



# ***Data Integration Techniques for Designing an ODS***

***by Angelo Bobak***



### **Module 0. About the Course (4 min)**

- *About the Author*
- *Course Objectives*
- *Audience and Pre-requisites*
- *Course Structure*

### **Module 1. Introduction to Operational Data Stores (45 min)**

- *Overview*
- *What is an ODS*
  - ODS Defined
  - Push ODS
  - Pull ODS
  - Hybrid ODS
  - Internal Architecture
  - ODS Layers and Their Role
  - Why is an ODS Important
  - Benefits of an ODS
  - Disadvantages of an ODS
  - ODS Data Model Example
- *Master Data and The ODS*
  - Master Data Categories
  - Master Data Examples
  - Customer Master Data
  - Address Master Data
  - Product Master Data
  - Using the Product Master
  - Geographical Master Data
  - Calendar Master Data
- *Data Quality and the ODS*
  - Overview
  - Data Profiling
  - What to Check For?
  - Data Cleansing
  - Data Quality Reporting
  - Data Quality Scorecard
  - Additional Scorecards
- *Loading the ODS*
  - Overview
  - Staging Layer
  - Profiling Layer
  - Cleansing Layer
  - Integration Layer
  - Export Layer
  - Financial ODS Application Example
- *ODS and Data Warehouse Architectures*
  - Overview
  - Primary Similarities and Differences
  - How They Complement Each Other
  - How Each Uses Master Data
  - Why Is Data Quality Important to Both



- Data Granularity

### **Module 2. The Theory of Schema Integration (58 min)**

- *Overview*
- *Data Integration Pioneers*
- *Schema Integration Types*
  - Overview
  - Type 1 Binary Schema Integration
  - Type 2 Binary Schema Integration
  - Type 3 Tertiary Schema Integration
  - Type 4 N-ary Schema Integration
- *Schema Integration Process*
  - Overview
  - Data Flow Diagram for Schema Integration
  - Defining the Integration Sequence
  - Identifying Tables to Integrate
  - A Simple Example
  - Identifying Columns to Integrate
  - identifying Data Conflicts
- *Resolving Data Conflicts*
  - Overview
  - Data Naming Conflict Example
  - Data Type Conflict Example 1
  - Data Type Conflict Example 2
  - Data Type Conflict Example 3
  - Data Semantic Conflict Example
  - Entity vs. Column Conflict Example
  - Entity vs. Column Conflict Resolution
  - PK/FK Conflict Example 1
  - PK/FK Conflict Resolution
  - PK/FK Conflict Example 2
  - PK/FK Conflict Example 3
  - Other Structural Conflicts Example 1
  - Other Structural Conflicts Example 2
  - Other Structural Conflicts Example 3
- *Profiling Data*
  - Overview
  - An Example
  - Example Data
  - Understanding the Quality and Nature of Data
  - Pre-Integration Cleansing Opportunities
- *Defining the ETL Specifications to Merge Data*
  - Overview
  - Source Dictionary Part 1
  - Source Dictionary Part 2
  - Unique Attribute Report
  - Common Attribute Report Part 1
  - Common Attribute Report Part 2
  - Conflict Resolution Reports
  - Customer and Customer Location Final Integrated Model
  - Data Mapping Specifications Template



## DI-03: Data Integration Techniques for Designing an ODS

- Data Flow Diagrams
- Data Mapping Specifications Template
- Data Flow Diagrams
- Process Hierarchy Diagrams
- Process Dependency Diagrams
- Data Flow Templates
- *Tracking Data Lineage*
  - Overview
  - Tracking Sources
  - Logging Transformations
  - Identifying the Final Destination Part 1
  - Identifying the Final Destination Part 2
  - Logging as Part of the ETL Process
  - Preserving Data Conversion History
- *Schema Integration ETL Tools*
  - What to Look For
  - Vendors

### **Module 3. ODS Maintenance (30 min)**

- *Overview*
- *Adding New Sources*
  - Overview
  - Scenario 1
  - Scenario 2
  - Scenario 3
  - Scenario 4
- *Adding New Destinations*
  - Adding New Export Views
  - Getting Data From Interim Data Integration Layers
- *Modifying Existing Sources*
- *Modifying Existing Destinations*
- *Retiring Old Sources*
  - Retiring Old Sources Part 1
  - Retiring Old Sources Part 2
- *Retiring Existing Destinations*
- *Managing Security and Access*
  - Outbound
  - Inbound
  - Example Security Tables
  - Some References
- *Monitoring and Managing Storage Capacity*
- *Monitoring Performance*
- *Physical Design Techniques to Increase Performance*
- *Key Project Roles and Responsibilities*
  - The Project Manager
  - The Business Analyst
  - The Data Architects
  - The Data Modeler
  - The ETL Designer
  - The Data Quality Report Developer



- Testers

### **Module 4. Case Study (38 min)**

- *Overview*
- *Databases To Integrate*
  - Introduction
  - Customer London Model
  - Customer Turin Model
  - Product London Model
  - Product Turin Model
  - Order London Model
  - Order Turin Model
  - Complete London Model
  - Complete Turin Model
- *Data Dictionaries*
  - Overview
  - London Customer Data Dictionary
  - Turin Customer Data Dictionary
  - London Product Data Dictionary
  - Turin Product Data Dictionary
  - London Order Data Dictionary
  - Turin Order Data Dictionary
  - Table Relationship Data Dictionary
- *Tools You Will Need*
  - Integration Maps
  - Common/Unique Table Reports
  - Common Attribute Reports
  - Unique Attribute Reports
  - Conflict Resolution Reports
  - ETL Specifications
  - Sample Accompanying Document
  - Interim Integration Data and Process Models
  - Detailed Task Descriptions
- *Performing the Integration*
  - Overview
  - Identify Candidate Tables and Integration Sequence
  - Common Table Report
  - Unique Table Report
  - Column Inventory Report
  - Common Column Report
  - Unique Column Report
  - Conflict Resolution Report
  - ETL Specification
  - Interim Schema Model
  - Data Flow Analysis
  - Data Flow Diagram
  - Creating New Address and Location Tables
- *Concluding Remarks*