Key things

Tuesday, 5 August 2025   11:36 PM

- Storage of vedio's -
  - Object storage service for storing and retrieving any amount of data from the web.
    - Scalable - **Automatically handles unlimited data** and traffic without manual provisioning.
    - Resilient - **11 nines (99.999999999%) durability** by replicating data across multiple Availability Zones.
    - Secure - **End-to-end encryption** and **fine-grained access control** using IAM, bucket policies, and monitoring tools.
- Better DB for large number of call's

Kafka

Kafka is a MQ + DB which is event based streaming for real time logs or click events, this can be consumed by batch for real time analytics and stored in hadoop and elastic search can be used to do fuzzy logic search on these events.

Kafka -> Batch consumer -> Hadoop -> Elastic search (Visualisation)

Elasticsearch is Better for "Last Number of Calls" Use Cases

Criteria	Elasticsearch	RDBMS (e.g., MySQL, PostgreSQL)
Data Model	Document-based, optimized for logs/events	Relational, normalized schema
Query Speed	Very fast for recent/time-based queries using **inverted index + time-based sorting**	Slower as rows grow unless well-indexed
Scalability	Horizontally scalable by default (sharding, replication)	Vertical scaling often required; complex horizontal scaling
Write Throughput	High ingestion rate – ideal for real-time logs or event data	Write-heavy workloads can affect performance
Time-Based Retrieval	Optimized with **timestamp-based indices** (e.g., last 100 API calls)	Requires ORDER BY timestamp DESC LIMIT N which slows with large datasets
Indexing	Data indexed at write time for fast reads	Indexing must be manually tuned
Use Case Fit	Designed for search, logs, analytics	Designed for transactional operations (CRUD)

- Video Storage in Multiple Resolutions
  - Video is uploaded (typically to **Amazon S3**).
  - A **Lambda function or EventBridge** event triggers **AWS Elemental MediaConvert**.
  - MediaConvert transcodes the video into multiple resolutions (e.g. 1080p, 720p, 480p, 360p).
  - Each version is optimized for different devices or network speeds.
  - Each resolution variant is stored again in **S3**, usually organized by folders or prefixes like /1080p/, /720p/, etc.
  - These videos are served using **adaptive bitrate streaming (ABR)** via **HLS or DASH protocols**, allowing seamless switching between resolutions during playback.

- To see any adult content or figure out the cast in a vedio:
  - **Amazon Rekognition** is a deep learning-based image and video analysis service.
  - **Detects faces**, **labels**, **objects**, **text**, **celebrities**, **inappropriate content**, etc.
  - Can track people, count them, detect emotions, and more.
  - JSON-based metadata is generated with time-stamped labels and detections.
  - This metadata is stored (often in S3 or DynamoDB) and used for:
    - Video search
    - Content moderation
    - Thumbnail generation
    - Auto-tagging
    - Ad placement

  NOTE : **AWS Step Functions** to **orchestrate** all the services you mentioned — like **S3 upload**, **MediaConvert**, and **Amazon Rekognition** — and even **run them in parallel**.

- How to deliver across different location in the world
  - **CDN (Content Delivery Network)** plays a **crucial role in delivering video content efficiently** to users across

- Auto-tagging
- Ad placement

NOTE : **AWS Step Functions** to **orchestrate** all the services you mentioned — like **S3 upload**, **MediaConvert**, and **Amazon Rekognition** — and even **run them in parallel**.

AWS Step Functions is a **serverless orchestration service** that lets you coordinate multiple AWS services into a **workflow** using **state machines**. It's ideal for building **complex, event-driven pipelines**.

- How to deliver across different location in the world
  - **CDN (Content Delivery Network)** plays a **crucial role in delivering video content efficiently** to users across different geographic locations
  - A **CDN** is a globally distributed network of **edge servers** that **cache and serve content** (like images, videos, HTML, CSS, JS) **closer to the end user**.
    On AWS, the primary CDN service is **Amazon CloudFront**.