

This curriculum is designed to help participants achieve intermediate-level expertise in **data ingestion, integration, and pipeline development**. It focuses on real-world **data slinger** (data wrangler) skills, progressing from foundational concepts to advanced live project exposure.

Week	Module	Topics Covered	Deliverables
1	Introduction to Data Slinger Role	Overview of data ingestion and integration; understanding the data lifecycle.	Assignment: Identify data sources for ingestion.
2	Data Ingestion Basics	File-based ingestion (CSV, JSON, XML); API-based ingestion techniques.	Mini-project: Create a simple file ingestion pipeline.
3	Databases for Data Integration	Working with relational databases (PostgreSQL, MySQL); basic SQL queries for data ingestion.	Hands-on: Write SQL scripts to ingest sample datasets.
4	NoSQL and Streaming Ingestion	Introduction to NoSQL (MongoDB, DynamoDB); streaming ingestion with Kafka.	Assignment: Ingest real-time data using Kafka.
5	Data Transformation Fundamentals	Data wrangling with Pandas; handling missing data, normalization, and data quality checks.	Hands-on: Build a transformation script for messy data.
6	Cloud Data Ingestion	Using AWS S3, Azure Blob Storage, and Google Cloud Storage for data ingestion.	Mini-project: Ingest data from cloud storage services.
7	Workflow Orchestration Tools	Introduction to Apache Airflow; building and scheduling ETL pipelines.	Hands-on: Create an Airflow DAG for data ingestion.
8	Advanced File and API Ingestion	Handling large files; paginated and secured API ingestion.	Project: API-based pipeline for multi-page data ingestion.
9	Error Handling and Retry Mechanisms	Implementing error handling; retry mechanisms in data pipelines.	Assignment: Build fault-tolerant ingestion pipelines.
10	Data Validation and Quality Assurance	Techniques for data validation; data profiling with Great Expectations.	Hands-on: Validate ingested data using predefined schemas.
11	Advanced SQL and Query Optimization	Writing complex SQL for transformations; indexing and optimization techniques.	Assignment: Optimize SQL queries for large datasets.
12	Data Lineage and Observability	Introduction to data lineage; monitoring tools like Datadog and Open Lineage.	Project: Implement lineage tracking for an ingestion flow.

Week	Module	Topics Covered	Deliverables
13	ETL vs. ELT Patterns	Differences between ETL and ELT; designing scalable data pipelines.	Mini-project: Build an ELT pipeline with SQL transformations.
14	Data Lake Ingestion	Understanding Data Lakes; partitioning strategies; ingesting structured and unstructured data.	Hands-on: Create and query a Data Lake using S3 or Azure.
15	Batch vs. Real-Time Data Processing	Batch ingestion using Spark; real-time streaming with Kafka and Spark Streaming.	Mini-project: Process real-time sensor data.
16	Data Governance and Compliance	GDPR, HIPAA; role-based access control (RBAC) for data ingestion.	Assignment: Apply data masking and encryption techniques.
17	Data Integration Tools	Working with tools like Talend, Informatica, and dbt (data build tool).	Hands-on: Build a simple pipeline using dbt.
18	End-to-End Pipeline Development	Combining ingestion, transformation, and validation processes into a single workflow.	Project: Develop an end-to-end pipeline using Airflow.
19	MLOps Integration in Pipelines	Introduction to MLOps workflows; connecting pipelines to machine learning models.	Hands-on: Automate feature extraction for ML pipelines.
20	Data Security in Pipelines	Secure transmission (SSL/TLS); managing secrets (AWS Secrets Manager, Azure Key Vault).	Assignment: Secure a pipeline with encrypted connections.
21	Advanced Monitoring and Troubleshooting	Setting up monitoring dashboards; troubleshooting common ingestion issues.	Hands-on: Build a pipeline monitoring dashboard.
22	Cloud-Native Ingestion Patterns	Serverless ingestion pipelines with AWS Lambda, Azure Functions, and Google Cloud Functions.	Project: Deploy a serverless ingestion pipeline.
23	Scalability in Data Pipelines	Techniques for scaling pipelines; partitioning, sharding, and concurrency handling.	Assignment: Optimize an existing pipeline for scalability.
24	Certification Prep	Overview of relevant certifications (AWS Data Analytics, GCP Data Engineer); practice tests.	Certification preparation.

### Weeks 25-36: Live Project Exposure

Week	Activity	Objectives	Deliverables
25	Live Project Kick-off	Introduction to a real-world project; understanding requirements and datasets.	Project plan and architecture design.

Week	Activity	Objectives	Deliverables
26	Data Source Analysis	Profiling and validating live data sources.	Report on data profiling and validation findings.
27	Data Ingestion Setup	Building ingestion pipelines for multiple data sources.	Ingestion pipelines for structured and unstructured data.
28	Workflow Orchestration Implementation	Scheduling and monitoring ingestion workflows.	Configured workflows in Airflow or a similar tool.
29	Data Transformation Development	Writing data transformation scripts for real-world requirements.	Transformation scripts implemented and tested.
30	Data Validation and Quality Assurance	Implementing quality checks and validations in pipelines.	Validated datasets with logs of quality checks.
31	Real-Time Data Processing	Building streaming ingestion for real-time data (e.g., Kafka-based pipelines).	Real-time data ingestion and processing flows.
32	Monitoring and Error Handling	Implementing robust monitoring, logging, and error recovery mechanisms.	Dashboard and alerting setup for the pipeline.
33	Deployment Preparation	Preparing the pipeline for deployment; final testing and debugging.	Deployment-ready pipeline codebase.
34	Pipeline Deployment	Deploying the pipeline to production environment.	Deployed and functional pipeline.
35	Documentation and Handover	Documenting the pipeline architecture, workflows, and configurations.	Comprehensive documentation for the live project.
36	Project Review and Retrospective	Reviewing the project outcomes, identifying areas for improvement.	Final project review report and retrospective document.