TITLE –BIG DATA

DURATION

40 Hours

OVERVIEW

Master the Hadoop ecosystem using HDFS, MapReduce, Yarn, Pig, Hive, Kafka, HBase, Spark, Knox, Ranger, Ambari, Zookeeper.
In this course you will learn Big Data using the Hadoop Ecosystem.

The course is aimed at Software Engineers, Database Administrators, and System Administrators that want to learn about Big Data. Other IT professionals can also take this course, but might have to do some extra research to understand some of the concepts.

You will learn how to use the most popular software in the Big Data industry at moment, using batch processing as well as realtime processing. This course will give you enough background to be able to talk about real problems and solutions with experts in the industry. Updating your LinkedIn profile with these technologies will make recruiters want you to get interviews at the most prestigious companies in the world.

The course is very practical

TARGET AUDIENCE

- You will need to have a background in IT. The course is aimed at Software Engineers, System Administrators, DBAs who want to learn about Big Data
- Knowing any programming language will enhance your course experience

- The course contains demos & Exercises you can try out on your own machine. To run the Hadoop cluster on your own machine, you will need to run a virtual server. 8 GB or more RAM is recommended.

APPROACH

This course contains examples and demonstrations and will walk you through everything you need to know step by step, by providing you with helpful tips along the way.

COURSE ROADMAP

SECTION ONE– Introduction to BIG DATA

- What is Big Data
- Example of Big Data
- Big Data Enabling Technologies
- Hadoop Stack for Big Data

SECTION TWO- HADOOP DFS - MAP REDUCE

- Hadoop Distributed File System (HDFS)
- Hadoop MapReduce 2.0
- MapReduce Examples

SECTION THREE- APACHE SPARK

- Parallel Programming with Spark
- Introduction to Spark
- Spark Built-in Libraries
- Design of Key-Value Stores

SECTION Four-Data Placement - CAP THEOREM

- Data Placement Strategies
- CAP Theorem
- Consistency Solutions
- Design of Zookeeper
- CQL (Cassandra Query Language)

SECTION Five-SPARK STREAMING

- Design of HBase
- Spark Streaming and Sliding Window Analytics
- Sliding Window Analytics
- KAFKA

SECTION Six- BIG DATA MACHINE LEARNING

- Big Data Machine Learning
- Machine Learning Algorithm K-means using Map Reduce for Big Data Analytics
- Parallel K-means using Map Reduce on Big Data Cluster Analysis

SECTION SEVEN- BIG DATA ANALYTICS

- Decision Trees for Big Data Analytics
- Big Data Predictive Analytics

SECTION Eight- Case Study - PagerRank Algorithm

- Parameter Servers
- PageRank Algorithm in Big Data
- Spark GraphX & Graph Analytics
- Case Study: Flight Data Analysis using Spark GraphX