

Big Data Engineering Professional

With Cloud Computing

Course Details

About the Course:

In this course, the participants will learn everything from data infrastructure setup, maintenance, data modeling to writing ETL (Extract Transform Load), ELT (Extract Load Transform) jobs, building data pipelines and Scheduling, monitoring all these critical jobs both on prem & cloud.

About the Trainers:

A Team of Trainers with 30+ years of overall combined industry experience. Currently working on big data engineering, data science and related projects.

What is the prerequisite?

Basic computer knowledge, passion to work with data systems

Duration:

180 Hours / 24 Weeks
-20 Weeks Training
4 Weeks Project

Options to Complete the Course in different tracks.

<u>Track</u>	<u>Duration</u>	<u>Class Time</u>	<u>Days</u>
Regular	24 Weeks	90 minutes	Mon-Fri
Weekend	24 Weeks	4 hours	Sat-Sun

Content@glance

<u>SKILL</u>	<u>DETAILS</u>
Linux	Introduction, Architecture, Installation, basic commands, admin command and utilities
PostgreSQL & Mysql Databases	Introduction, Architecture, Installation, database, users, schema creation, basic administration SQL: SQL to work with Relational Databases (Data Warehouse). Data Warehouse: Introduction, Building data warehouse database, designing star schema, star schema vs snowflake schema. Data Modeling a Key skills for data engineering
Python	Introduction, programming fundamentals, Variables, data types, operators, data structures, conditional statement (if), iteration/loop (while,for), functions, file management, exceptions handling. Numpy, pandas, matplotlib libraries, Object Oriented Programming (OOPs), importing data, web scraping, data wrangling, text manipulation using regex module. Building ETL (Extract Transform & Load) data pipelines.

Data Structures & Algorithms	For writing better & optimized code.
Cloud Computing for Data Engineering Services	AWS - Fundamentals & Data engineering services. Data Ingestion - AWS Kinesis Firehose Snowball Data Integration - AWS Glue Data Storage - AWS S3
Data Warehouse Vs Data Lake Vs Data Lakehouse on Cloud	Amazon Redshift, Microsoft Azure Synapse Analytics (ASA), Google Big Query
Big data - Spark (Pyspark)	Introduction, Architecture, Spark Installation, Spark Configuration, Spark Context, Using Spark Shell, working with RDDs (Resilient Distributed Datasets). Building ETL Pipelines: the bread & butter of data engineering
Spark (Sparksq)	Why Sparksq? Spark SQL overview, features, Spark SQL libraries, Querying using Spark SQL, Adding Schema to RDDs, Caching
Apache Airflow	Introduction, Authoring, Scheduling data pipelines jobs. (Automating ETL workflow)
Docker & Kubernetes	Containers with Orchestration service
Data Architectures	Introduction to Data Fabric & Data Mesh
Deployment	From Model Testing to Deployment with MLOps
Version Control System (VCS)	Git & Github
Certification	Preparation, Guidance, Mock Test
Job Readiness Program	Communication@workplace (Oral & Written), Presentation skills, Resume Building, Interview facing tips, Mock Interviews
Capstone Project	

Two days Workshops [Sat & Sun]

Big Data - Hadoop [Hive, Impala]

NOSQL Databases
MongoDB, Cassandra

Data Warehouse Vs Data Lake Vs Data Lakehouse on Cloud
Amazon Redshift, Azure Synapse Analytics, Google Big query

Building Real time Data Pipelines with Kafka Connect and Spark Streaming

Internship Opportunity for every Student!
Challenging projects to gain real time experience.