YouTube Trending Video Analytics Project

Overview

This project builds an end-to-end data pipeline to analyze trending YouTube videos. It uses AWS Lambda for automated data extraction, AWS Glue for ETL, Athena for querying, and Tableau for interactive visualizations. The goal is to uncover insights such as top trending videos, engagement levels, country-wise performance, and channel effectiveness.

Tools & Technologies

- AWS Lambda: Scheduled data extraction from YouTube API
- Amazon S3: Storage for raw and processed data
- AWS Glue: ETL jobs to clean and transform JSON into Parquet
- AWS Glue Crawler: Automatic schema discovery
- Amazon Athena: SQL querying over S3
- Tableau: Dashboard creation and visualization

Step-by-Step Project Workflow

1. Data Extraction with AWS Lambda

A Python script scheduled in AWS Lambda every 6 hours fetches trending video data from the YouTube API. It stores the data as JSON in an S3 bucket under the 'raw/' folder.

2. Data Transformation with AWS Glue

A Glue Job reads the JSON data from 'raw/', flattens nested structures, cleans the data, and converts it into Parquet format. The output is stored in the 'processed/' S3 folder.

3. Schema Discovery with AWS Glue Crawler

The crawler scans the 'processed/' folder and creates a table in the Glue Data Catalog, making the data available for Athena queries.

4. Querying with Amazon Athena

Athena uses SQL to query the structured Parquet files in S3. A table named 'processed' in the 'youtube_trending_db' database is queried for analysis.

5. Visualizations in Tableau

Tableau is connected to Athena using the ODBC driver. Five main worksheets are created:

- Top Trending Videos
- Views by Country
- Views Over Time
- Engagement Bubble
- Channel Performance

6. Dashboard Creation

The sheets are arranged in a dashboard with filters for country and published date. Interactive charts and color-coded metrics are used for clarity.

7. Export & Sharing

The final dashboard is exported as a PDF and optionally uploaded to Tableau Public for sharing.