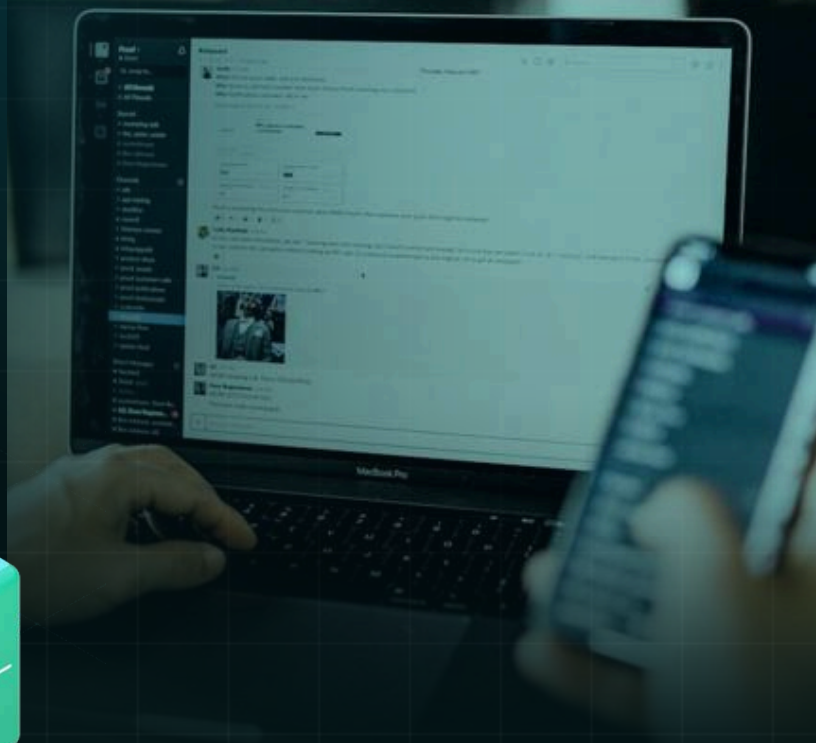
✦ DATA SCIENCE & ANALYTICS · INTERMEDIATE

Data Engineering Fundamentals

Move data reliably.





80	8	SAM-DE-F
HOURS	MODULES	CERTIFICATION

✦ INSIDE THIS BROCHURE

Contents

Program overview 03

Who it's for 04

Learning outcomes 05

Program runway 06

Detailed curriculum 07

Tools & platforms 10

Projects & capstone 11

Certification & method 12

Learning path 13

FAQs 14



AT A GLANCE

Level	Intermediate
Duration	80 Hours
Modules	8
Certificate	SAM-DE-F

✦ PROGRAM OVERVIEW

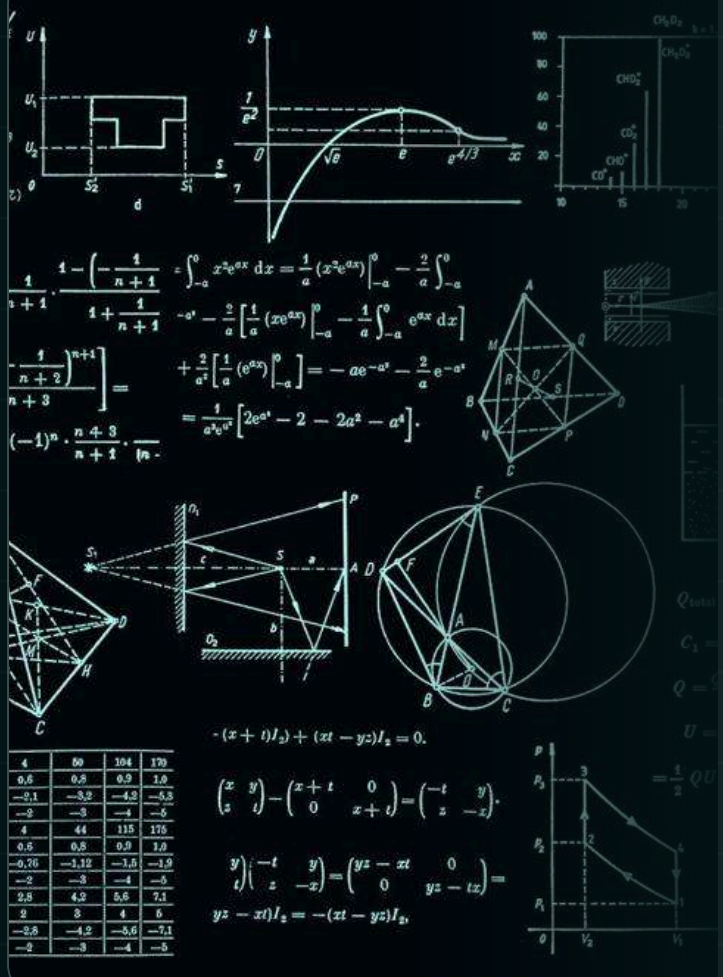
Move data reliably.

Foundations of data engineering — databases, ETL/ELT, data modelling and pipeline basics.



Reliable analytics and AI depend on data that is collected, modelled and moved correctly. This intermediate program builds data engineering foundations: relational and NoSQL databases, data modelling, ETL and ELT pipelines in Python and SQL, transformation with dbt, cloud warehouses, orchestration with Airflow, containerisation with Docker and pipeline monitoring basics.

LEVEL	DURATION	CERTIFICATION	DELIVERY
Intermediate	80 Hours	SAM-DE-F	Instructor-led · Live virtual · Blended



✦ WHAT YOU'LL PRACTISE

Design data models	01
Build ETL/ELT pipelines	02
Work with SQL and NoSQL stores	03
Schedule and monitor jobs	04

✦ WHO IT'S FOR

Built for people ready to grow.

01

Business professionals

Data literacy, Excel and SQL foundations

02

Analysts

Power BI, Tableau and Python for analysis

03

Data engineers

Pipelines with Spark, Kafka and Airflow

04

Data scientists & architects

ML, NLP, computer vision and platform design

PREREQUISITES

Basic Python programming	01
Working SQL knowledge	02
Familiarity with command line basics	03

ROLES YOU CAN TARGET

- Data Engineer (Junior) ↗
- ETL Developer ↗
- Analytics Engineer ↗
- Data Pipeline Developer ↗
- BI Developer ↗

✦ LEARNING OUTCOMES

What you'll be able to do.

01 Design normalised and dimensional data models

02 Build ETL and ELT pipelines with Python and SQL

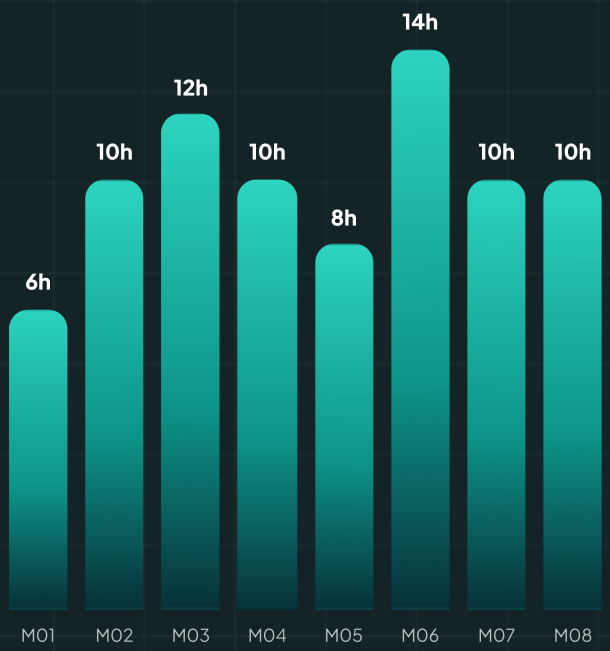
03 Work effectively with PostgreSQL and NoSQL data stores

04 Transform warehouse data using dbt models and tests

05 Schedule and monitor pipelines using Apache Airflow

06 Package pipeline components reproducibly with Docker

HOURS PER MODULE



✦ PROGRAM RUNWAY

Your journey, module by module.

8
MODULES · 80 HRS

01	**Data Engineering Landscape** Role of the data engineer, Data lifecycle and architecture, OLTP vs OLAP systems	6 hrs
02	**Linux, Git and Docker Essentials** Linux command line and shell scripting, Git workflows and branching, Docker images and containers	10 hrs
03	**Relational Databases and Advanced SQL** PostgreSQL architecture basics, DDL, constraints and indexes, Transactions and isolation	12 hrs
04	**Data Modelling** Normalisation and normal forms, Dimensional modelling, Star and snowflake schemas	10 hrs
05	**NoSQL and File Formats** MongoDB document model, Key-value stores like Redis, CSV, JSON, Parquet and Avro	8 hrs
06	**Building ETL/ELT Pipelines in Python** Extracting from APIs and databases, Transforming with Pandas, Loading with SQLAlchemy	14 hrs
07	**Cloud Warehouses and dbt** Cloud warehouses: BigQuery, Snowflake, Redshift, ELT vs ETL trade-offs, dbt models and materialisations	10 hrs
08	**Orchestration and Monitoring** Apache Airflow concepts, Writing DAGs and tasks, Scheduling and backfills	10 hrs
★	**Capstone · Containerised End-to-End Data Pipeline** Assessment, capstone review and certification	

✦ DETAILED CURRICULUM

The syllabus.

Modules 01–03

MODULE 01

Data Engineering Landscape

6 hours

Hands-on lab
Map the data flow of a sample business from source systems to dashboards.

TOPICS COVERED

- Role of the data engineer
- Data lifecycle and architecture
- OLTP vs OLAP systems
- Batch and streaming overview
- Modern data stack components

MODULE 02

Linux, Git and Docker Essentials

10 hours

Hands-on lab
Run PostgreSQL and a Python app together using Docker Compose.

TOPICS COVERED

- Linux command line and shell scripting
- Git workflows and branching
- Docker images and containers
- Docker Compose for local stacks
- Environment and secrets basics

Python · Linux · Git · Docker

MODULE 03

Relational Databases and Advanced SQL

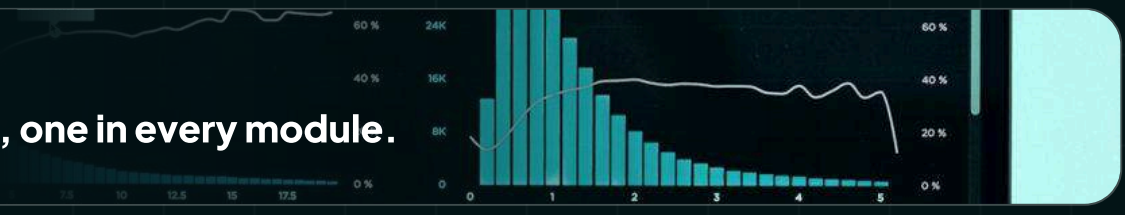
12 hours

Hands-on lab
Design and index a PostgreSQL schema, then optimise slow queries.

TOPICS COVERED

- PostgreSQL architecture basics
- DDL, constraints and indexes
- Transactions and isolation
- Advanced SQL: CTEs and windows
- Query plans and optimisation
- Stored procedures and views

PostgreSQL



LEARN BY DOING

8 guided labs, one in every module.

✦ DETAILED CURRICULUM

Syllabus, continued.

Modules 04–06

MODULE 04

Data Modelling

10 hours

Hands-on lab
Design a star schema with a slowly changing customer dimension.

TOPICS COVERED

- Normalisation and normal forms
- Dimensional modelling
- Star and snowflake schemas
- Slowly changing dimensions
- Data vault overview

MODULE 05

NoSQL and File Formats

8 hours

Hands-on lab
Load JSON events into MongoDB and convert them to Parquet files.

TOPICS COVERED

- MongoDB document model
- Key-value stores like Redis
- CSV, JSON, Parquet and Avro
- Object storage: S3 and equivalents
- Choosing the right store

MongoDB · Redis

MODULE 06

Building ETL/ELT Pipelines in Python

14 hours

Hands-on lab
Build an incremental pipeline from a public API into PostgreSQL.

TOPICS COVERED

- Extracting from APIs and databases
- Transforming with Pandas
- Loading with SQLAlchemy
- Incremental and idempotent loads
- Error handling and logging
- Testing pipeline code

Python · pandas · PostgreSQL



LEARN BY DOING

8 guided labs, one in every module.

✦ DETAILED CURRICULUM

Syllabus, continued.

Modules 07–08

MODULE 07

Cloud Warehouses and dbt

10 hours

Hands-on lab
Model raw tables into analytics-ready marts with dbt tests.

TOPICS COVERED

- Cloud warehouses: BigQuery, Snowflake, Redshift
- ELT vs ETL trade-offs
- dbt models and materialisations
- dbt tests and documentation
- Sources, seeds and snapshots

MODULE 08

Orchestration and Monitoring

10 hours

Hands-on lab
Schedule the ingestion and dbt pipeline in Airflow with failure alerts.

TOPICS COVERED

- Apache Airflow concepts
- Writing DAGs and tasks
- Scheduling and backfills
- Retries, alerts and SLAs
- Data quality checks
- Pipeline observability basics

Apache Airflow



CAPSTONE

Containerised End-to-End Data Pipeline

✦ YOUR STACK

Tools you'll build with.

Real platforms and frameworks — the same stack modern teams ship with, practised in cloud-hosted labs.

Python
Automation & scripting

PostgreSQL
Relational database

Apache Airflow
Pipeline orchestration

Docker
Containers & images

Linux
Server & CLI fundamentals

Git
Version control

pandas
Data wrangling

MongoDB
Document database

Redis
In-memory data store



LAB ACCESS

24×7 cloud labs for the length of the program

✦ HANDS-ON WORK

Projects that prove you can do it.

01

API to Warehouse Pipeline

Extract data from a public API, transform it with Python and load it incrementally into a PostgreSQL warehouse.

02

Dimensional Sales Model

Design and implement a star schema with slowly changing dimensions and dbt models with tests.

03

Scheduled Pipeline with Alerts

Orchestrate an end-to-end pipeline in Airflow with retries, data quality checks and automatic failure notifications to the team.

CAPSTONE PROJECT

Containerised End-to-End Data Pipeline

Build a Dockerised pipeline that ingests data from multiple sources, loads raw data, transforms it into a dimensional model with dbt, orchestrates runs with Airflow and feeds a simple dashboard, with documentation and tests.

Real-world scenario · Mentor review · Portfolio ready



CERTIFICATION

SAM-DE-F

Awarded on successful completion of the program assessments.

- MCQ assessments
- Practical lab evaluation
- Project & capstone review
- Final presentation



✦ HOW YOU LEARN

The SAM Learning Framework™

PHASE 01

S

Scenario

Real-world attack and use-case scenarios open every module.

PHASE 02

A

Apply

Hands-on lab work, CTF challenges and live projects.

PHASE 03

M

Master

Assessment, peer review and certification prep.

PHASE 04

M

Mentor

Post-program career guidance and support.

DELIVERY MODELS

- Instructor-Led (ILT)
- Virtual Instructor-Led
- Self-Paced Online
- E-Learning Licensing
- Blended Learning
- Corporate Onsite
- Bootcamp Intensive
- Hackathon / CTF
- Mentorship

✦ LEARNING PATH · DATA SCIENCE & ANALYTICS

Where this program fits.

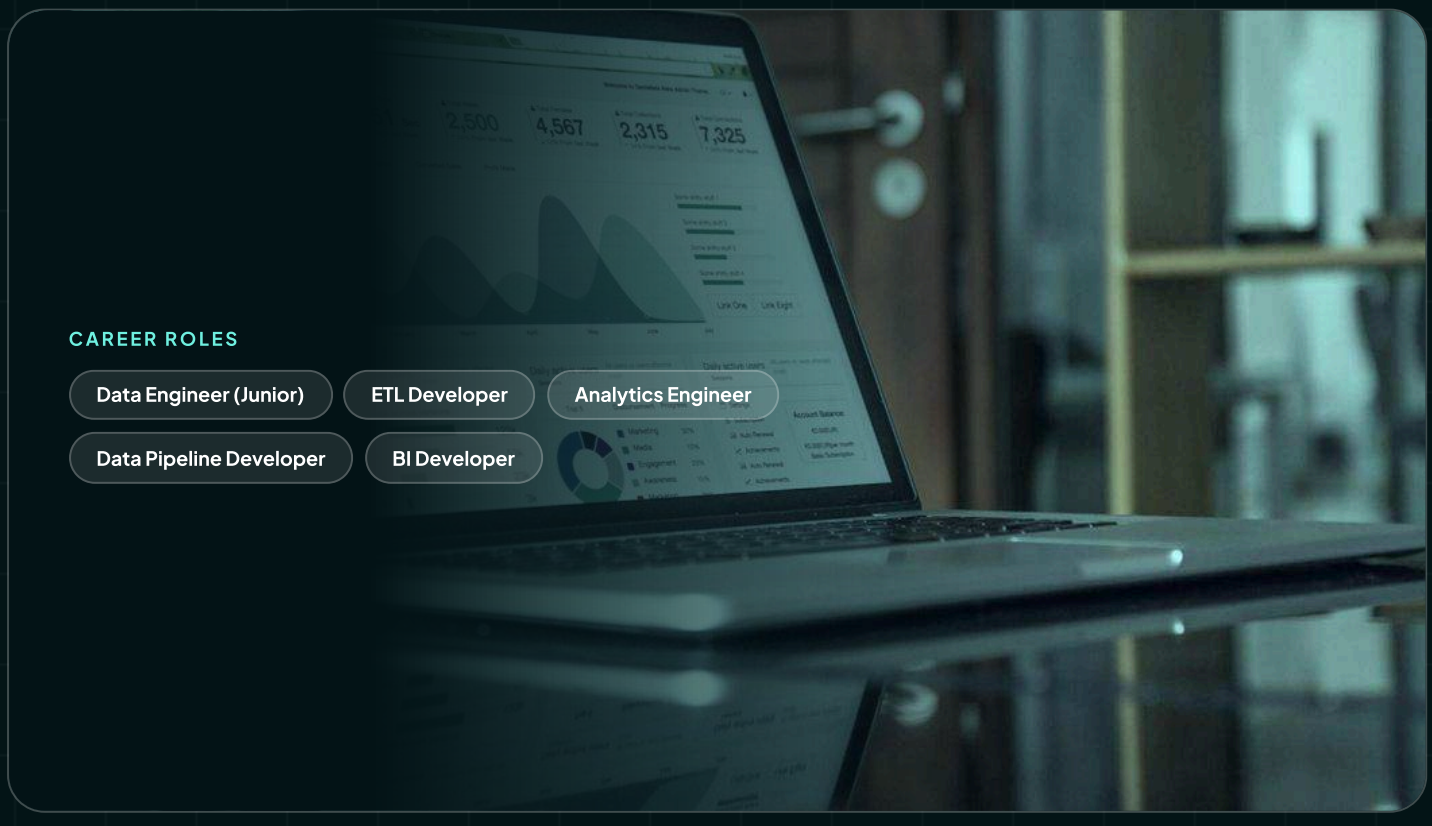
Beginner	Intermediate	Advanced	Expert
3 programs	You are here	4 programs	2 programs

BEFORE THIS

Data Literacy for Business Professionals	16 Hours
Excel for Data Analysis	24 Hours
SQL for Data Analysts	40 Hours

UP NEXT

Advanced Data Engineering (Spark/Kafka/Airflow)	96 Hours
Machine Learning for Data Scientists	80 Hours
NLP for Data Scientists	64 Hours
Computer Vision	64 Hours



CAREER ROLES

- Data Engineer (Junior)
- ETL Developer
- Analytics Engineer
- Data Pipeline Developer
- BI Developer

✦ FAQS

Questions, answered.

Q01

Do I need prior experience?

You should have basic Python programming skills and working SQL knowledge. No prior data engineering experience is required. Learners new to SQL are advised to complete SQL for Data Analysts before joining this program.

Q02

How are classes scheduled and delivered?

Sessions are instructor-led and run as live virtual or blended batches, with weekday and weekend options where available. Each module combines concept sessions with hands-on labs, and notes support revision between classes.

Q03

Will I receive a certificate?

Yes. Learners who complete the required labs, projects and final assessment receive the Samcommunity Academy SAM-DE-F certificate. It recognises the practical skills you demonstrated during the program rather than attendance alone.

Q04

How do I access labs and tools?

Labs run locally with Docker, PostgreSQL, Python, dbt and Airflow, with optional cloud warehouse free tiers. Setup guides are shared at the start, and instructors help resolve environment issues.

Q05

Is career support included?

You get guidance on portfolio projects, resume positioning and interview preparation for relevant roles. Pipeline projects are structured to be shared on GitHub as portfolio evidence. We do not guarantee jobs, placements or salaries.



✦ LET'S TALK

Start Data Engineering Fundamentals with Samcommunity.

INDIVIDUALS

academy@samcommunity.in

Batches, enrolment & career guidance

CORPORATE & INSTITUTIONS

corporate@samcommunity.in

Team batches & custom programs

Xylyn Samcommunity Private Limited · DPIIT-recognized startup · Surat, Gujarat, India