

Module Title : DW653G - BigInsights Analytics for Programmers

Duration : 2 days

Overview

This course is designed to aid programmers who are working with IBM's InfoSphere BigInsights. Writing programs that extract data from unstructured text can be a daunting task. You will learn how to create annotators through the use of IBM's Annotation Query Language (AQL). Analyzing data using Apache's Hadoop normally requires that MapReduce programs be written. You will learn how to use Jaql to create high level programs that are decomposed into Hadoop MapReduce programs. People familiar with the Hadoop technology are aware of other open source products that are used in this environment. This course will give you a foundation to program using the Apache Pig language. There is also overview of ZooKeeper, HBase, and Hive. One of the capabilities of the BigInsights Web Console is the ability to simplify the invocation of Hadoop applications. This course shows how to publish a text analytics application from the BigInsights development environment to a BigInsights server.

Training Paths that reference this course are:

- Streams & BigInsights

Objectives

- Describe the AQL data model
- Use AQL to create annotators
- List the different extractors used to create AQL views
- Describe how Jaql is used in a BigInsights environment
- Code Jaql core operators to manipulate JSON data
- Explain how to use SQL in Jaql
- Describe how Apache's Pig can be used in a BigInsights environment
- Use Pig core operators to analyze Hadoop data
- Explain how ZooKeeper can be used to manage barriers or queues
- Compare HBase to a Relational Database Management System
- Publish an application to a BigInsights server

Audience

This intermediate course is for anyone with a need to learn the IBM Annotation Query Language (AQL), Jaql or Pig.

Prerequisites

You should have attended:

- IBM DB2 OLAP Server Administration and Performance Tuning (DW610) or equivalent knowledge
- A programming background would be advantageous especially knowledge of SQL

Course Outlines

- Introduction to Programming for BigInsights
- Annotation Query Language
- AQL - Validating Results
- Jaql
- Introduction to Apache Pig
- Publish a Text Analytics Application to the Web Console