65/1, GST ROAD, GUINDY, CHENNAI.

BIG DATA ANALYTICS
USING HADOOP ECOSYSTEM

Introduction

Big Data – 3Vs – Role of Hadoop in Big data – Overview of other Big Data Systems – Requirements in Hadoop - Anatomy - Installation - Configuration - UseCases of Hadoop

HDFS

Design - Architecture – Data Flow – CLI Commands – Java API – Data Flow Archives – Data Integrity WebHDFS - Compression

MapReduce

Theory – Data Flow (Map – Shuffle - Reduce) - Programming [Mapper, Reducer] –Writables – InputFormat – Outputformat - Inherent Failure Handling – Magic of Shuffle Phase -FileFormats

Administration

Multi Node Cluster Setup using AWS Cloud machines –Hardware Considerations - Commands (fsck, job, dfsadmin) - Schedulers- RackAwareness Policy - Balancing - NameNode Failure and Recovery - commissioning and Decommissioning a Node

HBase

NoSQL vs SQL – CAP Theorem – Architecture - Installation – Configuration - Role of Zookeeper – Java Based APIs– Map reduce Integration – Performance Tuning

Hive

Architecture – Installation –Configuration – Hive vs RDBMS - Tables – DDL – DML – UDF – UDAF – Partitioning – Bucketing – MetaStore - Hive-Hbase Integration – Hive Web Interface – Hive Server(JDBC,ODBC) – File Formats (RCFile - ORCFile)

Pig

Architecture –Installation - Hive vs Pig - Pig Latin Syntax –Data Types –Functions (Eval, Load/Store, String, DateTime) - Joins - Pig Server –Macros- UDFs- Performance - TroubleShooting – Commonly Used Functions

Sqoop

Architecture , Installation, Commands(Import , Hive-Import, EVal, Hbase Import, Import All tables, Export)

Flume

Architecture, Configuration (Agents), Sources(Exec-Avro-NetCat), Channels(File,Memory,JDBC, HBase), Sinks(Logger, Avro, HDFS, Hbase, FileRoll), Contextual Routing (Interceptors, Channel Selectors)

Oozie

Architecture, Installation, Workflow, Coordinator, Action (Mapreduce, Hive, Pig, Sqoop)

Hadoop 2.0

HDFS Federation - High Availability – HDFS Snapshots - YARN aka MR2 – Tez – BackWard Compatibility for Hadoop 1.X

Duration: 32 Hours.