

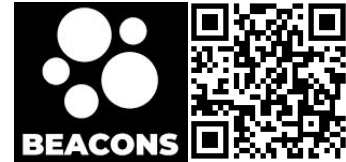
Taller Data Google

Introducción e intermedio I



Miguel Cotrina

- **Perfil Académico**
 - Ingeniero de software por la Universidad Tecnológica del Perú.
 - Maestría en ciencia de datos por la Universidad Ricardo Palma.
- **Perfil Profesional**
 - Arquitecto de datos en Clínica Internacional
 - Consultor e Instructor de Big Data, cloud, IA e IA Gen en empresas privadas y públicas



Agenda

- Introduccion a Google Cloud Platform (GCP)
- Bigquery en profundidad
- Aplicaciones de Bigquery en Marketing

Introduccion a Google Cloud Platform

Que es Google Cloud Platform



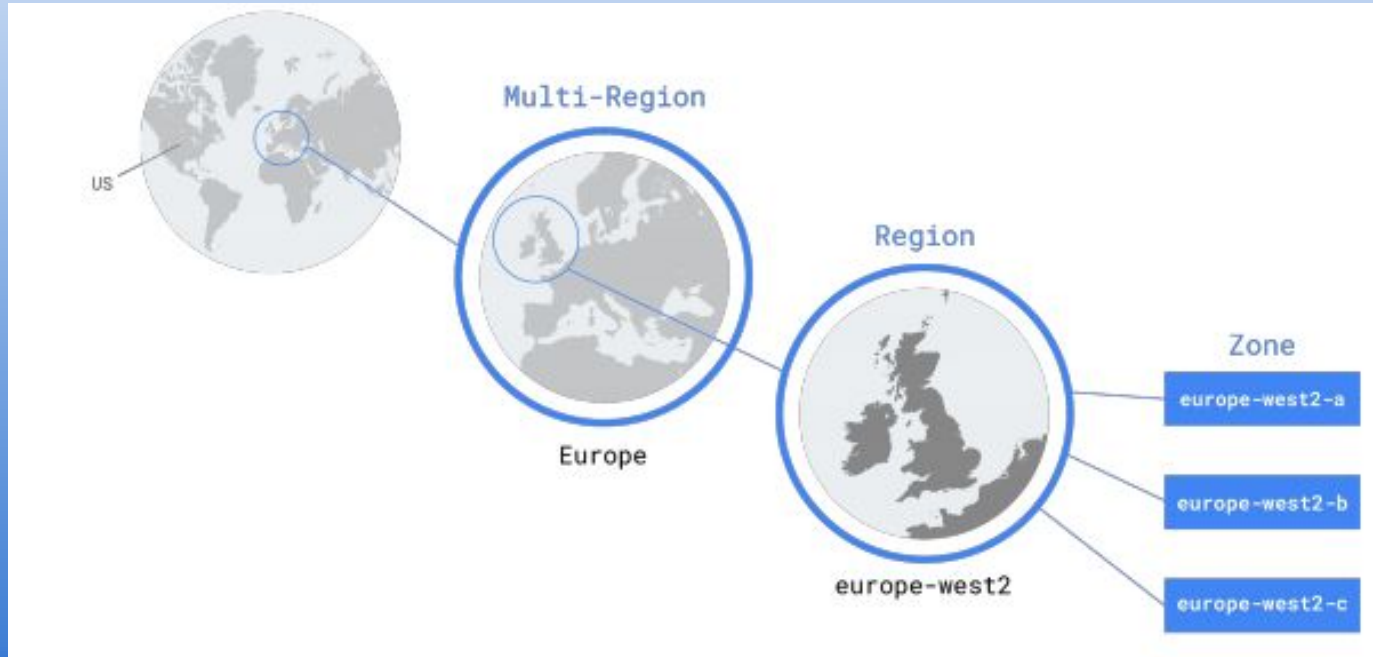
Google Cloud Platform

Google Cloud Platform (GCP) es la suite de servicios de computación en la nube que ofrece Google. Esta plataforma proporciona una amplia gama de herramientas y servicios que permiten a empresas y desarrolladores crear, desplegar y operar aplicaciones y servicios en la web. GCP abarca desde infraestructura como servicio (IaaS) hasta plataforma como servicio (PaaS) y software como servicio (SaaS), ofreciendo soluciones para diversas necesidades tecnológicas.

Google Cloud Platform



Infraestructura de seguridad distribuida



Definición y ventajas de la nube

Definición de la nube

La definición de la nube puede parecer poco clara, pero, básicamente, es un término que se utiliza para describir una red mundial de servidores, cada uno con una función única. La nube no es una entidad física, sino una red enorme de servidores remotos de todo el mundo que están conectados para funcionar como un único ecosistema.



Tipos de nube

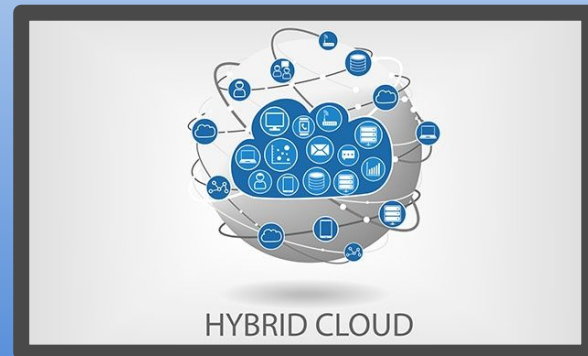
Nube privada On site



Nube publica



Nube Híbrida



Principales nubes



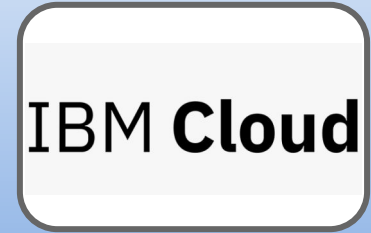
AWS
2006



GCP
2008



AZURE
2010



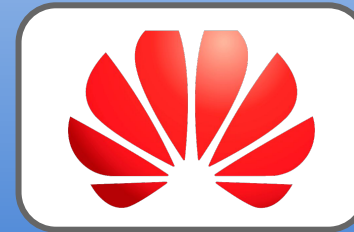
IBM
2011



Alibaba
2009



Oracle Cloud
2016



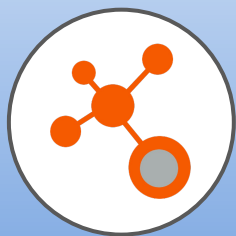
Huawei Cloud
2017

Ventajas de la nube



Autoservicio por demanda

Obtener recursos sin intervención del proveedor



Amplio acceso a red

Acceso desde cualquier sitio en cualquier momento



Recursos en común

Recursos compartido entre usuarios horizontalmente



Alta elasticidad

Incrementar o reducir la capacidad en segundos



Servicio medido

Paga solo lo que consumes sin contratos previos

Componentes principales de GCP y su arquitectura

Arquitectura de seguridad

Google encripta los datos en tránsito entre sus instalaciones y en reposo, lo que garantiza que solo se pueda acceder a ellos mediante roles y servicios autorizados con acceso auditado a las claves de encriptación.



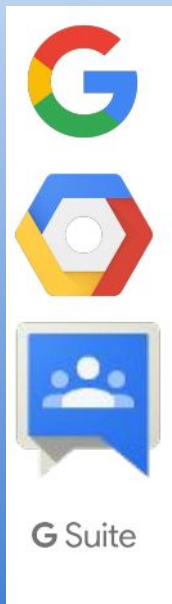
Arquitectura de accesos

IAM le permite otorgar acceso a recursos en la nube en niveles detallados, mucho más allá del acceso a nivel de proyecto



Arquitectura de accesos

Who: Representa a quien se le brinda los permisos



Google account or Cloud Identity user

test@gmail.com test@example.com

Service account

test@projectid.iam.gserviceaccount.com

G Suite Domain

example.com

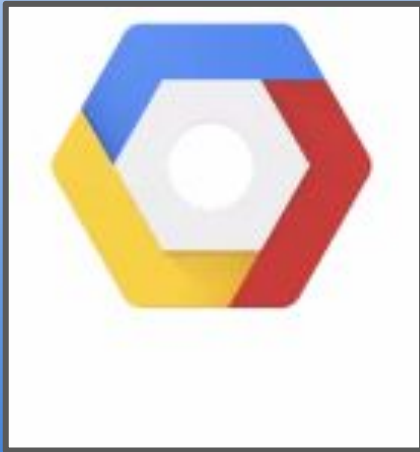
Google group

test@googlegroups.com

Arquitectura de accesos

Can do what: Representa que es lo que puede hacer(roles)

Primitivos



Predefinidos



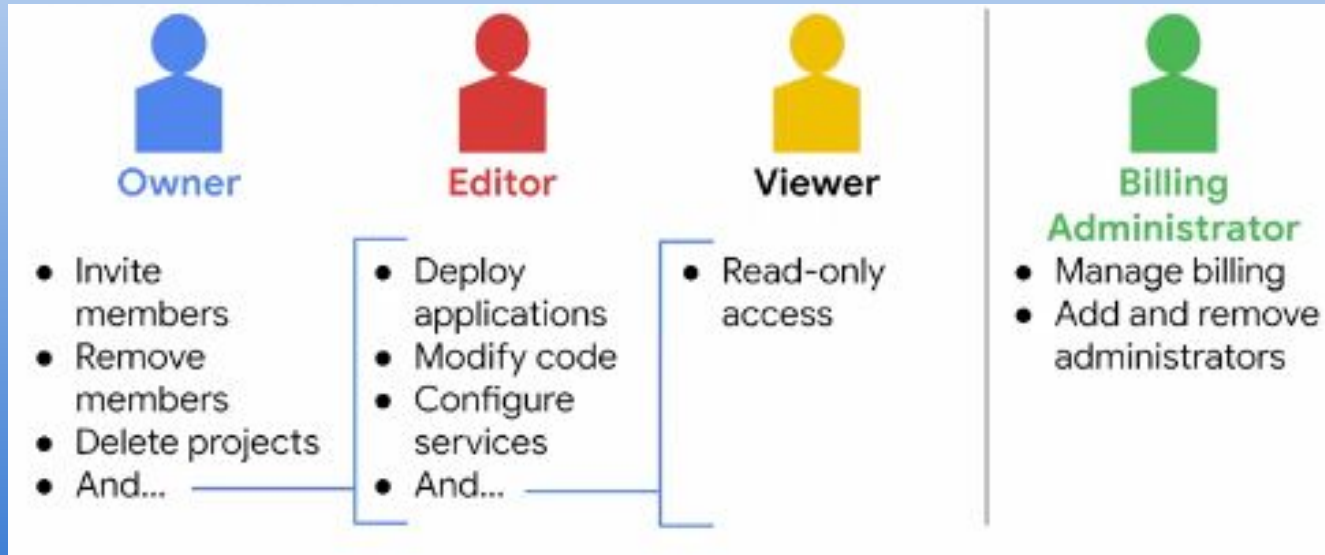
Personalizados



Arquitectura de accesos

Can do what: Representa que es lo que puede hacer(roles)

Primitivos



Arquitectura de accesos

Can do what: Representa que es lo que puede hacer(roles)

Predefinidos

Role Title	Description
Compute Admin	Full control of all Compute Engine resources (compute.*)
Network Admin	Permissions to create, modify, and delete networking resources, except for firewall rules and SSL certificates
Storage Admin	Permissions to create, modify, and delete disks, images, and snapshots

Arquitectura de accesos

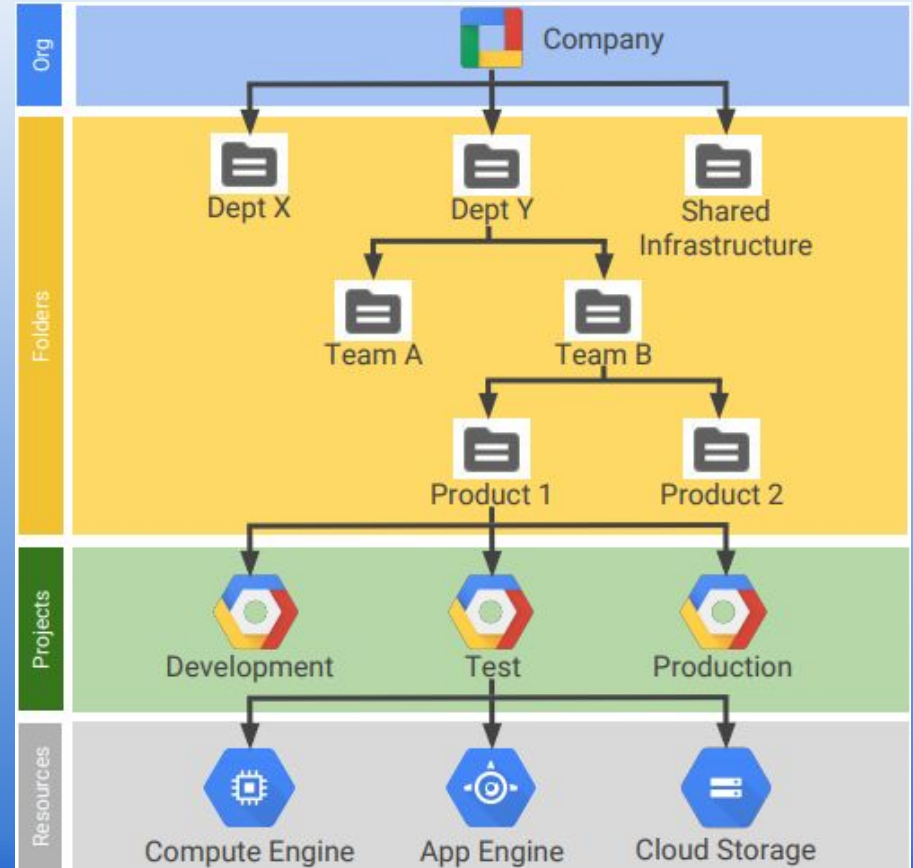
Can do what: Representa que es lo que puede hacer(roles)

Personalizados



Arquitectura de accesos

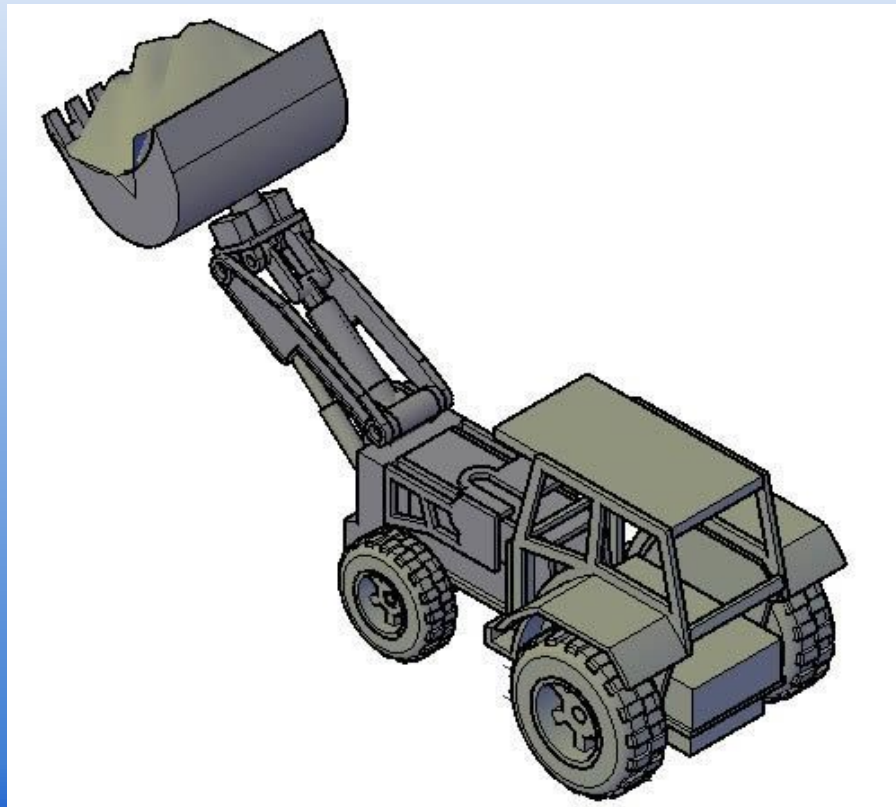
On which resource: Representa como gestionar los permisos sobre los componentes de gcp(recursos)



Casos de uso en diversas industrias

Predicción de fallas en palas mecánicas

Una empresa de maquinaria guarda registros la tensión de sus palas mecánicas en un datalake de GCP para calcular la dureza del material que excava. Con estos datos y añadiendo los datos históricos de fallas pueden predecir fallas futuras y reemplazar las palas mecánicas ahorrando el tiempo de conseguir el repuesto y realizar un cambio no previsto



Comportamiento de mercado farmaceutico y stock

Una empresa química distribuye medicamentos a las diversas farmacia de todo el país, sin embargo muchas veces se presentó el problema de agotamiento de stock debido a que muchos insumos son traídos del extranjero y luego del pedido puede tomar más de un mes la llegada.

La empresa solicitó a las farmacias le envíen un control de movimiento de productos de su marca, mediante esto pudo detectar las variaciones de comportamiento de consumo.



Puntos de venta y consumo al por menor

Cómo entender el consumo de las personas y poder detectar las compras recurrentes, si muchos clientes pagan en efectivo y no queda registro ni forma de identificarlos en la

Cómo entender el consumo de las personas y poder detectar las compras recurrentes, si muchos clientes pagan en efectivo y no queda registro ni forma de identificarlos en la siguiente compra



La solución es crear sistema de recompensas para que los usuarios se registren a cambio de recompensas de diversos tipos

Logro: Poder almacenar la información de compras por persona, poder entender la relación de productos y además realizar ofertas personalizadas

Como GCP transforma el marketing digital

Análisis de Datos y Segmentación de Audiencias

BigQuery: Plataforma de almacenamiento y análisis de datos que permite procesar grandes volúmenes de información para obtener insights valiosos sobre el comportamiento del cliente.

Segmentación Avanzada: Utilizando herramientas de GCP, las empresas pueden crear segmentos de audiencia basados en comportamientos, preferencias y valor de vida del cliente, optimizando así las campañas de marketing.

Aprendizaje Automático y Personalización

Vertex AI: Servicio que facilita la creación y despliegue de modelos de aprendizaje automático, permitiendo personalizar experiencias y recomendaciones para los clientes.

Recomendaciones Personalizadas: GCP permite analizar patrones de comportamiento para ofrecer productos o contenidos que se ajusten a las necesidades individuales de cada usuario, capturando mediante eventos, las diferentes actividades sobre su plataforma.

Integración y Automatización de Campañas

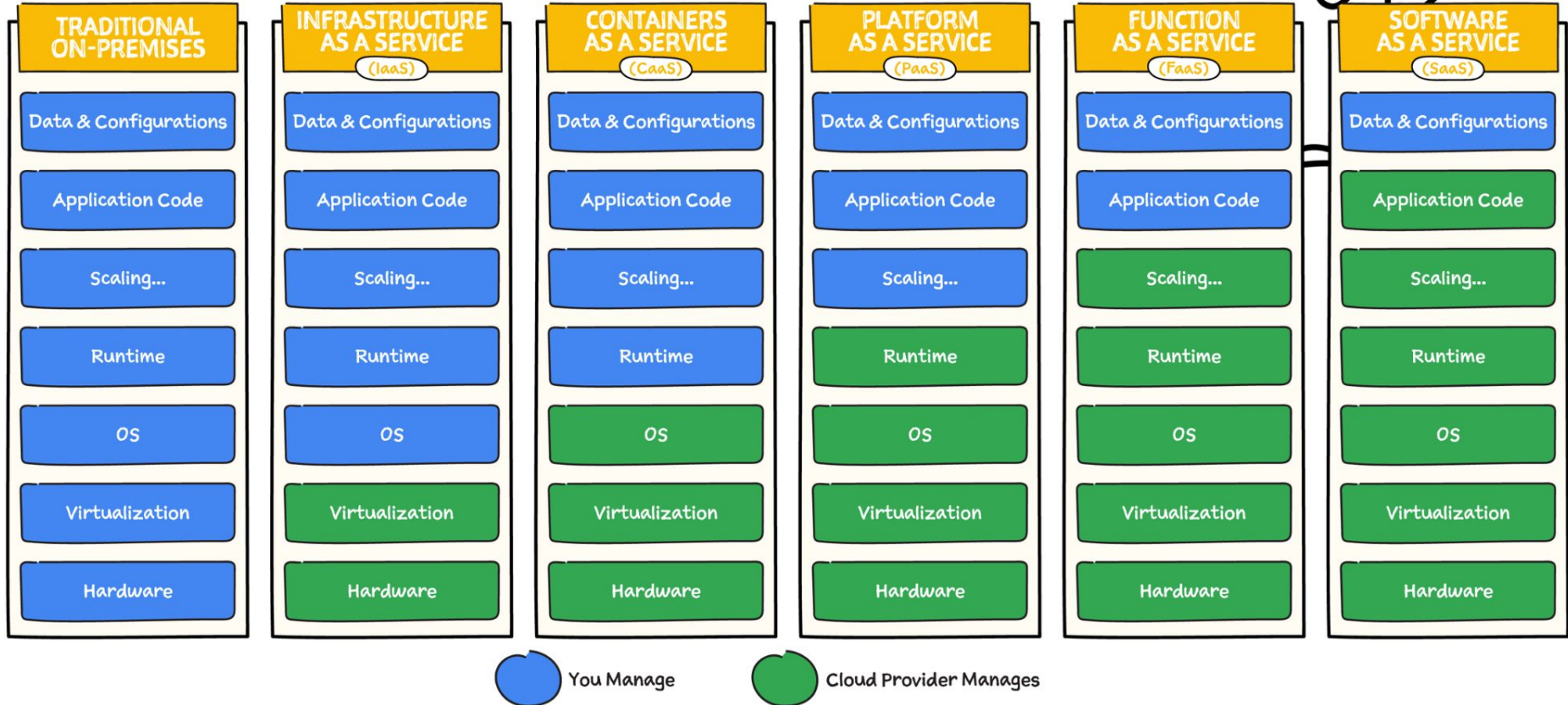
Google Marketing Platform: Plataforma que integra herramientas de publicidad y análisis, facilitando la gestión y medición de campañas en múltiples canales.

Automatización de Procesos: GCP permite automatizar tareas repetitivas, como la generación de informes y la gestión de campañas, liberando tiempo para actividades estratégicas, todo esto mediante componentes servlets

Servicios de Google Cloud clave en la industria



Wait... what is Cloud again?





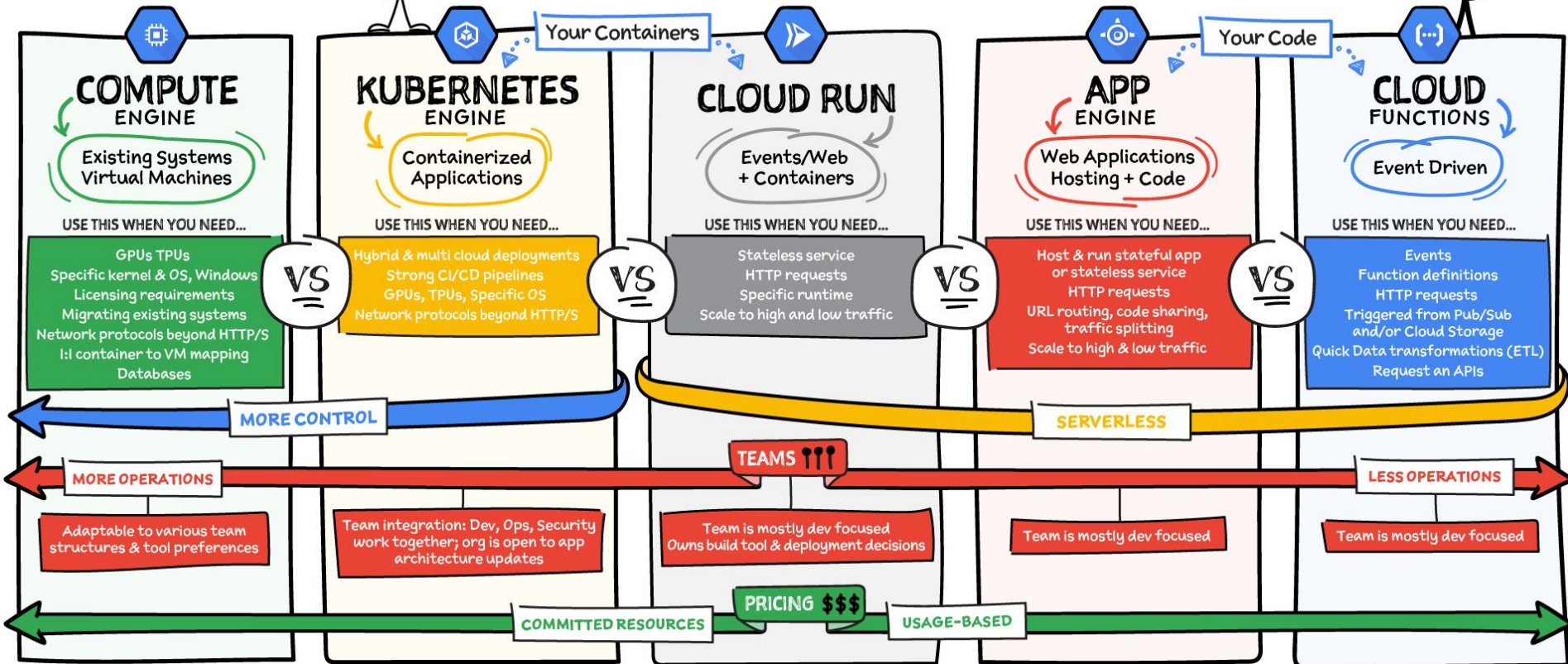
Where should I run my stuff?

IT DEPENDS...

#GCPSketchnotes

@PVERGADIA THECLOUDGIRL.DEV

PRO TIP: YOU CAN USE THEM TOGETHER





Which Database should I use?

#GCPSketchnotes @PVERGADIA THECLOUDGIRL.DEV
03.22.2023



RELATIONAL



Cloud SQL

Managed MySQL, PostgreSQL, SQL Server
SLA: 99.95%



AlloyDB

Managed PostgreSQL-compatible, with 100x faster analytics & 4x faster transaction queries
SLA: 99.99%

RELATIONAL & NON-RELATIONAL



Spanner

Multi-dialect (PostgreSQL, GoogleSQL) database, unlimited scaling
SLA: 99.999%

NON-RELATIONAL (NO SQL)



Firestore

Serverless, document database with built-in cross-client sync and offline caching
SLA: 99.999%



Bigtable

Key-value store for large scale, low latency workloads, Supports HBase APIs
SLA: 99.999%

IN MEMORY



Memorystore

Managed Redis & Memcached in memory cache & key-value store
SLA: 99.9%

BARE METAL



Oracle

Lift & shift oracle workloads to Google Cloud, run on certified, pre-configured and tuned infrastructure high performance and availability

Fully managed experience for low-touch administration with improved availability, security & governance, support

Good For:

- Fastest lift & shift migrations - OLTP workloads
- Managed experience
- Common API & control plane across database engines

- PostgreSQL workloads that need high performance & high availability or scale
- HTAP (hybrid transaction/analytical processing)
- Migrations off commercial databases

Highest scale and availability requirements without compromising on SQL capabilities, consolidating multiple databases for cost savings

- Rapid and cost-efficient application development (no need for middle tier)
- Easy aggregation from multiple data sources

- Cost-sensitive applications
- Need high throughput & consistent single-digit millisecond latencies irrespective of scale

- Sub-millisecond latencies for reads & writes
- Cache to improve app performance & throughput and reduce costs

Oracle applications and databases migrating to cloud

Use Case:



Web Frameworks



ERP



CRM



Ecommerce & Web



Operational Analytics



ERP, CRM



Financial Services



Ecommerce & Web



SaaS



Order & Inventory Management



Online Banking, payments & ledger



Gaming: player profiles & gameplay data



Electronic Medical Records



Catalog Metadata Management



Web & Mobile Apps



News Feeds, Social Chat, Influencer Engagements



Game Saves, Player Profiles



Retail Catalogs, Point of Sales



Real-Time Analytics (Personalization, Fraud detection)



IoT/Clickstream/Time Series



Feature Stores, Operational Data Hubs & Data Fabrics



Batch Unstructured Data Processing



Financial Markets, Crypto Ledgers



Database caching



Session Store



Jobs and Queues



Leaderboards



Fast Data Ingestion



Legacy Applications



Datacenter Exits



Which Storage Should I Use?

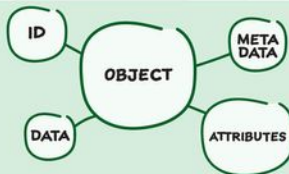
OBJECT STORE



Cloud Storage

For any app. Store any type & any amount of data for any duration, & retrieve it as often as needed

GOOD FOR:
Binary or object data,
blobs, unstructured data



Object Storage

DATA ↓ ↑ METADATA/
OBJECT ID

USE CASE:



Streaming videos



Images



Data analytics



Backups



Documents



Regulatory archives



Tape replacement



Websites



Genomics



Disaster recovery

BLOCK STORE



Persistent
Disk

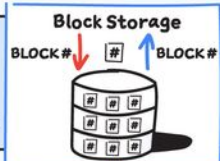


Fully integrated with
Compute Engine & GKE



Local
SSD

GOOD FOR:
✓ Block store for VMs
✓ Range of latency & performance options



GOOD FOR:
✓ Ephemeral block store for VMs
✓ Lowest latency
✓ Stateless workloads

USE CASE:



Disks for VMs



Share read-only data
across VMs



Rapid, durable backups
of running VMs



Storage for databases

USE CASE:



Flash-optimized databases



Hot caching layer for analytics



Application scratch disk



Scale out analytics



Media rendering

FILE STORE



Filestore

Fully managed, cloud-based
Network Attached Storage

GOOD FOR:
Shared file storage
(unstructured) data

USE CASE:



Media processing



Life sciences/
Genomics processing



Electronic Design Automation
(EDA)



Data Analytics



Application migrations



Machine Learning



Web content management



Financial Modeling

Laboratorios Fase 01

- Acceso a cuenta de correo y de servicios mediante permisos nativos, predefinidos y personalizados
- Despliegue de Cloud Storage y ciclo de vida
- Despliegue de Cloud function como eventdriver en python, para eventos en cloud storage
- Despliegue de cloud Run como microservicios

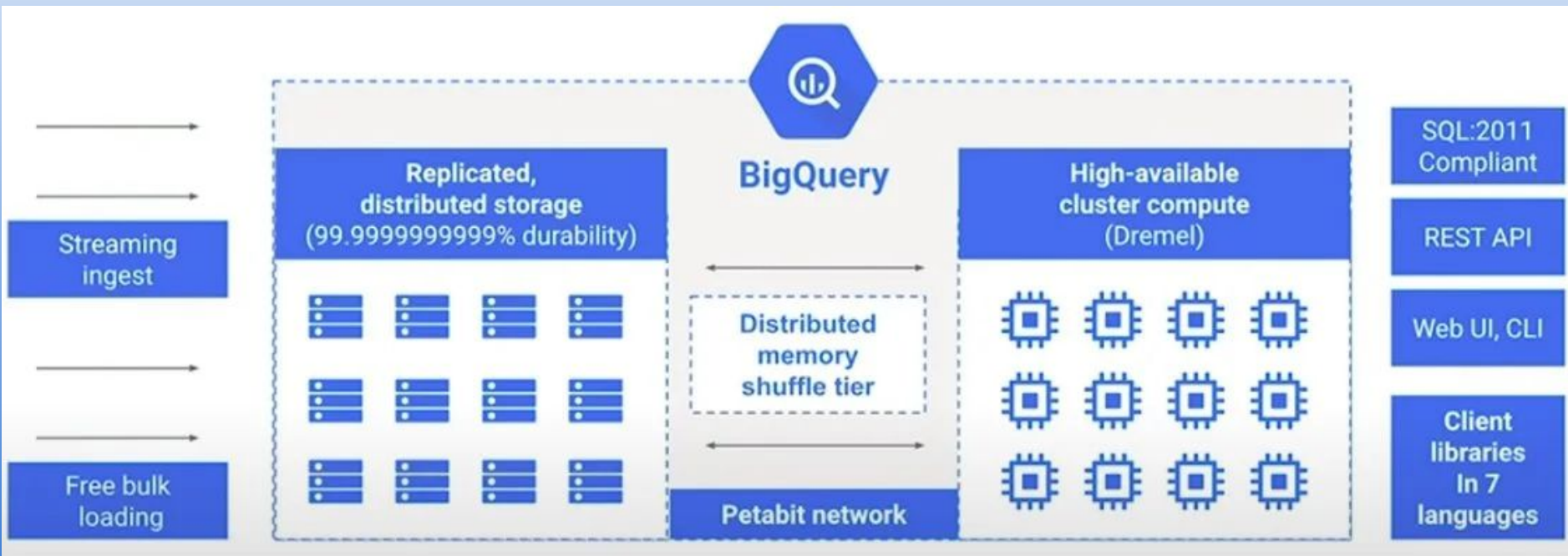
Recorrido sobre la plataformas de almacenamiento

Hands-On: Crear un acceso personalizado que permita visualizar el contenido de Cloud Storage a un compañero

Bigquery en profundidad

Arquitectura de BigQuery y su lugar dentro de GCP

Arquitectura de Bigquery



Bigquery Particionamiento

Una tabla particionada se divide en segmentos, denominados particiones, que facilitan la administración y la consulta de tus datos. Dividir una tabla grande en particiones más pequeñas puede mejorar el rendimiento de las consultas y ayudarlo a controlar sus costos, ya que se reduce la cantidad de bytes que lee una consulta.

Tipos de particiones

- Partición por rangos de números enteros
- Partición de columnas por unidad de tiempo
- Partición por tiempo de transferencia

Bigquery Particionamiento

nombre	Edad	Fecha

• • •

Sin particion

- Select nombre from tabla where Fecha = "dd/mm/yyyy"

Con particion

- Select nombre from tabla where Fecha = "dd/mm/yyyy"

Ejemplo x registros

			DIA 1
			DIA 2
			DIA 3
•	•	•	
•	•	•	
•	•	•	

Nota: Máximo de particiones 4000

Bigquery Clusterizacion

Las tablas agrupadas con clusters en BigQuery son tablas que tienen un orden de clasificación de columnas definido por el usuario mediante columnas agrupadas. Las tablas agrupadas pueden mejorar el rendimiento de las consultas y reducir los costos.

Cuando agrupas una tabla con varias columnas, el orden de las columnas determina qué columnas tienen prioridad cuando BigQuery ordena y agrupa los datos en bloques de almacenamiento.

Bigquery Clusterization

Orders table
Not clustered

Order_Date	Country	Status
2022-08-02	US	Shipped
2022-08-04	JP	Shipped
2022-08-05	UK	Canceled
2022-08-06	KE	Shipped
2022-08-02	KE	Canceled
2022-08-05	US	Processing
2022-08-04	JP	Processing
2022-08-04	KE	Shipped
2022-08-06	UK	Canceled
2022-08-02	UK	Processing
2022-08-05	JP	Canceled
2022-08-06	UK	Processing
2022-08-05	US	Shipped
2022-08-06	JP	Processing
2022-08-02	KE	Shipped
2022-08-04	US	Shipped

Orders table
Clustered by Country

Order_Date	Country	Status
2022-08-02	US	Shipped
2022-08-05	US	Shipped
2022-08-05	US	Processing
2022-08-04	US	Shipped
2022-08-04	JP	Shipped
2022-08-04	JP	Processing
2022-08-05	JP	Canceled
2022-08-06	JP	Processing
2022-08-05	UK	Canceled
2022-08-06	UK	Canceled
2022-08-06	UK	Processing
2022-08-02	UK	Processing
2022-08-06	KE	Shipped
2022-08-02	KE	Canceled
2022-08-04	KE	Shipped
2022-08-02	KE	Shipped

Orders table
Clustered by Country and Status

Order_Date	Country	Status
2022-08-05	US	Processing
2022-08-02	US	Shipped
2022-08-05	US	Shipped
2022-08-04	US	Shipped
2022-08-05	JP	Canceled
2022-08-04	JP	Processing
2022-08-06	JP	Processing
2022-08-04	JP	Shipped
2022-08-05	UK	Canceled
2022-08-06	UK	Canceled
2022-08-06	UK	Processing
2022-08-02	UK	Processing
2022-08-02	KE	Canceled
2022-08-06	KE	Shipped
2022-08-04	KE	Shipped
2022-08-02	KE	Shipped

Bigquery Particion y Clusterizacion

Puedes combinar la partición de tablas con el agrupamiento en clústeres de tablas para lograr una clasificación detallada a fin de optimizar más las consultas.

Cuando creas una tabla agrupada y particionada, puedes lograr una clasificación más detallada, como se muestra en el siguiente diagrama

Bigquery Particion y Clusterización

Orders table
Not Clustered; Not partitioned

Order_Date	Country	Status
2022-08-02	US	Shipped
2022-08-04	JP	Shipped
2022-08-05	UK	Canceled
2022-08-06	KE	Shipped
2022-08-02	KE	Canceled
2022-08-05	US	Processing
2022-08-04	JP	Processing
2022-08-04	KE	Shipped
2022-08-06	UK	Canceled
2022-08-02	UK	Processing
2022-08-05	JP	Canceled
2022-08-06	UK	Processing
2022-08-05	US	Shipped
2022-08-06	JP	Processing
2022-08-02	KE	Shipped
2022-08-04	US	Shipped

Orders table
Clustered by Country; Not partitioned

Order_Date	Country	Status
2022-08-04	JP	Shipped
2022-08-04	JP	Processing
2022-08-05	JP	Canceled
2022-08-06	JP	Processing
2022-08-06	KE	Shipped
2022-08-02	KE	Canceled
2022-08-04	KE	Shipped
2022-08-02	KE	Shipped
2022-08-05	UK	Processing
2022-08-06	UK	Canceled
2022-08-02	UK	Canceled
2022-08-06	UK	Processing
2022-08-02	US	Shipped
2022-08-05	US	Processing
2022-08-05	US	Shipped
2022-08-04	US	Shipped

Orders table
Clustered by Country; Partitioned by Order Date (Daily)

	Order_Date	Country	Status
Partition: 2022-08-02	2022-08-02	KE	Shipped
	2022-08-02	KE	Canceled
Clusters: Country	2022-08-02	UK	Processing
	2022-08-02	US	Shipped

	Order_Date	Country	Status
Partition: 2022-08-04	2022-08-04	JP	Shipped
	2022-08-04	JP	Processing
Cluster: Country	2022-08-04	KE	Shipped
	2022-08-04	US	Shipped

	Order_Date	Country	Status
Partition: 2022-08-05	2022-08-05	JP	Canceled
	2022-08-05	UK	Canceled
Cluster: Country	2022-08-05	US	Shipped
	2022-08-05	US	Processing

	Order_Date	Country	Status
Partition: 2022-08-06	2022-08-06	JP	Processing
	2022-08-06	KE	Shipped
Cluster: Country	2022-08-06	UK	Canceled
	2022-08-06	UK	Processing

Propósito de BigQuery como plataforma de análisis de datos a gran escala.

Propósito de BigQuery en el análisis de datos

Almacenamiento y procesamiento eficiente:

- Diseñado para almacenar y analizar grandes volúmenes de datos estructurados y semiestructurados.
- Facilita la ejecución de consultas complejas sin necesidad de gestionar infraestructura.

Análisis en tiempo real:

- Capacidad para procesar datos en streaming, permitiendo análisis y toma de decisiones basadas en datos actualizados al instante.

Beneficios de utilizar BigQuery

Rendimiento y velocidad:

- Procesa consultas en segundos, incluso en conjuntos de datos de gran tamaño.

Facilidad de uso:

- Utiliza SQL estándar, facilitando la adopción por parte de analistas y científicos de datos.

Integración con ecosistema de Google Cloud:

- Se conecta fácilmente con otras herramientas como Data Studio, Looker y servicios de aprendizaje automático.

Uso en proyectos de marketing análisis de datos de clientes, campañas y segmentos

Análisis de Datos de Clientes

Personalización de Experiencias:

- BigQuery permite procesar grandes volúmenes de datos de clientes, como preferencias, historial de compras y comportamientos.
- Uso de modelos predictivos para anticipar necesidades y diseñar experiencias personalizadas.

Creación de Perfiles Avanzados:

- Combina datos demográficos, geográficos y comportamentales en tiempo real.
- Ayuda a generar perfiles de clientes que mejoran la segmentación y las estrategias de fidelización.

Visualización de Insights:

- BigQuery se integra con herramientas como Looker o Google Data Studio para generar dashboards interactivos que muestran tendencias y patrones.

Optimización de Campañas

Medición del ROI en Tiempo Real:

- Analiza métricas como el costo por clic (CPC), tasa de conversión y retorno de la inversión (ROI).
- Permite realizar ajustes inmediatos en campañas para maximizar resultados.

Pruebas A/B a Gran Escala:

- Facilita la ejecución y análisis de experimentos para identificar estrategias más efectivas.
- Compara el rendimiento de diferentes segmentos de audiencia o variaciones creativas.

Previsión y Tendencias:

- Usa datos históricos para predecir resultados de campañas futuras.
- Detecta patrones que pueden influir en el éxito de las campañas, como temporalidades o comportamientos del mercado.

Segmentación Inteligente

Identificación de Audiencias Clave:

- Segmentación dinámica basada en datos como frecuencia de compra, gasto promedio y preferencias de productos.
- Genera grupos de clientes específicos para estrategias focalizadas.

Automatización de Estrategias:

- Con BigQuery y herramientas de marketing, las audiencias se actualizan automáticamente con nuevos datos.
- Aumenta la eficiencia y la relevancia de las campañas.

Integración Multicanal:

- Conexión con plataformas publicitarias como Google Ads para sincronizar segmentos.
- Mejora el alcance y la coherencia de mensajes en diferentes canales.

Laboratorio fase 02

Creación de tablas(UI, Query, Storage).

Ejecución de consultas y creación de reportes en Looker Studio



Aplicacion de Bigquery en Marketing

Segmentación de Clientes con BigQuery

- **Análisis de Datos Combinados:** BigQuery permite integrar datos de diversas fuentes (CRM, plataformas de ecommerce, redes sociales) para crear segmentos altamente personalizados.
- **Segmentación Dinámica:** Actualización automática de segmentos en función de los comportamientos de los usuarios, como compras, interacciones y navegación web.
- **Segmentos Avanzados:** Utilización de datos históricos, demográficos y psicográficos para crear grupos de clientes con necesidades y comportamientos similares.

Métricas de Campañas en Tiempo Real

- **Monitoreo Instantáneo:** BigQuery procesa grandes volúmenes de datos en tiempo real para proporcionar métricas clave como el CPC, CTR, conversiones y ROI al instante.
- **Optimización Automática:** Gracias a las consultas rápidas, las marcas pueden ajustar sus campañas en tiempo real, maximizando la efectividad y reduciendo el gasto innecesario.
- **Integración con Herramientas de Análisis:** Datos procesados se visualizan en plataformas como Google Data Studio y Looker, facilitando la toma de decisiones rápidas y basadas en datos.

¿Cómo vender BigQuery en un proyecto de marketing?

Enfoque en la Eficiencia y Escalabilidad

BigQuery se presenta como una herramienta imprescindible para proyectos de marketing que necesitan procesar grandes volúmenes de datos de forma eficiente. A medida que las marcas recopilan más datos sobre sus clientes y campañas, la capacidad de manejar esa información en tiempo real se vuelve crucial.

BigQuery, al ser una plataforma sin servidor, elimina la necesidad de gestionar infraestructura, lo que permite que el equipo de marketing se enfoque en lo que realmente importa: los datos y cómo utilizarlos para tomar decisiones más informadas.

Enfoque en la Personalización y Toma de Decisiones

Uno de los principales retos de cualquier estrategia de marketing es la capacidad de personalizar las campañas en función de los comportamientos y las preferencias de los clientes. Con BigQuery, esta personalización no solo es posible, sino que es escalable y manejable. La plataforma permite integrar y analizar datos provenientes de múltiples fuentes, como plataformas de ecommerce, redes sociales, y CRM, lo que permite crear perfiles completos y precisos de los clientes. Además, gracias a su capacidad para ejecutar consultas en tiempo real, BigQuery permite ajustar campañas y estrategias de marketing al instante.

Enfoque en la Integración y Visualización de Datos

BigQuery no es solo una herramienta para almacenar y procesar datos; es una plataforma que se integra perfectamente con otras herramientas del ecosistema de Google Cloud y terceros, como Google Ads, Looker y Google Looker Studio. Esta integración permite a los equipos de marketing consolidar todos sus datos en un solo lugar y luego visualizar esos datos de manera clara y accesible. Al hacerlo, se facilita la creación de dashboards personalizados que muestran el desempeño de campañas y la segmentación de clientes de manera intuitiva. Lo mejor de todo es que, dado que BigQuery está diseñado para trabajar con datos semiestructurados, las marcas pueden analizar datos provenientes de múltiples fuentes sin preocuparse por la estructura o el formato.

Enlaces de materiales

Material de laboratorios para:

Bigquery

Cloud SQL

Cloud Storage

Cloud Run

Cloud Function

<https://github.com/macespinoza/TallerGCPMaterialDidactico>

Muchas Gracias