SOLVIXE

CASE STUDY · Powered by AWS

How Solvixe and AWS Built a Real-Time JEE Question AI Enrichment Pipeline for eSaral

A serverless, decoupled content-processing pipeline that ingests JEE questions, enriches them with AI and persists results for downstream use.

CLIENT	ENGAGEMENT
eSaral	Serverless JEE question AI enrichment pipeline on AWS

SOLUTION AT A GLANCE

eSaral implemented a serverless, decoupled content-processing pipeline on AWS to ingest JEE questions, enrich them with AI and persist results for downstream use. The solution uses DynamoDB and Streams to capture new items, Lambda to route workload into subject-wise SQS queues and ECS Fargate workers to orchestrate Amazon Bedrock (explanations and embeddings) and Amazon SageMaker (custom inference) before storing enriched outputs into an Amazon S3 data lake. Everything is governed and monitored with CloudWatch, IAM and CI/CD.

KEY OUTCOMES DELIVERED

- **Real-time ingestion to AI enrichment** via DynamoDB, Streams, Lambda, SQS and ECS Fargate, enabling low-touch, scalable processing.
- **Higher resilience and throughput** with subject-wise SQS queues and DLQs to isolate failures and prevent pipeline back-pressure.
- **Accurate, explainable enrichment** by combining Bedrock for generation and embeddings and SageMaker endpoints for domain models.
- **Clean data lifecycle** in an S3 data lake (raw, processed, enriched, analytics) with versioning and Athena access for analysis.
- **Operational clarity and CI/CD velocity** using CloudWatch (metrics, logs, alarms) and CodeCommit, CodeBuild and CodePipeline for automated builds and deployments.
- **Defined DR posture** with Terraform-driven rebuild, DynamoDB PITR, S3 versioning and documented RTO and RPO targets.

ABOUT THE CLIENT AND THEIR MISSION

eSaral is one of India's most trusted online and offline coaching platforms for JEE and Classes 8 to 10, helping students achieve top ranks through live and offline classes, personal mentorship, expert doubt solving, comprehensive study material and advanced test series. Learning is delivered by Kota's experienced IITian and Doctor faculty, with a proven system built to produce a high selection ratio and admissions into IITs, NITs and AIIMS.

Driven by a mission to make high-quality exam preparation accessible and affordable, eSaral pairs a five-layered mentorship model with a large, pattern-based question bank and AI-powered test analysis. To keep enriching that question bank at scale, eSaral needed a reliable, automated pipeline to ingest large volumes of JEE questions and augment each one with explanations, embeddings and difficulty scoring.

ARCHITECTURE

eSaral needed to ingest large volumes of questions reliably, decouple enrichment steps to avoid bottlenecks and blend foundation models with custom ML, while ensuring least-privilege access, observability and a repeatable CI/CD path. Additional requirements included failure isolation with DLQs, quota-aware Bedrock and SageMaker usage, and a clear DR plan with verifiable end-to-end tests.

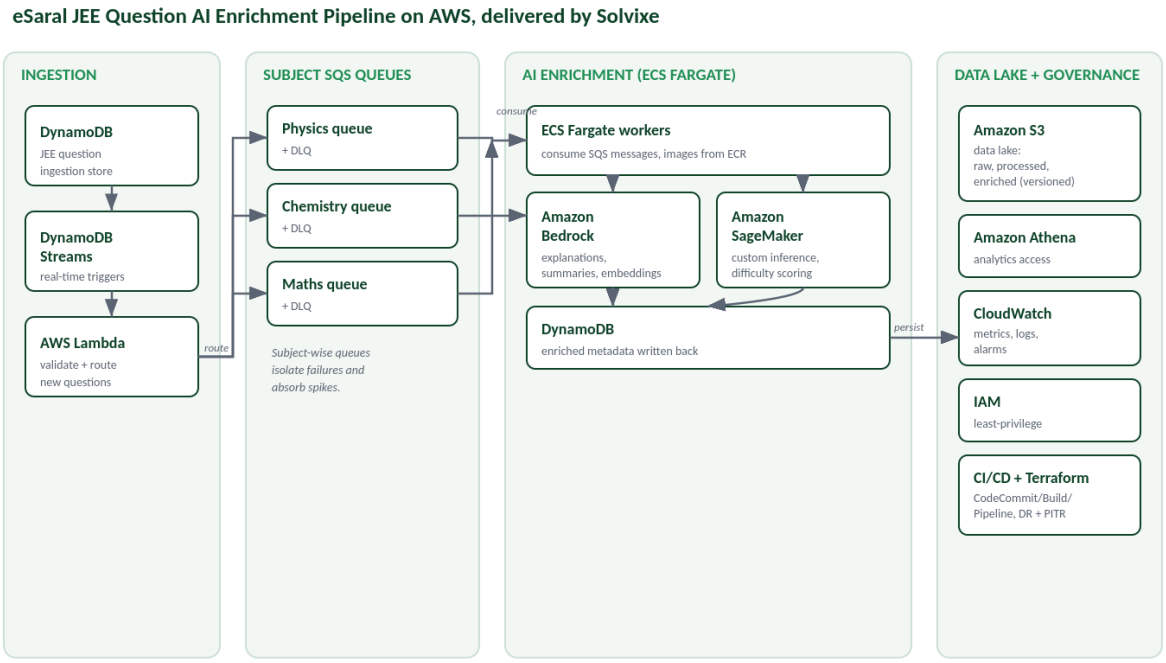

Figure 1: eSaral JEE question AI enrichment pipeline on AWS, delivered by Solvixe.

WHAT SOLVIXE BUILT

- Built a modular, event-driven JEE assessment pipeline on AWS using DynamoDB for question ingestion and DynamoDB Streams for real-time triggers.
- Used AWS Lambda to validate new questions and route them into subject-wise SQS queues (Physics, Chemistry and Maths) with DLQs for failure isolation and smooth scaling.
- Ran ECS Fargate workers (images in ECR) to consume SQS messages and orchestrate AI enrichment via Amazon Bedrock (explanations, summaries and embeddings) and Amazon SageMaker (custom inference such as difficulty scoring).
- Stored enriched metadata back in DynamoDB and persisted datasets in a versioned Amazon S3 data lake with clear raw, processed and enriched layers for auditability and analytics.
- Automated builds and releases with CI/CD using CodeCommit, CodeBuild and CodePipeline for Lambda packaging, container builds and ECS deployments.
- Monitored health and performance with CloudWatch (Streams iterator age, SQS backlog, ECS CPU and memory, Bedrock and SageMaker latency) and secured access with least-privilege IAM.
- Enabled fast recovery using Terraform-driven DR backed by DynamoDB PITR, S3 versioning and reusable model artifacts for quick SageMaker endpoint redeployments within defined RTO and RPO.

RESULTS AND IMPACT

- **Scalable, decoupled processing:** subject-wise SQS queues with DLQs absorb spikes, isolate failures and keep enrichment moving without blocking upstream ingestion.
- **Better quality with composable AI:** Bedrock enables high-quality generation (explanations and embeddings) while SageMaker serves custom inference such as difficulty, producing richer metadata for each question.

- **Faster, safer releases:** a first-class CI/CD pipeline automates builds, tests and blue/green style updates for Lambda and Fargate tasks, reducing manual errors and time-to-change.
- **Clear operations and governance:** CloudWatch dashboards and alarms track stream iterator age, queue depth, ECS utilization and model latencies, with runbooks defining thresholds and actions.

AWS SERVICES USED

AWS Service	Role in the Solution
Amazon DynamoDB + Streams	Ingests JEE questions and emits real-time triggers on new items to drive the enrichment pipeline.
AWS Lambda	Validates new questions and routes them into subject-wise SQS queues, with no infrastructure to manage.
Amazon SQS + DLQs	Subject-wise queues (Physics, Chemistry, Maths) decouple processing, absorb spikes and isolate failures.
Amazon ECS Fargate + ECR	Serverless workers consume queue messages and orchestrate AI enrichment from container images in ECR.
Amazon Bedrock	Generates explanations, summaries and embeddings for each question.
Amazon SageMaker	Serves custom inference such as difficulty scoring from domain models.
Amazon S3	Versioned data lake with raw, processed and enriched layers for auditability and analytics.
Amazon Athena	Serverless SQL analysis directly over the S3 data lake.
Amazon CloudWatch	Metrics, logs and alarms across Streams, SQS, ECS and model latencies.
AWS IAM	Least-privilege access across every service in the pipeline.
CodeCommit / CodeBuild / CodePipeline	CI/CD for Lambda packaging, container builds and ECS deployments.
Terraform + DynamoDB PITR	Terraform-driven DR with point-in-time recovery, S3 versioning and reusable model artifacts.

ABOUT SOLVIXE

Solvixe is a Select Tier AWS Partner dedicated to delivering end-to-end cloud and technology solutions that empower businesses to innovate and scale with confidence. Our expertise spans software development, cloud infrastructure, CI/CD and DevOps, Generative AI, managed services, and cloud migrations, enabling organizations to build secure, reliable and high-performing digital ecosystems.

CONTACT US

Ready to build or scale your AI pipeline on AWS? Contact the Solvixe team to get started.

support@solvixe.com · **solvixe.com**