



HADOOP TRAINING

HADOOP Training Course Syllabus

INTRODUCTION

- Big Data
- 3Vs
- Role of Hadoop in Big data
- Hadoop and its ecosystem
- Overview of other Big Data Systems
- Requirements in Hadoop
- Use Cases 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, Combiner, Partitioner]
- Writables
- Input Format
- Output format



- Streaming API

ADVANCED MAPREDUCE PROGRAMMING

- Counters
- CustomInputFormat
- Distributed Cache
- Side Data Distribution
- Joins
- Sorting
- ToolRunner
- Debugging
- Performance Fine tuning

ADMINISTRATION – Information required at Developer level

- Hardware Considerations – Tips and Tricks
- Schedulers
- Balancers
- NameNode Failure and Recovery

HBase

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

HIVE



- Architecture
- Tables
- DDL – DML – UDF – UDAF
- Partitioning
- Bucketing
- Hive-Hbase Integration
- Hive Web Interface
- Hive Server

OTHER HADOOP ECOSYSTEMS

- Pig (Pig Latin , Programming)
- Sqoop (Need – Architecture ,Examples)
- Introduction to Components (Flume, Oozie,ambari)