

SkillCred

Data Engineering Program

Design Data Pipelines. Build Warehouses. Power Analytics at Scale.

8

Weeks

96

Hours

4

Projects

6

Days/Week

HYBRID MODEL

2 Solo + 1 Pair + 1 Group Capstone

Dual Track

8W Standard + 4W Fast Track

Mentor Verified

Every project reviewed

PAT Certified

Score visible to all HRs

8-Week Syllabus

Mon-Sat | 2 Hrs/Day | 48 Sessions | 96 Contact Hours

01 SQL Mastery & Data Modeling

Advanced SQL, CTEs, window functions, normalization, star/snowflake schema

SOLO 1 START

02 Python for Data Engineering

Python OOP, file formats (Parquet, Avro), APIs, Pandas, pytest, CLI tools

SOLO 1 DELIVERY

03 ETL/ELT & Orchestration

ETL vs ELT, Airflow DAGs, operators, Great Expectations, idempotency, CDC

SOLO 2 START

04 Big Data with Spark

Spark RDDs, DataFrames, SparkSQL, PySpark, partitioning, optimization

SOLO 2 DELIVERY

05 Streaming & Real-Time

Kafka producers/consumers, Schema Registry, Spark Structured Streaming

PAIR START

06 Warehousing & dbt

Kimball vs Inmon, dbt transformations, testing, incremental models, BI

PAIR DELIVERY

07 Cloud Data Platforms

AWS S3/Glue/Redshift/Athena, data lake arch, Delta Lake, governance

CAPSTONE START

08 Capstone Pipeline & Demo

E2E pipeline, monitoring, data lineage, performance tuning, demo

CAPSTONE DELIVERY

DAILY 2-HOUR SESSION

0:00-0:10

Recap

0:10-0:40

Concept

0:40-1:30

Build Sprint

1:30-1:50

Standup

1:50-2:00

Wrap

Your 4 Portfolio Projects

2 Solo (individual) + 1 Pair (collaborative) + 1 Group Capstone (team)

SOLO 1 | INDIVIDUAL | Weeks 1-2

Difficulty:

Multi-Source Data Integration Pipeline

Extract from 4+ sources (REST API, CSV, PostgreSQL, web scraping), schema validation, data cleaning (dedup, nulls, type coercion), star schema design with fact/dimension tables, incremental load support, Great Expectations quality report. Handles 100K+ rows, <1% null rate post-cleaning.

Tech: Python, Pandas, SQLAlchemy, PostgreSQL, Requests, BeautifulSoup, Great Expectations | HR Signal: Junior Data Engineer hire bar

SOLO 2 | INDIVIDUAL | Weeks 3-4

Difficulty:

Orchestrated ETL with Airflow & Spark

Airflow DAG with 5+ tasks (extract validate transform load notify), PySpark transformations (joins, aggregations, window functions, UDFs), Great Expectations validation, SLA monitoring, backfill support, parameterized date processing. 7 consecutive daily runs, 1M+ rows in <2min.

Tech: Apache Airflow 2.x, PySpark, Great Expectations, PostgreSQL, Docker Compose, Slack | HR Signal: Data Engineer hire bar

PAIR PROJECT | 2-3 STUDENTS | Weeks 5-6

Difficulty:

Real-Time Streaming Analytics Platform

Kafka producer generating realistic events, Spark Structured Streaming for real-time aggregation, dbt-managed warehouse (staging/intermediate/mart layers), Grafana real-time dashboards, dead letter queue, schema registry. 500+ events/sec, E2E latency <10s, 100% dbt test coverage on marts.

Tech: Apache Kafka, PySpark Streaming, dbt Core, PostgreSQL/DuckDB, Grafana, Docker, Schema Registry

Student A: Kafka setup, producer/consumer, Spark streaming jobs, schema registry | Student B: dbt project (models, tests, docs), Grafana dashboards, dead letter queue

GROUP CAPSTONE | 5 STUDENTS | Weeks 7-8 | Pick 1 of 3

Difficulty:

Option 1: Enterprise Data Platform as a Service

Self-service data connectors, automated schema detection, Airflow-orchestrated pipelines, dbt transformations, data catalog with lineage tracking, RBAC, SQL query editor, data quality monitoring, usage analytics. Tech: Airflow, dbt, PostgreSQL, Great Expectations, React, FastAPI, Docker

Option 2: Real-Time Fraud Detection Pipeline

Real-time transaction ingestion via Kafka, feature engineering (velocity, amount patterns, geo-anomalies), ML fraud scoring, configurable rule engine, case management dashboard, alert routing. Tech: Kafka, PySpark, Scikit-learn, FastAPI, React, PostgreSQL, Docker

Option 3: Multi-Tenant Analytics Data Warehouse

Tenant onboarding with isolated pipelines, automated schema provisioning, shared ETL framework, dbt tenant-aware models, embeddable dashboards, usage-based billing tracking. Tech: Airflow, dbt, PostgreSQL (RLS), React, FastAPI, Docker

Module Ownership (all capstones): A: Connectors, ingestion | B: Orchestration, scheduling | C: Transformations, quality | D: Dashboard, catalog | E: API, deployment, monitoring

Two Career Pathways

Both available simultaneously – not mutually exclusive

Employment Track

1. Complete 4 verified projects
2. Take PAT certification exam
3. Portfolio published to HR Portal
4. HRs discover and shortlist you
5. Interview scheduled via platform
6. Offer extended – placement tracked

Startup Track

1. Build investor-grade capstone
2. Mentor verifies for Investor Portal
3. List project with pitch deck & demo
4. Investors discover your project
5. Pitch invitation and presentation
6. Startup mentorship and funding

Project Assessment Test (PAT)

Your certification score – visible to every HR on the platform

MCQs 20% | Coding 25% | Architecture 20% | Viva Defense 25% | Peer Review 10%

- 90–100** **Elite**
Top 5% – senior roles, startup founding
- 75–89** **Pro**
Job-ready with solid fundamentals
- 60–74** **Certified**
Entry-level – core concepts proven

Mentor Verification – The Trust Engine

No project reaches HR or Investor portal without rigorous mentor review.

Code Quality 25% | Functionality 25% | Architecture 20% | Docs 15% | Deployment 15%

Ready to Build Your Future?

4 verified projects | PAT certification | Two career pathways

[ENROLL NOW](#)

[DOWNLOAD SYLLABUS](#)