

Course Durations: 30 Hours

Course Mode: Online/Offline

About Company:

EduNextgen extended arm of Product Innovation Academy is a growing entity in education and career transformation, specializing in today's most in-demand skills. A platform with blended learning programs supported by in-trend technology platforms for learning. Engaging organizations for learning development objectives.

Training courses are designed and updated by renowned industry experts. Our blended learning approach combines online classes, instructor-led live virtual classrooms and virtual teaching assistance.

About The Course:

With businesses generating Big Data at a rapid pace, analyzing the data to leverage meaningful business insights is the need of the hour. The demand for Analytics Skill is going up steadily but there is a huge deficit on the supply side. In spite of Big Data Analytics being a 'Hot' job, there is still a large number of unfilled jobs across the globe due to shortage of required skill. A McKinsey Global Institute study states that the US will face a shortage of about 190,000 Data Scientists and 1.5 million Managers and Analysts who can understand and make decisions using Big Data by 2018.

This is an extensive program designed to cover most important modules of Big Data required by today's Industry and as well help you achieve Certifications from Hortonworks and Cloudera. The program is bundled with modules like MapReduce, Hive, Pig, Hbase, Zookeeper, Oozie, Scoop, Impala and Flume. And to support your learning path there are resources like our Cloud Labs, Industry Grade Projects, Assignments, Use Cases and more with our world class Technical Support post learning.

Why This Course:

- Cover Hadoop Components i.e. MapReduce, Hive, Impala, Pig etc.
- Hands on Experience and Use Case
- Project Execution in Different Domain Data Sets and Components i.e. MapReduce, Pig, Hive etc.
- Pre-Installed Hadoop Environment (Plug and Play)
- Cloud Lab
- Live Support (24x7)

Participants will get the Access to:

- LMS Access
- Cloud Lab
- 50+ Assignments
- 100+ Quizzes
- Pre-Installed Hadoop Environment (Plug and Play)
- 5+ Industry Grade Projects in Different Domain
- Live Support via one to one Screen Sharing, Mail and Call
- Course Completion Certificate

Batch Schedule:

Weekend: 3 Hours per day (Online), 4 Hours per day (Offline)

Weekday: 2 Hours per day (Online), 2 Hours per day (Offline)

Course Curriculum

Module 1: Introduction to Big Data and Hadoop (3 Hours)

This module will help you to understand about the Big Data and Hadoop. Also, this module will cover the Big Data Characteristics, Types of Data, Big Data Application, Hadoop Ecosystem, Hadoop Installation and work with Cloud Lab. Below topics are covered in this module:

- What is Big Data?
- Big Data Growth
- Why Need of More Data?
- Big Data Impacts
- Big Data Characteristics (5V's)
- Types of Data
- Industries who are making the Most of Big Data
- Big Data Applications
- Traditional Model and its Problems
- Hadoop: Introduction and Why Hadoop
- Hadoop Characteristics
- Hadoop Ecosystem
- Hadoop Installation/VM/Cloud Lab

Hands on/Programs/Practical:

- Hadoop Installation with Single node Cluster Setup
- How to Start Hadoop
- JPS Command
- How to work with Cloud Lab

Module 2: YARN and HDFS Architecture (3 Hours)

This module will help you to understand YARN and HDFS Architecture. Also, this module will cover Cluster, Nodes, Racks and HDFS Hands on. Below topics are covered in this module:

- What is Cluster?
- Details about: Blocks, Namenodes, Datanodes, Secondary NameNode
- Introduction to HDFS
- HDFS Federation
- HDFS High Availability
- Racks and Replication Factor
- Rack Awareness

- HDFS File Write Anatomy
- HDFS File Read Anatomy
- HDFS Architecture
- Hadoop 1 vs Hadoop 2
- YARN
- How YARN Runs an Application

Hands on/Programs/Practical:

- HDFS Commands: ls, mkdir, cat, put etc.
- Blocks
- Replication

Module 3: Hadoop: MapReduce Framework (5 Hours)

This module will help you to understand about MapReduce, how to write MapReduce Program and its Framework. Also, this module will cover MapReduce Feature, Phases, Traditional vs MapReduce Solution, Joins Counters etc. Below topics are covered in this module:

- What is MapReduce?
- MapReduce – Features
- MapReduce Phases: Map, Shuffle, Reduce and Combiner
- How Hadoop runs a MapReduce job
- Input Splits
- Traditional vs MapReduce Solution
- Understanding MapReduce Paradigm
- Distributed Cache
- Joins Operations in MapReduce
- Counters
- Custom Input Format
- Secondary Sort
- Total Order Sort
- Testing MapReduce with MRUnit

Hands on/Programs/Practical:

- How to Write and Execute Word Count Program
- How to work with Partition and Combiner
- Join Operations
- Shorting
- Testing with MRUnits
- Project Execution in Automobiles Domain via MapReduce

Module 4: Data Transferring using Sqoop and Flume (2 Hours)

This module will help you to understand how to Transfer Data to Hadoop. This module will take deep drive on problem with Data Loading into Hadoop, Sqoop and its Feature, Flume, its Feature and how to work with Flume. Below topics are covered in this module:

- Issues with Data Load into Hadoop
- Introduction to Sqoop
- Features of Sqoop
- Sqoop: Installation Connectors
- Limitations of Sqoop
- Understanding of FLUME
- Data Flow in Flume
- FLUME Features
- FLUME: Installation

Hands on/Programs/Practical:

- Data Transfer into Hadoop using Sqoop
- Data Transfer into Hadoop using Flume

Module 5: Structure Data Analysis with Hive (4 Hours)

This module will help you to understand how to work with Structure Data. Also this module cover Hive, its limitation, Compare with other Database, Table, Index, Joins, Hive UDF. Below topics are covered in this module:

- Hive: Introduction
- Hive: Limitation, Architecture and Components
- RDBMS vs Hive
- Traditional Database and Hive
- Meta Store in Hive
- Hive Data Types
- Partitions and Buckets
- Tables in Hive
- Indexes and View
- Joins in Hive
- Sub Queries
- Embedding Custom Scripts
- Hive Built-in Function
- Hive UDF
- Hive ETL: Loading JSON, XML, Text Data

Hands on/Programs/Practical:

- Hive Commands: Create Database, Table, Insert Data into Table etc.
- Data Loading into Hive Table using Local and Hadoop
- Execute Hive Queries

- Join Operations
 - Project Execution in Automobiles Domain via Hive
-

Module 6: Impala vs Hive (1 Hours)

This module will help you to understand what difference between Impala and Hive. Also this module help you to understand Impala with Running Example. Below topics are covered in this module:

- Impala Introduction
- Impala Feature
- Impala Architecture
- Impala-Shell
- Difference between Impala and Hive

Hands on/Programs/Practical:

- Data Loading into Hadoop and Execute Impala Commands
-

Module 7: Working with Pig (3 Hours)

This module will help you to understand what PIG is and how we can work with PIG. Also this module help you to understand the difference between PIG and MapReduce, Hive. Below topics are covered in this module:

- Introduction to Pig
- Pig Background and Advantages
- MapReduce vs Pig
- Hive vs Pig
- Pig Components
- Pig Data Types
- Pig Operators
- Pig UDF

Hands on/Programs/Practical:

- How to Start Pig Shell
 - How to work with Pig in MapReduce Mode
 - Pig: Local Mode
 - Data Loading into Pig
 - Pigs Command and Dump
 - Join Operations in Pig
 - Project Execution in Entertainments Domain via Pig
-

Module 8: Introduction to Hbase and Zookeeper (3 Hours)

This module will help you to understand what NoSQL is and how to work with this. Also this module cover HBase, its Feature, Architecture, Zookeeper. Below topics are covered in this module:

- Introduction to NoSql
- Introduction to HBase
- Why Hbase?
- Industries who use HBase
- Hbase Unique Features
- Storage Mechanism in Hbase
- HBase: Architecture, Components and MemStore
- Data Flow in HBase
- HBase vs RDBMS
- HBase vs HDFS
- HBase vs Hive
- HBase Advantage and Limitations
- Hbase Shell
- Introduction to ZooKeeper
- How to work with ZooKeeper

Hands on/Programs/Practical:

- How to Start HBase
- HBase Command: Status, Version, Whoami etc.
- Tables Managements Commands: Create, List, Describe, Disable, Drop etc.
- Data Manipulation Commands: Count, Put, Get, Delete etc.
- How to Start Zookeeper and work with this

Module 9: Oozie Overview (2 Hours)

This module will help you to understand the Oozie Workflow. Also this module cover Oozie, its workflow and Feature. Below topics are covered in this module:

- Understanding of Oozie
- Oozie Workflow
- Why Oozie?
- Oozie Feature

Hands on/Programs/Practical:

- Oozie: Setup and Handson

Module 10: Advanced Project Execution (2 Hours)

Understand Project: What is the Project, Dataset, Use Case Execution and Output

Hands on/Programs/Practical: Step by step Project Execution