Serverless Social Media Analytics using AWS Kinesis

A Project Based Learning Report Submitted in partial fulfilment of the requirements for the award of the degree

of

Bachelor of Technology

in the Department of Computer Science & Engineering

Cloud Based AI/ML Speciality (22SDCS07A)

Submitted by

2210030074 : Mikkilineni Namitha

Under the guidance of

Ms. P. Sree Lakshmi

Department of Computer Science and Engineering

Koneru Lakshmaiah Education Foundation, Aziz Nagar

Aziz Nagar – 500075

March - 2025.

Introduction

Social media platforms generate an immense volume of real-time data, including posts, comments, likes, shares, and reactions. Analyzing this data in real time is crucial for businesses, marketers, and organizations to understand customer behavior, detect emerging trends, and respond proactively to potential issues. This project focuses on designing a serverless social media analytics system using AWS Kinesis to efficiently ingest, process, and analyze live streaming data. The use of a serverless architecture ensures that resources scale automatically, reducing operational costs while providing high availability and reliability. By leveraging AWS services, the system will process large-scale data without requiring complex infrastructure management. The solution enables businesses to extract meaningful insights from social media interactions, track audience engagement, and improve decision-making. Additionally, the system can detect anomalies, such as sudden spikes in user engagement or potential security threats, helping organizations maintain a proactive approach toward user interactions and online presence

AWS Services

To ensure real-time, scalable, and cost-effective data processing, the project integrates several AWS services:

1. **AWS Kinesis**: Facilitates real-time ingestion of streaming social media data, ensuring low-latency data capture.
2. **AWS Lambda**: Performs serverless, event-driven processing of streaming data, automatically scaling based on incoming events.
3. **Amazon S3**: Acts as a centralized data lake for storing raw and processed social media data efficiently.
4. **Amazon DynamoDB**: Provides a fast and scalable NoSQL database for storing structured insights and metadata.
5. **Amazon QuickSight**: Enables interactive dashboards and real-time data visualization for users to gain meaningful insights.

Project Purpose and Expected Outcome

The goal is to develop a real-time social media analytics system that helps businesses monitor brand engagement, track sentiment, and detect trends efficiently. Using AWS serverless architecture, the system dynamically processes large volumes of data, providing instant insights into audience behavior. It ensures scalability, automates analytics, and helps optimize marketing strategies. Additionally, it detects anomalies like spam trends or unusual engagement patterns, enabling proactive responses.

Methodology

Architecture and Workflow

The architecture consists of Amazon Kinesis for real-time data streaming, AWS Lambda for processing and analytics, Amazon S3 for storage, Amazon DynamoDB for metadata management, and Amazon QuickSight for visualization. The workflow includes:

- Social media data is collected from platforms via APIs and sent to Amazon Kinesis Data Streams.
- Kinesis acts as a real-time ingestion layer, ensuring continuous data flow.
- AWS Lambda functions are triggered by Kinesis to process incoming data.
- Lambda applies filtering, transformation, and anomaly detection before forwarding the data.
- Processed data is stored in Amazon S3 for long-term analysis and archival.
- Metadata and indexing information are maintained in Amazon DynamoDB for quick lookups.
- Amazon Kinesis Data Analytics performs real-time aggregation and anomaly detection.
- Processed insights are sent to Amazon QuickSight for visualization and reporting.
- Amazon CloudWatch tracks data pipeline performance, Lambda execution times, and system health.
- AWS IAM ensures secure access control and authentication across the architecture.

AWS Services Interaction

- Amazon Kinesis streams social media data in real time to AWS Lambda for processing.
- AWS Lambda filters, aggregates, and transforms data before storing insights in Amazon S3 and DynamoDB.
- Amazon S3 retains raw and processed data for historical analysis, while DynamoDB enables efficient querying.
- Amazon QuickSight visualizes insights from S3 and DynamoDB through interactive dashboards.

Justification for AWS Service Selection

These services were selected due to their seamless integration, high scalability, and cost efficiency. AWS Kinesis ensures high-throughput streaming, while Lambda provides automatic event-driven processing, eliminating the need for manual intervention. QuickSight enhances the system by offering visual analytics, making it easier for stakeholders to interpret results.

Implementation

AWS Infrastructure Setup

- Create an Amazon S3 Bucket – Set up an S3 bucket to store raw and processed social media data with versioning and access control.
- Set Up an Amazon Kinesis Data Stream – Configure a Kinesis stream to ingest real-time social media data and define retention policies.
- Deploy AWS Lambda Functions – Write Lambda functions to process data from Kinesis, apply transformations, and trigger automated workflows.
- Set Up Amazon DynamoDB – Create a DynamoDB table to store metadata, optimize query performance, and enable auto-scaling.
- Configure Amazon QuickSight for Data Visualization – Connect QuickSight to S3 and DynamoDB to display analytics in interactive dashboards.
- Implement Amazon CloudWatch for Monitoring – Enable system monitoring, log tracking, and configure alerts for performance anomalies.

Security Policies, IAM Roles, and Access Controls

- Define IAM Roles & Policies – Assign restricted access roles to AWS services following the least privilege principle.
- Secure Kinesis and S3 Data – Enforce encryption and access controls to protect streaming and stored data.
- Enforce Lambda Execution Security – Restrict function permissions and store sensitive credentials securely.
- Enable Data Encryption and Backup – Encrypt DynamoDB and S3 data at rest and set up automated backups.

These security measures safeguard the system against data breaches while ensuring controlled and secure access.

Automation

Automation is key to ensuring the system remains efficient and self-sustaining:

- Automate Data Streaming & Processing – Configure Kinesis and Lambda for real-time data ingestion and transformation.

- Set Up Auto-Scaling for AWS Services – Enable automatic scaling for Kinesis, Lambda, and DynamoDB to handle varying workloads.
- Enable CloudWatch Logging & Alerts – Track performance metrics and configure alarms for failures.
- Implement Lifecycle Policies for Data Management – Automate data archiving and deletion for optimized storage management.