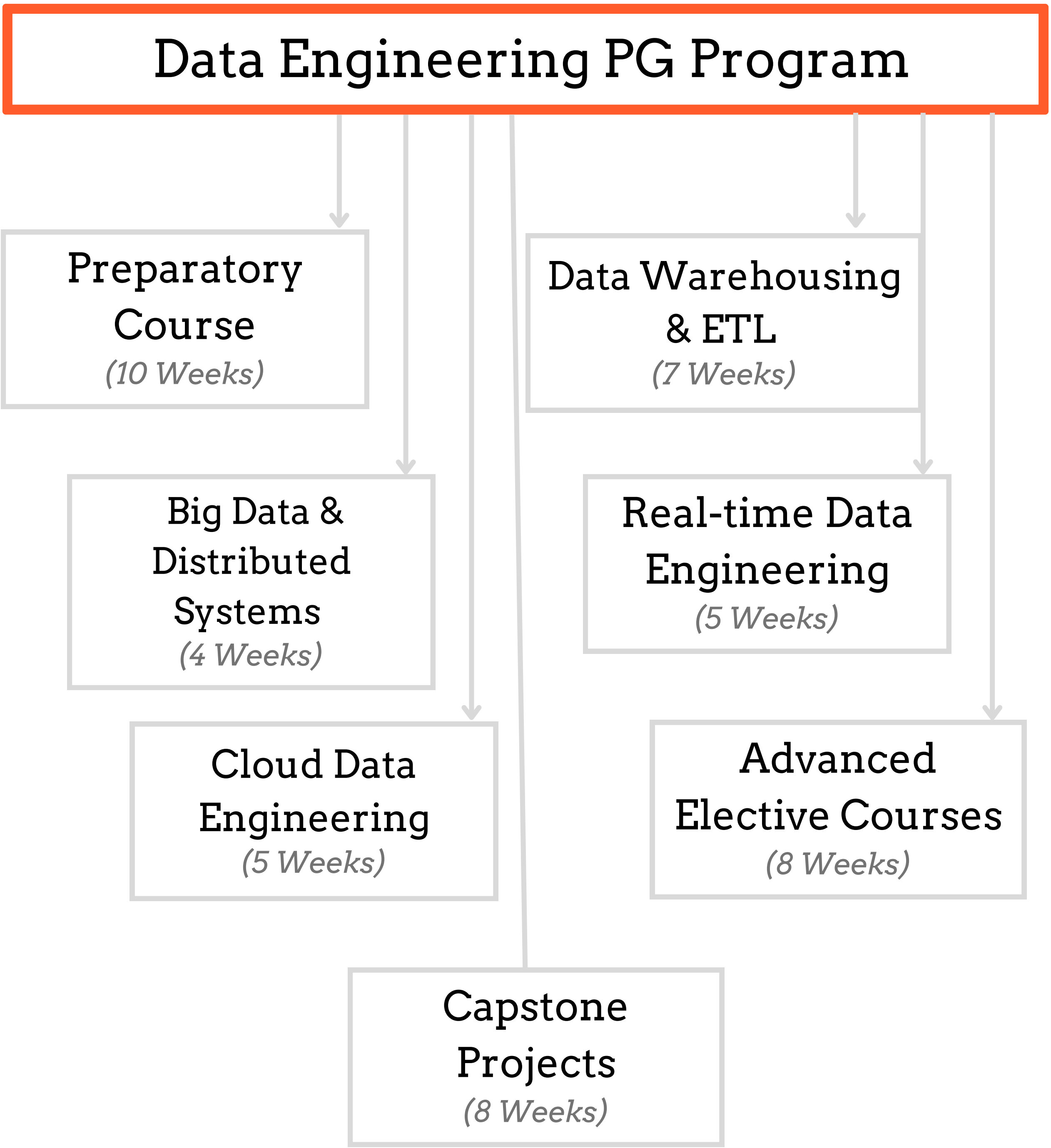
MERITSHOT

Program Delivery Plan

Data Engineering PG Program

A Complete timeline guide for the 12 Months Data Engineering PG Program offered at Meritshot

Course Outline



Data Engineering PG Program

Preparatory Course
(10 Weeks)

Big Data & Distributed Systems
(4 Weeks)

Cloud Data Engineering
(5 Weeks)

Data Warehousing & ETL
(7 Weeks)

Real-time Data Engineering
(5 Weeks)

Advanced Elective Courses
(8 Weeks)

Capstone Projects
(8 Weeks)

Preparatory Course

Build strong foundations in data tools, SQL, and programming.

Week 1-2	Weekends	Learn key Excel functions, Pivot Tables, Charts, Conditional Formatting, and tools like Scenario and What-if Analysis.
	Weekdays	Learn SQL basics including DDL, DML, SELECT queries, and key concepts like tables, keys, and constraints, along with joins, subqueries, and aggregations.
Week 3-5	Weekends	Learn Python fundamentals including data types, control flow, file handling, exceptions, and list/dictionary comprehensions.
	Weekdays	Work with Pandas for data handling, clean and transform datasets, and access data using basic APIs.
Week 6-8	Weekends	Create interactive dashboards using Power BI and Tableau with DAX, slicers, and calculated fields.
	Weekdays	Connect BI tools to databases, build data models, and create a sales and marketing dashboard.
Week 9-10	Weekends	Learn Git basics and collaborate using GitHub with branching.
	Weekdays	Use Git in VSCode and collaborate on a mini team project.

Data Warehousing & ETL

Learn to model, build, and automate data pipelines and warehouses.

Week 1-2	Weekends	Dimensional Modeling: Star, Snowflake, Fact vs Dimension Data Lake vs Warehouse
	Weekdays	Data Modeling with dbt (data build tool)
Week 3-4	Weekends	*ETL Concepts: Extraction (CSV/API/DB), Transformation, Load, Tools: Talend, Apache NiFi, Informatica (basics)
	Weekdays	Work with Pandas for data handling, clean and transform datasets, and access data using basic APIs.
Week 5–6	Weekends	Data Warehouse: Amazon Redshift, Snowflake, BigQuery Data Partitioning, Clustering, Performance Optimization
	Weekdays	Batch Ingestion: Scheduling Jobs with Apache Airflow
Week 7	Weekends	End-to-End ETL Pipeline: Retail Sales Analysis
	Weekdays	Use dbt + Airflow to automate data loading and transformation

Data Warehousing & ETL

Learn to model, build, and automate data pipelines and warehouses.

Week 1-2	Weekends	Dimensional Modeling: Star, Snowflake, Fact vs Dimension Data Lake vs Warehouse
	Weekdays	Data Modeling with dbt (data build tool)
Week 3-4	Weekends	*ETL Concepts: Extraction (CSV/API/DB), Transformation, Load, Tools: Talend, Apache NiFi, Informatica (basics)
	Weekdays	Work with Pandas for data handling, clean and transform datasets, and access data using basic APIs.
Week 5–6	Weekends	Data Warehouse: Amazon Redshift, Snowflake, BigQuery Data Partitioning, Clustering, Performance Optimization
	Weekdays	Batch Ingestion: Scheduling Jobs with Apache Airflow
Week 7	Weekends	End-to-End ETL Pipeline: Retail Sales Analysis
	Weekdays	Use dbt + Airflow to automate data loading and transformation

Big Data & Distributed Systems

Handle massive datasets using Hadoop and Spark.

Week 1-2	Weekends	Hadoop Ecosystem: HDFS, YARN, Hive, Pig (Intro), Spark Basics: SparkContext, RDDs, DataFrames
	Weekdays	Data Transformations in PySpark
Week 3-4	Weekends	Spark SQL: Filtering, Joins, Window Functions, Data Caching, Partitioning, Broadcasting
	Weekdays	Hands-on Project: Analyzing server logs using Spark

Real-time Data Engineering

Build real-time data pipelines with Kafka and Spark Streaming.

Week 1-2	Weekends	Kafka Architecture: Producers, Consumers, Brokers, Topics, Kafka Streams and Connectors
	Weekdays	Stream vs Batch Processing
Week 3-5	Weekends	Spark Streaming with Kafka integration Sliding Windows, Stateful Streams
	Weekdays	Use Case: Real-time fraud detection in transaction data, Monitoring with Kafka Manager and Grafana

Cloud Data Engineering

Use cloud-native tools for storage, compute, and orchestration.

Week 1-2	Weekends	AWS: IAM, S3, EC2, Lambda GCP: BigQuery, Cloud Storage, Cloud Functions
	Weekdays	Serverless ETL with AWS Glue or GCP Dataflow, Use Athena or BigQuery for SQL on Data Lake, Hands-on: Build Cloud ETL from S3 to Redshift
Week 3-4	Weekends	Introduction to Data Manipulation Functions, Statistical Transformations, and Feature Engineering.
	Weekdays	Case Study on Data Cleansing and Enrichment for a Job Portal
Week 5	Weekends	Orchestrating Cloud Pipelines with MWAA / Cloud Composer
	Weekdays	Monitoring and alerting best practices

Advanced Elective Courses

Choose 2 of the following electives:

Week 1	DevOps for Data Engineering * Docker, Containerization, Docker Compose * CI/CD Pipelines: GitHub Actions, Jenkins * Logging and Monitoring with ELK Stack
Week 2	* Encryption: Symmetric, Asymmetric * Role-Based Access, Row-Level Security * Data Lineage and Auditing
Week 3	*** MLflow Basics** *** Feature Stores and Model Registry** *** Deployment of ML Pipelines**

Capstone Projects

Solve real-world data engineering challenges with guidance. Project Tracks

Week 1-3	Ecommerce- * Clickstream Ingestion with Kafka * Building a Recommendation Pipeline with Spark
Week 4-6	2. Finance: * Ingesting and cleaning daily trades and transactions * Building a Data Lakehouse with Delta Lake
Week 7-9	3. Healthcare: * Real-time Patient Data Ingestion via IoT * Data Aggregation and BI Dashboards
Week 10-12	4. Logistics: * GPS-based Vehicle Tracking Data Pipeline * Predictive Maintenance with streaming analytics

Industry Projects and Case Studies

Deliver Real-time Product Recommendations at Scale

Utilize scalable data pipelines and streaming technologies to process user activity logs in real time, enabling product recommendation systems to deliver timely and relevant suggestions.

Design and implement ETL workflows and real-time data ingestion frameworks to support ML-driven personalization, improving latency and data freshness for ecommerce platforms.

Skills: Apache Kafka, Apache Spark, Data Pipelines, ETL, AWS/GCP, SQL, NoSQL, Real-time Processing

Streamline Patient Health Records for Real-time Insights

Enable healthcare providers to access up-to-date patient data by building a unified, secure, and scalable data platform.

Integrate disparate health data sources (EHRs, labs, wearable devices) into a centralized data lake using batch and stream processing, enabling real-time monitoring and analytics for patient care and operational efficiency.

Skills: Apache NiFi, Spark Streaming, Hadoop, Data Lakes, HIPAA-compliant data handling, SQL, Parquet/Avro

Industry Projects and Case Studies

Deliver Real-time Product Recommendations at Scale

Utilize scalable data pipelines and streaming technologies to process user activity logs in real time, enabling product recommendation systems to deliver timely and relevant suggestions.

Design and implement ETL workflows and real-time data ingestion frameworks to support ML-driven personalization, improving latency and data freshness for ecommerce platforms.

Skills: Apache Kafka, Apache Spark, Data Pipelines, ETL, AWS/GCP, SQL, NoSQL, Real-time Processing

Streamline Patient Health Records for Real-time Insights

Enable healthcare providers to access up-to-date patient data by building a unified, secure, and scalable data platform.

Integrate disparate health data sources (EHRs, labs, wearable devices) into a centralized data lake using batch and stream processing, enabling real-time monitoring and analytics for patient care and operational efficiency.

Skills: Apache NiFi, Spark Streaming, Hadoop, Data Lakes, HIPAA-compliant data handling, SQL, Parquet/Avro

Industry Projects and Case Studies

Build a Scalable Financial Transactions Pipeline

Create robust ETL pipelines to handle high-volume financial data for fraud detection, reporting, and compliance.

Design ingestion and transformation workflows that validate, enrich, and store transaction data efficiently for downstream analytics and alerting systems in banking and fintech environments.

Skills: Apache Airflow, Kafka, Snowflake, SQL, Python, Data Quality Checks, Data Warehousing

Optimize Logistics Operations with Real-time Shipment Tracking

Develop a data infrastructure to collect, process, and store live GPS and sensor data from delivery fleets.
Use stream processing and cloud storage to feed real-time dashboards, route optimizations, and predictive maintenance models that improve efficiency in supply chain management.

Skills: Apache Flink, AWS Kinesis or GCP Pub/Sub, Redshift/BigQuery, Docker, Data APIs, Time-series Databases

Industry Projects and Case Studies

Real-time Inventory Monitoring for Retail Chains

Track inventory across multiple warehouses and stores by building a centralized realtime data platform.

Ingest and process transactional and sensor data to prevent stockouts, forecast demand, and automate replenishment decisions.

Skills: Kafka, Delta Lake, Spark Structured Streaming, Azure Data Factory, SQL, BI Tools

Automated Data Warehouse for Marketing Analytics

Build a modern data warehouse to unify marketing campaign data from multiple platforms

Design ETL pipelines to collect and normalize data from Google Ads, Facebook, and CRMs for campaign performance tracking and ROI analysis.

Skills: Airbyte, dbt, BigQuery, SQL, Looker, Data Modeling, APIs

IoT Data Pipeline for Smart Agriculture

Process high-frequency sensor data from farms to monitor soil, weather, and irrigation patterns.

Develop edge-to-cloud streaming architecture that enables real-time dashboards and alerts for crop health and yield optimization.

Skills: MQTT, Apache Flink, InfluxDB, AWS IoT Core, Grafana, Time-series Analytics