IBM ZZ850G - InfoSphere BigMatch v11.4 for Apache Hadoop

Live Online oder Präsenz

Dauer : 2 Tage (14 Stunden)

Nr. : 30076

Preis : 1.700,00 € netto
2.023,00 € inkl. 19 % MwSt.

Inhouse-Paket : Auf Anfrage

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.

Wer sollte teilnehmen:

Zielgruppe

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.

Voraussetzungen

Prerequisites

This course has no pre-requisites.

Trainingsprogramm

Course Outline

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

Objective

1. Understand the capabilities of the Probabilistic Matching Engine
2. Understand how the Probabilistic Matching engine is used with Big Insights to solve certain use cases.
3. 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.
4. Create a project and data model using the Big Match Console
5. Configure the HBase tables that will be used in a Big Match solution
6. Configure an algorithm using he Big Match console that includes Standardization, Comparison and Bucketing functions.
7. Set up Strings for Anonymous value, Equivalency values, Frequency values, and character maps using the Big Match console
8. Set up and run the Weight Generation process
9. Evaluate and set thresholds for the algorithm
10. Deploy a new algorithm to Big Match
11. Evaluate Entity results and reconfigure algorithm based on evaluation. E.g. Large Buckets, Large Entities, Member not belonging to any buckets, etc

Schulungsmethode

presentation, discussion, hands-on exercises

Hinweis

Further seminars from the IBM area.