

Data Quality and Master Data Management with SQL Server 2012

2 Days DW-DQMDM12-401-EN

Description

Data is the key asset of any organization. Companies want to gather information from the data, so they start business intelligence (BI) projects. However, most of the BI projects also have to deal with data quality problems. Data quality can represent a tremendous obstacle to the success of any BI project. Line of business (LOB) applications will certainly suffer from poor data quality as well.

Every company has some data that is used everywhere, in every process, in every transaction, such as customer or product data. This type of data is referred to as *master data*. People in charge of managing master data are usually referred to as *data stewards*. The processes and activities required to maintain master data is collectively referred to as master data management (MDM). In this course, you are going to learn about typical MDM problems and solutions. You will also become familiar with Microsoft tools for data quality and MDM, including Data Quality Services (DQS) and Master Data Services (MDS).

Target Audience

This course is intended for data stewards, BI application developers, and advanced administrators.

Prerequisites

Before attending this seminar, it is recommended that the students have at least moderate experience with transactional applications, data warehousing, reporting, and online analytical processing. The attendees should be familiar with the Transact-SQL language. Knowledge of a .NET language like C# or VB.NET is welcome as well.

Course Objectives

Upon the completion of this course, the students will be able to:

- Describe what master data is
- Define MDM activities
- Understand data quality problems
- Install DQS

Microsoft Partner

Gold Business Intelligence
Gold Data Platform

- Create a DQS knowledge base and DQS projects
- Deploy MDS
- Manage master data with MDS
- Profile data with SQL Server tools (for instance, to find areas of low data quality)
- Control data quality over time
- Merge data from multiple independent sources by using identity mapping
- De-duplicate master data

Course Summary Outline

Day 1

Module 01: Introduction to Data Quality and Master Data Management

- Data quality
- Data models
- Master data management
- SQL Server suite and demo databases

Lab 01: Checking for Data Quality Issues and Defining Master Data

Module 02: Installing and Maintaining Data Quality Services

- Data Quality Services architecture
- Installing DQS
- Administering DQS

Lab 02: Installing and Administering DQS

Module 03: Installing and Implementing Master Data Services

- Master Data Services architecture
- Installing MDS
- Creating an MDS model

Lab 03: Installing MDS and Creating and Populating an MDS Model

Module 04: Managing Master Data with MDS

- Importing and exporting data
- MDS security
- Using the MDS Add-in for Excel

Lab 04: Importing and Exporting Data, Managing Security, and Using the MDS Add-in for Excel

Day 2

Module 05: Data Cleansing with DQS

- Creating a DQS knowledge base
- Creating a DQS project
- Using the DQS Cleansing transformation

Lab 05: Creating a DQS Knowledge Base and Project and Using the DQS Cleansing Transformation

Module 06: Data Profiling with SQL Server Tools

- Introductory profiling
- Measuring completeness
- Profiling accuracy

Lab 06: Using the SSIS Data Profiling Task

Module 07: Identity Mapping and De-Duplicating

- The problem
- DQS matching and SSIS fuzzy transformations
- MDS functions and a custom algorithm

Lab 07: DQS Matching and the SSIS Fuzzy Lookup Transformation

Module 08: Data Quality and Master Data Management with Excel

- Data profiling with Excel and PowerPivot
- Matching in Excel
- Using Excel Data Mining add-ins

Lab 08: DQS Matching and Data Profiling with PowerPivot