

BIG DATA SYLLABUS

Day 1 - Linux and Introduction

- 1.Linux fundamentals
- 2.Linux Commands
- 3.Introduction to Hadoop

Day 2 - Hadoop Architecture and HDFS

- 1.Distributed Systems
- 2.Hadoop Components
- 3.Hadoop Distributed File system

Day 3 - Sqoop and Scala Programming

- 1.Data Congestion using Sqoop
- 2.Scala Basics
- 3.Functions and Collection of Scala
- 4.OOP in scale

Day 4 - Spark FrameWork

- 1.Need of Apache Spark
- 2.Spark Architecture
- 3.Writing Spark Applications
- 4.Spark RDDs
- 5.RDD Transformations and Actions
- 6.Data through RDD
- 7.RDD Lineage
- 8.RDD Example

Day 5 - Query data using hive

- 1.Introduction to Hive
- 2.Hive Architecture and Query Language

Day 6 - Spark SQL and Data Frames

- 1.Introduction to Spark SQL
- 2.SQL Queries and Creating DataSet
- 3.Spark SQL Usecase
- 4.Data Frames in Spark
- 5.Introduction to Data Warehousing

Day 7 - Data Warehousing

- 1.Data Warehouse and Architecture
- 2.Dimensions and Facts
- 3.Normalization and Schemas
- 4.Modeling and Data Model
- 5.ETL
- 6.Models,ER Diagram and Attribute

Day 8 - Real Time Data Processing

- 1.Introduction to Real-time Processing
- 2.Introduction to Flume
- 3.Flume Architectutre
- 4.Introduction to Apache Kafka

Day 9 - Kafka and Spark Streaming

- 1.Configuring Kafka Multi-node Cluster
- 2.Spark Streaming Overview