

ENTERPRISE PROGRAM BRIEF

PySpark for Data Engineering

Use PySpark to build reliable notebook transformations.

VNode ITes

B01

DURATION

24 hours

LEVEL

Bridge

FORMAT

Private cohort

AUDIENCE PROFILE

Built for these roles

SQL-oriented data engineers

OVERVIEW

Executive overview

Use PySpark to build reliable notebook transformations. Proposed instructor-led duration; delivery subject to environment and lab validation.

READINESS

Prerequisites

- SQL and data engineering fundamentals
- SQL and Data Engineering Foundations, or equivalent demonstrable experience

PROGRAM OUTCOMES

Capabilities your teams will gain

- Deliver a parameterised PySpark transformation notebook
- Match reference SQL and verify duplicates, late records and replay
- Quarantine invalid data and protect configured connections
- Explain execution, quality and performance decisions

ENTERPRISE CUSTOMIZATION

Enterprise customization

Tailor this program to your organization's priorities: Use PySpark to build reliable notebook transformations.

- Private cohort scheduling
- Equivalent-experience entry review

CURRICULUM

Curriculum roadmap

1

MODULE 1 · 6 HOURS

Python-to-Spark bridge

Read and inspect typed source data in a notebook.

- Python syntax essentials
- Spark execution
- DataFrames
- Explicit schemas

2

MODULE 2 · 6 HOURS

Transformations and joins

Translate analytical SQL transformations into PySpark.

- Expressions
- Joins
- Aggregations
- Window functions

3

MODULE 3 · 6 HOURS

Incremental writes and quality

Implement an incremental write and test duplicate and malformed inputs.

- Delta writes
- Merge concepts
- Partition choices
- Quality and reruns

4

MODULE 4 · 6 HOURS

Notebook engineering assessment

Deliver a parameterised notebook with reconciliation and failure reporting.

- Parameters
- Functions
- Error handling
- Performance basics

PREPARATION

Environment prerequisites

Fabric capacity and isolated workspace Author and consumer personas Source files and notebook runtime Data-agent labs: paid F2+ or qualifying P1+ capacity, tenant AI settings and source access as documented; cohort capacity sized separately Delivery depends on environment and lab validation. This outline does not include step-by-step lab instructions.

CAPSTONE

Capstone

Deliver a parameterised notebook with reconciliation and failure reporting. Assessed transformation notebook Evidence of relevant access, quality and risk controls Demonstration and limitations record

ASSESSMENT

Assessment

Scenario-based practical assessment Deliver a parameterised PySpark transformation notebook 30% Match reference SQL and verify duplicates, late records and replay 25% Quarantine invalid data and protect configured connections 25% Explain execution, quality and performance decisions 20% Proposed VNode outline-level assessment; no vendor accreditation or completed lab validation implied.

DELIVERABLES

Participant deliverables

- Assessed transformation notebook
- Evidence of relevant access, quality and risk controls
- Demonstration and limitations record

SCOPE

Scope boundaries

Step-by-step lab instructions; Runnable solution code; Packaged source datasets; Executed environment validation

DELIVERY

Delivery options

Count teaching hours, not calendar days. Exclude breaks. Split into 4-hour or 8-hour teaching sessions; schedule complex project work in contiguous blocks. Proposed hours exclude breaks and are subject to pilot validation.

PRIVATE DELIVERY

Plan a private cohort for your team

Scope delivery format, role mix, and rollout timeline with our solutions team.

[Plan Private Delivery](#)
[Book advisory call](#)