

Hadoop and Big data Training



Course Description:

Organizations use their data to support and influence decisions and build data-intensive products and services, such as recommendation, prediction, and diagnostic systems. The collection of skills required by organizations to support these functions has been grouped under the term 'data science'. This statistics and data analysis course will attempt to articulate the expected output of data scientists

Interested in increasing your knowledge of the Big Data landscape? This course is for those new to data science and interested in understanding why the Big Data Era has come to be. It is for those who want to become conversant with the terminology and the core concepts behind big data problems, applications, and systems. It is for those who want to start thinking about how Big Data might be useful in their business or career. It provides an introduction to one of the most common frameworks, Hadoop, that has made big data analysis easier and more accessible -- increasing the potential for data to transform our world!

Hadoop Big Data Training course helps you learn the core techniques and concepts of Big Data and Hadoop ecosystem. It equips you with in-depth knowledge of writing codes using MapReduce framework and managing large data sets with HBase. The topics covered in this course mainly includes- Hive, Pig and setup of Hadoop Cluster.

Course Prerequisites:

Knowledge of programming in C++ or Java or any other Object Oriented Programming language is preferred, else you can enroll for our Java course free of cost to acquire the necessary skills to learn Hadoop.

Target Audience:

This course is specially designed for the B.Tech /B.E(CSE/IT/EEE/ECE/Mech) and all other IT related Graduates and Post Graduate students. Mission Professionalism has conquered the job scenario and companies seek for well qualified, professional and skilled manpower. Quality Education and Performance Oriented Training is our motto.

What Student/Professionals Will Learn?

- Understand Big Data and Hadoop ecosystem
- Work with Hadoop Distributed File System (HDFS)
- Write MapReduce programs and implementing HBase
- Write Hive and Pig scripts

COURSE-CONTENT

Module 1:-

- Virtual Box/VM Ware
 - Basics
 - Installations
 - Backups
 - Snapshots
- ClouderaVM
 - Installations
- Hadoop
 - Why Hadoop
 - Scaling
 - Distributed Framework
 - Hadoop v/s RDBMS

- Brief history of Hadoop
- Problems with traditional large-scale systems
- Requirements for a new approach
- Anatomy of a Hadoop cluster
- Other Hadoop Ecosystem components
- Setup Hadoop
 - Pseudo mode
 - Cluster mode
 - Installation of Java
 - Hadoop
 - Configurations of Hadoop
 - Hadoop Processes (NN, SNN, JT, DN, TT)
 - Temporary directory
 - UI
 - Common errors when running Hadoop cluster
 - Solutions

Module 2:-

- HDFS- Hadoop Distributed File System
 - HDFS design and architecture
 - HDFS concepts
 - Interacting HDFS using command line
 - Dataflow
 - Blocks
 - Replica

Module 3:-

- MapReduce
 - Developing MapReduce application
 - Phases in MapReduce framework
 - MapReduce input and output formats
 - Advanced concepts
 - Sample applications
 - Combiner
- Writing a MapReduce Program
 - The MapReduce flow

- Examining a sample MapReduce program
- Basic MapReduce API concepts
- Driver code, Mapper, Reducer
- Hadoop's streaming API
- Using Eclipse for rapid development
- Hands-on exercise
- New MapReduce API
- Common MapReduce Algorithms
 - Sorting and Searching
 - Indexing
 - Term Frequency – Inverse Document Frequency
 - Word Co-occurrence
 - Hands-on exercise
- Writing advance map reduce programs
 - Building multivalued writable data
 - Accessing and using counters, Partitioner - Hashpartitioner, Hands on Exercises

Module 4:-

- Hadoop Programming Languages HIVE
 - Introduction, Installation
 - Configuration
 - Interacting HDFS using HIVE
 - MapReduce programs through HIVE
 - HIVE commands, Loading
 - Filtering
 - Grouping
 - Data types
 - Operators
 - Joins
 - Groups
 - Sample programs in HIVE
- Pig
 - Basics, Configuration
 - Commands, Loading
 - Filtering
 - Grouping

- Data types
- Operators
- Joins
- Groups
- Sample programs in PIG
- HBase
 - What is HBase
 - HBase architecture
 - HBase API
 - Managing large data sets with HBase
 - Using HBase in Hadoop applications.

Module 5:-

- Integrating Hadoop into the Enterprise Workflow
 - Integrating Hadoop into an Existing Enterprise
 - Loading Data from an RDBMS into HDFS by Using Sqoop
 - Managing Real-Time Data Using Flume.

INTEGER Innovation will provide:

- Training Slides taught during training by trainers
- Programmatic Examples
- Assignments of each topic in a module
- Demos executed during training session.
- Software's and installation guide (for future help)
- E-books for further reading in depth
- Reference links
- 24X7 online support for any queries or doubts.