BIG DATA WITH HADOOP

COURSE DESIGN

High-quality videos, slides, hands-on examples, quizzes, automated assessments, case studies, and real-world projects.

COURSE MATERIAL

Lifetime access to cutting-edge self-paced learning content.

LIVE SESSIONS RECORDINGS

45+ hours of long duration videos from the industry expert to help you with the course content.

SUPPORT

Email support to answer your queries and we've also launched Discussions - a Q&A site for Artificial Intelligence, Machine Learning, Deep Learning, Big Data & Data Science professionals.

CERTIFICATE

Earn a certificate in Big Data with Hadoop.

BIG DATA WITH HADOOP & SPARK - COURSE SYLLABUS

INTRODUCTION

- What is Big Data?
- Why Now?
- Big Data Use Cases
- Various Solutions
- Overview of Hadoop Ecosystem
- Spark Ecosystem Walkthrough
- Quiz

FOUNDATION &

ENVIRONMENT

- Understanding the CloudxLab
- CloudxLab Hands-On
- Hadoop & Spark Hands-on
- Quiz and Assessment
- Basics of Linux - Quick Hands-On
- Understanding Regular Expressions
- Quiz and Assessment
- Setting up VM (optional)

ZOOKEEPER

- Why Do we need it?
- Understanding Data Model
- Hands-On
- Quiz & Assessment
- How does election happen - Paxos Algorithm?
- Use cases
- When not to use
- Quiz & Assessment

HDFS

- Why HDFS or Why not existing file systems?
- Understanding the architecture
- Quiz
- Advance HDFS Concepts (HA, Federation)
- Quiz
- Hands-on with HDFS (Upload, Download, SetRep)
- Quiz & Assessment
- Data Locality (Rack Awareness)

DATA FORMATS & MANAGEMENT

- InputFormat and InputSplit
- JSON
- XML
- AVRO

- How to store many small files - SequenceFile?
- Parquet
- Protocol Buffers
- Comparing Compressions
- Understanding Row Oriented and Column Oriented Formats - RCFile?

YARN

- Computing - Why not existing tools?
- MapReduce 1.0
- Resource Management: YARN Architecture
- Advance Concepts - Speculative Execution
- Quiz

MapReduce BASICS

- Why MapReduce?
- Understanding MapReduce Framework
- Quiz
- Example 0 - Word Frequency Problem - Without MR
- Example 1 - Only Mapper - Image Resizing
- Example 2 - Word Frequency Problem
- Example 3 - Temperature Problem
- Example 4 - Multiple Reducer
- Example 5 - Java MapReduce Walkthrough
- Quiz

MapReduce ADVANCED

- Example 6 - Secondary Sorting (Word Recommendation)
- Example 7 - Partitioner
- Concept - Associative & Commutative
- Quiz
- Example 8 - Combiner
- Example 9 - Hadoop Streaming
- Example 10 - Adv. Problem Solving - Anagrams
- Example 11 - Adv. Problem Solving - Same DNA
- Example 12 - Adv. Problem Solving - Similar DNA
- Example 12 - Joins - Voting

- Limitations of MapReduce
- Quiz

ANALYZING DATA WITH PIG

- Why Pig?
- Basic Structure of Pig Latin
- Getting Started
- Example - NYSE Stock Exchange
- Concept - Lazy Evaluation

PROCESSING DATA WITH HIVE

- Why Hive?
- Hive Architecture Overview
- Getting Started
- Loading Data in Hive (Tables)
- Example: Movielens Data Processing
- Advance Concepts: Views
- Connecting Tableau and HiveServer 2
- Connecting Microsoft Excel and HiveServer 2
- Project: Sentiment Analyses of Twitter Data
- Advanced - Partition Tables
- Understanding HCatalog & Impala
- Quiz

NOSQL AND HBASE

- Case Study: The days before NoSQL
- What is NoSQL?
- CAP Theorem
- HBase Architecture - Region Servers etc
- Hbase Data Model - Column Family Orientedness
- Getting Started - Create table, Adding Data
- Adv Example - Google Links Storage
- Concept - Bloom Filter
- Comparison of NOSQL Databases
- Quiz

IMPORTING DATA WITH SQOOP AND FLUME, OOZIE

- Sqoop Overview
- Import From MySQL to HDFS, Hive, HBase
- Exporting to MySQL from HDFS
- Concept - Unbounding Dataset Processing or Stream Processing
- Flume Overview: Agents - Source, Sink, Channel
- Example 1 - Data from Local network service into HDFS
- Example 2 - Extracting Twitter Data
- Quiz
- Example 3 - Creating workflow with Oozie

OTHER Topics/Content

- Java Essential
- Linux Basics
- Spark On Cluster
- Adv Spark Programming
- Hands-on videos

PROJECTS INCLUDED

- Hive - Sentiment Analysis
- Processing the NSE (National Stock Exchange) data with Hive for various insights
- Doing Analytics on the MovieLens data to generate the movie ratings

The Big Data with Hadoop & Spark course is compatible with the following certifications:

- Cloudera Certified Professional (CCP): Data Engineer
- Cloudera Certified Associate (CCA): Hadoop Developer
- Hortonworks Certified Developer (HDPCD)

Please feel free to email your queries to
reachus@cloudxlab.com

Regards, The CloudxLab Team