

Big Data Syllabus

Week One

Day 1

- Introduction
- Linux Environment
- Setting Up Linux Environment
- Skills Evaluation

Day 2

- Development Principles
- Setting Up Scala Environment
- Scala Development Principles

Day 3

- Setting Up Python Environment
- Python Development Principles
- Study Cases For Big Data

Day 4

- Introduction to Big Data
- Hadoop Overview and History
- Overview of the Hadoop Ecosystem

Week Two

Day 1

- Introduction, and install Hadoop
- HDFS: What it is, and how it works
- Installing Dataset into HDFS

Day 2

- YARN explained
- MapReduce: What it is, and how it works
- MapReduce example

Day 3

- Integrating DBMS with Hadoop
- Setting up MySQL Server and creating & loading datasets into Mysql
- Sqoop Architecture
- Writing the Sqoop Import Commands to transfer data from RDBMS to HDFS/Hive

Day 4

- Flume explained
- Set up Flume and publish logs with it.
- Set up Flume to monitor a directory and store its data in HDFS

Day 5

- What is Hive?
- Use Hive to find the most popular game
- How Hive works
- Skills Evaluation

Week Three

Day 1

- ZooKeeper explained
- Kafka explained
- Setting up Kafka, and publishing some data.
- Publishing web logs with Kafka

Day 2

- Introduction to Apache NiFi
- Installing And Configuring NiFi
- Important Concepts: FlowFile, Processor and Connector

Day 3

- Why Spark?
- Immutable Collections in Spark
- Datasets and Spark 2
- Databricks

Day 4

- Consuming Data from Stream Platforms
- Coding Consumer with Scala
- Coding Consumer with Python

Day 5

- Data Pipeline Building with Nifi/Kafka/Spark
- Skills Evaluation

Week Four

Day 1

- Introducing Ambari
- Setting Up Ambari
- Managing Ambari

Day 2

- Introducing Cloudera Manager
- Setting Up Cloudera Manager
- Managing Cloudera Manager

Day 3

- Why NoSQL?
- What is HBase
- Use HBase to import data at scale.
- Cassandra DB

Day 4

- DWH Concepts
- Mesos explained
- Cassandra explained
- Hue Explained

Day 5

- AWS
- Environment Creation AWS
- S3
- Lambda
- Cloud Formation
- IAM
- VPC

Week Five

Day 1

- AWS EMR
- How to Get Data Into Amazon EMR
- Plan and Configure Clusters
- Manage Clusters

Day 2

- Oozie explained
- Set up a simple Oozie workflow

Day 3

- MICROSOFT HDInsight
- Understanding Microsoft big data solutions
- Big Data Architecture MSFT

Day 4

- Elasticsearch
- Solr And Lucene

Day 5

- Flink explained
- Apache Impala Explained

Week Six

Day 1

- Architecture Definition
- Data Selection

Day 2

- Architecture Set Up
- Cluster Building

Day 3

- Developing Class (Scala/Python/Java)

Day 4

- Cluster Evaluation

Day 5

- Summary
- Final presentation