

AWS Data Engineering Masterclass

One AWS platform: batch, streaming, warehouse and orchestration

Glue

EMR

Redshift

Kinesis

MWAA

Iceberg

Trainer Venu Katragadda · Live online · Recordings for life

₹28,000

COURSE SYLLABUS

AWS Data Engineering Masterclass

Build a production data platform on AWS — batch, streaming, warehouse and orchestration — and walk into the DEA-C01 exam ready.

100 hours LIVE INSTRUCTION	10 modules STRUCTURED PATH	3 projects PORTFOLIO READY	₹28,000 EMI AVAILABLE
-------------------------------	-------------------------------	-------------------------------	---------------------------------

Level	Beginner to Advanced
Mode	Live, instructor-led online sessions. Every session recorded; lifetime access.
Trainer	Venu Katragadda — 14+ years in industry (L&T Infotech, Capgemini, Brillio, Manthan Systems). 1,200+ professionals trained.
Batch timings	Morning 6:00–7:30 AM IST · Morning 7:30–9:00 AM IST · Evening 7:30–9:00 PM IST · Weekend (Sat & Sun)
Certification	AWS Certified Data Engineer – Associate (DEA-C01)
Enquiries	+91-8500002025 / +91-9247159150 (WhatsApp) · info@databrickstraining.com

What you will be able to do

Every outcome below is demonstrated in a lab or a graded project — not just discussed in a slide.

- Design a lakehouse on S3 with Glue Catalog, Lake Formation permissions and Iceberg tables
- Build ETL with Glue (Spark and Python shell), EMR Serverless and Step Functions
- Model and tune Redshift: distribution and sort keys, RA3, Serverless, Spectrum, materialised views
- Stream with Kinesis Data Streams, Firehose, MSK (Kafka) and Flink/Managed Service for Apache Flink
- Orchestrate with Amazon MWAA (Airflow) and monitor cost with CUR + Cost Explorer
- Pass AWS Certified Data Engineer – Associate

Who this masterclass is for

- Data engineers standardising on AWS
- ETL developers migrating from on-prem
- Cloud engineers adding data skills
- DEA-C01 candidates

Prerequisites

- SQL and basic Python
- No prior AWS experience needed — Module 1 covers IAM, VPC and S3 from zero
- An AWS account (free tier); we show how to set a \$20 budget alarm before you touch anything

Tools and technologies

Amazon S3 · AWS Glue · Glue Data Catalog · Lake Formation · Amazon EMR · EMR Serverless · Amazon Redshift · Amazon Athena · Kinesis · Amazon MSK · Managed Flink · AWS Lambda · Step Functions · Amazon MWAA (Airflow) · DMS · Apache Iceberg · Terraform · CloudWatch

Curriculum — 100 hours across 10 modules

Sessions run live in small batches. Roughly 60% of each session is hands-on lab work.

MODULE 1 AWS foundations for data engineers

10 hrs

- IAM: users, roles, policies, trust relationships, least privilege, permission boundaries
- VPC essentials: subnets, security groups, endpoints (why Glue needs them), NAT cost traps
- S3 deep dive: storage classes, lifecycle, versioning, encryption (SSE-S3/KMS), S3 Tables
- CLI, SDK (boto3), CloudFormation vs Terraform vs CDK
- Cost control: budgets, alarms, tagging, Cost Explorer, the CUR
- Lab: secure landing zone with budget guardrails

MODULE 2 Data lake on S3: catalog, formats and Iceberg

10 hrs

- Glue Data Catalog, crawlers, partitions, partition projection
- Parquet vs Iceberg vs Delta vs Hudi on AWS — current support matrix
- Apache Iceberg on AWS: snapshots, time travel, compaction, S3 Tables managed Iceberg
- Lake Formation: LF-tags, row and column filters, cross-account sharing
- Athena v3: engine, CTAS, workgroups, cost per TB scanned, query tuning
- Lab: convert a 300 GB Parquet lake to Iceberg and cut Athena spend 60%

MODULE 3 AWS Glue in depth

12 hrs

- Glue Spark jobs, Python shell jobs, Ray jobs; worker types and DPU sizing
- DynamicFrame vs DataFrame; resolveChoice, relationalize, bookmarks
- Glue Studio, notebooks, interactive sessions; Glue Data Quality (DQDL)
- Triggers, workflows, blueprints; Glue Streaming
- Debugging: continuous logging, Spark UI, metrics, common OOM patterns
- Lab: incremental ingestion with job bookmarks and DQ rules that fail the build

MODULE 4 Amazon EMR and Spark on AWS

10 hrs

- EMR on EC2 vs EMR Serverless vs EMR on EKS — decision framework and pricing
- Cluster sizing, instance fleets, spot strategy, managed scaling
- Bootstrap actions, configurations, EMRFS consistency, S3A tuning
- Running the PySpark jobs from your own repo on EMR Serverless
- Lab: same job on Glue vs EMR Serverless — benchmark and cost report

MODULE 5 Streaming: Kinesis, MSK and Flink

12 hrs

- Kinesis Data Streams: shards, on-demand mode, enhanced fan-out, KCL/KPL
- Firehose: buffering, dynamic partitioning, transformation Lambdas, format conversion
- Amazon MSK: Kafka on AWS, MSK Serverless, Connect, Glue Schema Registry
- Managed Service for Apache Flink: SQL and DataStream API, checkpointing, exactly-once
- Kafka fundamentals: topics, partitions, consumer groups, offsets, compaction, idempotent producers
- Lab: end-to-end IoT stream with replay and a poison-message DLQ

MODULE 6 Amazon Redshift

12 hrs

- Architecture: leader/compute nodes, slices, RA3 managed storage, Redshift Serverless
- Distribution styles (KEY/EVEN/ALL/AUTO), sort keys, compression encodings
- COPY and UNLOAD, Redshift Spectrum, zero-ETL from Aurora, data sharing
- Workload management (WLM/Auto WLM), concurrency scaling, query monitoring rules
- Materialised views, auto-refresh, federated queries; STL/SVL system tables
- Lab: model a 1B-row fact table and take a report from 90 s to 4 s

MODULE 7 Databases, ingestion and migration

8 hrs

- DynamoDB for data engineers: partition keys, streams, export to S3
- RDS/Aurora, zero-ETL integrations, Aurora → Redshift
- AWS DMS: full load + CDC, task tuning, validation, common failure modes
- AppFlow and third-party SaaS ingestion; Airbyte/Fivetran on AWS
- Lab: Oracle → S3 CDC pipeline with DMS and schema drift handling

MODULE 8 Orchestration with Airflow (MWAA) and Step Functions

10 hrs

- Airflow core: DAGs, operators, TaskFlow API, XComs, sensors, pools, SLAs
- Airflow 2.x vs 3.x changes; deferrable operators and the triggerer
- Amazon MWAA: sizing, plugins, requirements.txt, DAG deployment via S3 + CI
- AWS Step Functions: state machines, Map/Distributed Map, error handling, cost vs Airflow
- EventBridge scheduling and event-driven pipelines; Lambda patterns and limits
- Lab: production DAG with backfill, retries, alerting and data-quality gates

MODULE 9 Security, observability and cost engineering

8 hrs

- KMS, encryption in transit/at rest, Secrets Manager, cross-account roles
- CloudWatch metrics/logs/alarms, CloudTrail, Glue/EMR observability
- PII handling: Macie, tokenisation, masking with Lake Formation
- FinOps for data: right-sizing, spot, storage tiering, query cost attribution
- Lab: build a cost dashboard for the platform you just built

MODULE 10 Architecture, capstone and DEA-C01 prep

8 hrs

- Reference architectures: batch lakehouse, streaming, hybrid; well-architected data lens
- Choosing between Glue / EMR / Redshift / Athena — a decision tree you can defend
- Capstone presentation and architecture review
- DEA-C01 exam guide walkthrough, 130 practice questions, two timed mocks
- AWS data engineer interview questions and resume framing

Hands-on projects

You leave with work you can demo in an interview. Each project is code-reviewed and returned with written feedback.

PROJECT 1 Serverless retail lakehouse

S3 + Glue + Iceberg + Athena + QuickSight, with Lake Formation row-level security and an MWAA DAG orchestrating the whole thing.

PROJECT 2 Streaming IoT platform

MSK → Managed Flink → S3/Redshift with schema registry, dead-letter handling and a real-time dashboard.

PROJECT 3 On-prem Oracle → AWS migration

DMS full load + CDC into S3, Glue transformation to Iceberg, Redshift serving layer, and a cutover runbook.

Frequently asked questions

Will AWS charge me during labs?

Labs are designed for the free tier plus a few dollars. We set budget alarms in the first session and every lab ends with a teardown checklist. Expect roughly \$15–\$40 total across the course if you follow the teardown steps.

Does this replace the AWS Solutions Architect course?

No — it is the data specialisation. You'll learn the IAM/VPC/S3 you need, but not EC2 fleet management or hybrid networking in depth.

Is Kafka covered or only Kinesis?

Both. Module 5 teaches Kafka fundamentals properly (partitions, consumer groups, offsets, compaction, exactly-once) and then applies them on Amazon MSK, alongside Kinesis for comparison.

Which certification does this map to?

AWS Certified Data Engineer – Associate (DEA-C01). Note that the older Data Analytics – Specialty (DAS-C01) has been retired — do not buy prep material for it.

How to enrol

- Book a free demo. Sit in on a live session with the current batch — no card, no commitment. WhatsApp +91-9247159150 or call +91-8500002025.
- Pick your batch. Weekday morning, weekday evening or weekend. All times IST.
- Pay the fee (₹28,000, instalments and EMI available) and get your workspace, material and calendar invites.
- Start building. Labs from session one; recordings posted the same day.

What is included

- 100 hours of live instruction by Venu Katragadda
- Lifetime access to all session recordings
- 3 code-reviewed portfolio projects
- Preparation for: AWS Certified Data Engineer – Associate (DEA-C01)
- Resume review, LinkedIn optimisation and one mock interview with written feedback
- Free re-attend of any future batch of the same masterclass
- Doubt-clearing sessions and a private student group

Refund policy in brief

When you cancel	Refund
Before the batch starts	Up to 90%
Within 5 days of start, max 2 classes attended	Up to 70%
After 7 days, or more than 5 classes attended	No refund — free transfer to a later batch

Full policy: databrickstraining.com/refund-policy/

Talk to us

Call	+91-8500002025 · +91-9247159150
WhatsApp	+91-9247159150 — usually answered within the hour
Email	info@databrickstraining.com
Website	databrickstraining.com/courses/aws-data-engineering-masterclass/
Foundation tier	databrickstraining.in — shorter 70–80 hour courses if this one is too advanced
Institute	Sreyobhilashi IT · Udyam UDYAM-TS-02-0334643 · 112, Samshraya Homes, opp. PF Office, Balanagar, Kukatpally, Hyderabad, Telangana 500037

Sreyobhilashi IT is an independent training provider and is not affiliated with, endorsed by or sponsored by Databricks, Amazon, Microsoft, Google, Snowflake, dbt Labs or Anthropic. Placement assistance is a best-effort service, not a guarantee of employment. Cloud platforms change frequently; confirm current pricing and service behaviour with the vendor's documentation.