



---

## InfoSphere BigMatch for Hadoop (v11.4)

**Duration: 2 Days**    **Course Code: ZZ850G**    **Delivery Method: Elearning (Self-paced)**

---

### 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 between two data sets that can be used to understand the full view of the member data. 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 between two data sets that can be used to understand the full view of the member data.

---

### Target Audience:

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.

---

### Objectives:

- Understand the capabilities of the Probabilistic Matching Engine
  - Understand how the Probabilistic Matching engine is used with Big Insights to solve certain use cases.
  - Understand the technical framework of the Big Match solution and how member data is derived, bucketed and compared to produce a complete entity from multiple data sets.
  - Create a project and data model using the Big Match Console
  - Configure the HBase tables that will be used in a Big Match solution
  - Configure an algorithm using the Big Match console that includes Standardization, Comparison and Bucketing functions.
  - Set up Strings for Anonymous value, Equivalency values, Frequency values, and character maps using the Big Match console
  - Set up and run the Weight Generation process
  - Evaluate and set thresholds for the algorithm
  - Deploy a new algorithm to Big Match
  - Evaluate Entity results and reconfigure algorithm based on evaluation. E.g. Large Buckets, Large Entities, Member not belonging to any buckets, etc
- 

### Prerequisites:

This course has no pre-requisites.

---

## Content:

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

---

## Further Information:

For More information, or to book your course, please call us on 353-1-814 8200

[info@globalknowledge.ie](mailto:info@globalknowledge.ie)

[www.globalknowledge.ie](http://www.globalknowledge.ie)

Global Knowledge, 3rd Floor Jervis House, Millennium Walkway, Dublin 1