

## Data Engineering Course Program





**Instructor:** Jitendra Mohanty (Cloud & DevOps Architect – Boeing, Bengaluru)

**Course Mode:** Live Classes + In-Office Recordings Access

**Placement Support:** 100% Job Assistance + Resume Building + Mock Interviews


### Who Should Attend?




This course is **designed for anyone who wants to break into or grow in the field of Data Engineering**, including:

1.  Fresh Graduates – B.Tech, B.Sc, BCA, MCA, M.Sc, etc.
2.  Software Developers looking to switch to Data Engineering
3.  Data Analysts & BI Developers wanting to scale up
4.  System Admins, DevOps, and Cloud Engineers
5.  Final-year students who want to become job-ready
6.  Non-CS Graduates with interest in data, databases, and cloud
7.  IT Professionals returning after a career break
8.  Professionals targeting **remote and international roles**



## Why This Course is Important (2025–2040 Outlook)

 **Data is the new oil.** Every business—from e-commerce to finance—is powered by data. With growing reliance on real-time analytics, machine learning, and cloud platforms:

-  **Demand for Data Engineers will grow 3x from 2025 to 2030** (NASSCOM, Gartner, LinkedIn Reports)
-  **Remote, hybrid, and onsite roles are booming** in India, US, UK, Germany, UAE, Canada, and more
-  From startups to top MNCs (Google, TCS, Capgemini, Amazon, Microsoft, etc.) are hiring skilled Data Engineers



## Job Opportunities & Salary Range

Role	Experience	Salary Range (India)
Data Engineer (Entry Level)	0–2 Years	₹7–12 LPA
Data Engineer (Mid-Level)	2–5 Years	₹12–20 LPA
Data Engineer (Abroad)	Any	\$80K–\$130K USD
Remote Global Projects	Any	₹15–30 LPA equivalent

✓ Remote roles and international placements are possible with skills in: **BigQuery, Spark, Kafka, Airflow, GCP, Docker, and Kubernetes.**

## What You'll Learn (Key Tools & Skills)

Domain	Tools & Tech You'll Learn
Data Warehousing	Google BigQuery
ETL & Orchestration	Apache NiFi, Apache Airflow
Big Data Processing	Apache Spark (PySpark), Apache Flink, Apache Beam
Databases	MySQL, PostgreSQL, Oracle, MongoDB, Cassandra, Redis, DynamoDB
Real-Time Streaming	Apache Kafka
Cloud Platform	Google Cloud Platform (GCP)
Containerization	Docker, Kubernetes
Monitoring & Logging	Prometheus, Grafana, ELK Stack
Data Governance	Great Expectations, Apache Griffin, Collibra
ML & Collaboration	Jupyter, TensorFlow, GitHub
Data Modeling	Erwin, Lucidchart


## Course Benefits

- ✓ 100% Job-Oriented
- ✓ Real-Time Projects
- ✓ Interview Preparation (DS, SQL, Cloud, System Design)
- ✓ Resume, LinkedIn & GitHub Portfolio Building
- ✓ 1:1 Mock Interviews
- ✓ Live + Recorded Sessions (Available at Office Only)
- ✓ Doubt Clarification and Mentorship from Working Architect
- ✓ Capstone Project (Batch + Real-Time Streaming)
- ✓ Certification + Career Guidance
- ✓ Tools & Techniques used in real-world company scenarios

## What Makes This Course Different?

- **Mentor from Industry:** Learn from Jitendra Mohanty – 15+ years of experience
- **Real-World Scenarios:** Build pipelines like those used at Amazon, Netflix, Boeing
- **Focus on Interview & Survival:** We don't just teach tools; we teach **how to crack interviews and perform confidently on the job**
- **Office-Only Recording Access:** Stay committed with monitored in-person access

## Final Outcome

 After this course, you will be able to:

- Build end-to-end data pipelines on GCP with Spark, Kafka, Airflow
- Handle batch and real-time data efficiently

- Design database models, monitor pipeline health
- Confidently clear interviews and excel in your job
- Apply for **remote, onsite, and international roles**



## **Weekly Module-Wise Breakdown:**

### **Month 1: Foundations of Data Engineering**

#### **Week 1: Introduction to Data Engineering & Git**

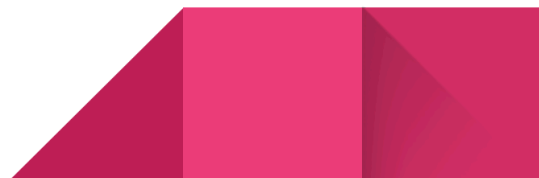
- Role of a Data Engineer
- Overview of Data Ecosystem
- Introduction to Git & GitHub
- Version Control & Best Practices

#### **Week 2: SQL Mastery for Data Engineers**

- MySQL, PostgreSQL: Basics to Advanced SQL
- Joins, Indexing, Views, Stored Procedures
- Introduction to Oracle DB

#### **Week 3: NoSQL Databases**

- MongoDB: Collections, Aggregation Framework



- Redis: Caching & Data Structures
- Cassandra: Columnar Data Storage
- DynamoDB Basics (AWS Overview)

#### **Week 4: Data Warehousing Concepts + Google BigQuery**

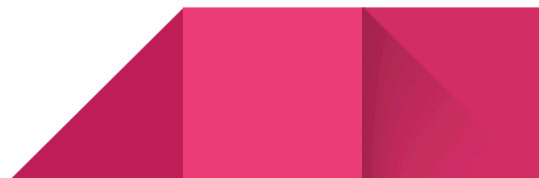
- OLTP vs OLAP, Star & Snowflake Schema
  - Google BigQuery: Setup, Datasets, SQL Interface
  - Data Partitioning & Clustering
  - Integration with BI Tools
- 

### **Month 2: ETL, Batch & Real-Time Processing**

#### **Week 5: ETL with Apache NiFi**

- Introduction to NiFi Architecture
- Creating Data Flows
- Connectors, Processors, Templates
- Real-Time Ingestion Scenarios

#### **Week 6: Orchestration with Apache Airflow**



- Airflow DAGs, Operators, and Schedulers
- Building ETL Pipelines
- Task Monitoring and Logging
- Integration with NiFi & Spark

### **Week 7: Apache Spark and PySpark**

- Spark Core & Spark SQL
- PySpark DataFrames, Transformations
- Writing ETL Jobs in PySpark
- Optimizing Spark Jobs

### **Week 8: Apache Flink & Beam**

- Introduction to Stream Processing
- Apache Flink: Event-time Processing, Windows
- Apache Beam Unified Model
- Dataflow (GCP) Execution with Beam

---

## **Month 3: Streaming, Cloud & Containerization**



## **Week 9: Real-time Data Streaming with Kafka**

- Kafka Architecture & Ecosystem
- Topics, Partitions, Producers, Consumers
- Kafka Connect, Schema Registry
- Kafka Integration with Spark/Flink


## **Week 10: Google Cloud Platform (GCP) for Data Engineers**

- GCP Services Overview (BigQuery, Dataflow, Cloud Storage, Pub/Sub)
- IAM Roles & Permissions
- Deploying Data Pipelines on GCP

## **Week 11: Docker for Data Engineers**

- Containerization Concepts
- Docker Images, Containers, Dockerfiles
- Docker Compose for Data Services

## **Week 12: Kubernetes Essentials**

- Kubernetes Architecture
  - Pods, Services, Deployments
- 



- Helm Charts
  - Orchestrating Data Pipelines with K8s
- 

## **Month 4: Data Modeling, Governance, ML Tools, & Job Prep**


### **Week 13: Data Modeling Tools**

- Dimensional Modeling
- Using Lucidchart & Erwin for ER Diagrams
- Translating Business Requirements into Data Models

### **Week 14: Data Quality, Monitoring & Logging**

- Data Validation with Great Expectations
- Data Quality with Apache Griffin
- Monitoring with Prometheus & Grafana
- ELK Stack for Centralized Logging

### **Week 15: ML Tools & Deployment for Data Engineers**

- Role of Data Engineers in ML Pipelines
  - Jupyter Notebooks for EDA
- 

- TensorFlow/PyTorch for Model Hosting
- Model Deployment on GCP using Docker/K8s

### **Week 16: Resume Building, Mock Interviews & Final Project**

- Creating a Job-Winning Resume
  - 1:1 Mock Interviews
  - Final Capstone Project Presentation
  - Interview Readiness + Placement Drive Support
- 

