



Master Cloud Big Data Engineering

Transform Your Career in 4 Months

Comprehensive training in Hadoop, Python, PySpark, and Azure. Learn industry-standard tools and real-world projects. Swipe to explore the complete curriculum.

Three-Phase Learning Journey

01

Phase 1: Foundation

Master Hadoop ecosystem, Linux commands, and Apache Hive for big data processing fundamentals.

02

Phase 2: Programming

Deep dive into Python and PySpark for advanced data engineering and transformation workflows.

03

Phase 3: Cloud Mastery

Azure Data Lake, Databricks, Synapse Analytics, and Delta Lake for enterprise-scale solutions.



Phase 1: Big Data Foundations

Hadoop Ecosystem

Architecture, YARN, MapReduce, and Hadoop 2.0 components for distributed data processing.

Linux Mastery

Essential commands, shell scripting, parameterization, and Hadoop command-line operations.

Apache Hive

HQL, partitioning, bucketing, joins, optimization, and ACID properties for data warehousing.



Hive Deep Dive

Table Management

- Internal and external tables
- Static and dynamic partitions
- Bucketing strategies
- Performance optimization

Advanced Features

- SCD Type 1 implementation
- Avro, Parquet, ORC formats
- XML and JSON processing
- Small file handling

Phase 2: Python & PySpark

Python Fundamentals

Data types, collections, functions, classes, inheritance, and error handling for data engineering.

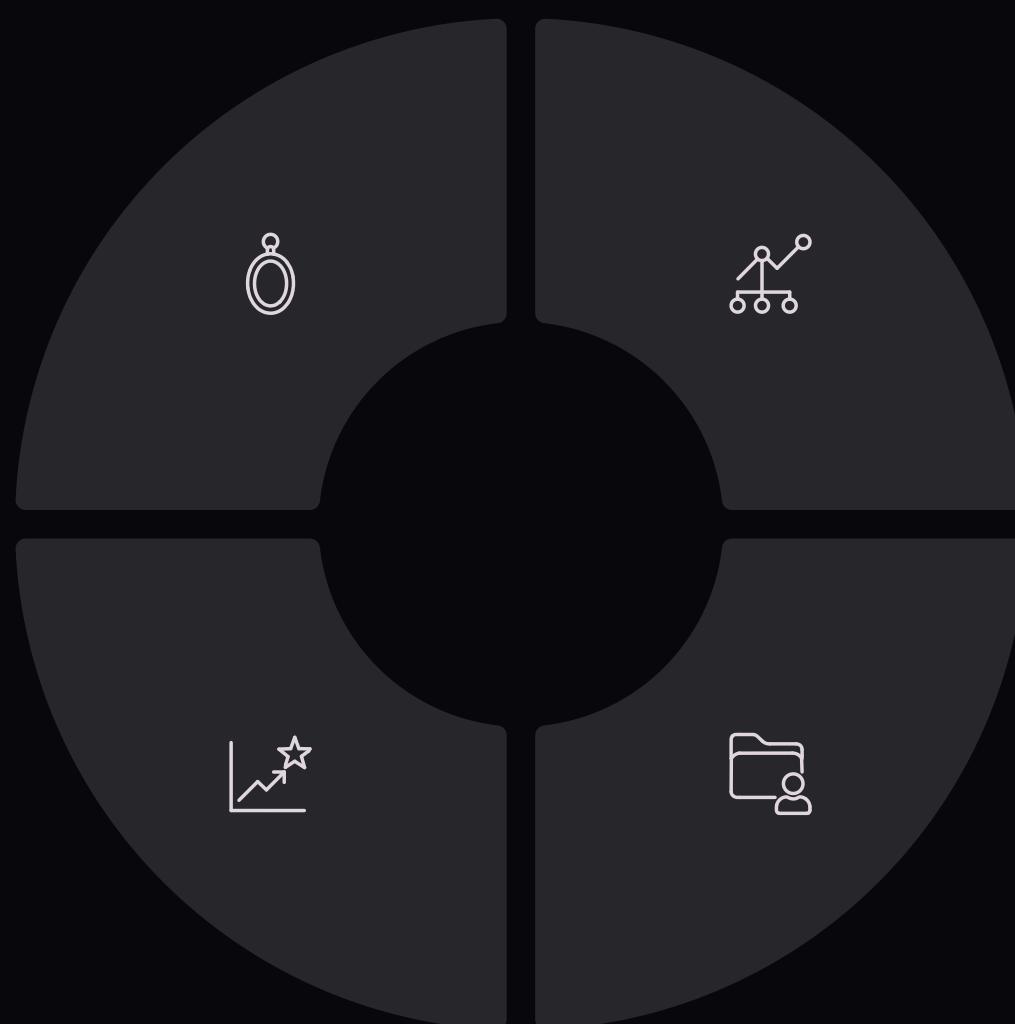
Spark Core

RDD operations, transformations, actions, DAG, lazy evaluation, and memory management techniques.

Spark SQL Mastery

DataFrames, datasets, joins, aggregations, window functions, UDFs, and performance tuning.

Spark Architecture Essentials



Memory Management

Caching, serialization, Tungsten optimization, and dynamic resource allocation.



Execution Modes

Client, cluster, standalone modes with broadcast variables and accumulators.



Data Operations

Repartition, coalesce, partition strategies, and shuffle optimization.



Performance

Broadcast joins, executor tuning, null handling, and memory exception resolution.

Phase 3: Azure Cloud Platform



Azure Data Lake Gen2

BLOB service, redundancy, access tiers, Storage Explorer, and mount points.



Azure Databricks

Clusters, notebooks, DBUtils, CLI deployments, and pipeline creation.



Synapse Analytics

Dedicated and serverless pools, pipelines, Spark analysis, and SQL operations.



Delta Lake

Version control, schema evolution, ACID transactions, and Unity catalog.



Real-World Project Experience



On-Premise Architecture

End-to-end pipeline with Sqoop, Hive, Spark, HBase, and RDBMS integration.



Azure Cloud Projects

Multi-cloud platform use cases with production-ready architecture patterns.

Career Support

Resume preparation, mock interviews, interview prep sessions, and weekly practice sessions.

Start Your Journey

4

Months

Complete training duration

30

Students

Limited batch enrollment

3

Phases

Structured learning path

Demo: November 2, 2025 at 9 AM IST

Start Date: November 8, 2025

Schedule: Weekends, 8:15 AM - 10:45 AM IST

Only