KKAR Technologies

....Career Through opportunities

Hadoop Development Course Content

1. Introduction to Big Data

- ✓ What is Big Data?
- ✓ Examples of Big Data
- ✓ Reasons of Big Data Generation
- ✓ Why Big Data deserves your attention
- ✓ Use cases of Big Data
- ✓ Different options of analyzing Big Data

2. Introduction To Hadoop

- ✓ What is Hadoop
- ✓ History of Hadoop
- ✓ How Hadoop name was given
- ✓ Problems with Traditional Large-Scale Systems and Need for Hadoop
- ✓ Understanding Hadoop Architecture
- ✓ Fundamental of HDFS (Blocks, Name Node, Data Node, Secondary Name Node)
- ✓ Rack Awareness
- ✓ Read/Write from HDFS
- ✓ HDFS Federation and High Availability

3. Starting Hadoop

- ✓ Setting up single node Hadoop cluster(Pseudo mode)
- ✓ Understanding Hadoop configuration files
- ✓ Hadoop Components- HDFS, Map Reduce
- ✓ Overview Of Hadoop Processes
- ✓ Overview Of Hadoop Distributed File System
- ✓ The building blocks of Hadoop
- ✓ Hands-On Exercise: Using HDFS commands

4. MapReduce-1(MR V1)

- ✓ Understanding Map Reduce
- ✓ Job Tracker and Task Tracker
- ✓ Architecture of Map Reduce
- ✓ Map Function
- ✓ Reduce Function
- ✓ Data Flow of Map Reduce
- ✓ How Map Reduce Works
- ✓ Anatomy of Map Reduce Job (MR-1)
- ✓ Submission & Initialization of Map Reduce Job
- ✓ Assigning & Execution of Tasks
- ✓ Monitoring & Progress of Map Reduce Job
- ✓ Hadoop Writable and Comparable
- ✓ Map Reduce Types and Formats
- ✓ Understand Difference Between Block and Input Split
- ✓ Role of Record Reader
- ✓ Different File Input Formats
- ✓ Map Reduce Joins

5. MapReduce-2(YARN)

- ✓ Limitations of Current Architecture
- ✓ YARN Architecture
- ✓ Application Master, Node Manager & Resource Manager
- ✓ Job Submission and Job Initialization
- ✓ Task Assignment and Task Execution
- ✓ Progress and Monitoring of the Job
- ✓ Failure Handling in YARN
- ✓ Task Failure

6. Hive

- ✓ Introduction to Apache Hive
- ✓ Architecture of Hive

- ✓ Installing Hive
- ✓ Hive data types
- ✓ Hive-HQL
- ✓ Types of Tables in Hive
- ✓ Partitions
- ✓ Parquet file
- ✓ Sequence file
- ✓ RC FILE
- ✓ ORC file
- ✓ SERD
- ✓ Buckets& Sampling
- ✓ Indexes
- ✓ Views
- ✓ Executing hive queries from Linux terminal
- ✓ Executing hive queries from a file
- ✓ Creating UDFs in HIVE
- ✓ Hands-On Exercise
- ✓ Hive with Hbase Integration
- ✓ Security

7. Pig

- ✓ Introduction to Apache Pig
- ✓ Install Pig
- ✓ Architecture
- ✓ Data types
- ✓ Working with various PIG Commands covering all the functions in PIG
- ✓ Working with un-structured data
- ✓ Working with Semi-structured data
- ✓ Creating UDFs
- ✓ Hands-On Exercise
- ✓ Pig with Hbase Integration
- ✓ Pig with

8. Sqoop

- ✓ Introduction to SQOOP& Architecture
- ✓ Installation of SQOOP
- ✓ Import data from RDBMS to HDFS

- ✓ Importing Data from RDBMS to HIVE
- ✓ Exporting data from HIVE to RDBMS
- ✓ Hands on exercise

9. HBASE

- ✓ Introduction to HBASE
- ✓ Installation of HBASE
- ✓ Exploring HBASE Master & Region server
- ✓ Exploring Zookeeper
- ✓ CRUD Operation of HBase with Examples
- ✓ HIVE integration with HBASE
- ✓ Hands on exercise on HBASE

10. OOZIE

- ✓ Both standalone and cluster
- ✓ OOZIE meta data management

11. FLUME

- ✓ Hadoop with twiter data
- ✓ Twiter data with analytics

12. Spark

- ✓ introduction
- ✓ spark core
- ✓ spark streaming
- ✓ spark sql
- ✓ spark installation
- ✓ integration with hadoop

13. kafka

- ✓ introduction
- ✓ producer
- ✓ producer configuration
- ✓ broker
- ✓ multiple broker configuration
- ✓ consumer
- ✓ kafka client

14. FAQs, Real time Environment & Real time scenarios

- **REAL TIME PROJECT**

- ✓ We will be providing raw data & requirements for the project & you will have to work. Finally we will have one Project execution session where we will be explaining the steps for execution.