

HADOOP DEVELOPMENT WITH HORTON WORKS DISTRIBUTION PLATFORM

Unit 1: Understanding Hadoop

- ❖ What is Big Data?
- ❖ Characteristics Of Big Data
- ❖ Big Data Reality
- ❖ Most Common New Types of Data
- ❖ What is Hadoop?
- ❖ What Hadoop is not?
- ❖ Relational Databases vs. Hadoop
- ❖ Companies using Hadoop
- ❖ Big Data Opportunities
- ❖ What is Hadoop 2.0?
- ❖ What's new in Hadoop 2.0?
- ❖ The Hadoop Ecosystem
- ❖ Who is Hortonworks?

Unit 2: The Hadoop Distributed File System (HDFS)

- ❖ What are Hadoop Core Components?
- ❖ What is HDFS?
- ❖ HDFS Keywords
- ❖ HDFS Design Details
- ❖ HDFS Features
- ❖ HDFS Architecture
- ❖ HDFS Components in detail
 - Name Node
 - Data Node
 - Secondary Name Node
- ❖ Anatomy of Write
- ❖ Anatomy of Read
- ❖ Rack Awareness
- ❖ Relationship between NN and DN's
- ❖ Hadoop Modes and configuration files
- ❖ HDFS Commands
- ❖ **Hands on exercise HDFS Commands**

Unit 3: Hadoop 2.0 and YARN

- ❖ What is HDFS High Availability?
- ❖ Quorum Journal Manager
- ❖ Configuring Automatic Failover
- ❖ What is HDFS Federation?
- ❖ Multiple Federated NameNodes
- ❖ Multiple Namespaces
- ❖ What is YARN?
- ❖ The components of YARN
- ❖ Lifecycle of YARN Application
- ❖ A Cluster View Example

Unit 4: The Map Reduce Framework

- ❖ What is Map Reduce?
- ❖ Overview of MapReduce
- ❖ Understanding MapReduce
- ❖ The Key/Value pairs of MapReduce
- ❖ Anatomy of MapReduce job run
- ❖ Job Submission, Job Initialization, Task Assignment, Task Execution
- ❖ Shuffling and Sorting
- ❖ Partition & Combiner
- ❖ Hands on "WordCount" in MapReduce

Unit 5: Map/Reduce Programming - Java

- ❖ Job Dependency API discussion and Hands on
- ❖ Input Format API discussion and hands on
- ❖ Map Side Join and Reducer Side Join
- ❖ Optimizing MapReducer Jobs
- ❖ Monitoring Performance
- ❖ Data Compression
- ❖ Overview of RawComparators
- ❖ Bloom Filter

IT TRAINING & SERVICES

Regd. Office # 207, II floor, HUDA Maithrivanam, Ameerpet, Hyderabad 500 038.

Ph # +91 40 40310000, 23743392.

E-mail: enq@peerstech.com URL: www.peerstech.com

Unit 6: Hive

- ❖ Hive Services
 - Hive Shell
 - Hive Server
 - Hive Web Interface (HWI)
- ❖ Meta store
- ❖ Hive QL
- ❖ Working with Tables
- ❖ Working with Partitions
- ❖ User Defined Functions
- ❖ **Hands on exercises on Hive QL**

Unit 7: Pig

- ❖ Execution Types
- ❖ Grunt Shell
- ❖ Pig Latin
- ❖ Data Processing
- ❖ Loading and Storing
- ❖ Filtering
- ❖ Grouping & Joining
- ❖ Working with Functions
- ❖ User Defined Functions
- ❖ **Hands on exercises in Pig**

Unit 8: Sqoop

- ❖ Overview of Sqoop
- ❖ Tools in Sqoop
 - Import tool
 - Export tool
- ❖ **Hands on exercises on Sqoop**
 - Import Data from RDBMS
 - Export Data to RDBMS

Unit 9: HBase

- ❖ What is HBase?
- ❖ HBase Architecture
- ❖ HBase Features
- ❖ HBase Data Storage model
- ❖ HBase Families
- ❖ HDFS vs. HBase
- ❖ Terms and Daemons
 - HBase Regions
 - HBase Master
- ❖ HBase vs. RDBMS Overview
- ❖ How to access HBase Data?

- ❖ HBase API
- ❖ Runtime Modes
- ❖ HBase and Zookeeper
- ❖ **Hands on exercise on HBase**

Unit 10: Monitoring HDP2 Services

- ❖ What is Ambari?
- ❖ Ambari Web Interface

Introduction and overview of other Echo Systems:

- ❖ Flume
- ❖ Oozie
- ❖ Zoo keeper

Assistance on Installation of Single node cluster

IT TRAINING & SERVICES

Regd. Office # 207, II floor, HUDA Maithrivanam, Ameerpet, Hyderabad 500 038.

Ph # +91 40 40310000, 23743392.

E-mail: enq@peerstech.com URL: www.peerstech.com