

Hadoop In Action by UnniKrishna

Course Contents

Introduction to Hadoop and Architecture

Hadoop 1.0 Architecture

- ✓ Introduction to Hadoop & Big Data
- ✓ Hadoop Evolution
- ✓ Hadoop Architecture
- ✓ Networking Concepts
- ✓ Use cases - Where Hadoop fits into

Hadoop 2.0 Architecture

- ✓ Limitations on Hadoop 1.0 Architecture
 - ✓ Features of Hadoop 2.0 Architecture
 - ✓ HDFS Federation
 - ✓ High Availability of Name Node
 - ✓ YARN – Yet Another Resource Negotiator
 - ✓ Non MR applications on top of YARN
- ✚ Quiz on Architecture Concepts

Prerequisites for Hadoop Developer/Data Analysts/Admins

Linux

- ✓ Introduction to Linux
- ✓ Commands & Shell Scripts
- ✓ Vi& Vim editor features

- ✚ Case Study to develop a Shell Script

Java

- ✓ Introduction to OOPS & JAVA
 - ✓ Discussion on Object, Class & Methods
 - ✓ Features & Concepts of Core Java for developing MR jobs
 - ✓ Familiarizing Eclipse
- ✚ Case Study to develop a Java Code with the concepts learnt

Python

- ✓ Introduction to concepts of Python
 - ✓ How different is Python from other Programming Languages
 - ✓ Complex data Types in Python (Tuple, List, Dictionary)
 - ✓ Inbuilt Modules available in Python
 - ✓ File handling functions using Python
- ✚ Case Study to develop a Python Code with the concepts learnt

Cluster Installation

Hadoop Cluster Installation

- ✓ Types of Hadoop Cluster
 - ✓ Installing Pseudo Mode Cluster
 - ✓ Walk thru on inbuilt scripts, directories, configuration files and port numbers.
 - ✓ Discussion on Real Time Cluster Size
- ✚ Detailed documentation on Installation Procedure

Distributed File System - HDFS

HDFS Commands

- ✓ Introduction to HDFS Commands
 - ✓ Discussion on scenarios where specific commands are applicable
 - ✓ Introduction to Advanced HDFS Commands including fine tuning of cluster
- ✚ Detailed documentation on all the HDFS Commands

- ✚ Custom Script building using HDFS & Unix commands
- ✚ Quiz on HDFS Commands

Map Reduce - MR

Map Reduce using Java

- ✓ Introduction to Map Reduce Architecture
- ✓ Detailed discussion on different phases of MR
 - Mapper
 - Reducer
 - Splitting
 - Sorting
 - Shuffling
 - Combiner
 - Spilling
 - Partitioning
 - Merging
- ✓ Developing Map Reduce Application from Scratch using different use cases
- ✓ Discussion of difference between Old MR API & New MR API
- ✓ Introduction to different file formats and their internal features (Sequential, Binary etc.,)
- ✓ Developing MR code for Image Analytics
- ✚ Case Study on Map Reduce (Customer Sentiment Analyzer)

Map Reduce using Python – Streaming

- ✓ Developing Map Reduce Application using Python
- ✓ Discussion of different features available in Streaming
- ✚ Case Study on Map Reduce Streaming (Analytics on Temperature Datasets)
- ✚ Quiz on Map Reduce

Hadoop Eco System Components

Hive (Data Warehouse on top of HDFS)

- ✓ Introduction to Hive Architecture
- ✓ Configuring Hive Metadata store in different ways
- ✓ Basic Queries in Hive (DDL,DML)
- ✓ Advanced features of Hive
 - Partitioning
 - Bucketing
 - Sampling
 - Multi Table Load Queries
 - Serialize & De Serialize
- ✓ Dealing with different formats of data (Flat file, JSON, CSV etc.,)
- ✓ Query optimization using Hive.
- ✓ Developing User Defined Functions (UDF's) in Java & Python
- ✚ Case Study (Analytics on Telecom Datasets)
- ✚ Quiz on Hive

PIG (Data Flow Language)

- ✓ Introduction to Pig Latin
- ✓ Basic Commands in Pig
- ✓ Explanation advanced features of Pig with real time scenarios
- ✓ Different ways of using PigStorage
- ✓ Dealing with Unstructured data
- ✓ Developing Regular Expressions
- ✓ Developing User Defined Functions (UDF's) in Java & Python
- ✚ Case Study (Analytics on Books Datasets)
- ✚ Quiz on Pig

SQOOP (Import – Export utility)

- ✓ Introduction to Sqoop
- ✓ Basic Sqoop Commands
- ✓ Advanced Import Features

- ✓ **Advanced Export Features**
 - **Upsert calls**
 - **EVAL**
 - **Compressed formats**

- ✚ **Case Study (Analytics on Telecom Datasets)**
- ✚ **Quiz on Sqoop**

HBASE (Versioned Database)

- ✓ **Introduction to HBASE & NOSQL**
- ✓ **Basic difference in Row Oriented and Column Oriented storage**
- ✓ **Basic HBASE Commands**
- ✓ **Advanced HBASE Features**
 - **Versions**
 - **Compression Techniques**
 - **Bloom Filters**
 - **Sequential Scans**
- ✓ **Bulk Loads to HBASE Features**

- ✚ **Case Study on HBASE**
- ✚ **Quiz on HBASE**

Flume

- ✓ **Flume Architecture**
- ✓ **Configuring Flume Components**
 - **Source**
 - **Sink**
 - **Channel**
 - **Agents**
- ✓ **Building Flume Config files for different scenarios**
 - **Basic Config File building**
 - **Config file for connecting to different File Servers**
 - **Config file for connecting to Web Servers**

- ✚ **Quiz on Flume**

Scheduler (OOZIE & Autosys)

- ✓ **Introduction to Oozie**
- ✓ **Introduction to Autosys**
- ✓ **Using Schedulers for Batch Processing**

- ✚ **Quiz on OOZIE**

- ✚ **Finally this series of Practical Sessions ends with Quiz on entire course.**