

Hadoop Training Course

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, Coherency Model)

Hadoop Archives

Getting Started with Hadoop or running a sample program

Running a sample program

Pseudo Cluster Environment

Cluster specification

Hadoop Configuration (Environment Settings, Hadoop Daemon- Properties, Addresses and Ports)

Basic Linux and HDFS Commands

Setup a Hadoop Cluster

MapReduce

Hadoop Data Types

Functional-Concept of Mappers

Functional-Concept of Reducers

The Execution Framework

Concept of Partitioners

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

SQOOP

Database Imports

Working with Imported data

Importing Large Objects

Performing Exports

Exports- A Deeper look