NIELIT Virtual Academy

National Institute of Electronics and Information Technology, Chennai

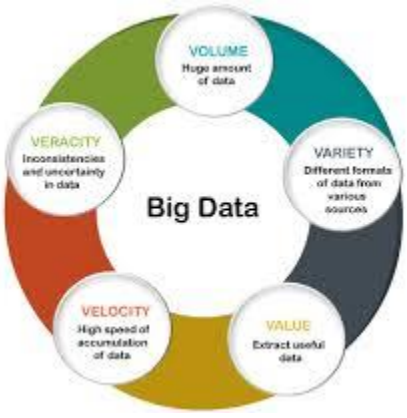
Autonomous Scientific Society of Ministry of Electronics & Information Technology (MeitY), Govt. of India, ISTE Complex, 25, Gandhi Mandapam Road, Chennai – 600025

Course Prospectus

NSQF Aligned

Mode: ONLINE (Blended)

Essentials of BIG DATA



Index

Course Prospectus

Course Name : Essentials of Big Data

NSQF Level: 05

Duration: 90 Hours

Last Date of Registration: 10-04-2025

Date of publishing Provisional Selection List: 11-04-2025

Course Start Date: 14-04-2025

Preamble:

The "Essentials of Big Data" upskilling course offers a comprehensive exploration of fundamental concepts and technologies in the realm of big data. Participants delve into key components such as Hadoop, Spark, and distributed computing, gaining a profound understanding of how to manage and analyze vast datasets efficiently. Through hands-on exercises and real-world case studies, learners develop the skills necessary to harness the power of big data technologies, enabling them to extract valuable insights and make informed decisions. This course is ideal for professionals seeking to enhance their expertise in the dynamic field of big data, equipping them with the knowledge to navigate large-scale data challenges and contribute to innovative solutions in various industries. Throughout the program, participants engage with practical scenarios, implementing big data technologies to solve complex problems.

Objective of the Course:

A course on the essentials of Big Data typically covers the foundational concepts, tools, and techniques used in the field. Here are some common course objectives:

1. Introduction to Big Data Concepts

- Understand the definition of Big Data and its characteristics (volume, velocity, variety, veracity, and value).
- Learn about the history and evolution of Big Data technologies.
- Recognize the impact of Big Data on various industries.

2. Big Data Architecture and Ecosystem

- Explore the architecture of Big Data systems, including data storage, processing, and analytics layers.
- Understand the key components of the Big Data ecosystem, such as Hadoop, Spark, and NoSQL databases.
- Learn about distributed computing, storage, and parallel processing models.

3. Data Collection and Management

- Learn how to collect data from various sources (e.g., social media, IoT devices, transactional data).
- Understand data cleaning, transformation, and preprocessing techniques.
- Explore techniques for managing unstructured, semi-structured, and structured data.

4. Big Data Tools and Technologies

- Gain hands-on experience with popular Big Data tools like Hadoop, Apache Spark, Hive, Pig, and HBase.
- Learn how to work with distributed data storage systems like HDFS and cloud storage options.
- Explore data processing techniques like batch processing and real-time streaming.

5. Data Analytics and Machine Learning

- Understand the role of analytics in Big Data and learn how to apply basic statistical techniques.
- Learn how to use machine learning algorithms for predictive analytics and data classification.
- Explore the integration of Big Data with AI and deep learning models.

6. Big Data Security and Privacy

- Understand the challenges related to data security, privacy, and ethical considerations in Big Data.
- Learn about encryption, data anonymization, and other privacy-preserving techniques.
- Explore legal and regulatory frameworks for handling Big Data.

Outcome of the Course:

The **outcomes** of a course on the **Essentials of Big Data** are designed to ensure students gain both theoretical knowledge and practical skills in handling and analyzing large datasets.

Course Structure

S.No	Module Name	Th.	Pr.	Total
1.	Big Data Hadoop Ecosystem	10	10	20
2.	Modeling Big Data	10	10	20
3.	Processing Big Data	10	15	25
4.	Streaming and Advanced Topics	10	15	25
	Duration (in Hours)	**40**	**50**	**90**

Course Fee Structure:(Including GST)

	General	SC/ST	Last Date of Registration	
Course Fee	Rs.4680/-	NIL	10-04-2025	

Eligibility

S. No.	Academic/Skill Qualification (with Specialization - if applicable)	Required Experience (with Specialization - if applicable)
1	➢ Pursuing Final Year B.Tech in any branch of Engineering , Or ➢ Pursuing Final Year MCA ,Or ➢ Pursuing Final Year B.Sc. in any branch of Sciences , Or ➢ Pursuing Final Year B.Sc. in IT/CS/Electronics/allied subjects Note : 1)Students should have relevant knowledge of the Data Analysis and programming concepts. 2)Students with the above entry requirements are eligible to take the course subject to clearing the written test comprising of Analytical Reasoning, Mathematics and English 3)Passout students in the above entry requirements are also eligible for the course.	-NA-

Number of Seats: 80(Eighty) – Total

Note: Seats are allocated based on the merit of the Qualification

How to Apply?

Candidates can apply online in our website http://nva.nielit.gov.in. Payment towards non-refundable registration fee can be paid through any of the following modes:

- ✓ **Payment Gateway**
- ✓ **Online transaction: Account No: 31185720641 Branch: Kottur (Chennai), IFSCode: SBIN0001669.**
- ✓ **GPAY/any UPI, Credit Card**

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: *10-04-2025*

Selection Criteria :

Selection of candidates will be based on their marks in the qualifying examination subject to eligibility and availability of seats.

- ✓ The first list of Provisionally Selected Candidates will be published on NIELIT Chennai website (www.nielit.gov.in/chennai/index.php) on **11-04-2025** by **5:00 PM**. In case of vacancy, an additional selection list will be prepared and the selection will be intimated by email only.
- ✓ Following documents of candidates will be verified:
  - Qualifying Degree (Consolidated Marksheet/Degree Certificate/Course Completion Certificate), 10^{th} and 12^{th} mark sheet.
  - One passport size photograph.
  - Self-attested copy of Govt. issued photo ID card.
  - AADHAR Copy
- ✓ Selected candidates are requested to upload the proof of remittance of fee on registration portal and also send the proof of remittance of fee as email to **trng.chennai@nielit.gov.in/dhirendra@nielit.gov.in/skjha@nielit.gov.in**

Admission:

All provisionally selected who have paid the fees (full or first instalment) andverified by accounts section of NIELIT Chennai will get a welcome message in his login id provided during registration.

Note-All Provisionally Selected Candidates have to visit NIELIT Chennai for Certificate Verification.

Otherwise their candidature will be cancelled without any intimation.

The credentials and URL for online portal will be shared through WhatsApp or email.

Discontinuing the course

- ✓ No fees under any circumstances, shall be refunded in the event of a student who have completed the process of admission or discontinuing the course in between. No certificate shall be issued for the classes attended.
- ✓ If candidates are not uploading consecutive 3 assignments within assigned time, then their candidature will be cancelled without any notice and all fees paid will be forfeited.
- ✓ If candidates are not appearing for any internal examinations/practical their candidature will be cancelled without any notice and all fees paid will be forfeited.

Course Timings:

This program is a practical oriented one and hence there shall be more lab than theory classes. The classes and labs are online cloud-based from **10 am to 01 pm** and Monday to Friday. In between **any 04 hours** can be fixed as your class timings according to the candidate's convenience and the faculty's availability and remaining student can do their lab.

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.

Placement:

Students who have completed the course successfully and qualified, Placement guidance and career counselling will be given to assist in their interviews.

Important Dates

Last Date of Registration: 10-04-2025
Display of Provisional Selection List: 11-04-2025

Course Start Date: 14-04-2025

Examination & Certification

- ✓ Final Certificates will be issued after successful completion of all the modules

including mini project. For getting certificate a candidate has to pass each module individually with minimum required marks of 50%.

NSQF Examination Pattern:

➢ Means of assessment:

S. No	Examination Pattern	Modules Covered	Duration in Minutes	Maximum Marks
1.	Theory 1: Essentials of Big Data	Module 1	90	100
4	Practical 1: Essentials of Big Data	Module1	120	60
6	Internal Assessment		-	20
7	Project/Presentation Assignment		-	20
Total				**200**

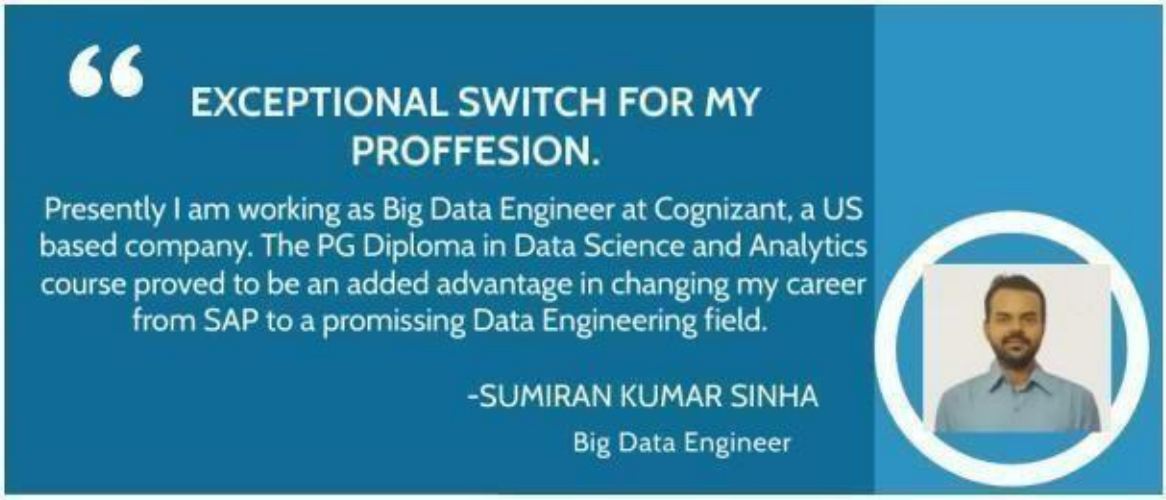
Examination Centre: NIELIT Chennai, Mode: Online

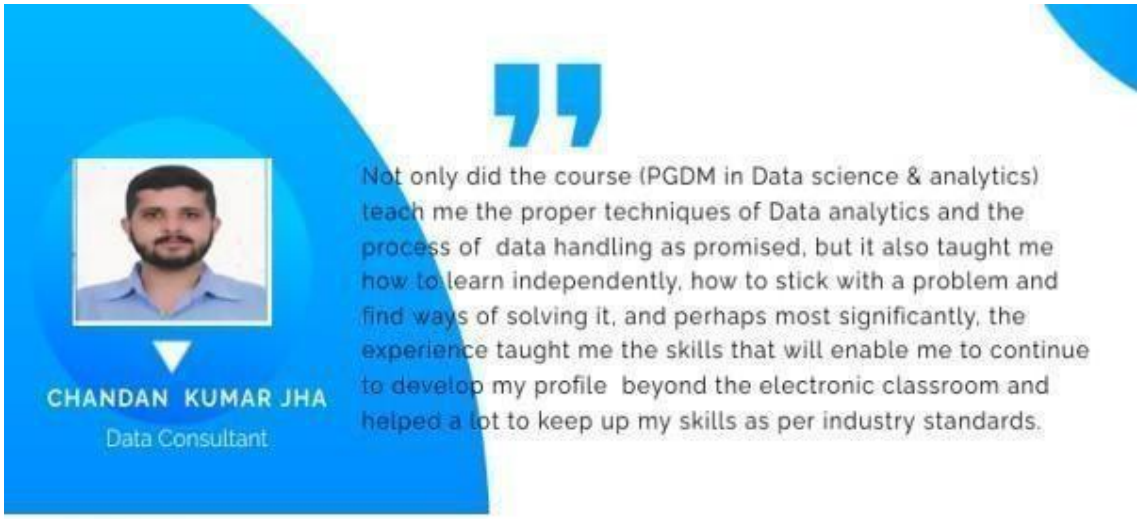
Grading Scheme

Following Grading scheme (on the basis of total marks) will be followed:

Grade	S	A	B	C	D
Marks Range (in %)	>=85%	>=75% and <85%	>=65% and <75%	>=55% and <65%	>=50% and <55%

Testimonials:




Not only did the course (PGDM in Data science & analytics) teach me the proper techniques of Data analytics and the process of data handling as promised, but it also taught me how to learn independently, how to stick with a problem and find ways of solving it, and perhaps most significantly, the experience taught me the skills that will enable me to continue to develop my profile beyond the electronic classroom and helped a lot to keep up my skills as per industry standards.
CHANDAN KUMAR JHA
Data Consultant

.....many more.

Detailed curriculum

Essentials of Big Data

- Introduction to Big Data
- Five V's of Big Data
- Attributes of Big Data
- Processing Big Data

Hadoop

- Introduction
- Hadoop Installation
- Components of Hadoop
  - HDFS (Hadoop Distributed File System)
    - Features
    - Architecture
      - NameNode (Master)
      - DataNodes (Slaves)
      - Secondary NameNode
  - Data Flow
    - File Write Workflow
    - File Read Workflow
  - Advantages & Limitations of HDFS
  - HDFS Commands
    - Basic Commands
    - File Operations
    - File Information
    - Administrative Commands
  - HDFS File Structure
    - File Splitting
    - Replication
  - YARN (Yet Another Resource Negotiator)
    - Features
    - YARN Architecture
      - ResourceManager
      - NodeManager
      - ApplicationMaster
      - Containers
    - YARN Workflow
    - YARN and its Processing Application
    - YARN MR Application Execution Flow
    - Resource Request
    - Resource Allocation
    - YARN Scheduling Models
      - Capacity Scheduler
      - Fair Scheduler
      - FIFO Scheduler
    - YARN Use Cases
      - Big Data Processing
      - Machine Learning
      - Real-Time Processing

- Map Reducer
  - Concepts of MapReduce
    - Map Phase
    - Shuffle and Sort
    - Reduce Phase
  - MapReduce Architecture
    - JobTracker (Master)
    - TaskTrackers (Workers)
  - MapReduce Workflow
  - Components of MapReduce
    - Mapper
    - Reducer
    - Combiner
    - Partitioner

H Base

- Introduction
- Features of HBase
- HBase Architecture
  - HBase Cluster
  - HMaster
  - RegionServer
  - ZooKeeper
- HBase Shell Basics
  - Creating a Table
  - Listing Tables
  - Adding Data to a Table
  - Retrieving Data
  - Deleting Data
  - Disabling and Dropping a Table
- HBase vs Traditional RDBMS
- Understanding HBase Configuration files and its configuration in Hadoop Eco System
- Understanding Conceptual View, Physical View, Namespace, Table, Row, Column
- Family, Cells, Data Model Operations, Versions, Sort Order, Column Metadata, Joins,
- ACID etc.
- HBase Constraints

Spark

- Introduction of Apache Spark: Need of Apache Spark, Feature of Apache Spark.
- Understanding concept of Spark Cluster Modes on YARN.
- Apache Spark Installation and Configuration
- Understanding Spark Cluster Modes on YARN
- Spark Applications
- The back bone of Spark - RDD (Resilient Distributed Dataset)
- Loading Data

- What is Lambda
- Using the Spark shell
- Actions and Transformations
- Associative Property
- Implant on Data
- Persistence
- Caching
- Loading and Saving data
- Operations of RDD
    - Challenges in Existing Computing Methods
    - Probable Solution & How RDD Solves the Problem
    - Introduction to RDD, its operations, Transformations & Actions
    - Data Loading and Saving Through RDDs
    - Key-Value Pair RDDs, Other Pair RDDs, Two Pair RDDs
    - RDD Lineage & RDD Persistence
    - Word Count Program Using RDD Concepts
    - RDD Partitioning achieve Parallelization.
    - Using Accumulators
    - Creating custom Accumulators
    - Using Broadcast variables
    - Passing Functions to Spark
- Data Frames and Spark SQL
    - Introduction to Spark SQL & its architecture
    - SQL Context in Spark SQL\

Kafka

- Introduction
- Features of Kafka
- Kafka vs Traditional Messaging Systems
- Concepts related to Kafka
    - Topic
    - Producer
    - Consumer
    - Broker
    - Partition
    - Offset
    - Zookeeper
        - Ensemble
        - Leader
        - Followers
        - Clients
    - ZooKeeper Data Model
        - znodes (ZooKeeper Nodes
        - Types of znodes
    - ZooKeeper Operations
        - Read Operations
        - Write Operations
        - Watchers

- APIs in Kafka
  - Producer API
  - Consumer API
  - Streams AP
  - Connector AP
- Various Topics related to Kafka
  - High Throughput
  - Scalability
  - Fault Tolerance
  - Decoupling
  - Real-Time Processing