

AWS Lambda Implementation for Amtex Systems



Executive Summary

Introduction:

Amtex Systems is a voice-first social media platform that requires a modern data-driven microservices solution for real-time data processing and transformation, along with AI/ML capabilities, to improve customer acquisition and the experience of users speaking multiple languages. With a serverless approach, CloudThat assisted the customer in building an end-to-end data engineering pipeline and machine learning solutions like Audio Transcription, Translation and Sentiment analysis by leveraging AWS services. The project entered production in December 2021.

Customer Challenge:

The client's user base was specific to 2 regional languages (Kannada and Malayalam), and they faced challenges with customer acquisition across other regions in India. They also had issues in real-time data processing, multi-lingual content creation, user interaction, and implementing automated AI/ML solutions efficiently in real-time, leading to suboptimal user experience and difficulties in acquiring new customers. Client wants to increase their userbase which is currently at 50k with only 25% engagement, and 5k active monthly users.

Solution

- AWS Lambda is implemented for building AI/ML solutions such as Audio Transcription, Text Translation, and Sentiment and Key-phrase analysis using Python as the programming language.
- Implemented cross-account access to the primary AWS account from Data Engineering account using Amazon SQS, AWS Lambda. User's metadata and social media insights are stored in Amazon DynamoDB.
- Streaming the data to Kinesis, which further pushes the data to multiple consumers like AWS Lambda, Amazon S3.
- AWS Lambda executes the Python codes for implementing Audio transcription, Translation, and Subtitles (SRT Format) leveraging Amazon Transcribe and Amazon Translate.
- AWS Lambda calls Amazon Comprehend APIs for Sentiment Analysis, Key-phrase extraction of the data and stored this data on Amazon DynamoDB.
- AWS Lambda to query the data using the Amazon Athena from S3 and store results in Amazon S3.
- AWS Lambda is integrated with Slack to provide the real-time updates to the user about the failed and processed tasks.
- Using AWS Lambda and Amazon API Gateway, insights pertaining to user posts, failed tasks, and processed tasks are extracted from Amazon DynamoDB. These insights are then fetched and seamlessly delivered to the client application.
- Insights related to Trending Hashtags are also fetched from Amazon DynamoDB with the help of AWS Lambda and Amazon API Gateway and served to the client application.

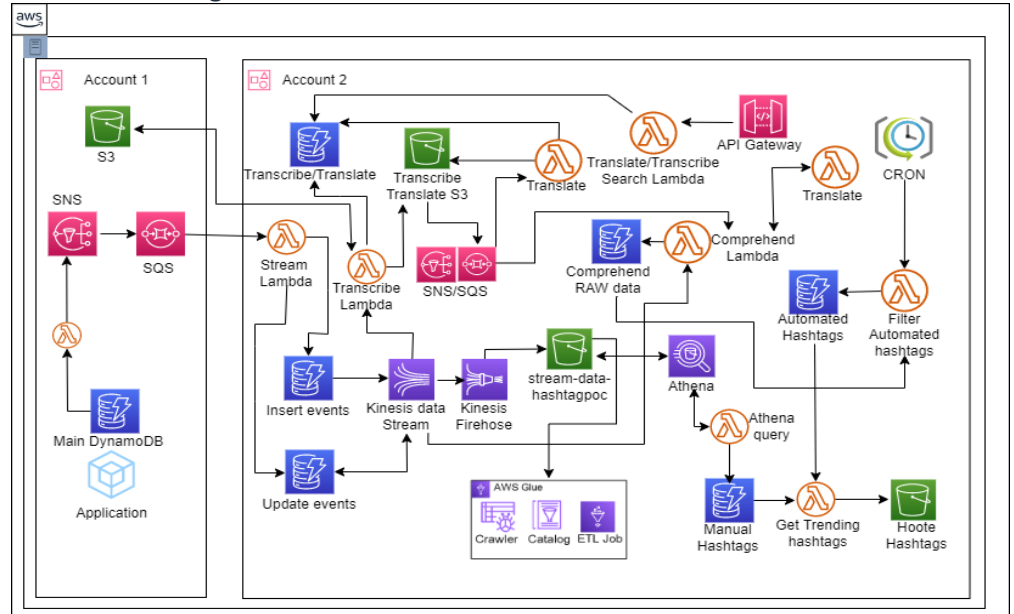
About Amtex Systems



The client's application is a dynamic social media platform catering to a diverse audience from urban and rural areas across India. They have emerged as a destination for individuals seeking to engage in unforced conversations in their native language. Their offerings are designed to provide a seamless and intuitive user experience, with a wide range of features that facilitate social interactions

- The AWS services leveraged in the solution are AWS Lambda, Amazon Kinesis Data Firehose, Amazon Kinesis Data Stream, AWS Glue Service, Amazon Athena, Amazon DynamoDB, Amazon IAM, Amazon SQS, Amazon Transcribe, Amazon Translate, Amazon SNS, Amazon Comprehend, Amazon CloudFront, Amazon CloudTrail, Amazon CloudWatch, and Amazon CloudFormation.

Architecture Diagram



The architecture diagram outlines a centralized data lake solution for data analysis and AI/ML solution. The raw data is transferred between two AWS accounts using Amazon SQS and AWS Lambda. The data is then processed using AWS Lambda which calls Amazon Transcribe to transcribe audio into text, which is further translated into SRT format and stored in Amazon DynamoDB and Amazon S3. Translated text is also sent to comprehend Lambda functions for Key Phrase and Sentiment analysis. Raw Data is also stored on Amazon S3 with proper prefixes. Using AWS Glue Crawler, ETL job is triggered to transform data stored in the Amazon S3 bucket. Amazon Athena is used to query the data and get the trending hashtags. Overall, the solution is serverless, and AWS Lambda is primarily used for automation.

Conclusion:

- Process huge workloads within couple of minutes
- The platform is now able to serve in 6 Indian languages (Hindi, English, Tamil, Telugu, Kannada, and Malayalam) as compared to only 2 languages (Kannada and Malayalam)
- The userbase has increased by 2X, resulting increase in user engagement to 55% and monthly active users 15K

About CloudThat

CloudThat is the official AWS (Amazon Web Services) Advanced Consulting Partner, AWS DevOps Competency Partner, and Microsoft Gold Partner, helping people develop knowledge of the cloud and help their businesses aim for higher goals using best-in-industry cloud computing practices and expertise. We are on a mission to build a robust cloud computing ecosystem by disseminating knowledge on technological intricacies within the cloud space. Our blogs, webinars, case studies, and white papers enable all the stakeholders in the cloud computing sphere.

