



Course Name IBM CE - Foundation Course in Big Data using IBM InfoSphere BigInsights
Course Code IMBDH2FIN
Course Duration 25 Hours
About Technology IBM InfoSphere BigInsights brings the power of Hadoop to the enterprise. Apache Hadoop is the open source software framework, used to reliably managing large volumes of structured and unstructured data.

BigInsights makes it simpler for people to use Hadoop and build big data applications. It enhances this open source technology to withstand the demands of your enterprise, adding administrative, discovery, development, provisioning, and security features, along with best-in-class analytical capabilities from IBM Research. The result is that you get a more developer and user-friendly solution for complex, large scale analytics.

InfoSphere BigInsights allows enterprises of all sizes to cost effectively manage and analyze the massive volume, variety and velocity of data that consumers and businesses create every day. InfoSphere BigInsights can help you increase operational efficiency by augmenting your data warehouse environment. It can be used as a query-able archive, allowing you to store and analyze large volumes of multi-structured data without straining the data warehouse. It can be used as a pre-processing hub, helping you to explore your data, determine what is the most valuable, and extract that data cost-effectively. It can also allow for ad hoc analysis, giving you the ability to perform analysis on all of your data.

About Course This hands-on training course is for those who want a basic understanding of IBM InfoSphere BigInsights. It will give you an overview of IBM's Big Data strategy as well as detailed information on Apache Hadoop.

Module 1

This module is designed to provide the students with a fundamental understanding of Business Intelligence and how a Data Warehouse supports this environment. From a Data Warehouse perspective, the course gives an overview of the mechanisms used to build a Data Warehouse. From a Business Intelligence perspective, it covers some analysis techniques such as OLAP and data mining.

Module 2

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 System administrators and developers.

- Students of:
- CS/IT/ ECE/EEE – 2nd / 3rd Year
 - MCA – 2nd/ 3rd Year
 - BCA/ BSC – 3rd year
 - MBA



Pre-requisites Prior to attending this course, knowledge of Linux would be beneficial.

Contents

After completing this course, you should be able to:

Unit 1 - Introduction to Big Data

Unit 2 - Introduction to InfoSphere BigInsights

Unit 3 - Apache Hadoop and HDFS Overview

Exercise 1- Exploring Apache Hadoop

Unit 4 - GPFS-FPO

Unit 5 - BigInsights Web Console Security

Unit 6 - Introduction to MapReduce

Exercise 2- MapReduce

Unit 7 - Adaptive MapReduce

Unit 8 - Setup, Configuration, and Administration of a Hadoop Cluster

Exercise 3- Hadoop Configuration

Unit 9 - Overview of Oozie

Exercise 4- Controlling Workloads with Oozie

Exercise 4- Controlling Workloads with Oozie continued

Unit 10 - Managing Job Execution

Unit 11 - Moving Data into Hadoop

Exercise 5 - Using Flume for Data Loading