Google Cloud

PROFESSIONAL CLOUD ARCHITECT

Altostrat Media Case Study

Company Overview

Altostrat is a prominent player in the media industry, with an extensive collection of audio and video content that comprises podcasts, interviews, news broadcasts, and documentaries. Their success in delivering premium content to a diverse audience requires a content management system that can keep pace with the dynamic media landscape.

Solution Concept

Altostrat seeks to modernize its content management and user engagement strategies using Google Cloud's generative AI. They want a platform that empowers customers with personalized recommendations, natural language interactions, and seamless self-service support. Simultaneously, they want to drive revenue growth through dynamic pricing, targeted marketing, and personalized product suggestions.

The seamless integration of AI-powered tools into their existing Google Cloud environment will enable Altostrat to efficiently manage their vast media library, enhance user experiences, and unlock new revenue streams. Google Cloud's generative AI will solidify their leadership in the media industry.

Existing Technical Environment

GKE

Cloud Storage

Altostrat's content management and delivery platform leverages GKE for scalability and high availability, essential for handling their vast media library. Their extensive media library, spanning various documents, audio and video formats, is stored in Cloud Storage. To gain valuable insights into user behavior, content consumption patterns, and audience demographics, Altostrat leverages BigQuery as their primary data warehouse. Additionally, they use Cloud Run functions for serverless execution of event-driven tasks such as video transcoding, metadata extraction, and personalized content recommendations.

BigQuery

CloudRun

Event Driven Processing

Cloud Storage upload events invoking Cloud Run functions for transcoding

GKE On prem Anthos

While Altostrat has made significant strides in cloud adoption, they also maintain some legacy on-premises systems for specific workflows like content ingestion and archival. These systems are slated for modernization and migration to Google Cloud in the near future. User management and authentication are currently handled through a combination of Google Identity and third-party identity providers. For monitoring and observability, Altostrat relies on a mix of native Google Cloud tools like Cloud Monitoring and open-source solutions like Prometheus, with alerts primarily delivered via email notifications.

IAM

Proxy

Stackdriver

Business Requirements

- Accelerate and enhance the reliability of operational workflows across all environments. [Google Cloud + On-premises] Anthos GKE on prem
- Simplify infrastructure management for rapid application deployment. Autopilot Cloud run

GKE
Cloud Build
Cloud Deploy

- Optimize cloud storage costs while maintaining high availability and scalability for media content. Cost Optimize by Gemini Cloud Recommendation
- Enable natural language interaction with the platform with 24/7 user support. Conversational Agents (Dialogflow) to create intelligent virtual agents
- Automatically generate concise summaries of media content. Utilize Natural Language AI for summarization:
- Extract rich metadata from media assets using NLP and computer vision. Natural Lang AI+Video AI + Vision AI pre-trained models for labels, objects, explicit content detection
- Detect and filter inappropriate content.
- Analyze media content to identify trends and extract insights. Analytics -> BigQuery
- Inform content strategy and decision-making with data.

masking Data sensitive: Securing AI (e.g., Model Armor, Sensitive Data Protection, and secure model deployment)

This solution leverages Video AI to extract insights from media content, enabling the identification of trends and extraction of valuable information.

Technical Requirements

- Modernize CI/CD for containerized deployments with a centralized management platform. Cloud Deploy Cloud Build
- Secure, high-performance hybrid cloud connectivity for data ingestion. Anthos networking : Dedicated

https://cloud.google.com/solutions/ccai-insights?

Conversational Insights to analyze customer interaction data, identifying key topics, sentiment, and common issues. This raw data is then visualized using Looker Studio, enabling users to create interactive dashboards and reports for actionable insights.

- Provide scalable, performant kubernetes environments both on-premises and in the cloud. GKE on prem
- Optimize cloud storage costs for growing media volumes.
- Design AI-powered detection of harmful content.
- Ensure that AI systems are auditable and their decisions can be explained.
- Leverage LLMs and conversational AI for personalized experiences and content virality.
- Develop advanced chatbots with natural language understanding to provide personalized assistance.
- Automated summarization for diverse media.

Vertex AI platform would be the central hub for developing, training, and deploying custom machine learning models specifically designed to detect harmful content.

Data lifecycle managem
Archival class for data accessed < 1

Vertex Explainable AI would be crucial for understanding why your AI systems make specific decisions, providing insights into model behavior and helping to debug and improve their performance.

Conversational Agents
Build personalized conversational AI

Design and build your chatbot using Conversational Agents (Dialogflow), defining intents, entities, and fulfillment logic to handle natural language understanding and personalized responses

Develop summarization models using Vertex AI

Use Workflows to Orchestrate summarization pipelines
Utilize Workflows to design and define the steps for the summarization process. This could include steps for input processing, calling the Vertex AI platform summarization models, and output handling.

Generic summaries : Start with Natural Language API for general

precisely tailored to your summarization or domain-specific summaries : Vertex AI build your own model

Executive Statement

At Altostrat, we are embracing the next frontier of artificial intelligence to revolutionize our content strategy. By harnessing the power of generative AI, we will create an unparalleled user experience by empowering our audience with intelligent tools for content discovery, personalized recommendations, and seamless interaction. Reliability and cost management are our top priorities. This strategic initiative will deepen engagement, foster customer loyalty, and unlock new revenue streams through targeted marketing and tailored content offerings. We see a future where AI-driven innovation is central to our business, leading to greater success for our company and delivering exceptional value to our customers.

Sample Question 1:
Altostrat wants to automate metadata extraction and content moderation for uploaded media. What is the most efficient solution?
A. Export media to on-premises GPU servers running custom CV/NLP models for analysis.
B. Use Google Cloud's Video AI and Vision AI APIs to analyze each media file on upload, extracting labels and detecting inappropriate content.
C. Train and deploy a custom TensorFlow model on GKE for video and image analysis using Altostrat's own data.
D. Use Cloud Functions triggered on uploads that call the Video Intelligence API for metadata, and a third-party API for content moderation.

D adds Unnecessary complexity
Mixing native + third-party increases risk and cost

A is High CapEx, heavy opex violets most

C is Requires: Training data, Model tuning

when you alreayd have managed server why do manual work

B If the use case is common AI capability (labels, moderation, OCR, speech, vision)

Use Google's pre-trained AI APIs