CASE STUDY

Enhancing Data Management Efficiency in IP Law Firm Operations

CONTATA

OVERVIEW

The client is a Minnesota-based intellectual property (IP) law firm offering strategic counseling services to multinational corporations, middle-market businesses, startups, universities, and individuals.

CHALLENGES

The project aimed at providing comprehensive operational support to the client's team for seamless collection, processing, and transformation of raw data into user-friendly formats. The client sought a versatile system capable of generating customized data presentations tailored to the specific requirements of diverse use cases within the team. With a continuous influx of raw data files, ensuring timely delivery of the latest data daifferentials via both visual representations and APIs was also crucial.

INDUSTRY
- Legal
- IP

SERVICES
- Data Management
- Data Ingestion
- Data Orchestration

TECHNOLOGY
- Apache Beam Python SDK
- Google Dataflow
- Pub/Sub
- BigQuery
- Postgre SQL

SOLUTION

Our solution employed a meticulously crafted data ingestion pipeline designed to ensure continuous availability of data, prioritizing ingestion and processing of latest datasets. Following are some of the key tools and technologies, which were central to our solution:

- Apache Beam Python SDK
- Google Dataflow
- Pub/Sub and Cloud Functions
- BigQuery
- Postgre SQL

The Apache Beam Python SDK was integrated into the data infrastructure specifically for executing batch processing tasks, operating within the Google Dataflow framework. This SDK served as the foundational component of our ingestion pipeline, facilitating seamless handling and batch processing of large-scale data. By leveraging the expressive features of Python alongside the scalability and fault tolerance mechanisms of Google Dataflow, our ingestion pipeline efficiently orchestrated data ingestion, transformation, and storage processes, ensuring prompt delivery of accurate and up-to-date data for downstream analysis and application purposes.

We leveraged Google Dataflow's capabilities to efficiently unzip incoming raw files, which typically contained thousands of files in each zip archive. It orchestrated the ingestion of data from these files into a unified centralized view. By distributing this task across various resources, Google Dataflow ensured its timely completion, optimizing our data processing workflows.

The integration of Pub/Sub and Cloud Functions was also a pivotal aspect of our data ingestion pipeline, operating within a microservices architecture paradigm. This integration enabled event-driven data processing, ensuring our pipeline executed processing only when triggered by relevant events, thereby enhancing cost-effectiveness and operational efficiency.

We utilized BigQuery to maintain tables directly as external tables, facilitating real-time visibility of data as soon as it was dumped in Google Cloud Storage by our ingestion pipeline in JSON format. This approach minimized unnecessary computing of resources typically associated with ingesting latest data. Moreover, BigQuery was utilized for storing logging information drilled down to each file level, facilitating efficient debugging processes.

To store operational data, we combined the capabilities of BigQuery with Cloud SQL (PostgreSQL), ensuring a comprehensive and scalable solution for managing both analytical and operational data within the ecosystem.

BENEFITS

Real-Time Data Availability
The solution ensured the availability of the latest data in real-time and provided up-to-date information, empowering the client to make timely and informed decisions.

Customized Views for Varied Use Cases
Customized views ensured that each team member has access to the relevant information needed for their tasks and responsibilities.

Scalability and Efficiency
The solution provided a scalable and efficient way to manage data within the client's ecosystem, facilitating seamless operations and analytics.

About Contata

Contata Solutions is a trusted leader in technology and digital innovation. Through our work in data engineering, data analytics, machine learning, marketing automation and app development, we deliver solutions that address complex problems in ways that are simple, insightful and impactful.

Our promise and value proposition to our customers is simple: we leverage our deep technical expertise and global presence to bring software products and data-driven decision capabilities to life.

Founded in 2000, Contata is a privately-held company headquartered in Minneapolis that serves clients globally from offices in the United States and India.