

Modern Data Engineering with Databricks, Spark & Snowflake

Engineer modern data platforms with Spark, Databricks and Snowflake across batch, streaming, lakehouse, warehouse and governance.

**8 weeks**

DURATION

92 hours

GUIDED LEARNING

2 projects

PORTFOLIO



ABOUT THE ACADEMY

Practical learning. Work you can demonstrate.

Brightnest AI Academy brings live instruction, guided practice and career preparation together for individuals and teams.

23

FOCUSED COURSES

6

FLAGSHIP PROGRAMMES

Noida

ACADEMY LOCATION

Our approach

1

Learn

Understand concepts through live instruction and current examples.

2

Build

Apply learning through labs, assignments and guided projects.

3

Explain

Document decisions, results, trade-offs and limitations.

4

Prepare

Connect completed work with CVs, interviews and workplace conversations.

PEOPLE BEHIND THE ACADEMY

Meet Ranjeet Kumar



ACADEMY ADVISER

Ranjeet Kumar

Innovation & Growth Leader

15+ years

Experience in data, artificial intelligence and machine learning leadership.

Areas of expertise

- Data science and machine learning solution leadership
- AI product strategy and scalable data platforms
- Technology team leadership and mentorship
- Practical AI and data adoption for business transformation

Education

Master of Information Systems, Anna University, Chennai
Bachelor of Computer Science Engineering

COURSE OVERVIEW

Build your next set of skills

Design end-to-end data platforms spanning batch and streaming pipelines, Spark, Databricks, lakehouse architecture, Snowflake, cloud warehousing, orchestration and governance. Progress from Modern Data Platform Architecture to Modern Data Engineering Capstone through guided labs, assessed projects, and portfolio evidence.

Who this is for

SQL/Python practitioners moving into enterprise data pipelines, lakehouse and warehouse engineering.

Before you begin

Learners should understand SQL and Python fundamentals. Basic data modelling or analytics experience is recommended.

Modern data engineering requires dependable ingestion, transformation, orchestration, quality, streaming, warehousing, and governance across the full platform lifecycle.



LEARNING OUTCOMES

What you will be able to do

01

Design a target architecture for a multi-source analytics platform with SLAs and data quality expectations.

02

Build distributed transformations on a multi-file dataset and inspect execution behaviour.

03

Build bronze-silver-gold Delta tables with data quality checks.

04

Create an incremental pipeline with retries, checkpoints and data quality gates.

05

Build a streaming pipeline that aggregates events and writes curated tables.

06

Tune one Spark and one Snowflake workload, quantify performance/cost changes and document trade-offs.

07

Build an end-to-end enterprise data platform using Databricks/Spark and Snowflake with curated BI-ready outputs.

CURRICULUM

Course curriculum

9 modules · 8 weeks · 92 hours of guided learning

01 Modern Data Platform Architecture

6 hours

TOPICS

OLTP versus analytics · Warehouse/lake/lakehouse · Medallion architecture · Batch versus streaming · Data contracts · Governance · Platform tradeoffs

PRACTICAL WORK

Design a target architecture for a multi-source analytics platform with SLAs and data quality expectations.

Tools: Architecture diagrams, Cloud object storage concepts

Assessment: Architecture case

Portfolio evidence: Modern data platform blueprint

02 Apache Spark Engineering Foundations

10 hours

TOPICS

Spark architecture · DataFrames · Transformations/actions · Partitions · Joins · Shuffle · Spark SQL · Performance intuition · Structured data processing

PRACTICAL WORK

Build distributed transformations on a multi-file dataset and inspect execution behaviour.

Tools: Apache Spark, PySpark, Spark SQL

Assessment: Spark lab

Portfolio evidence: PySpark data pipeline



CURRICULUM

Course curriculum

9 modules · 8 weeks · 92 hours of guided learning

03 Databricks Lakehouse and Delta Engineering

12 hours

TOPICS

Workspaces · Notebooks/repos · Delta tables · ACID · Schema enforcement/evolution · Medallion layers · Catalog/governance concepts · Jobs

PRACTICAL WORK

Build bronze-silver-gold Delta tables with data quality checks.

Tools: Databricks, Delta Lake, Unity Catalog concepts

Assessment: Lakehouse lab

Portfolio evidence: Databricks medallion pipeline

04 Batch ETL/ELT and Pipeline Orchestration

10 hours

TOPICS

Incremental ingestion · CDC concepts · Idempotency · Scheduling · Dependencies · Retries · Parameters · Testing · Pipeline observability

PRACTICAL WORK

Create an incremental pipeline with retries, checkpoints and data quality gates.

Tools: Databricks workflows/Lakeflow concepts, Airflow optional, dbt optional

Assessment: Pipeline review

Portfolio evidence: Production-style batch pipeline



CURRICULUM

Course curriculum

9 modules · 8 weeks · 92 hours of guided learning

05 Streaming Data Engineering

10 hours

TOPICS

Event streaming concepts · Windows · Watermarks · State · Late data · Exactly-once concepts · Stream-batch unification · Real-time KPIs

PRACTICAL WORK

Build a streaming pipeline that aggregates events and writes curated tables.

Tools: Spark Structured Streaming, Kafka concepts

Assessment: Streaming lab

Portfolio evidence: Real-time pipeline demo

06 Snowflake Data Warehousing Engineering

10 hours

TOPICS

Snowflake architecture · Databases/schemas · Warehouses · Stages · File loading · Streams/tasks concepts · Snowpark awareness · Security · Cost/performance

PRACTICAL WORK

Load, transform and optimise a warehouse workload in Snowflake with role-based access.

Tools: Snowflake, SQL, Snowpark concepts

Assessment: Warehouse lab

Portfolio evidence: Snowflake analytics pipeline



CURRICULUM

Course curriculum

9 modules · 8 weeks · 92 hours of guided learning

07 Analytics Engineering, Quality and Governance

8 hours

TOPICS

Dimensional models · Transformation layers · dbt-style testing/documentation · Lineage · Metadata · Data quality SLAs · Access controls · Cataloging

PRACTICAL WORK

Create tested curated models with documentation, lineage and ownership rules.

Tools: SQL, dbt concepts, Data catalog/governance tooling

Assessment: Quality checkpoint

Portfolio evidence: Governed analytics layer

08 Performance, Cost and Reliability Engineering

8 hours

TOPICS

Partitioning/clustering · File sizing · Caching · Query plans · Autoscaling · Warehouse sizing · Spark tuning · SLA/SLOs · Observability · Cost controls

PRACTICAL WORK

Tune one Spark and one Snowflake workload, quantify performance/cost changes and document trade-offs.

Tools: Databricks/Spark, Snowflake monitoring

Assessment: Optimisation report

Portfolio evidence: Performance and cost benchmark

CURRICULUM

Course curriculum

9 modules · 8 weeks · 92 hours of guided learning

09 Modern Data Engineering Capstone

18 hours

TOPICS

Ingestion · Batch/stream processing · Lakehouse · Warehouse · Orchestration · Quality · Governance · Monitoring · Documentation

PRACTICAL WORK

Build an end-to-end enterprise data platform using Databricks/Spark and Snowflake with curated BI-ready outputs.

Tools: Databricks, Spark, Snowflake, SQL, Git

Assessment: Capstone architecture and demo

Portfolio evidence: Enterprise data engineering portfolio project

APPLIED PORTFOLIO

Projects you will build

Two portfolio projects supported by guided module work.

PORTFOLIO PROJECT 1

Lakehouse-to-Warehouse Data Platform

Ingest raw data, process with Spark/Databricks, curate in Snowflake and serve analytics.

DELIVERABLES

Pipeline code · Architecture · Quality tests · Lineage · Performance report

PORTFOLIO PROJECT 2

Real-Time Customer Event Platform

Process streaming events into curated real-time and historical tables.

DELIVERABLES

Streaming pipeline · Medallion tables · Monitoring · BI-ready outputs

LEARNING EXPERIENCE

Learn, practise and present

- Live instructor-led sessions that connect concepts to real workplace decisions.
- Guided labs and workshops in every module.
- Assignments, checkpoints and practical feedback.
- Portfolio documentation, demonstrations and capstone review.
- Access to recordings and LMS resources according to the published batch policy.
- Career preparation based on completed work and target roles.

Tools you will work with

Architecture diagrams · Cloud object storage concepts · Apache Spark · PySpark · Spark SQL · Databricks · Delta Lake · Unity Catalog concepts · Databricks workflows · Lakeflow concepts · Airflow · dbt · Spark Structured Streaming · Kafka concepts · Snowflake · SQL

CAREER PREPARATION

Build confidence for interviews

A structured process to connect your learning with your next career step.

01**Role clarity**

Identify roles that fit your background, interests and current skills.

02**Profile readiness**

Review your CV, LinkedIn, GitHub and project summaries.

03**Project proof**

Prepare to demonstrate completed work and explain your decisions.

04**Interview practice**

Practise technical questions, project walkthroughs and interviews.

05**Opportunity updates**

Receive relevant job openings and employer requirements, subject to availability and learner readiness.

Career support helps you prepare. Employment outcomes depend on experience, project quality, interview readiness and market conditions.

LEARNER JOURNEYS

Companies where our learners have been hired

Across the Brightnest AI Academy learner community



Learner employment history across the academy; not a placement guarantee for this course. Names and logos identify employers; they do not imply endorsement or a hiring partnership.

BEFORE YOU ENROL

Your questions, answered

Is the Modern Data Engineering course suitable for beginners?

This course progresses from intermediate to advanced level. Learners should understand SQL and Python fundamentals. Basic data modelling or analytics experience is recommended.

What will I build during the course?

You will complete guided labs in every module and build two portfolio projects: Lakehouse-to-Warehouse Data Platform and Real-Time Customer Event Platform. Deliverables include working files or code, documentation, testing or evaluation evidence, and a final presentation.

Which tools and platforms are covered?

Key tools include Architecture diagrams, cloud object storage concepts, Apache Spark, PySpark, Spark SQL, Databricks, Delta Lake, and Unity Catalog concepts. Additional platforms are introduced in relevant modules through practical tasks, and the toolset may evolve as industry practice changes.

How long does the course take?

The course includes approximately 92 guided learning hours across 9 modules, delivered over 8 weeks.

Which career paths can this course support?

The curriculum supports the development of skills relevant to roles such as Data Engineer, Analytics Engineer, Databricks/Spark Engineer, and Cloud Data Engineer. Career outcomes depend on prior experience, project quality, interview readiness and market conditions; employment is not guaranteed.

BEFORE YOU ENROL

Your questions, continued

Will I receive mentor and career support?

The course includes live instruction, lab support, assignment feedback, project reviews and career preparation covering portfolio development, CV writing, LinkedIn profile improvement, and interview guidance.

FEES & ADMISSIONS

Plan your next step

COURSE FEE

INR 35,000 + GST

EMI and easy instalments available

Duration	8 weeks
Level	Intermediate-Advanced
Learning format	Live mentor-led learning
Batch starts	Expected Saturday / Sunday starts

Check the live Batches page for the next expected weekend. Confirm the exact date, class times and availability with admissions.

Book a free demo class >

+91 73030 11395

brightnestaiacademy.com/batches

A-56, Sector-64, Noida, Uttar Pradesh - 201301