Data Quality for Data Stewards

by Arkady Maydanchik, Olga Maydanchik and Dave Wells

Module 0. About the Course (9 min)

Module 1. Data Quality Basics (64 min)

- *Quality Defined*
  - What Is Quality?
  - What Isn't Quality
  - Quality Vs Qualities
  - Some Examples
- *Data Quality Defined*
  - Defect Free
  - Conforming to Specifications
  - Suited to Purpose
  - Meet Customer Expectations
  - Defining Data Quality
  - Qualities of Data
- *Common Causes of Data Quality Problems*
  - Manual Data Entry
  - Data Integration
    - Initial Data Conversion
    - System Consolidation
    - Batch Feeds
    - Real-time Interfaces
  - Data Manipulation
    - Data Processing
    - Data Cleansing
    - Data Purging
  - Data Decay
    - Changes Not Captured
    - System Upgrades
    - New Data Uses
    - Loss of Expertise
    - Process Automation
- *Dimensions of Data Quality*
  - Data Quality of Multi-Dimensional
  - Content and Quality
  - Structure and Quality
  - Time and Quality
  - Business and Quality
  - Usage and Quality
  - Presentation and Quality

Module 2. Data Quality Management (40 min)

- *Data Quality Projects*
  - Common Data Quality Projects
  - Assessment Projects
  - Data Cleansing Projects
  - Process Improvement Projects
- *Building-In Data Quality*
  - Data Quality and System Architecture
  - Data Quality and IT Projects
  - Application Development Projects

    - Data Conversion and Migration Projects
    - Data Warehousing Projects
    - Business Analytics Projects
- *Data Quality and Big Data*
    - Data Quality and Big Data
    - Judging the Quality of Big Data
    - A Framework to Judge Quality
    - Data Characteristics and Data Usefulness
    - Data Processing and Characteristics
    - Big Data Quality and Data Analysis
    - Aligned with the Analytics Lifecycle
- *Data Quality Programs*
- *Data Quality Profession*

Module 3. Introduction to Data Quality Assessment (48 min)

- *Why Assess Data Quality?*
    - Role of Assessment in Data Quality Programs
    - Assessment Objectives
- *Business Value of Data Quality*
    - Planning Data Cleansing
    - Improving Data Collection
    - Improving Data-Driven Business Processes
    - Supporting Data Migration and Data Integration
- *Data Quality Assessment Approaches*
    - Data Validation Against Trusted Source
    - Sample Data Validation
    - Rule-Driven Approach
- *Project Team*
- *Project Steps*
    - Defining Project Scope
    - Loading Data to Staging Area
    - Gathering General Metadata
    - Data Profiling
    - Designing Data Quality Rules
    - Implementing Data Quality Rules
    - Fine-tuning Data Quality Rules
    - Building Data Quality Scorecard
- *Data Quality Rules Overview*
    - Attribute level View of Data
    - Attribute Domain Constraints
    - Entity Level View of Data
    - Relational Integrity Constraints
    - Subject Level View of Data
    - Attribute Dependency Constraints
    - Advanced Business Rules
    - Time Dependent View of Data
- *Recurrent Data Quality Assessment*

Module 4. Data Quality Scorecard (46 min)

- *What is a Data Quality Scorecard*

- *DQ Scorecard Case Study #1*
  - DQ Scorecard Goal
  - RWA Calculation Process
  - Solution Steps
  - Field Scores
  - Field DQ Score vs EAD
  - List of Rule Affecting DQ Score
  - Rule Error List
  - Business Process DQ Score
  - DQ Score Decomposition By Instrument Type
  - DQ Score Decomposition By Asset Delinquency
  - Case Study #1 Summary
- *DQ Scorecard Case Study #2*
  - Catastrophe Insurance - Loss Reserves
  - Catastrophe Risk Modeling Process Explained
  - DQ Scorecard Goal
  - Data Quality Factor vs Structure Insured Value
  - The Cost of Poor Quality Data
  - Data Quality Factor Analytics
  - Rules for Data Element
  - Business Case for Data Quality
  - Case Study #2 Summary
- *Typical DQ Scorecard Components*

Module 5. Root Cause Analysis (53 min)

- *The Nature of Cause and Effect*
  - Correlation
  - Coincidence
  - Confounding Variables
  - Contributing Variables
  - Influence Examples 1 & 2
  - Complexity and Complications
- *Cause and Effect Misconceptions*
  - Simple vs Complex
  - Linear vs Circular
  - Close vs Distant
  - Immediate vs Delayed
  - Ordered vs Non-ordered
- *The Purpose of RCA*
  - Root Cause Analysis Defined
  - Root Cause Defined
  - Root Cause Criteria
  - Root Cause Criteria Examples
- *The Process of RCA*
  - Five Steps of RCA
  - The Problem Description
  - Data Gathering
  - Casual Modeling
  - Root Cause Identification
  - Recommendations

- *A First Look at Cause-Effect Models*
  - Five Why's
  - Fishbone Diagrams
  - Causal Loop Models
- *Verifying Cause and Effect Conclusions*
  - A Short Reasoning Exercise
  - Exercise Review
  - Nonsensical Thinking
  - Nonsense and Distortion
  - Seeing What We Want to See
  - Logical Fallacies
  - Rethinking Cause and Effect
  - Systems Thinking
  - Critical Thinking
  - Bias of Facts and Logic
  - Evidence and Assumptions
  - Lateral Thinking
- *Practical Application*
  - Beyond RCA
  - Implement Solutions
  - Measure and Monitor
  - Watch for Side Effects
  - Correct Existing Issues

Module 6. Ensuring Data Quality in Data Integration (41 min)

- *Data Integration Basics*
  - Data Integration Defined
  - Common Applications
  - Data Integration Approached
  - Comparison of Integration Approaches
- *Data Quality Perspective*
  - Pros and Cons of Data Federation
  - Data Federation Quality Solution
  - Pros and Cons of Data Propagation
  - Data Propagation Quality Solution
- *Data Propagation Overview*
  - Propagation Frequency and Data Latency
  - Real-Time Interfaces
  - Batch Interfaces
  - Using Information Bus
  - Using Staging Area
  - Interface Spider-Web Challenge
  - Information Integration Hub
- *Real-Time Interfaces*
  - Types of Data Quality Problems
  - Challenges in Quality Monitoring
  - Near-Real-Time Interfaces
  - Role of Data Quality Assessment
- *Batch Interfaces*
  - Causes of Data Quality Problems

- Data Quality Management
- Batch Monitors
- Change Monitors
- Error Monitors