



Course Details

This 40-hour Instructor led training course, part of the Big Data program at AiQuest, is aimed at teaching Big data Application development and analysis in Apache Hadoop using Apache ecosystem tools like Pig, Hive and Spark. Students will learn the details of Hadoop, YARN, Hadoop Distributed File System (HDFS), MapReduce and deep-dive into practical lab sessions in Pig programming, Hive programming and Spark to perform data analytics in Big Data. Data Ingestion techniques using Sqoop and Flume, and workflow definitions using Oozie are also covered in this course. This training course is best suited for certification aspirants (HDPCD and CCA developer certifications) and prepares Hadoop developers for real-world challenges.

Prerequisites

Students should be familiar with any programming language or scripting. SQL and basic Unix knowledge is helpful but not necessary. No prior Hadoop knowledge or experience is required.

Course Outline

Big Data Overview We want to bring corporate quality and industry standard training to individuals seeking a career in Big Data. Our courses are modelled based on extensive industry experience and cater to current Industry needs to provide relevant practical experience and real-time working knowledge. Our elite courses cover core concepts in Big Data as offered by corporate solution partners- Horton Works, Cloudera and Pivotal

For those looking to certify, the course has been designed specifically to help take the certification examination with ease. Also, the courses are designed with an ideal theory to practical ratio of 60:40, ensuring learning conceptual knowledge backed by practical applicable skills relevant for the work force.

The courses are delivered by professional trainers who offer corporate trainings to companies and are working as consultants and architects on Big Data projects.

- Introduction to Apache Hadoop
- Hadoop Overview
- Hadoop ecosystem projects overview
- Apache Hadoop file storage



- HDFS overview
- HDFS Architecture
- Apache Hadoop Data Processing framework
- MapReduce Overview
- MapReduce Architecture
- YARN overview
- YARN Architecture
- Demo on MapReduce Jobs
- Data Ingestion
 - HDFS
 - Understand HDFS commands
 - Move file between HDFS and Local File system
 - Apache Sqoop
 - Architecture Overview
 - Sqoop programming
 - Sqoop programming and free-form query
 - Import & export RDBMS data using sqoop
 - Demonstration on Sqoop Import and export from RDBMS
 - Exercises – Lab
 - Apache Flume
 - Architecture Overview
 - Demonstration on HDFS commands on cluster
 - Demonstration on Flume log file capture
 - Exercises – Lab
- Data Transformation
 - Apache Pig
 - Overview
 - Data types in pig
 - Pig modes
 - Pig programming
 - Pig user defined function(UDF)
 - Pig TEZ MapReduce engine
 - Demonstration on Pig Programming
 - Exercises – Lab
- Data Analysis
 - Apache Hive



- Hive architecture
- Data types in Hive
- Hive programming
- Hive advanced programming
- Partition, bucketing, Joins
- Hive User defined function (UDF)
- Demonstration on Hive
- Exercises – Lab
- Apache Spark
 - Spark architecture overview
 - Spark programming
 - Demonstration on Spark programming
 - Exercises – Lab
- Apache HCatalog overview
 - Access hive tables from Pig
 - Access Pig scripts from Hive query's
 - Demonstration on HCatalog

Course Duration

40 hours (20 hours theory and concepts; 20 hours practical labs and demos)

This course is spread over 4 weekends (Saturday and Sunday) 6.30 AM to 10.30 AM Eastern Standard Time (GMT- 4:00 hrs). Course could be customised to the needs of participants.

Cost

Fee per participant varies based on the course delivery method and extent of customisation. The cost includes training, material and cloud-based lab fees.

Note: Please inquire with us for ongoing promotions and early bird prices.

Registration and Enquiries

Please contact AiQuest at connect@aiquestinc.com or call us on 514-910-6785.

Visit us at www.aiquestinc.com



About AiQuest

The mission of Ai Quest (AiQ) is to help organisations and knowledge workers to explore and realize their true potential in the Artificial Intelligence (AI) landscape. The true potential of Big Data in the AI realm goes beyond implementing new technologies and having appropriate data analytics. The strategy must include well trained resources, right performance measures that affect the corporate performance, exploiting existing technological resources to maximize the value and continuous investment in corporate training.

Course design philosophy

We want to bring corporate quality and industry standard training to individuals seeking a career in Big Data. Our courses are modelled based on extensive industry experience and cater to current Industry needs to provide relevant practical experience and real-time working knowledge. Our elite courses cover core concepts in Big Data as offered by corporate solution partners-Horton Works, Cloudera and Pivotal

For those looking to certify, the course has been designed specifically to help take the certification examination with ease. Also, the courses are designed with an ideal theory to practical ratio of 60:40, ensuring learning conceptual knowledge backed by practical applicable skills relevant for the work force.

The courses are delivered by professional trainers who offer corporate trainings to companies and are working as consultants and architects on Big Data projects.