

# Data Quality Fundamentals

Dave Wells



## DQ-01 Data Quality Fundamentals

### Module 0: About the Course (2 min)

### Module 1: Data Quality Concepts (32 min)

- What is Data Quality?
  - What is Quality?
  - What isn't Quality?
  - What is Data Quality – Defect Free Data
  - What is Data Quality – Conforming to Data Specifications
  - What is Data Quality – Data Suited Purpose
  - What is Data Quality – Meeting Data Customer / Consumer Expectations
  - From Expectations to Assessment
- Dimensions of Data Quality
  - Data Quality Dimensions
  - Data Quality Criteria
  - Why and How are Dimensions Useful?
  - Data Correctness
  - Data Integrity
  - Data Usability
  - Data Objectivity
  - Detecting Data Quality Defects
- Module Summary

### Module 2: Data Quality Context (52 min)

- Causes of Data Quality Defects
  - What Causes Data Quality Defects?
  - Data Collection Errors & Deficiencies
  - Data Processing Errors & Deficiencies
  - Data Corruption, Intrusion, and Loss
  - Data Decay
  - Data Bias and Distortion
  - Defects Propagation
- People, Processes, and Data Quality
  - Data Governance and Data Quality
  - Data Policies and Data Quality Management
  - Data Stewardship and Data Quality
  - Data Stewardship and Metadata Quality
  - Data Literacy and Data Quality
  - Data Architecture and Data Quality
  - Data Quality as a Program
  - Developing & Evolving a Data Quality Program
- Module Summary



## DQ-01 Data Quality Fundamentals

### Module 3: Exploring Data (41 min)

- Exploring Data
  - Data Browsing and Data Gazing
- Statistical Data Profiling
  - The Statistical Data Profiling Process
  - Column Profiling – Extracting Values Metadata
  - Column Profile Metadata
  - Column Profile Analysis
  - Table Profiling – Examining Dependencies
  - Table Profile Metadata & Analysis
  - Cross-Table Profiling – Examining Overlaps
  - Cross-Table Profile Metadata & Analysis
- Visual Data Profiling
  - Visual Profiling Techniques
  - Univariate Analysis – Distribution of Values
  - Univariate Analysis – Central Tendency & Variation
  - Bivariate Analysis – Two Quantitative Variables
  - Bivariate Analysis – Categorical & Quantitative
  - Bivariate Analysis – Two Categorical Variables
  - Behavior Over Time Analysis – Single Variable Behaviors
  - Behavior Over Time Analysis – Related Variable Behaviors
- Module Summary

### Module 4: Discovering Data Quality Rules (36 min)

- Discovering Data Quality Rules
  - Sources of Data Quality Rules
- Discovering Correctness Rules
  - Discovering Correctness Rules from Business Rules
  - Discovering Correctness Rules from Data Profiles
  - Discovering Correctness Rules from Reference Data
- Discovering Integrity Rules
  - Discovering Integrity Rules from Reference Data
  - Discovering Integrity Rules from Data Models
  - Discovering Integrity Rules from Data Profiles
  - Discovering Integrity Rules from Business Rules
- Discovering Usability Rules
  - Discovering Usability Rules from Business Processes
  - Discovering Usability Rules from Observations
- Discovering Objectivity Rules
  - Discovering Objectivity Rules from Data Profiles
  - Discovering Objectivity Rules from Observations
  - Discovering Objectivity Rules from Standards
- Module Summary



## DQ-01 Data Quality Fundamentals

### Module 5: Managing Data Quality (52 min)

- DQ Management Overview
  - Data Quality Management Process
  - The DQ Management Stack
  - The Stack and the Process Together
  - Data Quality Metadata
  - Data Quality Management Tools
- Measurement
  - Measuring to Manage Quality
  - Hard and Objective Measures
  - Soft and Subjective Measures
  - Assessment and Reporting
  - Data Quality Scorecard Pragmatics
  - Data Quality Scorecard Variety
  - Data Quality Scorecard: Scope & Requirements
  - Data Quality Scorecard: Detailed Scores for CDEs
  - Data Quality Scorecard: Summary Scores for Compliance Management
- Issue Management
  - Issue Management in Context
  - Issue Management Activities & Questions
- Root Cause Analysis
  - The *What* and *Why* of RCA
  - The Nature of Cause and Effect
  - The Root Cause Analysis Process
  - Five Whys
  - Fishbone Diagraming
  - Causal Loop Modeling
  - What Next After Root Cause Analysis?
- Quality Improvement
  - Detecting Data Quality Defects
  - Correcting Data Quality Defects
  - Data Cleansing
  - Process Improvement
  - Process Improvement Example
- Quality by Design
  - Building in Data Quality
- Module Summary



### Module 6: Building a Data Quality Program (24 min)

- Program Framing
  - Program Components and Activities
  - DQ Program Lifecycle
- Program Foundations
  - The Building Blocks of a Data Quality Program
  - Program Charter
  - Organization & Roles
  - Resources
  - Assessment
  - Putting the Pieces Together
- Program Operations
  - Goal-Driven Operations
  - Goals
  - Metrics and SLAs
  - Tactics: Projects & Processes
  - Execution: Day-to-Day Activities
  - Stakeholder Engagement & Communication
  - Feedback & Adjustments
- Program Oversight & Governance
  - Program Monitoring
  - Program Maturity
  - Program Evolution & Sustainability
- Module Summary