

# Northcoders



## Upskill in

## Data Engineering

## Fundamentals

## Curriculum & Prerequisites



# What's Covered

Learn to orchestrate tasks as you would in a real production environment: scheduling jobs, implementing observability, and performing optimisations.

This module bridges theory and practice, preparing you to work with enterprise-scale data pipelines and analytics platforms.

## Key Topics:

- Orchestration
- Pandas (Python Data Analysis Library)
- Query Optimisation
- Data Warehouse Design
- Data Architecture
- AWS Athena & Glue
- Apache Spark
- Databricks
- ELT (Extract, Load, Transform)



# Prerequisite Skills ✓

This course is designed to build on existing knowledge.

To join, you'll need to know:

## **Programming & Development**

- Proficiency in Python
- SQL fundamentals (SELECT, WHERE, JOINS, GROUP BY, aggregates)
- Comfort working in a terminal/command line environment

## **Data and Database Concepts**

- Understanding of relational databases and basic database design
- Knowledge of data types, schemas, and table relationships
- Experience writing SQL queries (SELECT, JOIN, WHERE clauses)
- Basic data formats (CSV, JSON, Parquet)



## **Cloud & Infrastructure**

- Cloud platform awareness (general cloud computing concepts)
- Basic understanding of distributed computing

## **Development Environment**

- Python 3.x installed and configured
- PostgreSQL server installed and configured
- IDE setup

## **Also Nice to Have:**

- Version control basics (Git)
- Understanding of API concepts and data retrieval
- Exposure to workflow automation or scheduling tools
- Familiarity with Jupyter notebooks or interactive Python environments

