

Training: IBM
 IBM InfoSphere QualityStage Essentials



TRAINING GOALS:

This course teaches how to build QualityStage parallel jobs that investigate, standardize, match, and consolidate data records. Students will gain experience by building an application that combines customer data from three source systems into a single master customer record.

- List the common data quality contaminants
- Describe each of the following processes:

§ Investigation

§ Standardization

§ Match

§ Survivorship

- Describe QualityStage architecture
- Describe QualityStage clients and their functions
- Import metadata
- Build and run DataStage/QualityStage jobs, review results
- Build Investigate jobs
- Use Character Discrete, Concatenate, and Word Investigations to analyze data fields
- Describe the Standardize stage
- Identify Rule Sets
- Build jobs using the Standardize stage
- Interpret standardization results
- Investigate unhandled data and patterns
- Build a QualityStage job to identify matching records
- Apply multiple Match passes to increase efficiency

- Interpret and improve match results
- Build a QualityStage Survive job that will consolidate matched records into a single master record
- Build a single job to match data using a Two-Source match
- Data Analysts responsible for data quality using QualityStage
- Data Quality Architects
- Data Cleansing Developers

CONSPECT:

1. Data Quality Issues

- Listing the common data quality contaminants
- Describing data quality processes

2. QualityStage Overview

- Describing QualityStage architecture
- Describing QualityStage clients and their functions

3. Developing with QualityStage

- Importing metadata
- Building DataStage/QualityStage Jobs
- Running jobs
- Reviewing results

4. Investigate

- Building Investigate jobs
- Using Character Discrete, Concatenate, and Word Investigations to analyze data fields
- Reviewing results

5. Standardize

- Describing the Standardize stage
- Identifying Rule Sets
- Building jobs using the Standardize stage
- Interpreting standardize results
- Investigating unhandled data and patterns

6. Match

- Building a QualityStage job to identify matching records
- Applying multiple Match passes to increase efficiency
- Interpreting and improving Match results

7. Survive

- Building a QualityStage survive job that will consolidate matched records into a single master record

8. Two-Source Match

- Building a QualityStage job to match data using a reference match

REQUIREMENTS:

Participants should have:

- Familiarity with the Windows operating system
- Familiarity with a text editor

Helpful, but not required, would be some understanding of elementary statistics principles such as weighted averages and probability.

Difficulty level

