Python Data Engineering Academy: Spark, Databricks & Azure

A 288-hour academy for data engineers: Python, PySpark, Databricks, streaming, governance and orchestration on Azure.

DURATION	FORMAT	LEVEL	LABS
288 hours	Onsite, online or hybrid	Intermediate	Cloud sandbox included

About this programme

An immersive academy that takes graduate and early-career cohorts from Python fundamentals to a governed lakehouse in production. Engineers build reliable batch, incremental and streaming pipelines with PySpark and Databricks, design medallion architectures with data contracts and lineage, and orchestrate and monitor them on Azure. GitHub Copilot and agentic AI are used in every module, including a final block on building data agents with human checkpoints. Weekly assessments and a rubric-scored capstone give L&D a clear readiness view.

Who it is for: Graduate and early-career data cohorts, and developers moving into data engineering

What your team will be able to do

- Write clean, tested and packaged Python, and build secure FastAPI services
- Build reliable batch and incremental ETL and ELT pipelines with Pandas and PySpark
- Tune Spark on Databricks and Delta Lake, and build streaming pipelines that handle late and duplicate data
- Design medallion lakehouse architectures with data contracts, lineage and Unity Catalog governance
- Orchestrate and monitor pipelines with Airflow, Azure Data Factory and Kafka
- Use GitHub Copilot and agents responsibly across data engineering work

Before you start

- Basic programming in any language
- School-level mathematics; SQL helps but is taught from the start

Course outline

We send the module-by-module outline with your proposal, adapted to your team's tools and goals.

Assessment and certification

How learning is checked

- A short knowledge check on day one and again at the end, so you can see each person's progress
- Graded hands-on labs in every module, reviewed by the trainer
- A final exercise that applies the programme to a problem like your team's own

Certificate and exams

- Participants who complete the programme receive a Cognitive Networks AI certificate of completion.

Formats and cohort size

- Onsite: At your office or a venue you choose, in full days.
- Online, live: Instructor-led over video, usually in half-day sessions so people can keep up with work.
- Hybrid: Some days in a room together, the rest online. Useful when a team works across cities.

Cohort size: 15 to 25 people per trainer. Larger teams run as parallel cohorts on the same outline, so everyone learns the same thing.

Pricing: We quote per cohort, based on duration, format and number of cohorts. There is no per-seat price list: tell us your team and we send a proposal within one business day.

Bring this programme to your team

Request a proposal at cognitivenetworks.ai/contact?programme=python-data-engineering-academy

reachus@cognitivenetworks.ai · Bangalore +91 72043 05303 · Dallas +1 (303) 827-7415