



About the Company

We aim to bridge the gap between academic learning and industry expectations through practical, industry-oriented education. We are building a community of industry professionals passionate about sharing their expertise and making high-quality, industry-relevant education accessible to aspiring professionals.

Benefits of Completing the Course

- **Production-Grade Portfolio:** Build and deploy an entire end-to-end cloud data ecosystem hosted live on GitHub.
- **Modern Tech Stack Mastery:** Gain hands-on operational experience with SQL, Python, Snowflake, dbt, Databricks, PySpark, AWS S3, and Apache Airflow.
- **Real-World Engineering:** Solve actual enterprise challenges involving multi-source data extraction, relational modeling, and distributed cloud processing.
- **Career Acceleration:** Position yourself strongly for high-demand roles such as Junior Data Engineer, Analytics Engineer, and Cloud Data Engineer.

Month-by-Month Course Overview

Month 1: Core & Advanced SQL Analytics

- **Module 1: Core SQL & Relational Databases:** Relational database fundamentals and schema design; Data Definition Language (DDL) and Data Manipulation Language (DML); Basic table joins (Inner and Left joins); Querying e-commerce datasets for operational metrics.
- **Module 2: Advanced SQL & Complex Query Structures:** Complex joins including Full and Cross joins; Subqueries and nested query refactoring; Common Table Expressions (CTEs) for modular scripting; Writing readable, multi-stage query pipelines.
- **Module 3: Analytical SQL & Window Functions:** Analytical SQL window function architecture; Ranking functions (ROW_NUMBER, RANK, DENSE_RANK); Offset functions (LEAD, LAG) for historical tracking; Calculating month-over-month business growth and running totals.
- **Module 4: Query Optimization & Performance Tuning:** Database indexing strategies and maintenance; Reading and interpreting SQL execution plans; Identifying and fixing performance bottlenecks; Tuning resource-heavy, slow-running queries.

Month 2: Python Core, APIs & Data Wrangling

- **Module 1: Python Core for Data Engineers:** Python data structures (Lists, Dictionaries, Sets, Tuples); Environment setup and virtual environment management; Control flow, iteration loops, and comprehension patterns; Writing modular, reusable functions for extraction.
- **Module 2: APIs & Heterogeneous Data Formats:** REST API request lifecycle and authentication; Parsing nested JSON payloads programmatically; Reading and writing CSV and Parquet file formats; Pulling live structured data from external endpoints.
- **Module 3: Scalable Data Wrangling with Pandas:** Pandas DataFrame filtering, conditional selection, and slicing; Grouping, aggregating, and reshaping tabular datasets; Merging, joining, and concatenating multiple data frames; Handling missing values, duplicates, and data type cleaning.
- **Module 4: Resilient Coding & Error Handling:** Exception handling using try/except/finally blocks; Implementing structured script logging; Establishing reliable database connections via Python; Writing fail-safe data ingestion routines.

Month 3: Data Modeling, Snowflake, dbt & First Capstone

- **Module 1: Data Modeling & Architecture:** OLTP vs. OLAP database architecture principles; Dimensional modeling and schema design; Star schema and Snowflake schema architecture; Designing conceptual data models for enterprise reporting.
- **Module 2: Snowflake Cloud Data Warehouse:** Cloud data warehouse architecture and storage models; Provisioning compute clusters and managing virtual warehouses; Staging files and bulk loading using COPY INTO commands; Cloud-native data warehousing management.
- **Module 3: Analytics Engineering with dbt (Data Build Tool):** ELT (Extract, Load, Transform) methodology and workflow; Building modular SQL models inside dbt; Staging and data mart layer configuration; Automated data testing and documentation generation.
- **Module 4: First Half Capstone: Automated Analytics Pipeline:** Building a robust Python script to extract live API data; Loading raw JSON payloads into Snowflake data warehouse; Transforming raw data via dbt into clean Star schemas; Preparing analytics-ready tables for BI reporting.

Month 4: Big Data & PySpark Fundamentals

- **Module 1: Big Data Concepts & Cluster Provisioning:** Distributed systems architecture and core processing principles; Hadoop Ecosystem versus Apache Spark; Databricks workspace and cluster provisioning; Managing distributed compute environments.
- **Module 2: PySpark Fundamentals & DataFrames:** Resilient Distributed Datasets (RDDs) vs. Spark DataFrames; Understanding Spark's lazy evaluation engine; Core transformations and action operations; Executing basic data operations across 10M+ rows.
- **Module 3: Advanced PySpark Analytics:** Advanced window functions in distributed environments; Working with complex types (Arrays, Structs, Maps); Parsing nested JSON arrays using custom PySpark logic; Writing and applying User Defined Functions (UDFs).
- **Module 4: PySpark Performance Optimization:** Partitioning strategies and shuffle minimization; Caching and persistence mechanisms in memory; Diagnosing and fixing Out-Of-Memory (OOM) errors; Optimizing slow-running distributed query execution.

Month 5: AWS Cloud Infrastructure & Apache Airflow

- **Module 1: AWS Cloud Basics & Storage:** AWS Console navigation and core service overview; Identity and Access Management (IAM) roles and security policies; Amazon S3 bucket architecture and object storage; Programmatic S3 management using Python's boto3 library.
- **Module 2: Airflow Architecture & DAG Design:** Apache Airflow core architecture and scheduler mechanics; Designing Directed Acyclic Graphs (DAGs); Managing task dependencies and execution order; Writing basic DAGs to execute sequential Python tasks.
- **Module 3: Advanced Airflow Orchestration:** Utilizing PythonOperator and BashOperator; Conditional branching and dynamic pipeline scheduling; Managing Airflow Connections, Variables, and XComs; Building intelligent, reactive workflow automation.
- **Module 4: Pipeline Monitoring & Production Reliability:** Setting up secure connections to S3 and Snowflake; Implementing automated failure alerts and notifications; Troubleshooting task failures and handling retries; Production monitoring of orchestrated pipelines.

Month 6: Lakehouse Architecture, Version Control & Final Capstone

- **Module 1: Lakehouse Architecture & Medallion Design:** Data Lakes versus Data Warehouses architecture; Principles of modern Lakehouse systems; Implementing Medallion Architecture data flows; Orchestrating Bronze, Silver, and Gold data layers.
- **Module 2: Version Control & Engineering Best Practices:** Git workflow for data engineers (branching, commits, merges); Pull request reviews and collaboration practices; Structuring pipeline code into clean GitHub repositories; Maintaining reproducible repository documentation.
- **Module 3: Interview Preparation & System Design:** Mock technical interviews for data engineering roles; SQL and Python coding problem walkthroughs; System design whiteboarding for scalable data architectures; Portfolio presentation and resume optimization.
- **Module 4: Final Capstone: The Enterprise Cloud Lakehouse:** Deploying a fully automated production pipeline stack; Landing raw API data into AWS S3 Bronze layer via Airflow; Transforming and cleaning data with Databricks PySpark into Silver Parquet files; Loading pristine reporting tables into Snowflake Gold layer for BI consumption.

Learning Journey Visual Roadmap (Step-by-Step Process Flow Layout)

1

Month 1 Roadmap: Core & Advanced SQL Analytics
Deliverable: Complex SQL joins, CTEs, Window functions, and query performance optimization

2

Month 2 Roadmap: Python Core, APIs & Data Wrangling
Deliverable: Robust Python ingestion scripts, REST API parsers, and Pandas data cleaners

3

Month 3 Roadmap: Snowflake, dbt & First Capstone
Deliverable: Star schema models, dbt transformation layers, and Automated Analytics Pipeline

4

Month 4 Roadmap: Big Data & PySpark Fundamentals
Deliverable: Databricks clusters, PySpark DataFrames, nested JSON parsing, and OOM tuning

5

Month 5 Roadmap: AWS Cloud Infrastructure & Airflow
Deliverable: AWS S3 storage integration, Airflow DAGs, and automated task orchestration

6

Month 6 Roadmap: Lakehouse Architecture & Final Capstone
Deliverable: Medallion architecture (Bronze-Silver-Gold), Git versioning, and Enterprise Cloud Lakehouse

Course Pricing & Investment

Course Name	Original Price	Discount	Price After Discount
First 3 Months (Core Data Engineering)	₹50,000	30%	₹35,000
Next 3 Months (Advanced Data Engineering)	₹50,000	30%	₹35,000
6-Month Course (Full Course)	₹70,000	30%	₹49,000