



# ***Data Quality Assessment***

***by Arkady Maydanchik***



### **Module 0. About the Course (9 min)**

### **Module 1. 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*

### **Module 2. 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*

### **Module 3. 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*



## DQ-05: Data Quality Assessment

- 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

### **Module 4. 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

### **Module 5. Building Data Quality Scorecard (46 min)**

- *School Report Card Example*
- *A First Look at DQ Scorecard*
- *Introduction to 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



## DQ-05: Data Quality Assessment

- *Recurrent Data Quality Assessment*
  - What to Do After Data Quality Is Measured?
  - Overview
  - Rule Maintenance
  - Trends in Data Quality Scores
  - Subject-Level Data Quality Trends
  - Atomic-Level Data Quality Trends