

□

InfoSphere BigMatch v11.4 for Apache Hadoop  
Information

**Length:** 16.0 Hours  
**Ref:** ZZ850G □  
**Delivery method:** Classroom  
**Price:** AUD

Overview

The IBM InfoSphere Big Match on Hadoop course will introduce students to the Probabilistic Matching Engine (PME) and how it can be used to resolve and discover entities across multiple data sets in Hadoop.

Students will learn the basics of a PME algorithm including data model configuration, standardization, comparison and bucketing functions, weight generation, and threshold.  
During the exercises, the student will work on a large use case, where they will apply their knowledge of Big Match to discover relationships be two data sets that can be used to understand the full view of the member data.

Public

The course is designed for a technical audience that will be setting up a custom algorithm for the Probabilistic Matching Engine to use Big Match on Apache Hadoop to compare, match and/or search member records across multiple data sets.

Prerequisites

This course has no pre-requisites.

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. Introduction to Big Match for Apache Hadoop
  - What is Big Match
  - How Big Match Works
  - Big Match Components
  - Big Match Architecture
2. Big Match Data Model Definition
  - Members
  - Attribute Types
  - Member Attributes
  - Sources
  - Information Sources
3. PME Algorithm
  - Standardization
  - Bucketing
  - Comparison Functions
4. Bucket Analysis
  - Bucket Optimization
  - Bucket Concerns
5. Weights
  - String Weights
  - Numeric Weights
  - Multi-dimensional Weights
  - Troubleshooting Weights
6. HBase Tables
  - HBase concepts
  - Big Match commands
  - Big Match Tables (.pmebktidx, .pmemdmidx, .pmeentidx)
  - Best Practices
7. BigMatch Applications
  - PME Derive
  - PME Compare
  - PME Link
  - PME Analysis

