

AWS Data Engineering

2 months · live + recorded · AWS Certified Data Engineer Associate — the fastest path into a Rs 12-30 LPA data engineering role at India's product startups and global SaaS companies.

DURATION

2 mo

MODULES

6

ROLES UNLOCKED

6+

SALARY BAND

5-35 LPA

| **ABOUT THIS PROGRAMME**

AWS is the default cloud for Indian product startups and global SaaS in 2026, and the AWS Certified Data Engineer Associate (DEA-C01) is the fastest path into a Rs 12-30 LPA data engineering role. This programme takes you through the entire AWS data stack — S3, Glue, EMR, Redshift, Kinesis, Lambda, Step Functions, Athena, and the Lake Formation governance layer. You'll build a real lakehouse on AWS with ingestion, transformation, governance, and BI consumption. Every module ships a deliverable, every concept lands on a real exam question.

| **BUILT FOR**

- Engineers targeting AWS-first product companies + startups
- Existing data engineers migrating from on-prem to AWS
- Final-year CS students preparing for the DEA-C01 certification

OUR COMMITMENT

Job-focused training. Real instructors. Lifetime access. We don't sell certificates — we get you hired.

WHY NOW · 2026 MARKET SIGNAL

AWS dominates Indian product-startup infra — 60%+ market share. The AWS Certified DEA-C01 is the single highest-ROI certification a fresher can target in 2026.

01

1,800+ open AWS DE roles on Naukri this month

02

DEA-C01 adds median Rs 3-5 LPA at the offer table

03

AWS spending by Indian product cos grew 45% in 2025

TECHNOLOGY STACK

Technologies you'll master

The exact stack senior engineers + consultants are billing for in 2026 — nothing taught for show, everything taught because it lands on a job posting.

- AWS Glue — serverless ETL + Data Catalog
- Amazon EMR — managed Spark/Hadoop clusters
- Amazon Redshift — cloud data warehousing
- Amazon S3 + Lake Formation — lakehouse foundation
- Amazon Athena — serverless SQL on S3
- Amazon Kinesis — real-time data streaming
- AWS Lambda + Step Functions for orchestration
- Amazon MWAA — Managed Apache Airflow
- Iceberg + Hudi for open table formats
- AWS IAM + KMS for security and encryption

WHY THIS PROGRAMME

What makes this different

- ✓ Complete AWS DEA-C01 and Solutions Architect Associate certification prep
- ✓ Lake Formation governance and cross-account data sharing patterns
- ✓ End-to-end project: Kinesis → S3 → Glue → Redshift → QuickSight
- ✓ Real-time pipelines with Kinesis Data Streams, Firehose, and Lambda
- ✓ Glue PySpark deep-dive with bookmarks, partitioning, and dynamic frames
- ✓ Cost optimization workshop — when to use Athena vs Redshift vs EMR
- ✓ Redshift performance: sort keys, dist keys, RA3 nodes, materialized views
- ✓ Mock interviews with engineers currently working at top AWS-first companies

CAREER OUTCOMES

By the end, you will be able to

- ✓ Architect a serverless lakehouse on S3 + Glue + Athena with full governance
- ✓ Build Redshift warehouses optimized for sub-second analytics at TB scale
- ✓ Ship real-time pipelines via Kinesis with exactly-once processing semantics
- ✓ Write production PySpark in Glue with proper error handling and monitoring

- ✓ Implement Lake Formation policies for column and row-level security
- ✓ Estimate AWS data costs and optimize 40%+ through smart service selection
- ✓ Pass AWS DEA-CO1 and SAA-CO3 certifications
- ✓ Interview for AWS Data Engineer roles paying Rs 8-28 LPA

| SALARY OPPORTUNITIES

Job roles + earning potential

Indian-market salary bands for 2026, sourced from active listings on Naukri, LinkedIn, and Instahyre. Bands assume the certifications and projects this programme delivers.

	JOB ROLE	EXPERIENCE	SALARY
01	Data Engineer (AWS)	Fresher to 2+ Years	Rs 5-9 LPA
02	ETL Developer (AWS Glue/Redshift)	Fresher to 3 Years	Rs 6-12 LPA
03	Cloud Data Engineer	2 to 4 Years	Rs 8-14 LPA
04	Big Data Engineer (PySpark/Hadoop/Spark)	2 to 5 Years	Rs 10-16 LPA
05	AWS Solutions Associate (Data Focus)	3 to 5 Years	Rs 12-18 LPA
06	Data Warehouse Engineer (Redshift/Snowflake)	3 to 6 Years	Rs 12-20 LPA
07	AWS Data Engineer / Data Consultant	4 to 7 Years	Rs 15-25 LPA
08	Senior Data Engineer / Lead	6+ Years	Rs 20-35 LPA

| CAREER PATHS

Roles you can target

Roles ordered from entry → senior. The same role titles you'll see on Naukri, LinkedIn, and Instahyre filtered to the tech stack you'll master.

- AWS Data Engineer
- Cloud Data Engineer (AWS)
- Big Data Engineer
- Data Platform Engineer
- Senior Data Engineer (AWS)
- AWS Solutions Architect — Data

Emerging — high-demand roles for 2026

Roles that didn't exist five years ago, paying premium because the talent pool is small.

- AWS Lakehouse Engineer
- AI/ML Data Engineer (AWS)
- Real-Time Streaming Engineer

| CAREER TRAJECTORY

Where this programme leads, in 5 years

Median 2026 salary bands for the role progression this curriculum unlocks. Numbers are conservative — top performers cross the upper band earlier.

YEAR 1

AWS Data Engineer

Rs 7-14 LPA

Ship first production lakehouse.
Earn DEA-C01.

YEAR 3

Senior AWS Data Engineer

Rs 15-25 LPA

Own pipeline architecture for a
business function.

YEAR 5

Principal Data Engineer

Rs 28-45 LPA

Set platform strategy. Mentor 10+
engineers.

| COURSE CURRICULUM

What you'll build, module by module

Every module is taught hands-on with real datasets, instances, or codebases — not slides. By the end, you'll have built, not just watched.

Python

- What is Python?
- Why Python for Data Engineering?
- Installing Python and Setting Up Environment (IDEs, ...)
- Running Python Scripts and Notebooks
- Basic Syntax and Indentation
- Variables and Data Types (int, float, str, bool, None ...)
- Type Casting and `type()` function
- Arithmetic, Comparison, Logical Operators
- Membership (`in`, `not in`) and Identity Operators
- Operator Precedence and Associativity
- `if`, `elif`, `else` Statements
- `while` and `for` Loops
- Loop Control: `break`, `continue`, `pass`
- List Comprehensions (important for Glue transform...)
- Defining and Calling Functions
- Parameters and Return Values
- Lambda Functions (used heavily in PySpark)
- `map()`, `filter()`, `reduce()` (from `functools`)
- Lists, Tuples, Sets, Dictionaries
- CRUD operations on each data structure
- Iterating through collections
- Common built-in functions (`len`, `sum`, `sorted`, `zip`...)
- String Manipulation and Formatting
- `split()`, `join()`, slicing, and regex intro (`re` module)
- Introduction to `datetime` and `time` modules (for p...)
- Try-Except Blocks
- Catching Specific Exceptions
- `finally` and `else` in error handling
- Importance in ETL pipeline robustness
- Classes and Objects
- Constructors (`__init__`)
- `self` keyword
- Simple inheritance and method overriding

Data Warehouse

- What is Data Warehousing?
- OLTP vs OLAP
- Data Warehouse Architecture (Single-tier, Two-tier...)
- Components of a Data Warehouse
- ETL vs ELT in Data Warehousing
- What is Data Modeling?
- Conceptual, Logical, and Physical Data Models
- Key Data Modeling Concepts: Entities, Attributes, R...
- Primary Keys, Foreign Keys, and Constraints
- Normalization & Denormalization
- Choosing the Right Model for Analytical Workloads
- Introduction to Dimensional Modeling
- Fact Tables vs Dimension Tables
- Star Schema: Concepts & Design
- Snowflake Schema: When to Use It?
- Slowly Changing Dimensions (SCD) (Types 0, 1, 2, 3, ...)
- Handling Hierarchies & Aggregations
- Overview of ETL & ELT Processes
- Common ETL Challenges & Solutions
- Data Quality & Data Governance in ETL
- Change Data Capture (CDC) Strategies
- Traditional Data Warehouses vs Cloud Data Wareh...
- Introduction to Data Lakes & Data Lakehouses
- Overview of Modern DW Platforms: Snowflake, Big...

PySpark

- What is PySpark?
- PySpark vs Pandas vs Dask
- PySpark Architecture & Execution Model
- Setting up PySpark in Google Colab
- Introduction to SparkSession & DataFrames
- Reading & Writing Data (CSV, JSON, Parquet, Avro)
- Understanding Schema Inference & Defining Schem...
- Basic Transformations: `select()`, `filter()`, `withColu...
- Handling Nulls & Missing Data (`fillna()`, `dropna()`, `...`)
- Column Operations: `cast()`, `alias()`, `when()`, `case()...
- Working with Date & Time Functions (`current_date...`)
- Grouping & Aggregations (`groupBy()`, `agg()`, `pivo...
- Joins in PySpark (inner, left, right, full)
- Window Functions (Row Number, Ranking, Lead/La...
- Exploding & Flattening Nested Data (`explode()`, `ar...
- Working with UDFs (User-Defined Functions)
- Broadcasting & Skew Handling
- Understanding Spark Execution Plan (`explain()`, `ca...
- Catalyst Optimizer & Tungsten Execution
- Partitioning & Bucketing Strategies
- Repartitioning & Coalescing
- Optimizing Shuffle Operations
- Performance Tuning Parameters (`spark.conf.set()`)

PySpark Assignment Problem

- Statements 1 – Hands-On Coding PySpark Assignm...
- Statements 2 – Hands-On Coding

Amazon Web Services (AWS)

- Setting up AWS Account and Configuring IAM Roles...
- Creating S3 Buckets, Uploading Data, and Configur...
- Implementing IAM Best Practices for Secure Data A...
- Setting Up AWS Glue Crawler to Discover Metadata
- Creating and Querying AWS Glue Catalog Tables
- Schema Evolution & Handling Semi-Structured Dat...
- Integrating Glue Catalog with Athena & Redshift Sp...
- Writing SQL Queries on S3 Data Using Athena
- Optimizing Queries with Partitioning & Bucketing
- Using Iceberg Tables in Athena for Time-Travel Qu...
- Performance Optimization: Query Federation & Co...
- Setting Up AWS Glue Job with PySpark
- Transforming & Cleaning Raw Data Using PySpark i...
- Handling Schema Drift in Glue ETL Pipelines
- Writing Processed Data to S3, Redshift, and RDS
- Configuring AWS Glue Job to Ingest Data from RES...
- Using AWS Lambda to Trigger Glue Jobs on Event S...
- Handling Real-Time Data Streams in PySpark
- Writing Ingested Data to Iceberg Tables in Athena
- Setting Up an Amazon Redshift Cluster
- Loading Data from S3 to Redshift Using COPY Com...
- Performance Tuning with Sort & Distribution Keys
- Running Complex Analytical Queries in Redshift
- Creating S3, IAM Roles, Glue Jobs, and Redshift Usi...
- Automating Data Pipeline Deployment Using Cloud...
- Managing Stack Updates & Rollbacks

Athena Assignment & Problem Statements

- Statements 1 – Hands-On Coding Redshift Assignm...
- Statements 2 – Hands-On Coding Glue PySpark Ass...
- Statements 3 – Hands-On Coding

I INSTRUCTED BY

Trained by working practitioners

Not academics. Not pre-recorded YouTubers. Engineers and consultants who still ship production code at top companies — the exact people you want to learn from.



Mr. Anuj S

Industry Trainer

A Data Engineering professional with over 11+ Yrs of experience in building scalable data pipelines, distributed systems & cloud-native architectures. Has extensive expertise in Apache Spark, Hadoop, Hive, Kafka, and programming with Python, Java, SQL. Anuj brings real-world project knowledge into the classroom, helping learners master modern data engineering practices including streaming ETL, Data Lakehouse design, and ML pipeline integration. Approved trainer by Raj Cloud Technologies.

APPROVED TRAINER · RAJ CLOUD TECHNOLOGIES

| WHAT'S INCLUDED

Everything you get beyond the curriculum

The sessions are the tip. Here's the full value stack a single fee unlocks — mentorship, community, career support, and lifetime updates.

- + Lifetime access to recordings — re-watch any session, forever
- + Private community access — get unblocked in minutes, not days
- + 1:1 mentor reviews on every portfolio project you build
- + Weekly doubt-clearing sessions led by the instructor
- + Resume + LinkedIn rewrite by a working hiring manager
- + Two full mock interviews with real-time feedback
- + Interview question bank — 50+ scenario-based questions
- + Real-world datasets + playground accounts you keep
- + Verifiable course completion certificate
- + Lifetime curriculum updates — when modules refresh, you get them

| ENROLMENT

Pick your tier and enrol

Two tiers — pick what fits your budget. Both include lifetime access to recordings and a verifiable certificate.

MOST POPULAR

LIVE + RECORDED

Rs 27,499

- ✓ All live sessions + Q&A
- ✓ Lifetime recordings
- ✓ Doubt-clearing sessions
- ✓ Resume + interview prep
- ✓ Certificate of completion

RECORDED ONLY

Rs 19,249

Save Rs 8,250

- ✓ Lifetime access to recordings
- ✓ Self-paced learning
- ✓ Certificate of completion
- ✗ No live sessions
- ✗ No live Q&A

HOW TO ENROL

Three steps to your seat

01 Book a free demo

30-min walk-through with the instructor — ask anything before you commit.

02 Pick your tier + enrol

Live + Recorded or Recorded-only. UPI, card, EMI — all accepted.

03 Start on day one

Onboarding kit, course portal access, and your first live session calendar invite.

| HOW WE STACK UP

Compared to your alternatives

Free tutorials and generic bootcamps both have their place. Here's exactly where they fall short — and where the fee here pays for itself.

WHAT YOU GET	FREE YOUTUBE TUTORIALS	GENERIC BOOTCAMPS	RAJ CLOUD
Live practitioner instructors	×	×	✓
Real datasets + playground accounts	×	partial	✓
1:1 mentor reviews on your portfolio	×	×	✓
Mock interviews with feedback	×	partial	✓
Resume + LinkedIn rewrite	×	×	✓
Job referral network access	×	×	✓
Lifetime updates + recordings	partial	×	✓
Verifiable certificate	×	✓	✓

| QUESTIONS ANSWERED

Common questions, answered upfront

Q. Do I need prior experience to start?

Most students begin with little to no experience. We assume only basic computer literacy. The first two weeks ramp you up to the level the rest of the curriculum needs — no one is left behind.

Q. What if I miss a live session?

Recordings are uploaded within 24 hours of every session. Weekly doubt-clearing sessions catch you up on missed material. Your community is active 24/7. Nothing gets lost, only delayed.

Q. Is there a job placement guarantee?

We don't promise placement — we engineer outcomes. Resume + LinkedIn rewrite, two mock interviews, a referral network, and portfolio reviews are all included. The work happens; the offer is yours to earn.

Q. Can I pay in instalments?

Yes. UPI EMI, credit-card EMI, and a 3-month no-cost split-pay are all supported. Email admissions and we'll set up a plan that fits your cash flow.

Q. Does the certificate hold weight in interviews?

Yes. It's dated, verifiable via our website, and signed by the trainer. Alumni reference it on LinkedIn and inside offer-stage interview rounds — but the portfolio you build matters more, and that's the point.

Talk to admissions

EMAIL

info@rajcloudtech.com

PHONE

+91 93804 87284

WEB

rajcloudtech.com

OUR COMMITMENT

Job-focused training. Real instructors. Lifetime access. We don't sell certificates — we get you hired.