

Data Engineering

Job-Ready Online Program · Course Duration 60-65 Hours

Free Foundation — 15 Hours

Paid Job-Ready Program — 48 Hours

Total — 63 Hours

In short: This Data Engineering program is a 60-65 Hours, fully online, job-ready course. You start with 15 hours of free foundations (Python, SQL, Linux & shell, Git) and continue into a 48-hour paid program covering advanced SQL, data modeling and warehousing, ETL/ELT pipelines, big data with Apache Spark, orchestration with Airflow, streaming with Kafka, analytics engineering with dbt, cloud data platforms on AWS and GCP, and Docker & CI/CD — finishing with an end-to-end capstone data platform, a portfolio, and full interview prep. On completion you are ready for Data Engineer, Analytics Engineer, and ETL Developer roles.

Free Foundation — 15 Hours

Build your core Python, SQL, and Linux foundations before advancing to the job-ready Data Engineering program.

Module 1: Programming Foundations with Python

5 Hours

LEARNING OBJECTIVES

- Write and run Python programs confidently
- Use core programming logic to prepare for data pipelines

TOPICS

- Python setup, syntax, variables, and data types
- Conditionals, loops, functions, and modules
- Lists, dictionaries, and file handling for data processing

HANDS-ON

- Build Python scripts that read, clean, and transform data files

Assessment: Quiz + Coding Exercises

Module 2: SQL & Relational Database Fundamentals

4 Hours

LEARNING OBJECTIVES

- Query relational databases with confidence
- Understand how data is stored and related

TOPICS

- SELECT, WHERE, GROUP BY, and ORDER BY
- Joins, keys, and relationships
- Aggregations, subqueries, and table design basics

HANDS-ON

- Write SQL queries against a real relational database

Assessment: Quiz + Lab

Module 3: Linux, Shell & Data Formats**4 Hours****LEARNING OBJECTIVES**

- Work comfortably in the Linux command line
- Handle the file formats data engineers use daily

TOPICS

- Linux navigation, permissions, and shell scripting
- Automating tasks with cron and shell pipelines
- Working with CSV, JSON, Parquet, and APIs

HANDS-ON

- Automate a data-download and preprocessing task with a shell script

Assessment: Quiz + Lab**Module 4: Version Control with Git & GitHub****2 Hours****LEARNING OBJECTIVES**

- Track pipeline code and configs with Git
- Collaborate and showcase work on GitHub

TOPICS

- Repositories, commits, and branches
- Merging, pull requests, and conflict resolution
- Building a professional data engineering GitHub profile

HANDS-ON

- Push your first pipeline projects to GitHub

Assessment: Quiz + Lab**Paid Job-Ready Program — 48 Hours**

Master advanced SQL, data modeling and warehousing, ETL/ELT, Apache Spark, Airflow, Kafka, dbt, cloud data platforms, and CI/CD — everything you need to land your first data engineering job.

Module 5: Advanced SQL & Analytical Query Engineering**5 Hours****LEARNING OBJECTIVES**

- Write advanced, performant analytical SQL
- Optimize queries for large datasets

TOPICS

- Window functions and common table expressions (CTEs)
- Advanced joins, set operations, and pivoting
- Indexing, query plans, and performance tuning

HANDS-ON

- Optimize a slow analytical query on a large table

Assessment: Project Milestone

Module 6: Data Modeling & Data Warehousing**5 Hours****LEARNING OBJECTIVES**

- Design data models for analytics at scale
- Build and query modern data warehouses

TOPICS

- Normalization, star and snowflake schemas
- Fact and dimension tables, slowly changing dimensions
- Cloud warehouses: Snowflake, BigQuery, and Redshift

HANDS-ON

- Design and load a dimensional data warehouse

Assessment: Project Milestone**Module 7: Building ETL & ELT Pipelines with Python****5 Hours****LEARNING OBJECTIVES**

- Design reliable ETL and ELT data pipelines
- Ingest, transform, and load data from many sources

TOPICS

- ETL vs ELT patterns and pipeline architecture
- Extracting from APIs, files, and databases
- Transformations, idempotency, and data validation

HANDS-ON

- Build a Python ETL pipeline from source to warehouse

Assessment: Project Milestone**Module 8: Big Data Processing with Apache Spark****6 Hours****LEARNING OBJECTIVES**

- Process large-scale data with distributed computing
- Write efficient Spark transformations

TOPICS

- Spark architecture, RDDs, and DataFrames
- PySpark transformations, actions, and Spark SQL
- Partitioning, caching, and performance optimization

HANDS-ON

- Process a multi-gigabyte dataset with PySpark

Assessment: Project Milestone

Module 9: Workflow Orchestration with Apache Airflow**5 Hours****LEARNING OBJECTIVES**

- Schedule and orchestrate complex data pipelines
- Monitor, retry, and manage pipeline dependencies

TOPICS

- DAGs, tasks, operators, and scheduling
- Dependencies, retries, backfills, and SLAs
- Monitoring, logging, and alerting

HANDS-ON

- Orchestrate a multi-step ETL pipeline with Airflow

Assessment: Project Milestone**Module 10: Streaming Data with Apache Kafka****4 Hours****LEARNING OBJECTIVES**

- Build real-time data pipelines
- Understand event-driven data architecture

TOPICS

- Kafka topics, producers, consumers, and partitions
- Stream processing and delivery guarantees
- Batch vs streaming trade-offs

HANDS-ON

- Build a real-time streaming pipeline with Kafka

Assessment: Lab**Module 11: Cloud Data Platforms (AWS & GCP) & Data Lakes****5 Hours****LEARNING OBJECTIVES**

- Build data lakes and pipelines in the cloud
- Use managed data services on AWS and GCP

TOPICS

- Data lakes on S3 / Google Cloud Storage
- AWS Glue, Athena, EMR and GCP Dataflow, BigQuery
- Lakehouse architecture and file formats (Parquet, Delta)

HANDS-ON

- Build a cloud data lake and query it serverlessly

Assessment: Project Milestone

Module 12: Analytics Engineering with dbt & the Modern Data Stack**4 Hours****LEARNING OBJECTIVES**

- Transform warehouse data with dbt
- Apply software engineering practices to analytics

TOPICS

- dbt models, sources, refs, and materializations
- Testing, documentation, and data quality
- The modern data stack and ELT workflow

HANDS-ON

- Build and test a dbt transformation project

Assessment: Lab**Module 13: Docker, CI/CD & Deploying Data Pipelines****5 Hours****LEARNING OBJECTIVES**

- Containerize and deploy data pipelines
- Automate testing and delivery with CI/CD

TOPICS

- Docker images, containers, and Docker Compose
- CI/CD pipelines with GitHub Actions
- Data quality monitoring, logging, and observability

HANDS-ON

- Containerize a pipeline and deploy it via CI/CD

Assessment: Project Milestone**Module 14: Capstone Data Platform Project, Portfolio & Job Readiness****4 Hours****LEARNING OBJECTIVES**

- Ship a complete, portfolio-ready data platform
- Prepare for the data engineering job search

TOPICS

- End-to-end capstone: ingestion, warehouse, orchestration
- Portfolio, GitHub, and resume optimization
- Data engineering interview, SQL, and system-design prep

HANDS-ON

- Build and present an end-to-end cloud data platform
- Mock interviews and resume review

Assessment: Capstone Project + Mock Interview

Free Foundation

15 Hours

Paid Job-Ready Program

48 Hours

Total Course**63 Hours**Enrol now — 50% OFF · Book a free demo at [progmatrix.com/courses/data-engineering](https://www.progmatrix.com/courses/data-engineering)