BIG DATA ANALYTICS AND HADOOP

DATA WILL TALK TO YOU IF YOU'RE READY TO LISTEN!!!

2 Days
80% hands -on session

The New Natural Resource

INDIA LABS

MAKING INDIA FUTURE READY!!!

www.htindialabs.com

/htindia.labs.delhi

+91 9911330807,
+91 7065657373

htindialabsworkshops@gmail.com,
info@htindialabs.com

INDIA LABS

MAKING INDIA FUTURE READY!!!

ABOUT BIG DATA HADOOP

Big data is a popular term used to describe the exponential growth and availability of data, both structured and unstructured. And big data may be as important to business – and society – as the Internet has become Hadoop is 100% open or free source, and pioneered a fundamentally new way of storing and processing data. Instead of relying on expensive, proprietary hardware and different systems to store and process data, Hadoop enables distributed parallel processing of huge amounts of data across inexpensive, industry-standard servers that both store and process the data, and can scale without limits. With Hadoop, no data is too big. And in today's hyper-connected world where more and more data is being created every day, Hadoop's breakthrough advantages mean that businesses and organizations can now find value in data that was recently considered useless. The students would get to work on a Real Life Project on Big Data Analytics and gain hands on project.

TOPICS TO BE COVERED:

Session 1: BigData

- How Big is this Big Data ?
- Definition with Real Time Examples
- How BigData is generated with Real Time
- Generation Use of BigData-How Industry is utilizing
- BigData Traditional Data Processing Technologies
- Future of BigData!!!

Session 2: Hadoop

- Why Hadoop?
- What is Hadoop?
- Hadoop vs RDBMS, Hadoop vs BigData
- Brief history of Hadoop
- Apache Hadoop Architecture
- Problems with traditional large-scale systems
- Requirements for a new approach
- Anatomy of a Hadoop cluster
- Hadoop Setup and Installation

Session 3: Hadoop Ecosystem

- Brief Introduction about Hadoop EcoSystem (MapReduce, HDFS, Hive, PIG, HBase).

Session 4: HDFS

- Concepts & Architecture
- Data Flow (File Read , File Write)
- Fault Tolerance
- Shell Commands
- Java Base API
- Data Flow Archives
- Coherency
- Data Integrity
- Role of Secondary NameNode
- HDFS Programming Basics

Session 5: MapReduce

- Theory
- MapReduce Architecture
- Data Flow (Map – Shuffle - Reduce)
- MapRed vs MapReduce APIs
- MapReduce Programming Basics
- Programming [Mapper, Reducer, Combiner, Partitioner]

Session 6: HIVE & PIG

- Architecture
- Installation
- Configuration
- Hive vs RDBMS
- Tables
- Partitioning & Bucketing
- Hive Web Interface
- Why Pig
- Use case of Pig

Session 7: HBase

- RDBMS Vs NoSQL
- HBase Introduction