

# CloudPandith

## Azure Databricks for Beginners | PySpark, Unity Catalog & ETL

Real-Time Projects | Interview Ready Training

[www.cloudpandith.com](http://www.cloudpandith.com) | 8904424822 | [cloudpandith@gmail.com](mailto:cloudpandith@gmail.com)

### ● Section 1: Introduction to Azure Databricks & Setup

What you learn: Platform basics, workspace, clusters, compute setup

1. Introduction to Azure Databricks
2. Databricks Workspace, Clusters, Compute & Free Edition Setup

### ● Section 2: Notebooks & Execution Framework

What you learn: Notebook execution, job orchestration, parameterization

3. Notebook Run Options (Run All, Jobs, Interactive vs Job Clusters)
4. Parameterized Notebooks using dbutils Widgets
5. %run vs dbutils.notebook.run (Interview Scenarios, Jobs vs Manual Run)

### ● Section 3: Security & Secrets Management

What you learn: Secure credential handling using Azure services

6. Azure Key Vault Integration with Databricks (Secrets Management)

### ● Section 4: Unity Catalog & Data Governance

What you learn: Data governance, ADLS access, centralized metadata control

7. Unity Catalog Hands-On (ADLS Gen2 Access using Managed Identity)
8. Unity Catalog Classroom Practice (ADLS Access & Governance)
9. Migration to Unity Catalog (Governance & Medallion Architecture)

[www.cloudpandith.com](http://www.cloudpandith.com) | 8904424822 | [cloudpandith@gmail.com](mailto:cloudpandith@gmail.com)

## 🌀 Section 5: Python for Data Engineering

What you learn: Python fundamentals for PySpark & ETL development

10. Python Fundamentals for PySpark
11. Python Strings for Data Engineering (ETL Use Cases)
12. List vs Tuple in Data Engineering
13. Set & Dictionary for Data Engineering
14. Python Functions, Lambda, Map, Filter, Reduce
15. Modules and Packages in Python

## 🌀 Section 6: Spark Architecture & Internals

What you learn: Spark execution model, DAG, partitions, performance flow

16. Spark Architecture & Job Execution Flow
17. Spark Architecture (DAG, Partitions, Parallelism, Spark UI)

## 🌀 Section 7: PySpark Core Concepts

What you learn: Transformations, actions, lazy evaluation

18. PySpark Transformations and Actions (Lazy Evaluation)
19. Lazy Evaluation Deep Dive (Transformations vs Actions, Narrow vs Wide)

## 🌀 Section 8: DataFrame Operations & Data Processing

What you learn: Data ingestion, transformations, data cleaning

20. Read CSV Files in Databricks (Schema, Delimiters, Metadata Handling)
21. DataFrame Transformations (select, filter, distinct, orderBy)
22. Advanced Transformations (Null Handling, Aggregations, Union vs UnionByName)

## 🌀 Section 9: Joins & Data Integration

What you learn: Combining datasets and integrating external systems

23. PySpark Joins (Inner, Left, Right, Full, Semi, Anti Joins)

24. Read & Write Azure SQL Database using PySpark JDBC

## 🌀 **Section 10: Cluster Management & Performance Optimization**

What you learn: Cluster tuning, policies, scheduling

25. Databricks Cluster Policies (Runtime, Auto Scaling, Auto Termination, Scheduling)

## 🌀 **Section 11: End-to-End ETL Pipeline Project**

What you learn: Real-world data pipeline design and execution

26. Azure SQL to ADLS Gen2 and ADLS to SQL ETL Project (PySpark End-to-End Pipeline)

## 🌀 **Section 12: Advanced PySpark Concepts**

What you learn: Custom logic, UDFs, interview-level concepts

27. PySpark UDF Explained (Real-Time Examples + Interview Questions)