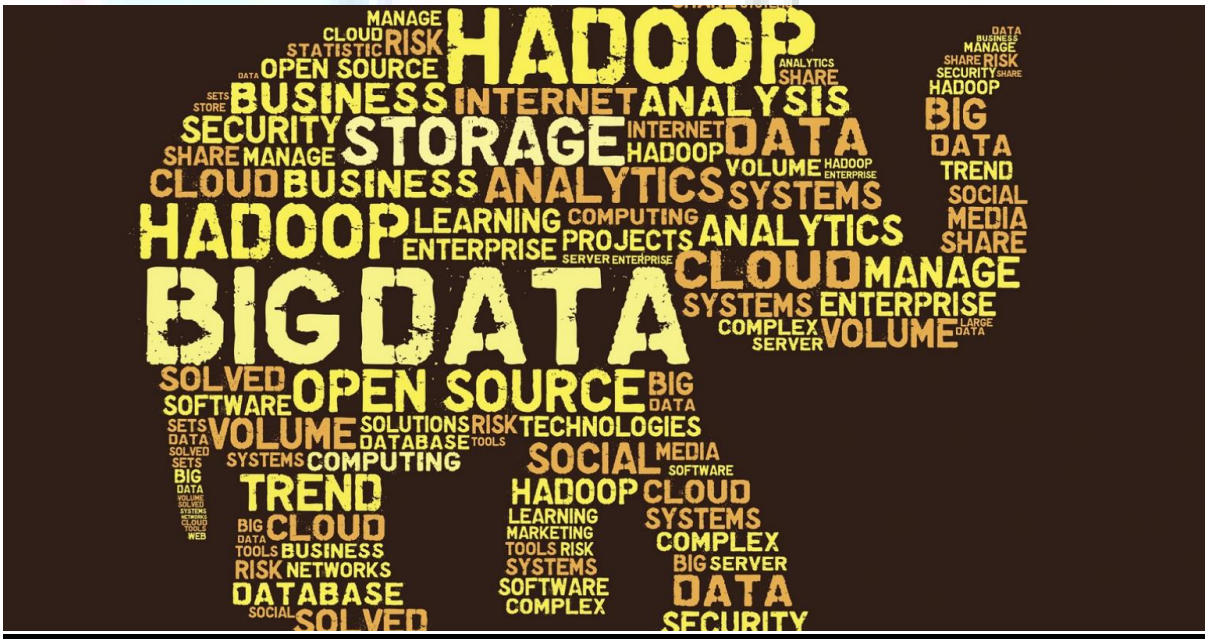
राष्ट्रीय इलेक्ट्रॉनिकी एवं सूचना प्रौद्योगिकी संस्थान, चेन्नई
National Institute of Electronics & Information Technology, Chennai

Ministry of Electronics & Information Technology
Government of India

नेशनल इंस्टीट्यूट ऑफ इलेक्ट्रॉनिक्स एंड इंफॉर्मेशन टेक्नोलॉजी, चेन्नई
National Institute of Electronics and Information Technology, Chennai
Autonomous Scientific Society of Ministry of Electronics & Information Technology (MeitY), Govt. of India
ISTE Complex, 25, Gandhi Mandapam Road, Chennai - 600025

Course Prospectus

Certificate course in Hadoop Administration



Index

Course Prospectus

Name of the Group	:	Data Science
Course Name	:	Certificate course in Hadoop Administration
Duration	:	6 weeks - 120 Hours (04 Hours per Day)
Course Coordinator	:	Sanjeev Kumar Jha, Mobile: 7765803105
Start Date	:	22 - 06 - 2020
Last Date of Registration	:	21 - 06 – 2020

Objective of the Course:

In present scenario Hadoop Administration is a highly valuable skill for anyone working at companies with Hadoop Clusters to store and process data. Almost every large company like Google, Amazon, Facebook etc. uses Hadoop in some way. It is clear that if the companies are using Hadoop for storing, analyzing and processing data then there will be a requirement for Hadoop Administrator.

The **Certificate course on Hadoop Administration** 120 Hours program offered by NIELIT Chennai provides with all the skills in order to successful work as a Hadoop Administrator and also provides expertise in all the steps necessary to manage a Hadoop cluster. This course on Hadoop Administration will make participants expert in working with Hadoop clusters and deploy that knowledge on real world projects.

Outcome of the Course: After undergoing this course the participants will learn:

- ✓ Planning and Deployment of a Hadoop Cluster
- ✓ Loading of Data and Run Applications
- ✓ Configuration of a Hadoop Cluster
- ✓ Performance Tuning of Hadoop Cluster
- ✓ Management and Maintenance of Hadoop Cluster
- ✓ Monitoring a Hadoop Cluster

- ✓ Troubleshooting a Hadoop Cluster

Target Audience:

- ✓ Software engineers and programmers who want to understand the administration of larger Hadoop ecosystem.
- ✓ Hadoop Developers and Java Developers who want to be a Hadoop Administrator and have interest in Big Data Field.
- ✓ Project, program, or product managers who want to understand how to manage, monitor and troubleshoot the Hadoop Cluster
- ✓ Linux / Unix Administrator, Data analysts and database administrators who are curious about Hadoop Administration part and how it relates to their work.
- ✓ Database Administrator, System architects who need to understand the components available in the Hadoop ecosystem, and how they manage and monitor them.

Requirements

- ✓ You will need access to a PC/Laptop with an Internet connection.
- ✓ You must have at least 8GB of free RAM on your system.
- ✓ A basic familiarity with the Linux command line will be very helpful.

Course Structure

S.no	Topic	Duration (in weeks) via Online mode
1	Module 1 : Introduction and Setup of HDFS	1 week
2	Module 2 : Hadoop Admin Commands and Mapreduce	1 week
3	Module 3 : Pig	1 week
4	Module 4 : Hive	1 week
5	Module 5 : Hbase	1 week
6	Module 6 : Apache Spark	1 week
Total		6 weeks

Other Details:

Course Fees: Rs. 2,000/- (Including GST) (Non-Refundable)

However the above Training fee shall be refunded on few special cases as given below:

1. If course postponed and new date is not convenient for the student.

2. If course cancelled

Payment Schedule

The Fee has to be paid in one instalment as given below.

Installment No	Last Date of Payment	Amount (in Rs.)
1	21 - 06 - 2020	Rs.2,000/-

Eligibility

- ✓ M.E./M.Tech/B.E./B.Tech/DOEACC B Level/B.Sc./Any Master Degree with Knowledge of Mathematics/Statistics and Computer Programming.
- ✓ Candidates who have appeared in the final semester examination and awaiting results may also apply. However, they have to submit proof of passing all semester examination/final degree at the time of completion of the course. Otherwise no certificate will be issued.

How to Apply

Candidates can apply online in our website http://14.139.173.196/reg. Payment be paid through any of the following modes:

- ✓ Online transaction: Account No: 31185720641 Branch: Kottur (Chennai), IFS Code: SBIN0001669.
- ✓ Pay through Nationalized Bank Debit Card (Service charges applicable)
- ✓ DD drawn from a nationalized bank (preferably SBI) in favour of "NIELIT Chennai" payable at Chennai.

Note: *The Institute will not be responsible for any mistakes done by either the bank concerned or by the depositor while remitting the amount into our account*

Last date of Registration: 21 - 06 - 2020

Selection Criteria of candidates

Selection is based on the first come basis (subject to fulfilling the eligibility criteria)

Admission Procedure

All interested candidates are required to fill the Registration form with the Course fees before 21 - 06 - 2020 with all the necessary following documents.

- Original and self-attested Copies of Proof of Age, Qualifying Degree (Consolidated Mark sheet & Degree Certificate/Course Completion Certificate), etc.
- One passport size photograph and one stamp size photograph for identity card.
- Self-attested copy of Govt. issued photo ID card

Note : Working days are from Monday to Friday.

Course Timings: 4 hours online daily (from Monday to Friday) during working hours

Mode of Training: Online

Discontinuing the course

✓ No fees under any circumstances, shall be refunded in the event of a student discontinuing the course. No certificate shall be issued for the classes attended.

Course Contents:

Introduction to Big Data

Topics: Big Data Introduction, Attributes of Big Data, Types of data, other technologies vs Big Data, Compare Hadoop vs traditional systems, Limitations and Solutions of existing Data.

Configuring Hadoop Environment

Topics: Configuring Virtual Box for Hadoop, Basic Linux Commands used in Hadoop, Vim Editor, Linux Configuration Files and Folders, Basic Concept of Shell Scripts, Adding User and Group in Linux, understanding sudo, Configuring Secure shell and connecting with other hosts using ssh.

Configuring Hadoop

Installing Java, Installing Hadoop, Configuring Java for Hadoop, Standalone mode, Pseudo Distributed Mode, Hadoop Path Setting, Fully Distributed, Monitoring the Cluster Health, Starting and Stopping the Nodes.

Hadoop: Eco System

Topics: Introduction to Hadoop Ecosystem, Hadoop Cluster, Pseudo Distributed mode, Type of clusters, Hadoop Ecosystem, Introduction of Pig, Hive, Oozie, HBase, Flume, SQOOP etc.

HDFS Architecture

Topics: Distributing Processing System, Core Components of Hadoop, HDFS Architecture, HDFS Design, HDFS role in Hadoop, Features of HDFS, Daemons of Hadoop and its functionality - Name node, Data node, Secondary Name Node**,** Job Tracker, Task Tracker, Anatomy of File Write, Anatomy of File Read, Network Topology, Heartbeat Signal, How to Store the Data into HDFS, How to Read the Data from HDFS, CLI commands, Format NameNode and Data Node, Changing Conf files, Checking the file system, Creating a Directory and putting data

Hadoop Admin commands

Topics: Hadoop Admin Commands and Browser interface

Hadoop Cluster Setup and Management

Topics: Introduction, Adding Machines to your VMBox, Setting Apache Hadoop Cluster Setup, writing data to cluster and checking replication status, Hadoop Configurations and Daemon Logs, Decommissioning and commissioning nodes in Apache Hadoop Cluster, Balancing a Cluster, Managing Software Packages with Apache Hadoop.

Hadoop MapReduce

Topics: HDFS Java API, Overview of MapReduce Framework, MapReduce Architecture, learn about Job tracker and Task tracker, use cases of MapReduce, Anatomy of MapReduce Program, Basic MapReduce API Concepts, Writing MapReduce Driver, Mappers, and Reducers in Java, Unit Testing MapReduce Programs etc, MapReduce Programming using Python.

Working with Pig and HIVE

Pig:

Topics: Installation, Architecture, Datatypes (scalar, complex), Running Pig (interactive, Batch), Pig Operators – Local, Store, Dump, Distinct, Filter, For Each, generate, Limit, Union, join, order by, Describe, Group by, Avg Default UDFs available (Built in function) REG EX, EXPLAIN, Parallel processing, Custom UDF

Hive:

Topics: Installation, Hive Services, Architecture, Comparing Hive to traditional Databases, Relational Data Analysis – (data types (primitive, complex) databases tables, create, alter, delete Hive Schema & Data storage Loading data into Hive views Storing query results (store), Text processing - Built in functions, string functions, regular expressions, Managed vs External Tables.

HBase

Topics: HBase Introduction, HBase architecture, HBase vs RDBMS (fixed Vs flexible schema), Master and Region servers, HBase commands

Apache Spark

Topics: Introduction of Apache Spark, Features, Data Source, Data Sets, Data Frames, Spark Configuration and Cluster Modes, Spark Transformers, Spark ML Lib, Spark ML Pipeline, Spark Algorithms for Machine Learning, Integration with Hadoop etc

Location and how to reach: NIELIT Chennai is located at Gandhi Mandapam Road, Kotturpuram, Chennai (Landmark: Opp. To Anna Centenary Library).

Address:

National of Electronics and Information Technology

ISTE Complex, No. 25, Gandhi Mandapam Road, Chennai – 600025

E-mail: trng.chennai@nielit.gov.in/Phone: 044-24421445

Contact Person: Sanjeev Kumar Jha, Mobile: 7765803105

Course enquiries

Students can enquire about the various courses either on telephone or by personal contact between 9.15 A.M. to 5.15 P.M. (Lunch time 1.00 pm to 1.30 pm) Monday to Friday.