Azure Data Engineer (DP-203)

Design and build secure, optimised data storage and pipelines on Azure, per the DP-203 skills.

DURATION	FORMAT	LEVEL	LABS
45 hours	Onsite, online or hybrid	Advanced	Cloud sandbox included

About this programme

Covers the data engineering skills in Microsoft's DP-203 outline: designing Data Lake storage and serving layers, building batch and streaming pipelines with Data Factory, Synapse, Databricks and Stream Analytics, and securing, monitoring and tuning them in production.

Who it is for: Data engineers and ETL developers working on Azure

What your team will be able to do

- Design partitioned, cost-effective storage on Azure Data Lake Storage Gen2
- Model star schemas and serving layers for analytics
- Build batch and streaming pipelines with Data Factory, Synapse and Databricks
- Secure data with encryption, RBAC, row- and column-level security and masking
- Monitor, troubleshoot and tune pipelines for performance and cost

Before you start

- Working SQL and Python
- Experience building data pipelines
- Azure fundamentals (AZ-900 level)

Course outline

1. Design and implement data storage

- Azure Data Lake Storage Gen2: hierarchical namespace, ACLs and SAS
- Partitioning strategies for files and analytical workloads
- File types, compression, sharding and folder structures
- Data lifecycle management and archiving

2. Design the serving layer

- Star schemas and dimensional hierarchies
- Data models for reporting and operational dashboards
- Designing for efficient querying: pruning and distribution

3. Develop batch processing

- Azure Data Factory: linked services, datasets, triggers and pipelines
- Synapse pipelines and mapping data flows
- Apache Spark in Azure Synapse and Azure Databricks
- Transformations with T-SQL and Spark
- Handling, validating and scheduling batch loads

4. Develop stream processing

- Real-time processing with Azure Stream Analytics
- HTAP with Azure Synapse Link for Cosmos DB and SQL
- Ingesting batch and streaming data at volume

5. Secure data

- Encryption at rest and in transit
- Azure RBAC, row-level and column-level security
- Firewalls, private endpoints and network security groups
- Auditing, data masking and regulatory compliance

6. Monitor and optimise

- Logging and monitoring with Azure Monitor and Log Analytics
- Query and Spark tuning: indexes, caching, skew and spill
- Troubleshooting failed Spark jobs and pipeline runs
- Optimising storage, compute and network costs

Sample schedule

6 days onsite, or 12 half-day sessions online. We fit days and session lengths to your team's calendar.

Day 1	8 h	Design and implement data storage
Day 2	8 h	Design the serving layer; Develop batch processing
Day 3	8 h	Develop batch processing (continued); Develop stream processing
Day 4	8 h	Develop stream processing (continued); Secure data
Day 5	8 h	Secure data (continued); Monitor and optimise
Day 6	5 h	Monitor and optimise (continued)

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
- A timed practice exam in the format of the real one, with a review of every answer

Certificate and exams

- Participants who complete the programme receive a Cognitive Networks AI certificate of completion.
- Prepares your team for the Microsoft DP-203 exam. Exam vouchers are not included: each participant books the exam with Microsoft.

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=azure-data-engineer

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

Microsoft, Azure and the exam names are trademarks of Microsoft. Cognitive Networks AI is not affiliated with Microsoft. This programme prepares your team for the exam. It is not an official course of Microsoft, and the exam is booked separately.