


Data Quality: Advanced Techniques

 now.informatica.com/Data-Quality-Advanced-Techniques-ILT.html

Course Overview

Learn to leverage advanced techniques when utilizing Developer to profile, cleanse, standardize, de-duplicate and consolidate data in an enterprise. Focused on creating and applying custom built Classifier and Probabilistic Models, utilizing advanced Parsing and Matching methods, refining Human Tasks and Workflows, automatically Associating and Consolidating matched records, applying Parameters in mappings and more. This course is applicable for all version 10 releases.

Objectives

After successfully completing this course, students should be able to:

- Perform Join Profiling.
- Create and apply Classification Models.
- Parse data using advanced techniques.
- Create and apply Probabilistic Models.
- Apply sophisticated Grouping and Matching techniques.
- Automatically Associate and Consolidate matched records.
- Refine Exception and Duplicate Record Workflows used to populate Analyst inboxes.
- Design, Implement and Test processes to manage updated exception/duplicate records.
- Appropriate DQ Parameters.
- Examine Performance considerations.
- Review CRM and Dashboard & Reporting Templates.
- Optionally/Time allowing:
 - Leverage Web Services to apply DQ mappings in Excel.
 - Perform Identity Matching.
 - Use the Universal ID store to match against master data.

Target Audience

- Developer

Prerequisites

- [Data Quality: Data Quality Management for Developers](#)

Agenda

Module 1: Course Introduction

- Course Introduction, Agenda and Overview

Module 8: Processing Updated Exception and Cluster Data

- How to process updated exception records

Module 2: Developer Review & Join Profiling

- A quick review of Informatica Developer
- Use Enterprise Discovery to create Join Profiles.
- Lab: Perform Join Profiling using an Enterprise Discovery Profile

Module 3: Standardizing and Classifying Data

- Review Standardization Techniques
- Build, refine and apply a Classifier Model
- Labs: Create, refine and apply Classifier Model

Module 4: Advanced Parsing Techniques

- What is Probabilistic Labeling and Parsing?
- Build, refine and apply a Probabilistic Model.
- Additional Parsing Techniques:
 - Build regular expressions.
- Labs: Build, refine and apply a Probabilistic Model
- Lab: Review an example of Advanced Parsing
- Lab: Generate and test Regular Expressions

Module 5: Grouping & Matching Data

- Additional Grouping Techniques
 - Using Composite keys
- Advanced Matching Techniques
 - Matched pairs outputs.
 - Working with Match Mapplets.
 - Manipulating the matched data using the Driver ID
 - Perform Dual Matching
- Lab: Create a Match mapping using Matched Pairs
- Lab: Create and update a Match Mapplet
- Lab: Manipulating Matched Data using the Driver ID
- Lab: Perform Dual Matching using a Master Dataset.

Module 6: Automatically Associate and Consolidate Matched Data

- Overview of the Consolidation Process

- How to process consolidated records
- Fields of Interest
- Lab: Create a mapping to process updated exception data
- Lab: Create a mapping to process consolidated data
- Lab: Update and deploy Exception and Cluster Workflows

Module 9: Analyst Tasks

- Update exception and duplicate records in Informatica Analyst
- Lab: Update records and push the Tasks through the Exception Process
- Lab: Update records and push the Tasks through the Consolidation Process

Module 10: Parameterization

- Explain the difference between System and User defined parameters
- Use Parameters in Data Quality mappings.
- Lab: Create a parameterized mapping
- Lab: Build and deploy an Application
- Lab: Create and execute parameter files

Module 11: Performance tips and tricks

- General Installation and Memory Information
- DQ Component Configuration
 - Service Settings
- DQ Transformations
 - Configuration Settings

Module 12: Optional - Data Quality at work

- Learn how Data Quality has been implemented in different projects.

Module 13: CRM and Dashboard and Reporting Templates

- Review the CRM and Dashboard and Reporting Templates that are available
- Lab: Review the CRM Template

Appendix:

- Module: DQ for Excel using Web Services
- Use Data Quality Web Services to execute DQ mappings on Excel Spread sheets.
- Lab: Use Web Services to execute mappings in Excel

- Use the Consolidation Transformation to consolidate matched data.
- Use the Association Transformation to link matched data ahead of Consolidation.
- Lab: Automatically Consolidate matched data.
- Lab: Perform multi-criteria Matching, Association and Consolidation.

Module 7: Task and Workflow Management

- Additional Task and Workflow functionality:
 - Permission settings for data access and editing
 - Notifications including Human Task Notification Variables
 - Setting Timeouts
 - Reviewing Tasks
 - Configuring Workflow Recovery
 - Lab: Update the Exception Workflow
 - Lab: Review the Consolidation Workflow
-
-
-

Module: Identity Matching -

- Match Data using Identity Matching
 - Use UID to match data against a Master Data Store
- Lab: Use Identity to match customer data
- Mixed Matching Workshop
- Lab: Universal ID, Create and load the Persistent Data Store
- Lab: Match and update new records to the Store.

[Back to Course Overview](#)

QUESTIONS?

Instructor Led | Data Quality | 3 Days