

□

IBM InfoSphere QualityStage Essentials v11.5  
Information

**Length:** 32.0 Hours  
**Ref:** KM213G  
**Delivery method:** Classroom  
**Price:** AUD

Overview

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.

Public

- Data Analysts responsible for data quality using QualityStage
- Data Quality Architects
- Data Cleansing Developers

Prerequisites

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.

Objective

Prior to enrolling, IBM Employees must follow their Division/Department processes to obtain approval to attend this public training class. Failure to follow Division/Department approval processes may result in the IBM Employee being personally responsible for the class charges.

GBS practitioners that use the EViTA system for requesting external training should use that same process for this course. Go to the EViTA site to start this process:

<http://w3.ibm.com/services/gbs/evita/BCSVTEnrl.nsf>

Once you enroll in a GTP class, you will receive a confirmation letter that should show:

- The current GTP list price
- The 20% discounted price available to IBMers. This is the price you will be invoiced for the class.

Topics

## 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