



2026 EDITION

Data Engineering *Training* Programs

Equipping data teams to build reliable pipelines, platforms and architectures

INSTRUCTOR-LED

ENTERPRISE READY

CUSTOMIZABLE

GLOBAL DELIVERY

37

COURSES

10

CATEGORIES

5,000+

EXPERT TRAINERS

Building *reliable* data foundations

Edstellar is an enterprise learning partner that helps organizations build future-ready workforces. Our 2,000+ instructor-led programs are delivered onsite or live online by 5,000+ practitioner trainers, and each one is tailored to your goals. This catalog brings together 37 data engineering programs that span spark, lakehouse and big data, orchestration and dataops, enterprise etl platforms, data architecture and modelling, and cloud data platforms.

Global leaders including Visa, ABB and Aditya Birla Group trust Edstellar to build capability at enterprise scale. Before any program starts, we look at the real capability gap and the business goal behind it, so every module ties back to outcomes your leaders can measure, not just attendance. The result is training your data teams put to work straight away.

Learning paths for *every* role



Data Architects & Leads

ARCHITECTURE & MODELLING

Design data architectures, mesh and models that scale across the organization.

START WITH

- Modern Data Architecture & Data Mesh
- Data Mesh & Streaming Architecture
- Dimensional Data Modeling



Big Data & Spark Engineers

LAKEHOUSE & BIG DATA

Build and run large-scale processing on Spark, Databricks and lakehouse engines.

START WITH

- Data Engineering with Apache Spark & Databricks
- PySpark
- Apache Iceberg for Data Lakehouse Engineering



Pipeline & ETL Developers

ORCHESTRATION & ETL

Orchestrate pipelines and integrate data across enterprise ETL platforms.

START WITH

- Apache Airflow
- Informatica (PowerCenter & IDMC)
- Azure Data Factory



Cloud Data Engineers

CLOUD & STREAMING

Build data platforms and streaming workloads on the major clouds.

START WITH

- Data Engineering on AWS (Redshift, Glue & Lake Formation)
- Microsoft Fabric & Azure Data Engineering
- Event Streaming with Apache Kafka

Why *choose* us

01

Practitioner trainers

5,000+ verified data engineers, architects and platform specialists who build and run pipelines in real production environments.

02

Customizable programs

Curriculum built around your data stack, cloud platforms, scale and engineering maturity.

03

Global delivery

Onsite or live virtual in any time zone, with in-person delivery across 1,000+ locations.

04

ISO 9001 & ISO 27001 certified

ISO 9001:2015 for quality and ISO 27001:2022 for information security.

How we *work*

01

Submit your brief

Share your training needs, team size and skill gaps.

02

Custom plan

Receive a tailored curriculum and trainer recommendations.

03

Approve & schedule

Finalize content, dates and delivery mode for your team.

04

Train & measure

Expert-led sessions with progress tracking and ROI reporting.

Course *categories*

01 Spark, Lakehouse & Big Data

8

03 Enterprise ETL Platforms

5

05 Cloud Data Platforms

4

07 Streaming & Event Data

2

09 Data Integration & ETL

1

02 Data Architecture & Modelling

5

04 Orchestration, dbt & DataOps

5

06 Data Quality & Testing

2

08 Data Integration

1

10 More Data Engineering Topics & Tools

4

37 courses in total · full list on the pages that follow

All 37 courses

01 Spark, Lakehouse & Big Data 8

Apache Beam	→
Apache Iceberg for Data Lakehouse Engineering	→
Data Engineering with Apache Spark & Databricks	→
Databricks	→
Delta Lake for Data Lakehouse Engineering	→
Hadoop Administration & Development	→
Hadoop Security	→
PySpark	→

02 Data Architecture & Modelling 5

Data Mesh & Streaming Architecture	→
Data Modelling for Analytics & AI	→
Denodo	→
Dimensional Data Modeling	→
Modern Data Architecture & Data Mesh	→

03 Enterprise ETL Platforms 5

Azure Data Factory	→
IBM DataStage	→
Informatica (PowerCenter & IDMC)	→
Informatica IDMC Cloud Data Integration	→
Intelligent Data Integration	→

04 Orchestration, dbt & DataOps 5

Apache Airflow	→
Apache Hama	→
Apache Hive	→
Apache Pig	→
DataOps & Pipeline Reliability	→

05 Cloud Data Platforms 4

Data Engineering & Analytics on Google Cloud (BigQuery)	→
Data Engineering on AWS (Redshift, Glue & Lake Formation)	→
Google BigQuery	→
Microsoft Fabric & Azure Data Engineering	→

06 Data Quality & Testing 2

Data Contracts & Data Quality Engineering	→
ETL Testing	→

07 Streaming & Event Data 2

Data Pipelines with Shell, Airflow, and Kafka	→
Event Streaming with Apache Kafka	→

08 Data Integration 1

Talend	→
--------	---

09 Data Integration & ETL 1

AWS Glue	→
----------	---

10 More Data Engineering Topics & Tools 4

CloverETL	→
DataLake	→
Fundamentals of Data Warehousing	→
HBase, Impala & Flume	→

Trusted by *enterprises*

“ Our team went from brittle nightly batch jobs to a governed lakehouse with dbt and orchestration that actually holds up. Data quality testing is now part of every pipeline, not an afterthought we scramble to fix.

VP OF DATA PLATFORM · GLOBAL TECHNOLOGY FIRM

“ The Spark and streaming modules cut our processing costs while handling far more event data. Our architects redesigned the whole ingestion layer around what they learned.

DIRECTOR OF DATA ENGINEERING · LEADING FINANCIAL SERVICES GROUP

“ Migrating to cloud data platforms used to worry us. The ETL and data modelling training gave the team a blueprint, and the cutover went off without a single missed SLA.

HEAD OF ANALYTICS ENGINEERING · FORTUNE 500 RETAILER

TRUSTED BY **Visa** **ABB** **Aditya Birla Group** & OTHER FORTUNE 500 ORGANIZATIONS

Let's *connect*

EMAIL contact@edstellar.com

WEB www.edstellar.com

US +1 682 297 4830

UK +44 151 453 4009

INDIA +91 636 483 3830

[Schedule a consultation →](#)



SCAN TO CONTACT US