BIGDATA-HADOOP SYLLABUS

Big Data Opportunities & Challenges

Introduction to 3V

BigData & Hadoop

OOPS & Java Fundamentals

Understanding Linux Commands

Linux commands required for Hadoop

Introduction to Hadoop

- Concept of Hadoop
- Distributed file system(HDFS)
- Design of HDFS
- Common challenges
- Best practices for scaling with your data
- Configuring HDFS
- Interacting with HDFS
- HDFS permission and Security
- Additional HDFS Tasks
- Data Flow – Anatomy of a File Read, Anatomy of a File Write and Coherency Model
- Hadoop Archives

Getting Started with Hadoop

Creating & Running your program

Pseudo Cluster Environment – Setting up Hadoop Cluster

- Cluster specification
- Hadoop Configuration (Environment Settings, Hadoop Daemon- Properties, Addresses and Ports)
- Basic Linux and HDFS Commands
- Setup a Hadoop Cluster

Map Reduce

- Hadoop Data Types
- Functional-Concept of Mappers
- Functional-Concept of Reducers
- The Execution Framework
- Concept of Partione
- Functional- Concept of Combiners
- Hadoop Cluster Architecture

- MapReduce types
- Input Formats (Input Splits and Records, Text Input, Binary Input, Multiple Inputs)
- OutPut Formats (TextOutput, BinaryOutPut, Multiple Output).
- Writing Programs for MapReduce

PIG

- Installing and Running Pig
- Grunt
- Pig's Data Model
- Pig Latin
- Developing & Testing Pig Latin Scripts
- Writing Evaluation
- Filter
- Loads & Store Functions

HIVE

- Hive Architecture
- Running Hive Comparison with Traditional Database (Schema on Read versus Write, Updates, Transactions and Indexes)
- HiveQL (Data Types, Operators and Functions)
- Tables (Managed and External Tables, Partitions and Buckets, Storage Formats, Importing Data)
- Altering Tables, Dropping Tables
- Querying Data (Sorting And Aggregating, Map Reduce Scripts, Joins & Subqueries & Views
- Map and Reduce site Join to optimize Query
- User Defined Functions
- Appending Data into existing Hive Table
- Custom Map/Reduce in Hive
- Perform Data Analytics using Pig and Hive

HBASE

- Introduction
- Client API- Basics
- Client API- Advanced Features
- Client API – Administrative Features
- Available Client
- Architecture
- MapReduce Integration
- Advanced Usage
- Advanced Indexing
- Implement HBASE

SQOOP

- Database Imports
- Working with Imported data
- Importing Large Objects
- Performing Exports
- Exports- A Deeper look

ZooKeeper

The Zookeeper Service (Data Modal, Operations, Implementation, Consistency, Sessions, States)

Oozie

- Workflow
- Coordinator
- Flume
- Concepts and Real time data streaming

Hadoop Admin

Roles and Responsibilities

Admin Activities

CCDH Certification Walkthrough

Certification Exam Preparation