

Google Cloud Platform

Para Big Data

Christian Barrios y Mayco Catellanos

Contenido

- | | | | |
|----|--------------|----|--------------|
| 1. | Introducción | 4. | Demo |
| 2. | Servicios | 5. | Conclusiones |
| 3. | Casos de Uso | | |



Google Cloud

Introducción

GCP es la plataforma de servicios en la nube de Google

Ofrece soluciones avanzadas en computación,
almacenamiento, análisis de datos e inteligencia
artificial

Permite escalar recursos según las necesidades del
proyecto

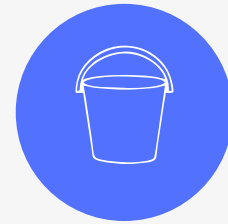
Garantiza una infraestructura global con alta
disponibilidad, baja latencia y seguridad robusta

Se actualiza constantemente, facilitando la adopción
de nuevas funcionalidades para mantener la
competitividad.

Servicios

Servicio GCP	Equivalente Open Source
Cloud Storage	HDFS / Hadoop
BigQuery	Hive
Pub/Sub	Kafka
Dataproc	Spark / Hadoop
Bigtable	Cassandra
Cloud Data Fusion	

Cloud Storage



Organización en Buckets:

Los datos se agrupan en contenedores llamados buckets.



Acceso mediante API y URL:

Los objetos almacenados son accesibles mediante una URL única y a través de APIs RESTful



Clases de Almacenamiento:

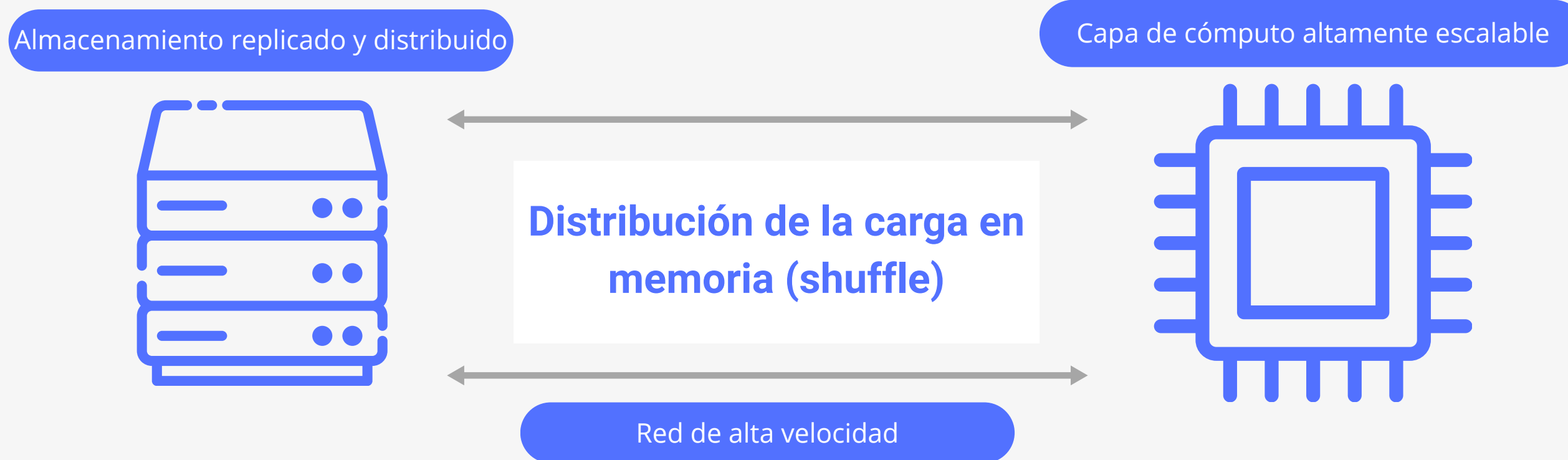
Cloud Storage ofrece diferentes niveles (Multi-Regional, Regional, Nearline y Coldline) que permiten optimizar costos y tiempos de acceso



Seguridad y Durabilidad:

Los datos se replican automáticamente en múltiples ubicaciones geográficas

Big Query



Almacenamiento Columnar

Los datos se guardan en formato columnar, lo que agiliza la lectura y reduce costos

Ejecución de Consultas SQL

Utiliza SQL estándar para ejecutar consultas sobre grandes volúmenes de datos

Escalabilidad y Distribución

BigQuery se encarga de distribuir las consultas en múltiples nodos

Integración y Automatización

Se integra con otros servicios de GCP, permitiendo cargar datos desde diversas fuentes

Pub/Sub

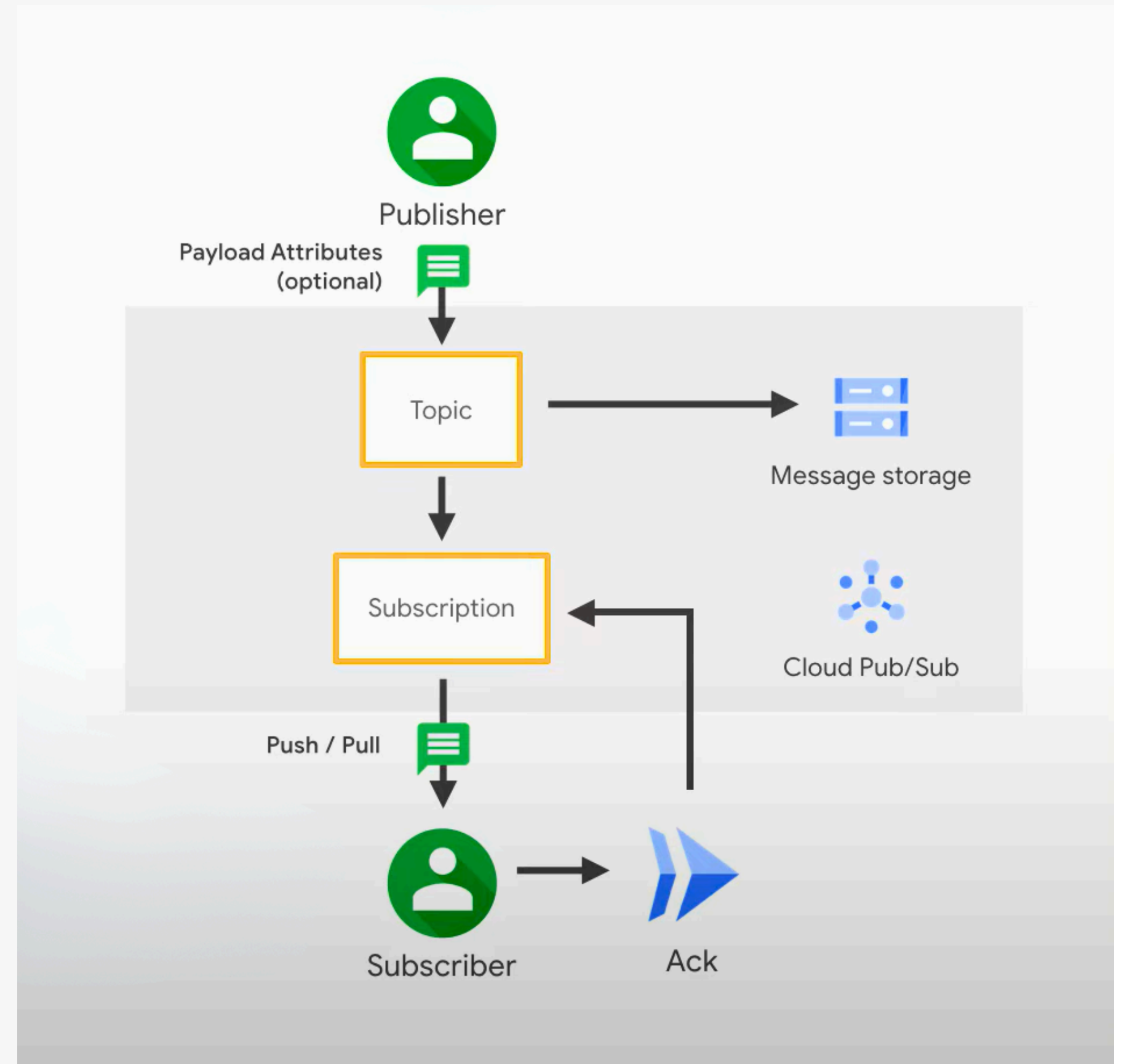
Publisher → Topic: El publisher envía mensajes a un topic

Message Storage: Pub/Sub almacena los mensajes en un sistema distribuido

Subscription: Para recibir los mensajes de un topic, se crea una suscripción

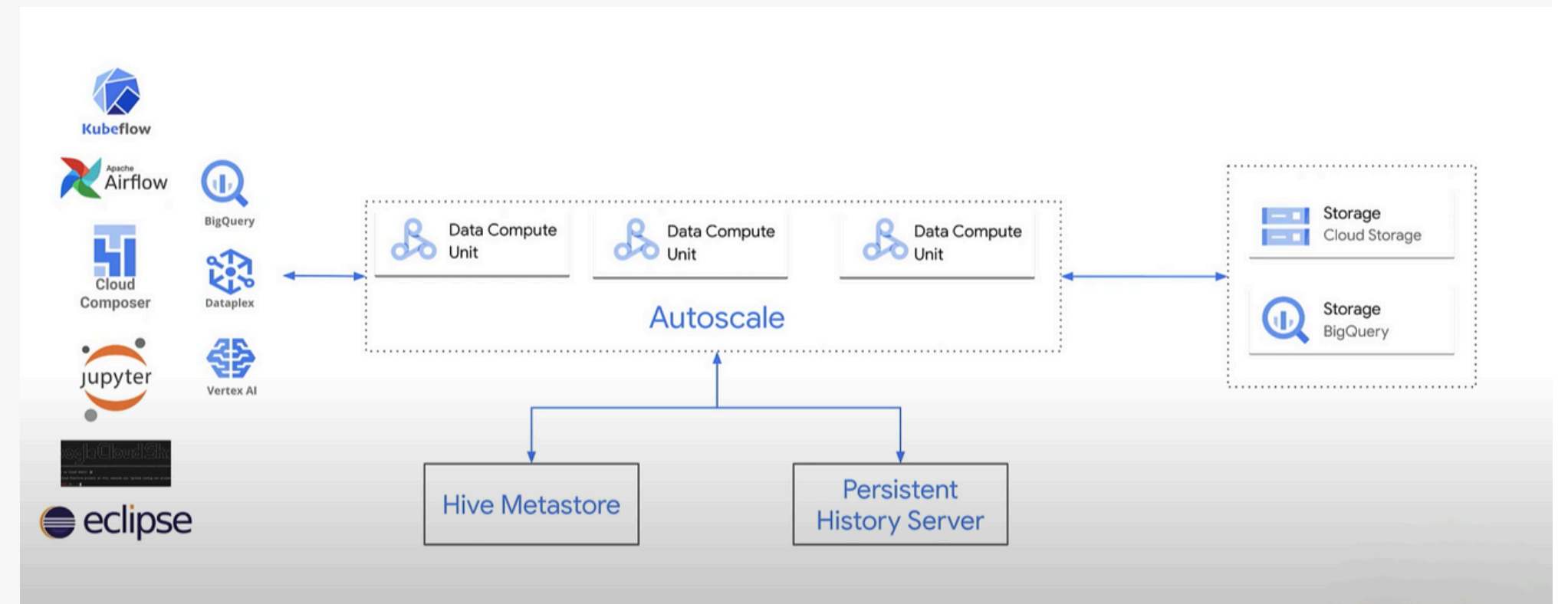
Subscriber: El suscriptor recibe los mensajes de la suscripción, ya sea mediante un pull (el suscriptor solicita los mensajes) o un push (Pub/Sub los envía automáticamente a una URL configurada)

Ack (Reconocimiento): Una vez que el suscriptor procesa el mensaje, envía una confirmación (ack)

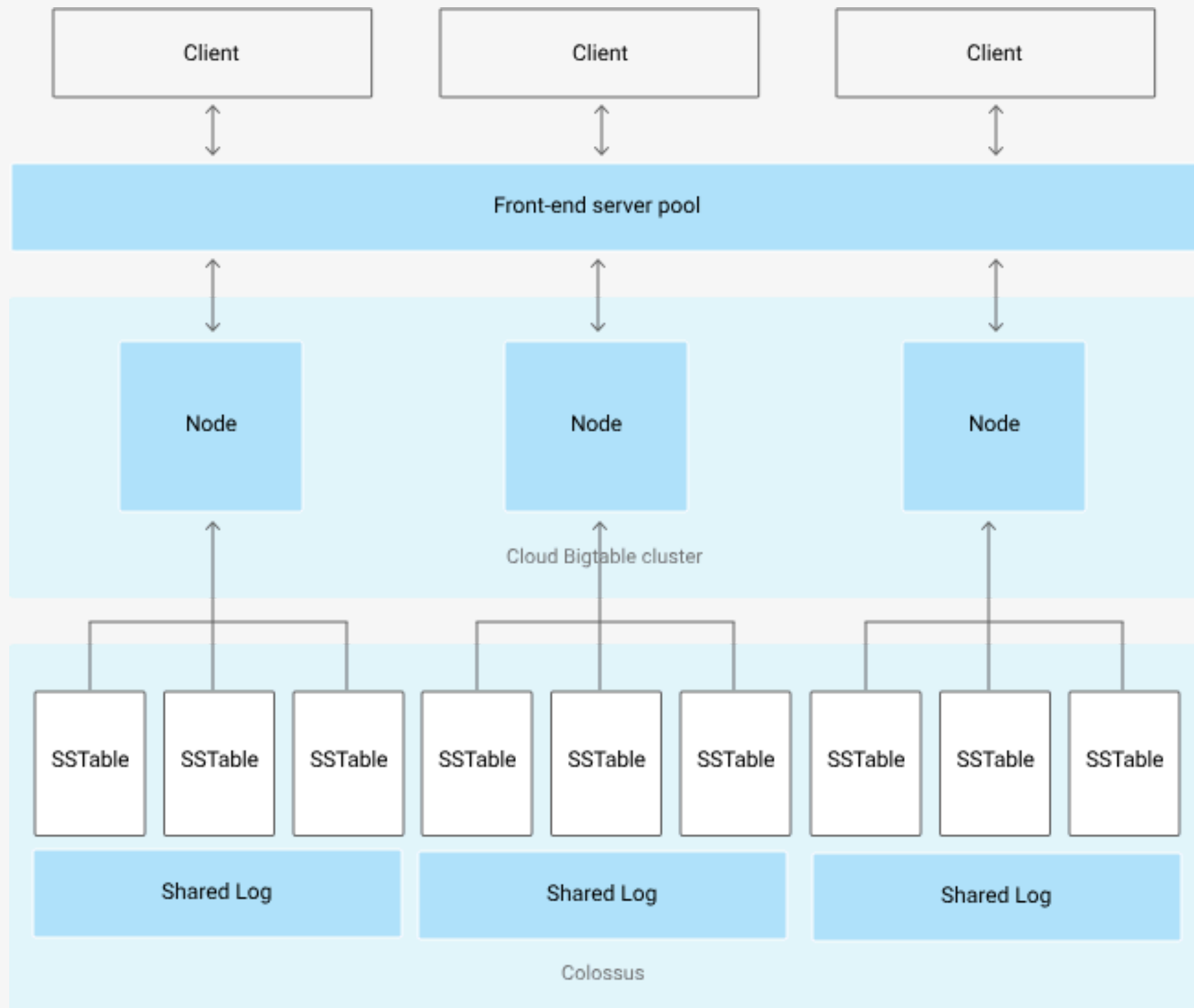


Dataproc

- **Servicio Administrado / Fully-managed:** Lanza y gestiona clústeres de Hadoop y Spark
- **Escalabilidad y Optimización:** Ajusta automáticamente los recursos según la carga de trabajo
- **Integración Nativa:** Se conecta fácilmente con servicios GCP
- **No Lock-In:** Compatible con estándares abiertos
- **Seamless AI/ML:** Integra bibliotecas de machine learning en Spark
- **Enterprise Security:** Ofrece controles de acceso robustos y cifrado de datos



Bigtable



Front-end Server Pool:

- Recibe solicitudes de clientes y las distribuye a los nodos adecuados.

Nodos en el Clúster:

- Cada nodo gestiona una parte de la tabla; se escala agregando o quitando nodos.

Almacenamiento Distribuido:

- Los datos se almacenan en SSTables en Colossus, con un shared log que garantiza la consistencia en las escrituras.

Baja Latencia y Alta Disponibilidad:

- La arquitectura distribuida ofrece respuestas en milisegundos y asegura disponibilidad continua mediante replicación en múltiples zonas.

Cloud Data Fusion

Integración sin código con interfaz gráfica.

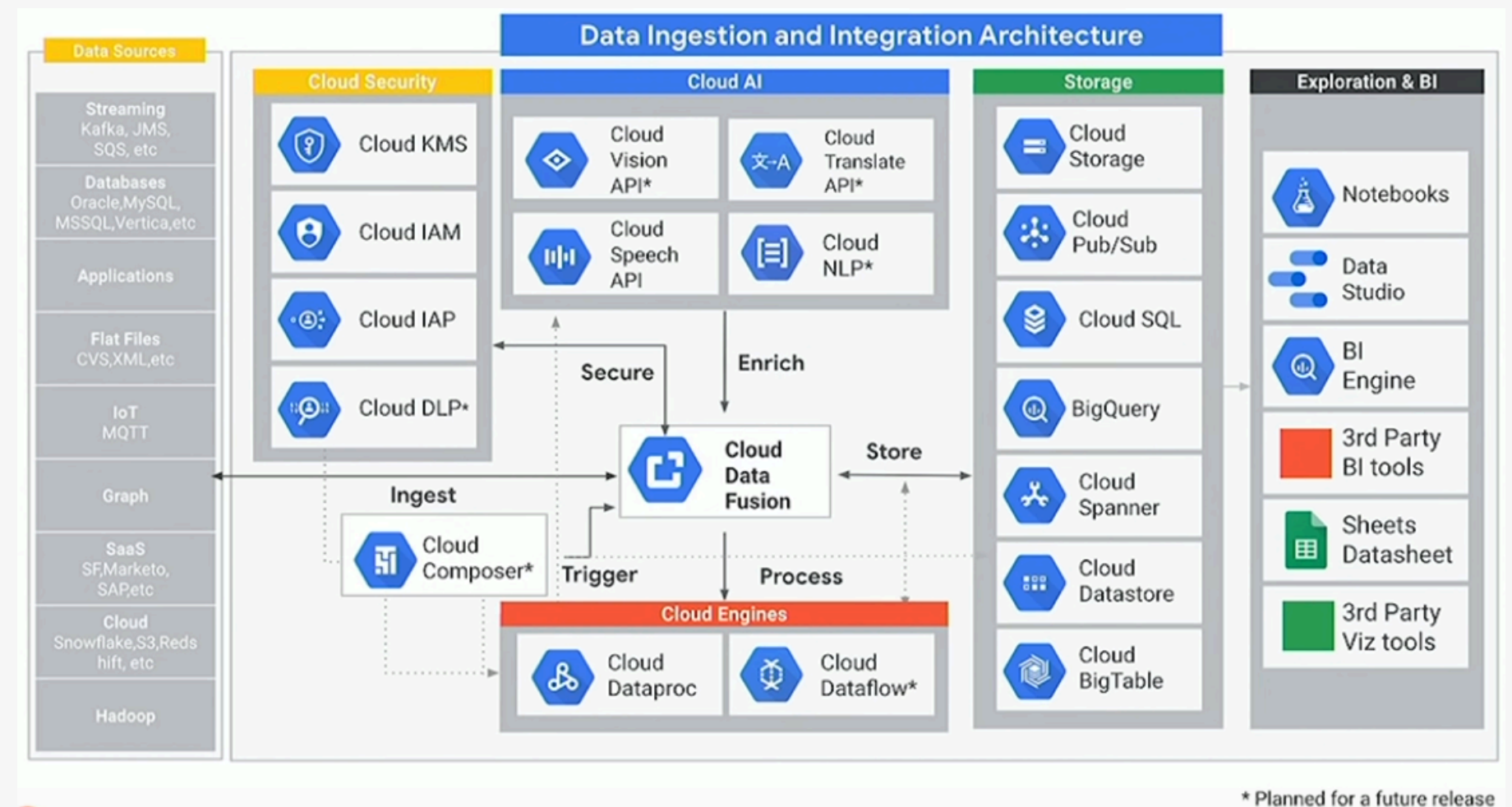
Biblioteca colaborativa para compartir transformaciones.

Arquitectura nativa de GCP: escalable, fiable y segura.

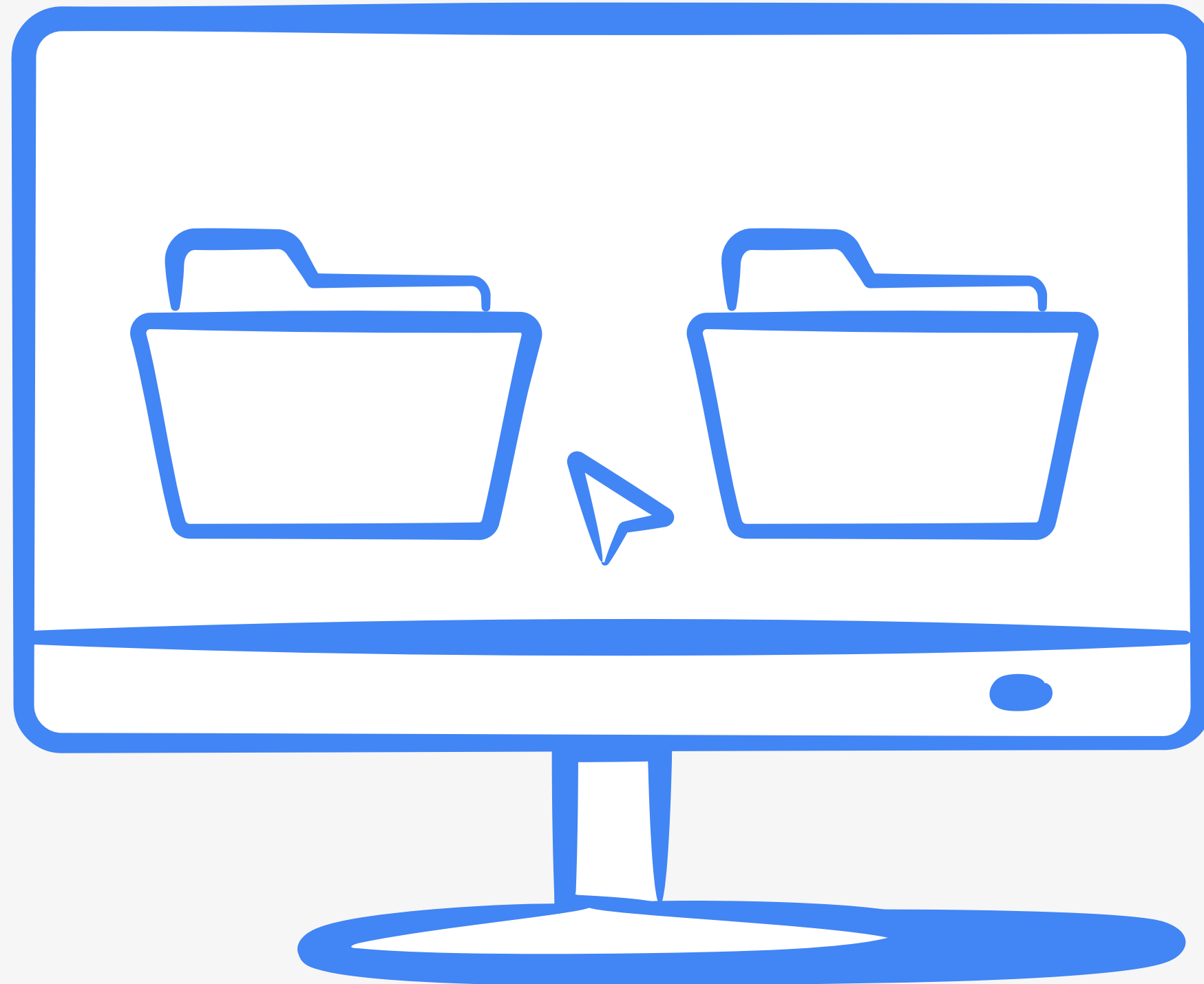
Soporta integración en tiempo real y por lotes.

Gestión de linaje y metadatos de datos.

Herramientas integrales y habilitación híbrida para entornos multinube.



Demo



Conclusiones

Escalabilidad e innovación, optimizando recursos según la demanda.

Integración nativa de servicios para facilitar pipelines y soluciones de IA.

Alta disponibilidad y robusta seguridad en una infraestructura global.

Administración simplificada mediante modelos serverless y servicios gestionados.

Gracias por su atención

Christian Barrios y Mayco Castellanos