

□

InfoSphere BigInsights Foundation
Information

Length: 2.5 Days
___Ref: DW612G-X
Delivery method: Classroom
Price: AUD

Overview

This training course is for those who want a foundation of IBM InfoSphere BigInsights. It will give you an overview of IBM's Big Data strategy as well as a more detailed information on Apache Hadoop. It presents concepts required by a system administrator to work with the Hadoop Distributed File System and concepts of MapReduce that are required by a developer. It gives an introduction to the scheduling capabilities of Hadoop and how to use Oozie to control workflows and use Flume to load data into HDFS.

Curriculum relationship:

- *Programming for InfoSphere Streams V3 with SPL (DW720)*
- *InfoSphere Streams Administration (DW730)*
- *IBM InfoSphere BigInsights Analytics for Business Analysts (DW640)*
- *IBM InfoSphere BigInsights Analytics for Programmer (DW650)*

Public

This intermediate training course is for those who want a foundation of IBM InfoSphere BigInsights and is designed for system administrators and developers.

Prerequisites

There are no prerequisites for this course. However, knowledge of Linux would be beneficial.

Objective

- Describe functions and features of InfoSphere BigInsights
- List the capabilities of Hadoop and HDFS
- Administer HDFS
- Describe the use of MapReduce
- Setup a Hadoop cluster
- Manage job execution
- Explain the Oozie workflows
- Describe some scenarios for loading data into HDFS

Topics

Day 1

- Unit 1 - Introduction to Big Data
- Unit 2 - Introduction to InfoSphere BigInsights
- Exercise 1 - Installing BigInsights
- Unit 3 - Apache Hadoop and HDFS
- Exercise 2- Exploring Apache Hadoop
- Unit 4 - GPFS-FPO
- Unit 5 - BigInsights Web Console

Day 2

- Exercise 3 - BigInsights Web Console
- Unit 6 - Introduction to MapReduce - continued
- Exercise 4 - MapReduce
- Unit 7 - Adaptive MapReduce
- Unit 8 - Setup and Configurations of BigInsights Clusters
- Exercise 5 - Hadoop Configuration

Day 3

- Unit 9 - Overview of Oozie
- Exercise 6 - Scheduling with Oozie
- Unit 10 - Managing Job Execution
- Unit 11 - Moving Data into Hadoop
- Exercise 6 - Using Flume for Data Loading