COURSE DETAILS & CURRICULUM

Mastering Hadoop and related tools: The course provides you with an in-depth understanding of the Hadoop framework including HDFS, YARN, and MapReduce. You will learn to use Pig, Hive, and Impala to process and analyze large datasets stored in the HDFS, and use Sqoop and Flume for data ingestion.

At the end of this Big Data Certification training course, you will be able to:
* Describe the Big Data landscape including examples of real world big data problems including the three key sources of Big Data: people, organizations, and sensors.
* Explain the V's of Big Data (volume, velocity, variety, veracity, valence, and value) and why each impacts data collection, monitoring, storage, analysis and reporting.
* Get value out of Big Data by using a 5-step process to structure your analysis. * Identify what are and what are not big data problems and be able to recast big data problems as data science questions.
* Provide an explanation of the architectural components and programming models used for scalable big data analysis.
* Summarize the features and value of core Hadoop stack components including the YARN resource and job management system, the HDFS file system and the MapReduce programming model.
* Install and run a program using Hadoop!

This course is for those who are relatively new to data science. No prior programming experience is needed, although the ability to install applications and utilize a virtual machine is necessary to complete the hands-on assignments.