SODBRIDGE ACADEMY
ADVANCED PROFESSIONAL PROGRAMME

TRACK: DATA ENGINEERING

Professional Training Curriculum

Programme Detail	Information
Duration	12 Weeks
Live Instruction	6 Hours Per Week
Total Live Instruction	72 Hours
Weekly Schedule	3 Sessions × 2 Hours
Programme Fee	₦250,000

1. Programme Overview

The Data Engineering track focuses on the practical development of data pipelines and systems for collecting, transforming, validating and preparing data for analytics and downstream applications.

Learners use Python, REST APIs, PostgreSQL and modern data transformation practices to build repeatable workflows, improve data quality, handle failures and automate pipeline execution.

The programme also introduces Azure Data Factory and cloud storage as an industry-relevant cloud integration pathway.

Emphasis is placed on building maintainable, reliable and well-documented data workflows rather than simply writing scripts that work once.

2. Learning Outcomes

By the end of the programme, learners should be able to:

- Develop maintainable Python scripts and modular data workflows.
- Extract data from files and REST APIs.
- Handle common ingestion challenges such as pagination, inconsistent formats and failed requests.
- Design relational database tables and load data into PostgreSQL.
- Develop ETL/ELT transformations using Python and SQL.
- Organise reusable SQL transformation models with dbt.
- Apply data validation, logging, retries and failure-handling practices.
- Automate pipeline execution.
- Understand cloud-based data ingestion using Azure Data Factory and cloud storage.
- Document pipeline architecture and operational considerations.

3. Entry Requirements

Applicants should have:

- Working knowledge of Python, including functions, modules, files and common data structures.
- Basic to intermediate SQL, including SELECT, JOIN, GROUP BY and filtering.
- Familiarity with relational databases.
- Basic familiarity with command-line or development environments.

- Willingness to practise independently between live sessions.

Applicants without sufficient Python or SQL foundations may be advised to complete preparatory training before joining the programme.

4. Tools and Technologies

The programme will primarily use:

- Python
- PostgreSQL
- SQL
- REST APIs
- dbt
- Git and GitHub
- GitHub Actions
- Microsoft Azure Data Factory
- Azure Storage

5. Programme Structure and Weekly Curriculum

The programme begins with one shared professional core week, followed by 11 weeks of Data Engineering specialisation, applied practice and capstone development.

WEEK 1 — SHARED PROFESSIONAL CORE

Focus: Professional project practices and engineering foundations.

Topics:

- Understanding professional data projects.
- Translating business requirements into technical objectives.
- Data source identification and assessment.
- Introduction to data quality and reliability.
- Git and GitHub fundamentals.
- Project structure and documentation.
- Reproducible development practices.
- Responsible handling of data and credentials.

WEEK 2 — ADVANCED PYTHON FOR DATA WORKFLOWS

Focus: Developing maintainable Python applications for data processing.

Topics:

- Modular Python code.
- Functions and classes where appropriate.
- Virtual environments and dependency management.
- Working with files and common data formats.
- Configuration management.
- Exception handling.
- Reusable scripts.
- Organising Python projects and modules.

WEEK 3 — APIs AND DATA INGESTION

Focus: Collecting data from external sources.

Topics:

- HTTP and REST API concepts.
- Requests and responses.
- Working with JSON data.

- API authentication concepts.
- Pagination.
- Rate limits and request management.
- Extracting data from APIs.
- Handling inconsistent source data and failed requests.

WEEK 4 — POSTGRESQL AND DATA MODELLING

Focus: Designing and managing relational data structures.

Topics:

- Relational database design.
- Primary and foreign keys.
- Constraints and data integrity.
- Normalisation concepts.
- Loading files and API data into PostgreSQL.
- SQL joins and aggregations.
- Transaction awareness.
- Designing tables for downstream use.

WEEK 5 — ETL/ELT PIPELINE DESIGN

Focus: Understanding and designing data workflows.

Topics:

- Introduction to ETL and ELT.
- Pipeline architecture and stages.
- Batch data ingestion.
- Staging and target layers.
- Incremental versus full data loads.
- Transformation strategies.
- Pipeline repeatability.
- Maintainability and design considerations.

WEEK 6 — DATA TRANSFORMATION WITH PYTHON AND SQL

Focus: Transforming raw data into reliable datasets.

Topics:

- Data cleaning and standardisation.
- Type conversion.
- Duplicate detection and removal.
- Joins and aggregations.
- Data transformation using Python.
- SQL-based transformations.
- Separating raw, staging and curated data.
- Structuring reusable transformation logic.

WEEK 7 — DBT FUNDAMENTALS

Focus: Managing SQL transformations systematically.

Topics:

- Introduction to dbt.
- dbt project structure.
- Sources and models.
- Model dependencies.
- Data testing.
- Documentation.

- Modular SQL transformations.
- Running and reviewing a dbt project.

WEEK 8 — RELIABILITY, TESTING AND DATA QUALITY

Focus: Building dependable pipelines.

Topics:

- Schema validation.
- Business-rule validation.
- Data quality checks.
- Logging and error reporting.
- Retry strategies.
- Idempotency.
- Pipeline failure handling.
- Audit fields and execution records.
- Monitoring concepts.
- Diagnosing failed pipeline runs.

WEEK 9 — WORKFLOW AUTOMATION AND ORCHESTRATION

Focus: Automating and coordinating data workflows.

Topics:

- Scheduling pipeline execution.
- Task dependencies.
- Parameterised workflows.
- Environment configuration.
- Basic automation using GitHub Actions.
- Introduction to orchestration concepts.
- Understanding when orchestration is necessary.
- Managing repeatable pipeline execution.

WEEK 10 — AZURE DATA FACTORY AND CLOUD STORAGE

Focus: Introduction to cloud-based data integration.

Topics:

- Introduction to Microsoft Azure.
- Azure resource and storage concepts.
- Introduction to Azure Data Factory.
- Linked services and datasets.
- Pipelines and activities.
- Copying data into cloud storage.
- Parameters and pipeline configuration.
- Monitoring pipeline execution.
- Connecting cloud ingestion to the wider data workflow.

WEEK 11 — CAPSTONE BUILD AND INTEGRATION

Focus: Developing an end-to-end data engineering solution.

Activities:

- Selecting a suitable file or API data source.
- Designing the pipeline architecture.
- Extracting and loading source data.
- Loading data into PostgreSQL and/or Azure Storage.
- Developing transformation workflows.
- Implementing data quality checks.

- Adding logging and failure handling.
- Automating pipeline execution.

WEEK 12 — OPERATIONALISATION AND CAPSTONE DEFENCE

Focus: Demonstrating engineering competence.

Activities:

- Demonstrating pipeline execution.
- Explaining the architecture and design decisions.
- Showing how failures are detected and handled.
- Discussing data quality and reliability.
- Presenting technical documentation.
- Explaining limitations and trade-offs.
- Identifying possible improvements.

6. Capstone Project

Learners will develop an end-to-end data pipeline demonstrating their ability to collect, store, transform and validate data.

The expected deliverables include:

1. A documented data source and pipeline objective.
2. A Python-based extraction workflow using a file or API.
3. A PostgreSQL database and/or Azure Storage destination.
4. Data transformation using Python and/or SQL.
5. Clean analytical tables or structured outputs.
6. Data quality checks and validation.
7. Execution logs and basic failure handling.
8. Repeatable or scheduled pipeline execution.
9. Technical documentation explaining the architecture.

The project should demonstrate not only that the pipeline works, but also how it can be maintained, tested and operated.

7. Assessment Approach

Learners will be assessed through:

- Python and SQL implementation exercises.
- API ingestion and database-loading tasks.
- Data transformation checkpoints.
- dbt implementation exercises.
- Data quality and reliability tasks.
- Pipeline automation exercises.
- Capstone project assessment.

The capstone assessment will consider architecture, reliability, correctness, automation, documentation and the learner's ability to explain operational trade-offs.

8. Learning Model

The programme follows three complementary stages:

LEARN

Instructor-led explanations, demonstrations and guided engineering practice.

BUILD

Practical weekly exercises that progressively develop data pipeline and workflow engineering skills.

APPLY

An end-to-end engineering project, review and presentation demonstrating the learner's ability to build and explain a data pipeline.

9. Cloud Technology and Cost Considerations

The programme introduces Microsoft Azure Data Factory and Azure Storage through guided practical exercises.

Azure resources may incur charges depending on usage, region and service configuration.

Where available, learners may be guided to use eligible trial credits or learning allowances.

Practical sessions will include guidance on resource configuration, cost controls and deleting resources after use.

Access to a free account or trial credit does not mean that every Azure service or resource is free.

10. Additional Information

- Live instruction is delivered over three 2-hour sessions each week.
- Learners should plan for approximately 4–6 additional hours of independent practice weekly.
- Learners should have access to a suitable computer and development environment.
- Public, synthetic or appropriately authorised data should be used for practical projects.
- Programme tools and cloud service availability may change; equivalent learning environments may be used where necessary.
- Programme details are subject to the final cohort calendar and published enrolment terms.

SODBRIDGE ACADEMY

Real-world learning through practical skills, projects and professional development.