**HADOOP TRAINING**

**Day Wise Plan**

**Day1:-**

**Introduction to BIG Data and Hadoop**

* Introduction to BIG Data
* Challenges associated with Big Data
* Requirement of BIG Data
* Introduction to Hadoop
* Utility of Hadoop
* Basic Components of Hadoop
* Introduction to Hadoop Ecosystems
* 5 daemons of Hadoop
* Overview of Hadoop Implementation

**Hadoop Distributed File Systems**

* Understanding HDFS Architecture
* Understanding How Namenode maintains the file system metadata
* Understand how data is stored in HDFS
* Understand the relationship between Namenode and Datanode
* Understand replication factor, under and over replication
* Namenode High Availability
* Rack Awareness
* Introduction and hands on to ***Hadoop fs*** commands
* File Reading and Writing over HDFS

**MapReduce**

* Overview of MapReduce
* Understand the architecture of MapReduce
* Understand the various phases of MapReduce through Word Count Paradigm
* Programming algorithms
* Combiners and Counters
* MapReduce Types and Formats

**Day2:-**

**YARN**

* Understand the architecture of YARN
* Understand the components of YARN Resource Manager
* Demonstrate the relationship between Node Managers and application masters.
* Demonstrate the relationship between Resource Managers and application masters.

**Hadoop Installation**

* Understand minimum hardware and software requirements
* CDH4 Hadoop Installation
* Understand complete deployment layout
* Understand how to configure and manage different services

**Commands Practice**

* Understand different configuration parameters.
* Practicing HDFS Commands
* Introduction to Hue- GUI Console

**Day3:-**

**PIG**

|  |
| --- |
| * Introduction to PIG * PIG Components * PIG Data Types * Functions and Macros * Writing Scripts in PIG * PIG Execution   **Hive**   * Introduction to Hive * Hive Architecture * HiveQL * Partitioning and Bucketing * Managed tables and External tables * Joining * Functions * SerDes   **Hbase** |

* Introduction to NoSQL Databases
* Why to use NoSQL Databases
* CAP Theorem
* HBase Concept
* HBase Architecture
* HBase Data Model, Bloom Filter, Block Cache, Schema Design

**Impala**

* Real Time Analysis
* Impala Overview
* Impala Architecture

**Day4:-**

**Practical Session**

* Practical Mini Projects on PIG
* Practical Mini Projects on Hive
* Practical Mini Projects on HBase
* Practical Mini Projects on Impala

**Basic Administrative Activities**

* Service Management
* Cluster Installation
* Commissioning and Decommissioning Nodes
* Adding more clusters
* Checking Alerts and Events

**Day5:-**

**SQOOP**

* Introduction to Sqoop
* Sqoop Tools and Commands
* Data Importing to HDFS, Hive, HBase etc
* Data Exporting and connectors

**Flume and Chukwa**

* Introduction to Flume
* Source, Channel and Sink
* Data Analysis using Chukwa
* Practical Use Cases

**Oozie**

* Data Level Scheduling
* Time Bound Scheduling
* Workflow, Coordinator and Bundle

**Hadoop Security and Monitoring**

* Why Hadoop Security Is Important
* Hadoop’s Security System Concepts
* What Kerberos Is and How it Works
* Securing a Hadoop Cluster with Kerberos

**Spark**

**EXIT TEST**