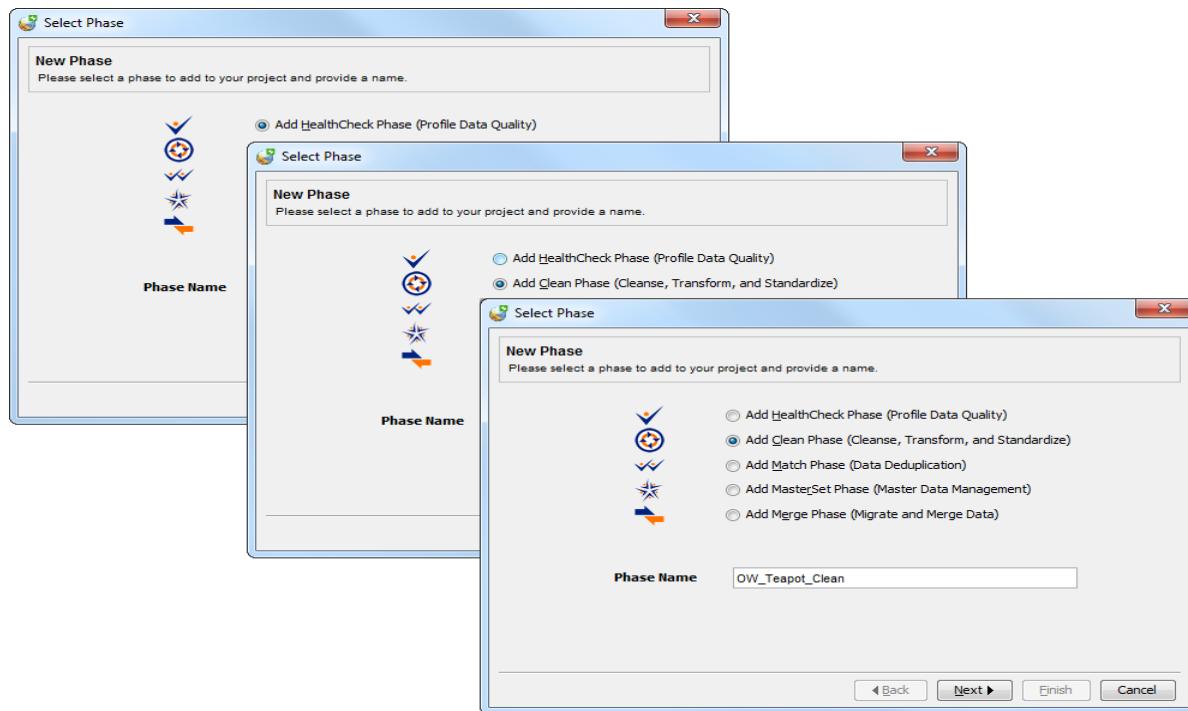
16. Select **File > Exit** to close the **Basic View Frame** window.

Resolving Data Quality Issues using the Clean Phase

Adding a Clean Phase

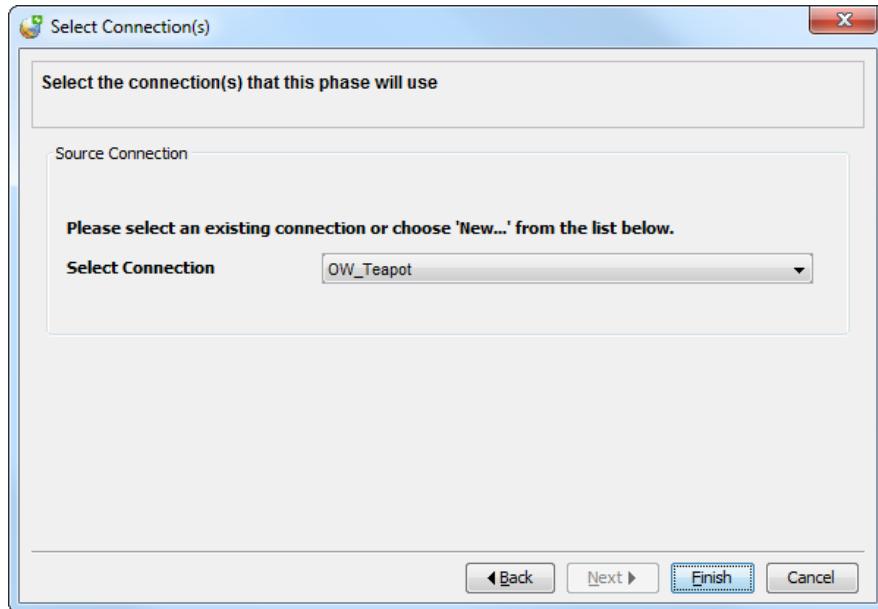
To add a Clean Phase:

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.

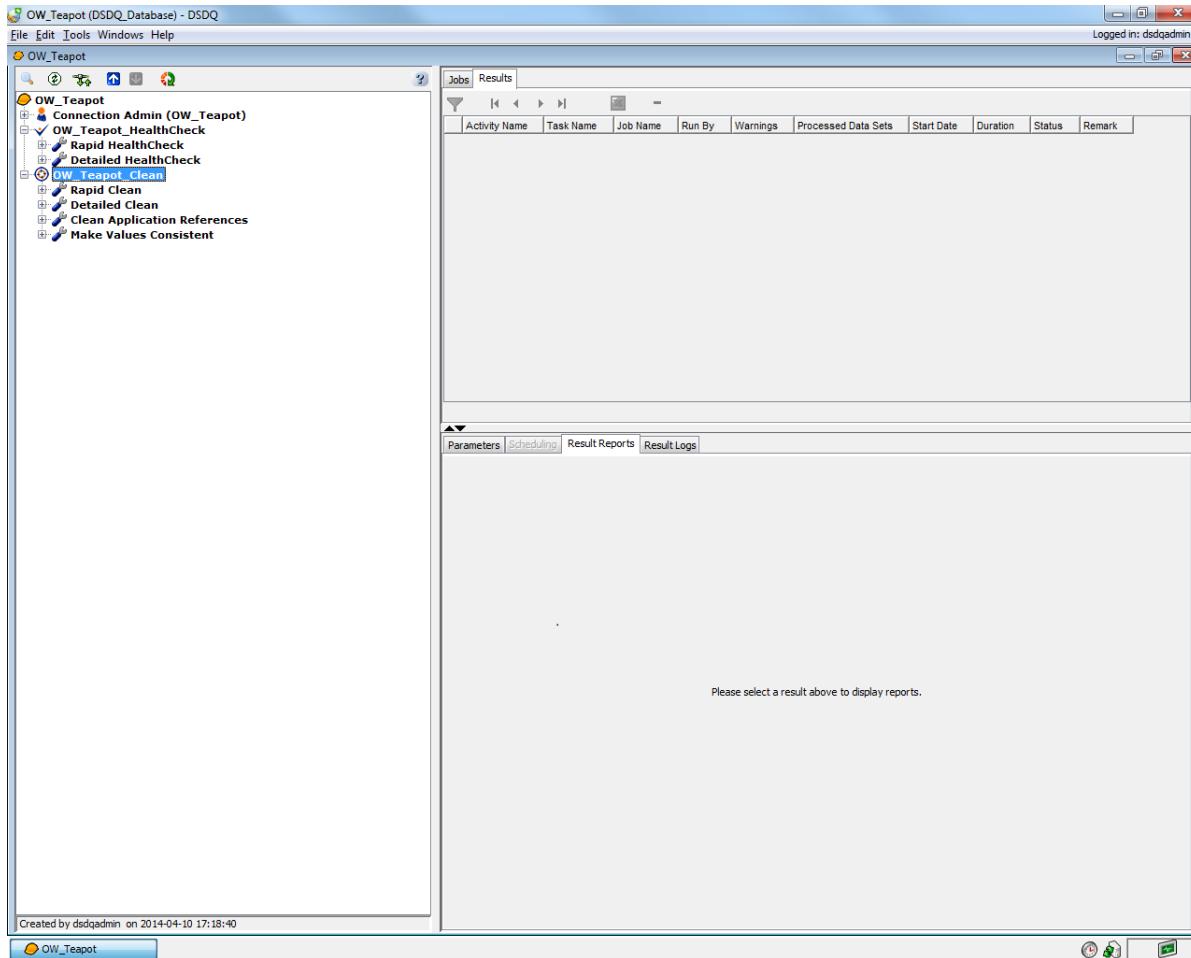
4. Click **Next** to continue.
The **Select Connection(s)** window appears.



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.



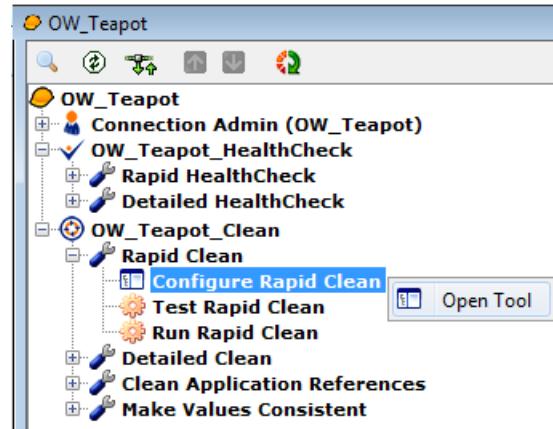
Rapid Clean Activity

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

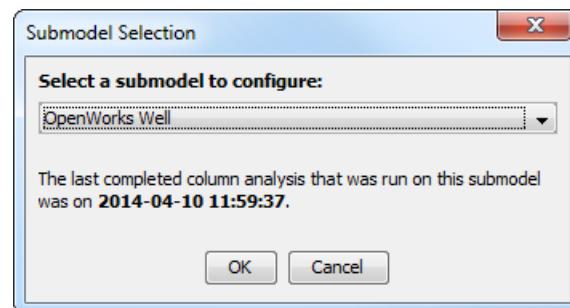
Exercise: Configuring the Rapid Clean Tool

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.



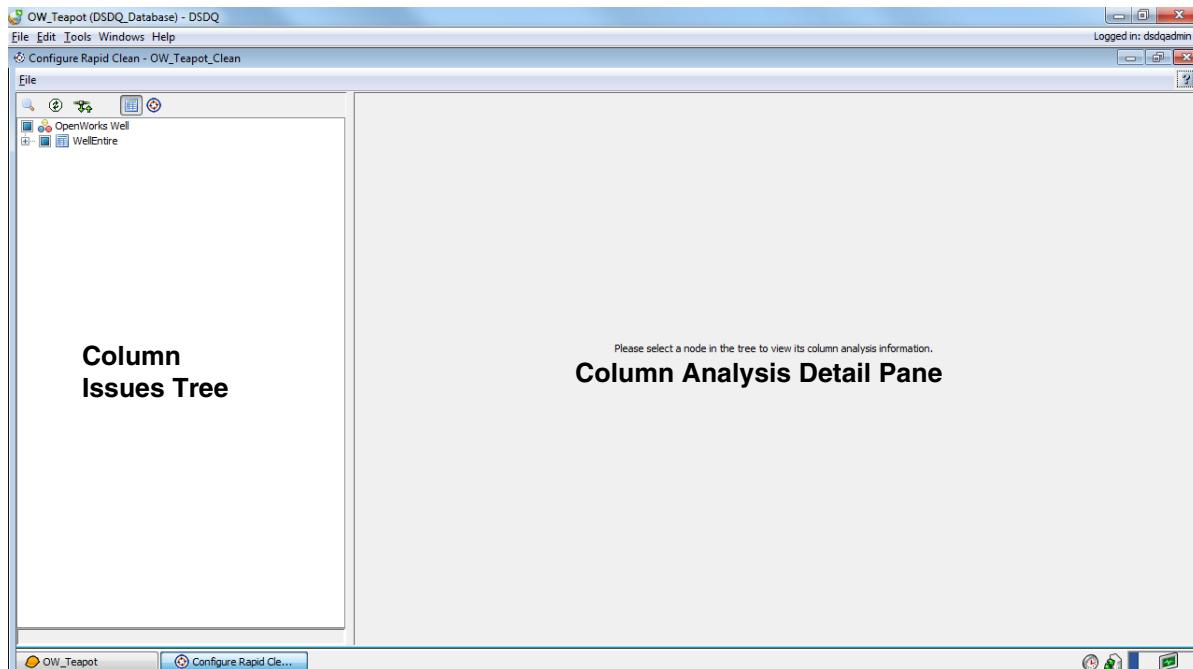
The **Submodel Selection** dialog box appears.



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.

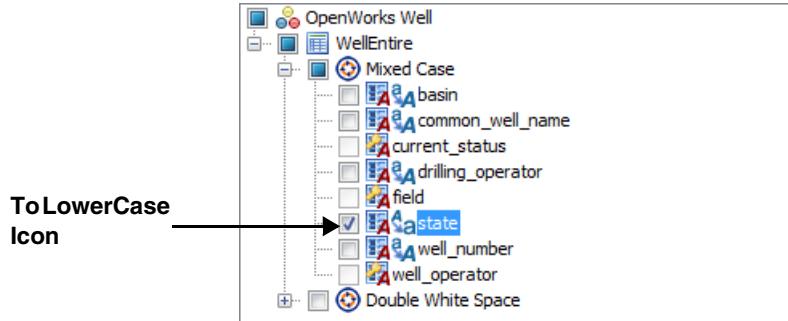
- Click the **State** column.

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

Table 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	0
WellEntire	current_status	STRING	1492	1492	100	2	1492	0	0	0	0
WellEntire	drilling_operator	STRING	1492	1492	100	33	49	0	0	0	0
WellEntire	field	STRING	1492	1492	100	9	50	0	0	0	0
WellEntire	remark	STRING	1308	1308	87	41	0	0	0	0	0
WellEntire	state	STRING	1492	1492	100	2	1492	0	0	0	0
WellEntire	well_number	STRING	1380	92	1254	131	0	0	0	0	0
WellEntire	wel_operator	STRING	1492	1492	100	22	114	0	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.

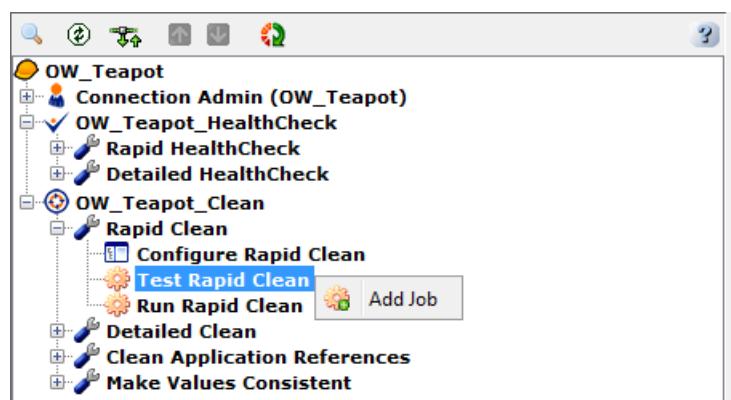


10. Select **File > Exit** to close the **Configure Rapid Clean** window.

Exercise: Running the Test Rapid Clean Task

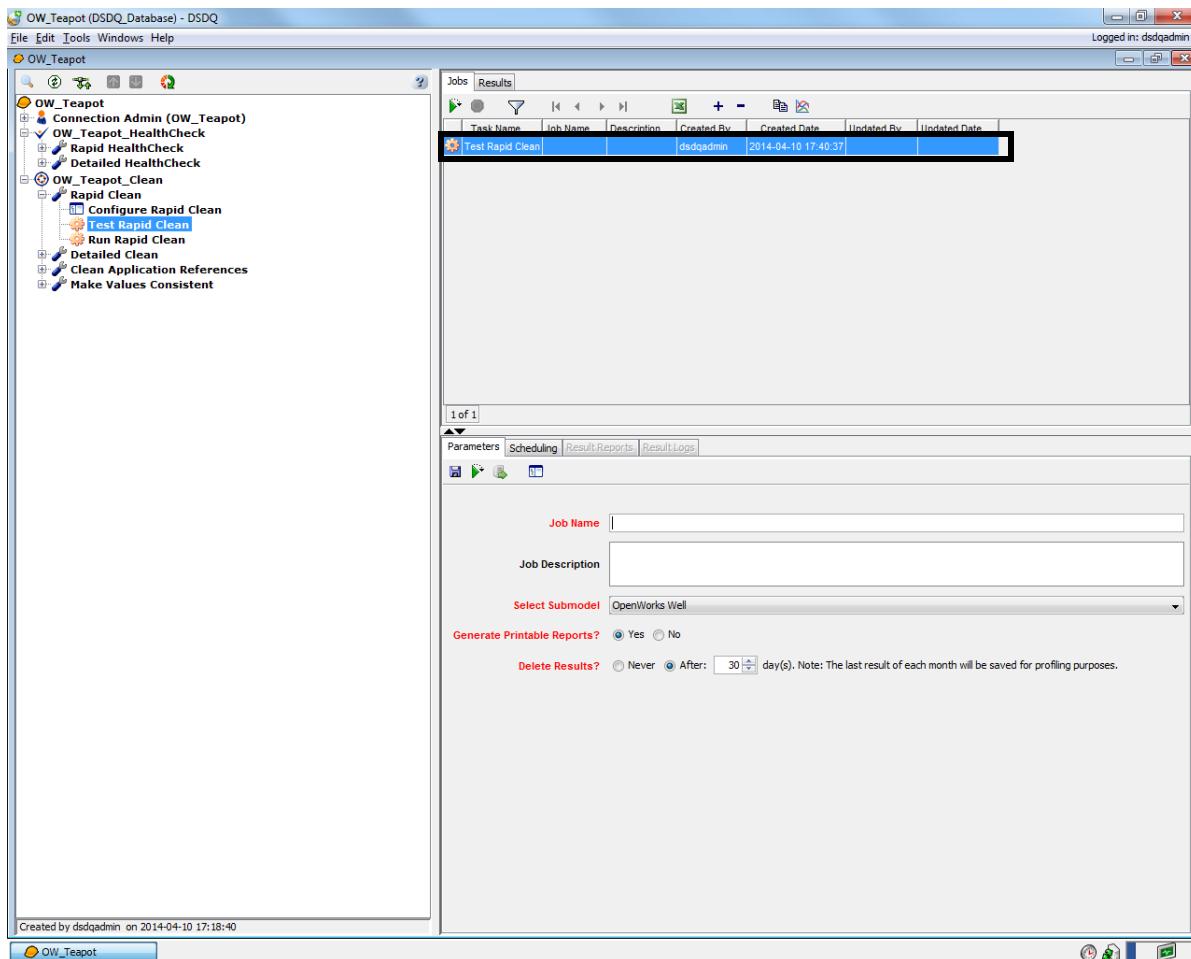
After issues have been selected to be cleaned using the **Configure Rapid Clean** Tool, the **Test Rapid Clean** Task is run to make sure that the expected results are seen before running the **Run Rapid Clean** Task to fix the entire dataset in the submodel. To run the **Test Rapid Clean** Task:

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.



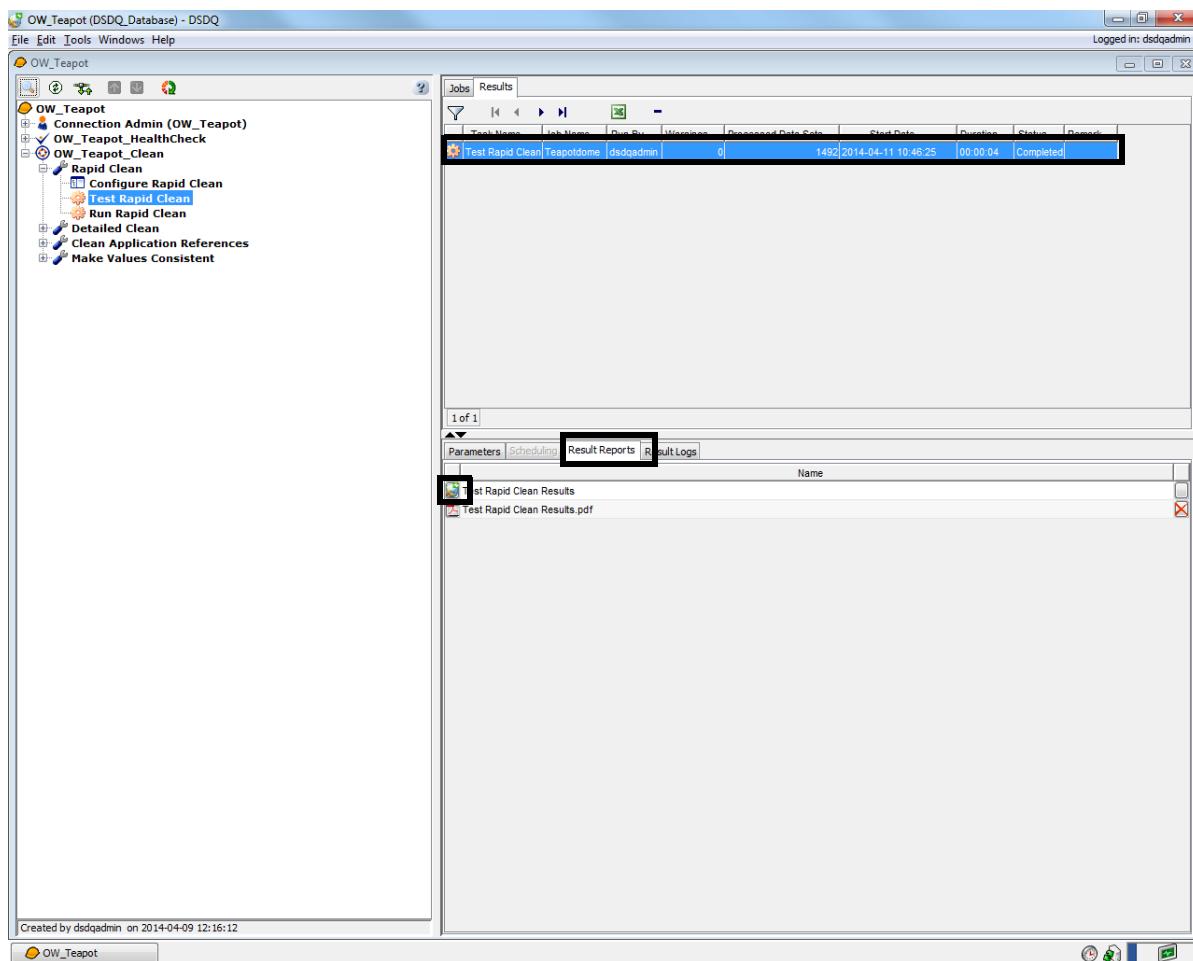
2. Enter **Teapotdome** in the **Job Name** field.
3. Enter **OW_Teapot State Name Conversion** in the **Job Description** field.
4. Select **OpenWorks Well** from the **Select Submodel** drop-down list.
5. Select the **Yes** option for **Generate Printable Reports**.
6. Select the **After** option for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

8. Click  to run the job.

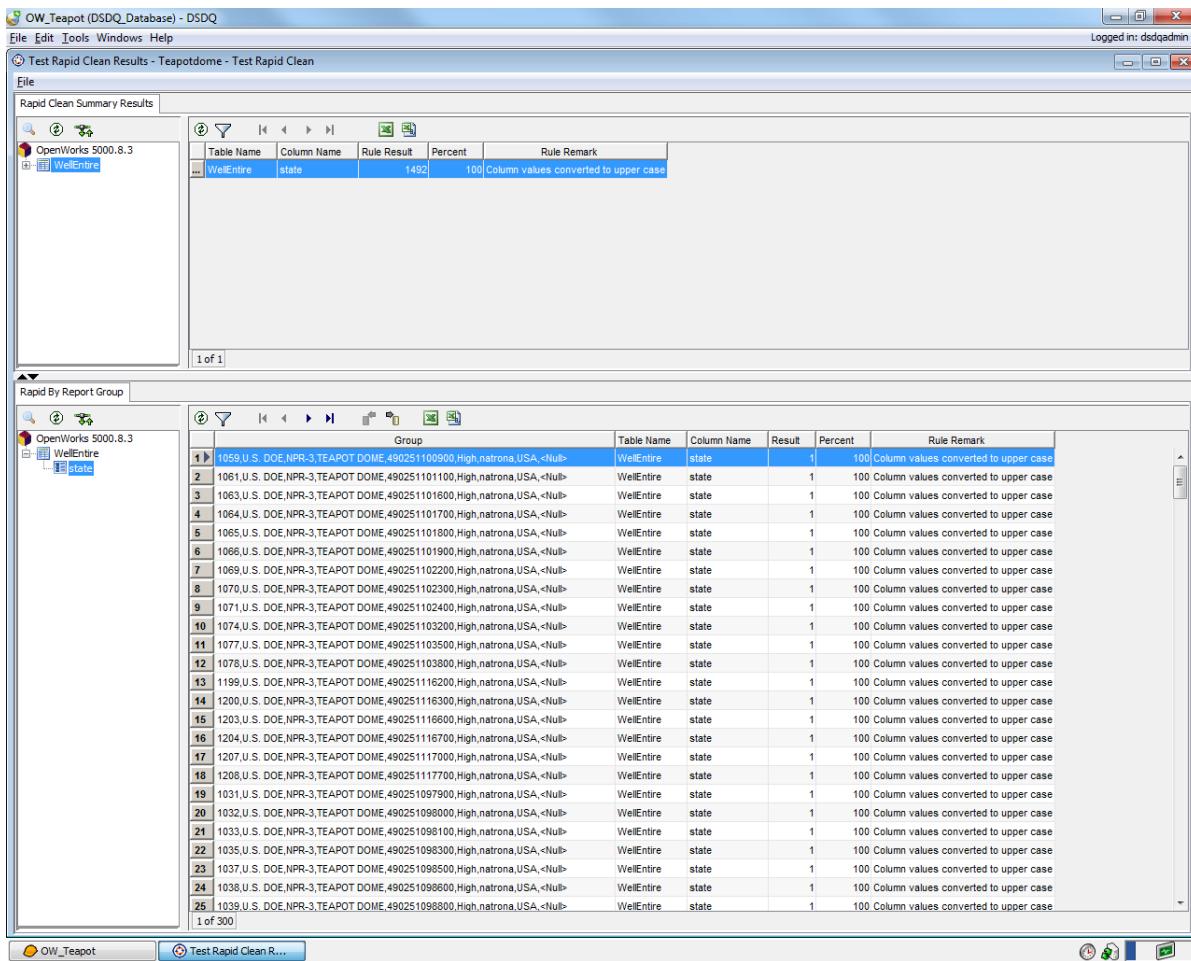
The **Test Rapid Clean** Task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

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.

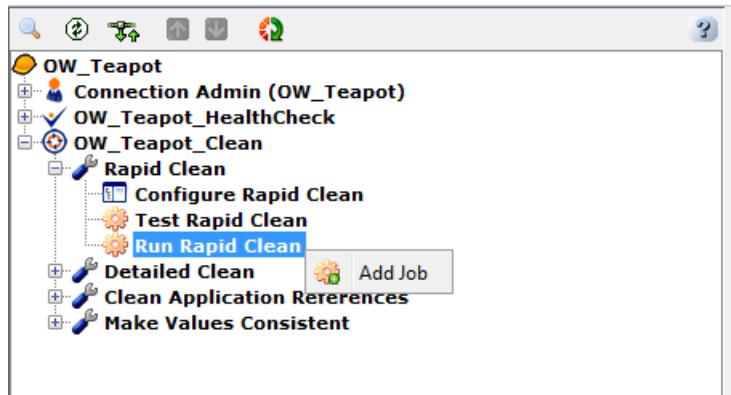


11. Select **File > Exit** to close the **Basic View Frame** window.

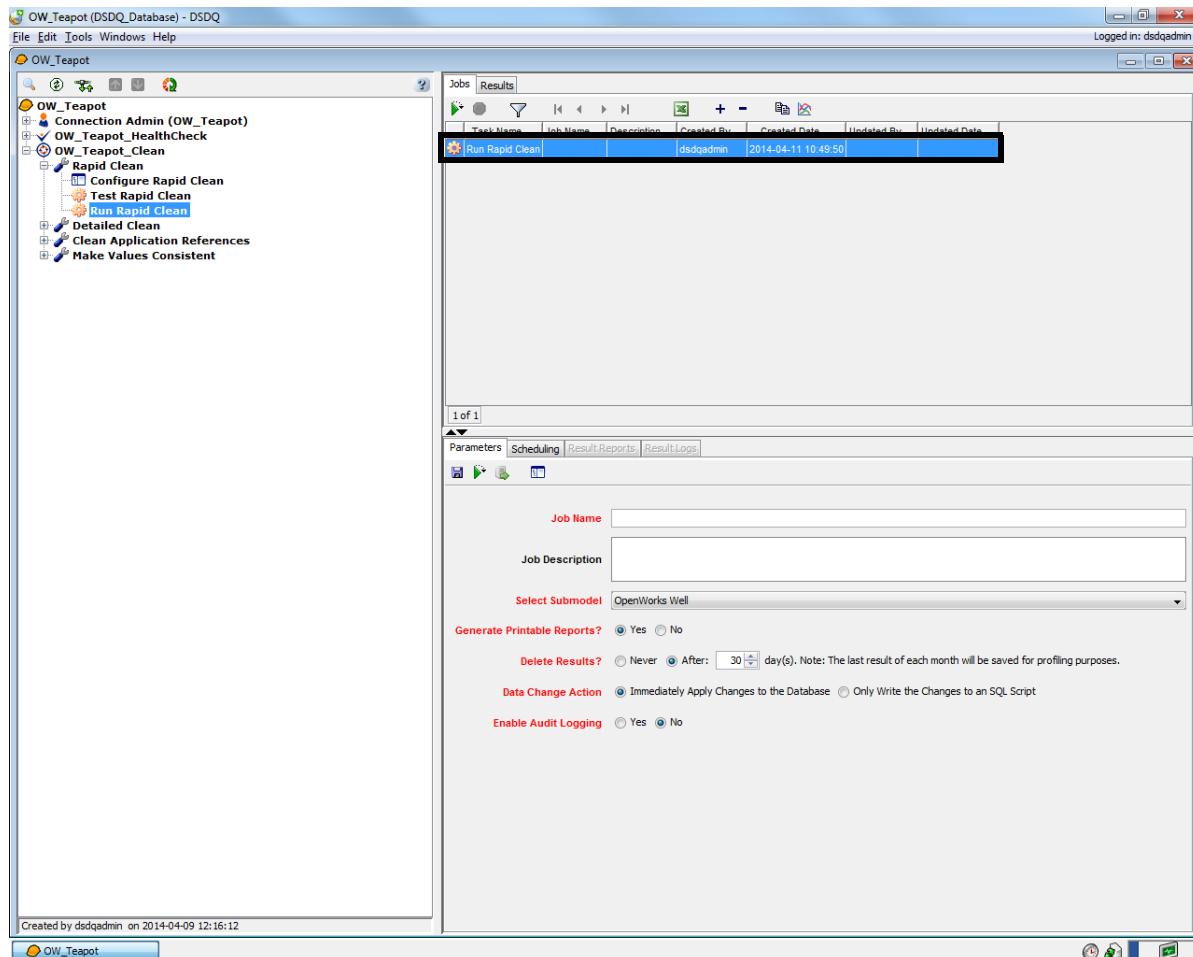
Exercise: Running the Rapid Clean Task

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**.



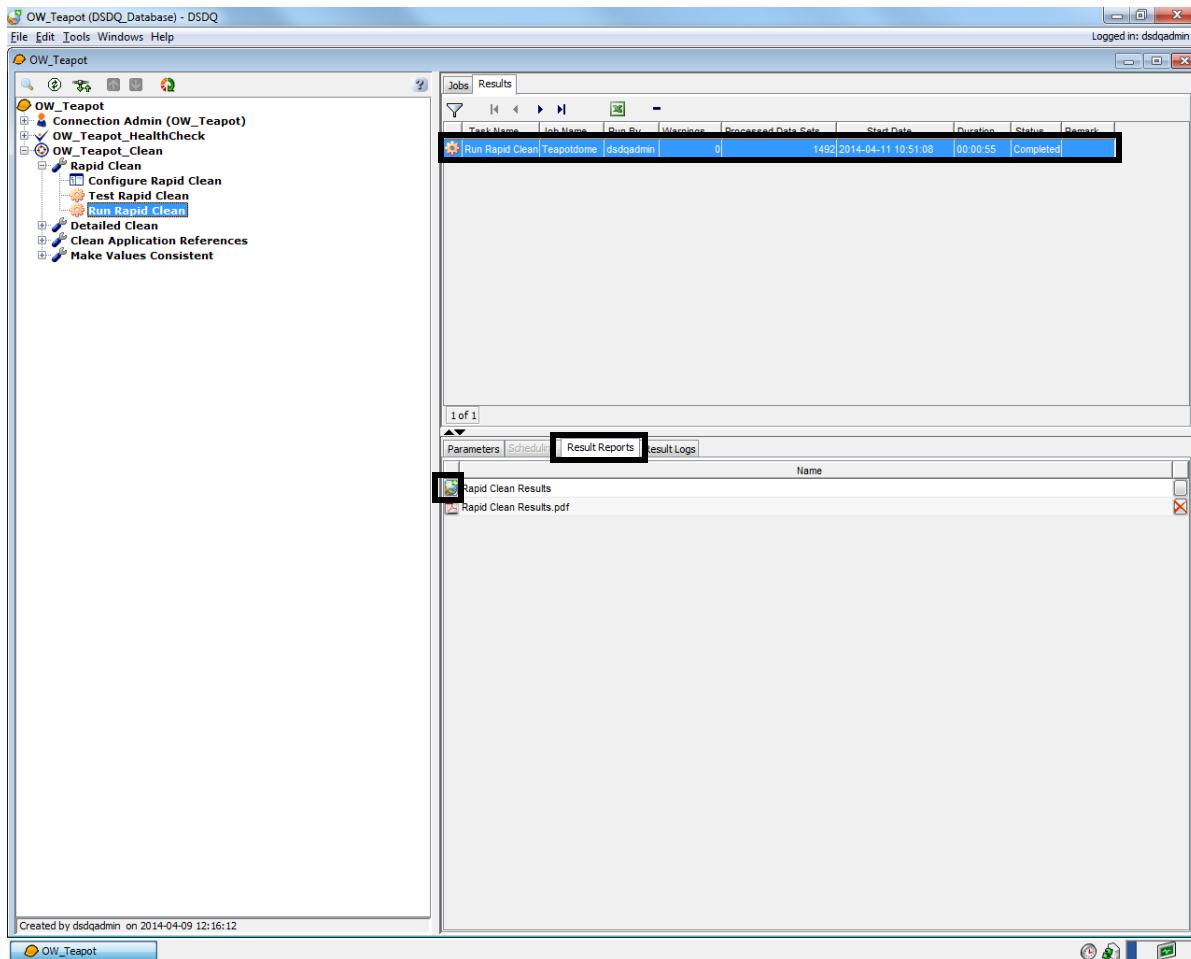
2. Enter **Teapotdome** in the **Job Name** field.
3. Enter **OW_Teapot Case Convert** in the **Job Description** field.

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.

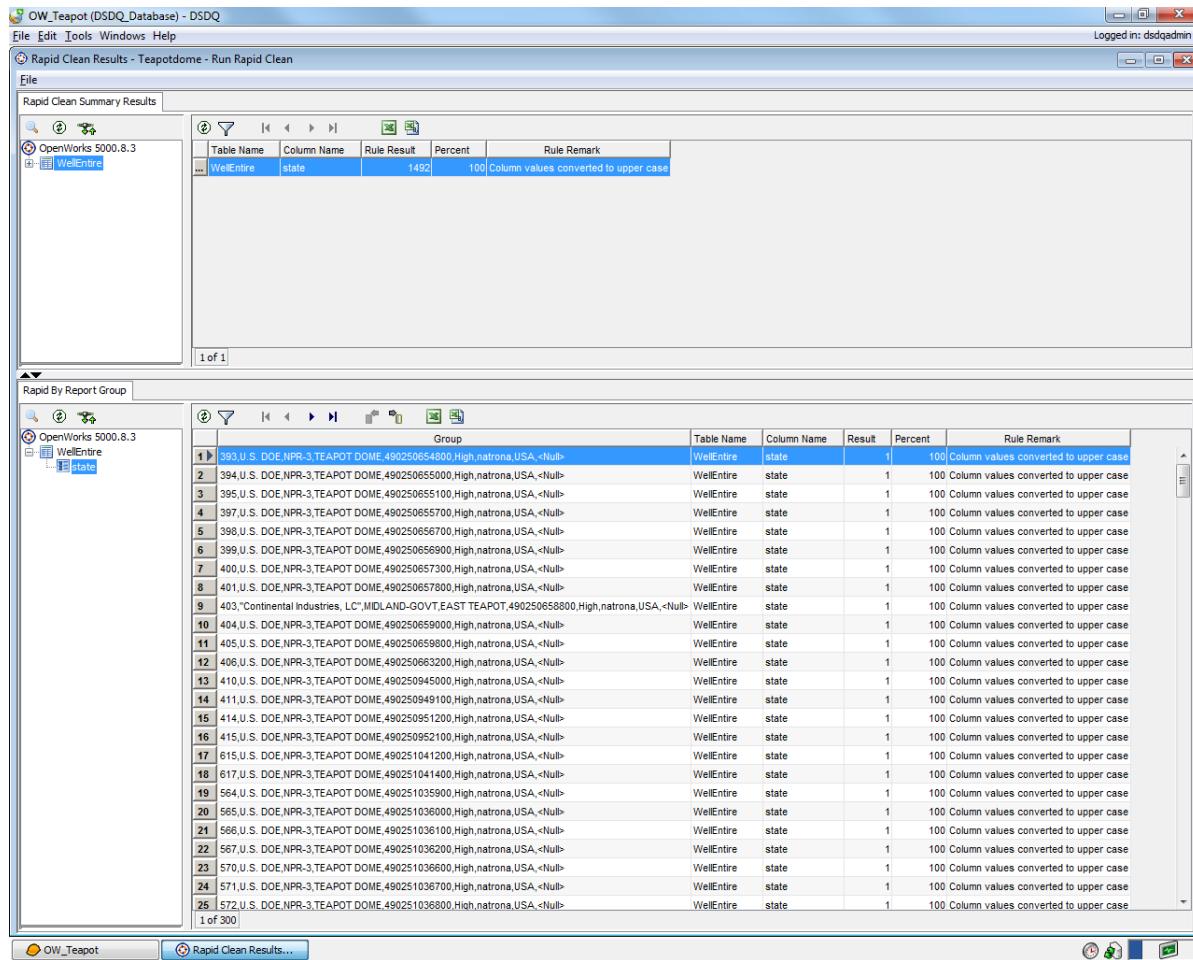
The **Run Rapid Clean Task** runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

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

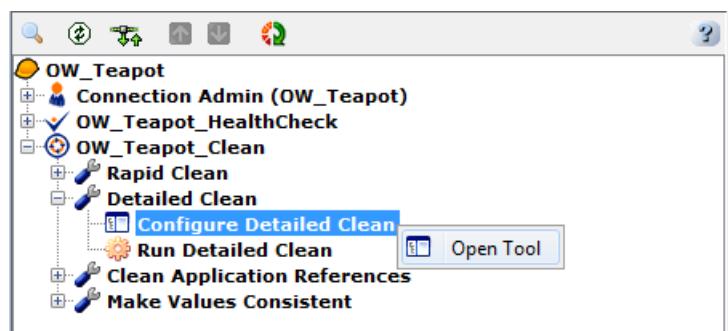
The **Detailed Clean** Activity helps in assigning columns to the clean requirements and testing service levels.

Exercise: Configuring the Detailed Clean Tool

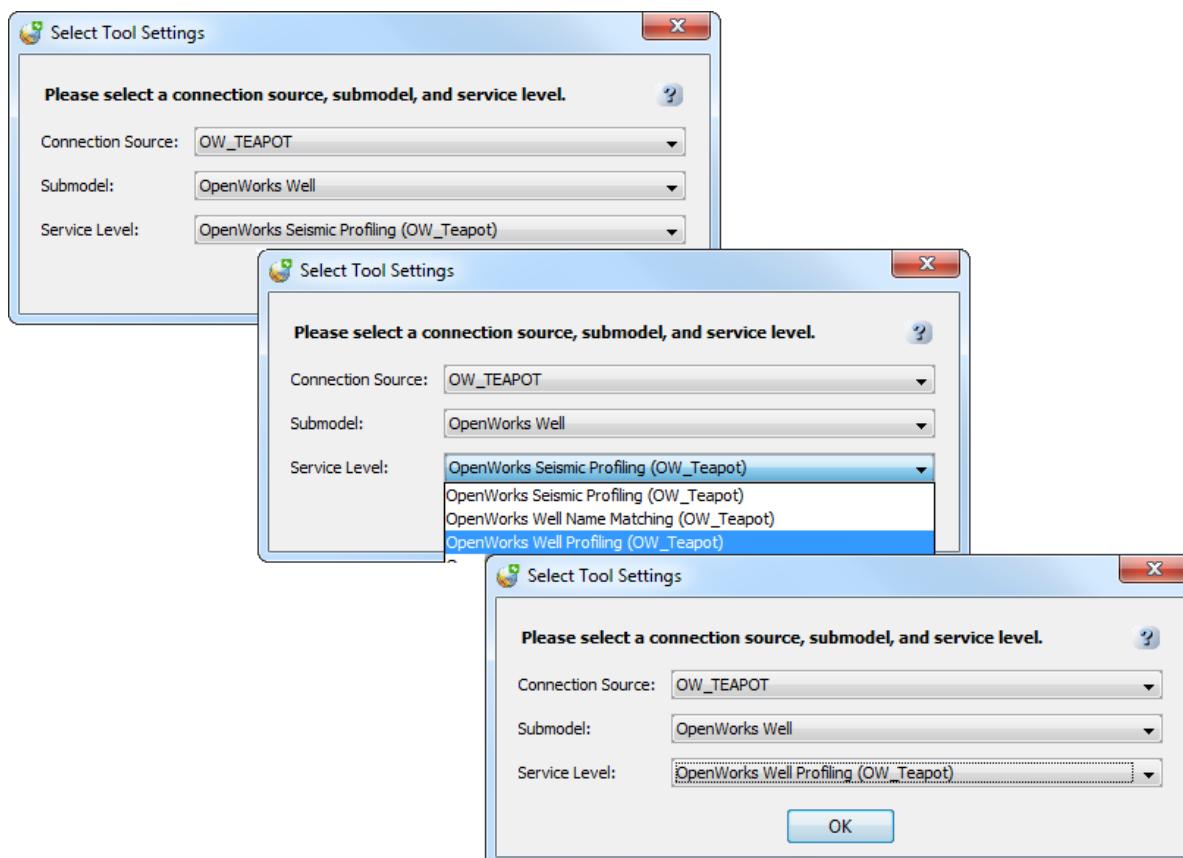
The **Configure Detailed Clean Tool** is used to configure service levels for testing prior to running the **Run Detailed Clean Task**. You can select which requirements in the service level to enable/disable, and when testing a service level, what subset of the total data to use. A

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.

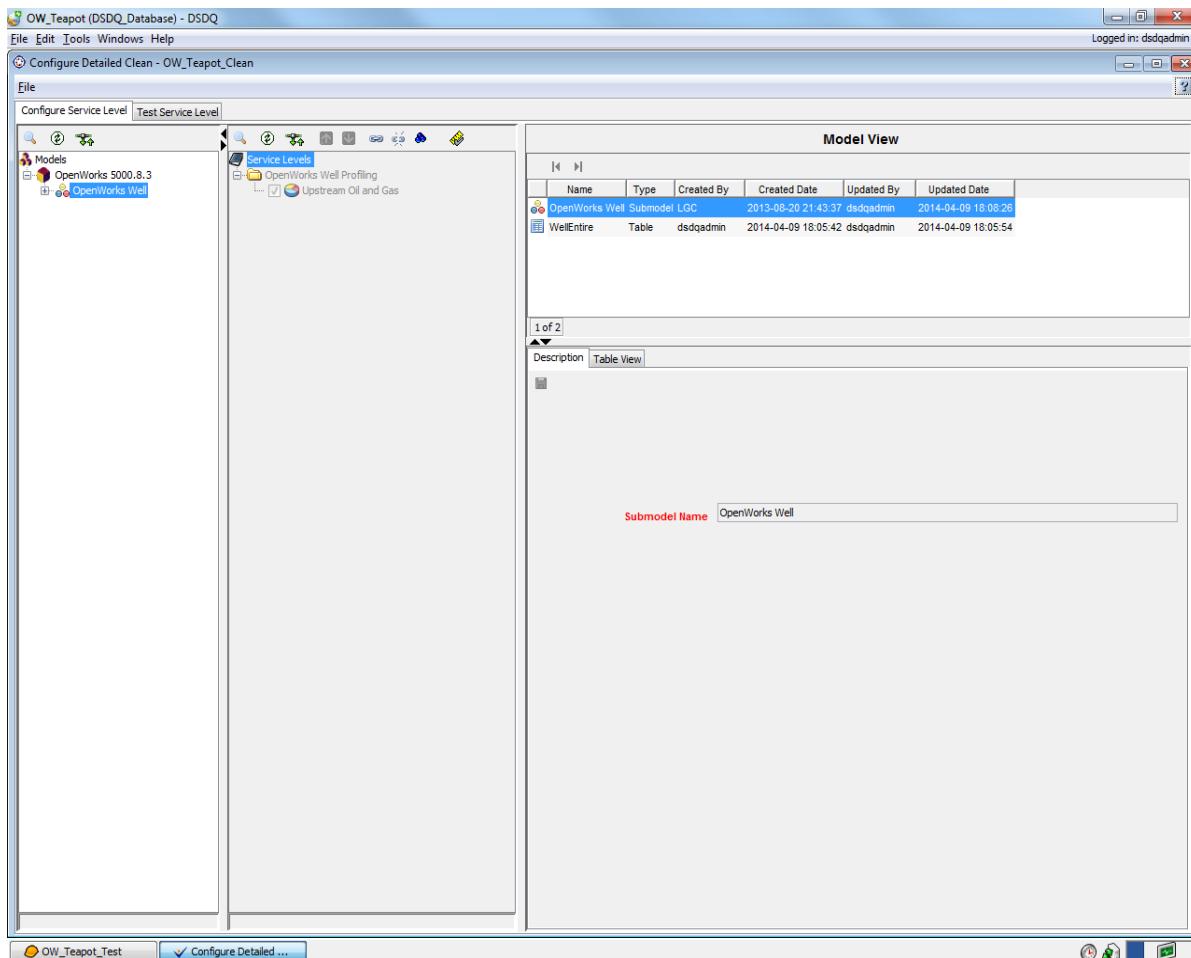


The **Select Tool Settings** window appears:



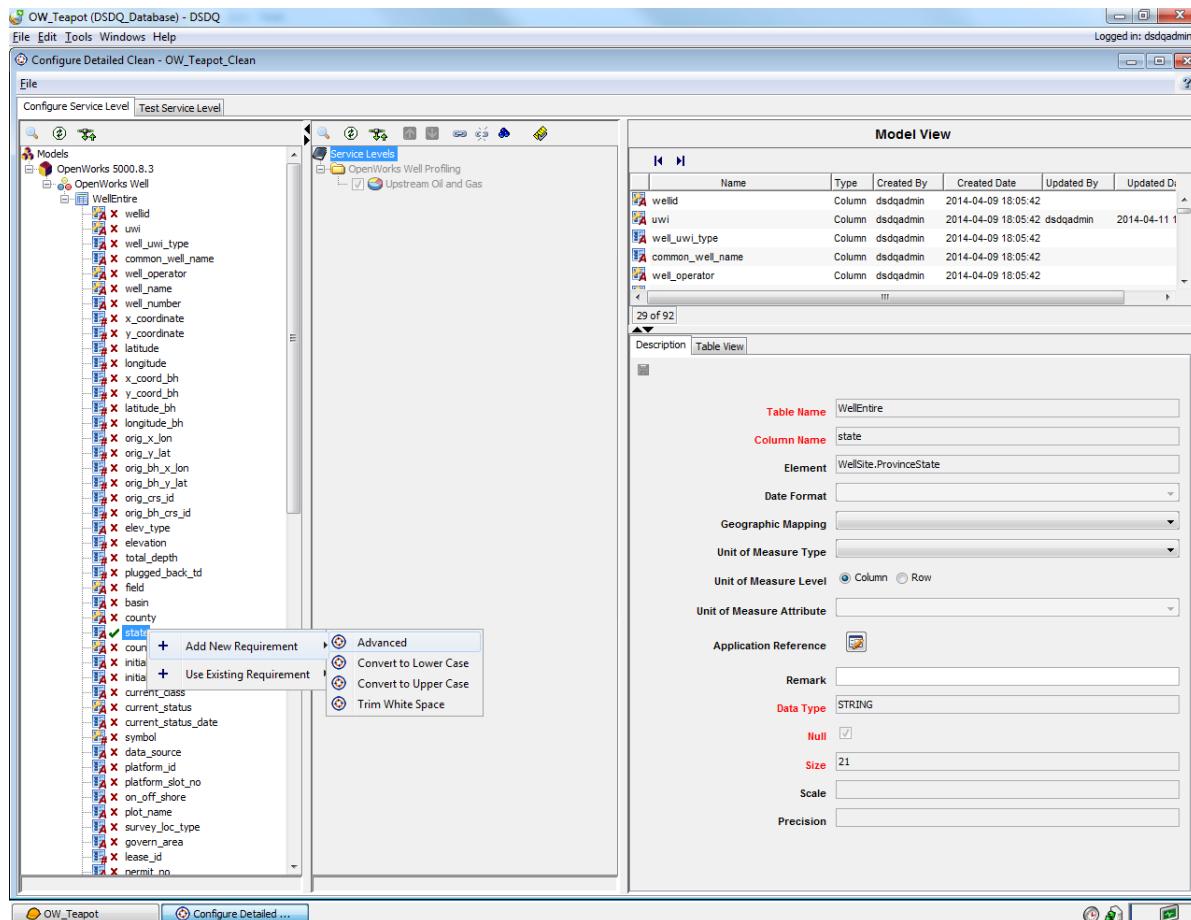
2. The **Connection Source** drop-down list is set to **OW_Teapot** by default.

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.

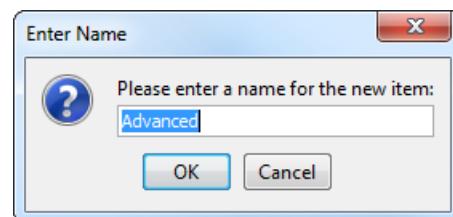


6. Click to expand the **OpenWorks Well** submodel in the Data Model Tree.
7. Click to expand the **WellEntire** table in the Data Model Tree.

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.

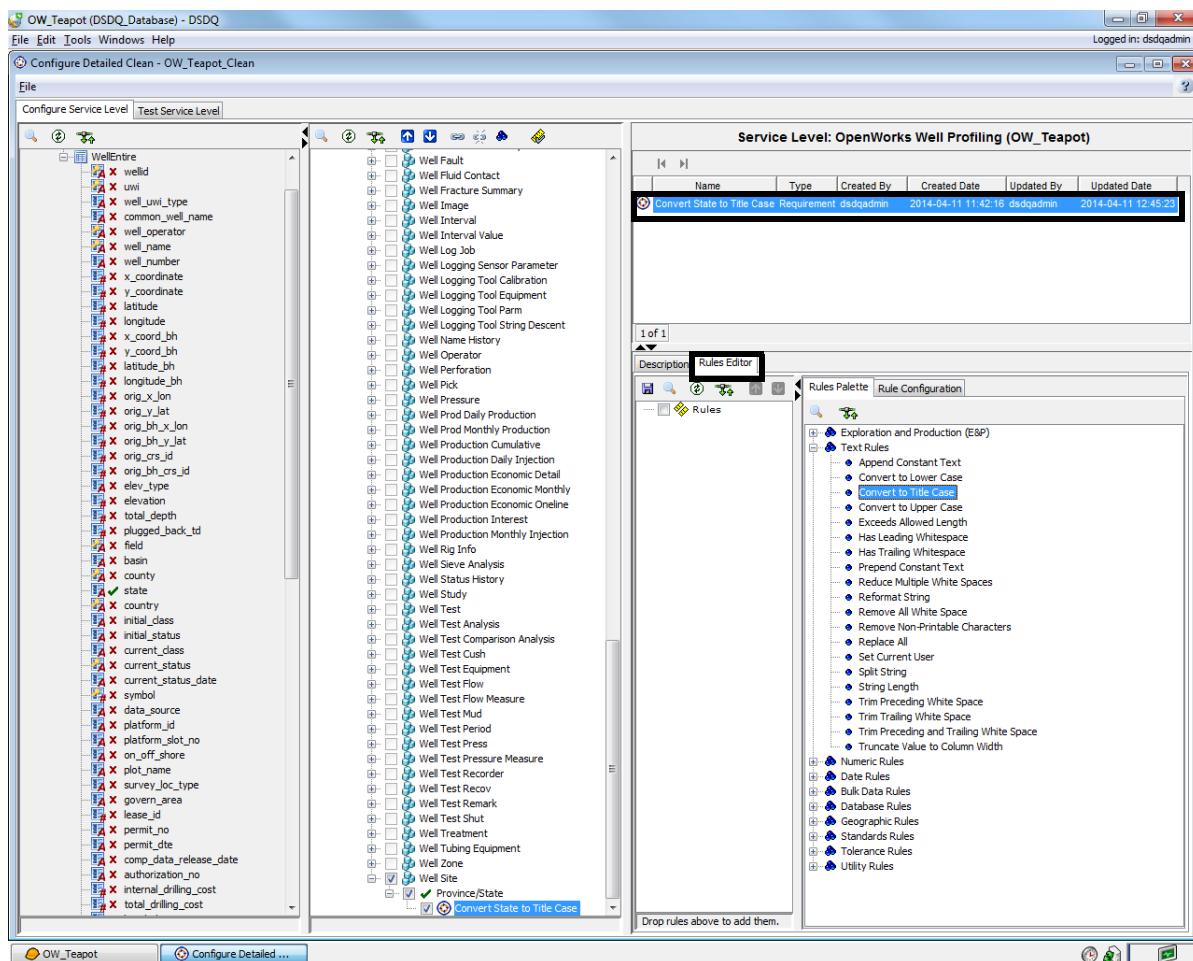


10. Click OK.

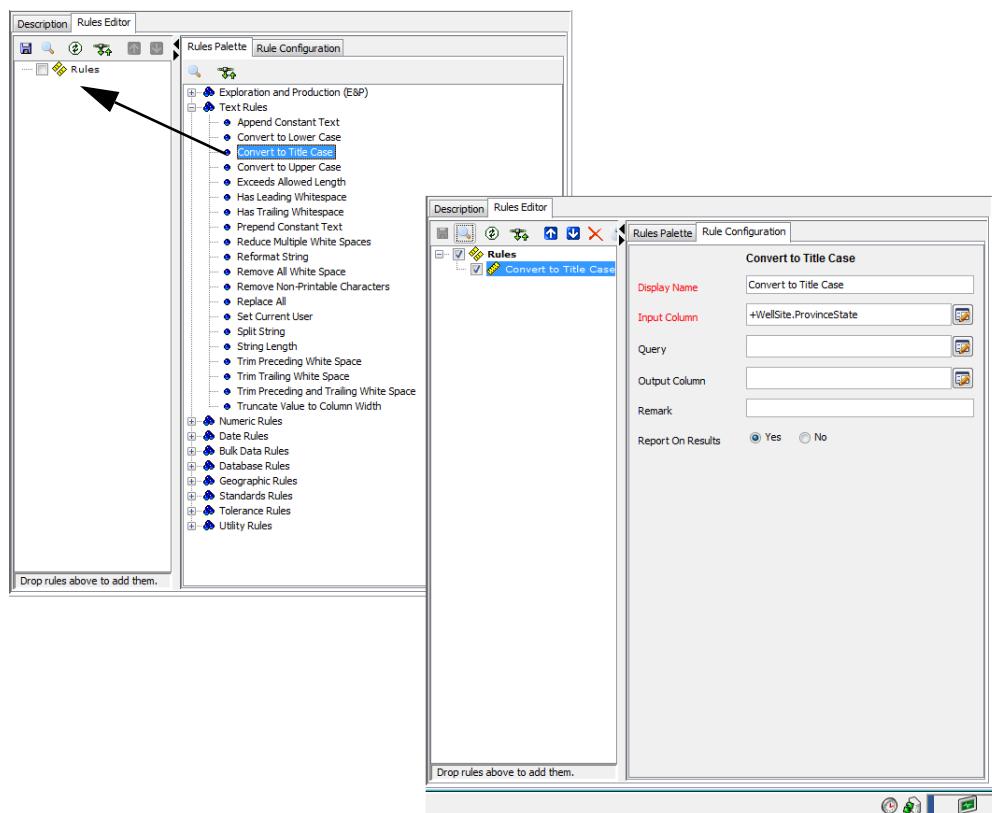
The **Convert State to Tile Case** requirement is added to the **State** column and displays in the **Model View Pane**.

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

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



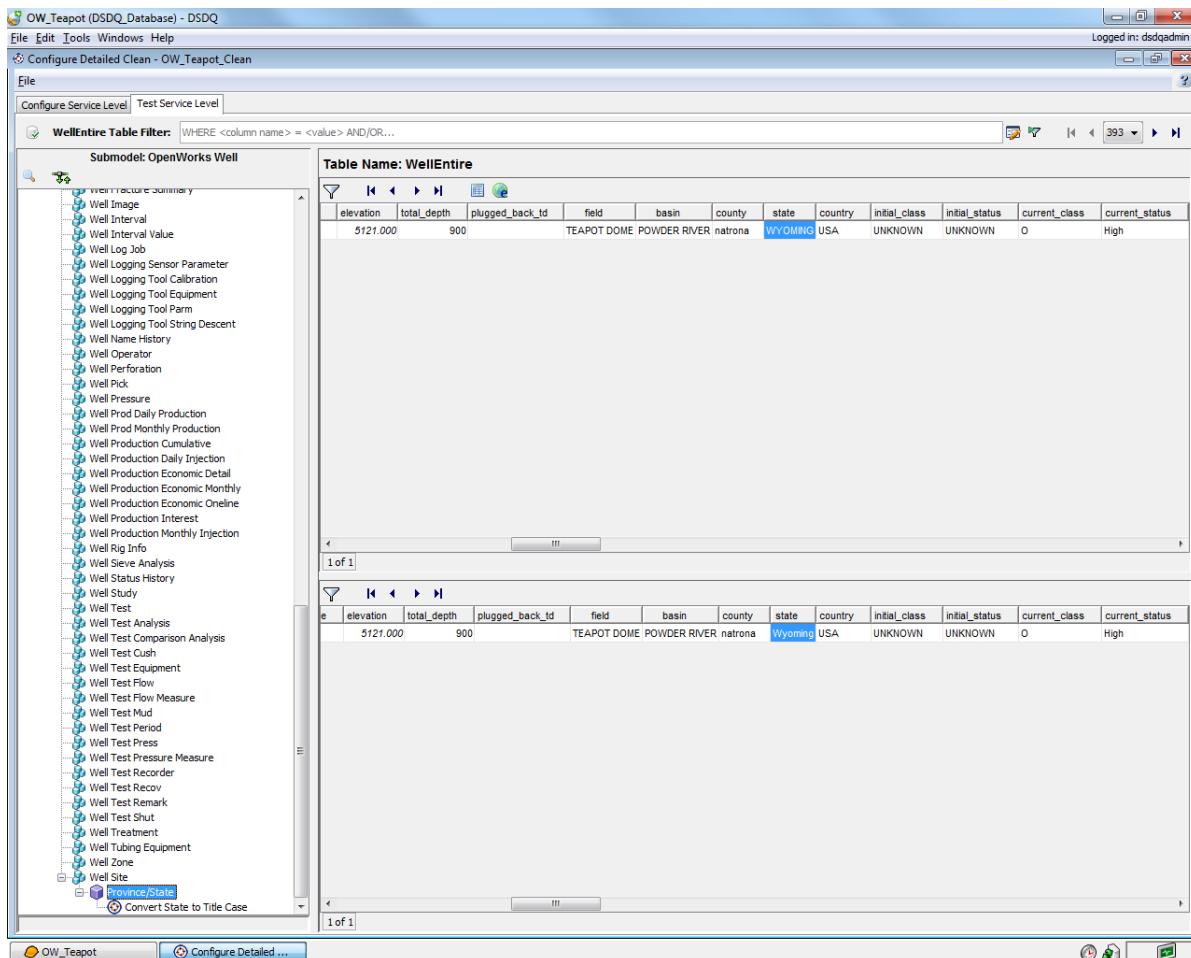
13. Drag and drop the **Convert to Title Case** onto the **Rules** area.



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.



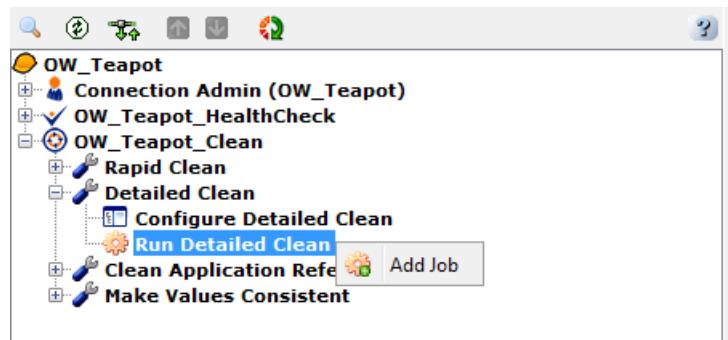
By looking at the columns that have been changed and temporary columns, you can verify that the behavior of the service level is correct prior to running the **Run Detailed Clean** Task.

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.

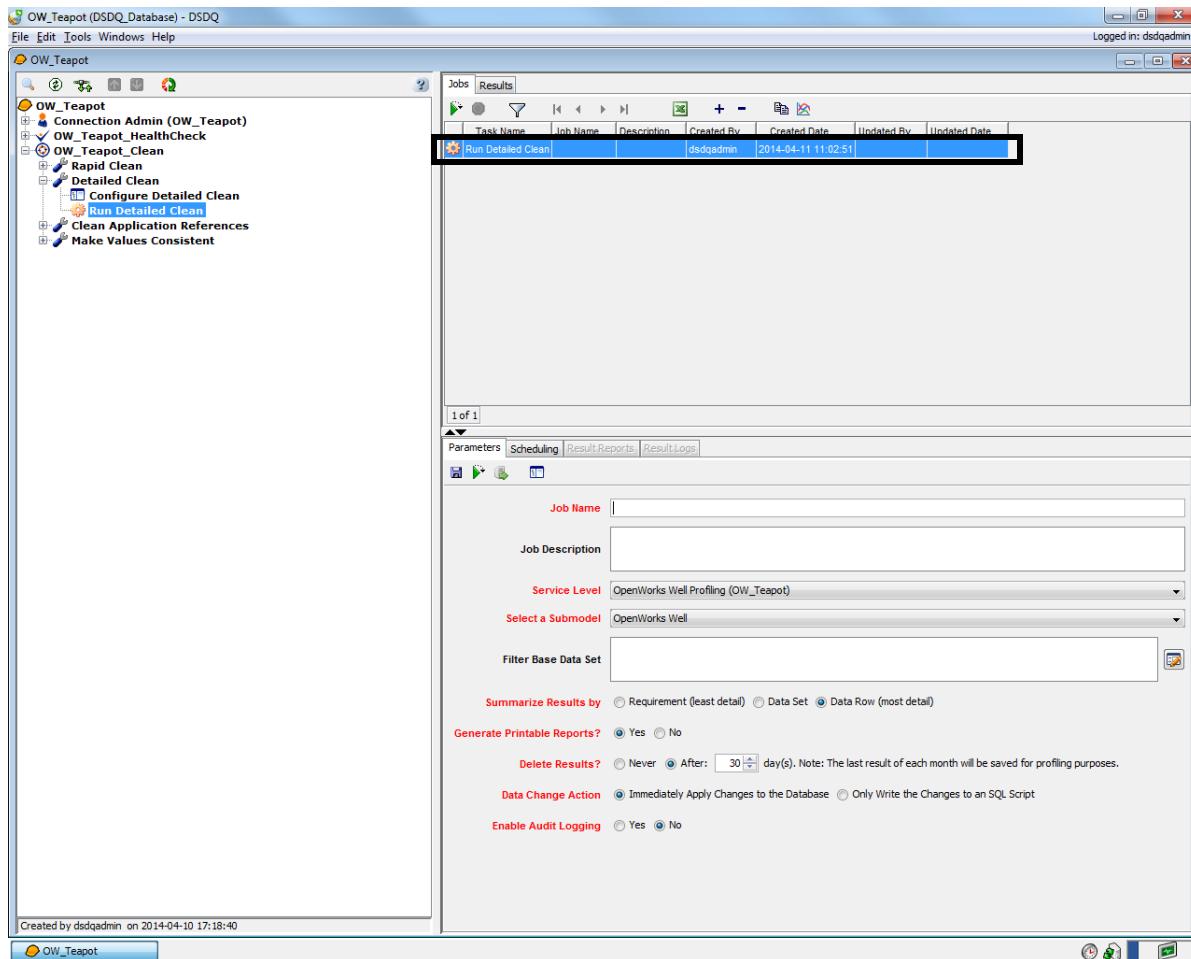
Exercise: Running the Detailed Clean Task

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**.



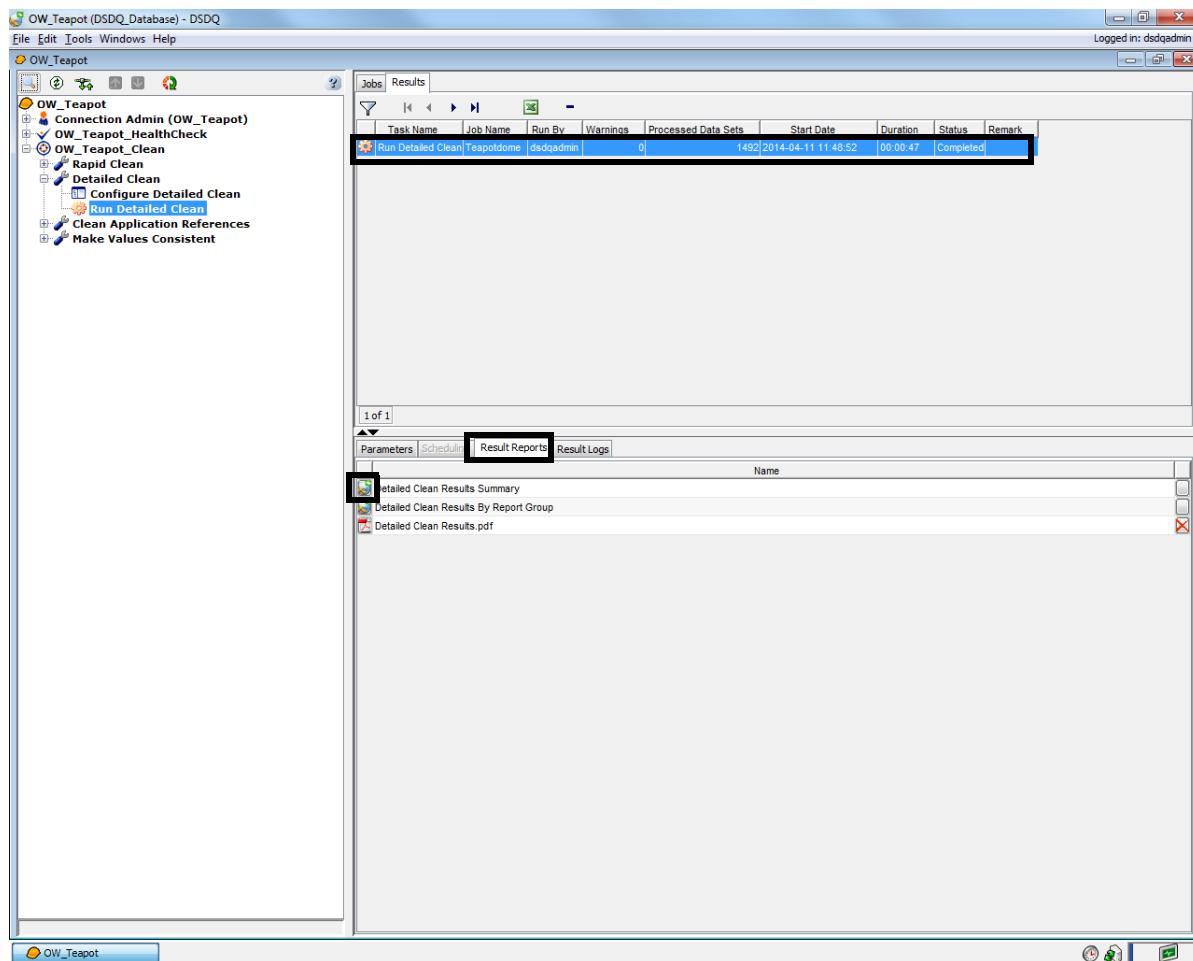
2. Enter **Teapotdome** in the **Job Name** field.

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.

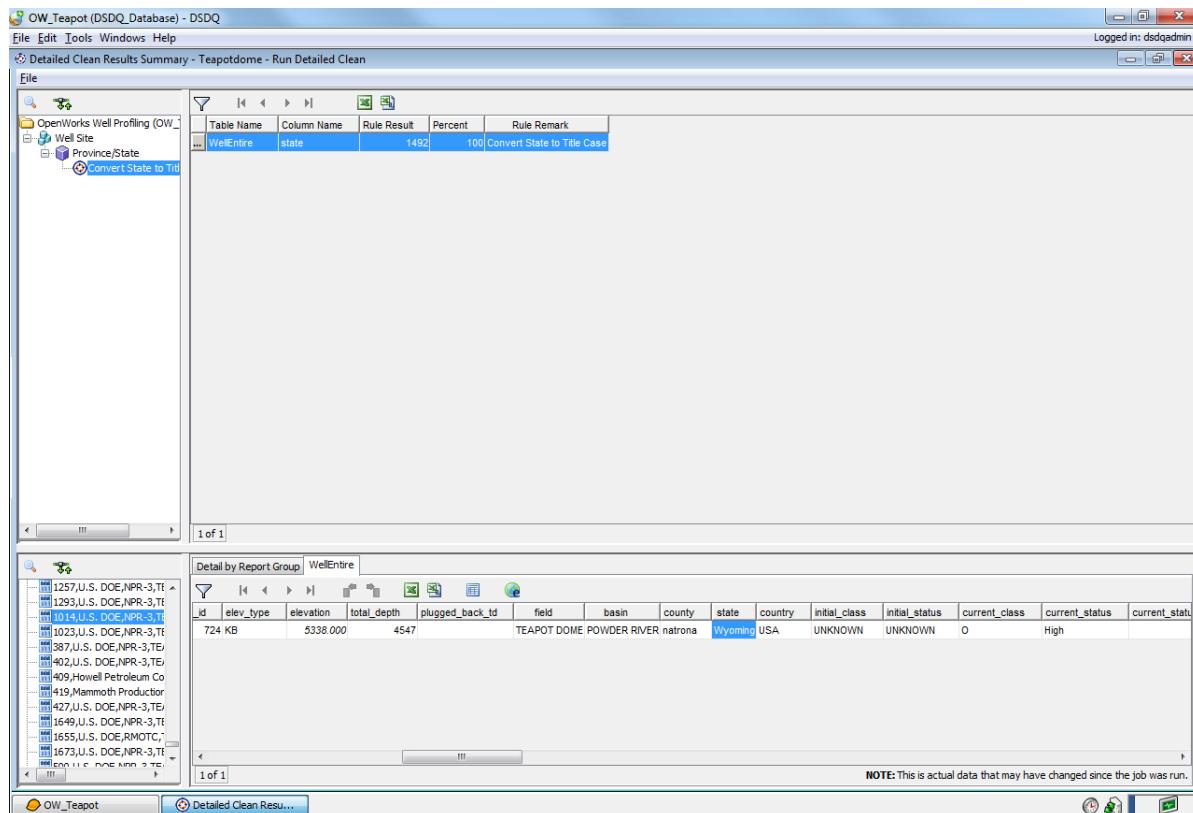
The **Run Detailed Clean** Task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

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.

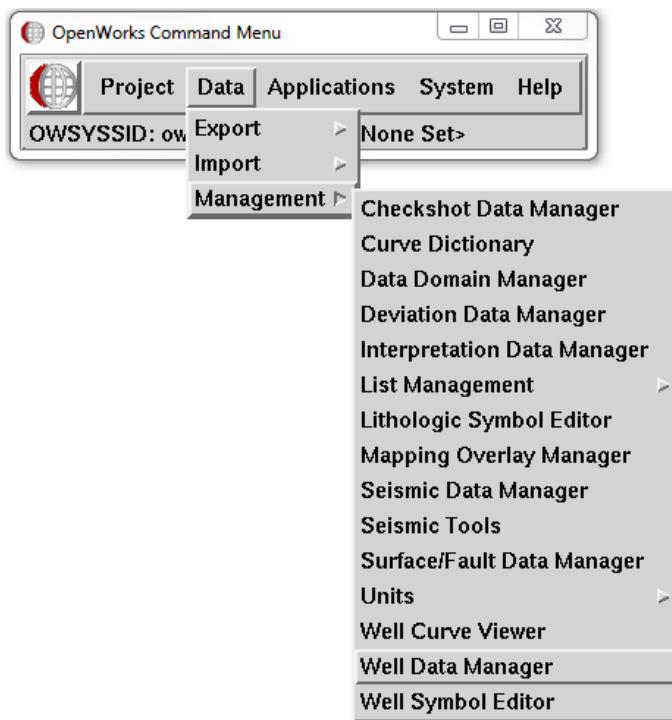


16. Select **File > Exit** to close the **Basic View Frame** window.

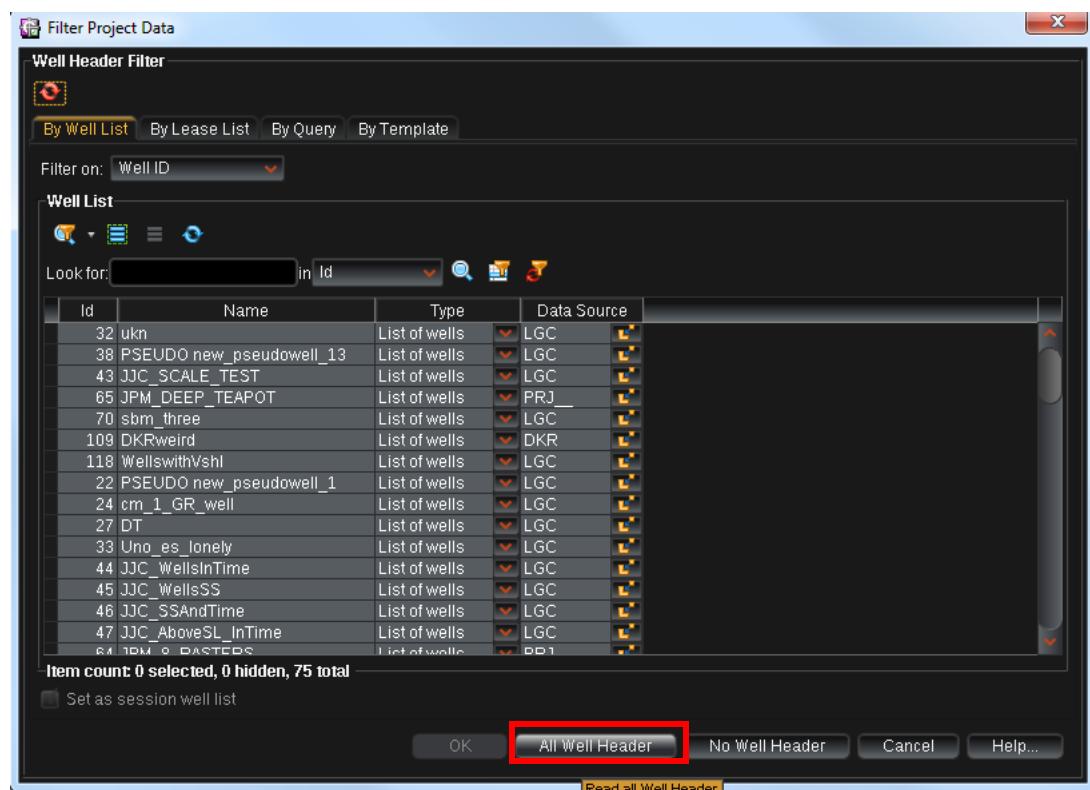
Validating Corrected Data in the OpenWorks Data Source

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.

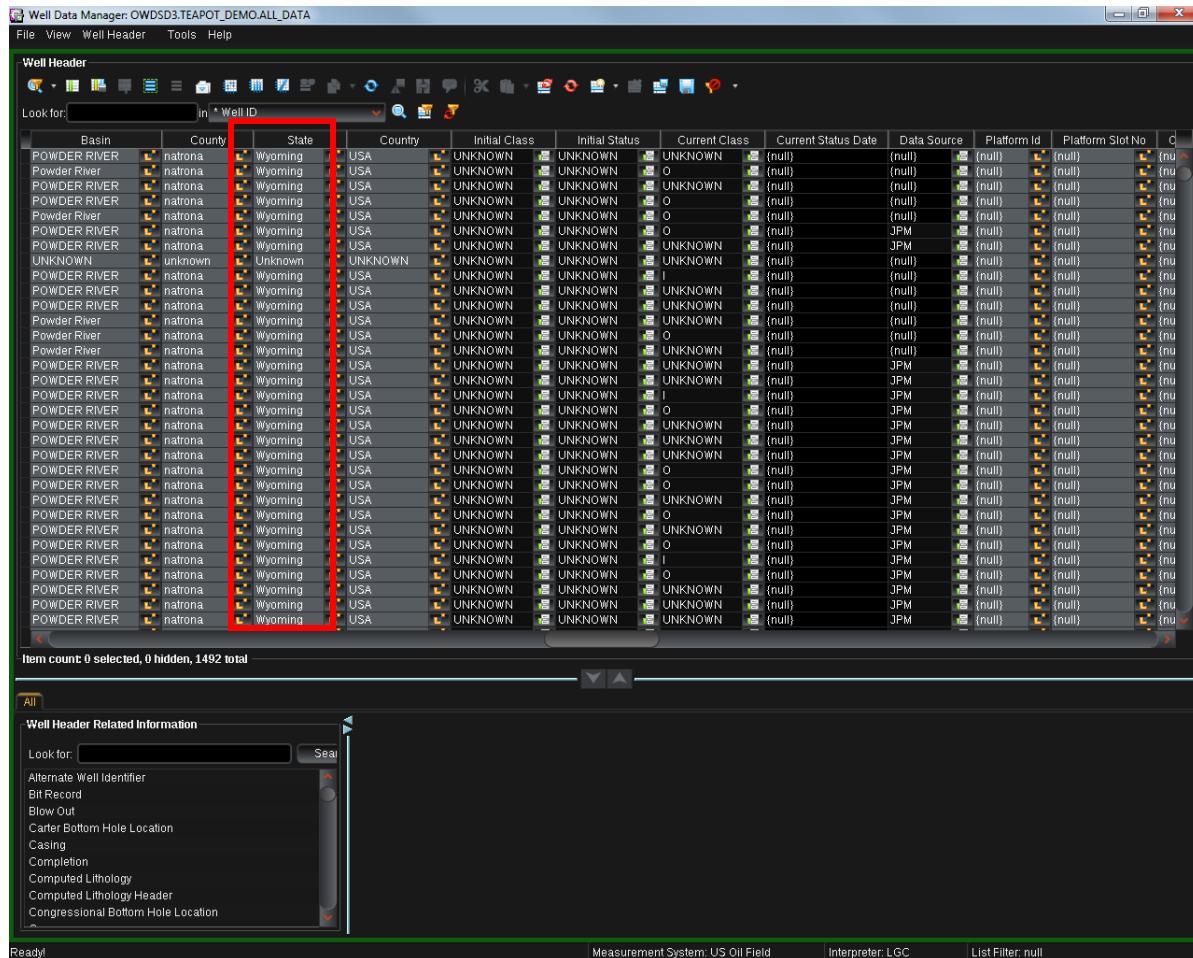


The Filter Project Data window appears.



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

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

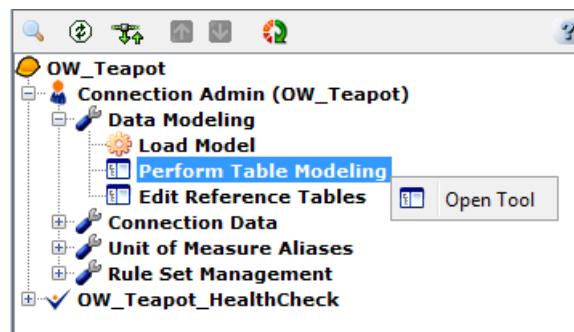


4. Select **File > Exit** to close the **Well Data Manager** window.

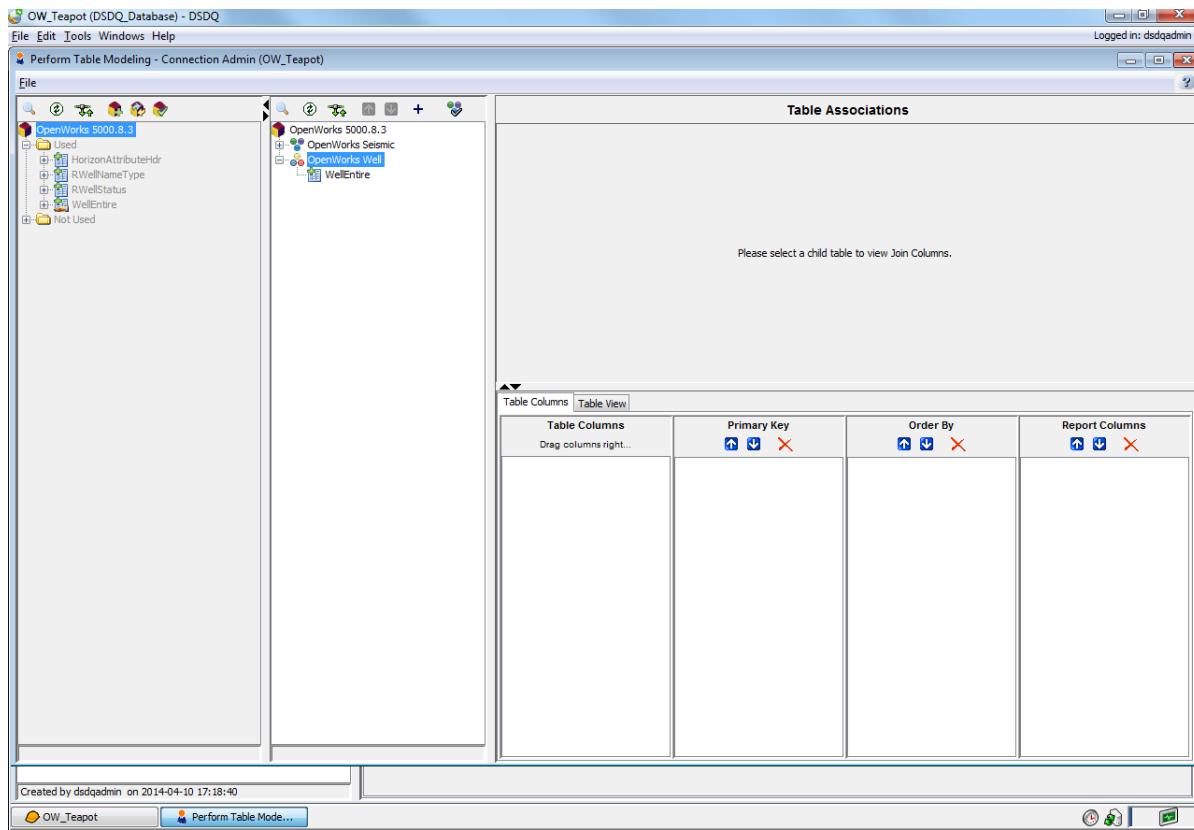
Viewing Results on the DSDQ Web Dashboard

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

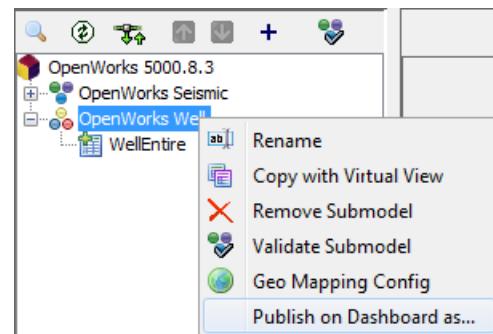
1. Click  on the DecisionSpace Data Quality Tree to expand **Connection Admin (OW_Teapot)**.
2. Click  on the **Data Modeling** Activity.
3. Double-click the **Perform Table Modeling** Tool on the DecisionSpace Data Quality Tree or right-click the **Perform Table Modeling** Tool and select **Open Tool**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



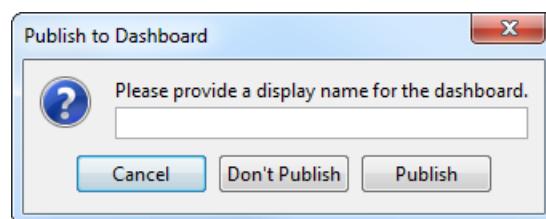
The **Perform Table Modeling - Connection Admin** window appears.



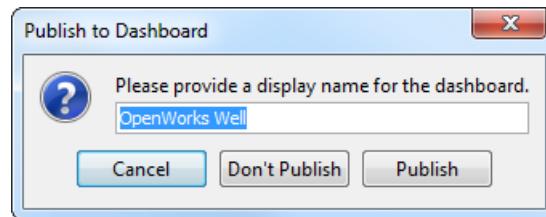
- Right-click the **OpenWorks Well** submodel from the Submodel Listing Tree and select the **Publish on Dashboard as...** option from the pop-up menu.



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

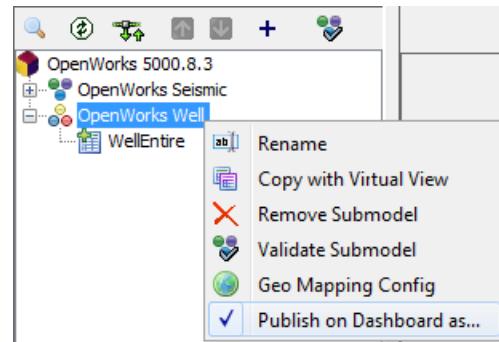


5. Enter **OpenWorks Well** in the **Please Provide a display name for the dashboard** field.



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.

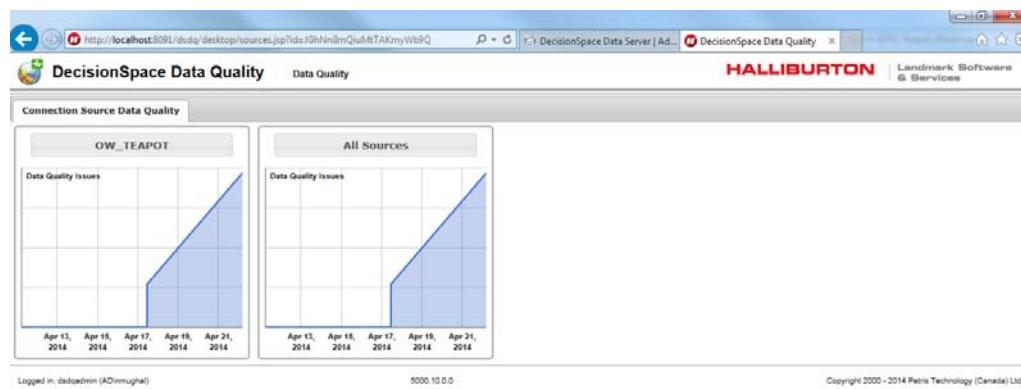


7. Select **File > Exit** from the menu bar to close the **Perform Table Modeling** window.
8. Enter **http://localhost:8091** in the address bar of the web browser. The **Please wait. Your browser will be redirected when ready.**



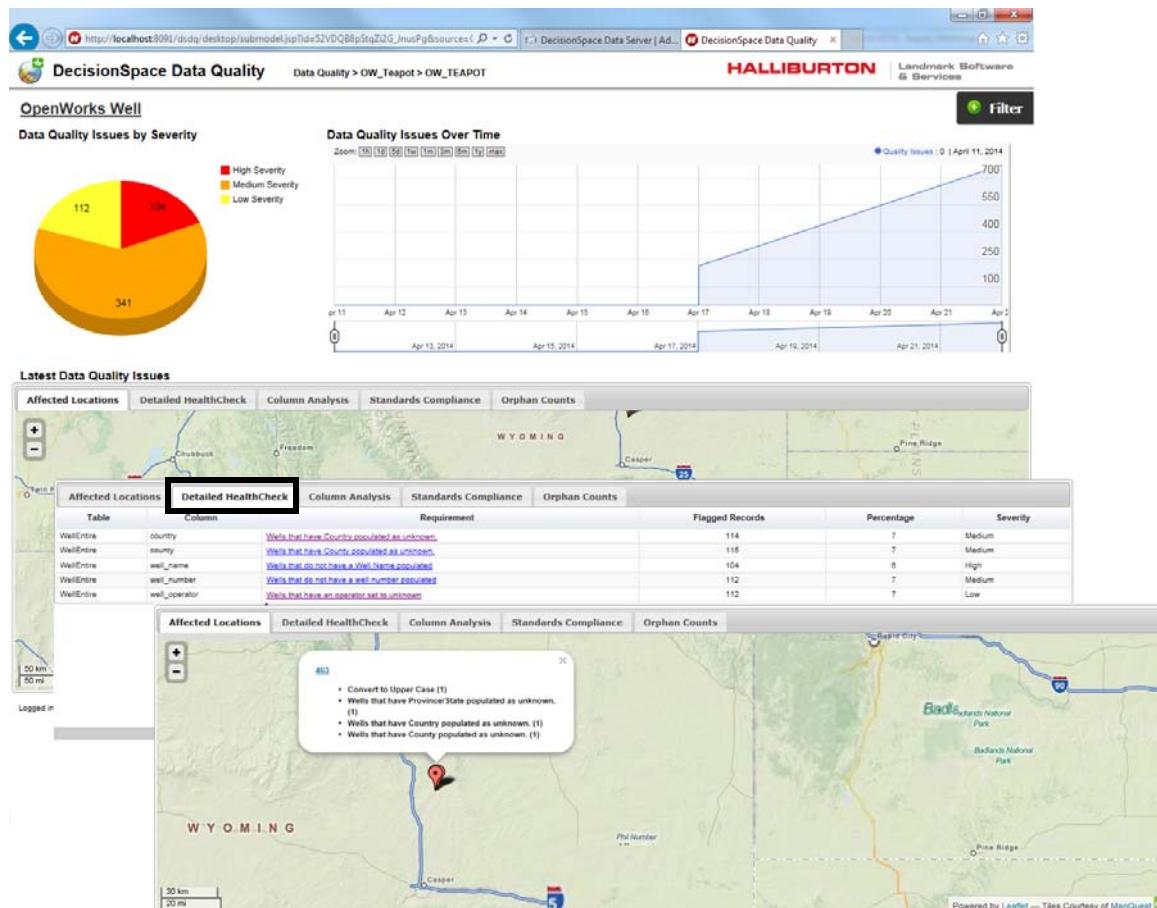
message displays in the web browser.

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.

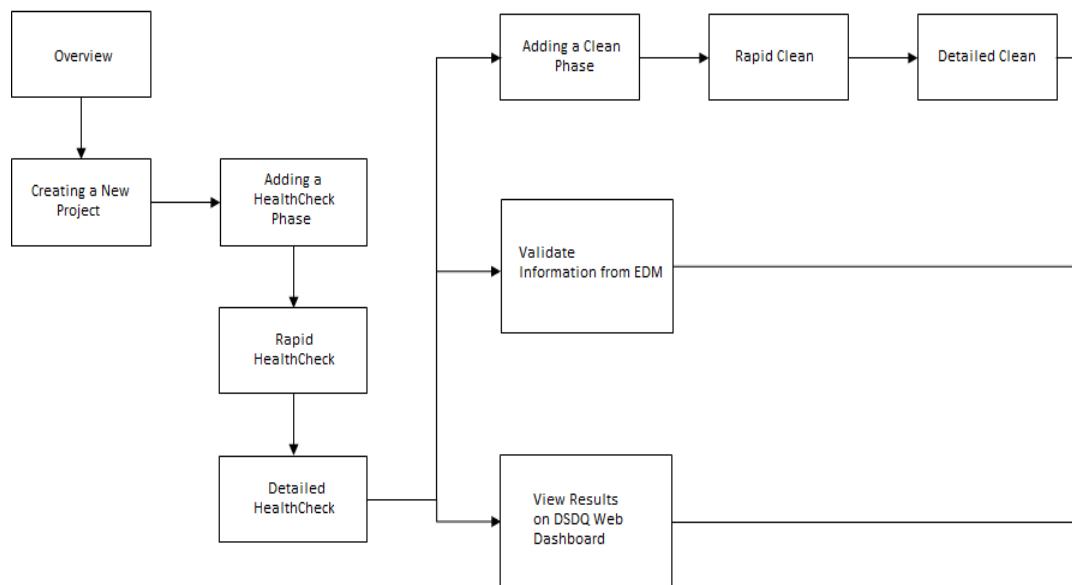
- Creating a New Project
- Evaluating data using the HealthCheck Phase
 - a) Rapid HealthCheck
 - Running Table Analysis on All Tables
 - Running Column Analysis on Columns
 - Running Table Analysis on Modeled Tables
 - Running Column Analysis on Modeled Tables
 - b) Detailed HealthCheck
 - Configuring 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



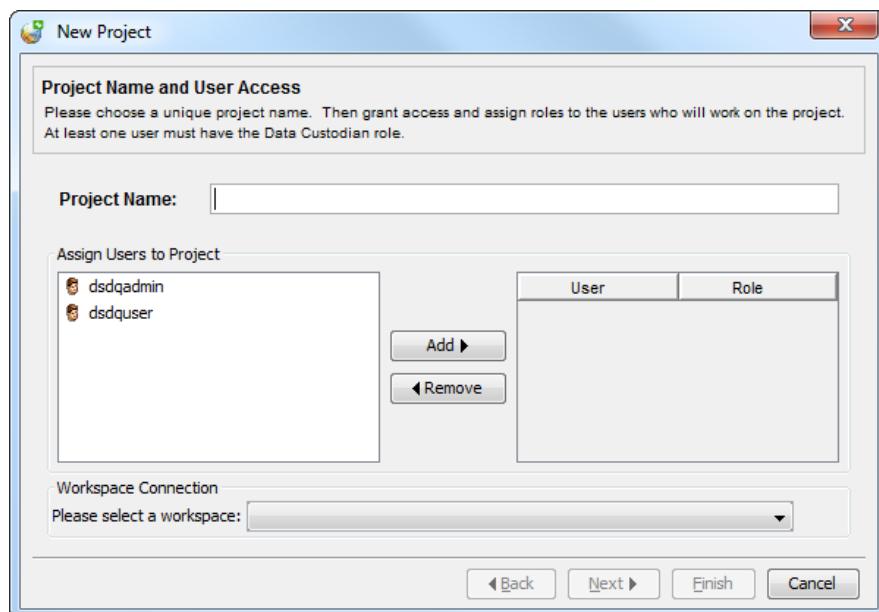
Creating a New Project

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).

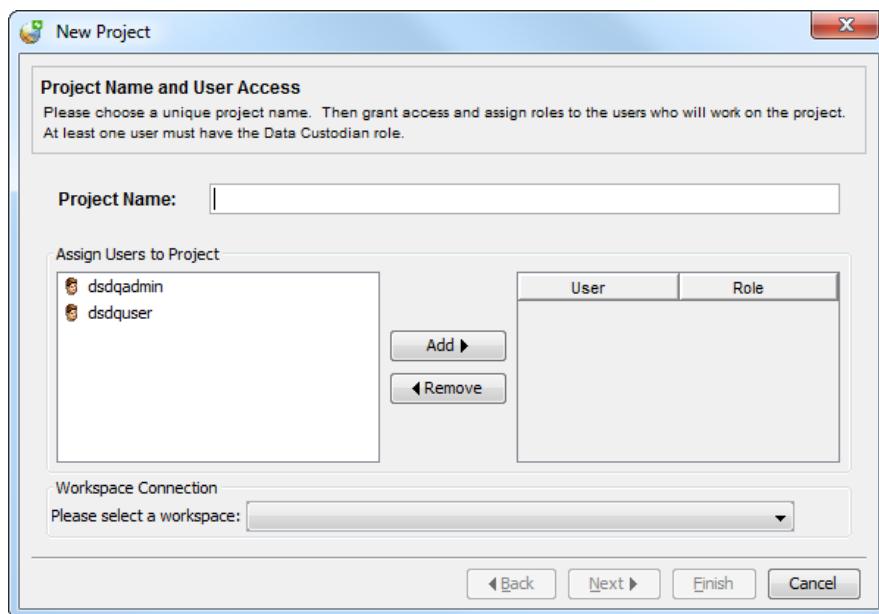
A new Project can be created:

- 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.



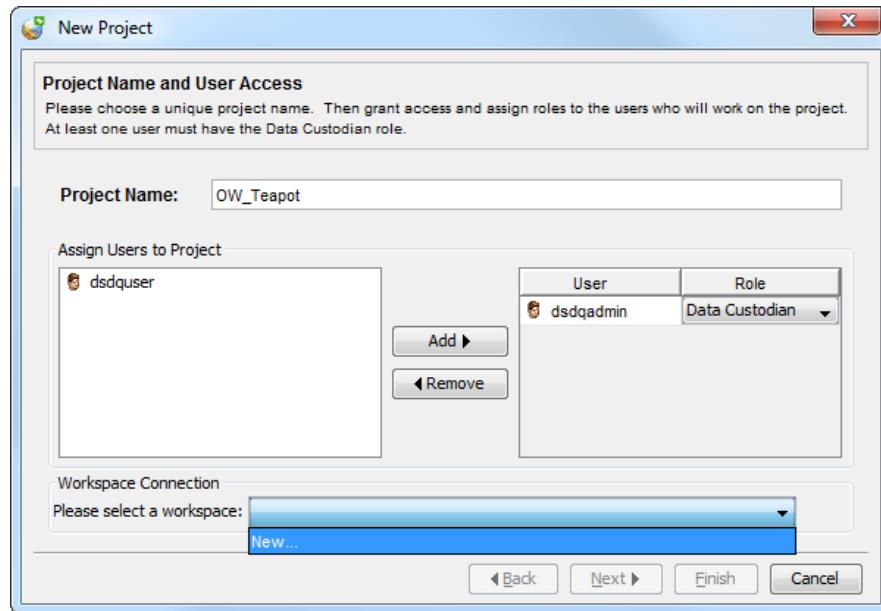
2. Enter **EDM_Teapot** in the **Project Name** field.

Note

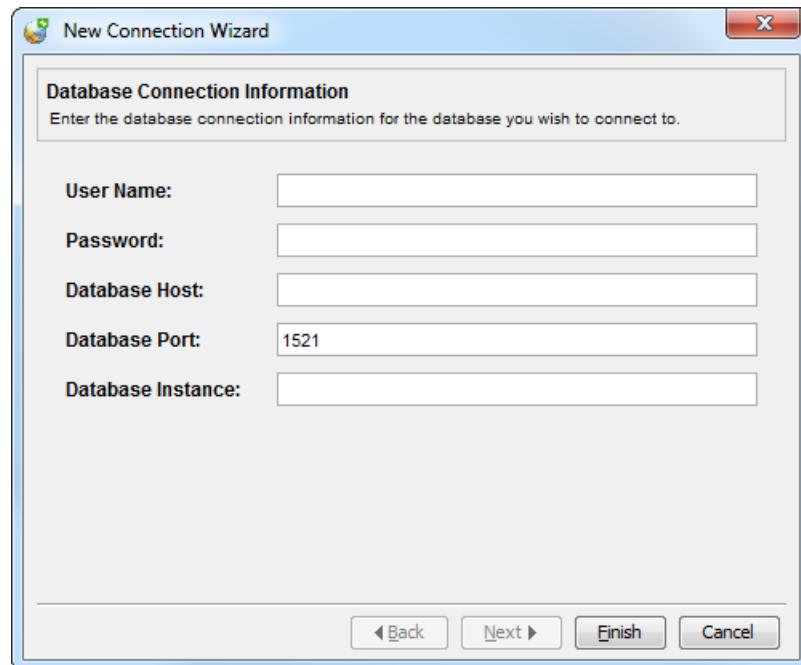
You will be using **EDM_Teapot** throughout this workflow.

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.

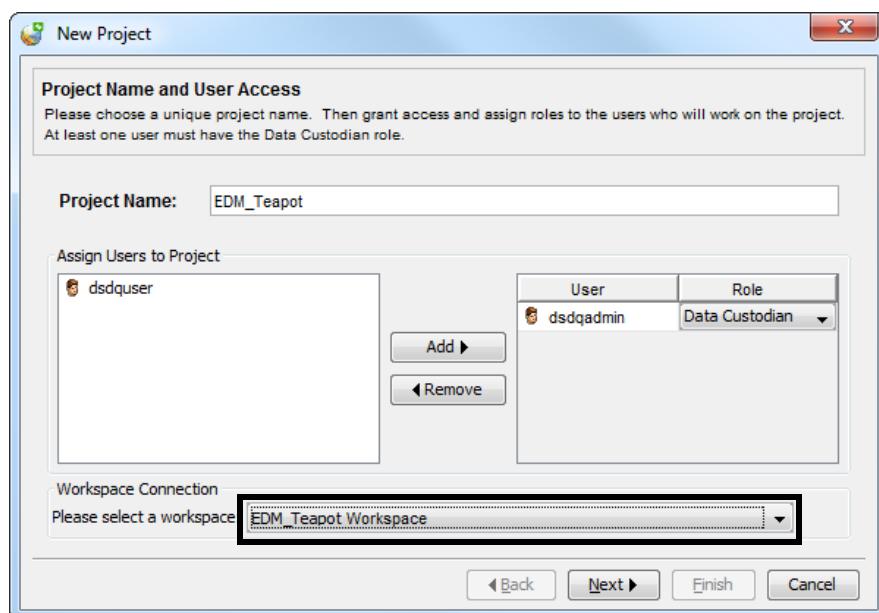


7. Enter **DSDQ_EDM_Teapot_WKSP** in the **User Name** field.

Note

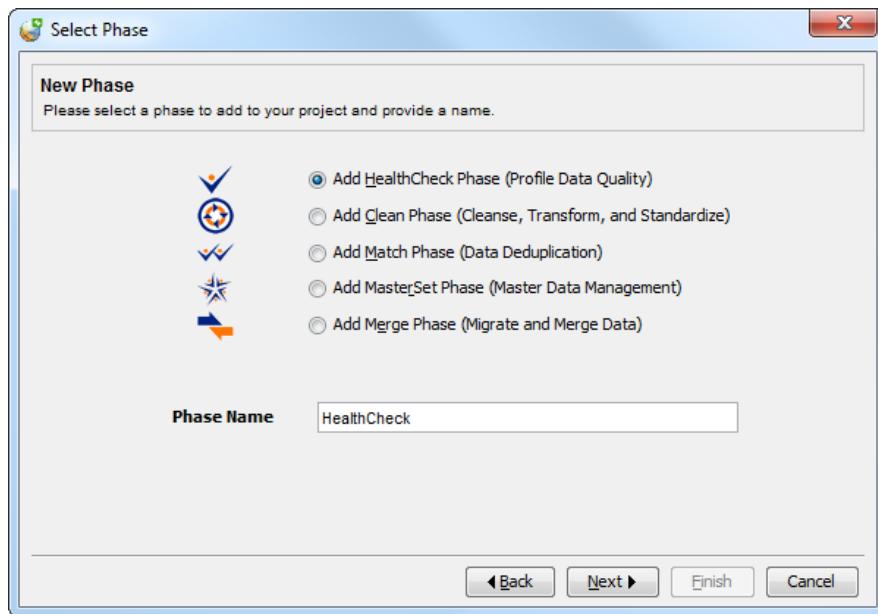
Make sure that the user and database connection have already been setup in Oracle (*reference: chapter 1, DSDQ Training Manual*).

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 **oradssdq10** 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.

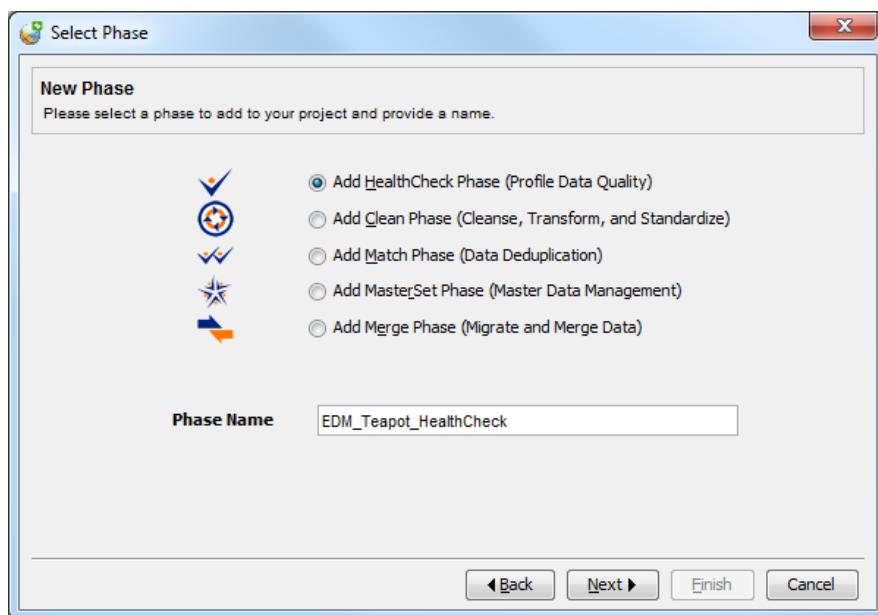


12. Click **Next** to continue.

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

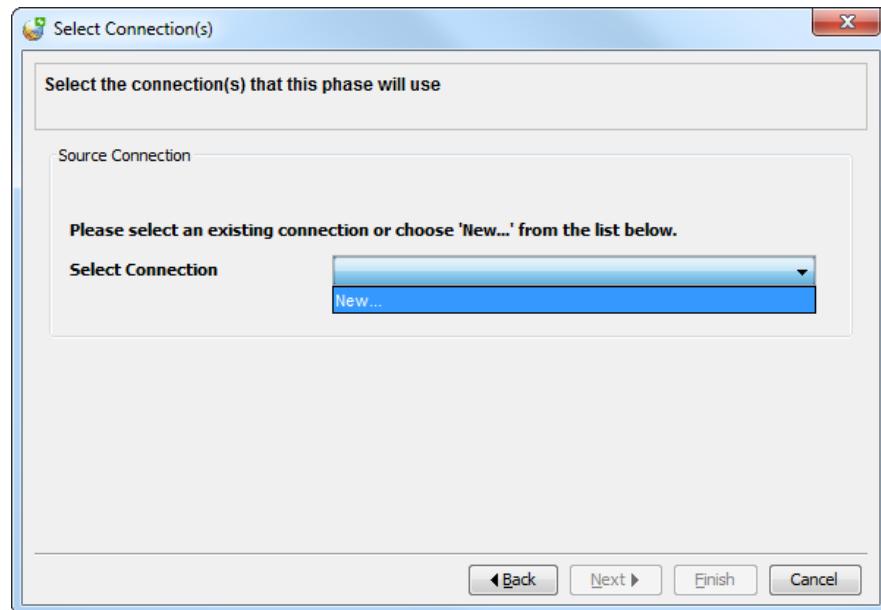


13. Enter **EDM_Teapot_HealthCheck** in the **Phase Name** field.

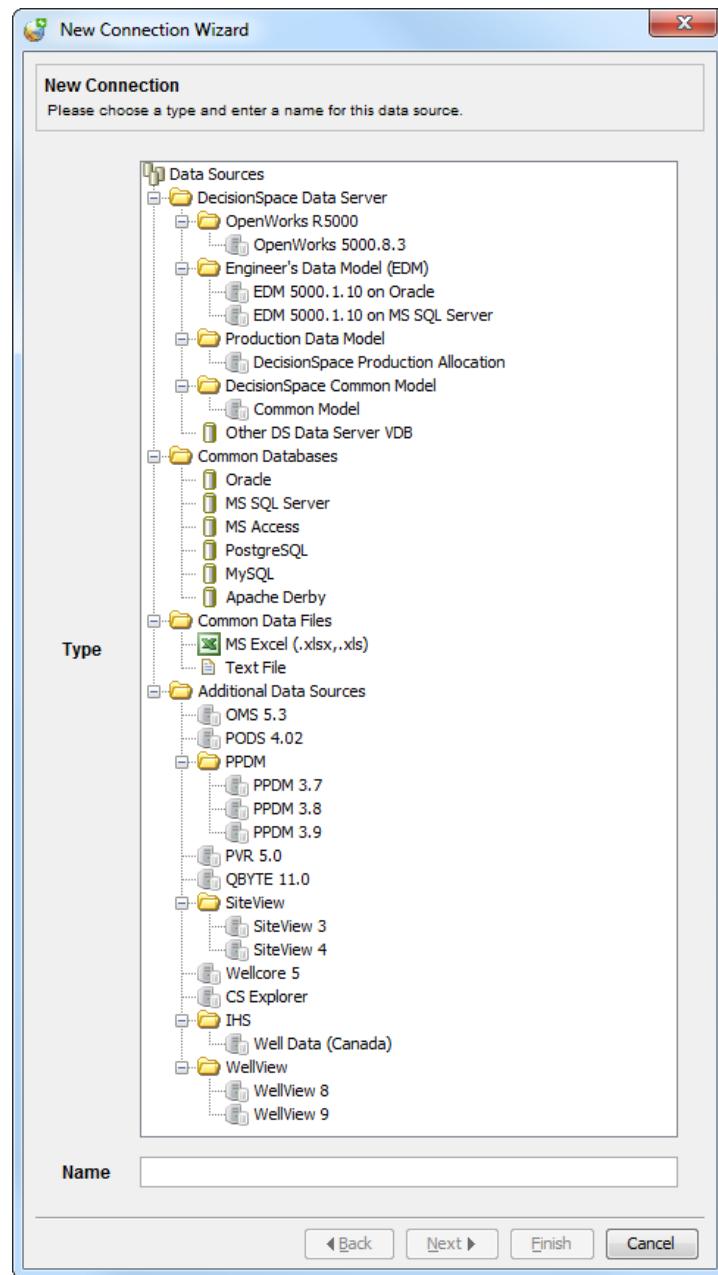


14. Click **Next** to continue.

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

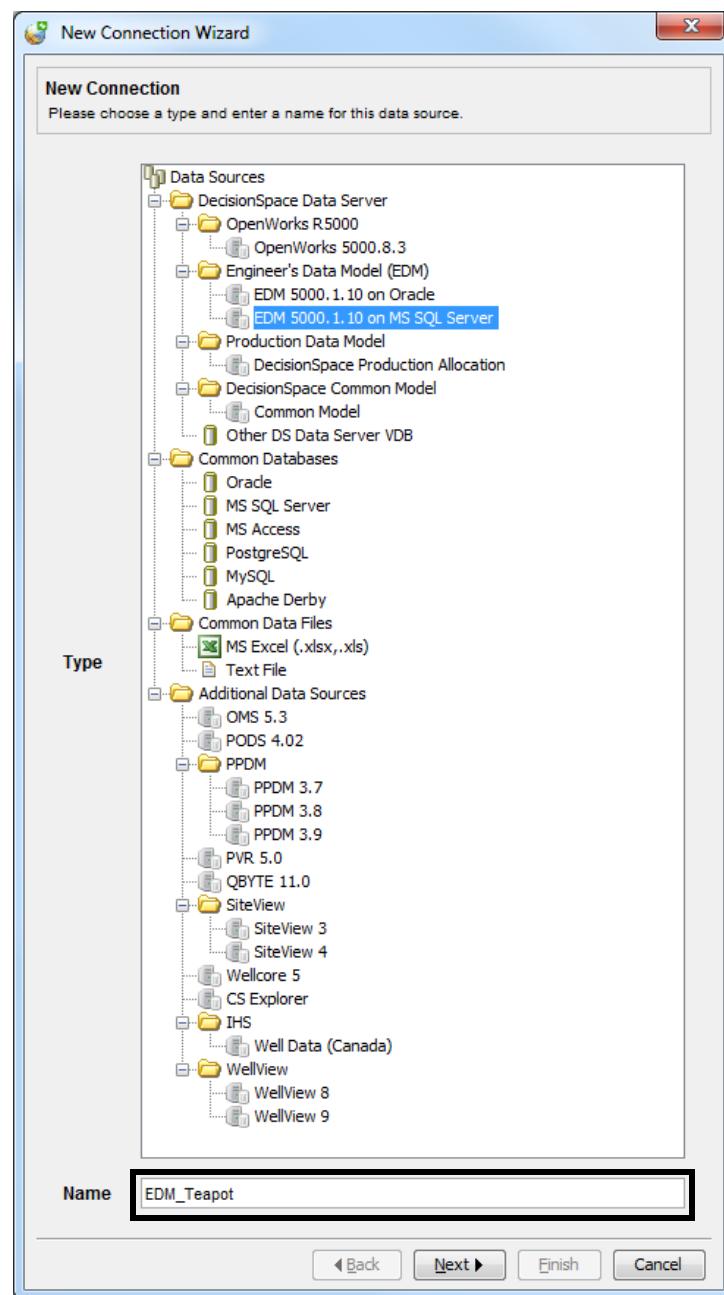


15. Select **New...** from the **Select Connection** drop-down list.
The **New Connection Wizard - New Connection** window appears.



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

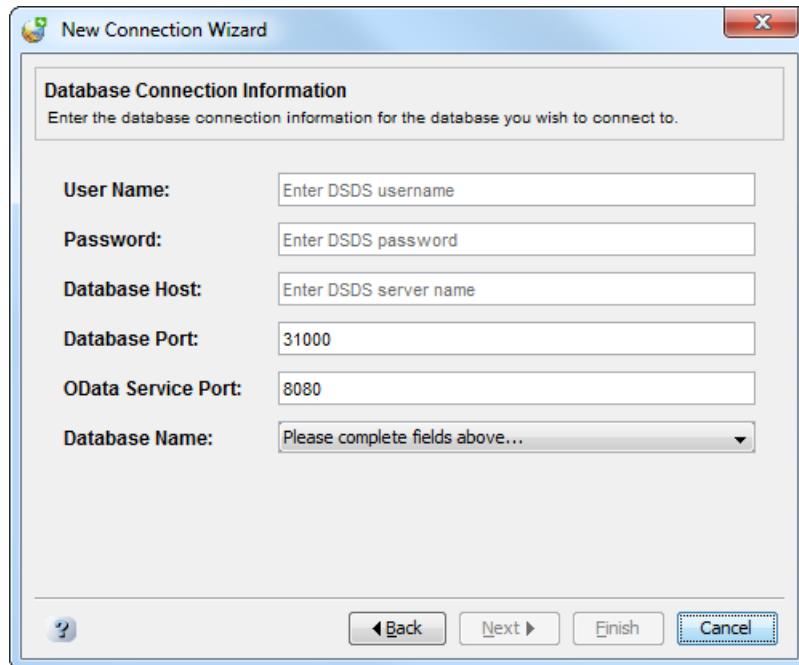
17. Enter **EDM_Teapot** in the **Name** field.



18. Click **Next** to continue.

The New Connection Wizard - Database Connection

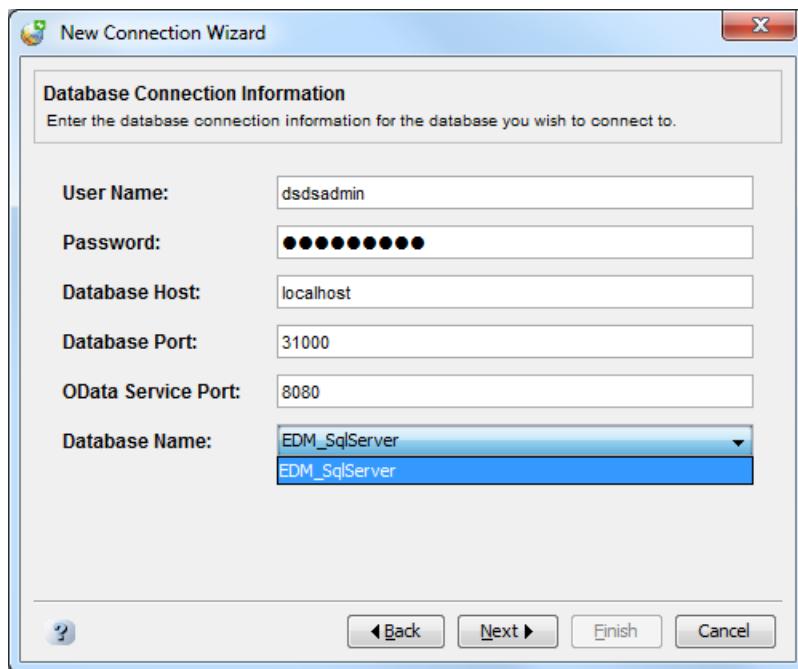
Information window appears displaying database connection information for DecisionSpace Data Server.



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**.

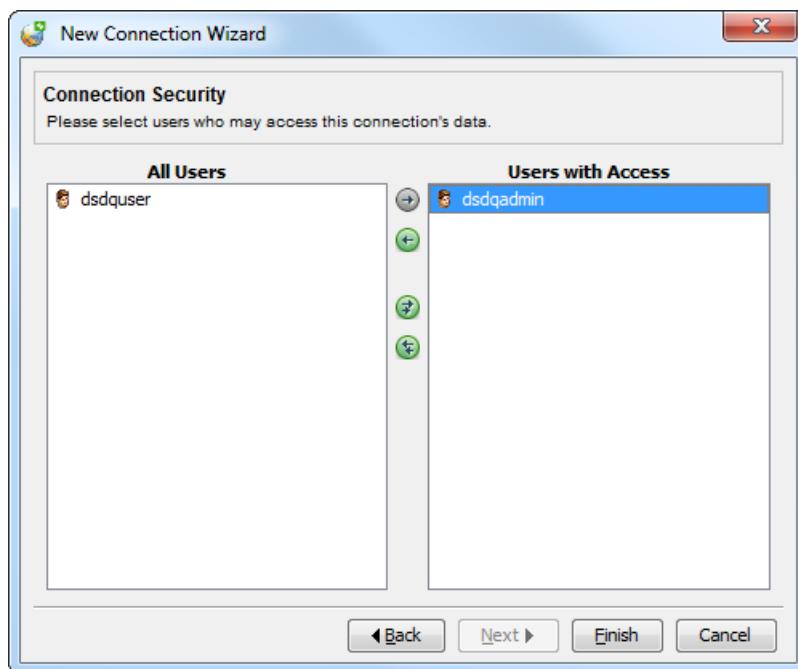
23. Accept the default **OData Service Port** value of **8080**.

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



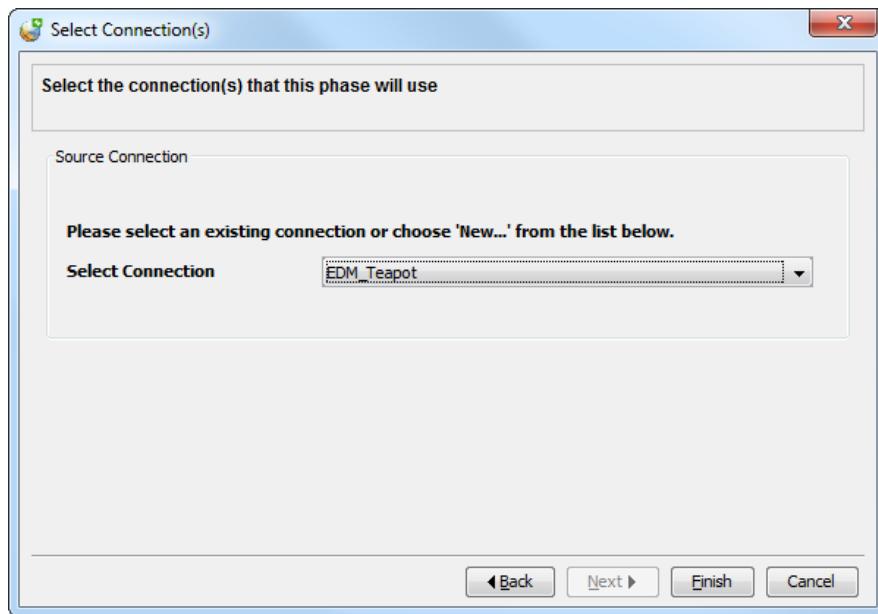
24. Click **Next** to continue.

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 showing "EDM_Teapot" expanded, with "Connection Admin (EDM_Teapot)" and "EDM_Teapot_HealthCheck" as children.
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, and 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 search, refresh, and other functions.

Evaluating Data using the HealthCheck Phase

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

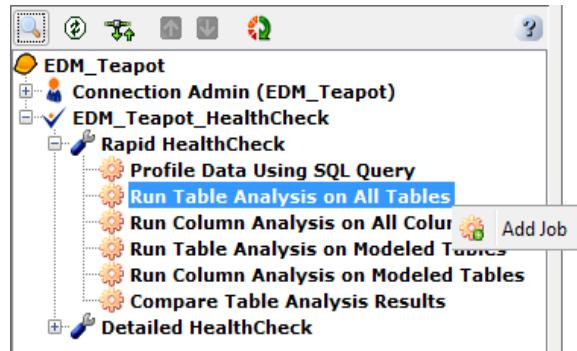
Rapid HealthCheck Activity

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

Exercise: Running Table Analysis on All Tables Task

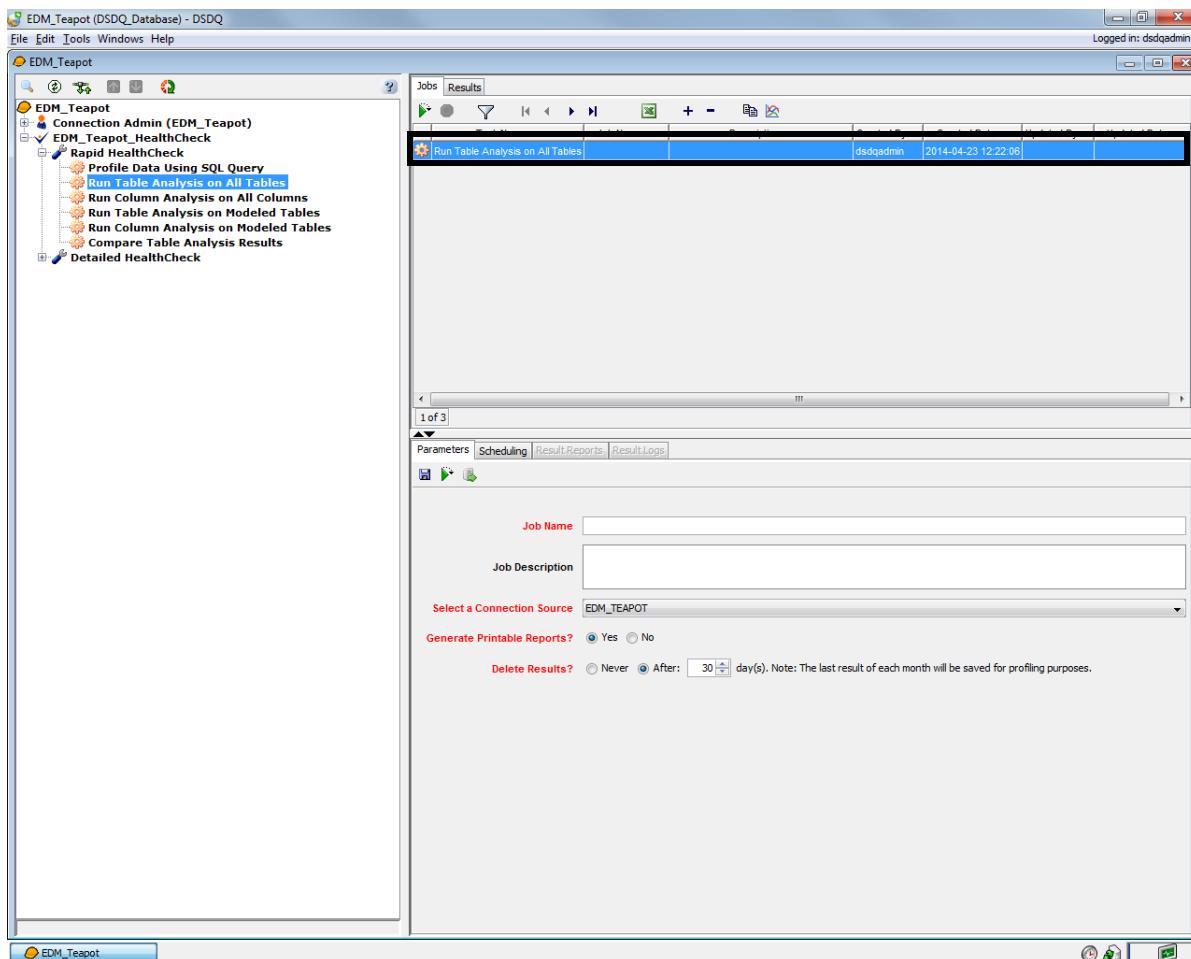
The **Run Table Analysis on All Tables** Task is used to analyze all the tables for issues and inconsistencies. In this particular exercise, we will analyze the tables and count the number of rows in them. Rows are counted when values are entered in them. To run Table Analysis on all tables:

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.

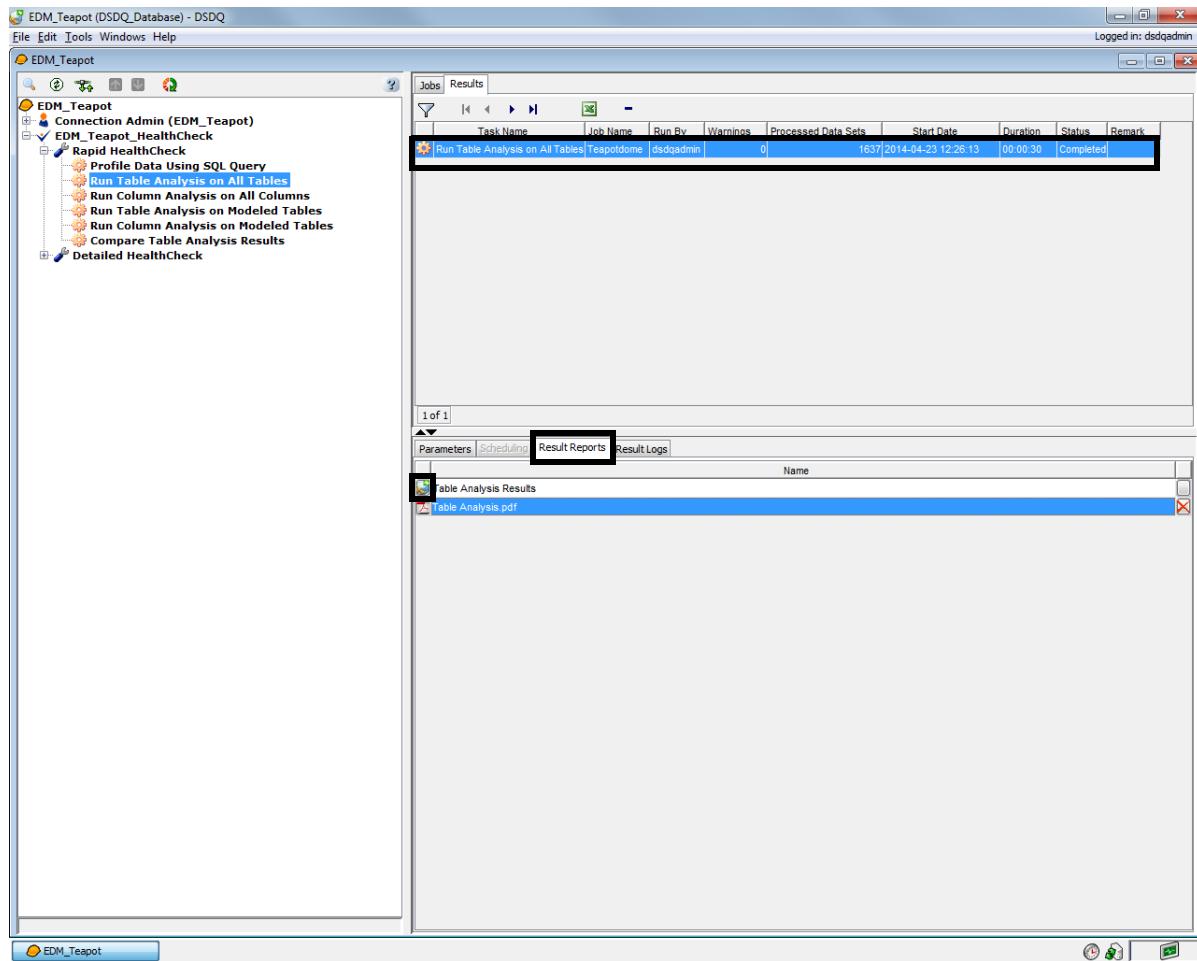


3. Enter **Teapotdome** in the **Job Name** field.
 4. Enter **Table Analysis on All Tables of EDM_Teapot** in the **Job Description** field.
 5. Select **EDM_Teapot** from the **Select a Connection Source** drop-down list.
 6. Select the **Yes** option for **Generate Printable Reports**.
 7. Select 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.

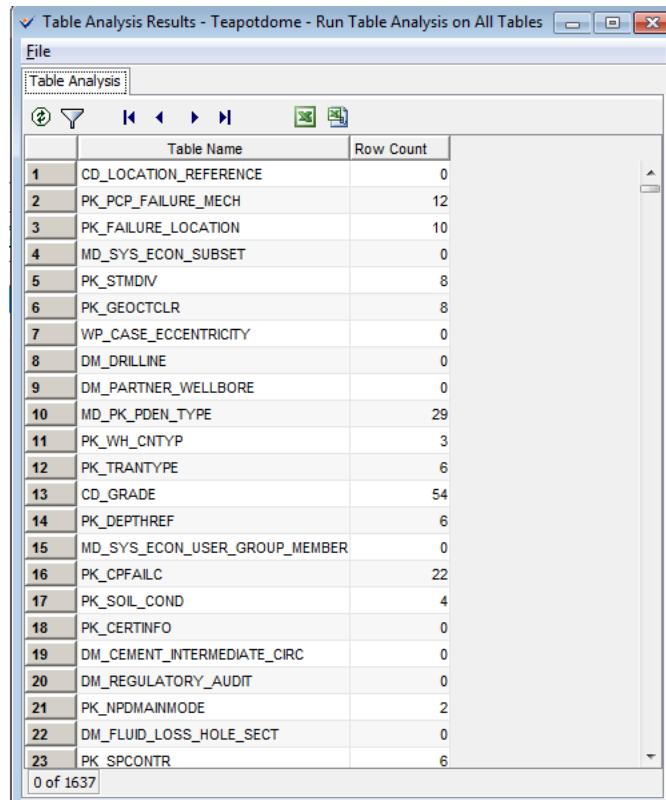


	Table Name	Row Count
1	CD_LOCATION_REFERENCE	0
2	PK_PCP_FAILURE_MECH	12
3	PK_FAILURE_LOCATION	10
4	MD_SYS_ECON_SUBSET	0
5	PK_STMDIV	8
6	PK_GEOCTCLR	8
7	WP_CASE_ECCENTRICITY	0
8	DM_DRILLINE	0
9	DM_PARTNER_WELLBORE	0
10	MD_PK_PDEN_TYPE	29
11	PK_WH_CNTYP	3
12	PK_TRANTYPE	6
13	CD_GRADE	54
14	PK_DEPTHREF	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

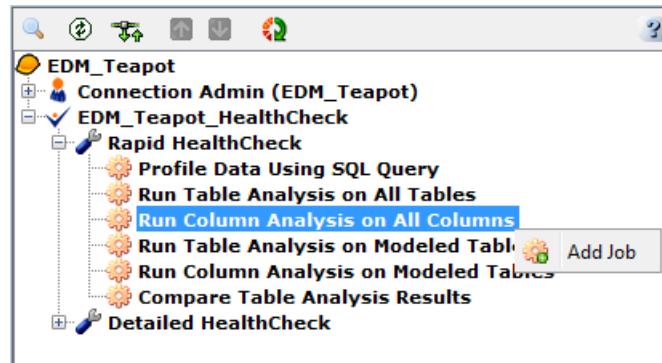
12. Select **File > Exit** to close the **Basic View Frame** window.

Exercise: Running Column Analysis on All Columns Task

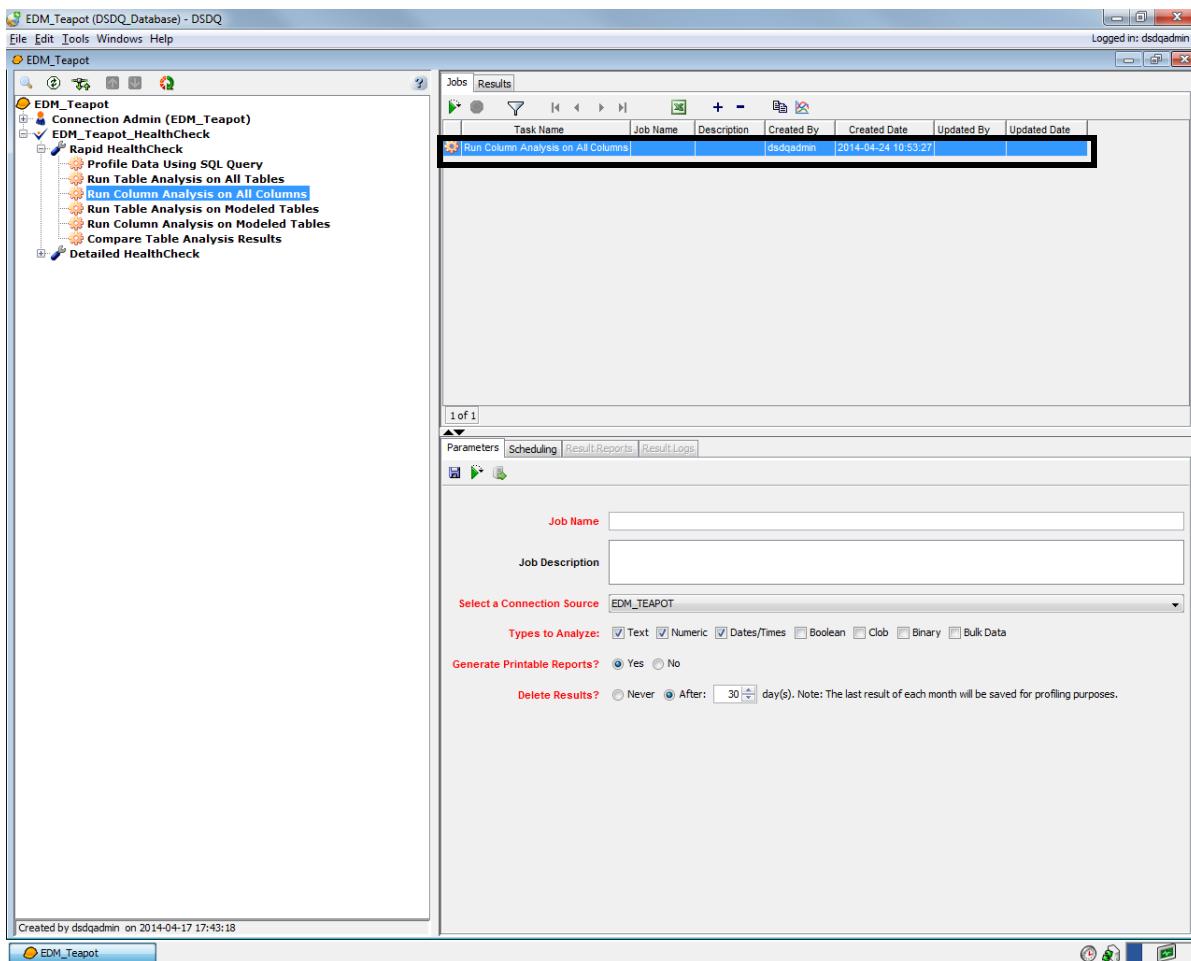
The **Run Column Analysis on All Columns** Task offers basic data profiling by checking the number of rows, number of not null values, unique values, percentage of row populated, minimum values, maximum values, number of values with mixed cases, number of non-printable characters, number of preceding, 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.



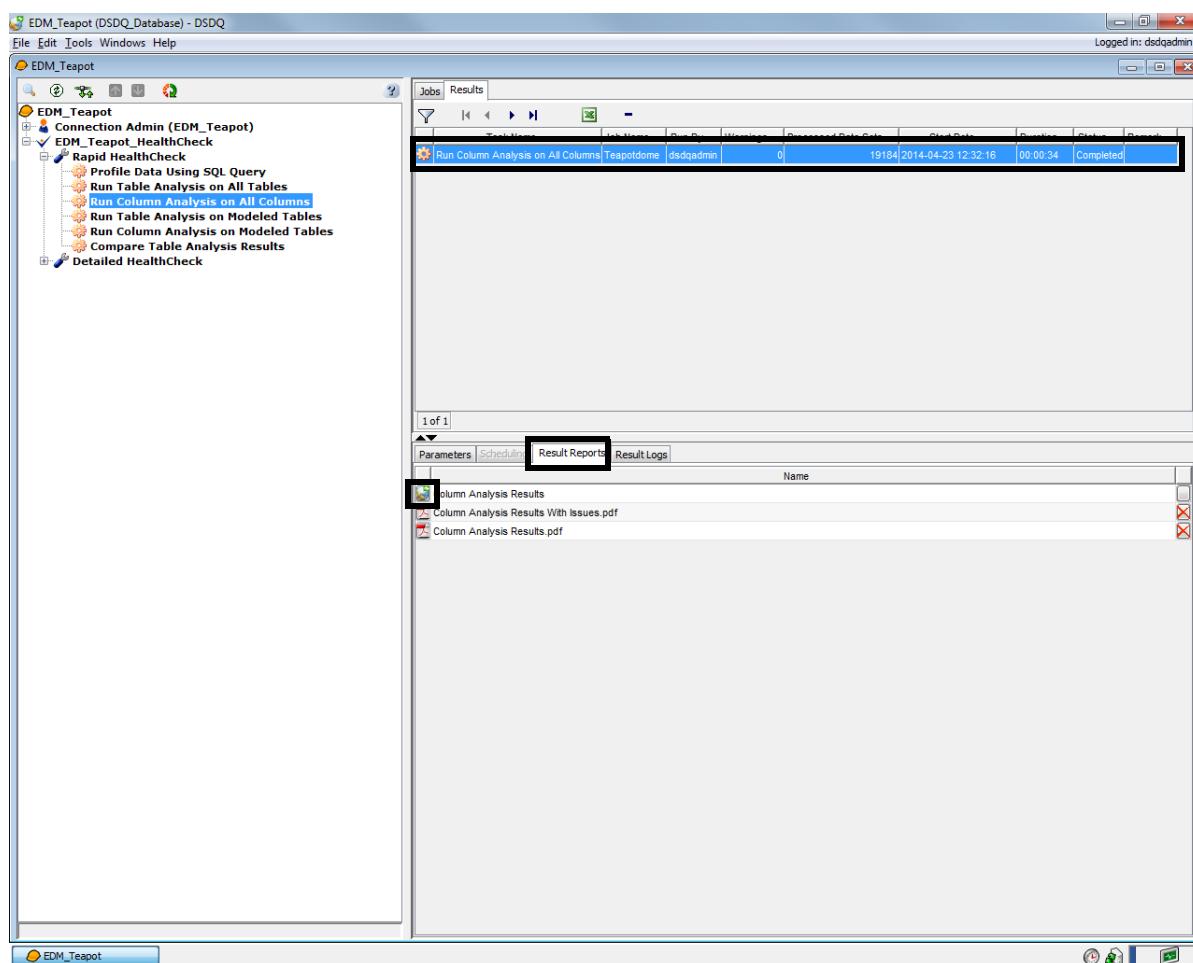
2. Enter **Teapotdome** in the **Job Name** field.
3. Enter **Column Analysis on All Columns of EDM_Teapot** in the **Job Description** field.

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.

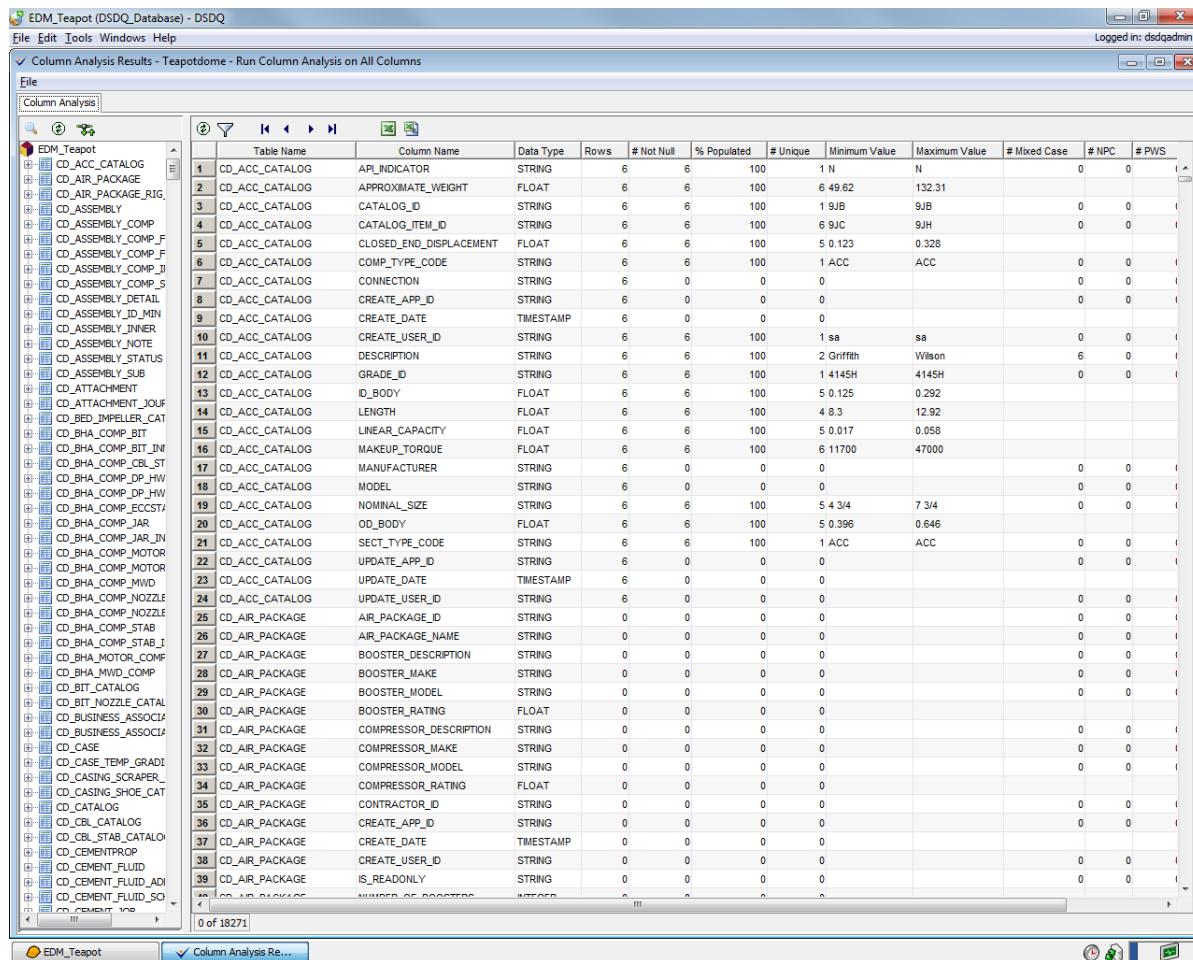


Table Name	Column Name	Data Type	Rows	# Not Null	% Populated	# Unique	Minimum Value	Maximum Value	# Mixed Case	# NPC	# PWS
CD_ACC_CATALOG	APLINDICATOR	STRING	6	6	100	1 N	N		0	0	
CD_ACC_CATALOG	APPROXIMATE_WEIGHT	FLOAT	6	6	100	6 49.62	132.31				
CD_ACC_CATALOG	CATALOG_ID	STRING	6	6	100	1 9JB	9JB	0	0		
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_BOARDS	INTEGER	0	0	0	0					

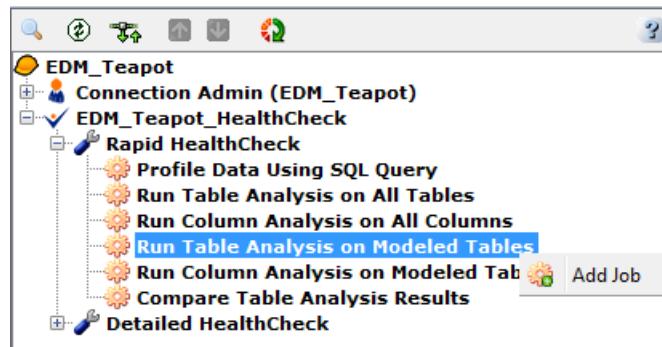
12. Select **File > Exit** to close the **Basic View Frame** window.

Exercise: Running Table Analysis on Modeled Tables Task

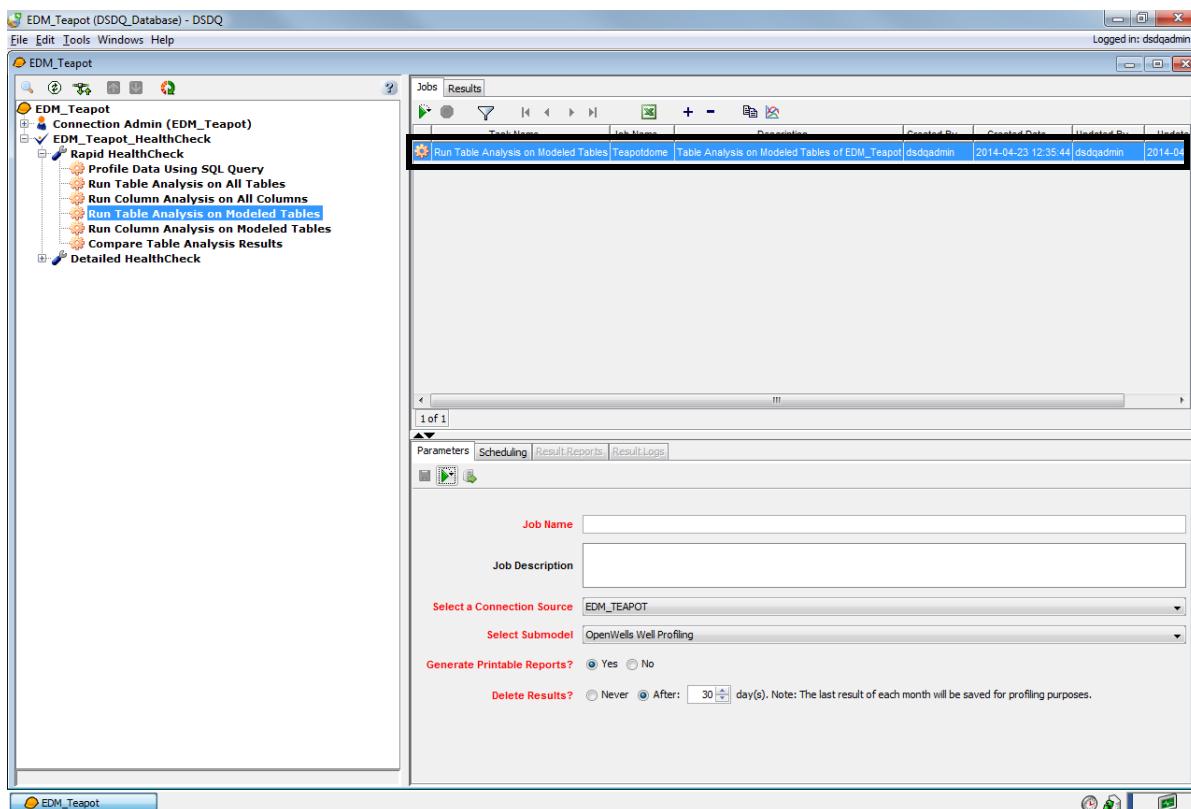
The **Run Table Analysis on Modeled Tables** Task runs only on tables that have been modeled in the **Perform Table Modeling** Tool (*reference: DecisionSpace Data Quality Training Manual, Chapter 4: Data Evaluation, Perform Table Modeling*). To run Table Analysis on all the modeled tables:

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.

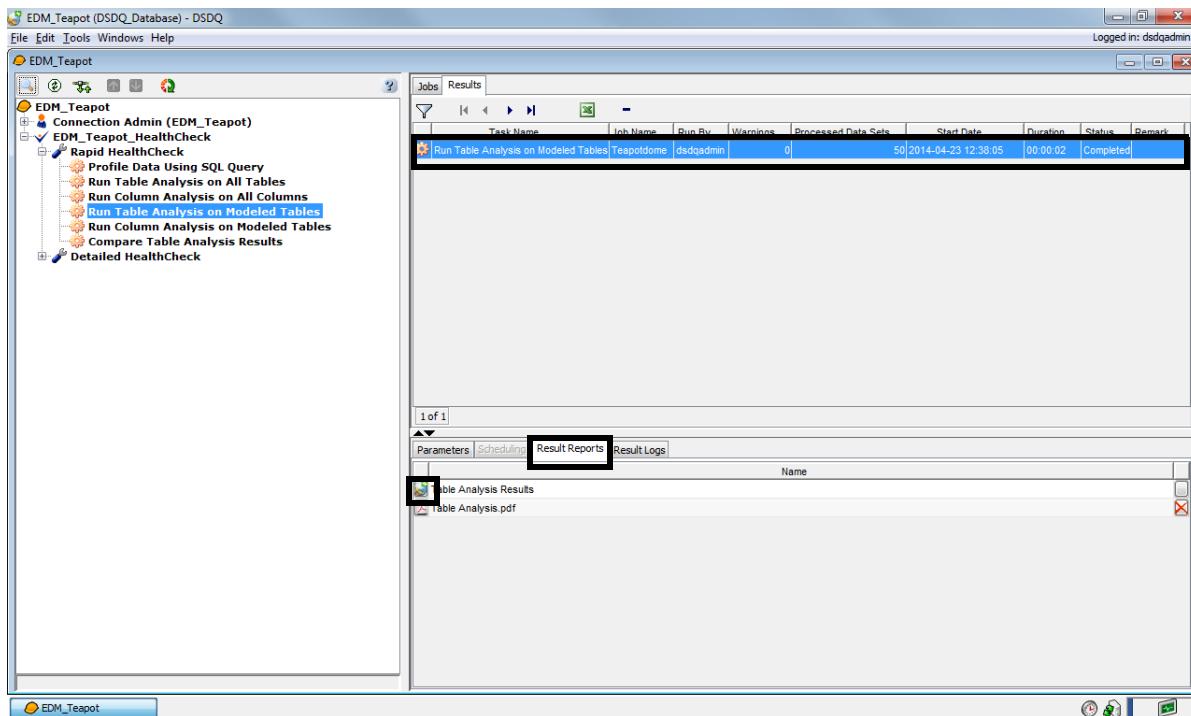


A new job is initiated and it displays on the **Job and Results Information Pane** on the right side of the DecisionSpace Data Quality Project window.

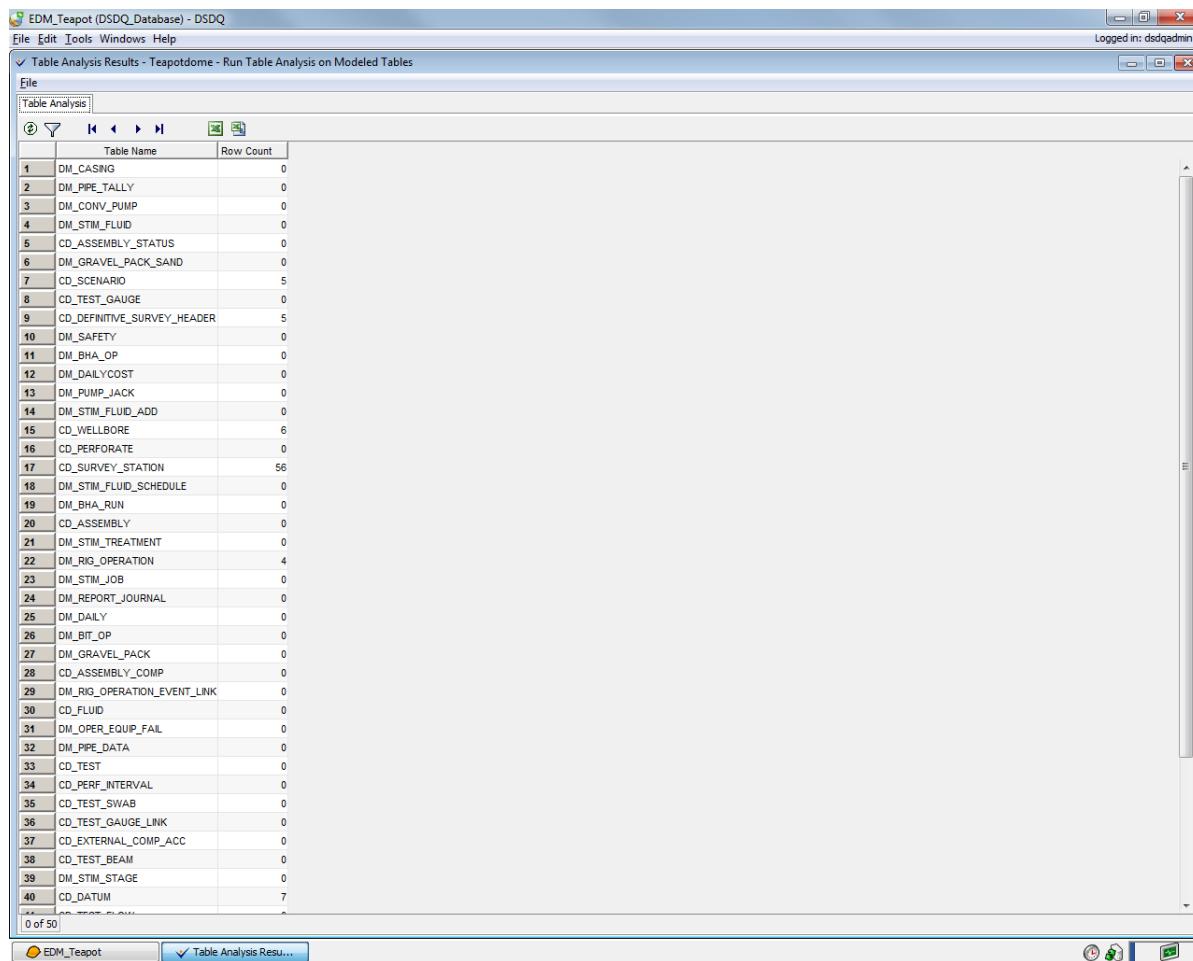


2. Enter **Teapotdome** in the **Job Name** field.
3. Enter **Table Analysis on Modeled Tables of EDM_Teapot** in the **Job Description** field.
4. Select **EDM_Teapot** from the **Select a Connection Source** drop-down list.

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.



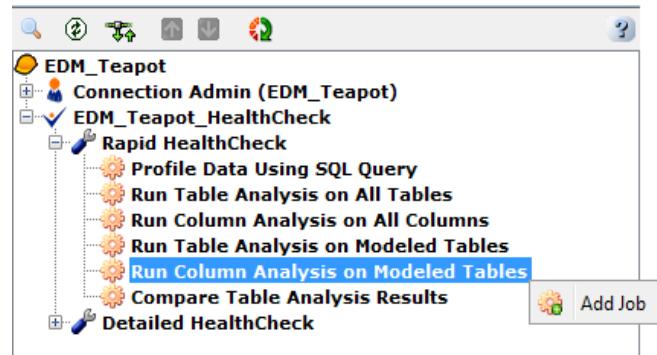
12. Select **File > Exit** to close the **Basic View Frame** window.

Exercise: Running Column Analysis on Modeled Tables Task

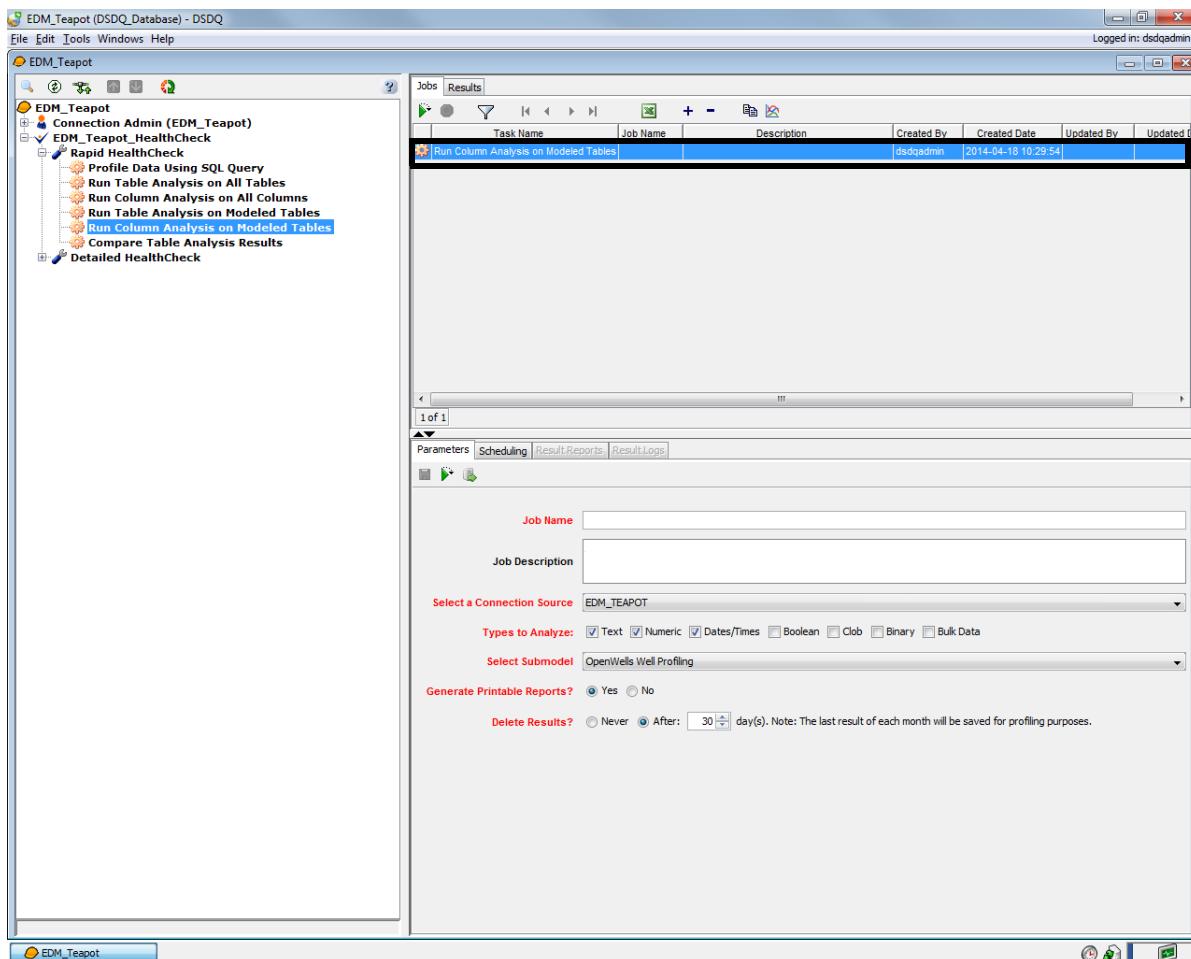
The **Run Column Analysis on Modeled Tables** Task runs only on tables that have been modeled in the **Perform Table Modeling** Tool (reference: *DecisionSpace Data Quality Training Manual, Chapter 4: Data Evaluation, Perform Table Modeling*). To run Column Analysis on all the modeled tables:

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**.

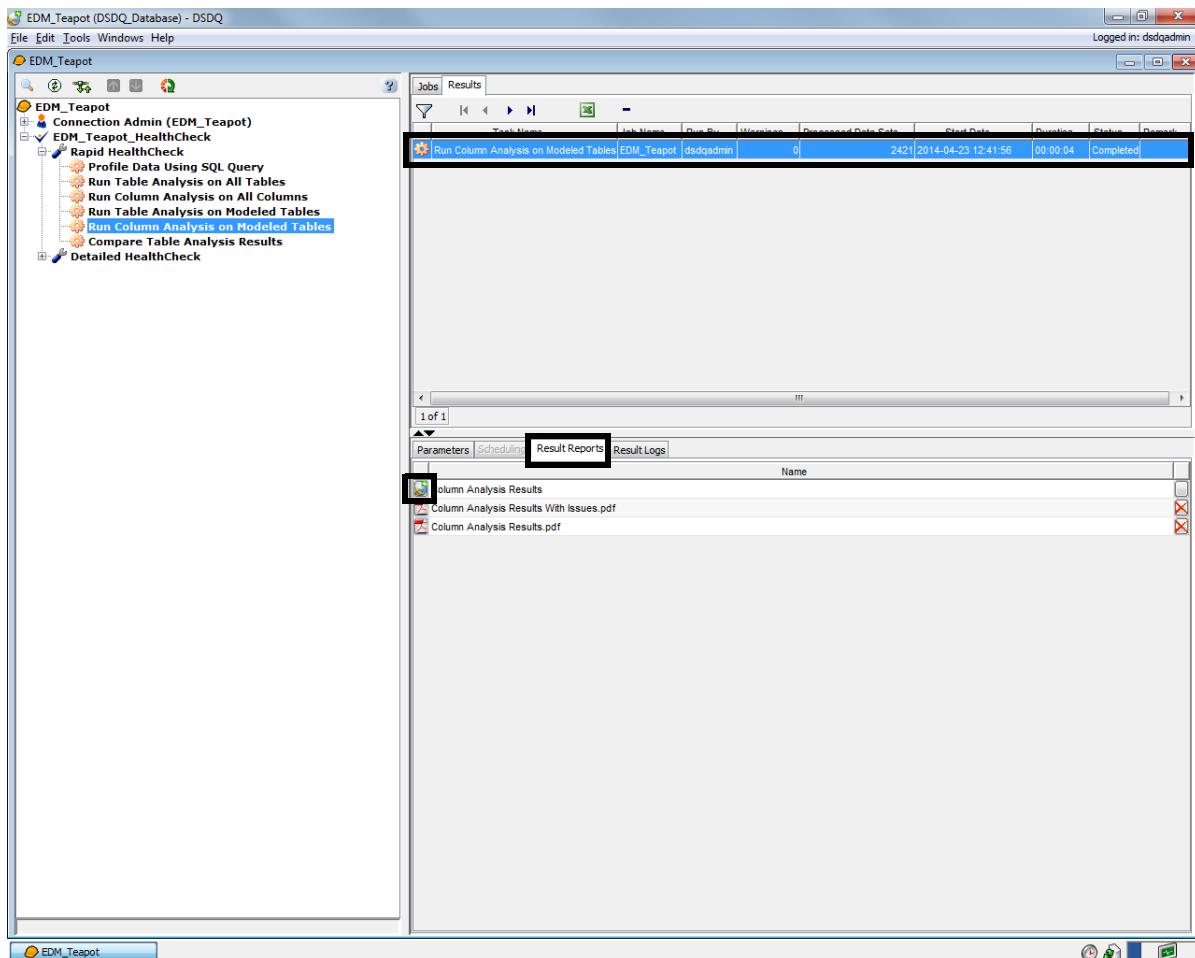
8. Click  to save changes in the **Parameters** tab.

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.

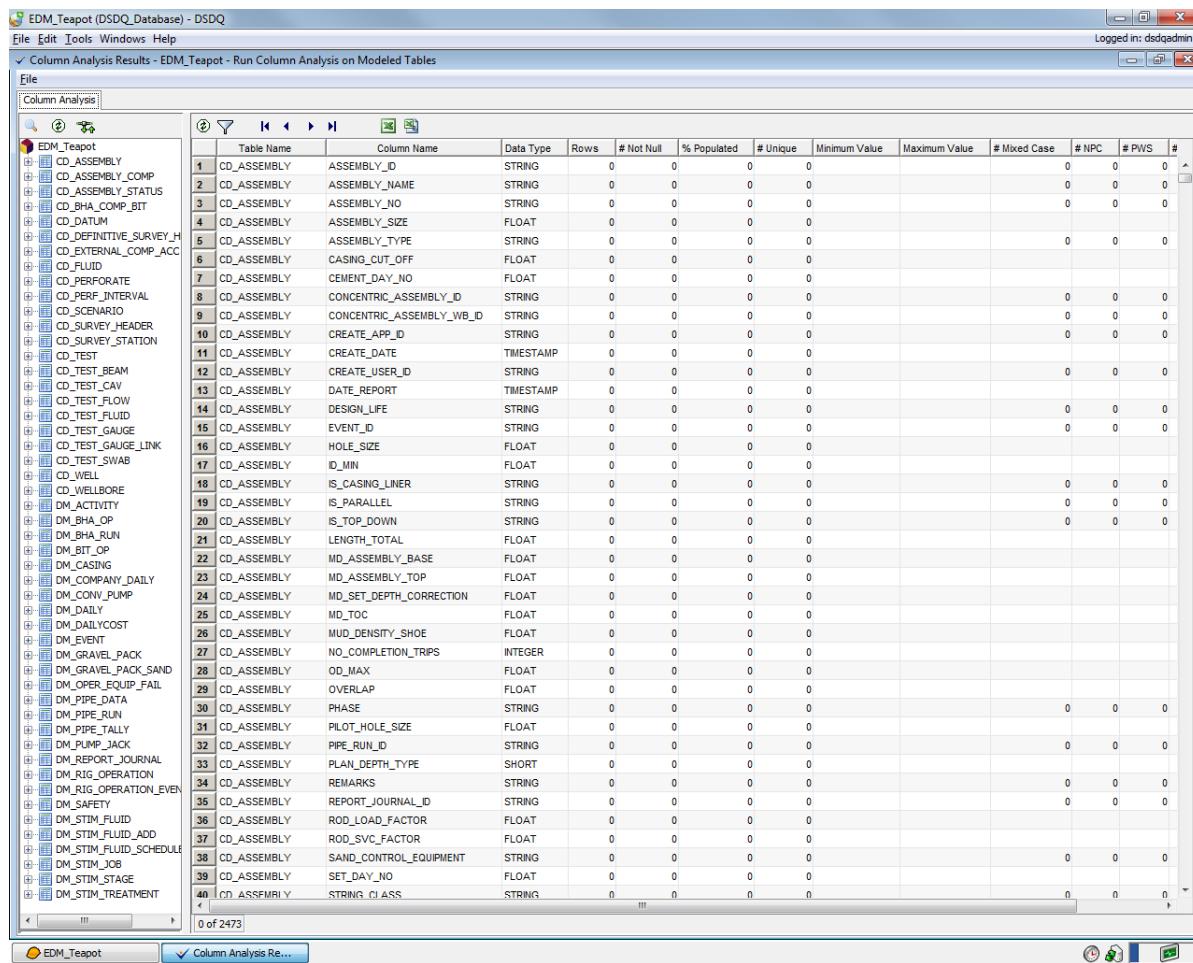


	Table Name	Column Name	Data Type	Rows	# Not Null	% Populated	# Unique	Minimum Value	Maximum Value	# Mixed Case	# NPC	# PWS	#
1	CD_ASSEMBLY	ASSEMBLY_ID	STRING	0	0	0	0			0	0	0	
2	CD_ASSEMBLY	ASSEMBLY_NAME	STRING	0	0	0	0			0	0	0	
3	CD_ASSEMBLY	ASSEMBLY_NO	STRING	0	0	0	0			0	0	0	
4	CD_ASSEMBLY	ASSEMBLY_SIZE	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

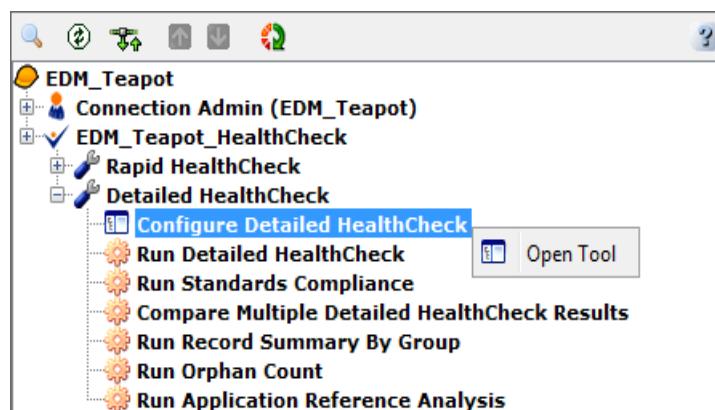
The **Detailed HealthCheck Activity** allows you to run business rules against the dataset to identify data problems.

Exercise: Configuring the Detailed HealthCheck Tool

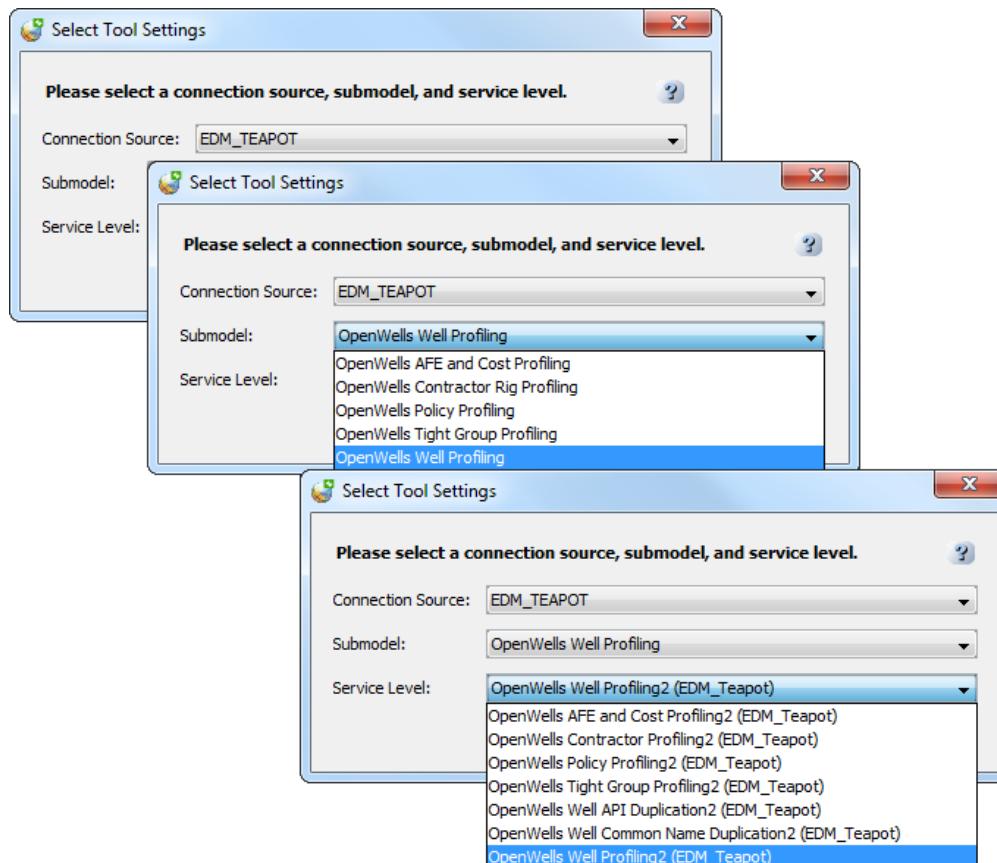
The **Configure Detailed HealthCheck Tool** configures service levels for testing prior to running the **Run Detailed HealthCheck Task**. You can select which requirements are to be enabled/disabled in the service level. You can also select subset of the total data to be used when testing a service level.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

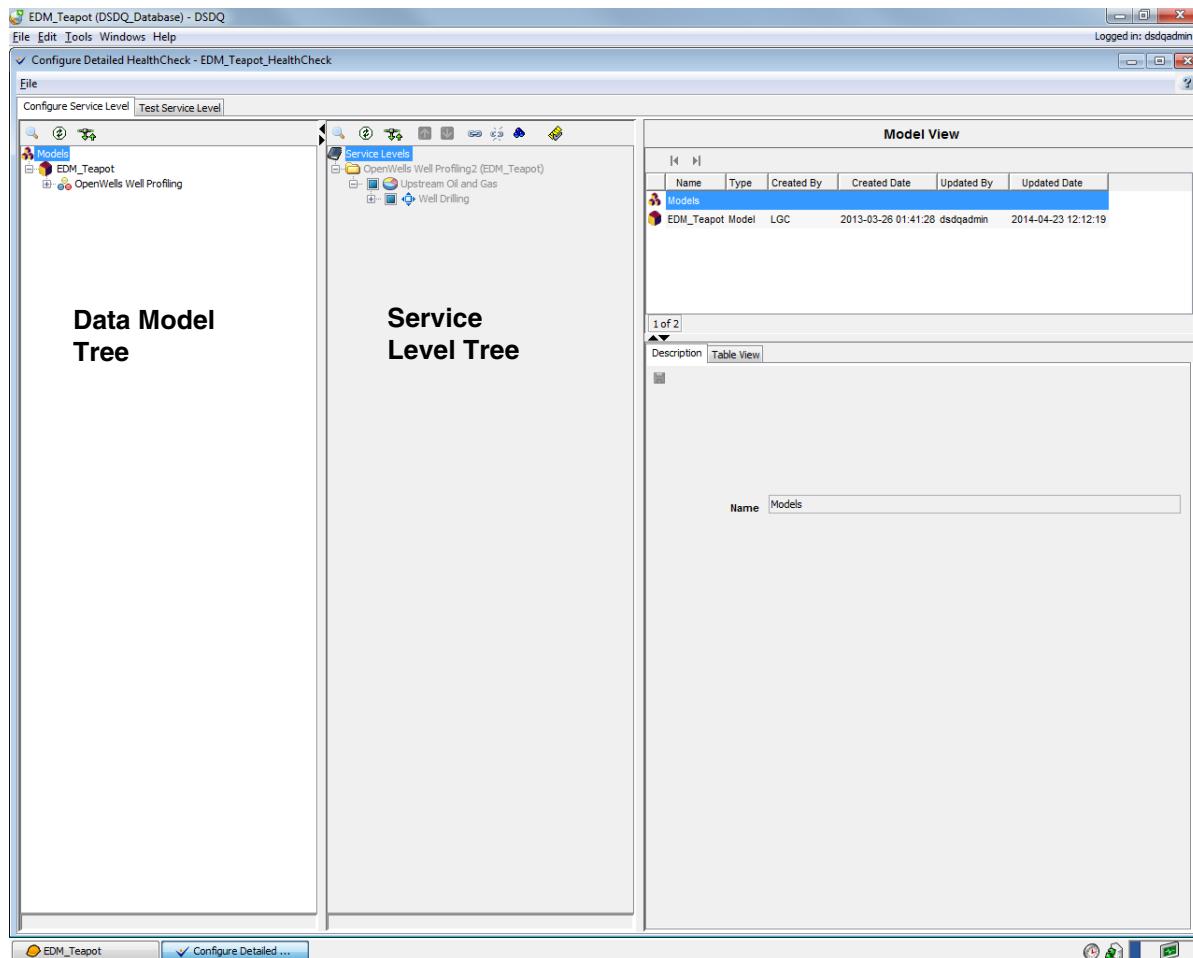


The **Select Tool Settings** window appears.



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**.

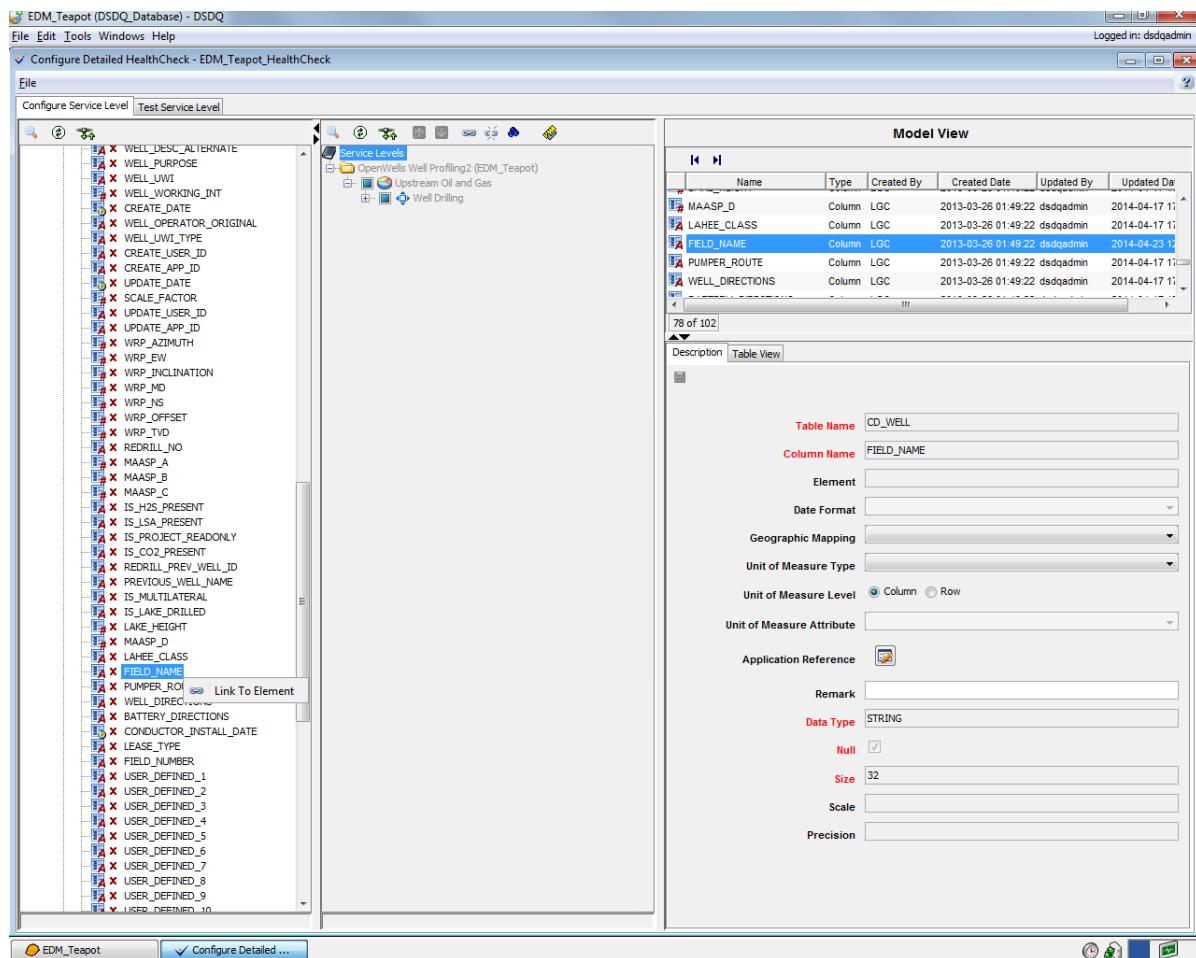
The **Configure Detailed HealthCheck** window appears.



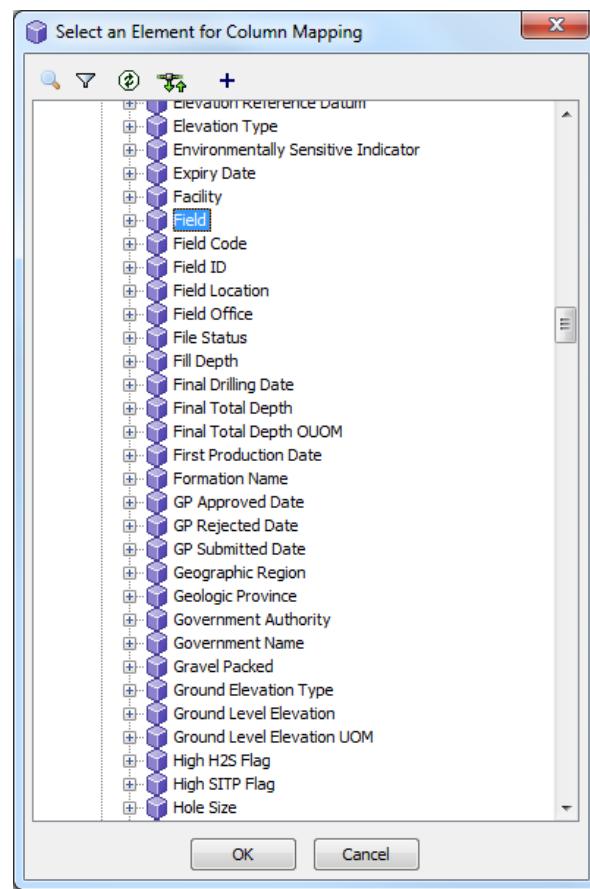
The Data Model Tree populates with the submodel selected in step 4. The Service Level Tree populates with the service level selected in step 5.

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.

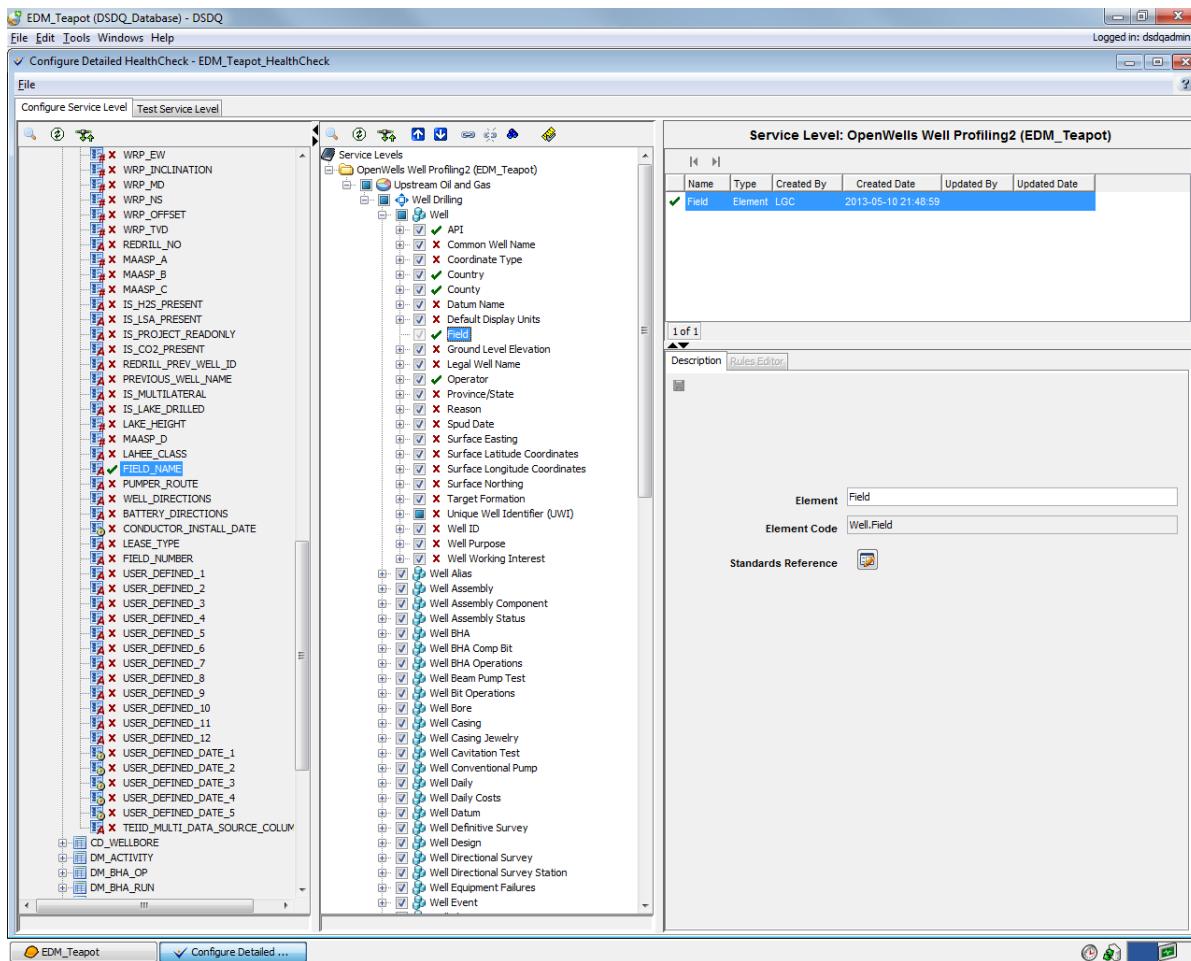


10. Select the **Field** element.

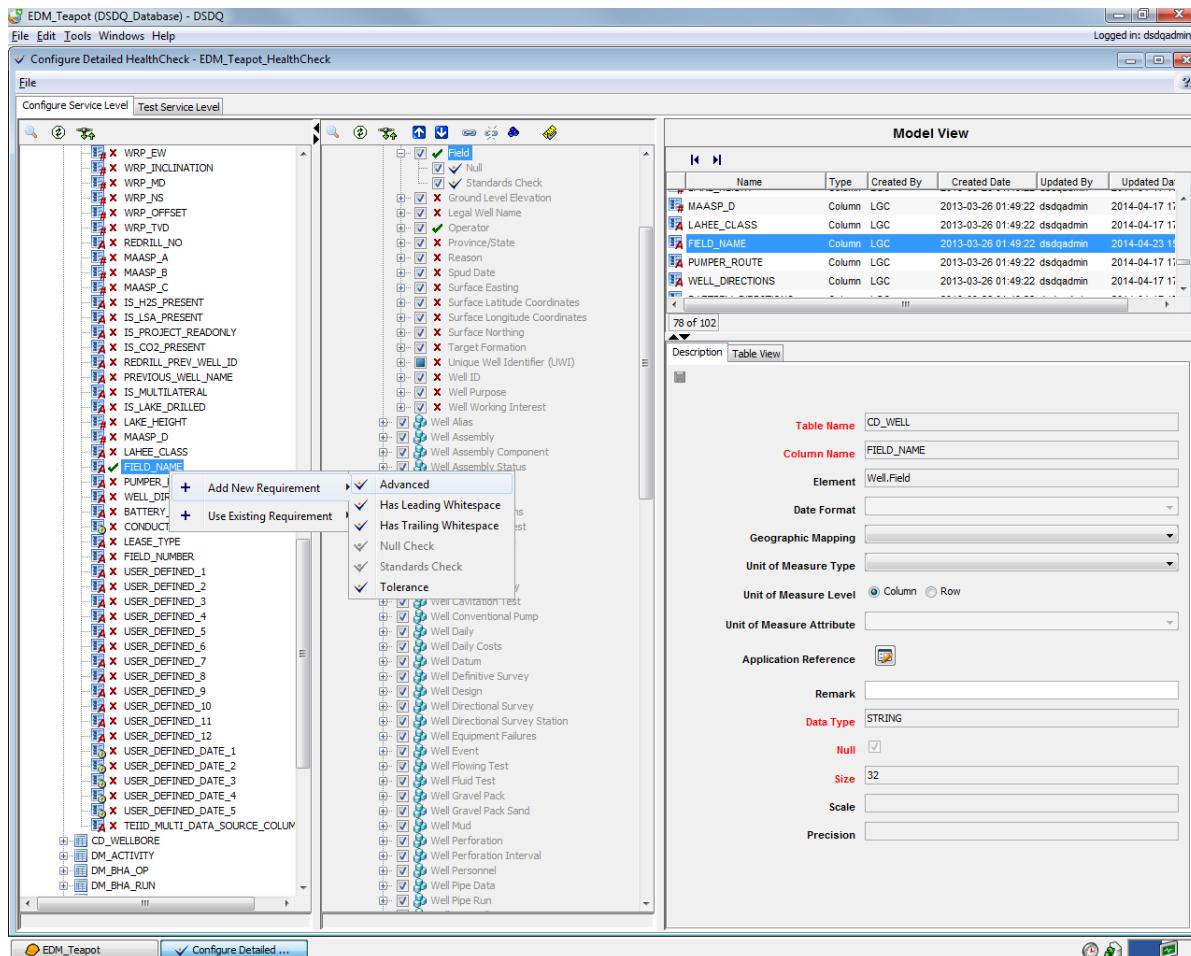
11. Click **OK**.

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.

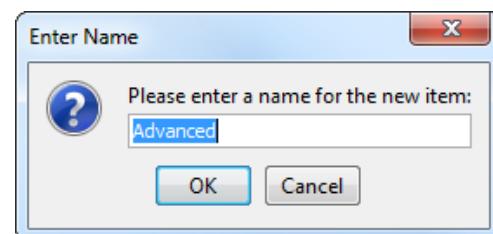
However, it is possible to link different tables' columns to the same element.



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



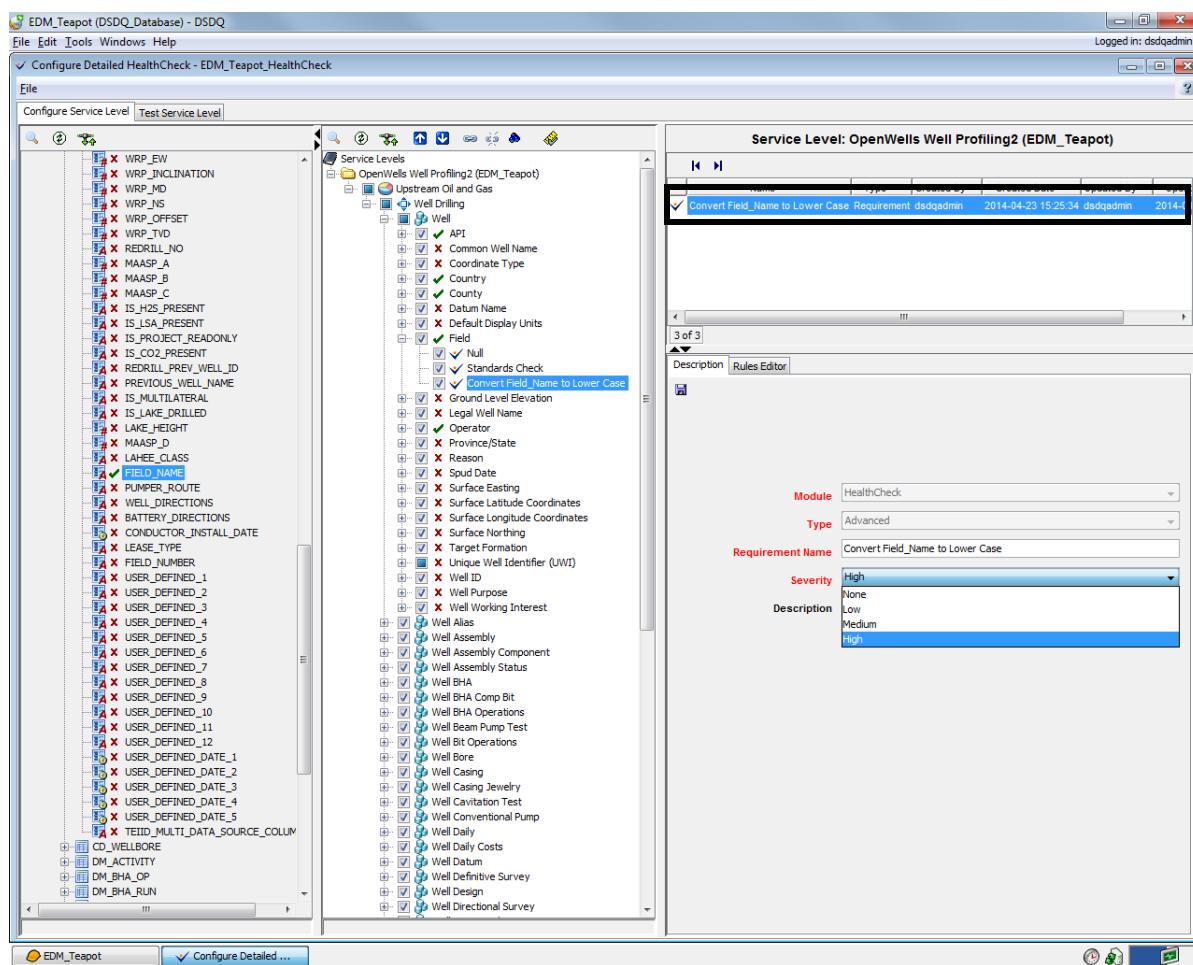
The Enter Name dialog box appears.



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

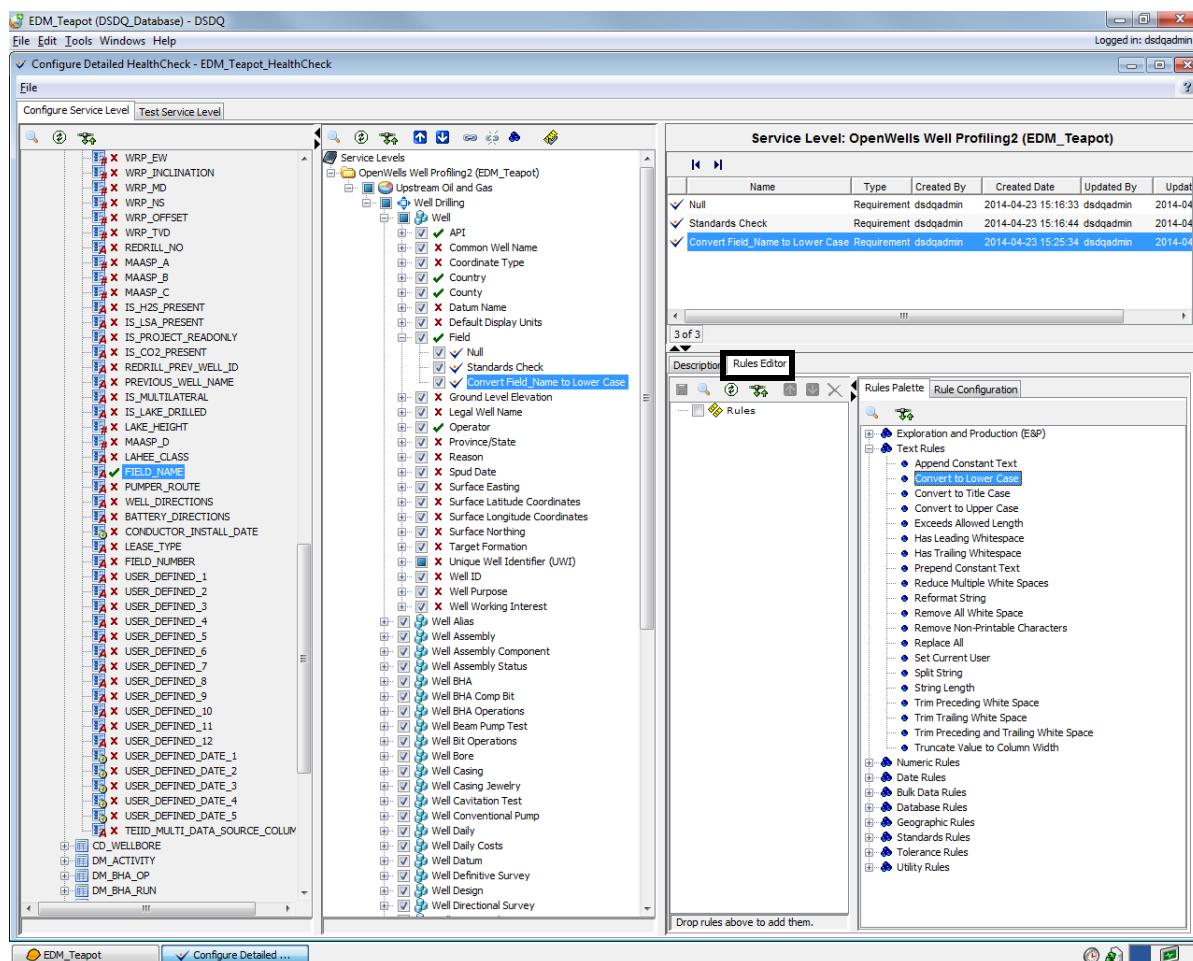


14. Click **OK** to add the requirement to the selected column.
 The **Convert Field_Name to Lower Case** requirement is added to the **Field** element.



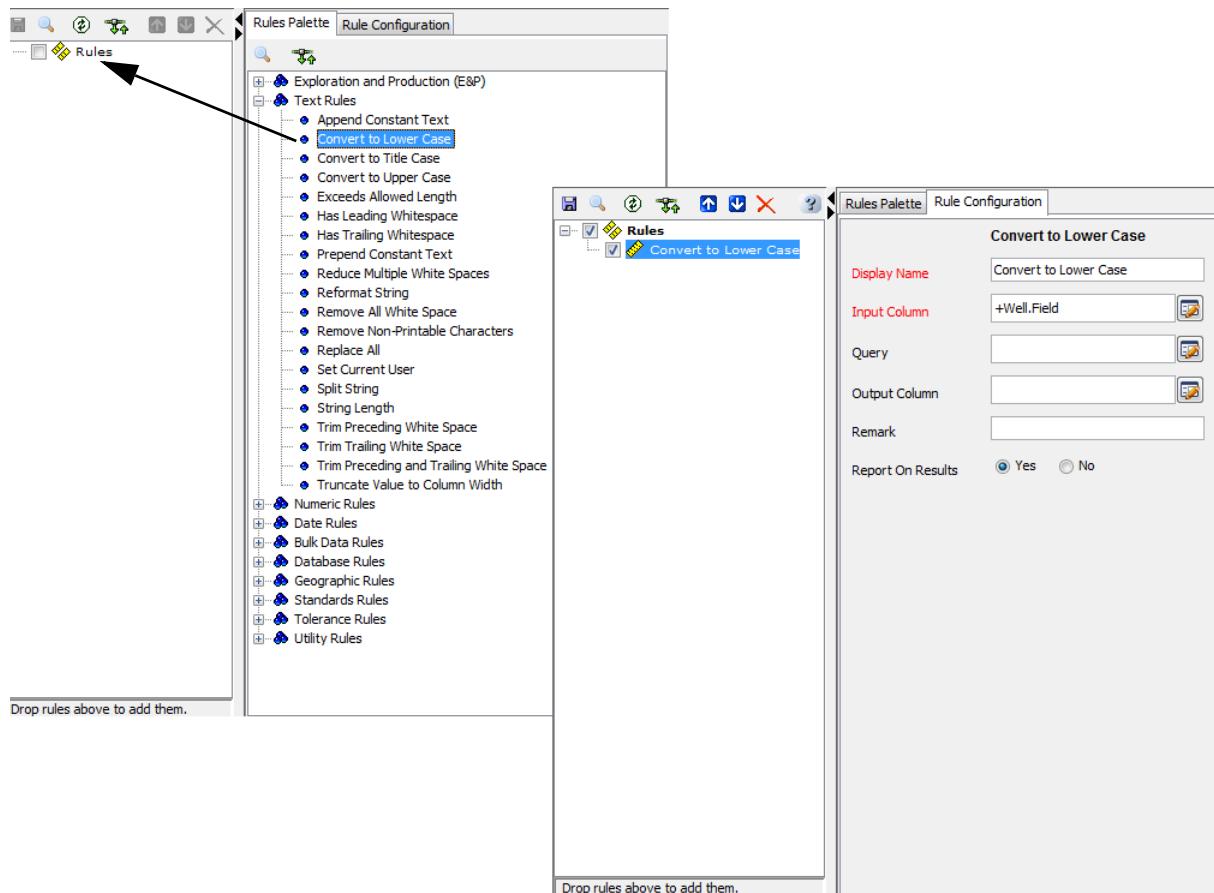
15. The **Module** and **Type** fields are disabled by default.
16. The **Requirement name** field populates with the name assigned in step **13**.
17. Select **High** from the **Severity** drop-down list.
18. Enter **Advance Rule for FIELD_NAME column** in the **Description** field.
19. Click to save 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.

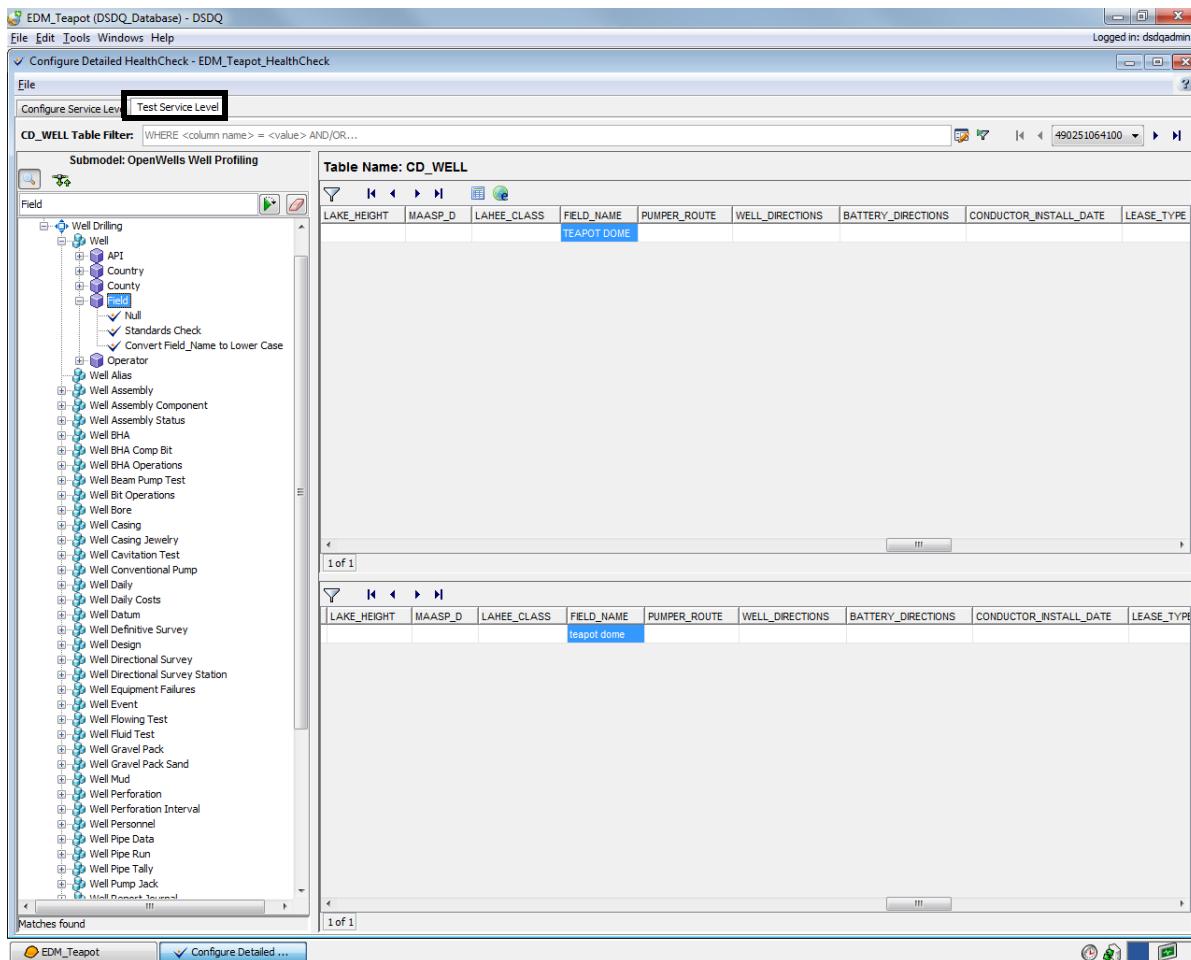
23. Drag and drop the **Convert to Lower Case** rule onto the **Rule** area.



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.

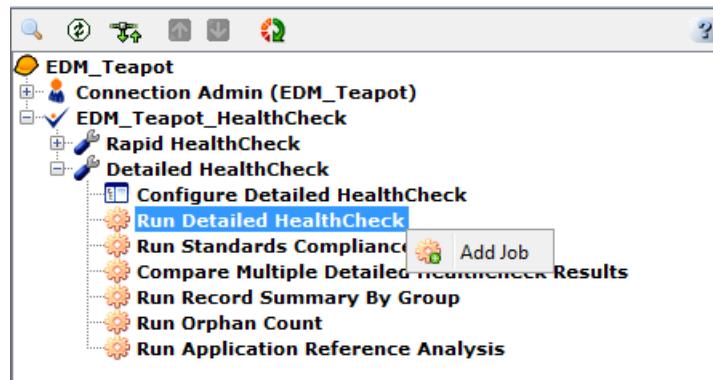
26. Click the **Next Data Set** button to test the next record.
27. Repeat step **26** to test all records.
28. Select **File > Exit** to close the **Configure Detailed HealthCheck** window.

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

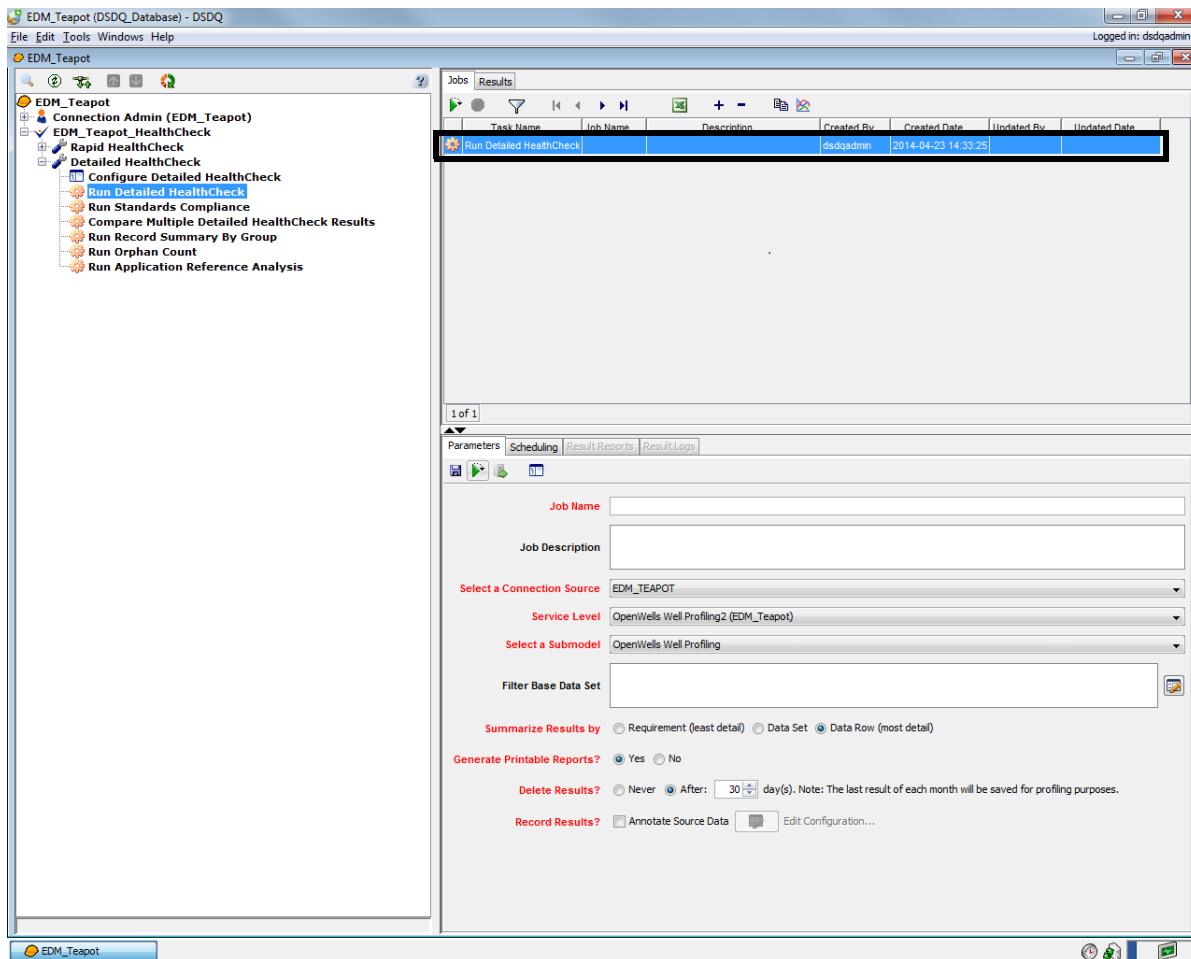
submodel table have been assigned elements from the desired service level. To run the Detailed HealthCheck task:

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**.

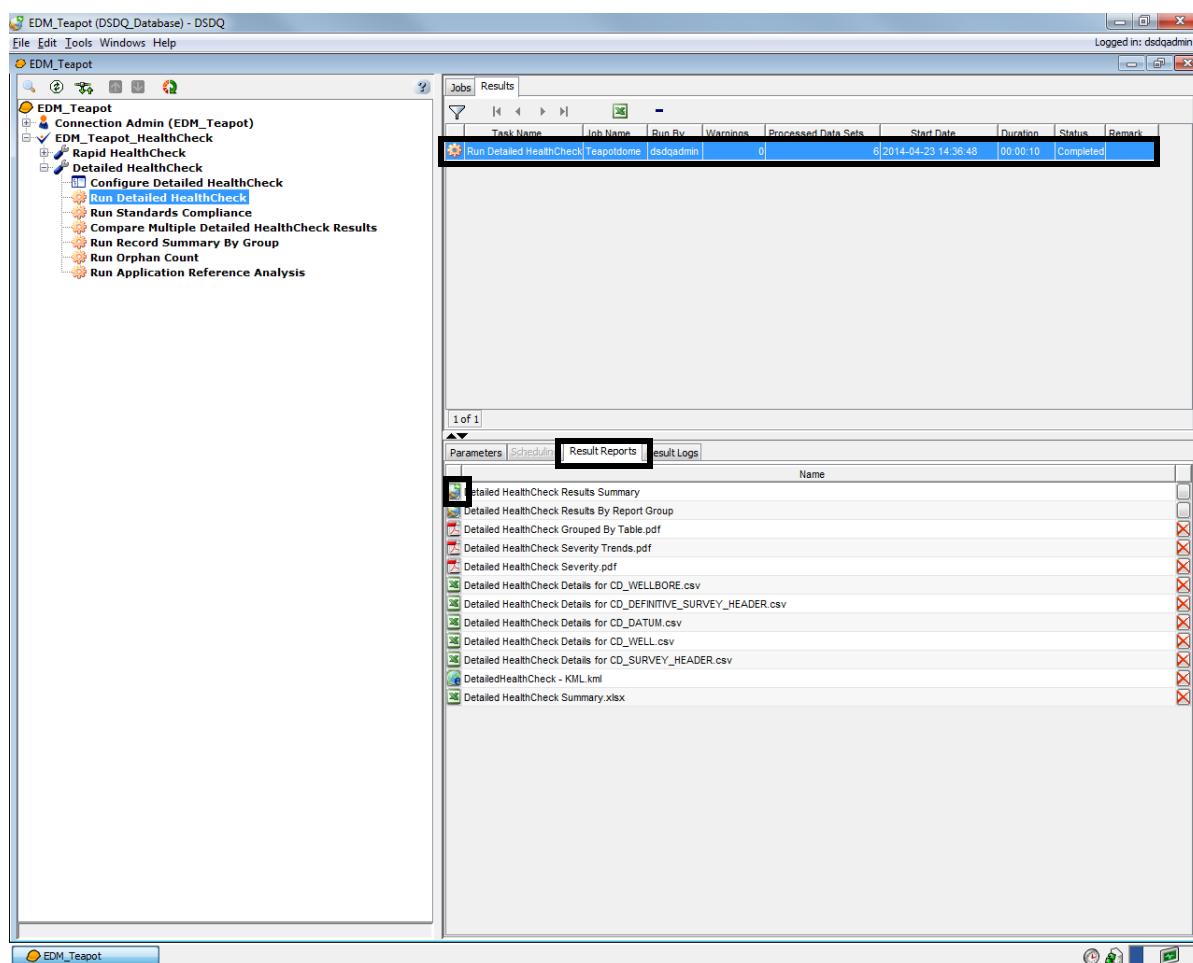
12. Click to save changes in the **Parameters** tab.

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.

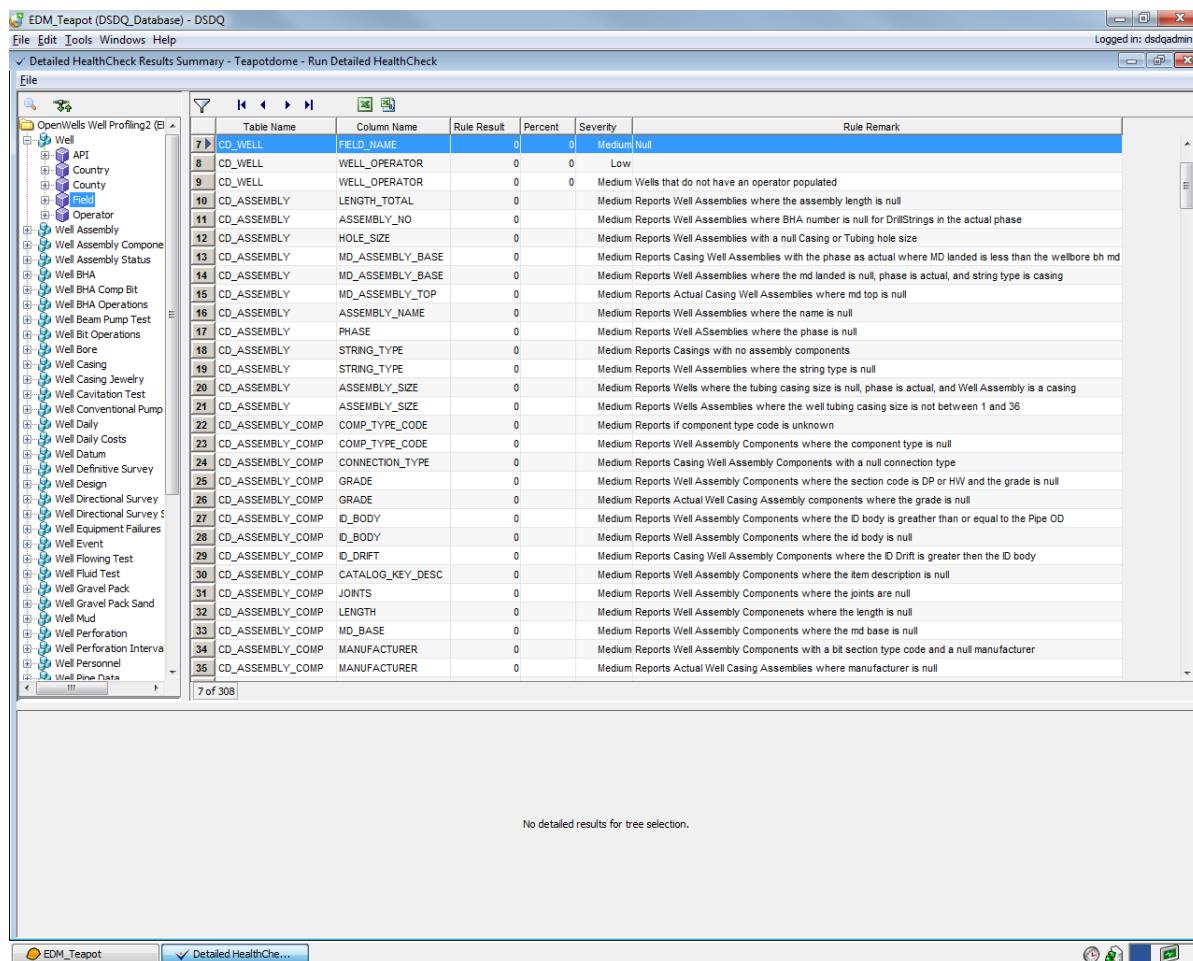


Table Name	Column Name	Rule Result	Percent	Severity	Rule Remark
CD_WELL	FIELD_NAME	0	0	Medium	Null
CD_WELL	WELL_OPERATOR	0	0	Low	Medium Wells that do not have an operator populated
CD_WELL	WELL_OPERATOR	0	0	Medium	Reports Well Assemblies where the assembly length is null
CD_ASSEMBLY	LENGTH_TOTAL	0	0	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	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	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	Actual Well Casing Assemblies where manufacturer is null

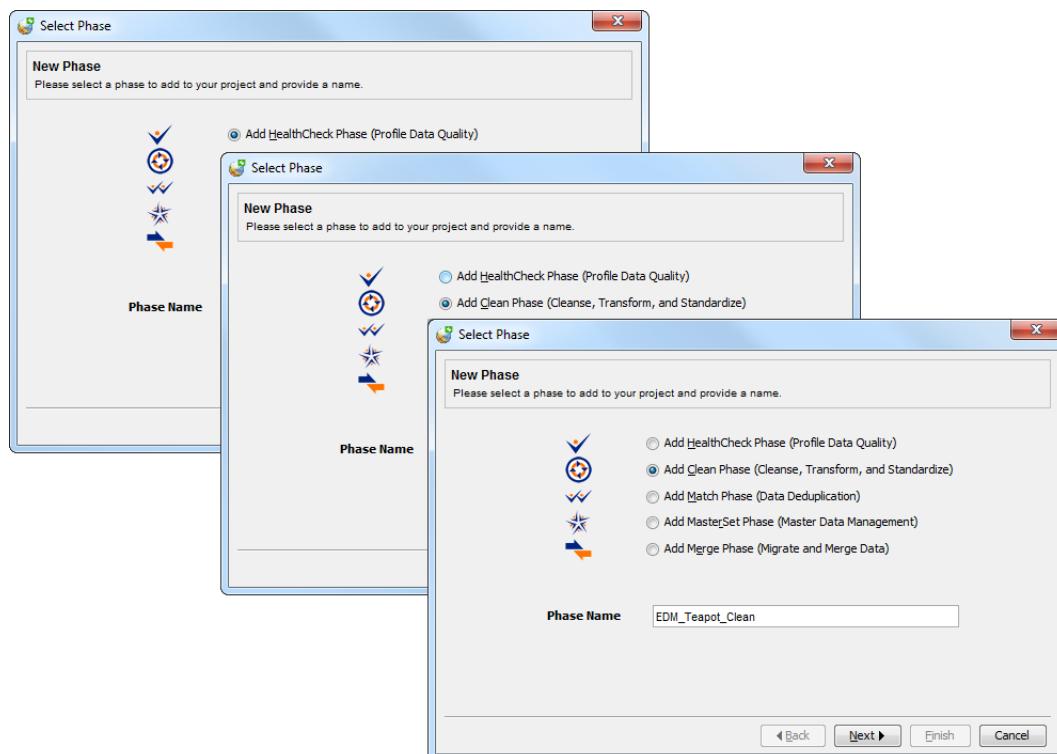
16. Select **File > Exit** to close the **Basic View Frame** window.

Resolving Data Quality Issues using the Clean Phase

Adding a Clean Phase

To add a Clean Phase:

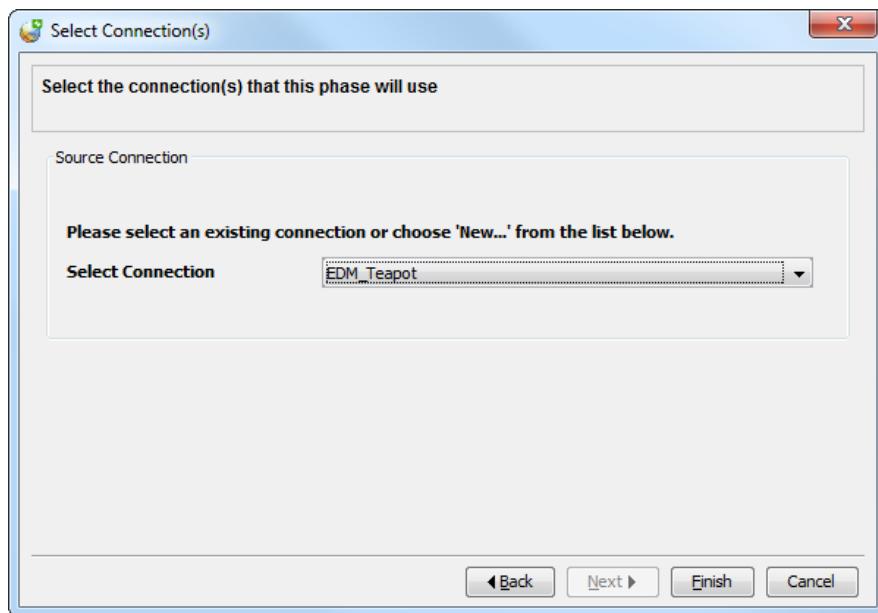
1. Click the **Add New Phase** button on the Project Toolbar. The **Select Phase** window appears with the **Add HealthCheck Phase (Report on Data Quality Profiling)** option selected by default.



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.

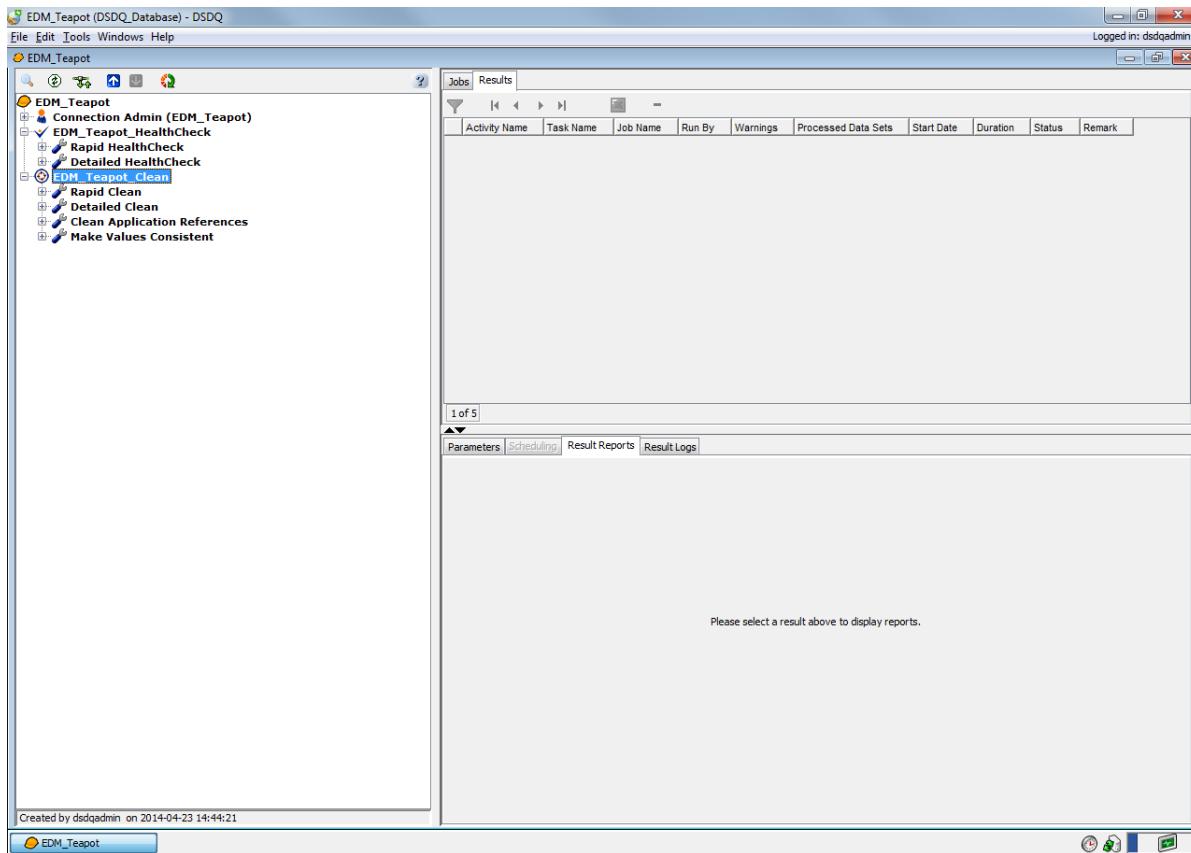
The **Select Connection(s)** window appears.



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.



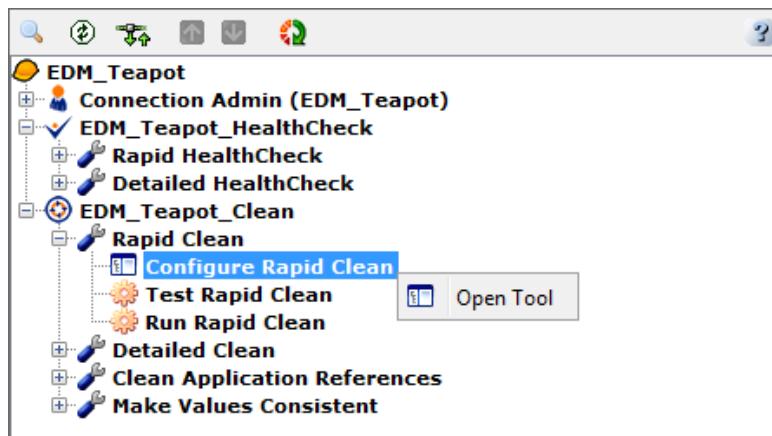
Rapid Clean Activity

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

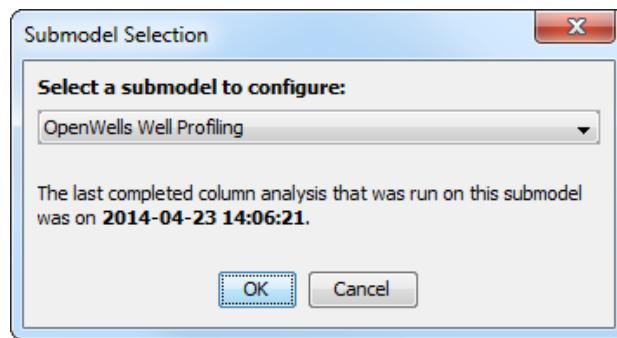
Exercise: Configuring the Rapid Clean Tool

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



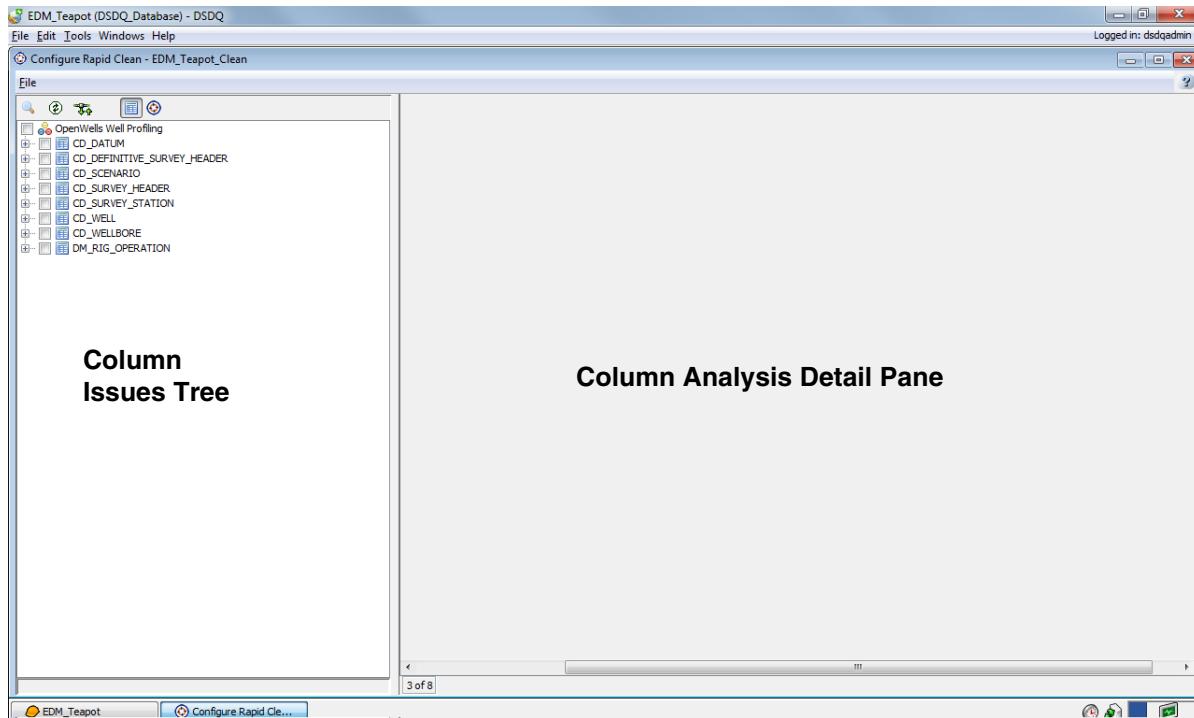
The **Submodel Selection** dialog box appears.



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.

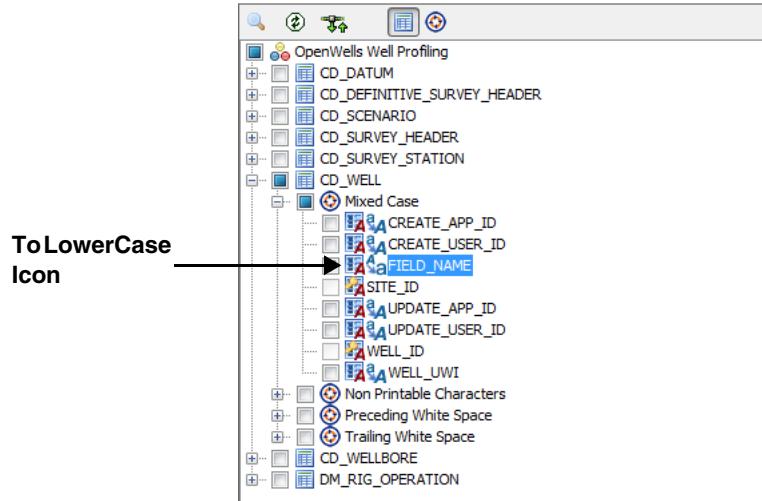
6. Click to expand the **Mixed Case** issues.

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.

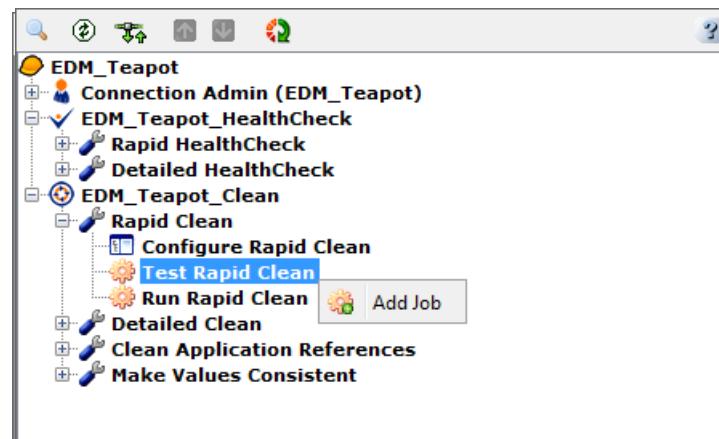


10. Select the **Preceding White Spaces** requirement.
11. Select **File > Exit** to close the **Configure Rapid Clean** window.

Exercise: Running the Test Rapid Clean Task

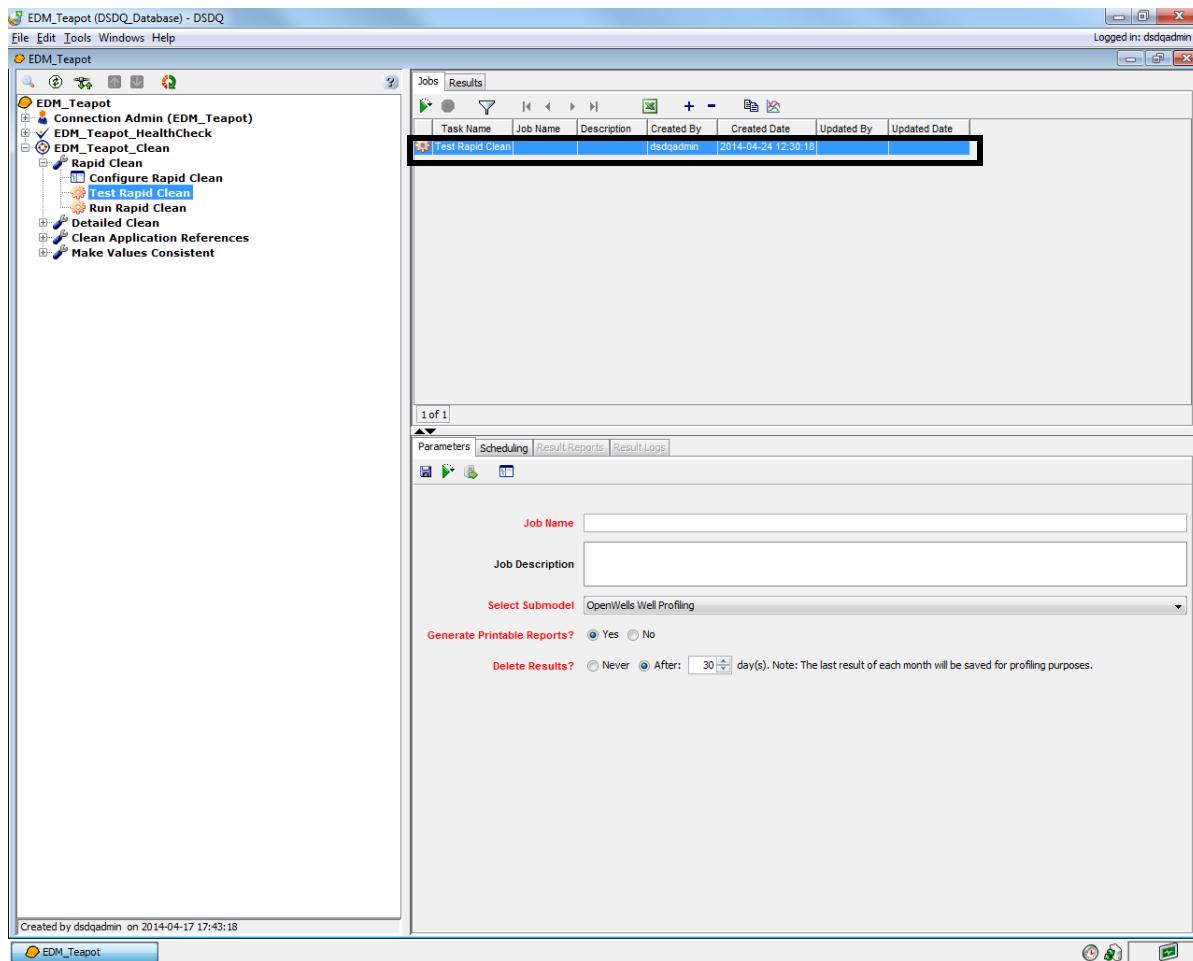
After issues have been selected to be cleaned using the **Configure Rapid Clean Tool**, the **Test Rapid Clean Task** is run to make sure that the expected results are seen before running the **Run Rapid Clean Task** to fix the entire dataset in the submodel. To run the **Test Rapid Clean Task**:

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.



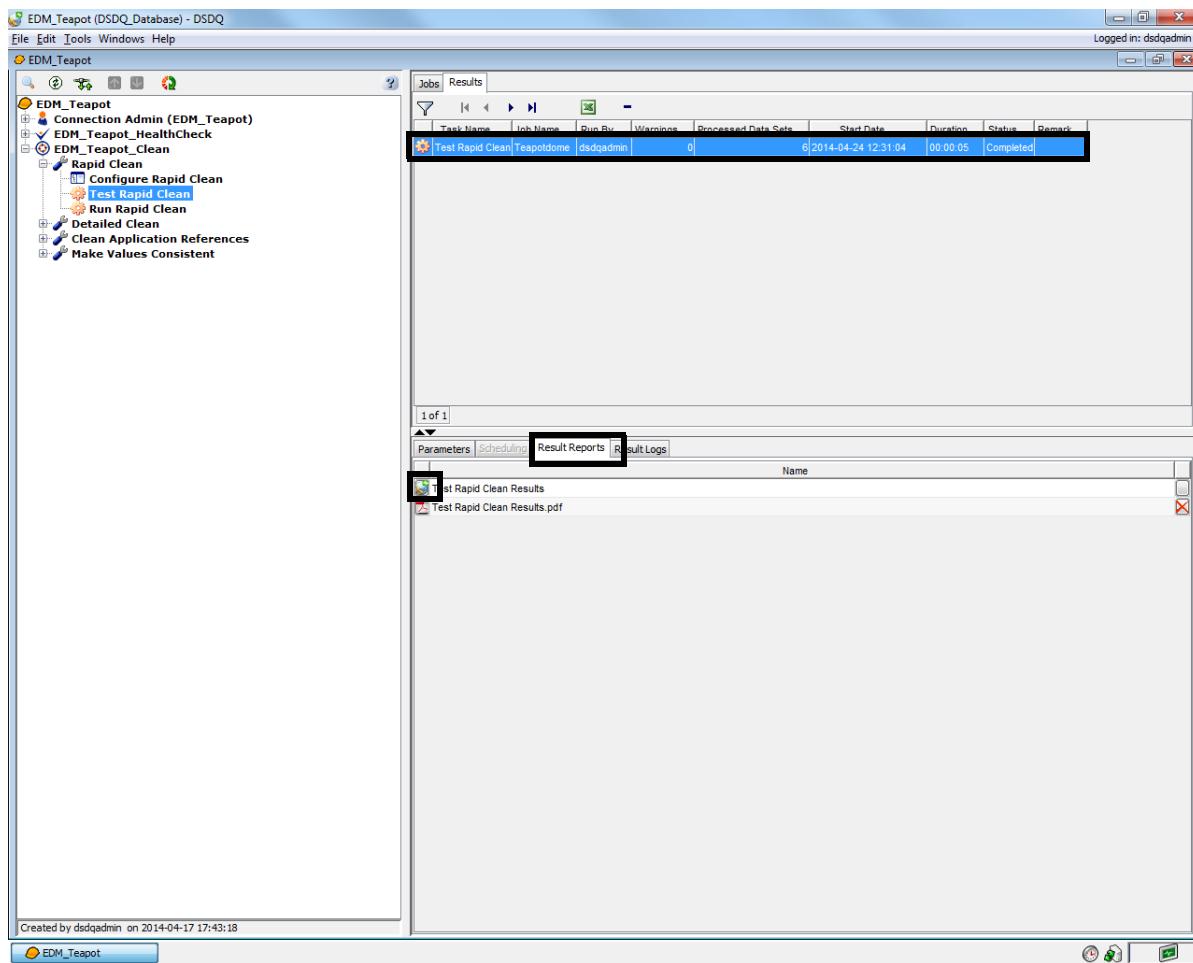
2. Enter **Teapotdome** in the **Job Name** field.
3. Enter **EDM_Teapot FIELD_NAME Name Conversion** in the **Job Description** field.
4. Select **OpenWells Well Profiling** from the **Select Submodel** drop-down list.
5. Select the **Yes** option for **Generate Printable Reports**.
6. Select the **After** 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.

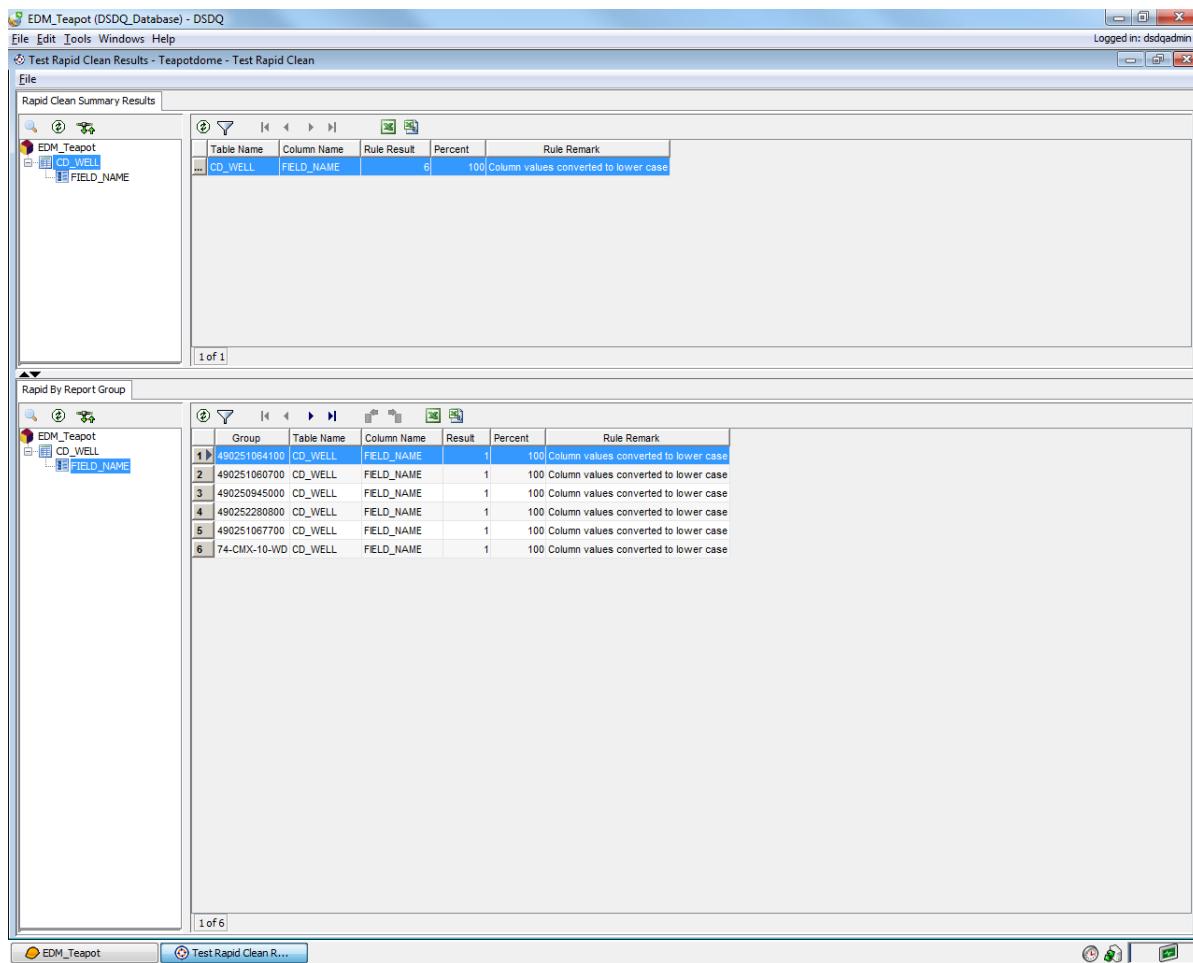
The **Test Rapid Clean** Task runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

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.

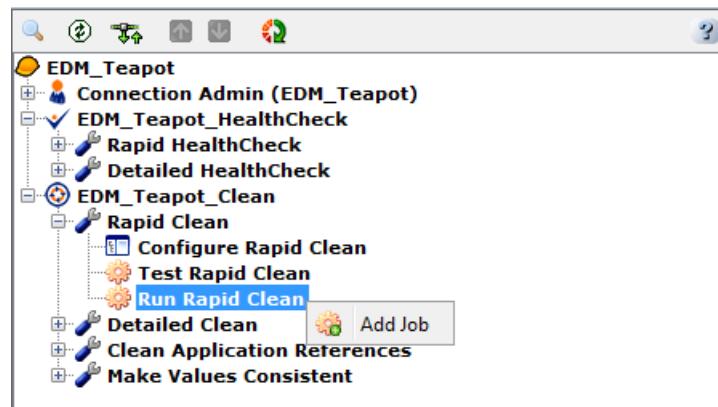


11. Select **File > Exit** to close the **Basic View Frame** window.

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. The 'Jobs' tab is selected, showing a table with one row:

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). The status bar at the bottom indicates 'Created by dsdqadmin on 2014-04-17 17:43:18'.

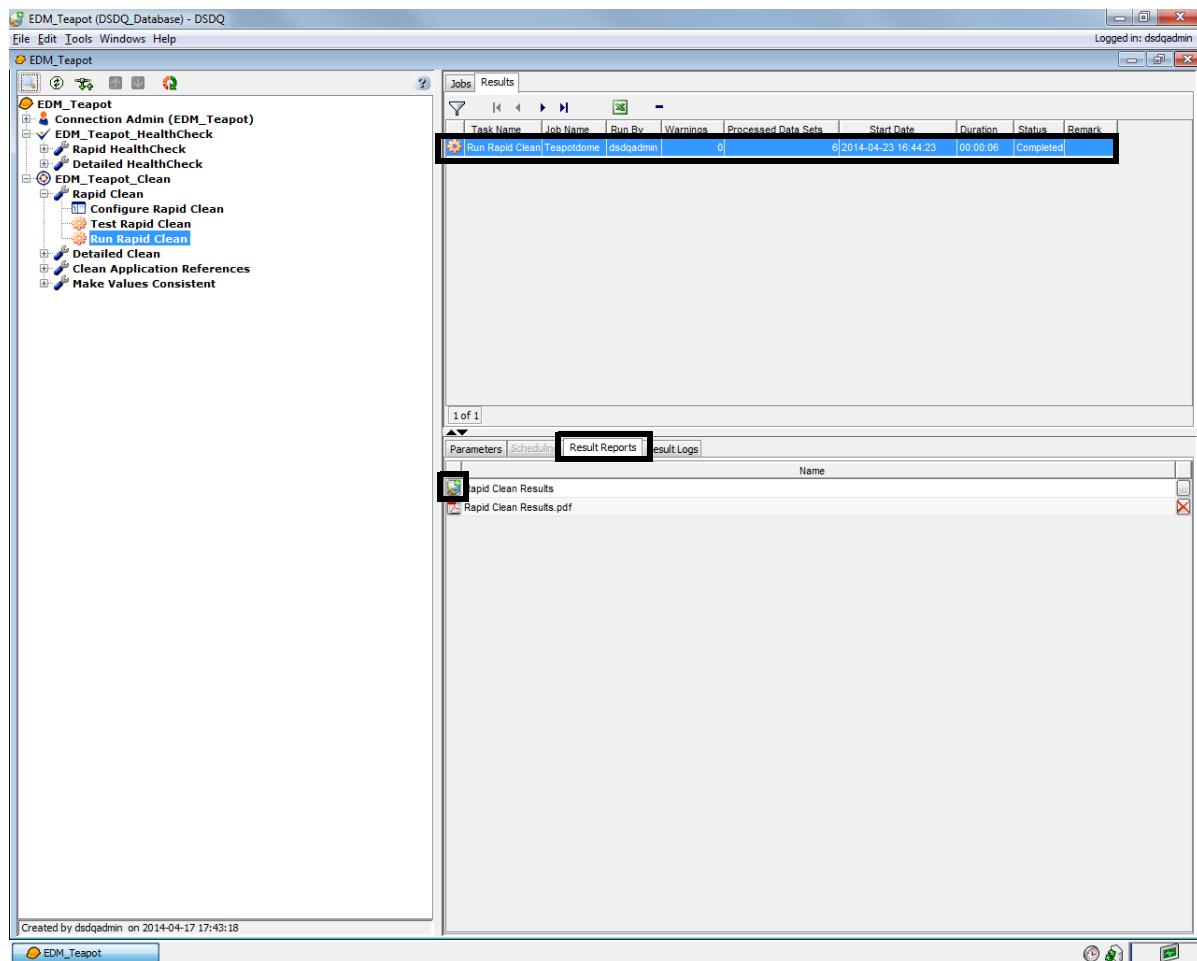
2. Enter **Teapotdome** in the **Job Name** field.
3. Enter **EDM_Teapot Case Convert** in the **Job Description** field.

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.

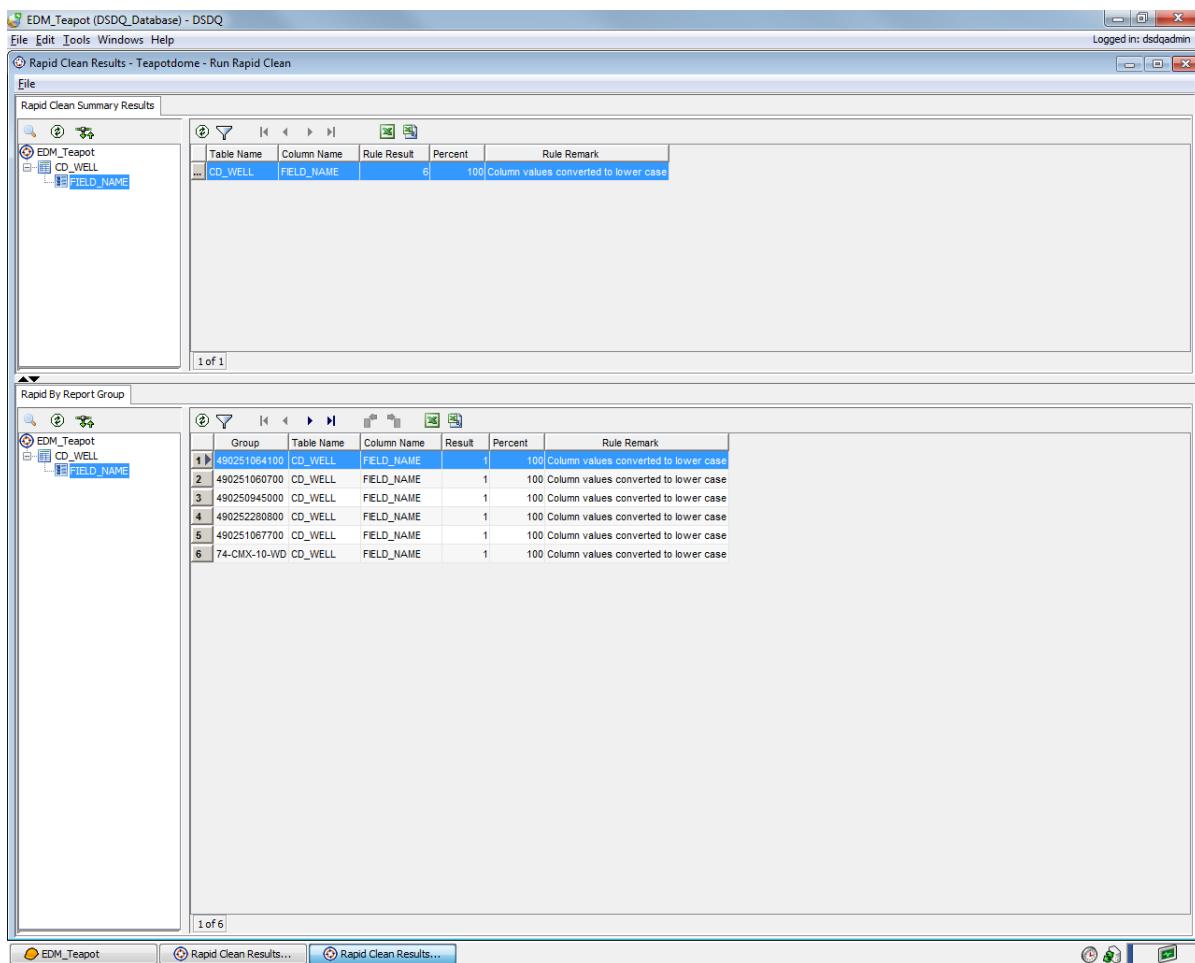
The **Run Rapid Clean Task** runs and displays results in the **Result Reports** tab on the **Jobs and Results Information Pane**.

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

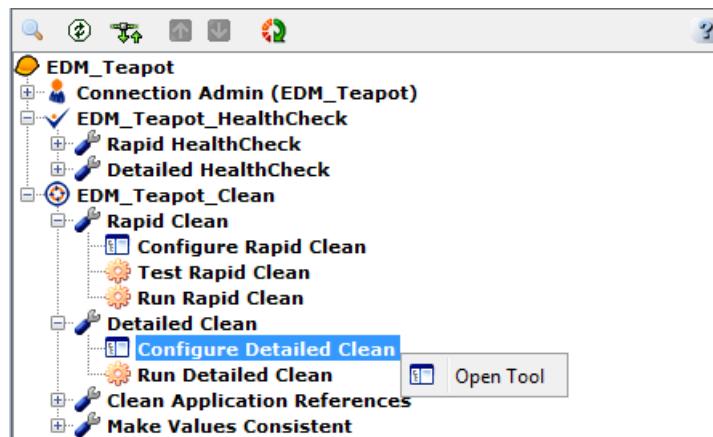
The **Detailed Clean** Activity helps in assigning columns to the clean requirements and testing service levels.

Exercise: Configuring the Detailed Clean Tool

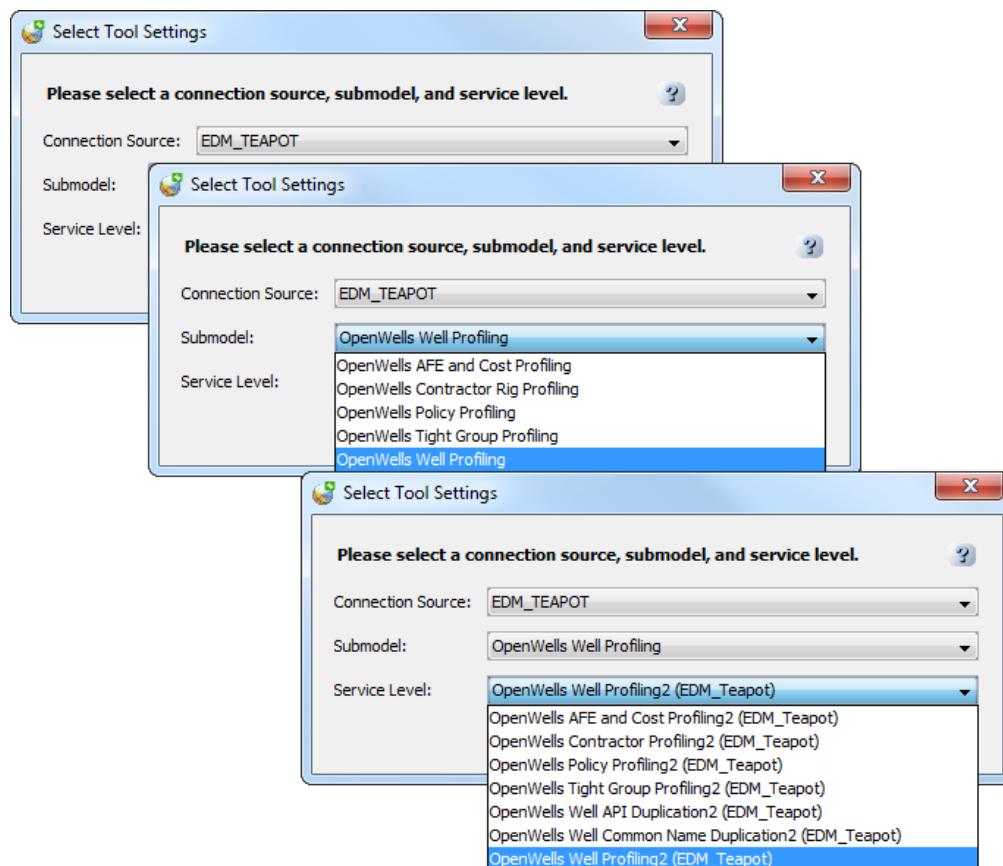
The **Configure Detailed Clean** Tool is used to configure service levels for testing prior to running the **Run Detailed Clean** Task. You can select which requirements in the service level to enable/disable, and when testing a service level, what subset of the total data to use. A

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.

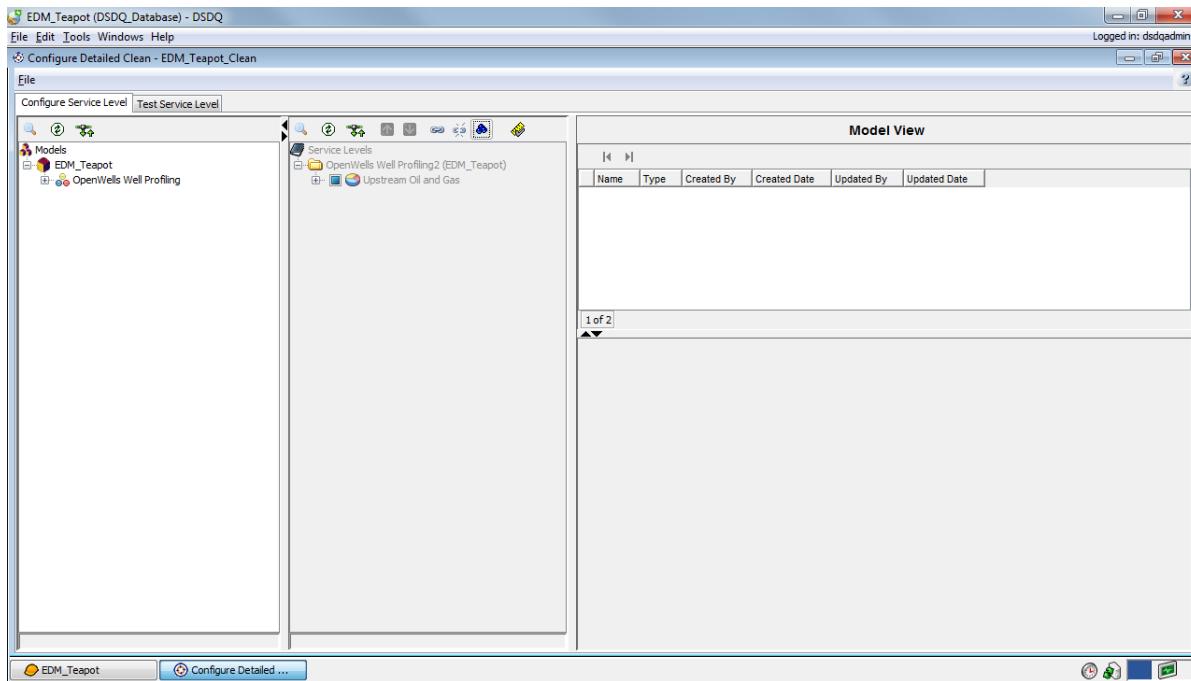


The **Select Tool Settings** window appears.

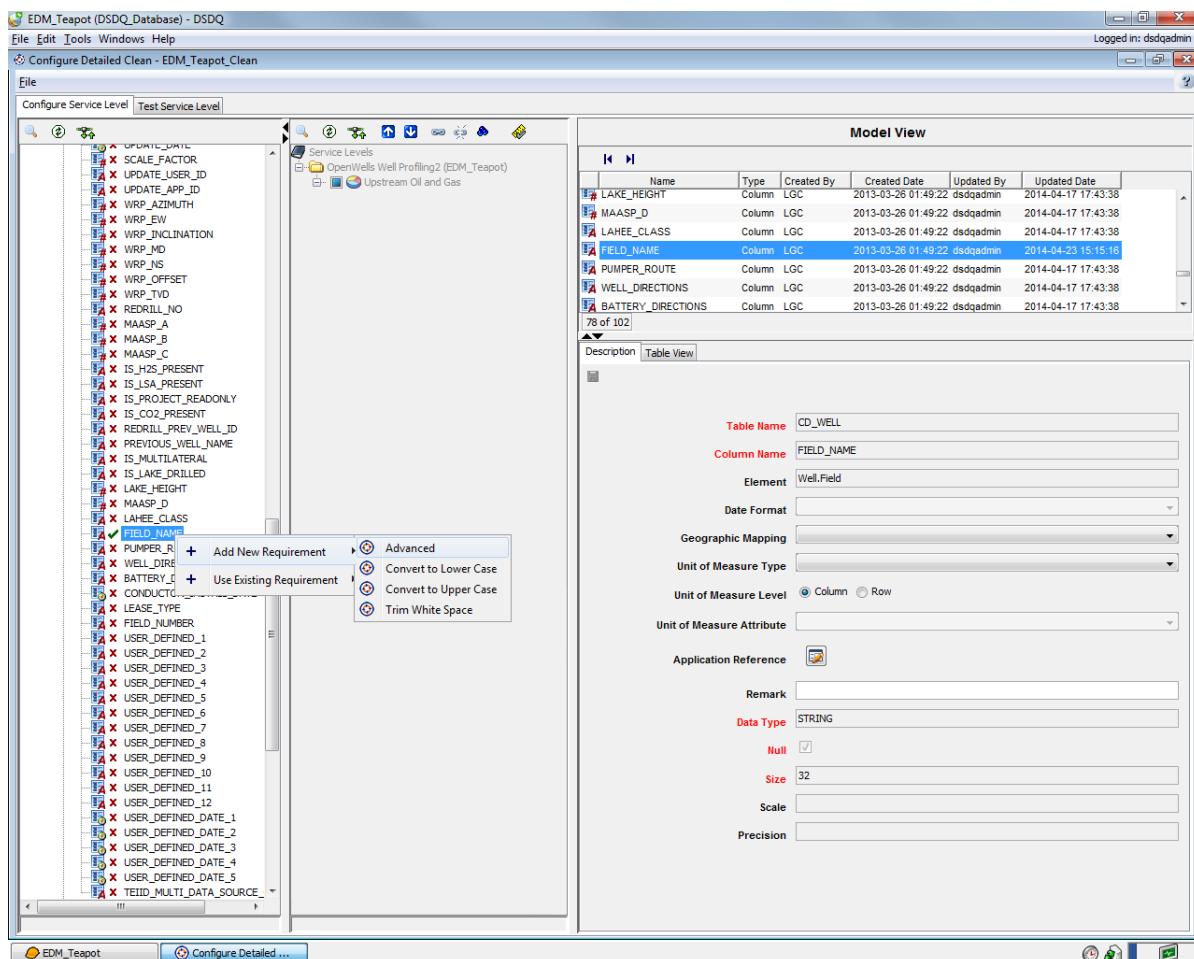


2. The **Connection Source** drop-down list is set to **EDM_Teapot** by default.
3. Select **OpenWells Well Profiling** from the **Submodel** drop-down list.
4. Select **OpenWells Well Profiling2 (EDM_Teapot)** from the **Service Level** drop-down list.
5. Click **OK**.

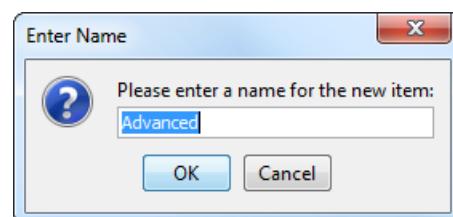
The **Configure Detailed Clean - EDM_Teapot_Clean** window appears.



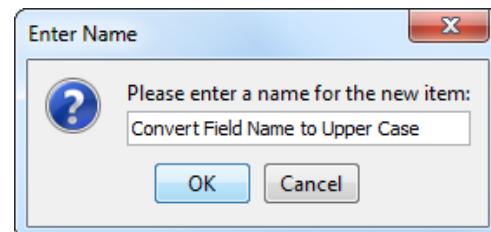
6. Click to expand the **OpenWells Well Profiling** submodel in the Data Model Tree.
7. Click to expand the **CD_WELL** table in the Data Model Tree.
8. Right-click the **FIELD_NAME** column in the Data Model Tree and select **Add New Requirement > Advanced** from the pop-up menu.



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.

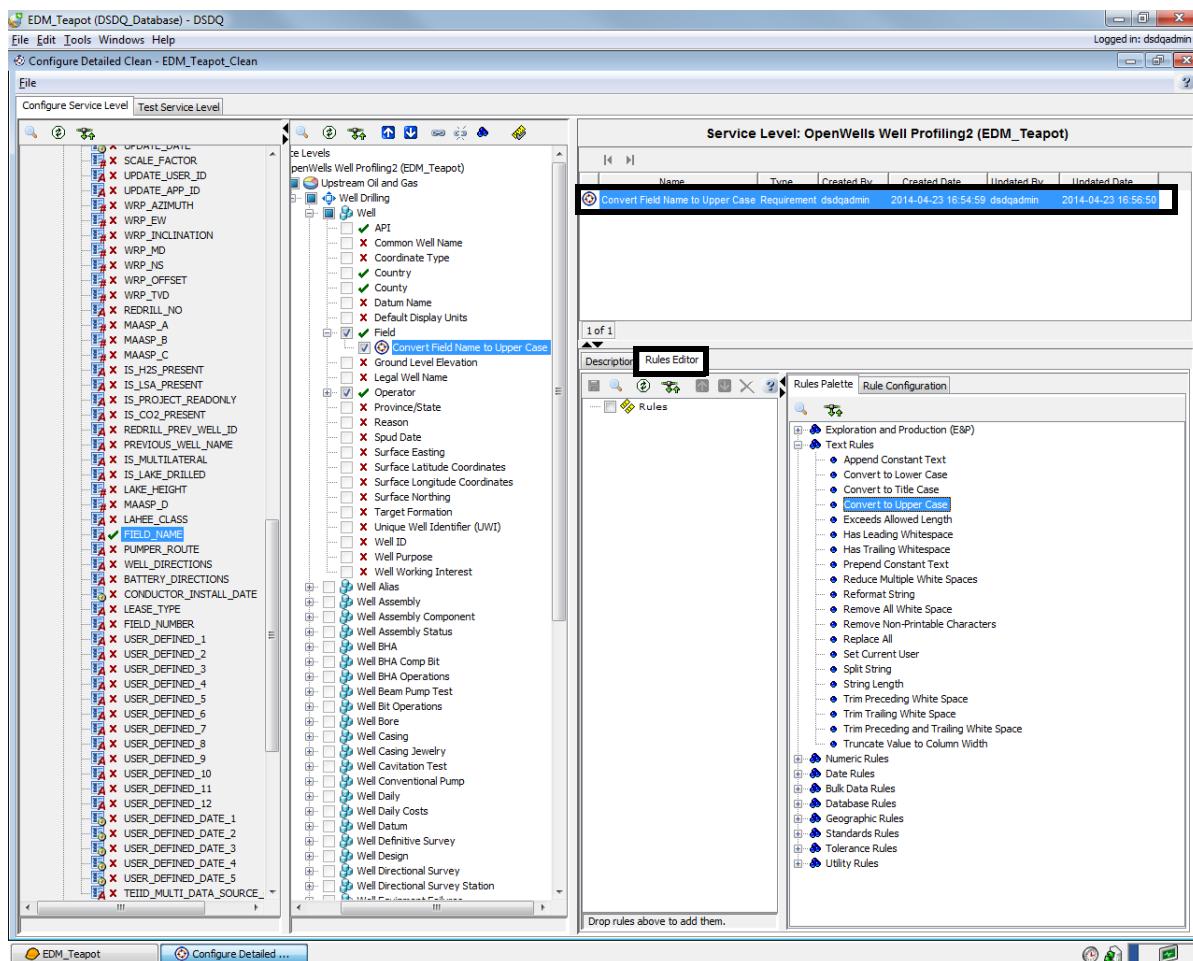


10. Click **OK**.

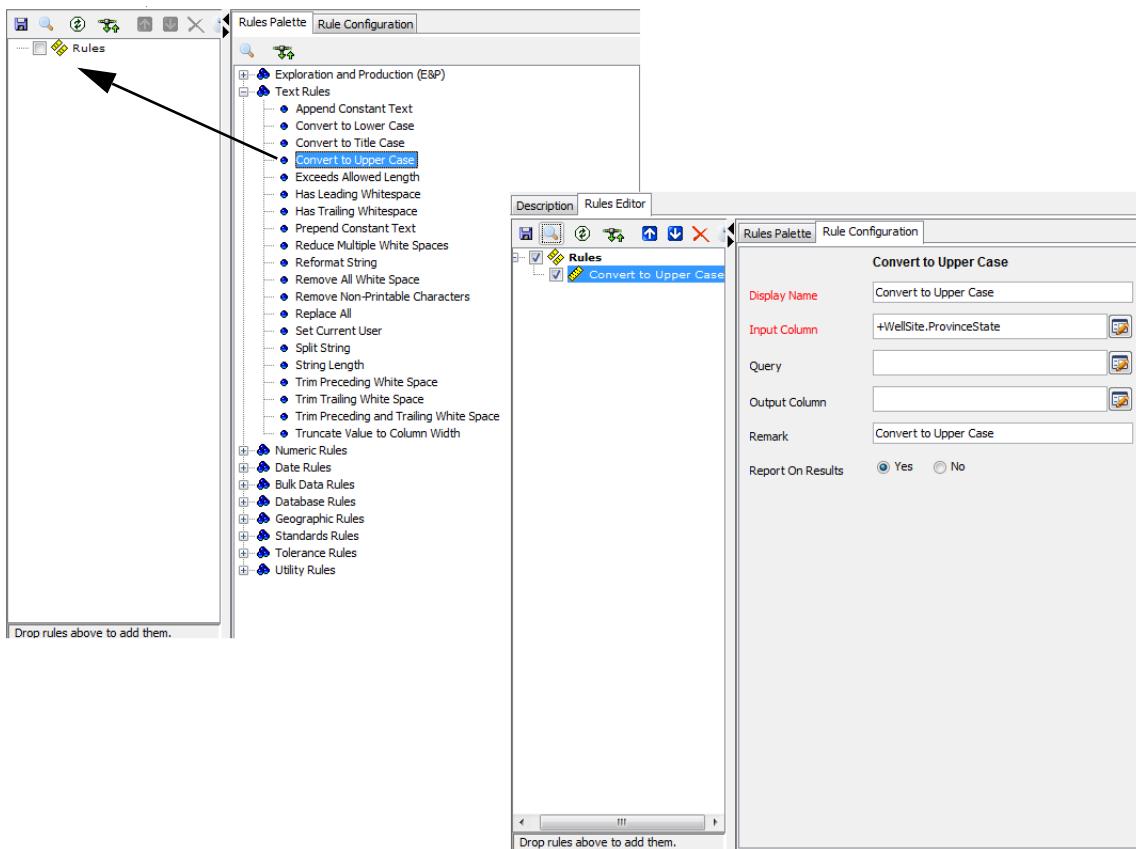
The **Convert Field Name to Upper Case** requirement is added to the **FIELD_NAME** column and displays in the **Model View Pane**.

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

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



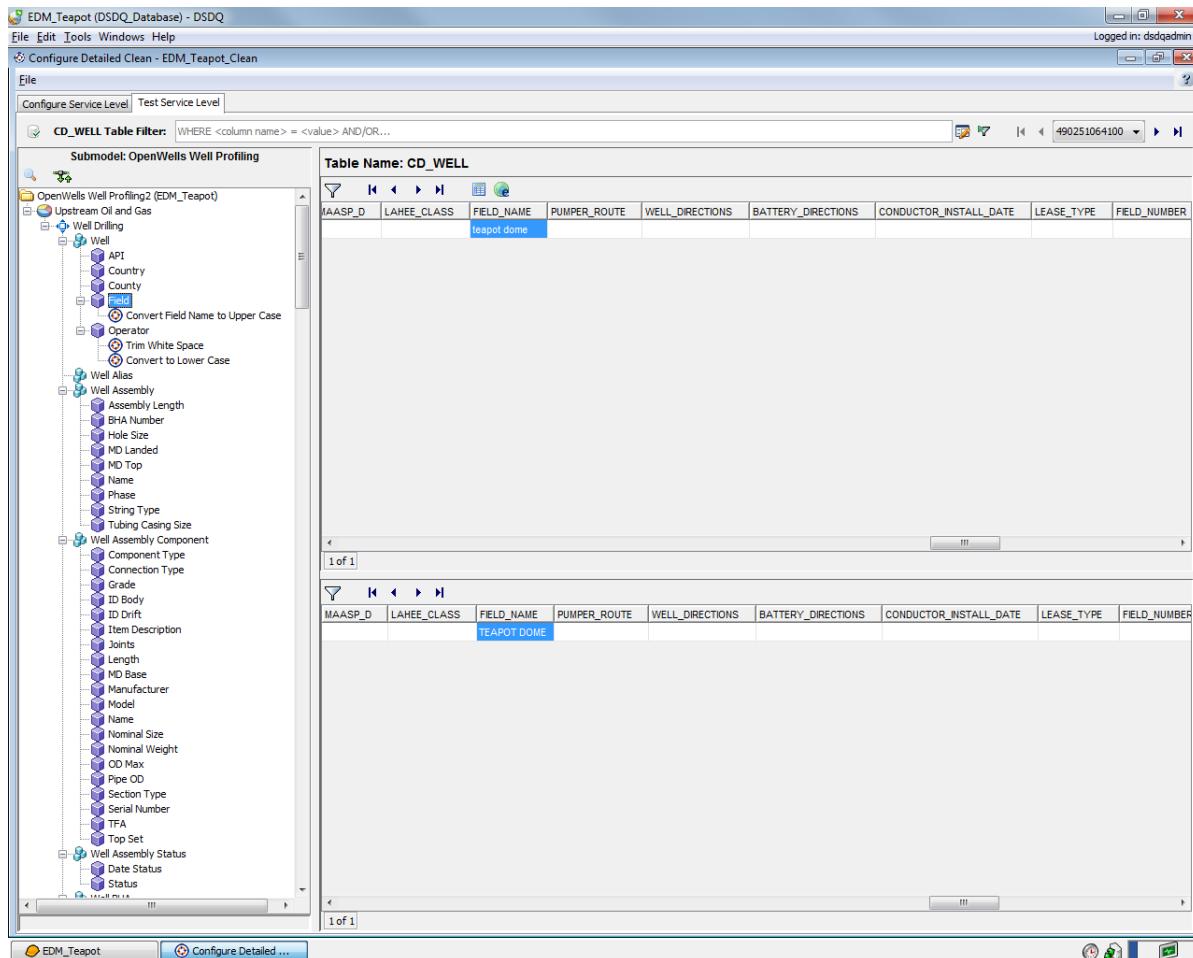
13. Drag and drop the **Convert to Upper Case** onto the **Rules** area.



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.



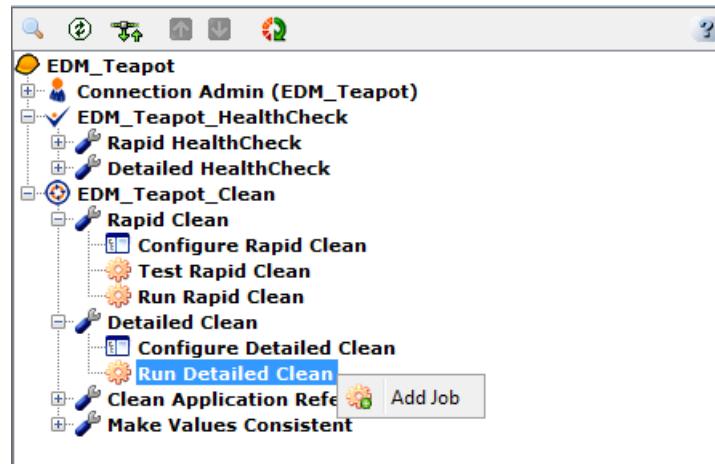
By looking at the columns that have been changed and temporary columns, you can verify that the behavior of the service level is correct prior to running the **Run Detailed Clean Task**.

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:

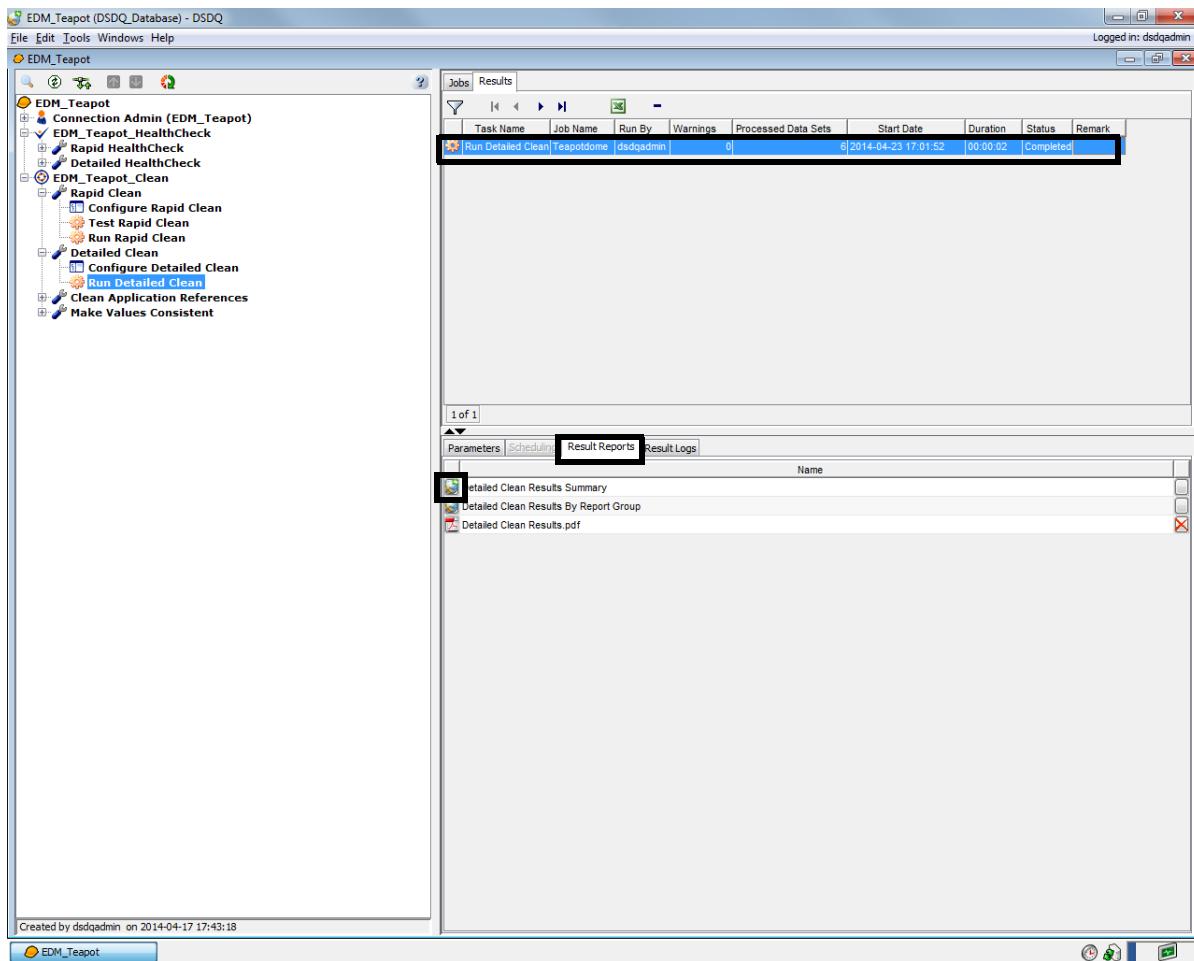
- Job Name:**
- Job Description:**
- Service Level:** OpenWells Well Profiling2 (EDM_Teapot)
- Select a Submodel:** OpenWells Well Profiling
- Filter Base Data Set:**
- Summarize Results by:** Requirement (least detail) Data Set Data Row (most detail)
- Generate Printable Reports?** Yes No
- Delete Results?** 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.

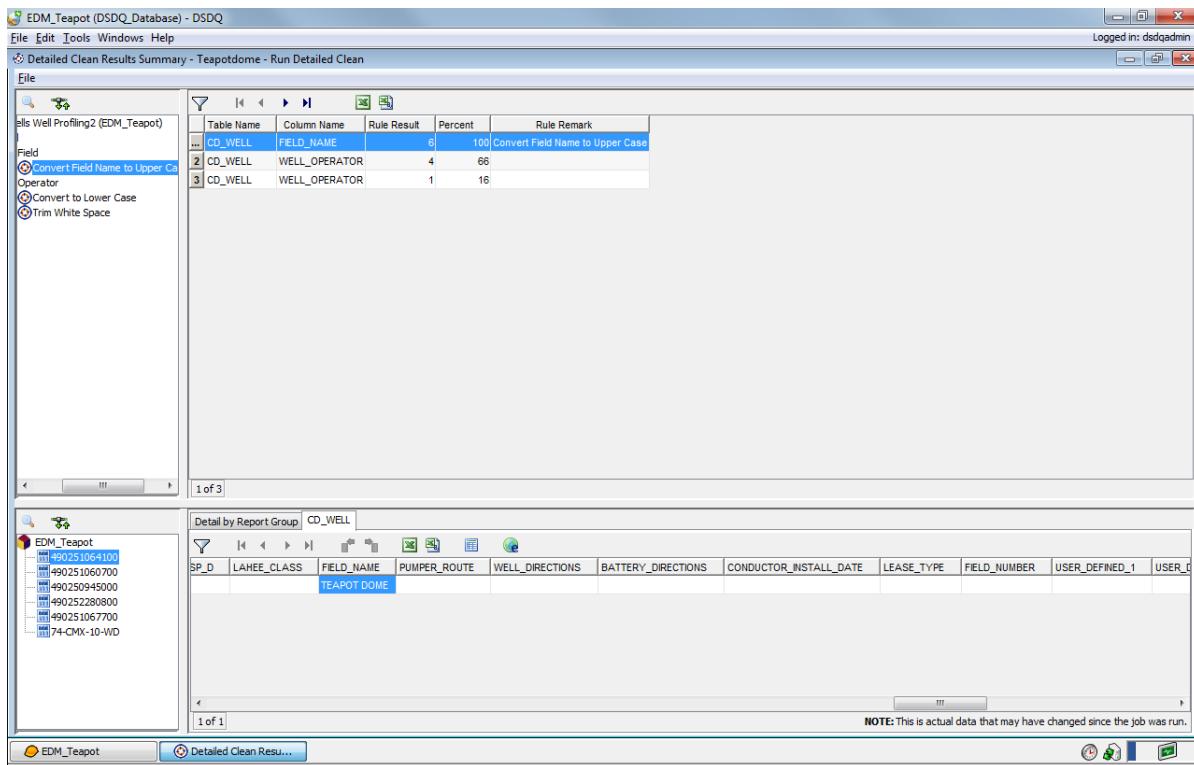
3. Enter **Run Detailed Clean for EDM_Teapot State to Upper Case** in the **Job Description** field.
4. Select **OpenWells Well Profiling2 (EDM_Teapot)** from the **Service Level** drop-down list.
5. Select **OpenWells Well Profiling** from the **Select a Submodel** drop-down list.
6. Do not set the 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.



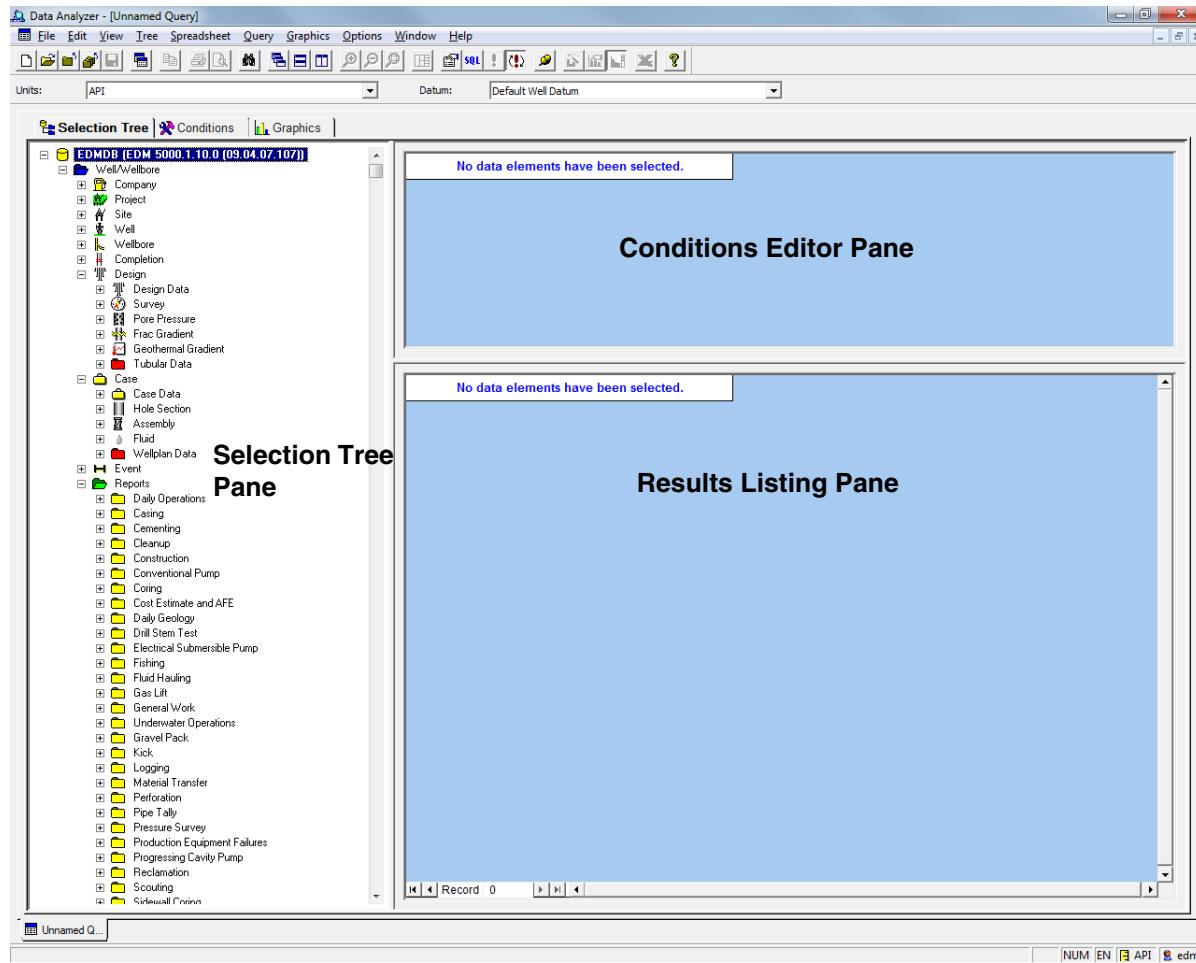
16. Select **File > Exit** to close the **Basic View Frame** window.

Validating Corrected Data in the EDM Data Source

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

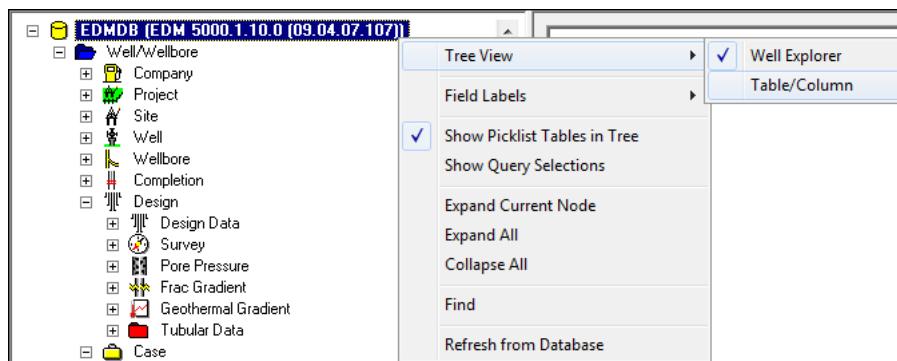
1. Double-click the Data Analyzer desktop icon to launch the Data Analyzer application.

The **Data Analyzer** window appears.

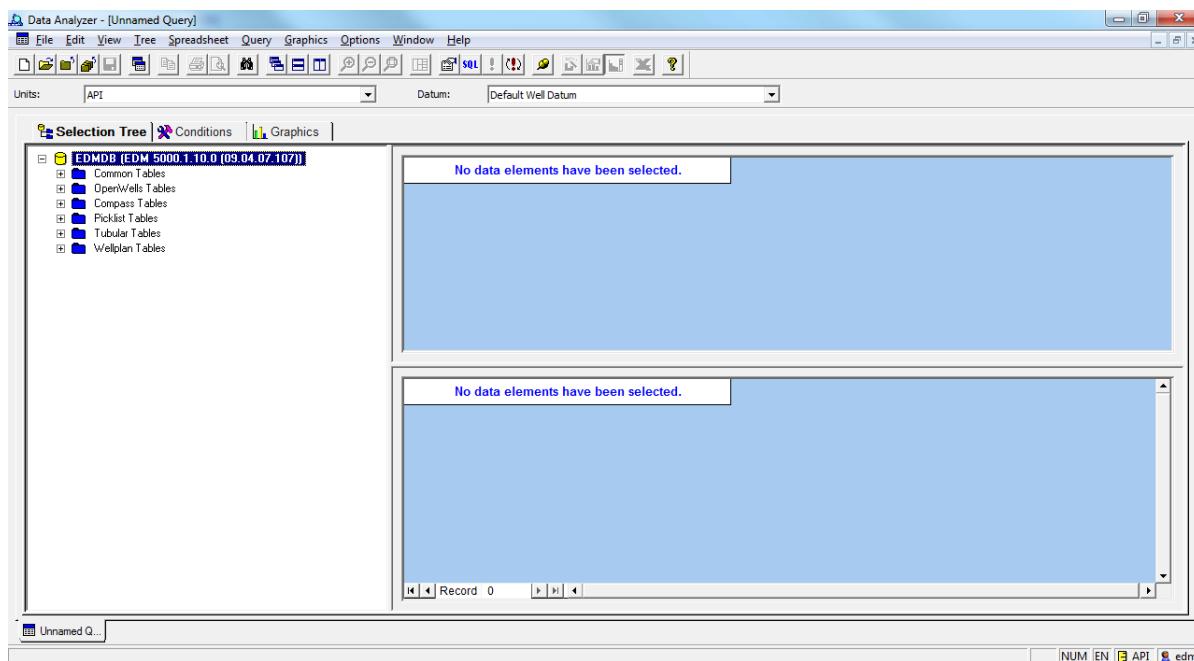


2. Right-click the **EDMDB (EDM 5000.1.10.0 (09.04.07.107))** data source in the **Selection Tree Pane** and select **Tree View > Table/**

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**.

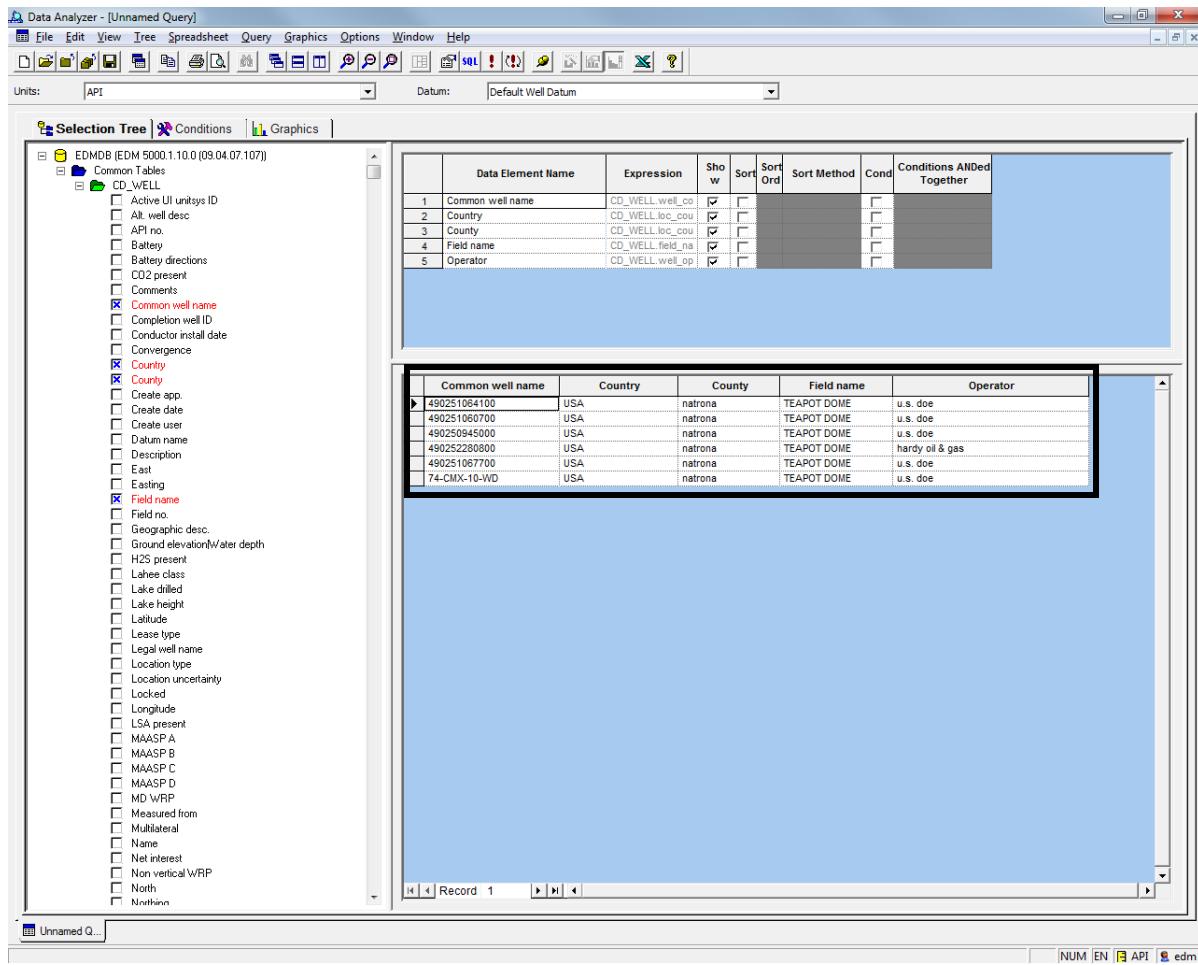
6. Repeat step 5 to select other elements.

The screenshot shows the Data Analyzer application window. On the left, the Selection Tree pane displays a hierarchical list of data elements from the 'EDMDB (EDM 5000.1.10.0 (09.04.07.107))' database, specifically under the 'CD_WELL' table. Several items are selected, indicated by a checked checkbox icon next to them. In the center, a grid table lists five 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"/>					

A message box at the bottom of the central pane states 'Query needs to be Run.' At the bottom of the window, the status bar shows 'Record 1' and navigation icons. The bottom right corner of the status bar includes buttons for 'NUM', 'EN', 'API', and '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 right pane consists of two tables. The top table is a query editor with columns: Data Element Name, Expression, Show, Sort, Sort Ord, Sort Method, Cond, and Conditions ANDed Together. It lists five rows corresponding to the fields selected in the Selection Tree. The bottom table is the Results Listing Pane, showing actual data records with columns: Common well name, Country, County, Field name, and Operator. The data shows several records for '490251064100' and '74-CMX-10-WD', all associated with 'USA', 'natrona', 'TEAPOT DOME', and 'u.s. doe'.

	Data Element Name	Expression	Show	Sort	Sort Ord	Sort Method	Cond	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"/>

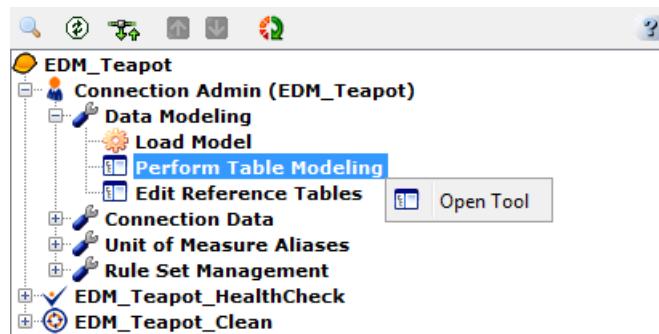
Common well name	Country	County	Field name	Operator
490251064100	USA	natrona	TEAPOT DOME	u.s. doe
490251060700	USA	natrona	TEAPOT DOME	u.s. doe
490250945000	USA	natrona	TEAPOT DOME	u.s. doe
490252280800	USA	natrona	TEAPOT DOME	hardy oil & gas
490251067700	USA	natrona	TEAPOT DOME	u.s. doe
74-CMX-10-WD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

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

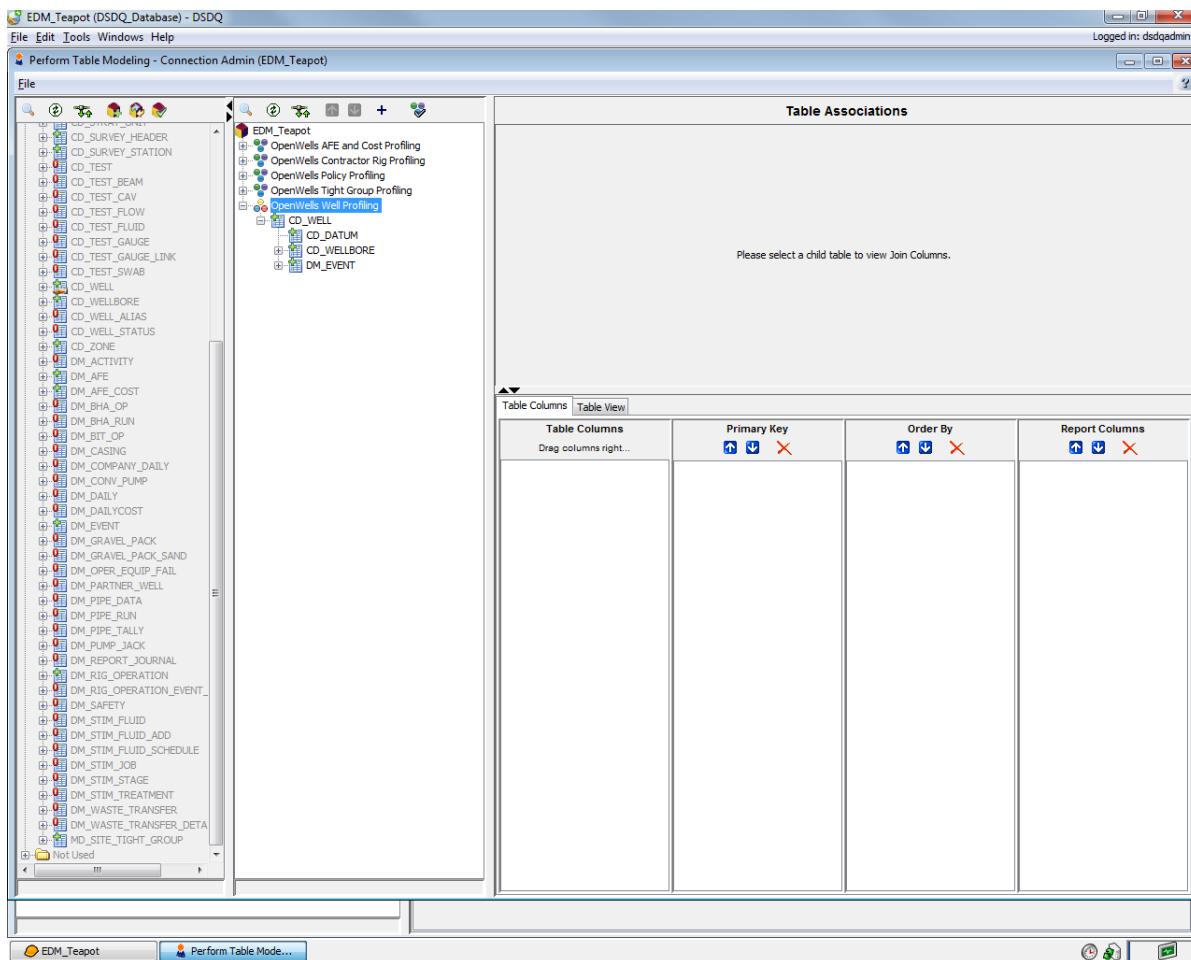
Viewing Results on the DSDQ Web Dashboard

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.

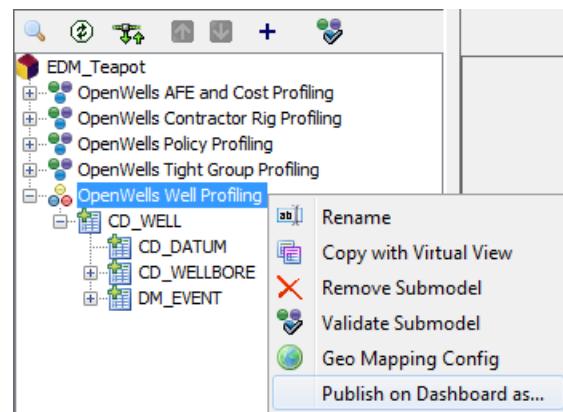


The **Perform Table Modeling - Connection Admin** window appears.

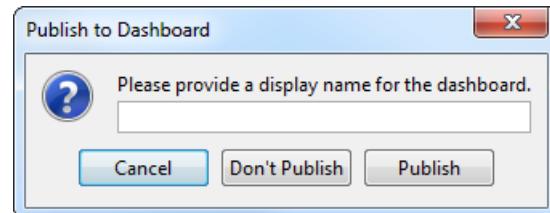


- Right-click the **OpenWells Well Profiling** submodel from the Submodel Listing Tree and select the **Publish on Dashboard as...**

option from the pop-up menu.



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



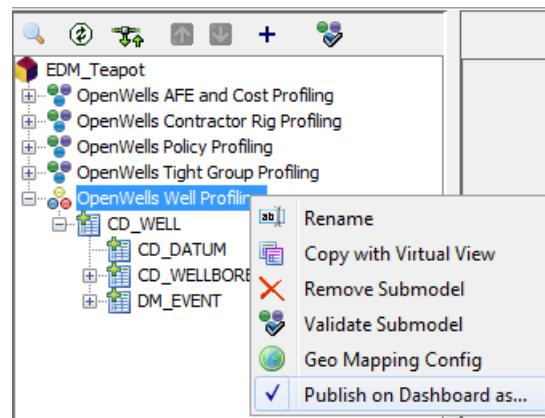
- Enter **OpenWells Well Profiling** in the **Please Provide a display name for the dashboard** field.



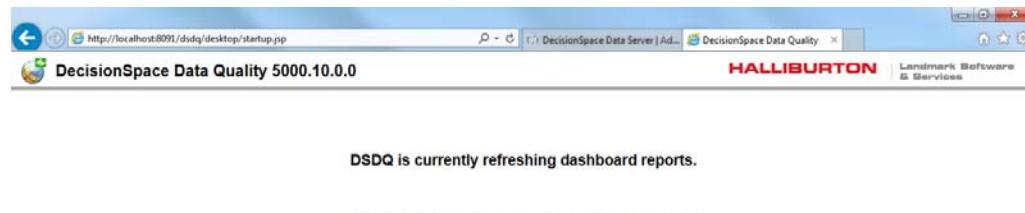
The **OpenWells Well Profiling** submodel is published to the Web Dashboard.

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

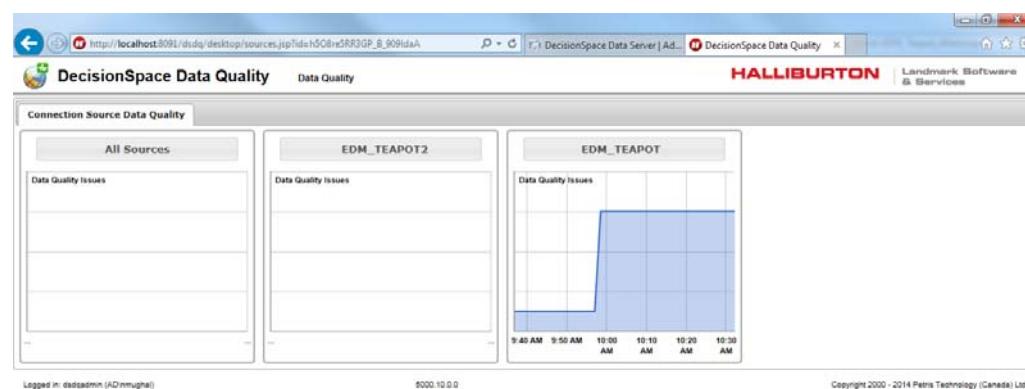
A checkmark appears on the right side of the **Publish on dashboard as...** option.



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.

