



AWS Data Engineering Training

Industry-Oriented Curriculum • Hands-On Labs • End-to-End Project



20 Modules	AWS Services	Core Skills	Capstone
Structured Learning	S3 • Glue • Kinesis • Lambda	SQL • Python • PySpark	Real-Time Data Pipeline

Designed from the supplied AWS Data Engineering syllabus and reorganized into a clearer learning path for practical training.

Program Overview

This curriculum takes learners from data engineering foundations and AWS fundamentals through storage, ETL, streaming, serverless processing, analytics, orchestration, observability, and a complete real-time project. The prerequisite track in Linux, SQL, Python, and PySpark is integrated into the learning sequence so learners can build pipelines rather than study services in isolation.

What You Will Be Able to Build

- Batch ingestion and transformation pipelines using Amazon S3 and AWS Glue.
- Serverless data-processing workflows using AWS Lambda and Step Functions.
- Streaming pipelines using Amazon Kinesis and PySpark concepts.
- Interactive analytics using Amazon Athena and warehouse workloads using Amazon Redshift.
- Operational workflows with IAM security, CloudWatch monitoring, and orchestration.
- An end-to-end project that connects ingestion, processing, storage, analytics, and monitoring.

Recommended Learning Path

Phase	Focus	Modules	Outcome
1	Foundations	1–6	AWS + Linux + SQL + Python + PySpark readiness
2	Core AWS Data Platform	7–12	Storage, ETL, query, streaming, serverless, warehouse
3	Operations & Orchestration	13–15	Monitoring and workflow automation
4	Capstone	16	End-to-end real-time project
5	Career Readiness	Integrated	Resume guidance, FAQs, mock interview

Module 1: AWS Data Engineering Introduction

Learning outcome: Understand the role of a data engineer and how AWS services fit into modern data platforms.

- What is Data Engineering and common data engineering technologies
- AWS Data Engineering overview
- Role of SQL, Linux, Python, and PySpark in data engineering
- Batch vs. streaming data pipeline concepts

Hands-on: Map a sample business requirement to a simple AWS data pipeline.

Module 2: AWS Fundamentals for Data Engineering

Learning outcome: Build the AWS foundation required to work safely with data services.

- AWS Cloud Computing introduction, AWS account sign-in, Free Tier and service overview
- AWS Global Infrastructure: Regions and Availability Zones
- EC2, IAM and default VPC fundamentals
- Basic cost-awareness and resource cleanup practices

Hands-on: Create/access an AWS environment and identify the services used in the course.

Module 3: Linux for AWS Data Engineering

Learning outcome: Use Linux confidently for cloud data engineering tasks.

- Linux filesystem architecture
- Linux installation/use on an EC2 instance
- Connecting to a Linux machine
- Basic commands, filters, files/directories and permissions

Hands-on: Connect to an EC2 Linux host and perform common file and permission operations.

Module 4: SQL for AWS Data Engineering

Learning outcome: Query and manipulate relational data used by pipelines and analytics workloads.

- SQL basics: DDL, DML, DQL, TCL and DCL
- Filtering, sorting and joins
- MySQL setup concepts
- Provisioning MySQL using Amazon RDS

Hands-on: Create a small relational dataset in RDS/MySQL and query it with joins and filters.

Module 5: Python for AWS Data Engineering

Learning outcome: Use Python as the scripting foundation for AWS automation and data processing.

- Python installation, interactive mode and script mode
- Editors/IDEs including Jupyter, PyCharm and VS Code
- Language fundamentals, indentation and data types
- Flow control, collections, modules, packages and libraries
- Functions, lambda functions, classes and objects

Hands-on: Write a Python script that reads, validates and transforms a sample dataset.

Module 6: PySpark for AWS Data Engineering

Learning outcome: Process distributed datasets using Spark concepts relevant to AWS workloads.

- PySpark foundation
- RDD programming: transformations and actions
- Spark SQL: DataFrames, tables, DSL and native SQL
- PySpark Streaming concepts
- Database integration and Amazon S3 integration

Hands-on: Transform a dataset from S3 using DataFrames and Spark SQL.

Module 7: Storage with Amazon S3

Learning outcome: Design and manage a durable data-lake storage layer.

- S3 buckets and objects using AWS Console
- Upload local datasets and manage objects
- Versioning and cross-Region replication
- Storage classes and Glacier overview
- S3 management using AWS CLI
- S3 integration with PySpark, Glue, Lambda and Athena

Hands-on: Create a structured S3 data lake area and load versioned source data.

Module 8: Data Processing with AWS Glue

Learning outcome: Build metadata-driven ETL pipelines with AWS Glue.

- AWS Glue components
- Crawlers, Catalog databases and tables
- Create and run Glue jobs and triggers
- Glue workflows and validation
- Spark History Server / Spark UI setup concepts
- IAM permissions required for Glue
- Managing the Glue Data Catalog and crawling multiple folders

Hands-on: Catalog S3 data, run a Glue ETL job and validate transformed output.

Module 9: Querying with Amazon Athena

Learning outcome: Query S3 data efficiently using serverless SQL.

- Athena and Glue Catalog integration
- Query Editor and database/table creation
- CTAS and data population
- Athena architecture and relationship with Hive
- Partitioned tables and partition-aware queries
- Insert into partitions and validate partitioning
- Drop tables and manage related data files

Hands-on: Create a partitioned Athena table and compare partition-aware queries.

Module 10: Real-Time Data Processing with Amazon Kinesis

Learning outcome: Understand and implement streaming ingestion patterns.

- Building a streaming pipeline using Kinesis
- Rotating logs and Firehose agent setup
- Create a Firehose delivery stream
- Pipeline planning and IAM permissions
- Configure, start and validate the agent
- Integrating streaming data with PySpark concepts

Hands-on: Stream sample events into an AWS destination and validate delivery.

Module 11: Serverless Compute with AWS Lambda

Learning outcome: Build event-driven data processing without managing servers.

- Lambda Hello World and local development setup
- Deploy a project to AWS Lambda
- Using third-party Python libraries
- S3 download/upload integration
- Incremental file validation
- Bookmarks/state maintained in S3
- Deployment and scheduling with Amazon EventBridge

Hands-on: Build an incremental S3 processing Lambda and schedule it with EventBridge.

Module 12: Data Warehousing with Amazon Redshift

Learning outcome: Load, query and manage analytical data in Amazon Redshift.

- Redshift introduction and cluster creation concepts
- Query Editor, information schema and saved queries
- Tables and CRUD operations
- COPY data from S3 to Redshift
- Load JSON data using an IAM role
- Redshift architecture and multi-node cluster concepts
- Databases, users and schemas
- PySpark integration with Redshift

Hands-on: Load curated S3 data into Redshift and run analytical SQL queries.

Module 13: Logging & Observability with Amazon CloudWatch

Learning outcome: Monitor pipeline health and troubleshoot AWS data workloads.

- CloudWatch role in data engineering observability
- Logs, metrics and alarms for pipeline components
- Monitoring failures and operational events
- Basic troubleshooting workflow for data pipelines

Hands-on: Define monitoring checkpoints and alerts for the capstone pipeline.

Module 14: Workflow Orchestration with AWS Step Functions

Learning outcome: Coordinate multi-step data workflows with retries and error handling.

- Step Functions introduction and setup
- States, tasks and state machines
- Integrating with AWS services
- Developing and deploying workflows
- Monitoring, logging and error handling
- Workflow design best practices
- Scaling, optimization and cost management

Hands-on: Orchestrate a serverless ETL workflow with failure handling.

Module 15: Apache Airflow for AWS Data Engineering

Learning outcome: Understand DAG-based orchestration for scheduled data workflows.

- Airflow role in data engineering orchestration
- DAGs, tasks, dependencies and scheduling
- Coordinating AWS-oriented pipeline steps
- Operational considerations for workflow monitoring

Hands-on: Design an Airflow DAG for a multi-stage AWS data pipeline.

Module 16: End-to-End Real-Time Project

Learning outcome: Combine the course services into a portfolio-ready AWS data engineering solution.

- Project requirements and pipeline planning
- Source ingestion to S3 / streaming layer
- Cataloging and ETL processing
- Query and warehouse serving layer
- Workflow orchestration and monitoring
- Validation, troubleshooting and project documentation
- Resume guidance, interview FAQs and mock interview support

Hands-on: Build and present an end-to-end AWS data engineering project with architecture, implementation flow and validation.

Real-Time Project:



Suggested Architecture Flow

Source Data → Amazon S3 / Amazon Kinesis → AWS Glue / AWS Lambda → Amazon Athena / Amazon Redshift → Analytics & Insights, coordinated with Step Functions or Airflow and monitored through CloudWatch.

- Ingest batch files and/or streaming events.
- Maintain raw and processed zones in S3.
- Catalog datasets and transform them with Glue/PySpark concepts.
- Implement incremental/serverless processing with Lambda where appropriate.
- Query curated data with Athena and load analytical data into Redshift.
- Add orchestration, logging, error handling and validation checkpoints.
- Document architecture, assumptions, test results and business insights.

Portfolio Deliverables

- Architecture diagram
- Source-to-target mapping
- ETL/streaming implementation notes
- SQL queries and validation results
- Monitoring and failure-recovery checklist
- README / project explanation for interviews

Thank You for Exploring AWS Data Engineering!

You've taken the first step toward building a career in **Cloud Data Engineering with AWS**.

With this course, you'll gain practical experience in building scalable data pipelines, processing large datasets, and working with industry-leading AWS data services.

Get Ready to Learn

-  Amazon S3 & Data Lakes
-  AWS Glue & ETL Pipelines
-  Amazon Redshift
-  Amazon EMR & Apache Spark
-  AWS Lambda & Event-Driven Processing
-  Amazon Kinesis for Streaming Data
-  AWS RDS & DynamoDB
-  IAM, Security & Data Governance
-  Real-Time Data Engineering Projects
-  Interview & Career Preparation

Start Your AWS Data Engineering Journey Today!

Learn • Build • Practice • Become Job-Ready

Reach Us for a FREE Demo

 **Call / WhatsApp:** +91 99514 40801

 www.sqlschool.com

SQL School

Learn Today. Lead Tomorrow.

For the visual page, I recommend a **dark AWS-style cloud background**, AWS/Data Engineering icons in the center, a large **“Thank You!”** headline, and the **Free Demo CTA + contact details** prominently at the bottom.

SQL SCHOOL

Premium Quality Training • Practical Learning • Career-Focused Projects