SYLLABUS FOR BOARD OF STUDIES ADOPTION — generated from the live catalogue record

Data Engineering and ETL Pipelines

S-03B · 3 days (configurable 1–5 days) · Informatics · Data & Analytics · Years: 2nd, 3rd, PG

CONTENTS

1. Overview 2
2. Hours and credit position 2
3. Course outcomes 2
4. Course content — modules 2
5. Assessment scheme (100 marks) 2
6. CO–PO mapping 2
7. Every participant receives 3
8. Fee 3
9. Designed by 3

Catalogue S-03B · version 2 · generated 2026-11-07

OVERVIEW

Ingestion, transformation, scheduling and monitoring of a working data pipeline.

Behind every analytics and ML team sits a data pipeline; data-engineering skills are scarce, well-paid, and rarely met in the standard curriculum.

HOURS AND CREDIT POSITION

Contact hours: 24 Guided project: 6 Self-study: 0 Notional total: 30

Credit claim: 1
30 notional hours ÷ 30 hours per credit (NSQF track, UGC Micro-credentials Guidelines §5.3.3) = 1 credits.

COURSE OUTCOMES

CO1. Distinguish batch and streaming approaches for a stated data requirement (K4)
CO2. Build an ingestion step that extracts data incrementally (K3)
CO3. Implement transformation logic that cleans and joins source data (K3)
CO4. Schedule a pipeline with dependencies and retry handling (K3)
CO5. Monitor a running pipeline and report its data quality (K4)

COURSE CONTENT — MODULES

Hours shown are the recommended format; module time scales proportionally when the engagement runs 1–5 days.

Module 1 — The data-engineering landscape (4h recommended)
Theory: Batch versus stream, warehouses versus lakes, the data-engineer role
Lab: Match approaches to requirements
Output: Approach comparison

Module 2 — Ingestion (4h recommended)
Theory: Sources, connectors, incremental extraction
Lab: Build an incremental ingestion step
Output: Ingestion job

Module 3 — Transformation (4h recommended)
Theory: Cleaning, joining, aggregation, data modelling
Lab: Transform and model source data
Output: Transformation job

Module 4 — Orchestration (4h recommended)
Theory: Scheduling, dependencies, retries with a workflow tool
Lab: Orchestrate the steps into a DAG
Output: Scheduled workflow

Module 5 — Storage and loading (4h recommended)
Theory: Warehouse loading, partitioning, file formats
Lab: Load the output into a warehouse
Output: Loaded dataset

Module 6 — Pipeline capstone (4h recommended)
Theory: Monitoring and data-quality checks end to end
Lab: Assemble a scheduled, monitored pipeline
Output: End-to-end pipeline

ASSESSMENT SCHEME (100 MARKS)

Continuous lab exercises — 40 marks
Capstone project & demonstration — 40 marks
Quiz & viva voce — 20 marks

CO–PO MAPPING

CO1 !' PO2 (strength 2)
Choosing an approach is problem analysis.

CO2 !’ PO3 (strength 2)
Building ingestion is solution development.
CO3 !’ PO3 (strength 2)
Transformation logic is developed solution.
CO4 !’ PO5 (strength 2)
Orchestration tooling is modern-tool usage.
CO5 !’ PO4 (strength 2)
Monitoring data quality is investigation.

EVERY PARTICIPANT RECEIVES

- Certificate of completion
- Course material
- LMS access
- Interview question bank
- Mock interview & viva practice
- Optional nasscom NSQF assessment (priced per candidate on your quotation)

FEE

¹850 per student, incl. GST
Excludes the optional nasscom assessment where offered. Final pricing on your college's quotation.

DESIGNED BY

Venkat Kurela — Founder & Director, Kurela Cognisive Private Limited