

DQ-05 Data Quality Assessment

DQ-05-00 About the Course (9 min)

DQ-05-01 Introduction (53 min)

- *Why Assess Data Quality*
- *Business Value of Data Quality Assessment*
- *Types of Data Errors*
- *Data Quality Assessment Approaches*
- *How Rule-Driven Approach Works*
- *Project Planning*
- *Project Steps*

DQ-05-02 Data Quality Rules Overview (63 min)

- *Attribute Domain Constraints*
 - Attribute Level View of Data
 - Attribute Profiling
 - Optionality Constraints
 - Format Constraints
 - Valid Value Constraints
 - Precision Constraints
 - Granularity Constraints
- *Relational Integrity Constraints*
 - Entity Level View of Data
 - Relational Data Models
 - Identity Rules
 - Reference Rules
 - Cardinal Rules
 - Inheritance Rules
- *Complex Data Relationships*
 - Subject Level View of Data
 - Redundant Attributes
 - Derived Attributes
 - Attributes with Related Domains
 - Attributes with Conditional Optionality
 - Advanced Business Rules

DQ-05-03 Rules for Historical Data (56 min)

- *Historical Data Overview*
 - Value Histories
 - Accumulator Histories
 - Event Histories
- *Timeline Constraints*
 - Currency Rules
 - Retention Rules
 - Continuity Rules
 - Granularity Rules
 - Advance Timeline Constraints
 - Timestamp Pattern Rules
- *Value Pattern Rules*
 - Constraints on Direction of Change
 - Constraints on Magnitude of Change
 - Constraints on Volatility of Change
- *Rules for Event Histories*
 - Event Dependencies
 - Event Conditions
 - Event-Specific Attribute Constraints
- *Rules for State-Dependent Objects*
 - State-Transition Models
 - State-Transition Constraints
 - State Continuity and Duration Constraints
 - Action-Specific Attribute Constraints
 - State-Specific Attribute Constraints

DQ-05-04 Finding Data Errors (76 min)

- *Discovering Data Quality Rules*
 - Data Profiling
 - Gathering Expert Knowledge
 - Investigating Data Relationships
 - Data Gazing
- *Implementing Data Quality Rules*
 - Selecting Relevant Rules
 - Choosing Optimal Rule Design
 - Rule Coding
- *Building Rule Catalogue*
 - Rule Listing
 - Error Groups
 - Rule Domains

- *Building Error Catalogue*
 - Error Cataloguing Basics
 - Referencing Erroneous Records
 - Using Error Messages
 - Record and Subject-Level Error Cataloguing
- *Fine-Tuning Data Quality Rules*
 - Handling False Positives
 - Handling False Negatives
 - Handling Uncertainty in Error Location

DQ-05-05 Aggregate Data Quality Scores (66 min)

- *School Report Card Example*
- *A First Look at DQ Scorecard*
- *Defining Aggregate Scores*
 - Defining Data Quality
 - Examples of Data Quality Definitions
 - What is an Aggregate Score?
 - Business-Driven Aggregate Scores
 - Scores Identifying Sources of Bad Data
 - Scores Related to Data Structure
 - Time Dimension of Aggregate Scores
 - Record-Level and Subject-Level Scores
 - Tabulated vs. Non-Tabulated Scores
- *Score Tabulation*
 - The Simplest Aggregate Score
 - Score for a Data Subset
 - Score for Two DQ Rules
 - Using Error Levels
 - Data Quality in Shades of Gray
 - Assigning Ratings to Errors
 - Handling Records with Multiple Errors
 - Assigning Weights to Fields
 - Assigning Weights to Data Subsets
 - Adjusting for Missing Records
 - Adjusting for False Negatives
 - A Word of Caution
 - Tabulating Subject-Level Scores

DQ-05-06 Building Data Quality Scorecard (61 min)

- *Basic Scorecard Example*
 - Score Objective
 - Relevant Data Elements
 - Relevant Data Quality Rules
 - Relevant Records
 - Top Level Scores
 - Decomposition by Error Type
 - Decomposition by Data Location
 - Lower Level Decompositions
 - Drill-Down Error Reports
 - Other Useful Error Reports
 - Integration with Data
 - Atomic Level Data Quality
 - Subject Browser
 - Integration with Other DQ Metadata
- *Recurrent Data Quality Assessment*
 - What to Do After Data Quality is Measured?
 - Rule Maintenance
 - Dynamic Data Profiling
 - Subject Level Data Quality Trends
 - Atomic Level Data Quality Trends
- *Database and Enterprise-Wide DQ Scorecard*
 - Scorecard Dimensions
 - Purposes of Use
 - Subject Classes
 - Data Quality Dimensions
 - Data Sources
 - Database Top Level Score
 - Time
 - Cross-Database Scores