

CLOUD ARCHITECTURE

FLICK

RICHIK MUKHERJEE

CLOUD ARCHITECTURE PORTFOLIO

Contents

Mission	2
Objective	2
Vision Diagram	2
Data Sources	2
Proprietary Content Repository	3
Live Streaming Feeds	3
Third-Party Vendors	3
External Metadata Providers	3
Cloud Architecture.....	3
Data Sinks	4
AI Systems	4
End Users	4
Marketing Systems	4
Business Intelligence (Power BI)	4
APIs for External Vendors	4
Proposed Pipeline	5
Strategy for Pipeline Failure	6

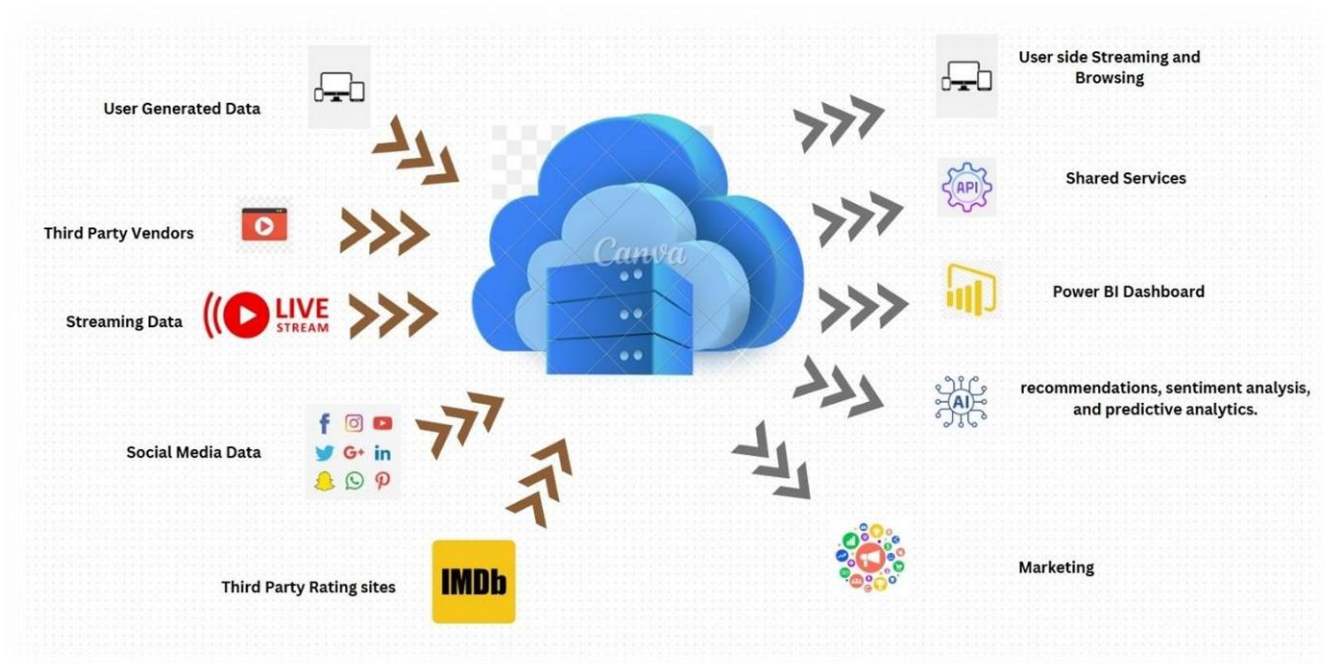
Mission

"To revolutionize entertainment consumption by providing a personalized, accessible, and innovative streaming platform that connects viewers with exceptional content from around the world, empowering users to discover, enjoy, and engage with stories that inspire, entertain, and transform their lives."

Objective

- Content Diversity and Quality
- Personalized User Experience
- Ensure seamless streaming across multiple devices and platforms
- User Engagement and Satisfaction
- Global Accessibility

Vision Diagram



Data Sources

The cloud architecture incorporates multiple data sources to provide a seamless streaming experience and enrich platform functionalities. The primary data sources include:

Proprietary Content Repository

- **Description:** A repository of movies and series produced by the business.
- **Purpose:** Serves as the foundation for on-demand and live-streamed content.

Live Streaming Feeds

- **Description:** Real-time feeds of live events or exclusive premieres.
- **Purpose:** Supports high-quality live content delivery for time-sensitive broadcasts.

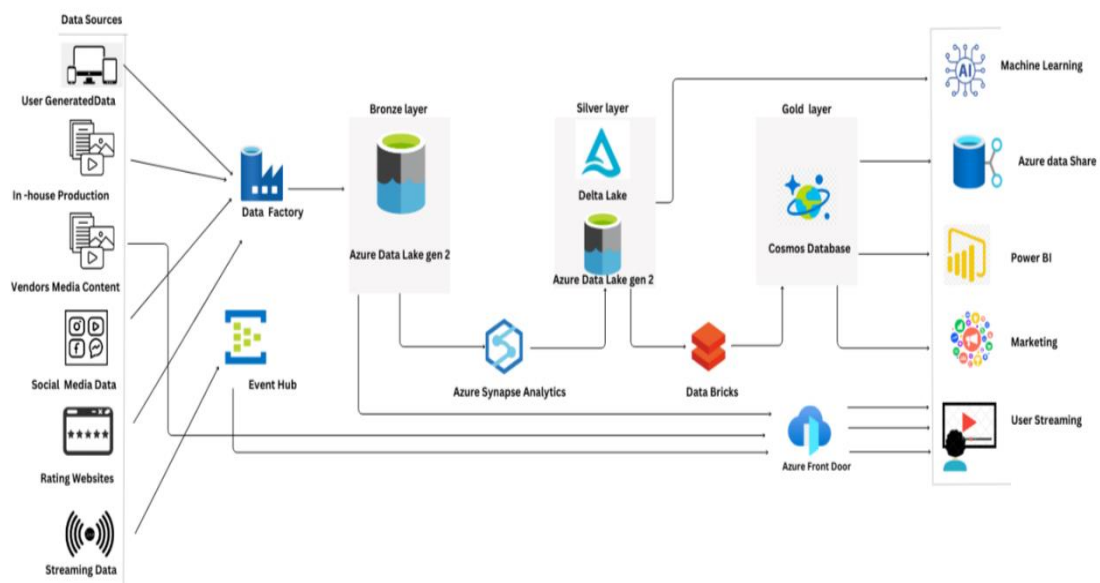
Third-Party Vendors

- **Description:** External providers offering additional movies, series, or live events.
- **Purpose:** Expands the content catalog with diverse and exclusive offerings.

External Metadata Providers

- **Examples:** IMDb ratings and reviews, Social Media Handles
- **Purpose:** Enriches the platform's content catalog by providing additional details and ratings for user engagement.

Cloud Architecture



Data Sinks

The data sinks are designed to distribute processed data effectively to end-users, applications, and stakeholders. The chosen sinks include:

AI Systems

- **Description:** Machine learning models for personalized recommendations, sentiment analysis, and predictive analytics.
- **Purpose:** Enhances user experience by providing tailored content suggestions and insights.

End Users

- **Description:** Subscribers accessing the platform via various devices (smartphones, smart TVs, tablets).
- **Purpose:** Delivers high-quality on-demand and live content through adaptive streaming technologies.

Marketing Systems

- **Description:** Internal marketing tools and CRM systems.
- **Purpose:** Utilize analytics to optimize campaigns, track user behavior, and improve customer retention strategies.

Business Intelligence (Power BI)

- **Description:** Analytical dashboards for internal teams.
- **Purpose:** Provides actionable insights on content performance, user engagement, and platform growth.

APIs for External Vendors

- **Description:** APIs designed for third-party vendors to access ratings, reviews, and performance metrics.
- **Purpose:** Facilitates partnerships and extends the platform's ecosystem by enabling data sharing with trusted partners.

Proposed Pipeline

BRONZE LAYER: RAW DATA STORAGE

- Data Ingestion – The bronze layer is utilized to ingest raw unstructured data from various sources as mentioned in the data sources above (live streaming, proprietary content, third party vendors and social media handles). The data is procured in the data factory from which is shifted to the Azure Data Lake Gen 2.
- Unstructured Data – The bronze layer stores the unstructured data in the rawest form without any type of transformation.
- Storage for future utilization – The data stays in the raw format i.e., JSON, CSV etc. This increases the flexibility in data scalability and accessibility important for future cleansing.
- Establishing foundation for future analysis – It serves as the base foundation for future data cleansing as well as transformation in the Silver Layer.

SILVER LAYER: DATA CURATION

- Data Cleansing and Normalization – The Silver Layer cleans and standardizes the raw unstructured data, it handles missing values, duplications and several other data bottlenecks.
- Transformation with Azure Synapse Analytics – Through the application of business and aggregation using Azure Synapse Analytics, the data is prepared and forwarded for advanced analytics.
- Storage in Delta Lake – The thoroughly cleansed data is carried forward in Delta Lake for supporting ACID transactions as well as time travelling. This stage ensures that the data is structured and ready for analysis.
- Prepared for Analysis – The Silver Layer prepares the data for thorough insights in the Gold Layer.

GOLD LAYER: REFINED DATA FOR ANALYSIS

- Data Transformation – The raw data retrieved from the Bronze Layer is thoroughly cleansed, transformed and enriched through the use of Azure Synapse Analytics as well as Databricks.
- Structured and Processed Data – The Gold Layer refines the data into usable format making it structured and enriched for a thorough analysis as well as preparing the data for reporting to the final users. It stores the data in the Delta Lake for ACID transactions, it ensures the data is consistent and reliable.
- Business Insights – It provides a high quality, structured data set for AI/ML modelling, Business Intelligence as well as reporting.
- Data Storage – The data is stored in Cosmos DB and Delta Lake, it enables thorough data scalability, faster accessibility for analysis and downstream applications.

Strategy for Pipeline Failure

- Early Monitoring – Emergency alert mechanisms to flag errors or failures in the pipeline to generate fast solutions.
- Retrying mechanisms – Automated retries listed for flagging and working on transient mishaps.
- Fallback and Graceful Degradation – Consistent maintenance of service availability during failures.
- Rollback and Immediate Recovery – Entrusting consistent data status as well as recovery from persistent failures.
- Scalable Resources – Scaling resources whenever needed for handling and ensuring doubled load or bottlenecks in the pipeline.
- Consistent Improvement – Curating and analyzing failures for improving process formation for pipeline.

**THANK
YOU**