

Data Engineering

Python for Data Engineering

- Python Fundamentals
- File Handling & Automation
- Working with APIs
- Data Processing with Pandas
- Logging & Error Handling

SQL for Data Engineering

- Database Fundamentals
- Advanced SQL Queries
- Joins & Subqueries
- Window Functions
- Stored Procedures
- Query Optimization

Data Processing & ETL

- ETL vs ELT Concepts
- Data Extraction Techniques
- Data Transformation Pipelines
- Data Loading Strategies
- Workflow Automation

Big Data Fundamentals

- Introduction to Big Data
- Hadoop Ecosystem

- Distributed Computing Concepts
- Batch vs Real-time Processing

Apache Spark

- Spark Architecture
- PySpark Basics
- DataFrames & RDDs
- Spark SQL
- Spark Streaming
- Performance Optimization

Data Pipelines & Orchestration

- Apache Airflow
- Workflow Scheduling
- DAG Creation
- Pipeline Monitoring
- Automation & Alerts

Data Warehousing

- Data Warehouse Concepts
- Star & Snowflake Schema
- OLTP vs OLAP
- Data Modeling Techniques
- Data Lake vs Data Warehouse

Microsoft Azure Data Engineering

- Introduction to Microsoft Azure
- Azure Data Factory (ADF)
- Azure Data Lake Storage
- Azure Synapse Analytics
- Azure Databricks
- Azure SQL Database
- Data Pipelines in Azure

AWS Data Engineering

- Introduction to Amazon Web Services
- Amazon S3
- AWS Glue
- Amazon Redshift
- AWS Lambda
- Amazon EMR
- Kinesis for Streaming Data
- Data Pipelines on AWS

Google Cloud Platform (GCP)

- Introduction to Google Cloud
- BigQuery
- Cloud Storage
- Dataflow
- Dataproc
- Pub/Sub Messaging
- Cloud Composer
- Data Pipelines on GCP

Streaming & Real-time Data

- Apache Kafka Basics
- Stream Processing Concepts
- Real-time Data Pipelines
- Event-driven Architecture

Data Governance & Security

- Data Security Fundamentals
- Access Control & IAM
- Data Encryption
- Data Quality & Validation
- Compliance Basics

DevOps & Deployment

- Docker Basics
- CI/CD for Data Pipelines
- Git & GitHub
- Infrastructure Basics

Monitoring & Optimization

- Pipeline Monitoring
- Logging & Debugging
- Performance Tuning
- Cost Optimization in Cloud

Real-world Projects

- End-to-End ETL Pipeline
- Cloud-based Data Warehouse
- Real-time Streaming Pipeline
- Data Lake Implementation
- Multi-cloud Data Engineering Project

Career Preparation

- Resume Building
- Portfolio Projects
- Interview Preparation
- Real-world Case Studies

Tools Covered

- Python
- Jupyter Notebook
- Google Colab
- Pandas
- MySQL
- Apache Hadoop
- Apache Spark
- PySpark
- Apache Airflow
- Apache Kafka
- Docker
- Git
- GitHub
- Microsoft Azure
- Amazon Web Services
- Google Cloud
- Azure Data Factory
- Azure Databricks
- Amazon Redshift
- BigQuery