



IBM BigInsights Foundation
Information

Length: 24.0 Hours
Ref: DW613G
Delivery method: Classroom
Price: AUD

Overview

This training course is for those who want a foundation of IBM BigInsights. This course consists of two separate modules.

The first module is IBM BigInsights Overview and it will give you an overview of IBM's big data strategy as well as a why it is important to understand and use big data. It will cover IBM BigInsights as a platform for managing and gaining insights from your big data. As such, you will see how the BigInsights have aligned their offerings to better suit your needs with the IBM Open Platform (IOP) along with the three specialized modules with value-add that sits on top of the IOP. Along with that, you will get an introduction to the BigInsights value-add including Big SQL, BigSheets, and Big R.

The second module is IBM Open Platform with Apache Hadoop. IBM Open Platform (IOP) with Apache Hadoop is the first premiere collaborative platform to enable Big Data solutions to be developed on the common set of Apache Hadoop technologies. The Open Data Platform initiative (ODP) is a shared industry effort focused on promoting and advancing the state of Apache Hadoop and Big Data technologies for the enterprise. The current ecosystem is challenged and slowed by fragmented and duplicated efforts between different groups. The ODP Core will take the guesswork out of the process and accelerate many use cases by running on a common platform. It allows enterprises to focus on building business driven applications.

This module provides an in-depth introduction to the main components of the ODP core --namely Apache Hadoop (inclusive of HDFS, YARN, and MapReduce) and Apache Ambari -- as well as providing a treatment of the main open-source components that are generally made available with the ODP core in a production Hadoop cluster.

IBM BigInsights v4 itself is built upon the ODP core and these other main open-source components. The relationships between the IBM Open Platform with Apache Hadoop and the BigInsights add-ons is covered briefly in Unit 1 - pro.

Public

This intermediate training course is for those who want a foundation of IBM BigInsights. This includes:

- Big data engineers
- Data scientist
- Developers or programmers
- Administrators who are interested in learning about IBM's Open Platform with Apache Hadoop.

This course consists of two separate modules. The first module is IBM BigInsights Overview and it will give you an overview of IBM's big data strategy as well as a why it is important to understand and use big data. The second module is IBM Open Platform with Apache Hadoop. IBM Open Platform (IOP) with Apache Hadoop is the first premiere collaborative platform to enable Big Data solutions to be developed on the common set of Apache Hadoop technologies.

Prerequisites

There are no pre-requisites for this course but knowledge of Linux would be beneficial.

Objective

Prior to enrolling, IBM Employees must follow their Division/Department processes to obtain approval to attend this public training class. Failure to follow Division/Department approval processes may result in the IBM Employee being personally responsible for the class charges.

GBS practitioners that use the EViTA system for requesting external training should use that same process for this course. Go to the EViTA site to start this process:

<http://w3.ibm.com/services/gbs/evita/BCSVTEnrl.nsf>

Once you enroll in a GTP class, you will receive a confirmation letter that should show:

- The current GTP list price
- The 20% discounted price is available to IBMers for GTP delivered courses only. This is the price you will be invoiced for the class.

Topics

(DW6A1)

- Unit 1: Introduction to Big Data
- Exercise 1: Setting up the lab environment
- Unit 2: Introduction to IBM BigInsights
- Exercise 2: Getting started with IBM BigInsights
- Unit 3: IBM BigInsights for Analysts
- Exercise 3: Working with Big SQL and BigSheets
- Unit 4: IBM BigInsights for Data Scientist
- Exercise 4: Analyzing data with Big R, Jaql, and AQL

- Unit 5: IBM BigInsights for Enterprise Management

(DW6B1)

- Unit 1: IBM Open Platform with Apache Hadoop
- Exercise 1: Exploring the HDFS
- Unit 2: Apache Ambari
- Exercise 2: Managing Hadoop clusters with Apache Ambari
- Unit 3: Hadoop Distributed File System
- Exercise 3: File access & basic commands with HDFS
- Unit 4: MapReduce and Yarn
- Topic 1: Introduction to MapReduce based on MR1
- Topic 2: Limitations of MR1
- Topic 3: YARN and MR2
- Exercise 4: Creating and coding a simple MapReduce job (Possibly a more complex second Exercise)
- Unit 5: Apache Spark
- Exercise 5: Working with Spark's RDD to a Spark job
- Unit 6: Coordination, management, and governance
- Exercise 6: Apache ZooKeeper, Apache Slider, Apache Knox
- Unit 7: Data Movement
- Exercise 7: Moving data into Hadoop with Flume and Sqoop
- Unit 8: Storing and Accessing Data
- Topic 1: Representing Data: CSV, XML, JSON, and YAML
- Topic 2: Open Source Programming Languages: Pig, Hive, and Other [R, Python, etc]
- Topic 3: NoSQL Concepts
- Topic 4: Accessing Hadoop data using Hive
- Exercise 8: Performing CRUD operations using the HBase shell
- Topic 5: Querying Hadoop data using Hive
- Exercise 9: Using Hive to Access Hadoop / HBase Data
- Unit 9: Advanced Topics
- Topic 1: Controlling job workflows with Oozie
- Topic 2: Search using Apache Solr
- No lab exercises