Learnomate Technologies is the Information technology company which provide training o**n different IT Technologies.**

Hadoop is an open-source software programming framework for storing a large amount of data and performing the computation.

Big data is a collection of large datasets that cannot be processed using traditional computing techniques. It is not a single technique or a tool, rather it has become a complete subject, which involves various tools, technqiues and frameworks.

It also Known as Hadoop big data or Apache Spark or Data Engineer into Industry.

DATA ENGINEERING SYLLABUS KEY POINTS

Module 1: Introduction to Data and Opportunities
- What is data? (Structured, Semi-structured, Unstructured)
- The Data Lifecycle (Capture, Store, Process, Analyze, Visualize)
- Big Data and its characteristics (Volume, Variety, Velocity)
- Career paths in Data Engineering
- Real-world use cases of Data Engineering

Module 2: Python for Data Engineering

Introduction to Python Programing
- Variables, Data Types, Operators
- Control Flow (if/else, loops)
- Functions

Data Structures in Python
- Lists, Tuples, Dictionaries, Sets

Libraries for Data Manipulation and Analysis
- NumPy (Numerical Computing)
- Pandas (Data Analysis)

Module 3: Databases
- **Introduction to Database Systems**
  - Relational Databases vs. NoSQL Databases

- **SQL Fundamentals (Structured Query Language)**
  - SELECT, INSERT, UPDATE, DELETE statements
  - JOIN operations (INNER JOIN, LEFT JOIN, etc.)
  - WHERE clauses and filtering data

Module 4: : MySQL

- Introduction to MySQL (a popular relational database)
- Creating and Managing Databases
- Working with Tables, Columns, and Data Types
- Writing SQL queries to retrieve, manipulate, and analyze data
- Hands-on Labs with MySQL workbench

Module 5: MongoDB

- Introduction to MongoDB (a popular NoSQL document database)
- JSON data format and working with documents
- CRUD operations (Create, Read, Update, Delete) in MongoDB
- Querying data using MongoDB Query Language
- Hands-on Labs with MongoDB Compass

Module 6: Big Data Technologies

- Introduction to Big Data Processing
- The need for distributed computing frameworks
- Apache Hadoop Ecosystem (HDFS, YARN, MapReduce) (High-Level overview)
- Apache Spark for large-scale data processing (Spark basics)

Module 7: Introduction to Cloud Platforms

- Benefits of using Cloud Platforms for Data Engineering
- Introduction to Microsoft Azure and Amazon Web Services (AWS)

Module 8: Azure Data Services

- Azure Data Factory (ADF) for ETL/ELT orchestration
- Creating and scheduling data pipelines with ADF
- Azure Synapse Analytics for data warehousing and big data analytics
- Azure Blob Storage for scalable data storage
- Azure Databricks for distributed data processing with Apache Spark
- Azure SQL Database: Managed relational database service

Module 9: AWS Data Services

- Introduction to AWS Services for Data Engineering
- Amazon S3 for object storage
- Amazon Redshift for data warehousing
- AWS Glue for ETL/ELT jobs
- Amazon EMR for distributed processing with Hadoop and Spark (High-Level overview)

Module 10: Introduction to Additional Technologies

- **Apache Kafka:** A distributed streaming platform for real-time data ingestion. (High-Level overview)

- **Apache Airflow:** A workflow orchestration tool for scheduling and managing data pipelines. (High-Level overview)

- **Snowflake:** A cloud-based data warehouse solution. (High-Level overview)

- **Informatica:** A commercial data integration platform for ETL/ELT processes. (High-Level overview)

- **Hive:** A data warehouse software framework for reading, writing, and managing large datasets stored in distributed storage systems like Hadoop.

Module 10: Data Visualization with Power BI

- Introduction to Power BI for data visualization
- Connecting Power BI to data sources (Azure Synapse, etc.)
- Creating reports and dashboards with interactive visuals
- Sharing insights with stakeholders

Module 11: Machine Learning Fundamentals

- Introduction to Machine Learning concepts
- Supervised vs. Unsupervised Learning
- Common Machine Learning algorithms (optional)
- Exploring Machine Learning libraries in Python (optional)

PROJECTS COVERD

- **ETL Data Pipeline on AWS EMR Cluster**
- **Modern ETL Data Pipeline using Informatica cloud**
- **Data Pipeline based on Messaging Using PySpark and Airflow**
- **Hive Project to build a data warehouse for e-Commerce**
- **Finance Complaint**
- **Aws Glue Data Pipeline**

TRAINING HIGHLIGHTS

- **Recording Access shared to students on Learnomate App**
- **Professional Resume building by Industrial working mentors**
- **Dedicated Support Team to solve issues [8 Am to 8 Pm]**
- **Placement assistance/Job requirement notification support/HR contacts**
- **Training Certificate: Receive a recognized certificate upon course completion**
- **LinkedIn, Naukri.com Profile: Enhance your online presence with professionally curated profiles.**
- **Flexible Learning Options: Choose between offline and online training to suit your schedule.**
- **Interview Preparation, Mock Interviews: Nail your interviews with our tailored preparation and mock interview sessions**
- **Real-time Scenarios Explained: Learn through practical examples to master real-world applications.**
- **❓ Doubt Sessions: Clarify your doubts through dedicated doubt-clearing sessions.**

LEARNOMATE TECHNOLOGIES

CONTACT DETAILS

If you required any further information, please fill free to contact us.

Learnomate Technologies Pvt. Ltd
(Sai Luxuria, Office No 15, 3rd Floor,Bhumkar Chowk,
Wakad, Pune, Maharashtra, 411057 India)

Learnomate HR Team Contact Details:

Call/WhatsApp: +91 7757062955
+91 7822917585
Email: info@learnomate.org

THANK YOU