

## **Data Engineering Course Syllabus**

### **Section 1: Introduction to Data Engineering**

- Overview of Data Engineering
- Data Science vs Data Engineering
- Data Engineering Infrastructure and Data Pipeline
- Data Architecture

### **Section 2: Data Storage and Management**

- Relational Databases
- SQL fundamentals
- Database design and normalization
- Query optimization
- NoSQL Databases
- MongoDB, Cassandra, Redis basics
- Use cases for NoSQL
- Cloud Storage Systems
- Distributed file systems (HDFS)

### **Section 3: Cloud Sources**

- Amazon Relational Database Service
- Microsoft Azure SQL Database
- Google Cloud SQL
- Concepts of Extra-Load, Extract-Load-Transform, or Extract-Transform-Load paradigms

### **Section 4: Data Integration and ETL Processes**

- ETL (Extract, Transform, Load) concepts
- Data pipelines
- Apache NiFi
- Informatica
- Talend
- Data ingestion techniques
- Data cleansing and transformation

## **Section 5: Big Data Processing**

- Big Data concepts
- Apache Hadoop and HDFS
- Map Reduce Programming
- Apache Spark
- Data Frames
- Spark SQL

## **Section 6: Data Streaming**

- Real-time data processing
- Apache Kafka
- Stream processing
- Apache Flink
- Apache Storm

## **Section 7: Data Warehousing**

- Data Warehousing concepts
- Amazon Redshift
- Google BigQuery
- Star and Snowflake Schema Design
- OLAP vs. OLTP

## **Section 8: Data Governance and Security**

- Data quality management
- Data privacy regulations
- Role-based access control
- Data encryption techniques

### **Section 9: Machine Learning for Data Engineers**

- ML workflows overview
- Feature engineering & data prep for ML
- Integration with TensorFlow & Scikit-learn

### **Section 10: Emerging Trends in Data Engineering**

- Data Mesh principles
- Serverless data pipelines
- Real-time analytics with Apache Pinot
- Ethical data engineering & sustainability

