



Course Name **Fundamental Course in Big Data Programming using IBM InfoSphere BigInsight**

Course Duration 50 Hours

About Technology IBM® InfoSphere® BigInsights™ is an analytics platform, based on open source Apache Hadoop, for analyzing massive volumes of unconventional data in its native format. The software enables advanced analysis and modeling of diverse data, and supports structured, semi-structured and unstructured content to provide maximum flexibility.

About Course 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. The student will learn how to create annotators through the use of IBM's Annotation Query Language (AQL). Analyzing data using Apache's Hadoop requires that map / reduce programs be written. The student will learn how to use Jaql to create high level programs that are decomposed into Hadoop map / reduce programs. People familiar with the Hadoop technology are aware of other open source products that are used in this environment. This course will give the student an overview of Apache Pig, ZooKeeper, and Map / Reduce.

Target Audience This course is intended for Programmers with the need to learn AQL and Jaql.
Students of:

- Under Graduate & Post Graduate Computer Science / Information Technology

Pre-requisites

- Fundamentals of Big Data & Hadoop using IBM InfoSphere Big Insight Course
- A programming background would be advantageous especially knowledge of Java & SQL

Contents

1. Introduction to Programming in BigInsights
2. Annotation Query Language
3. Create Views using Regular Expressions, Dictionaries, and Splits
4. Create Views using Part of Speech, Blocks, and Patterns
5. Create Views using Select Statements and Tables
6. Develop and AQL Extraction Application
7. Understanding Jaql
8. Working with Jaql
9. Understanding Apache's Pig