



Enterprise Computing Solutions - Education Services

TRAINING OFFERING

You can reach us at:

Arrow ECS, Nidderdale House, Beckwith Knowle, Harrogate, HG3 1SA

Email: educationteam.ecs.uk@arrow.com
Phone: 0870 251 1000



IBM BigInsights Foundation - SPVC

CODE: **LENGTH:** **PRICE:**

ZL1_2W613 24 Hours £720.00

Description

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.

If you are enrolling in a Self Paced Virtual Classroom or Web Based Training course, before you enroll, please review the **Self-Paced Virtual Classes and Web-Based Training Classes** on our Terms and Conditions page, as well as the system requirements, to ensure that your system meets the minimum requirements for this course.

<http://www.ibm.com/training/terms>

Objectives

IBM BigInsights Overview (DW6A1)

- Understand the purpose of big data and know why it is important
- List the sources of data (data-at-rest vs data-in-motion)
- Describe the IBM BigInsights offering
- Utilize the various IBM BigInsights tools including Big SQL, BigSheets, Big R, Jaql and AQL for your big data needs.

IBM Open Platform (IOP) with Apache Hadoop (DW6B1)

- List and describe the major components of the open-source Apache Hadoop stack and the approach taken by the Open Data Foundation.
- Manage and monitor Hadoop clusters with Apache Ambari and related components
- Explore the Hadoop Distributed File System (HDFS) by running Hadoop commands.
- Understand the differences between Hadoop 1 (with MapReduce 1) and Hadoop 2 (with YARN and MapReduce 2).
- Create and run basic MapReduce jobs using command line.
- Explain how Spark integrates into the Hadoop ecosystem.
- Execute iterative algorithms using Spark's RDD.
- Explain the role of coordination, management, and governance in the Hadoop ecosystem using Apache Zookeeper, Apache Slider, and Apache Knox.
- Explore common methods for performing data movement
- Configure Flume for data loading of log files
- Move data into the HDFS from relational databases using Sqoop
- Understand when to use various data storage formats (flat files, CSV/delimited, Avro/Sequence files, Parquet, etc.).
- Review the differences between the available open-source programming languages typically used with Hadoop (Pig, Hive) and for Data Science (Python, R)
- Query data from Hive.

- Perform random access on data stored in HBase.
- Explore advanced concepts, including Oozie and Solr

Audience

This intermediate course is for those who want a foundation of IBM BigInsights. This includes the following and those who are interested in learning about IBM's Open Platform with Apache Hadoop:

- Big data engineers
- Data scientist
- Developers or programmers
- Administrators

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. However, knowledge of Linux would be beneficial.

Programme

IBM BigInsights Overview (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

IBM Open Platform with Apache Hadoop (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, & YAML
- Topic 2: Open Source Programming Languages: Pig, Hive, & 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

Session Dates

On request. Please [Contact Us](#)

Additional Information

[This training is also available as onsite training. Please contact us to find out more.](#)