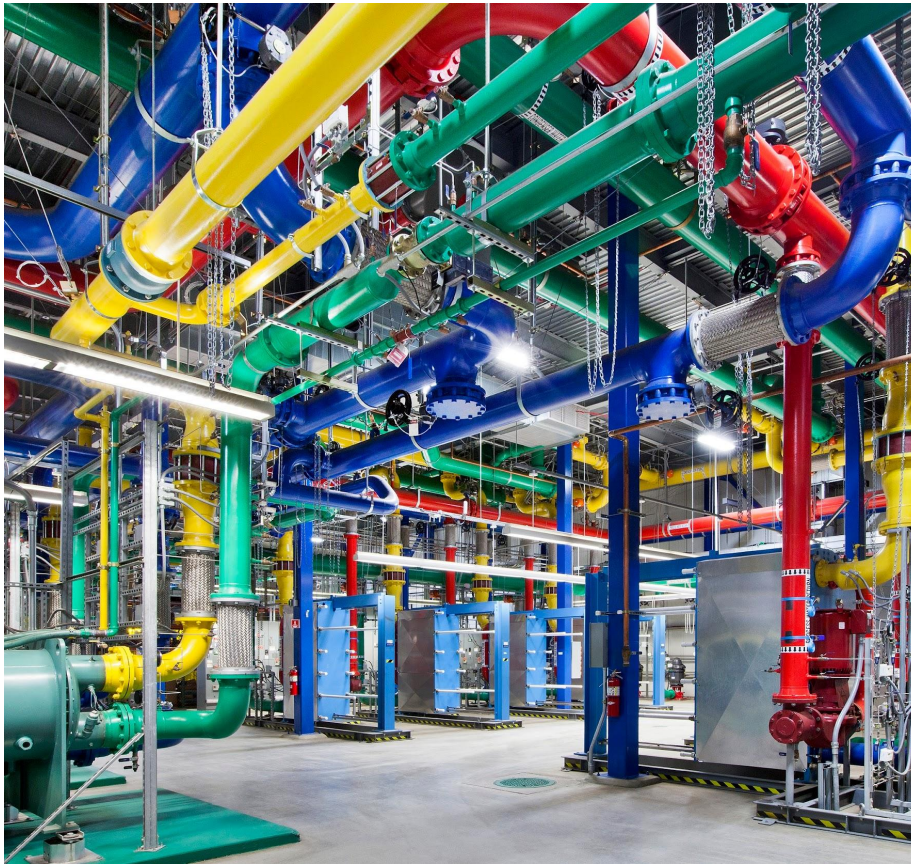


Google Cloud

Cristina Montserrat
cmontserrat@google.com

Ana López-Mancisidor
analopezm@google.com

Google Cloud Public Sector

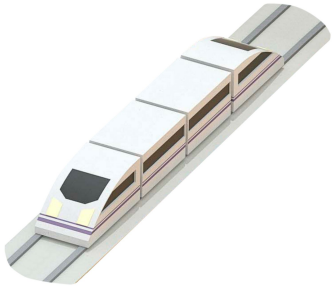
Google Cloud

Agenda

- 01. Introduction to Google Cloud
- 02. Compute and storage services
- 03. App Development in Google Cloud
- 04. BigData & Analytics
- 05. AI and ML
- 06. Geolocation
- 07. Google Cloud labs this afternoon

Google Cloud

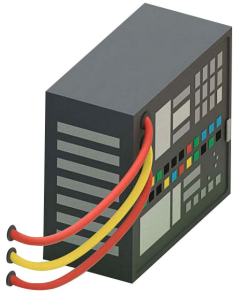
Main benefits of using Public Cloud services



Velocity & Innovation

Scalability

Security



Cost efficiency

Sustainability

Google Cloud

DEVELOPER'S CHEAT SHEET

Created by the Google Developer Relations Team
Maintained at https://4words.dev

Feedback? @pvergadia @GoogleCloudTech

COMPUTE

Cloud Functions	Event-driven serverless functions
App Engine	Managed app platform
Cloud Run	Serverless for containerized applications
Kubernetes Engine (GKE)	Managed Kubernetes/containers
Compute Engine	VMs, GPUs, TPUs, Disks
Bare Metal Solution	Hardware for specialized workloads
Preemptible VMs	Short-lived compute instances
Shielded VMs	Hardened VMs
Sole-tenant Nodes	Dedicated physical servers
VMware Engine	VMware on Compute Engine

STORAGE

Cloud Filestore	Managed NFS server
Cloud Storage	Multi-class multi-region object storage
Persistent Disk	Block storage for VMs
Local SSD	VM locally attached SSDs

DATABASE

Cloud Bigtable	Petabyte-scale, low-latency, non-relational
Cloud Firestore	Serverless NoSQL document DB
Cloud Memorystore	Managed Redis and Memcached
Cloud Spanner	Horizontally scalable relational DB
Cloud SQL	Managed MySQL, PostgreSQL, SQL Server
Database Migration Service	Migrate to Cloud SQL
DB Insights	SQL Inspector

DATA ANALYTICS

BigQuery	Data warehouse/analytics
BigQuery BI Engine	In-memory analytics engine
BigQuery ML	BigQuery model training/serving
BigQuery GIS	BigQuery geospatial functions/support
BigQuery DTS	Automated data ingestion service
Connected Sheets	Spreadsheet interface for BigData
Cloud Composer	Managed workflow orchestration service
Data Fusion	Graphically manage data pipelines
Dataflow	Stream/batch data processing
Dataprep by Trifacta	Visual data wrangling
Dataproc	Managed Spark and Hadoop
Datastream	Change data capture/replication service
Pub/Sub	Global real-time messaging
Data Catalog	Metadata management service
Data Studio	Collaborative data exploration/dashboarding
Looker	Enterprise BI and Analytics
Public Datasets	Hosted data in BigQuery/GCS

HYBRID AND MULTI-CLOUD

Anthos	Enterprise hybrid/multi-cloud platform
Anthos Clusters	Hybrid/on-prem Kubernetes Engine
Anthos Config Management	Policy and security automation
Anthos Service Mesh	Managed service mesh (Istio)
Cloud Run for Anthos	Serverless development for Anthos
GCP Marketplace for Anthos	Pre-configured containerized apps
Migrate for Anthos	Migrate VMs to Kubernetes Engine
Operations	Monitoring, logging, troubleshooting
Traffic Director	Service mesh traffic management
Apigee API Management	API management, development, security

AI/ML

Vertex AI	Managed platform for ML
AutoML	Custom low-code models AI Platform
Vertex AI Data Labeling	Data labeling by humans
Deep Learning VM Images (DLVM)	Preconfigured VMs for deep learning
Vertex AI Workbench	Jupyter-based environment for Data Science
Vertex AI Deep Learning Containers	Preconfigured containers for deep learning
Vertex AI Matching Engine	Vector similarity searches
Vertex AI Pipelines	Hosted ML workflows
Vertex AI Predictions	Autoscaled model serving
Vertex AI Training	Distributed AI training
Vertex AI Edge Manager	Deploy monitor edge inferences
Vertex Explainable AI	Understand ML model predictions
Vertex AI Feature Store	Managed ML feature repository
Vertex ML Metadata	Artifact, lineage, and execution tracking
Vertex AI Model Monitoring	Monitor models for skew/drift
Vertex AI Tensorboard	Managed TensorBoard for ML-experiment Visualization
Vertex AI Vizier	Black-box hyperparameter tuning
Cloud Speech-To-Text API	Convert audio to text
Cloud Talent Solutions API	Job search with ML
Cloud Text-To-Speech API	Convert text to audio
Cloud TPU	Hardware acceleration for ML
Cloud Translation API	Language detection and translation
Cloud Video Intelligence API	Scene-level video annotation
Cloud Vision API	Image recognition and classification
Contact Center AI	AI in your contact center
Dialogflow	Create conversational interfaces
Document AI	Analyze, classify, search documents
Recommendations AI	Create custom recommendations
Vision Product Search	Visual search for products

NETWORKING

Carrier Peering	Peer through a carrier
Direct Peering	Peer with GCP
Dedicated Interconnect	Dedicated private network connection
Partner Interconnect	Connect on-prem network to VPC
Cloud Armor	DDoS protection and WAF
Cloud CDN	Content delivery network
Cloud DNS	Programmable DNS serving
Cloud Load Balancing	Multi-region load distribution/balancing
Cloud NAT	Network address translation service
Cloud Router	VPC/on-prem network route exchange (BGP)
Cloud VPN (HA)	VPN (Virtual private network connection)
Network Service Tiers	Price vs performance/Berlin
Network Telemetry	Network telemetry service
Traffic Director	Service mesh traffic management
Google Cloud Service Mesh	Service-aware network management
Virtual Private Cloud	Software defined networking
VPC Service Controls	Security perimeters for API-based services
Network Intelligence Center	Network monitoring and topology

IDENTITY AND SECURITY

Access Transparency	Audit cloud provider access
Assured Workloads	Workload compliance controls
Binary Authorization	Kubernetes deploy-time security
Certificate Authority Service	Managed private CAs
Cloud Asset Inventory	All assets, one place
Cloud Audit Logs	Audit trails for GCP
Cloud DLP	Classify and redact sensitive data
Cloud HSM	Hardware security module service
Cloud EKM	External keys you control
Cloud IAM	Resource access control
Cloud Identity	Manage users, devices & apps
Cloud Identity-Aware Proxy	Identity-based app access
Cloud KMS	Hosted key management service
Cloud Resource Manager	Cloud project metadata management
Security Command Center	Security management & data risk platform
Cloud Security Scanner	App engine security scanner
Confidential Computing	Encrypt data in-use
Context-aware Access	End-user attribute-based access control
Event Threat Detection	Scans for suspicious activity
Managed Service for Microsoft Active Directory	Managed Microsoft Active Directory
Secret Manager	Store and manage secrets
Security Key Enforcement	Two-step key verification
Shielded VMs	Hardened VMs
Titan Security Key	Two-factor authentication (2FA) device
VPC Service Controls	VPC data constraints
Chronicle	Find threats from security telemetry
VirusTotal	Research/hunt for Malware
Risk Manager	Evaluate organization's security posture
reCAPTCHA Enterprise	Protection against bot/spam/abuse
BeyondCorp Enterprise	Zero trust secure access
Access Context Manager	Fine-grained, attribute based access-control
Web Security Scanner	Identifies web-app security vulnerabilities

OPERATIONS & MONITORING

Cloud Debugger	Live production debugging
Error Reporting	App error reporting
Cloud Logging	Centralized logging
Cloud Monitoring	Infrastructure and application monitoring
Cloud Profiler	CPU and heap profiling
Cloud Trace	App latency insights

DEVOPS CI/CD

Cloud Build	Continuous integration/delivery platform
Cloud Deploy	Deployment pipeline for GKE
Artifact Registry	Universal package manager
Cloud Source Repositories	Hosted private git repos
Container Registry	Private container registry/storage

APPLICATION INTEGRATION

Eventarc	Event-driven Cloud Run services
Cloud Scheduler	Managed cron job service
Cloud Tasks	Asynchronous task execution
Cloud Workflows	HTTP services orchestration
Pub/Sub	Global real-time messaging

API PLATFORM AND ECOSYSTEMS

API Analytics	API metrics
API Monetization	Monetize APIs
Apigee API Platform	Develop, secure, monitor APIs
API Gateway	Fully managed API Gateway
Apigee Hybrid	Manage hybrid/multi-cloud API environments
Apigee Sense	API protection from attacks
Cloud Endpoints	Cloud API gateway
Developer Portal	API management portal
Marketplace	Partner & open source marketplace
AppSheet	No-code App creation

INTERNET OF THINGS (IOT)

Cloud IoT Core	Manage devices, ingest data

GAMING

Google Cloud Game Servers	Orchestrate Agones clusters

HEALTHCARE

Cloud Healthcare API	Healthcare system GCP interoperability
Apigee Healthcare APIs	Healthcare system GCP interoperability
Healthcare Natural Language AI	Real-time insights from medical text
Cloud Life Sciences	Manage, process, transform biomedical data

RETAIL

Vision Product Search	Visual search for products
Recommendations AI	Create custom recommendations
Visual Inspection AI	Train/deploy models to detect defects

MANAGEMENT TOOLS

VM Manager	Manage OS VM Fleets
Cloud APIs	APIs for cloud services
Cloud Billing API	Programmatically manage GCP billing
Cloud Billing	Billing and cost management tools
Cloud Console	Web-based management console
Cloud Deployment Manager	Templated infrastructure deployment
Cloud Mobile App	iOS/Android GCP manager app
Private Catalog	Internal Solutions Catalog

DEVELOPER TOOLS

Cloud Code for IntelliJ	IntelliJ GCP tools
Cloud Code for VS Code	VS Code GCP tools
Cloud Code	Cloud native IDE extensions
Cloud Tools for Eclipse	Eclipse GCP tools
Cloud Tools for Visual Studio	Visual Studio GCP tools
Gradle App Engine Plugin	Gradle App Engine plugin
Maven App Engine Plugin	Maven App Engine plugin
Cloud SDK	CLI for GCP
Cloud Shell	Browser-based terminal/CLI

MIGRATION TO GOOGLE CLOUD

BigQuery Data Transfer Service	Bulk import analytics data
Cloud Data Transfer	Data migration tools/CLI
Google Transfer Appliance	Rentable data transport box
Storage Transfer Service	Online/on-premises data transfer
Migrate for Anthos	Migrate VMs to GKE containers
Migrate for Compute Engine	Compute Engine migration tools
Migrate from Amazon Redshift	Migrate from Redshift to BigQuery
Migrate from Teradata	Migrate from Teradata to BigQuery
Cloud Foundation Toolkit	Infrastructure as Code templates
KF	Cloud Foundry to Kubernetes

GOOGLE MAPS PLATFORM

Directions API	Get directions between locations
Distance Matrix API	Multi-origin/destination travel times
Geocoding API	Convert address to/from coordinates
Geolocation API	Derive location without GPS
Maps Embed API	Display iframe embedded maps
Maps JavaScript API	Dynamic web maps
Maps SDK for Android	Maps for Android apps
Maps SDK for iOS	Maps for iOS apps
Maps Static API	Display static map images
Maps SDK for Unity	Unity SDK for games
Maps URLs	URL scheme for maps
Places API	Rest-based Places features
Places Library, Maps JS API	Places features for web
Places SDK for Android	Places features for Android
Places SDK for iOS	Places features for iOS
Roads API	Convert coordinates to roads
Street View Static API	Static street view images
Street View Service	Street view for JavaScript
Time Zone API	Convert coordinates to timezone

WORKSPACE PLATFORM

Admin SDK	Manage Google Workspace resources
AMP for Email	Dynamic interactive email
Apps Script	Extend and automate everything
Calendar API	Create and manage calendars
Classroom API	Provision and manage classrooms
Cloud Search	Unified search for enterprise
Docs API	Create and edit documents
Drive Activity API	Retrieve Google Drive activity
Drive API	Read and write files
Drive Picker	Drive file selection widget
Email Markup	Interactive email using schema.org
Google Workspace Add-ons	Extend Google Workspace apps
Google Workspace Marketplace	Storefront for integrated applications
Gmail API	Enhance Gmail
Google Chats API	Conversational bots in chat
People API	Manage user's Contacts
Sheets API	Read and write spreadsheets
Slides API	Create and edit presentations
Task API	Search, read & update Tasks
Vault API	Manage your organization's eDiscovery

MOBILE (FIREBASE)

Cloud Firestore	Document store and sync
Cloud Functions for Firebase	Event-driven serverless applications
Cloud Storage for Firebase	Object storage and serving
Crashlytics	Crash reporting and analytics
Firebase A/B Testing	Create A/B test experiments
Firebase App Distribution	Trusted tester early access
Firebase Authentication	Drop-in authentication
Firebase Cloud Messaging	Send device notifications
Firebase Dynamic Links	Link to app content
Firebase Extensions	Pre-packaged development solutions
Firebase Hosting	Web hosting with CDN/SSL
Firebase In-App Messaging	Send in-app contextual messages
Firebase Performance Monitoring	App/web performance monitoring
Firebase Predictions	Predict user targeting
Firebase Realtime Database	Real-time data synchronization
Firebase Remote Config	Remotely configure installed apps
Firebase Test Lab	Mobile testing device farm
Google Analytics for Firebase	Mobile app analytics
ML Kit for Firebase	ML APIs for mobile

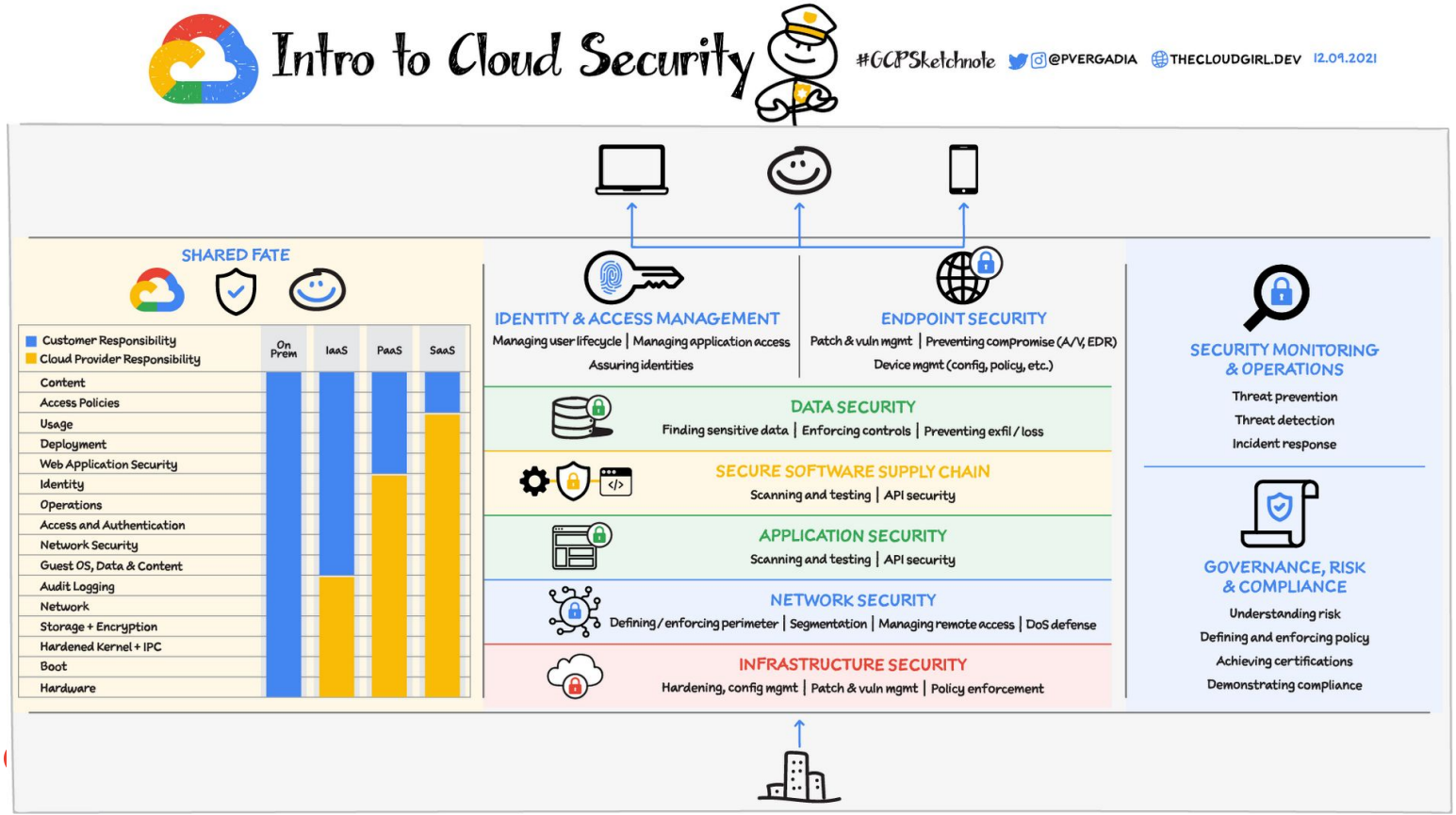
ADDITIONAL RESOURCES

Google Cloud Home Page	cloud.google.com
Google Cloud Blog	cloud.google.com/blog
Google Cloud Platform Podcast	gcppodcast.com
Kubernetes Podcast from Google	kubernetespodcast.com
Google Cloud Reader	podcasts.google.com
Google Cloud Open Source	opensource.google/projects/list#cloud
GCP Medium Publication	medium.com/google-cloud
Apigee Blog	apigee.com/about/blog
Firebase Blog	firebase.googleblog.com
Google Workspace Developers Blog	gsuite-developers.googleblog.com
Google Workspace GitHub	github.com/gsuitedevs
Google Workspace Twitter	twitter.com/gsuitedevs
Google Cloud Certifications	cloud.google.com/certification
Google Cloud System Status	status.cloud.google.com
Google Cloud Training	cloud.google.com/training
Google Developers Blog	developers.googleblog.com
Google Maps Platform Blog	mapsplatform.googleblog.com
Google Open Source Blog	opensource.googleblog.com
Google Security Blog	security.googleblog.com
Kaggle Home Page	www.kaggle.com
Kubernetes Blog	kubernetes.io/blog
Loglens and Network Map	cloud.google.com/about/locations
DORA - Software & Delivery Research	cloud.google.com/devops
Cloud Security Podcast	cloud.withgoogle.com/cloudsecurity/podcast
GCP Sketchnote	goo.gle/gcpsketchnote
Google Cloud Solutions Library	cloud.google.com/solutions
Google Workspace Solutions Gallery	developers.google.com/gsuite/solutions
Google Cloud Support Hub	cloud.google.com/support-hub
GCP Pricing	cloud.google.com/pricing
GCP Pricing Calculator	cloud.google.com/products/calculator
Qwiklabs Home Page	www.qwiklabs.com
Codelabs Home Page	codelabs.developers.google.com
Reddit - www.reddit.com/	
Googlecloud	/r/googlecloud
AppEngine	/r/AppEngine
BigQuery	/r/bigquery
Dataflow	/r/dataflow
Firebase	/r/Firebase
Google Apps Script	/r/GoogleAppsScript
Big Data/Data Analytics Comparisons	cloud.google.com/products/big-data
Compute Product Comparisons	cloud.google.com/products/databases
Database Product Comparisons	cloud.google.com/products/compute
Networking Product Comparisons	cloud.google.com/products/networking
Storage Product Comparisons	cloud.google.com/products/storage

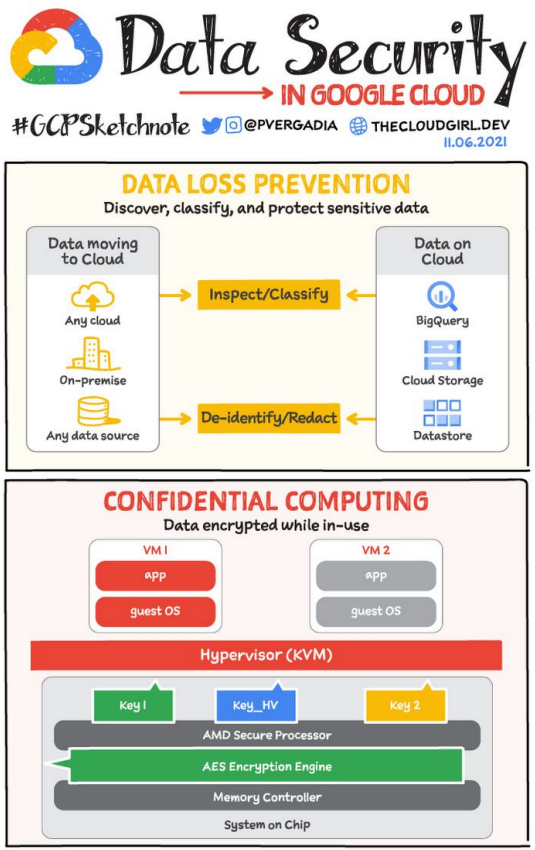
What makes Google Cloud different

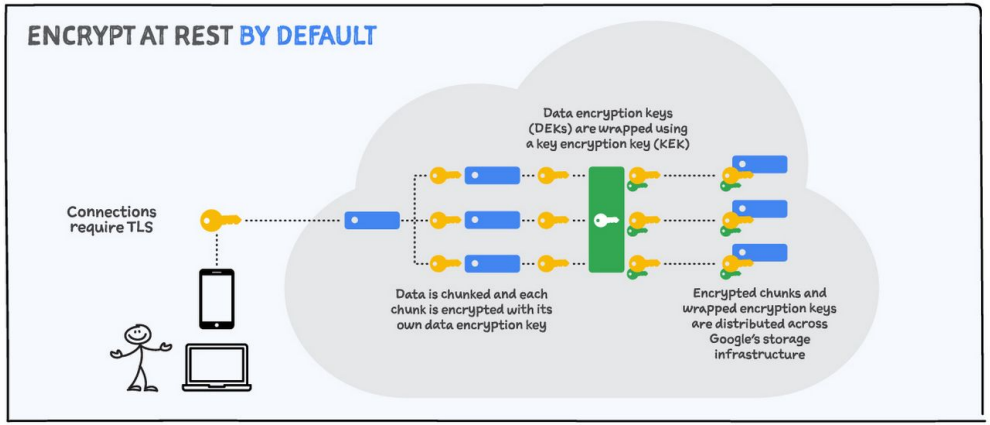
Best-in-class Security & Scalability	Security controls at different levels Compliance and soveraignity
Open Source No vendor lock-in	Enables choice Open source standards
Fully Managed No Ops	Ease of use with serverless. No operations
Embedded AI & ML	Intelligence and AI in Everything. Generative AI
Sustainability	Carbon neutral since 2007, and 24/7 carbon free announcement for 2030

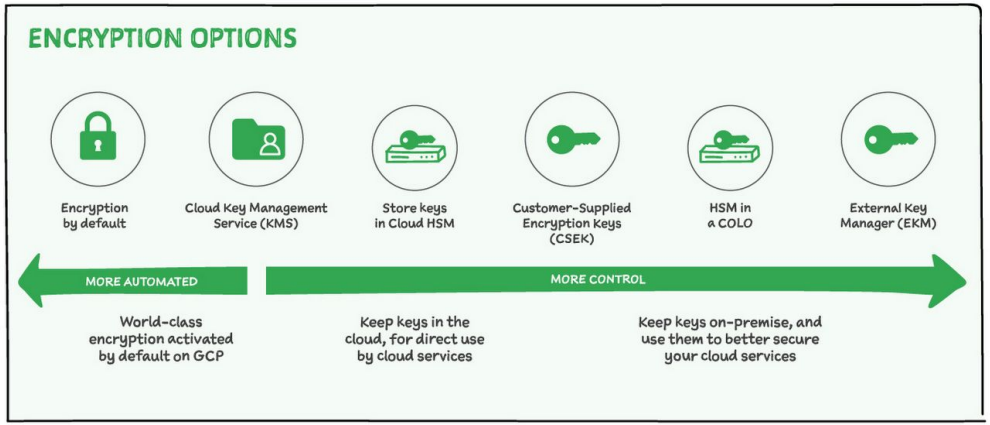
Cloud Security in Google Cloud



Data Security in Google Cloud

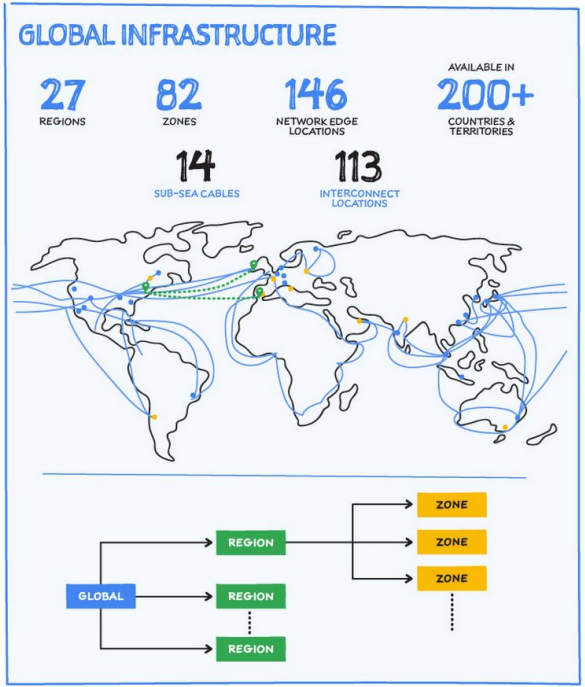


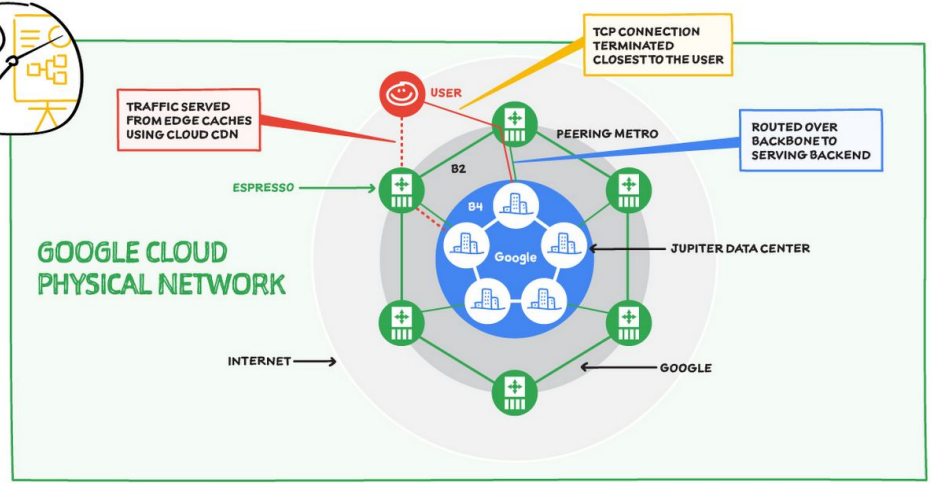


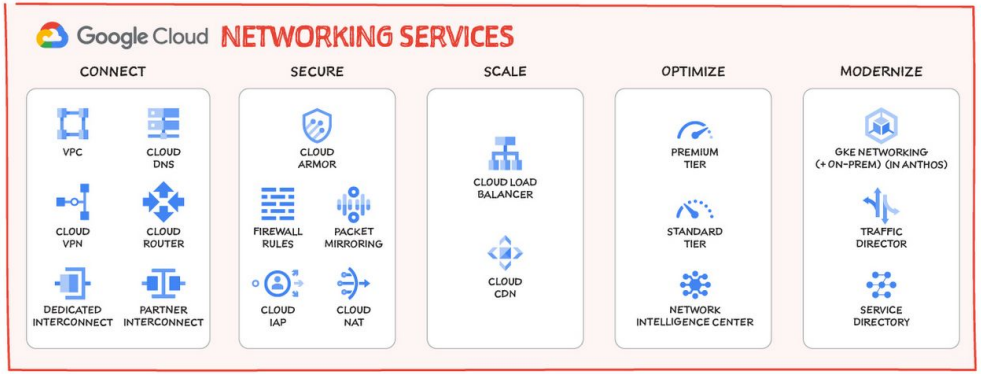


https://cloud.google.com/about/locations

Google Cloud Networking





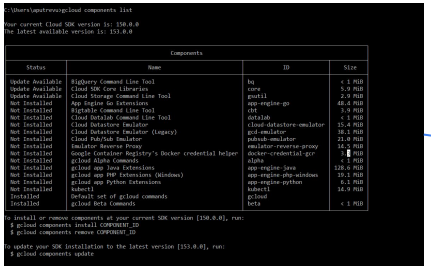


Accessing GCP

https://console.cloud.google.com

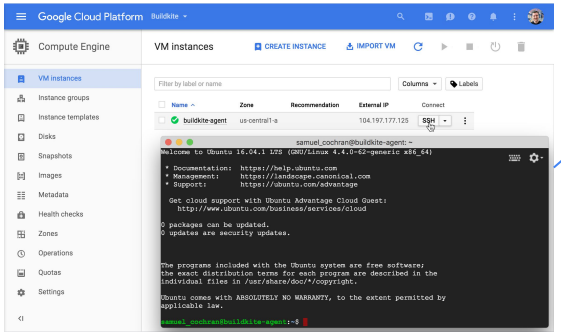
Mobile App

Google Cloud SDK



Restful APIs

Cloud Console & Shell



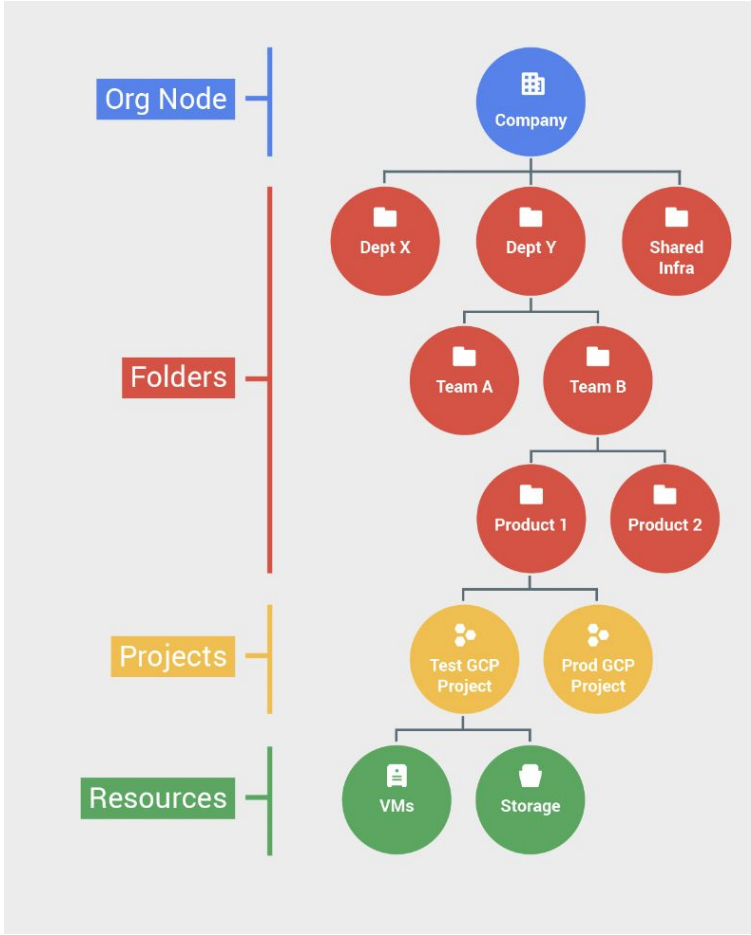

```
https://www.googleapis.com/compute/v1/projects/[PROJECT_ID
]/zones/[ZONE]/instances -d
'{
  "disks":[
    {
      "boot":"true",
      "initializeParams":{

"sourceImage":"https://www.googleapis.com/compute/v1/proje
cts/debian-cloud/global/images/debian-8-jessie-v20160301"
      }
    }
  ],
```

Google Cloud

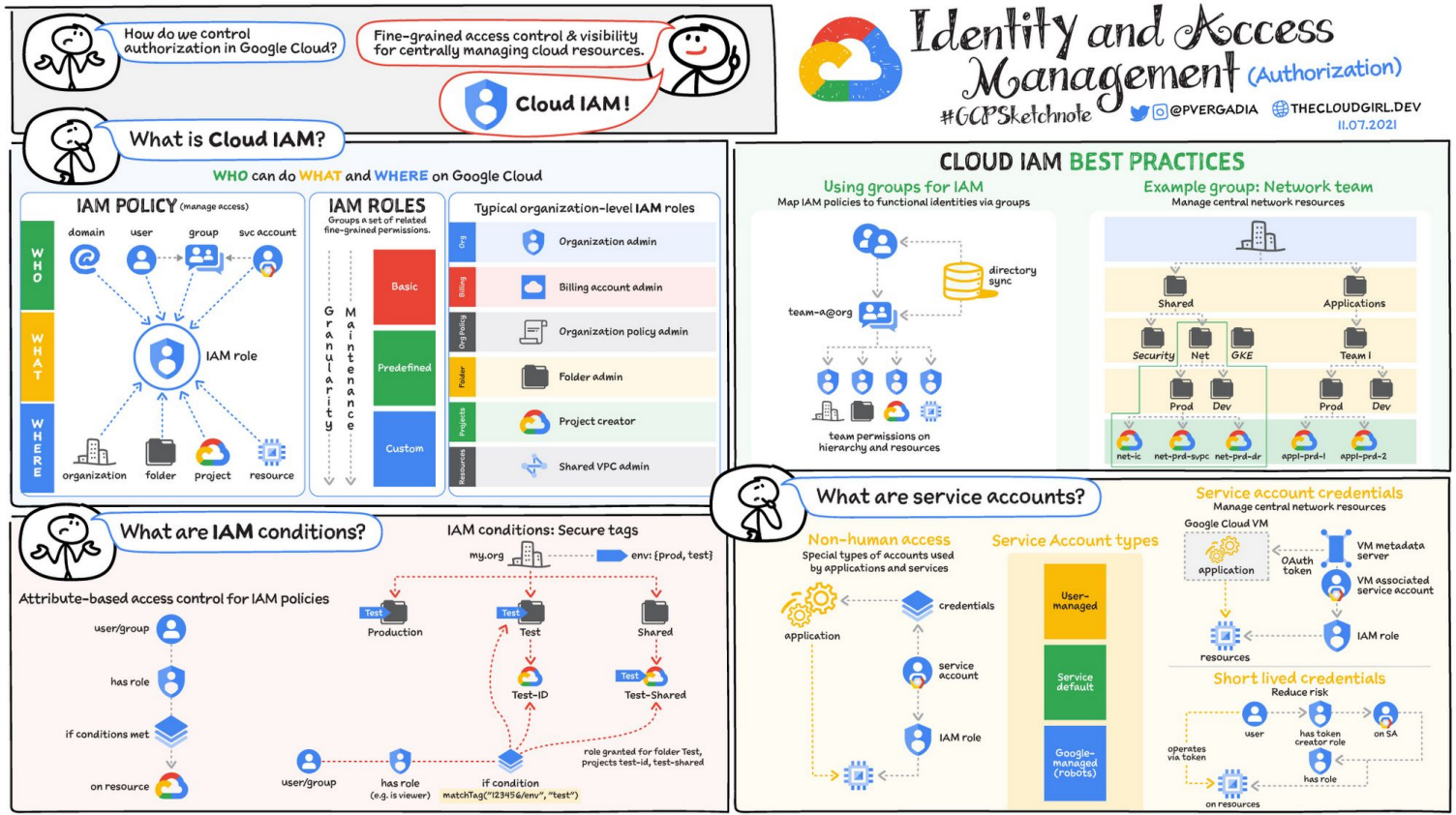
Resource Management in Google Cloud

- Resource Manager provides hierarchical grouping **to organize Cloud Platform resources**
- It manages 3 main resources containers:
  - An Organization
  - Folders
  - Projects
- Resources (VMs, databses, storage...) are created in the context of a **Project**

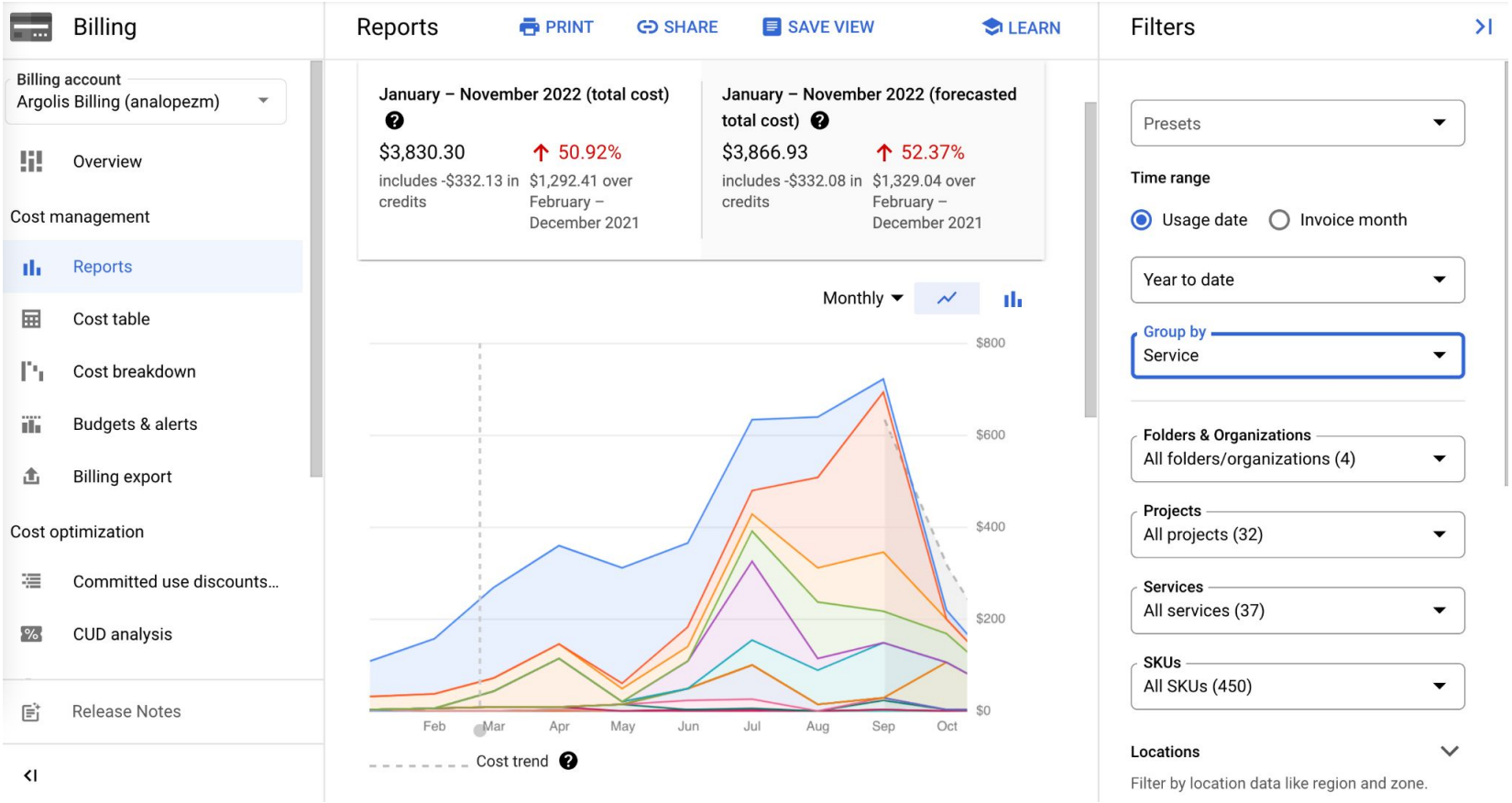


Google Cloud

User & IAM Access Management



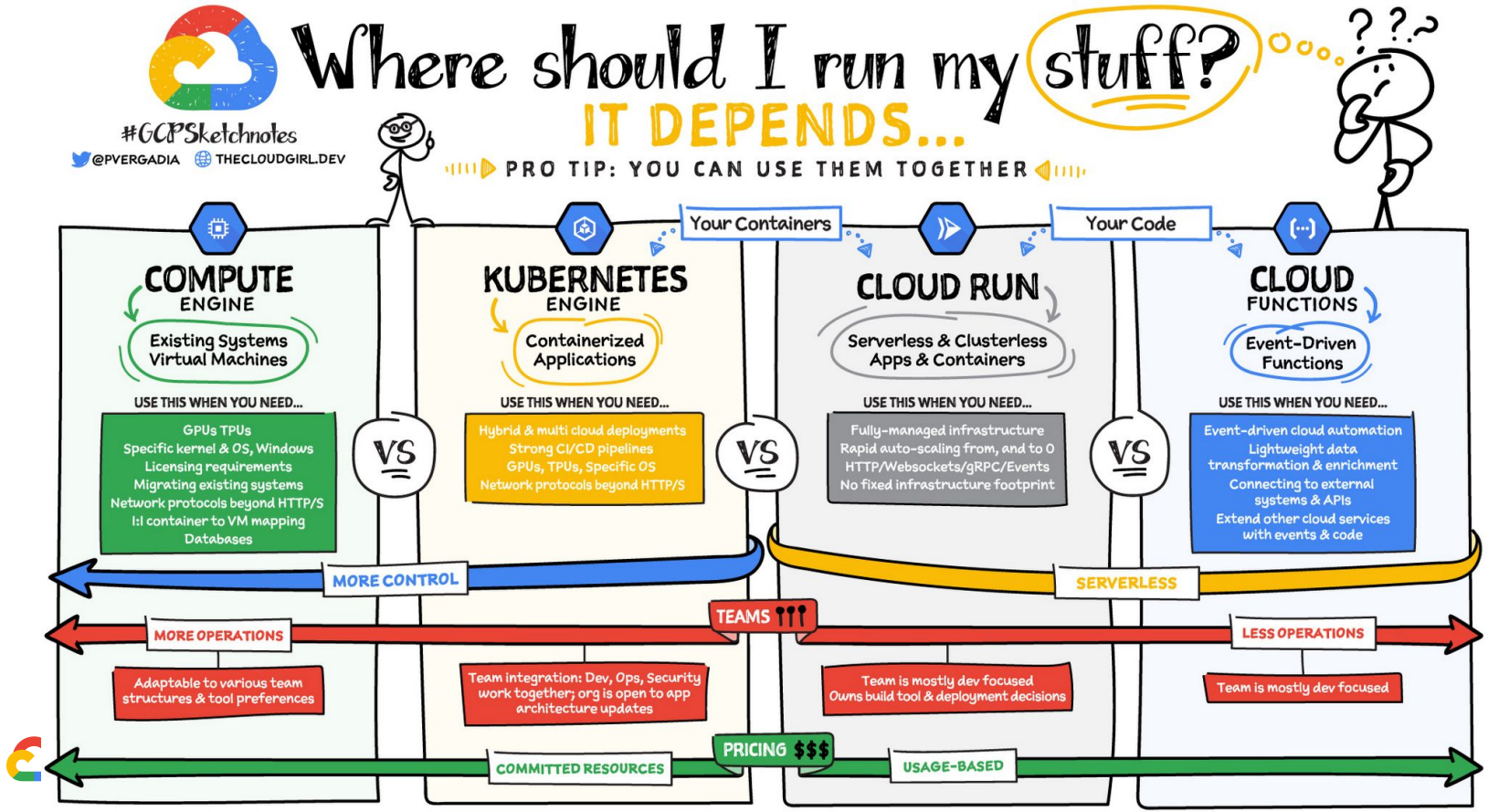
Control costs in GCP



Compute and Storage Services

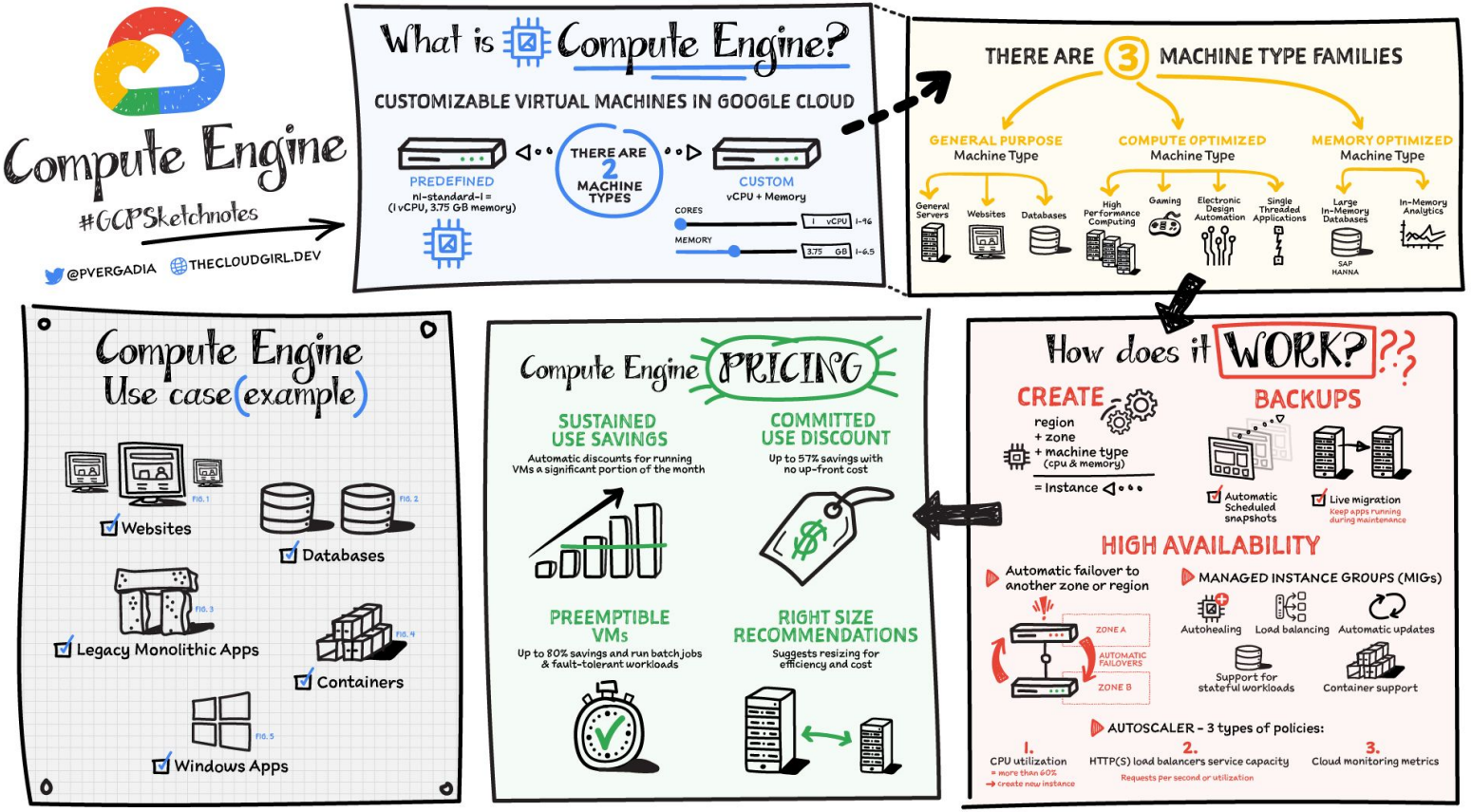
Google Cloud

Choices of compute options

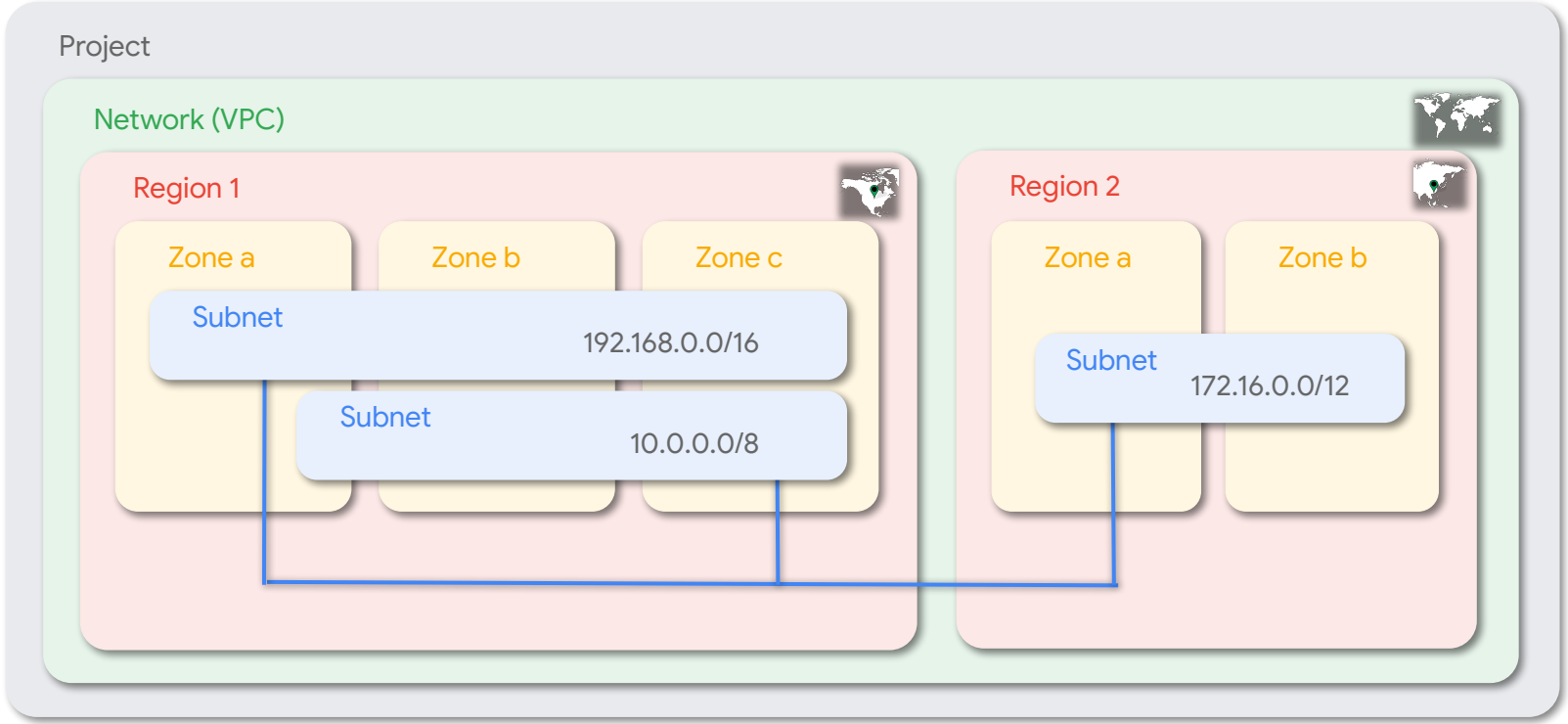


https://cloud.google.com/compute/docs/machine-resource

Google Compute Engine (GCE)

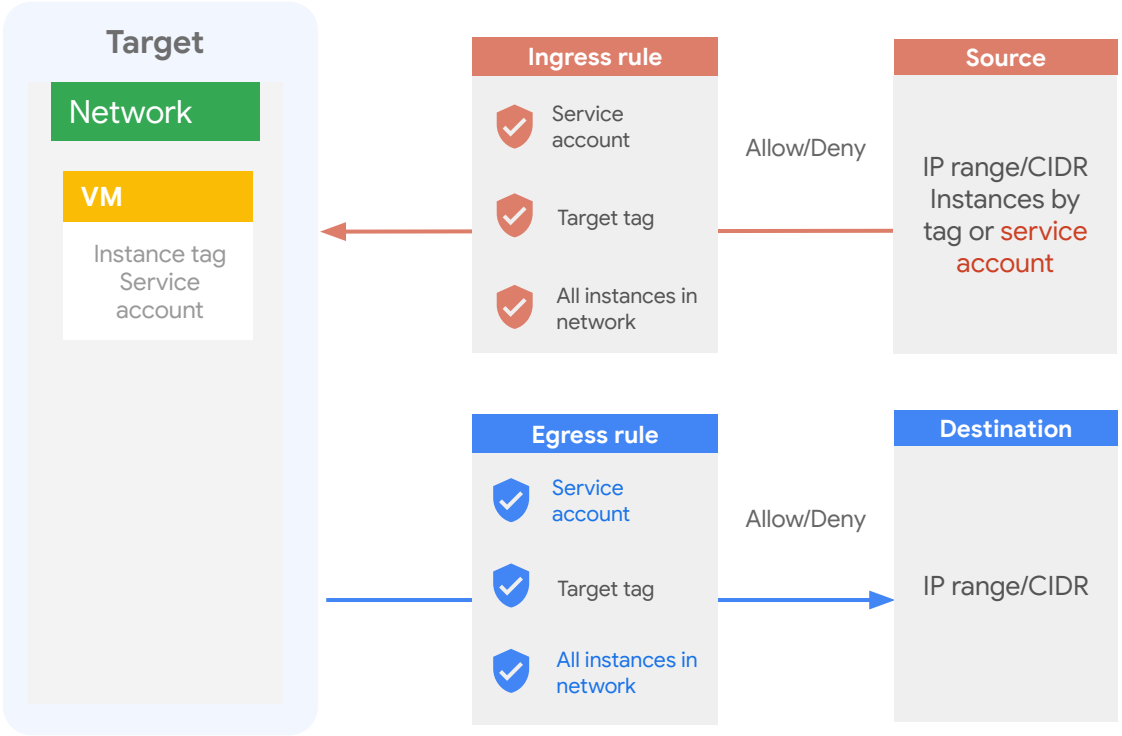


Google VPC



Google Cloud

VPC firewall to protect access



VPC firewall

- **Stateful** with connection tracking
- **Distributed**: enforced on underlying host

Control paths

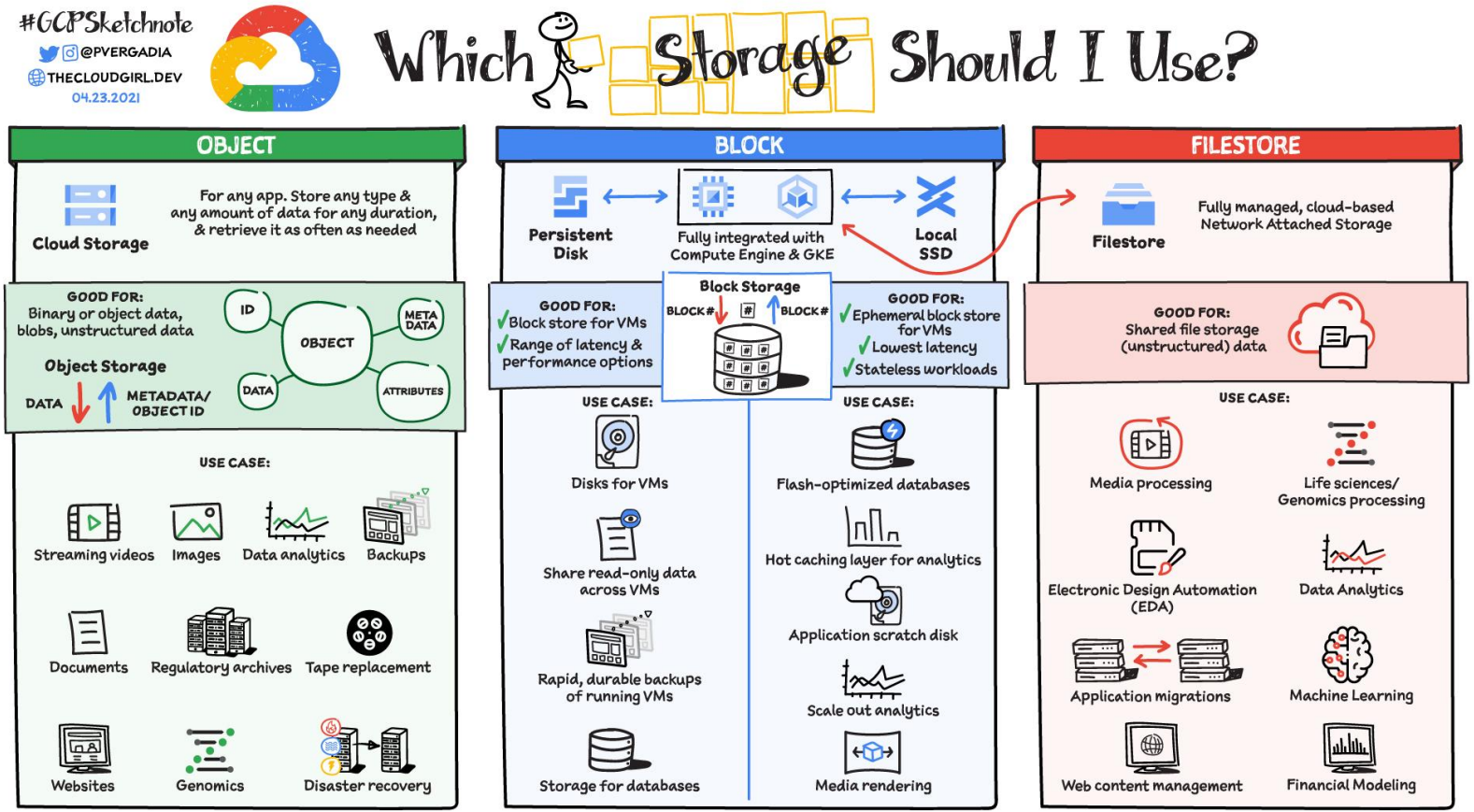
- VM <-> VM
- VM <-> Internet
- VM <-> On-prem

Implied rules

- Ingress deny
- Egress allow

Google Cloud

Which storage type?



Cloud Storage

#GCPSketchnote

@PVERGADIA THECLOUDGIRL.DEV

8.8.2020

What is Cloud Storage?

A GLOBAL, SECURE AND SCALABLE OBJECT STORE

GLOBAL

SECURE

SCALABLE

.png .jpg

.mpg .mov

.txt

other

How does it WORK?

WORKFLOW

Image

Bucket

Organization

Project

4 STORAGE CLASSES

Based on Budget, Availability and Access Frequency

STANDARD	NEARLINE	COLDLINE	ARCHIVE
Frequent access High Availability	Once a month	Once a quarter	Once a year

New Version >30 Days >90 Days

OBJECT LIFECYCLE MANAGEMENT

AUTO VERSIONING v1 v2 v3

Cloud Storage Use case example

STANDARD — Website + Streaming + Document

NEARLINE — Backups + Disaster Recovery

COLDLINE — Backups + Disaster Recovery

ARCHIVE — Long Term Archiving

STREAMING DATA

BATCH DATA

AI PLATFORM

BIGQUERY

SECURITY for Cloud Storage

- Encryption at rest
- Bring your own encryption key
  - CMEK - Customer Managed
  - CSEK - Customer Supplied

Cloud Storage PRICING

STANDARD	NEARLINE	COLDLINE	ARCHIVE
$$$$	$$$	$$	$

Automatic Redundancy Frequent Access

How to USE Cloud Storage

ONLINE TRANSFER	TRANSFER SERVICE	TRANSFER APPLIANCE
gsutil, API, UI <gsutil cp image.png gs://my-bucket>	Transfer data from other clouds & on-premise	Hardware for >100tb data transfer

image.png

API

Persistent Disk

#GCPSketchnotes

@PVERGADIA THECLOUDGIRL.DEV

3.29.2020

What is Persistent Disk?

HIGH PERFORMANCE BLOCK STORAGE

- ✓ Transparently Resize
- ✓ Quick backup

Persistent Disk

SSD/HDD storage

Compute Engine

Kubernetes Engine

✓ Supports Simultaneous readers

Which Block storage is right for your app?

LOWER COST ← → PERFORMANCE

Standard PD	Balanced PD	Performance (PD-SSD)	Extreme PD	Local SSD
HDD	SSD	SSD	SSD	Local SSD
Standard throughput	Best price per GB	Best price per IOPS	Best performance	Lowest Latency
Current 'cost sensitive'	Best fit for most workloads	Performance sensitive workloads	Uncompromising performance	For caches and hot layers

(How to pick) based on availability needs.

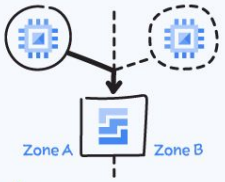
Local SSD

- ✓ Ephemeral
- ✓ Stateless workloads, or replication managed at app or database layer

Persistent Disk

- ✓ Durable, support snapshots
- ✓ Most workloads

Regional Persistent Disk



- ✓ Durable & Highly Available
- ✓ Mission-critical workloads with RPO/RTO near 0

Persistent Disk «Use case example»

Standard PD	Balanced PD	Performance (PD-SSD)	Extreme PD	Local SSD
Cost sensitive workloads	Most enterprise apps	Most databases	SAP HANA	Scale out analytics
Scale out analytics (Hadoop, Kafka)	LOB apps	Persistent cache	Oracle	Media rendering
	Boot disks	Scale-out analytics	Largest in-memory DBs	Other use cases where ephemeral scratch space is required
	Webserving			

Filestore #GCPSketchnote

@PVERGADIA THECLOUDGIRL.DEV

3.25.2021

We need a shared file storage!

I know just the tool!

Filestore!

What is Filestore?

FULLY MANAGED NETWORK ATTACHED STORAGE (NAS)

for Unstructured data (Files)

Filestore

How does Filestore work?

SHARED FILE STORAGE

Filestore

Mount on GCE or GKE instances

Compute Engine

Kubernetes Engine

Are there different capacity options?

TWO TIER

BASIC

Low-mid capacity
test/dev

SCALE UP ARCHITECTURE

EXAMPLE USE CASES

- Web content management
- Home directories
- Software development
- Content sharing
- Enterprise applications

HIGH SCALE

Large capacity,
High performance

SCALE OUT ARCHITECTURE

EXAMPLE USE CASES

- Electronic Design Automation (EDA)
- Media rendering
- Life sciences / genomics
- Manufacturing
- Data analytics
- Financial modeling

What are it's benefits & use cases?

BENEFITS →

- Low latency
- High & consistent Performance
- Resize - while in use

USE CASES

- Application migrations
- Web content management
- Media processing
- Electronic design arts (EDA)
- Data Analytics
- Home directories
- Life sciences/ Genomics processing
- Financial Modeling

#GCPSketchnote
@PVERGADIA
THECLOUDGIRL.DEV
03.30.2021

Options to move data to Google Cloud

GCS TRANSFER TOOLS

For small transfers upto a few TBs

TRANSFER SERVICE

Large-scale online transfer

Cloud data | On-premises

TRANSFER APPLIANCE

Large scale transfers from low bandwidth networks

BIGQUERY DATA TRANSFER SERVICE

Data from SaaS & 3rd party apps

HOW DO THEY WORK?

Upload data directly to GCS

TOOLS - GSUTIL, UI, JSON API

S3 / Azure Cloud

TOOLS
UI, Client Libraries, HTTP REST API

On-premises

TOOLS
On-premise agent with GUI

1. Request Appliance
2. Upload data SFTP, SCP or NFS
3. Ship to Google
4. GCS
5. Erase NIST 800 standard

S3
Teradata
RedShift
Google SaaS apps
GCS
3rd party

SOME COOL THINGS TO KNOW

GSUTIL - Fast multi threaded

rsync

Streaming

Scheduled & Incremental updates
Scale to 10's of Gbps

Data moves over Google's high-bandwidth network pipes

Minimize transfer time max available bandwidth performance optimizations

All SSDs
Minimal software
Performance features
Multiple network connectivity options

Automatic Schedule

Mirror or Append data

SECURITY

HTTPS encrypted TLS connections

Cloud sources accessed using private credentials

All bytes checksummed & encrypted in-flight

AES 256 encryption on the appliance
Tamper resistant, Ruggedized
Customer Managed Encryption Key (CMEK)

HTTPS encrypted TLS connections

USE CASES FOR DATA TRANSFER

Data Center Migration

Decommission Tape Libraries & Infrastructure

Machine Learning

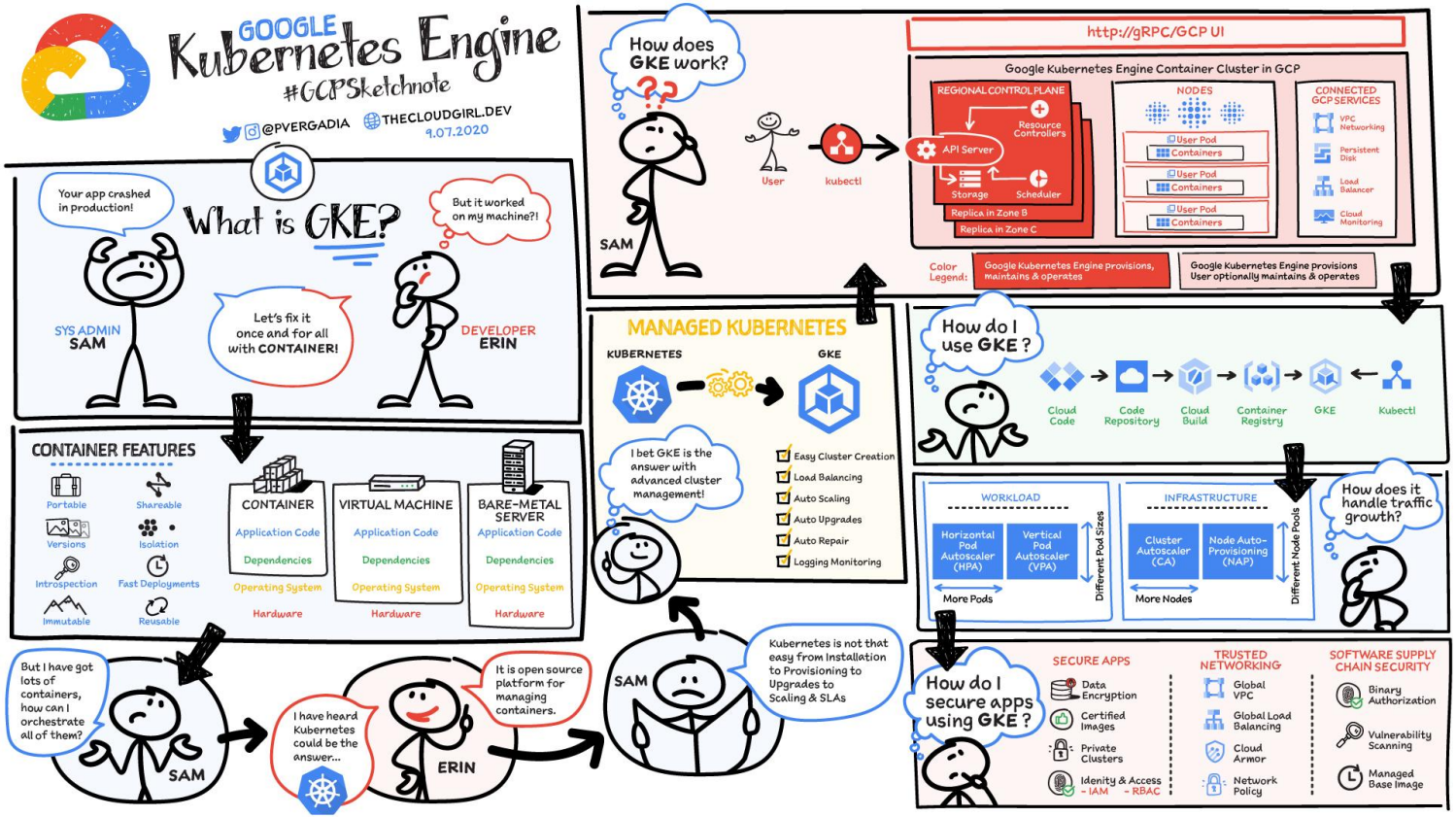
Storage & Delivery

Backup & Archival

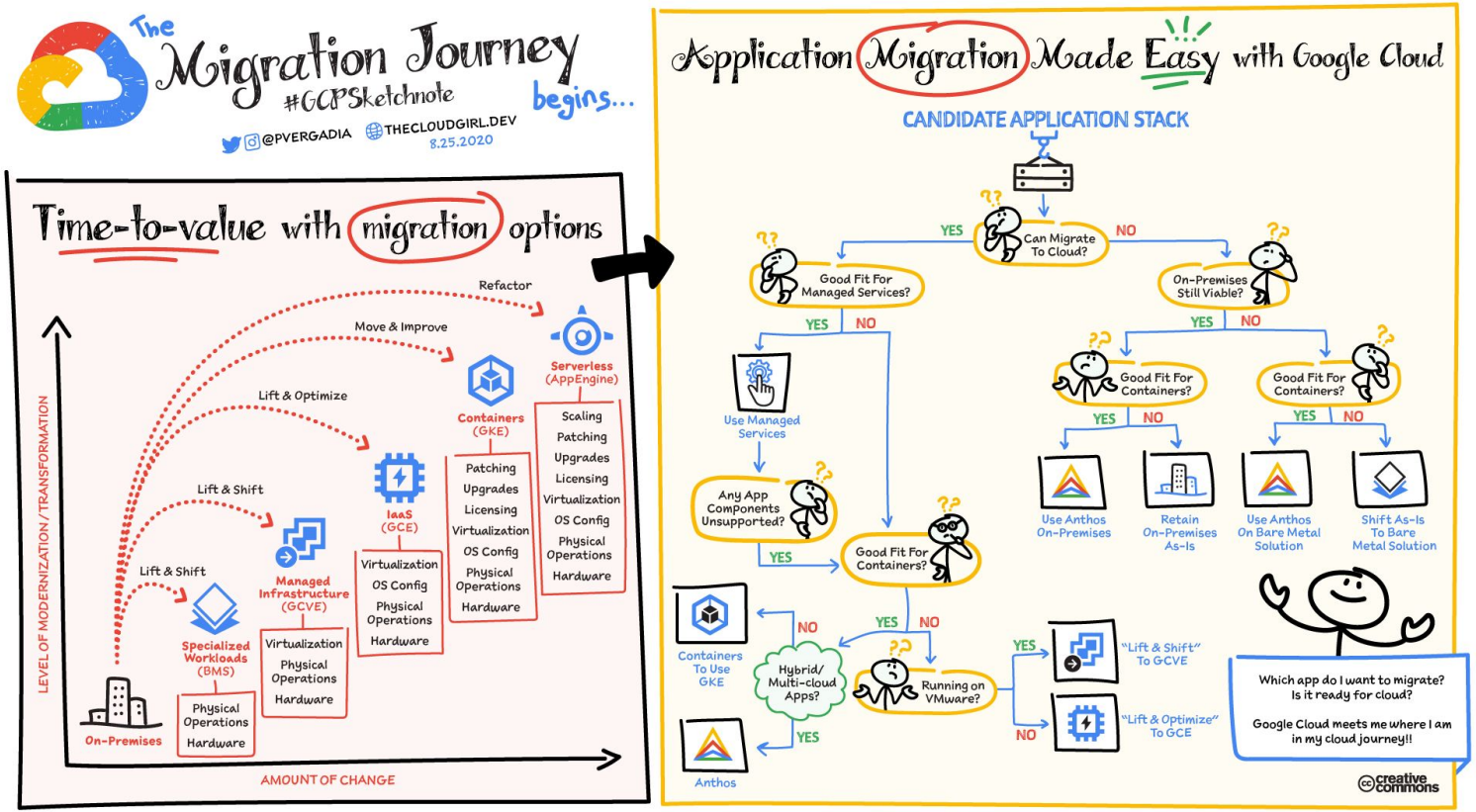
App Development in Google Cloud

Google Cloud

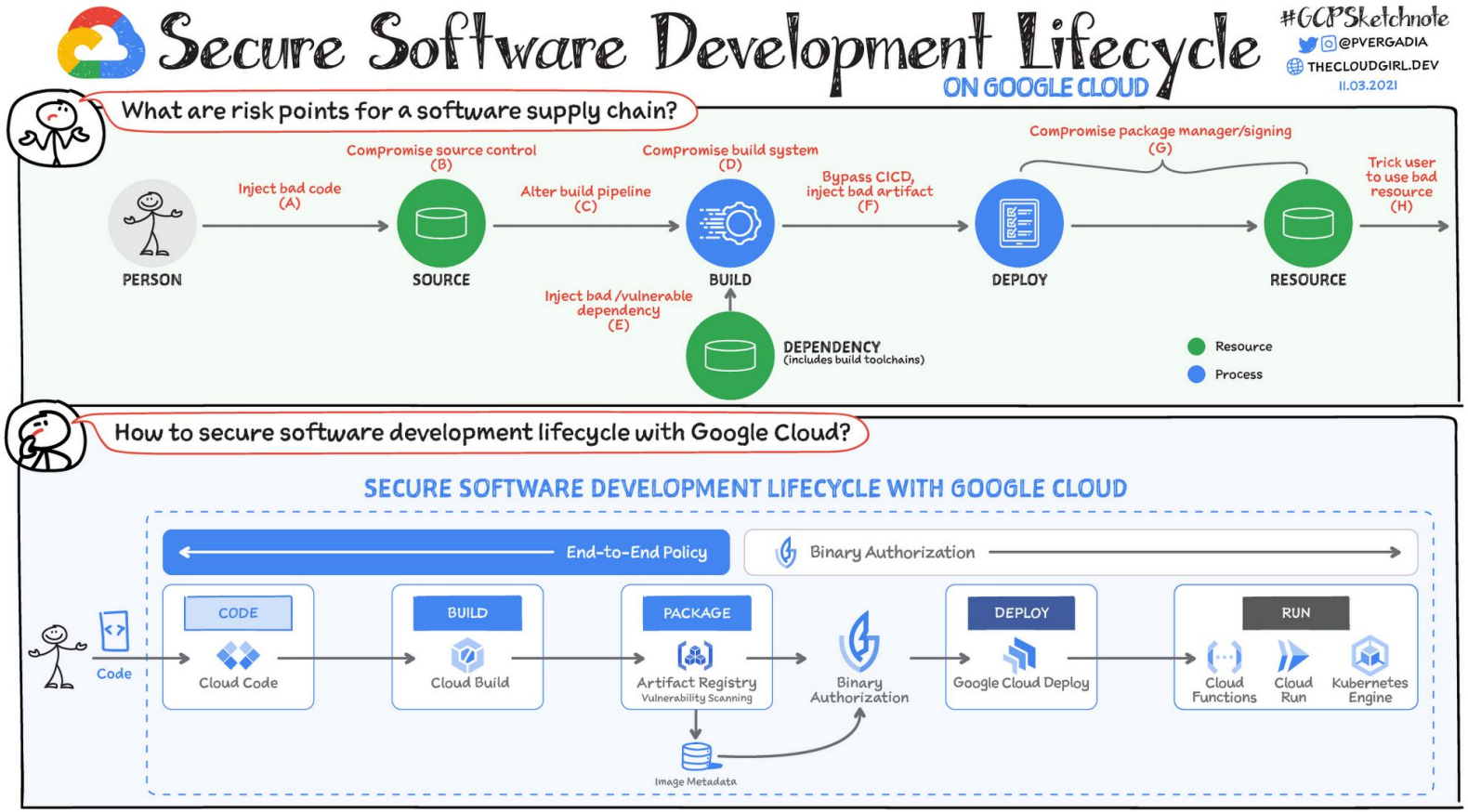
Google Cloud Kubernetes Engine



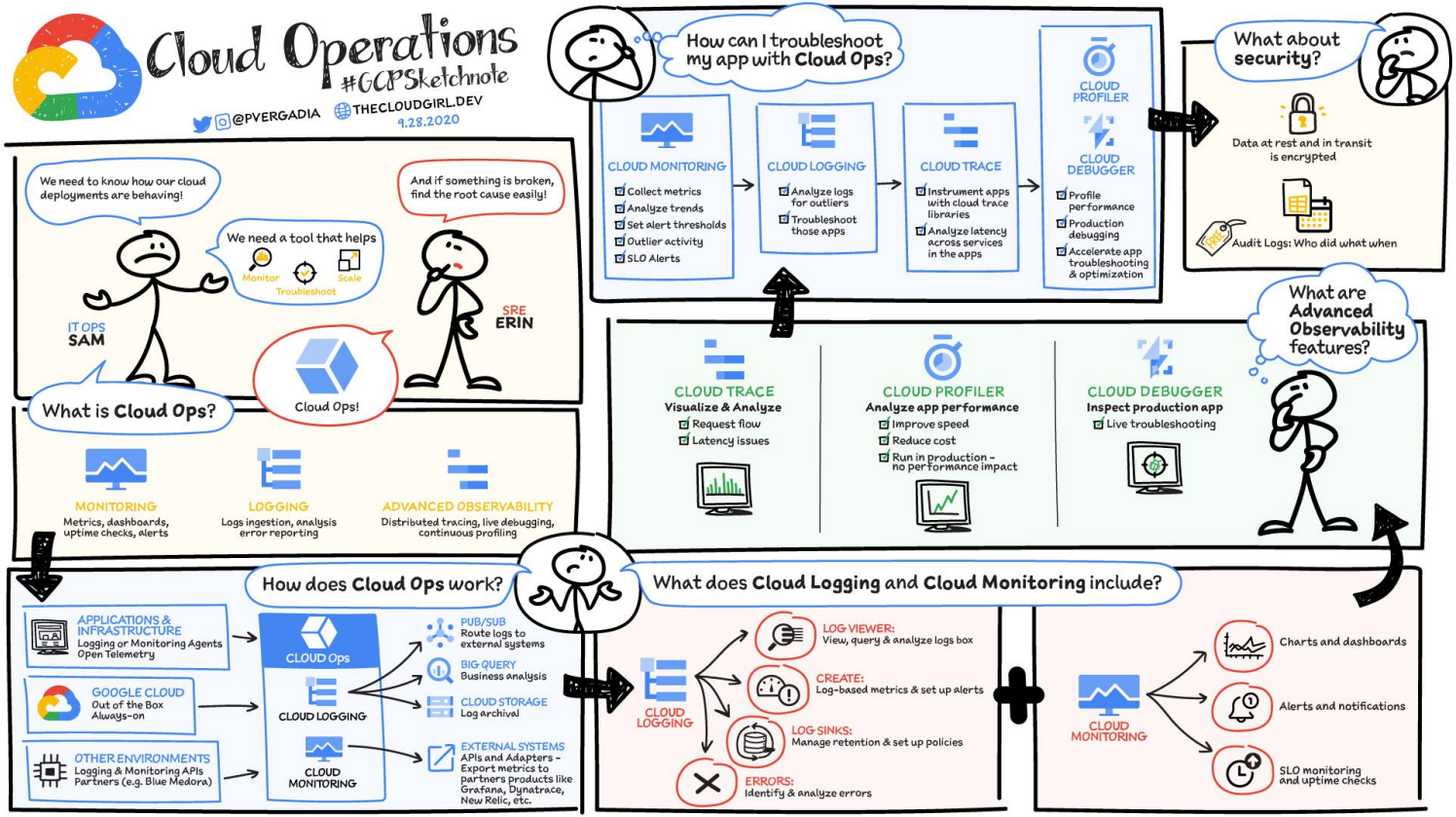
Migration to Google Cloud



CI / CD in Google Cloud



Cloud Operations

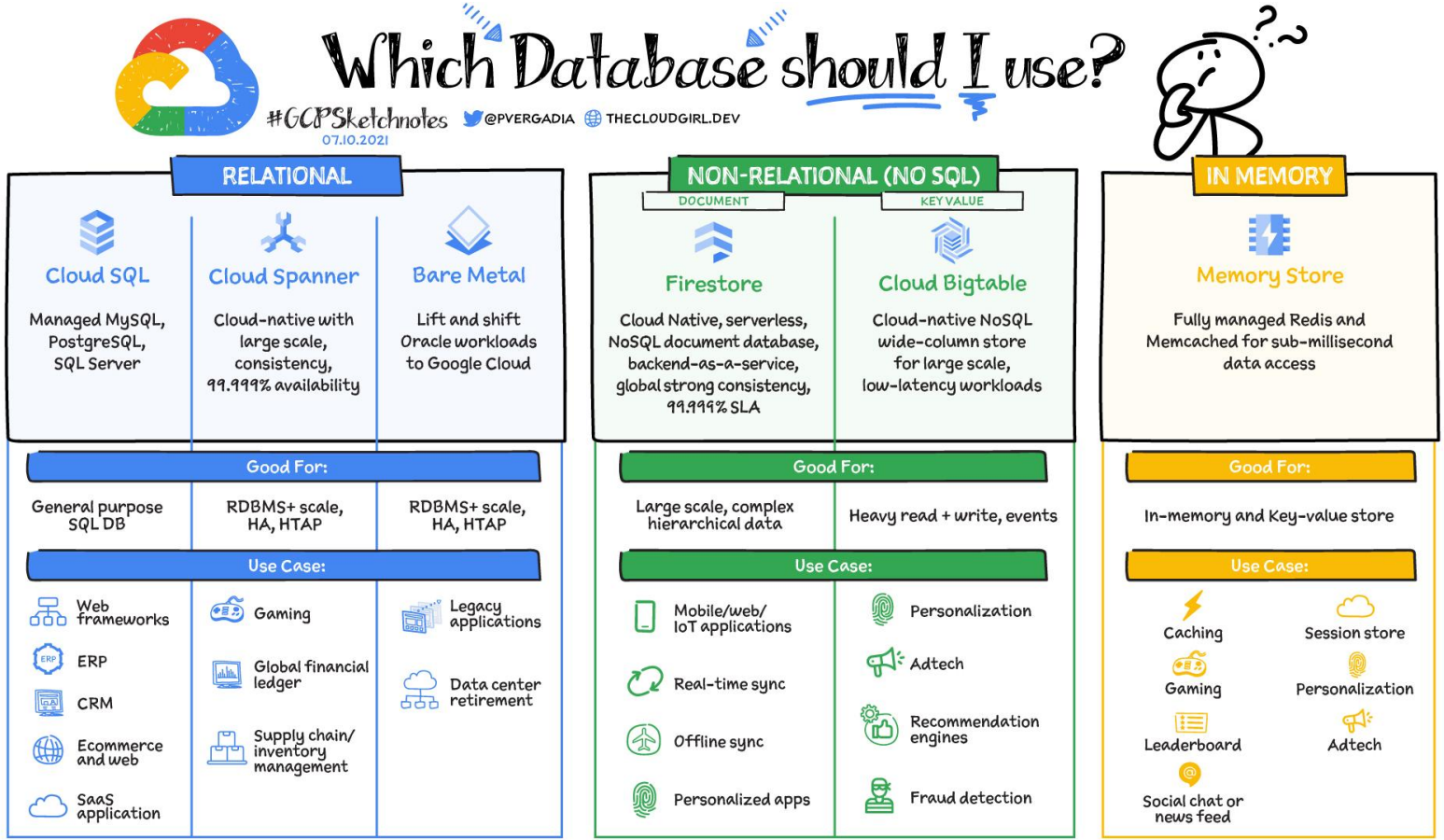


Google Cloud

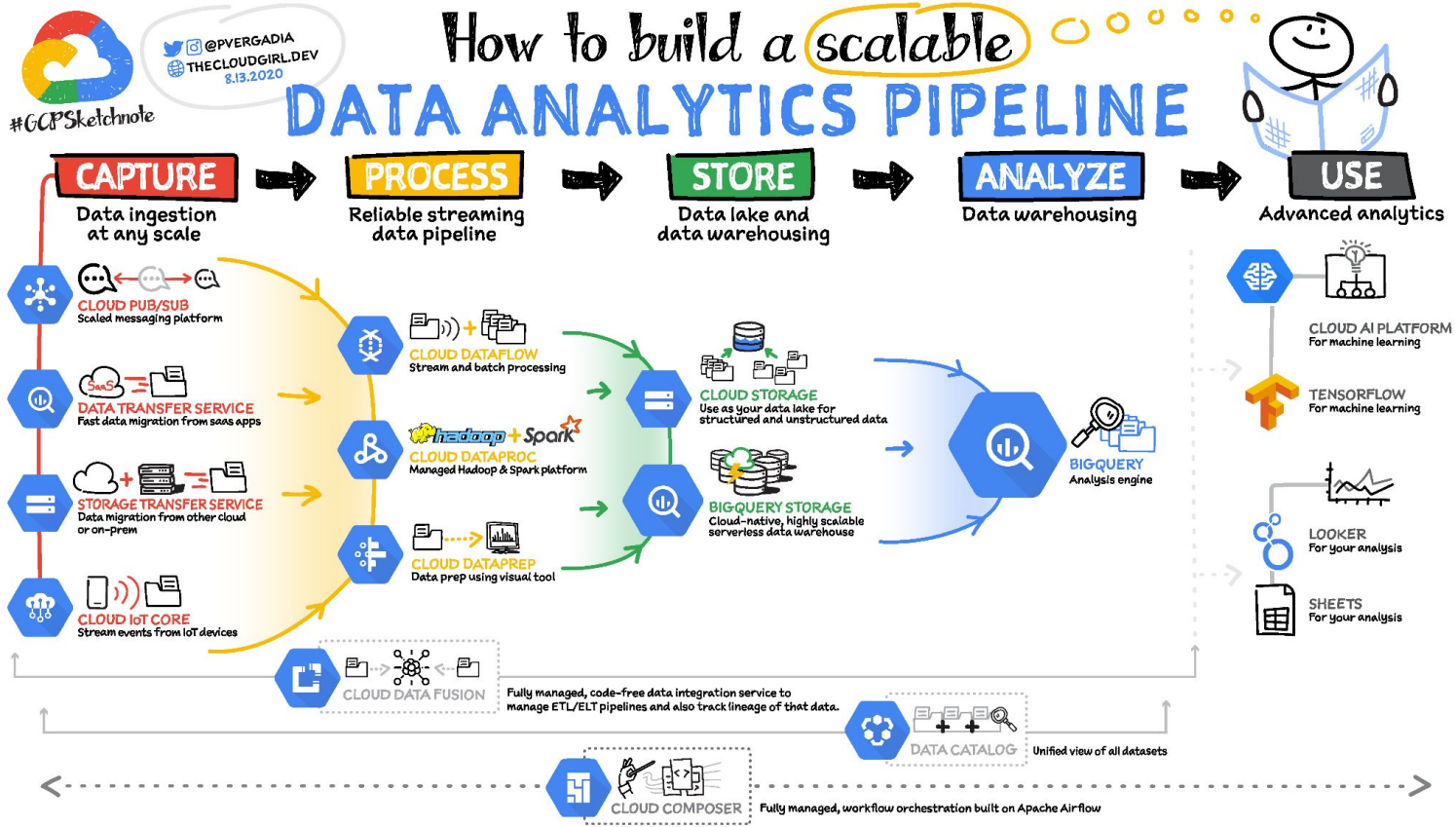
BigData & Analytics

04

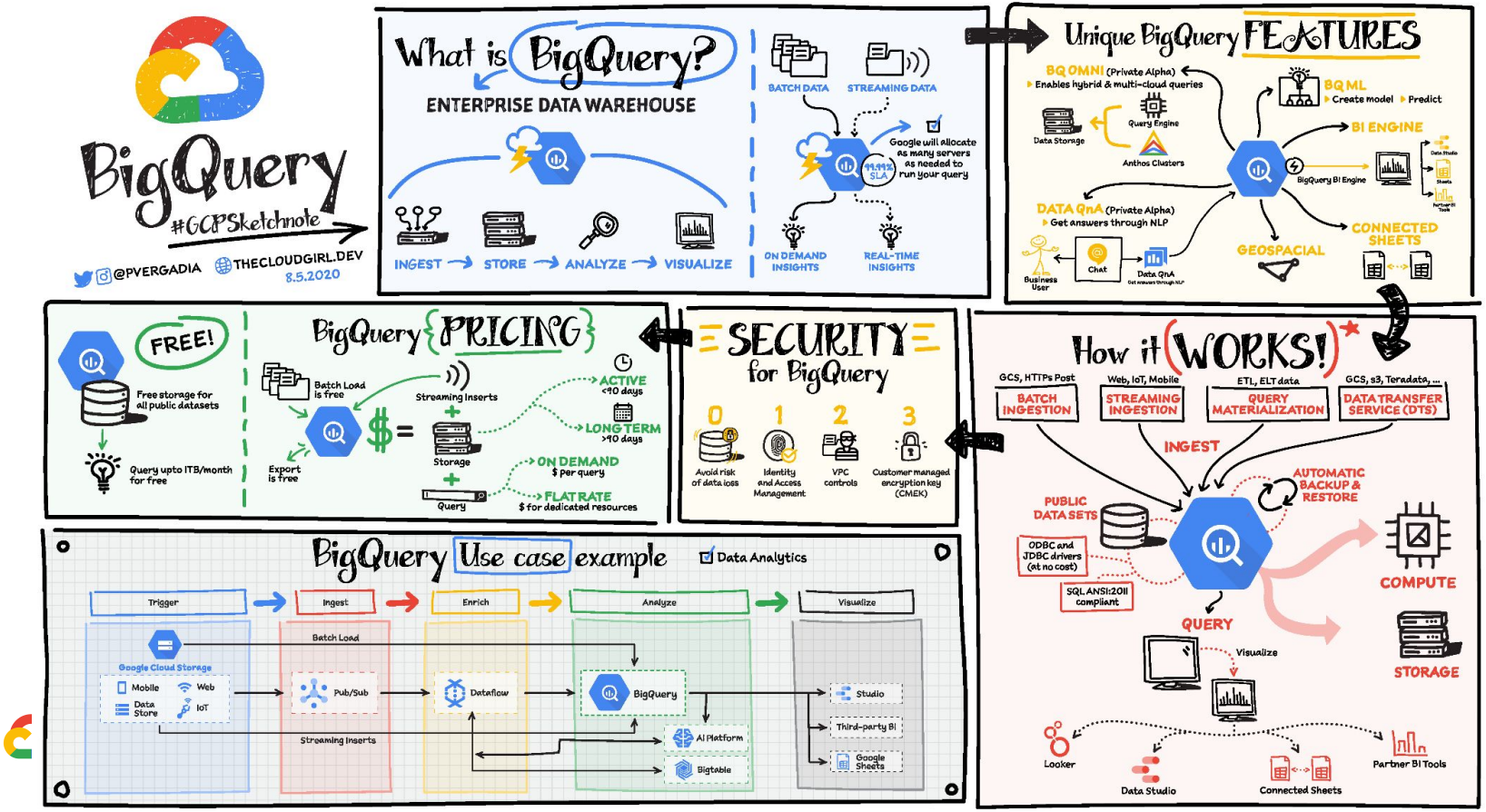
GCP Managed database services



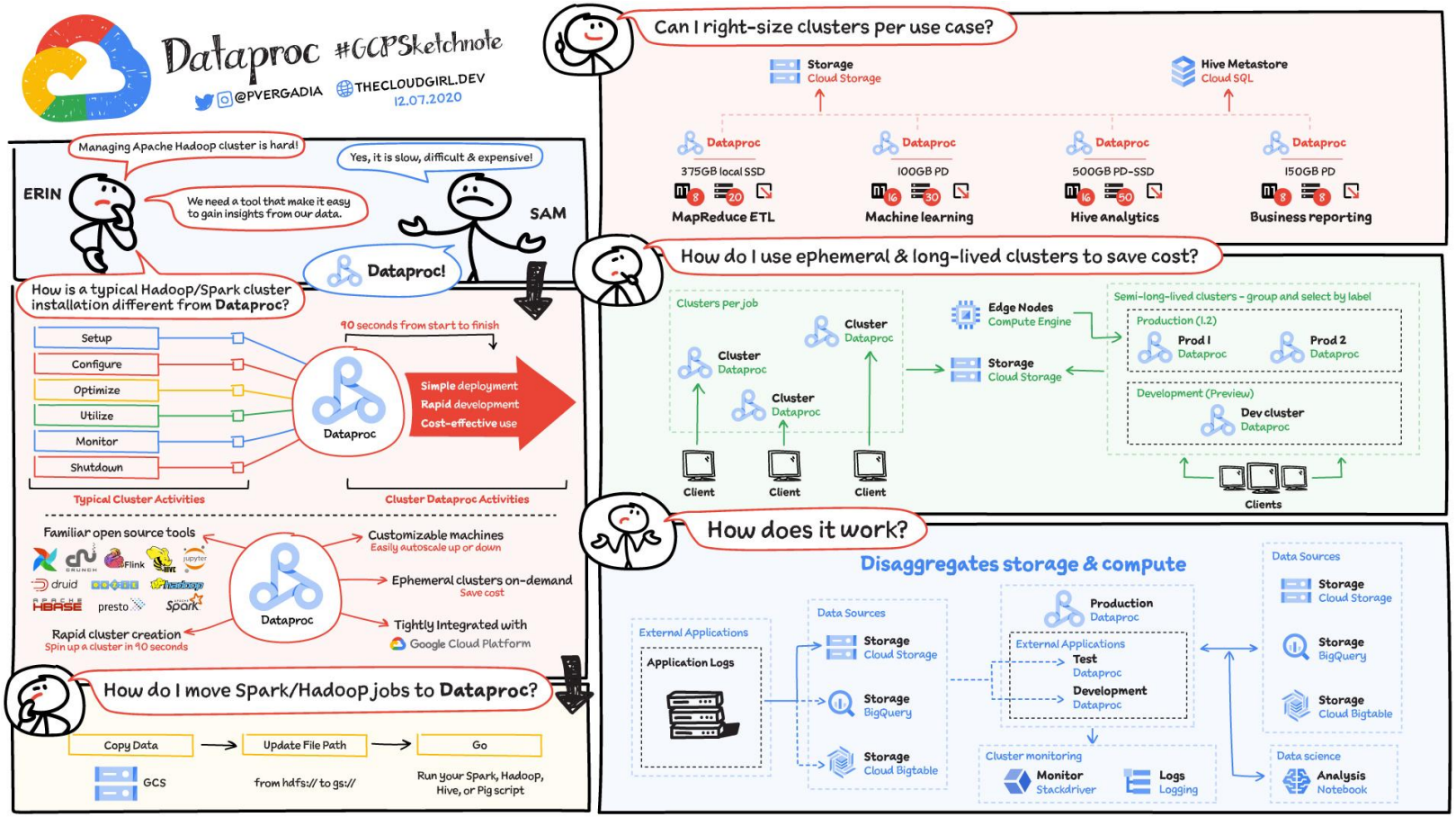
Data Analytics in Google Cloud



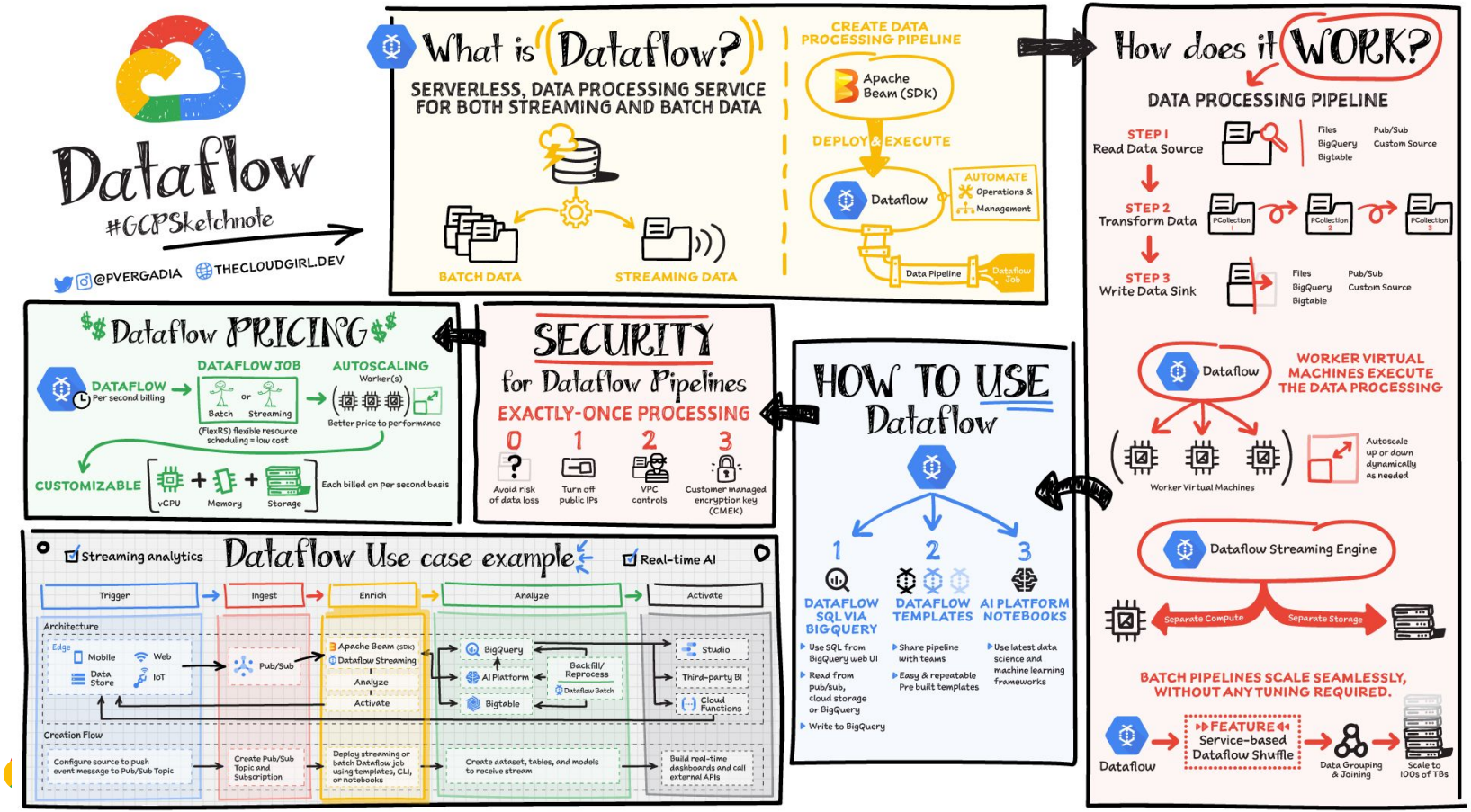
Google BigQuery - cloud datawarehouse SQL based



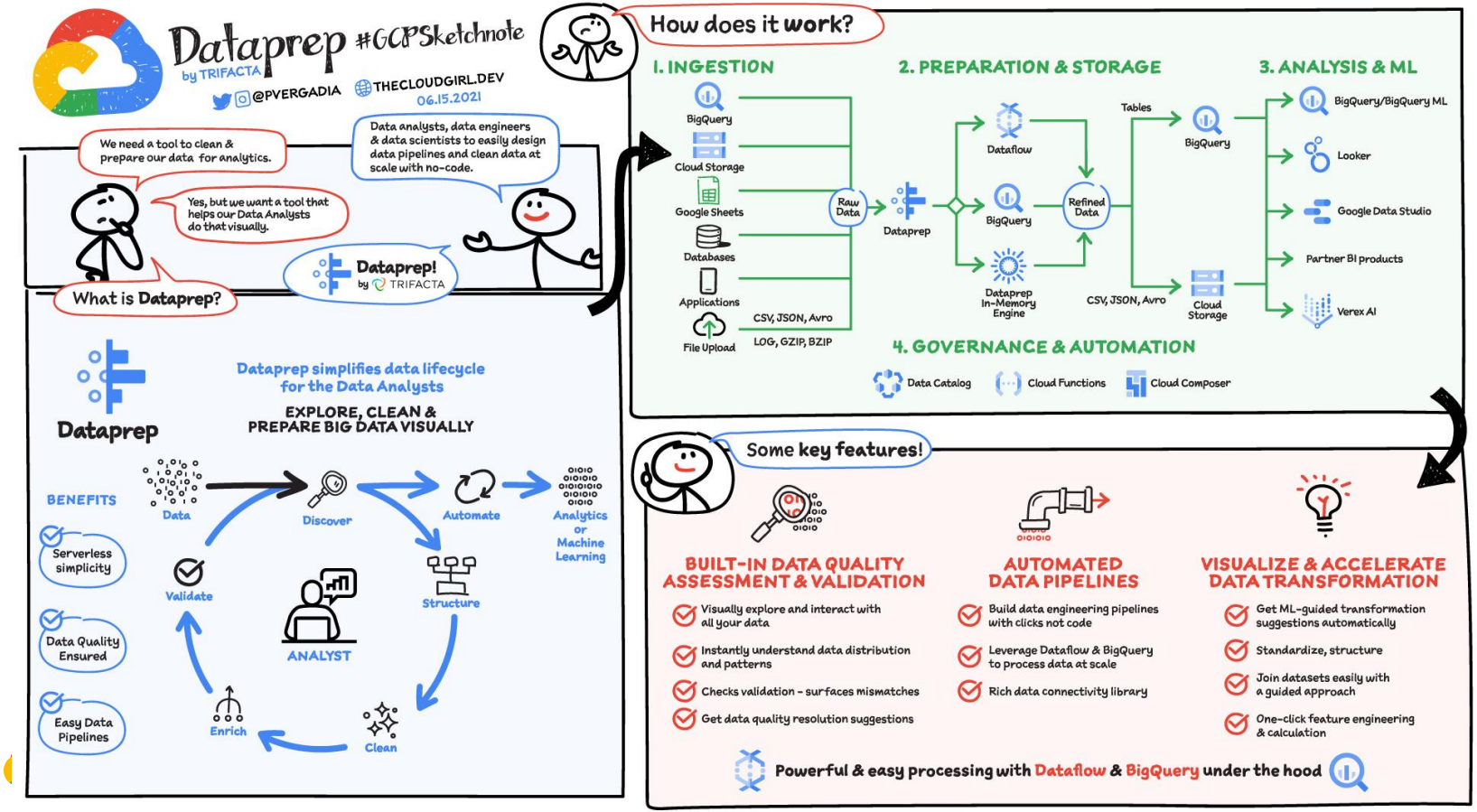
Cloud Dataproc - managed Hadoop/Spark cluster



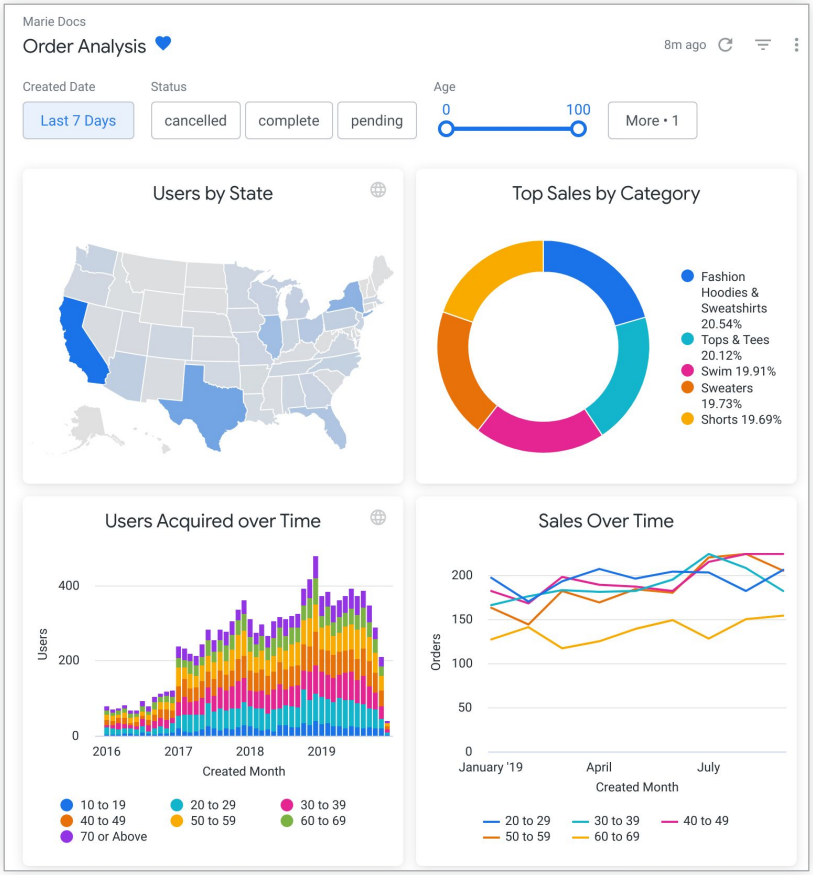
Cloud Dataflow - batch and stream processing

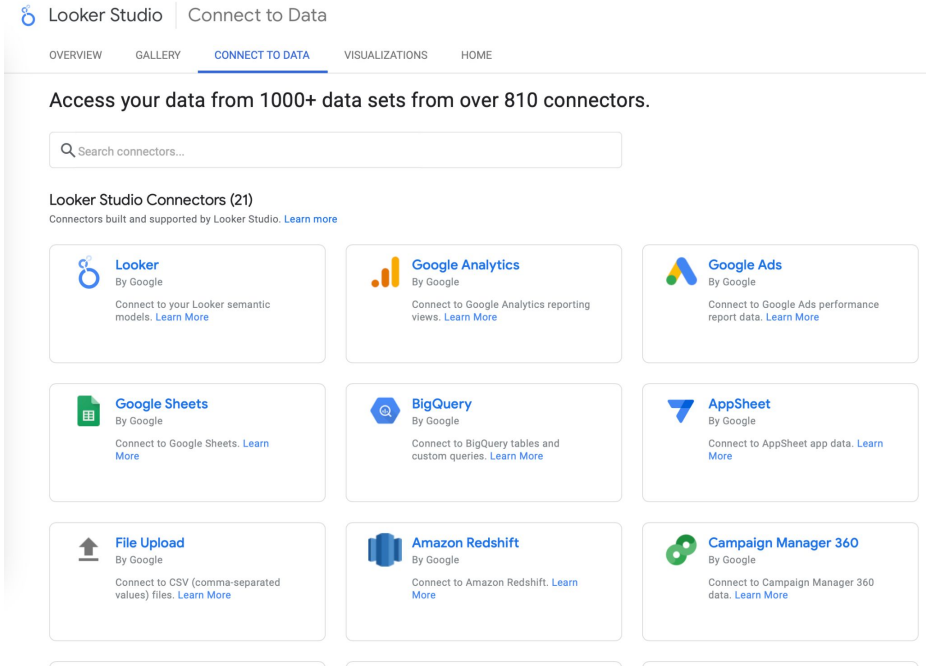


Cloud Dataprep - visually clean and prepare your data



Looker - Business Intelligence





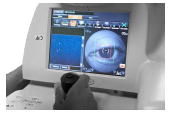
Artificial Intelligence and Machine Learning

Google Cloud

Google is the pioneer in AI

2015

Google DeepMind AlphaGo defeats Go champion



2016

Google's DeepMind helps detect eye disease

2017

Google invents Transformer kickstarting LLM revolution

2018

Google's groundbreaking large language model, BERT

T5

2019

Text-to-Text Transfer Transformer LLM 10B P Model Open Sourced

2020

Google LaMDA Model Trained to converse

2022

AlphaFold predicts structures of all known proteins

Bard

2023

A conversational AI Service powered by LaMDA.

3,000 Researchers
7,000 Publications

Responsible AI

Built & Tested for Safety

Privacy in design

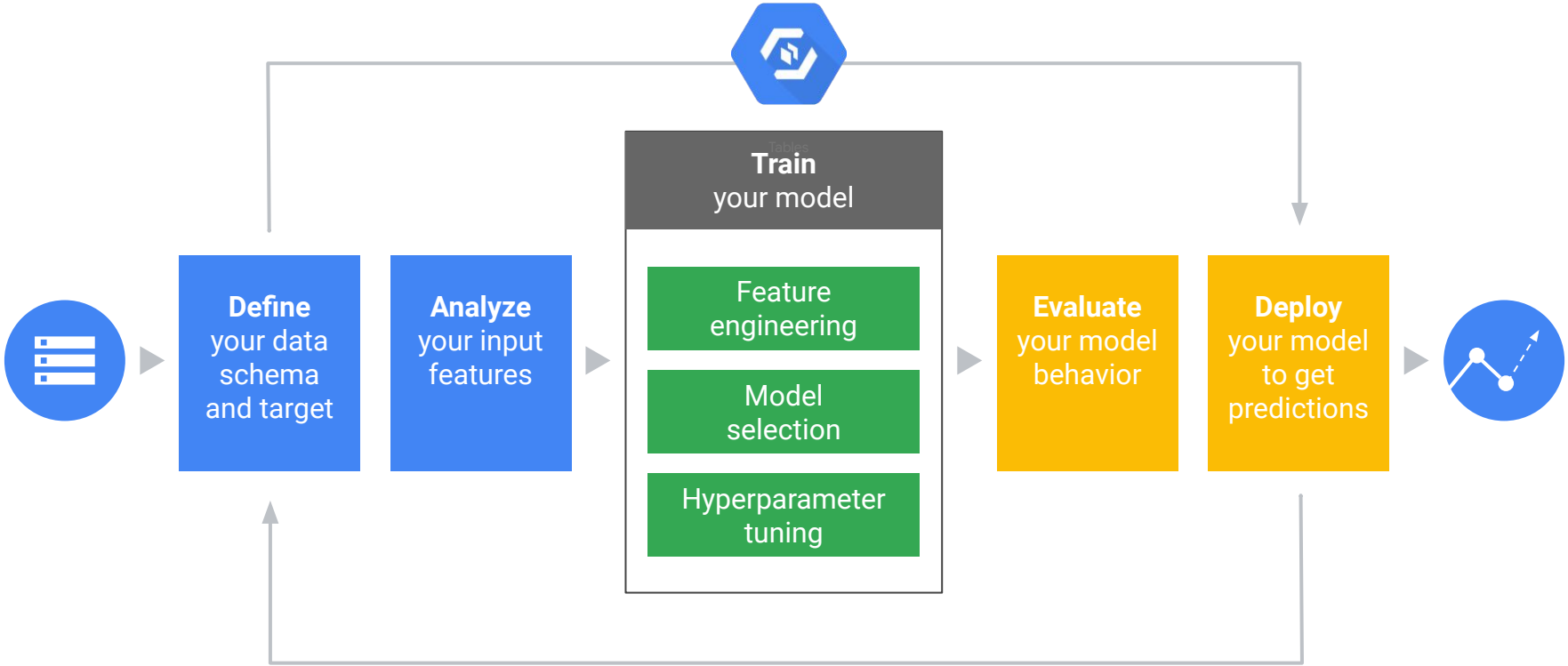
Upholds high scientific standards

Accountable to People

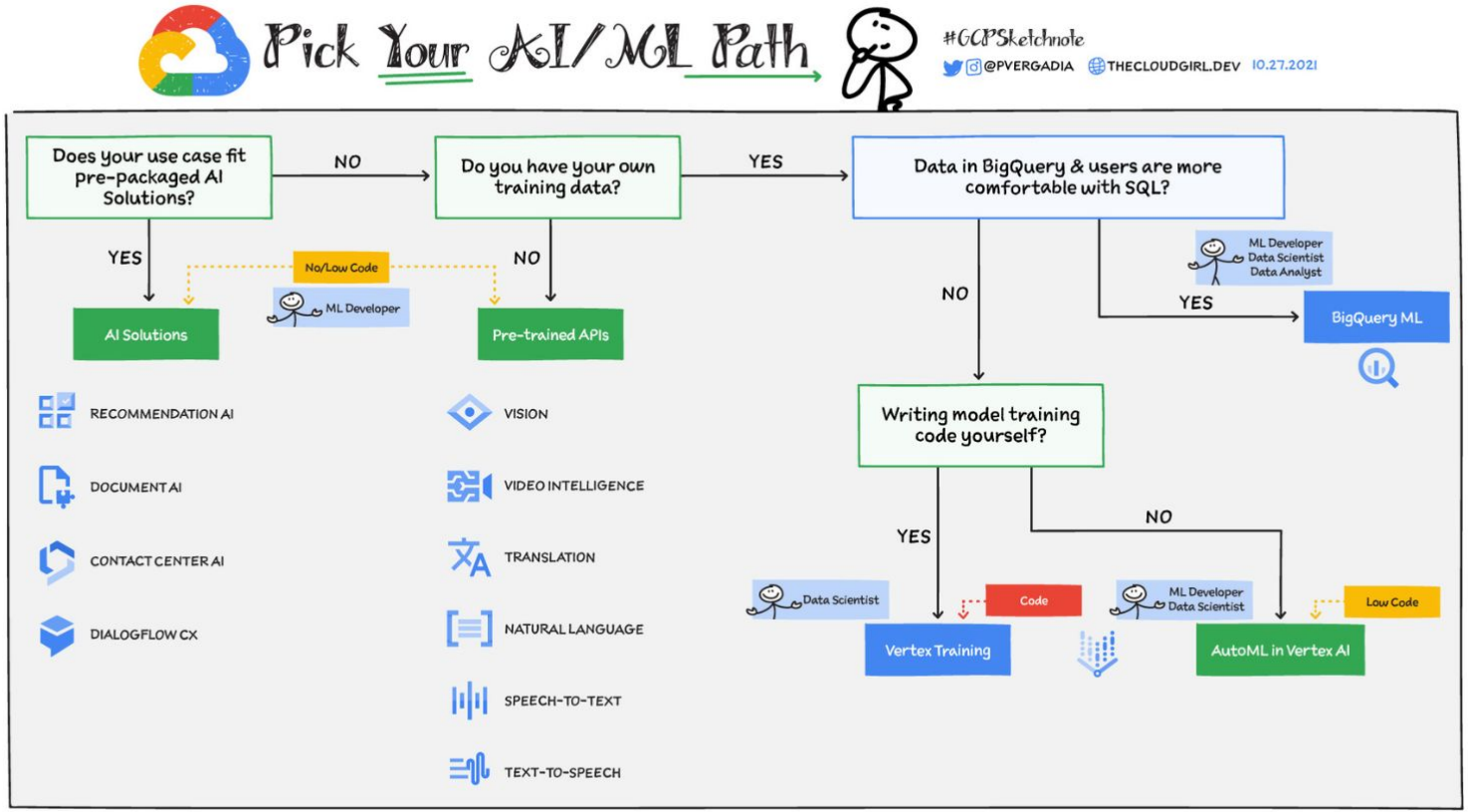
Socially Beneficial

Avoid creating unfair bias

End-to-end- Machine Learning lifecycle



Your AI/ML Path



Vertex AI Pre-Trained Models and Services

Best in class tools allowing customers to leverage Google's leadership in AI to solve common problems

Vision

- Vision
- AutoML Vision
- Video Intelligence
- AutoML Video Intelligence

Language

- Translation
- AutoML Translation
- Natural Language
- AutoML Natural Language

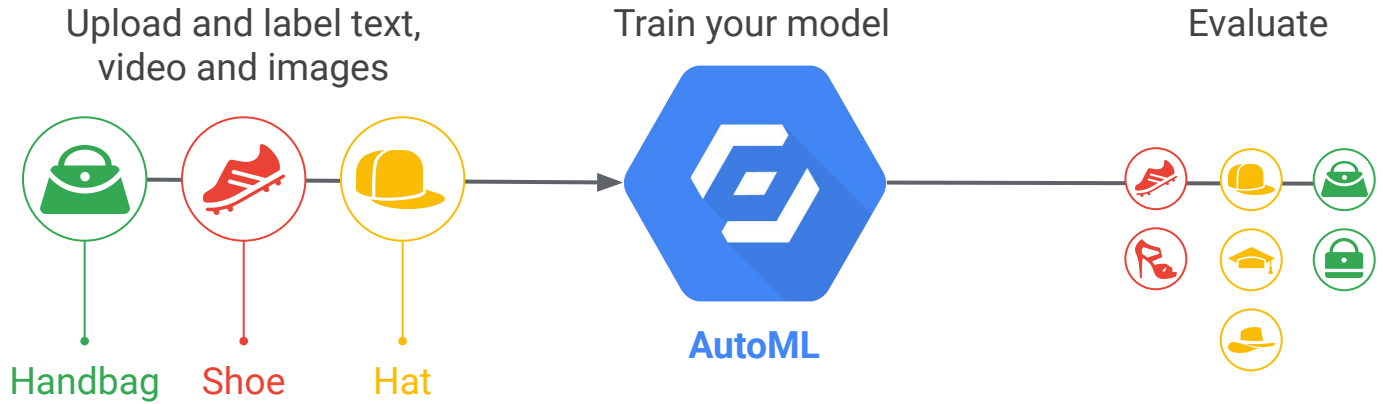
Conversation

- Dialogflow
- Speech-to-Text
- Text-to-Speech
- Media Translation API

AI Industry Solutions

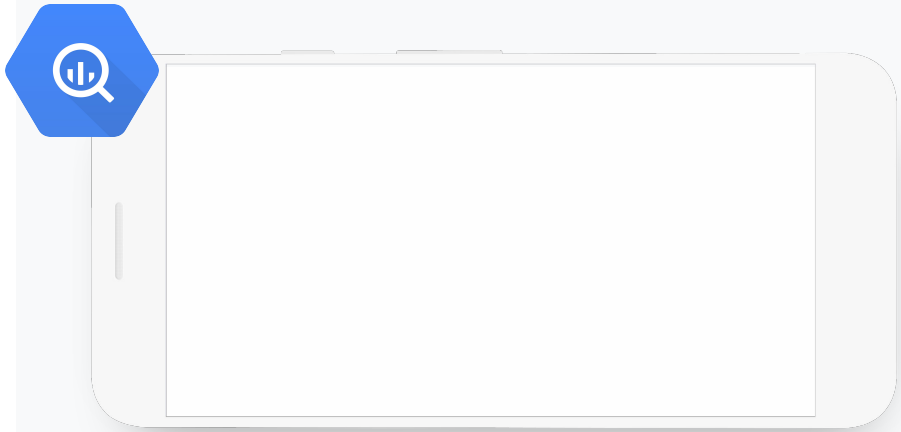
- Contact Center AI
- AutoML Tables
- Document AI
- Fleet routing API
- Healthcare API
- Translation Hub

Vertex AutoML



Google Cloud

BigQuery ML - build custom models with standard SQL



1. **Execute** ML initiatives without moving data from BigQuery
2. **Iterate** on models in SQL in BigQuery to increase development speed
3. **Automate** common ML tasks, and hyperparameter tuning

Google Cloud

Vertex AI Workbench

A hosted Jupyter notebook solution that makes it easy for Data Scientists to spin up JupyterLab; and gives DevOps teams the controls they need.

You can also link Colab with Compute Engine to eliminate restrictions

Google Cloud

Get started quickly: Latest data science and machine learning frameworks are pre-configured.

No learning curve: Uses the industry standard JupyterLab interface.

Scalable & cost-effective: Pick the hardware you need; and scale up and down easily.

Centrally managed: DevOps teams can easily manage and secure these environments.

GCP integration: It's easy to access and use GCP services from within your notebooks.

Easily build, train, and deploy models: Supports the full ML lifecycle through integration with the most popular ML frameworks and tools.

GenAI revolution started at Google and we continue to innovate

2017

Transformers

- Pathbreaking Neural Network Architecture
- Open Sourced by Google in 2017
- Started the revolution in Language Models

2018

BERT

(Bidirectional Encoder Representations from Transformers)

- World's first Language Model
- Open Sourced by Google in 2018
- SOTA on number of language benchmarks

2019

T5

(Text-to-Text Transfer Transformer)

- Large Language Encoder-Decoder Model
- 10-billion parameter model
- Open Sourced by Google in 2019

2020

LaMDA

(Language Model for Dialog Applications)

- Model trained on dialogue data
- Model could talk about virtually anything
- Published by Google in 2020

2021

Diffusion Models

- High Fidelity Image Generation Using Diffusion Models

PaLM

- (Pathways Language Model)
- Single model to generalize across domains
- 540-billion parameter, dense decoder model

2022

CALM

(Confident Adaptive Language Modeling)

- Accelerating the text generation of LMs

2023

Bard

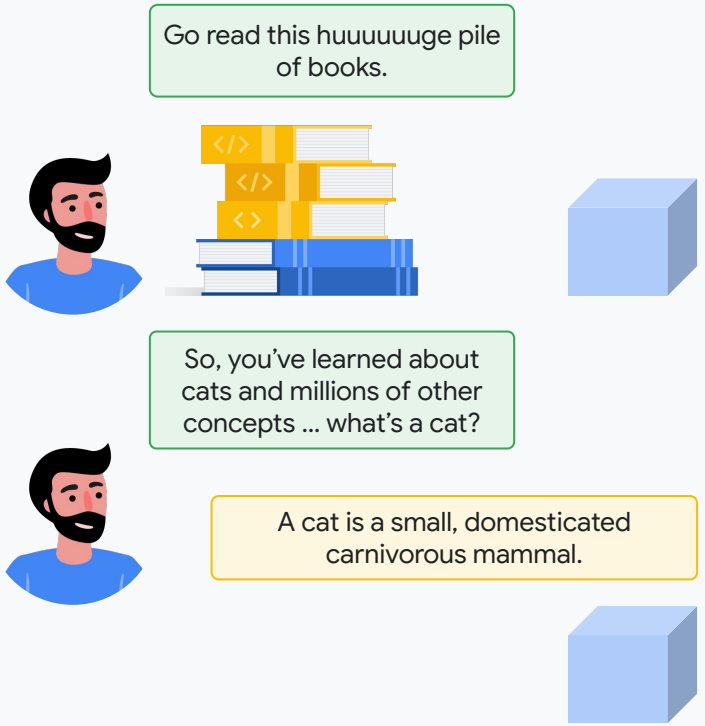
- A conversational AI service powered by LaMDA.

Vertex Gen AI

- Generative AI App Builder: Conversation AI, Enterprise Search, Foundation Models
- Vertex AI: Gen AI Studio, Gen AI APIs, Model Garden

GenAI and Large Language Models

- ML algorithms that can **recognize, predict, and generate** human languages
- Pre-trained on petabyte scale text-based datasets resulting in large models with **10s to 100s of billions of parameters**
- LLMs are normally **pre-trained on a large corpus of text** followed by fine-tuning on a specific task
- LLMs can also be called **Large Models** (includes all types of data modality) and **Generative AI** (a model that produces content)

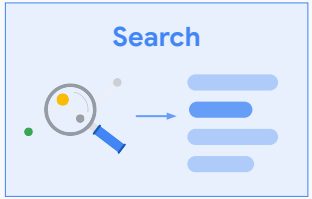


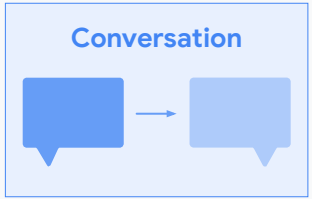
Why are large language models different?

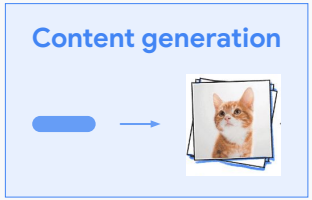
LLMs are characterized by **emergent abilities**, or the ability to perform tasks that were not included in their training examples.

LLMs contextual understanding of human language **changes how we interact** with data and intelligent systems.

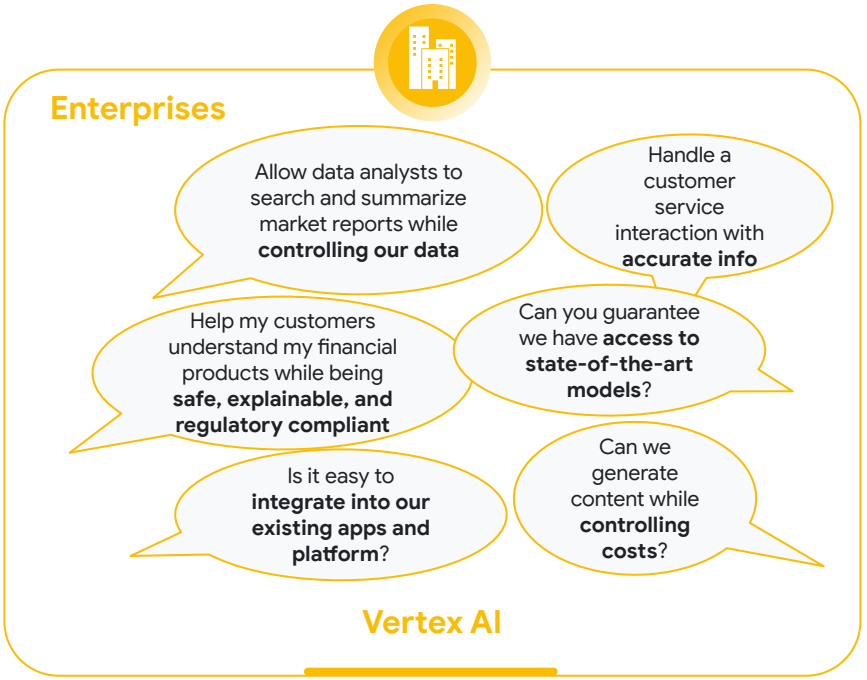
LLMs can find patterns and connections in **massive, disparate data corpora.**







Consumers & enterprises have different needs....



Foundation Models

Across a variety of model sizes to address use cases

Text

Understand and generate natural language

Examples: Summarize an analyst report, write a blog post

Code

Understand, generate, and auto-complete code

Examples: Write SQL code to complete a data analysis task, Finish this python function

Image

Generate, edit, and understand images

Examples: Ad campaigns with AI-generated visuals, image for website

Dialogue

Understand and generate spoken conversations.

Examples: 24x7 customer service chatbot, virtual assistant

Audio and Music Roadmap

Understand and generate audio and music.

Examples: Music for Youtube video, synthetic speaker for news broadcast

Video Roadmap

Generate, edit, and understand videos

Examples: Digital avatar, cutscene in a video game

Choice: Wide range of model sizes to ensure best price performance for a given app
Safety: Immediately leverage the built-in best practices, data governance, and tooling from Google Cloud AI
Innovation: Enjoy the state-of-the-art models commercialized from Google Research and Deepmind
Enterprise-ready: Built on Google infrastructure – high performance and scalability, low cost, global reach

And more models to come in the future ...

Cloud AI Portfolio

To support the needs of Generative AI centric enterprise development

Contact Center AI

Document AI

Discovery AI

Healthcare AI

Generative AI App Builder

Conversation AI

Foundation Models

Enterprise Search

Vertex AI

End-to-End ML Platform

Generative AI Studio | Generative AI APIs | Model Garden

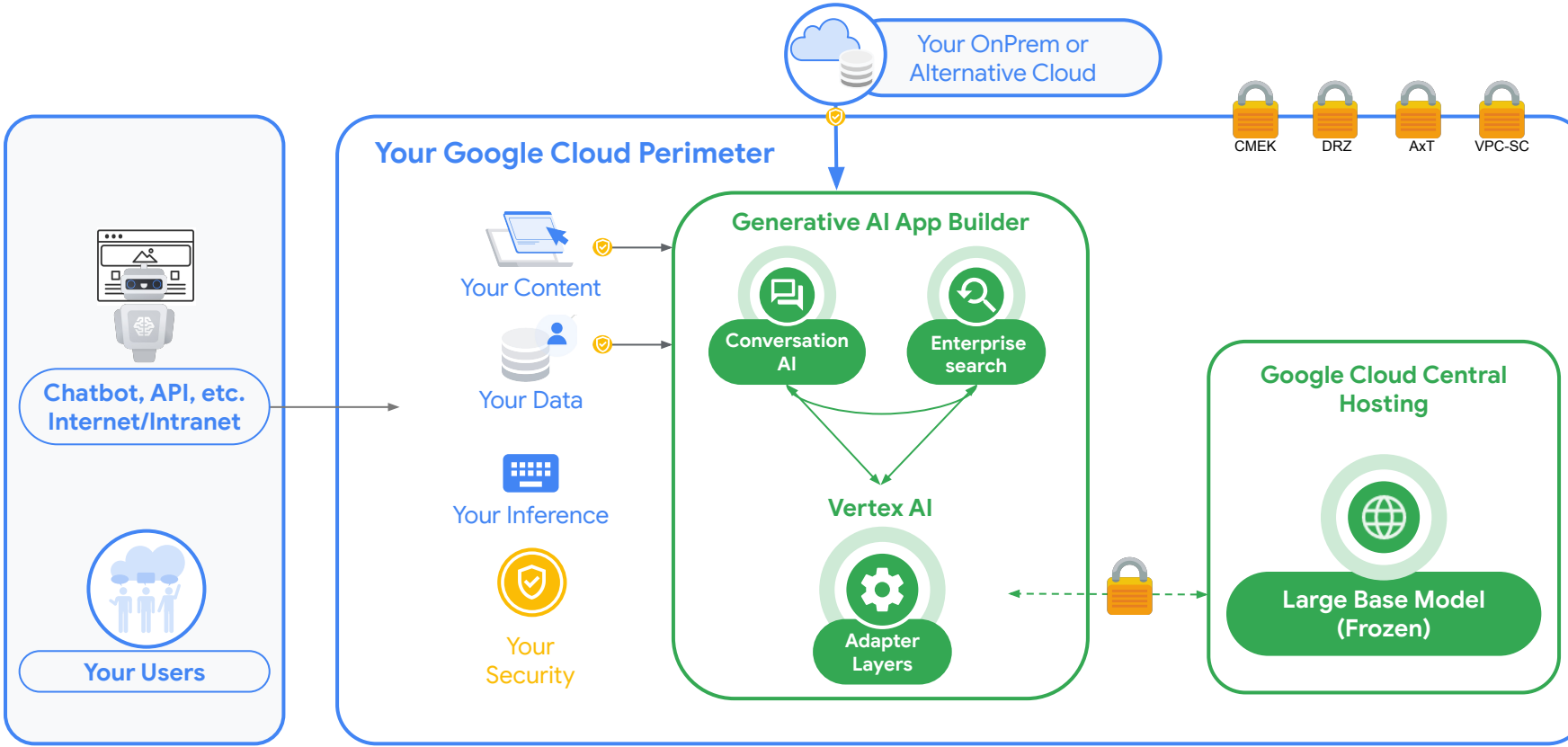
Google Cloud Infrastructure

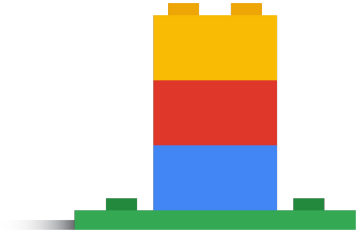
Business users

Developers

AI Practitioners

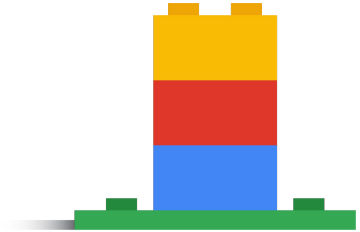
Your Data, Your Terms





Build, tune, and deploy foundation models with Vertex AI

Google Cloud makes it easy to access, customize, and deploy large models - opening the door for a new-era of applications that can create, recommend, troubleshoot, synthesize, analyze and engage in a natural way.



Build, tune, and deploy foundation models with Vertex AI

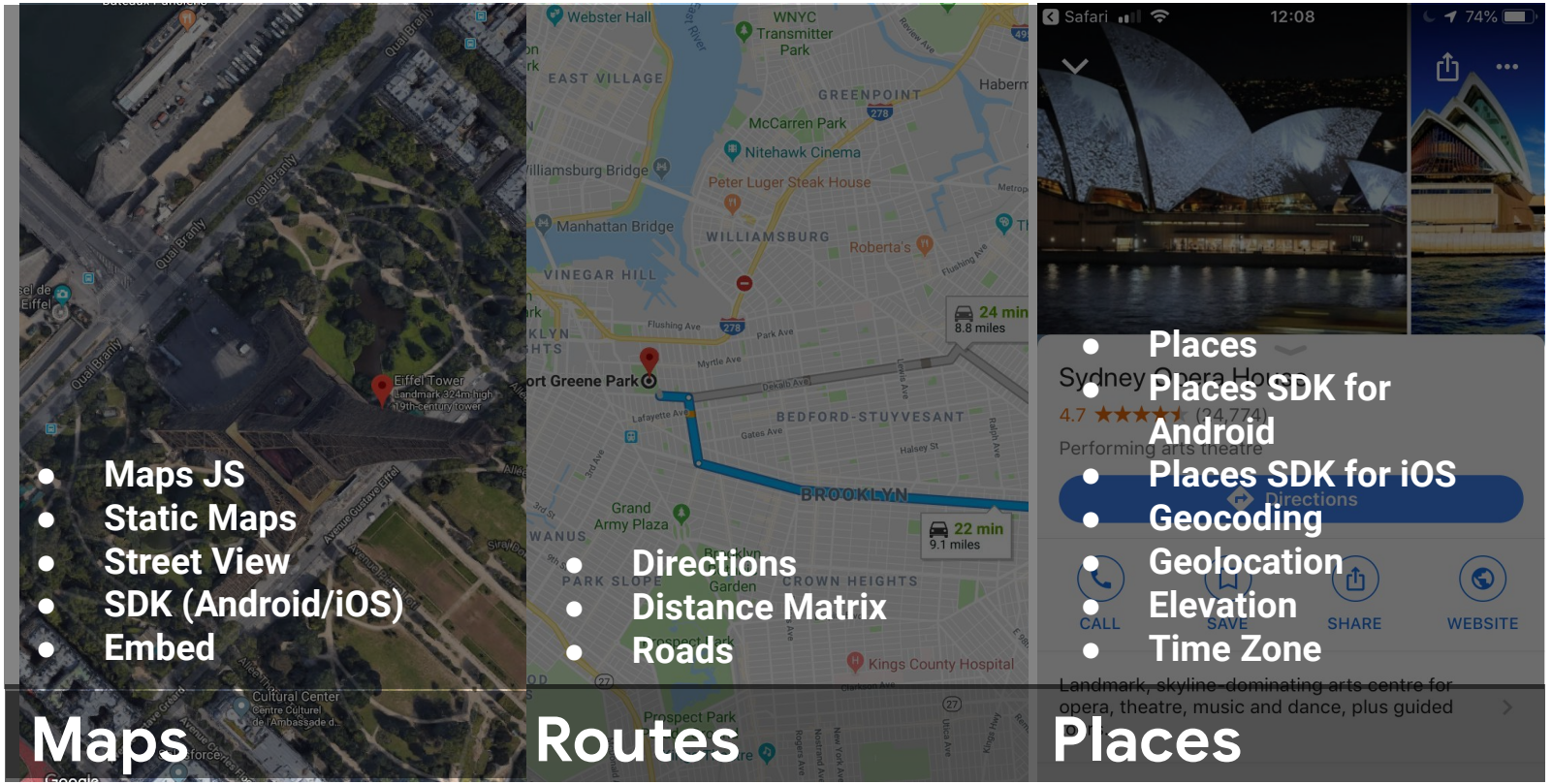
Google Cloud makes it easy to access, customize, and deploy large models - opening the door for a new-era of applications that can create, recommend, troubleshoot, synthesize, analyze and engage in a natural way.

Geolocation

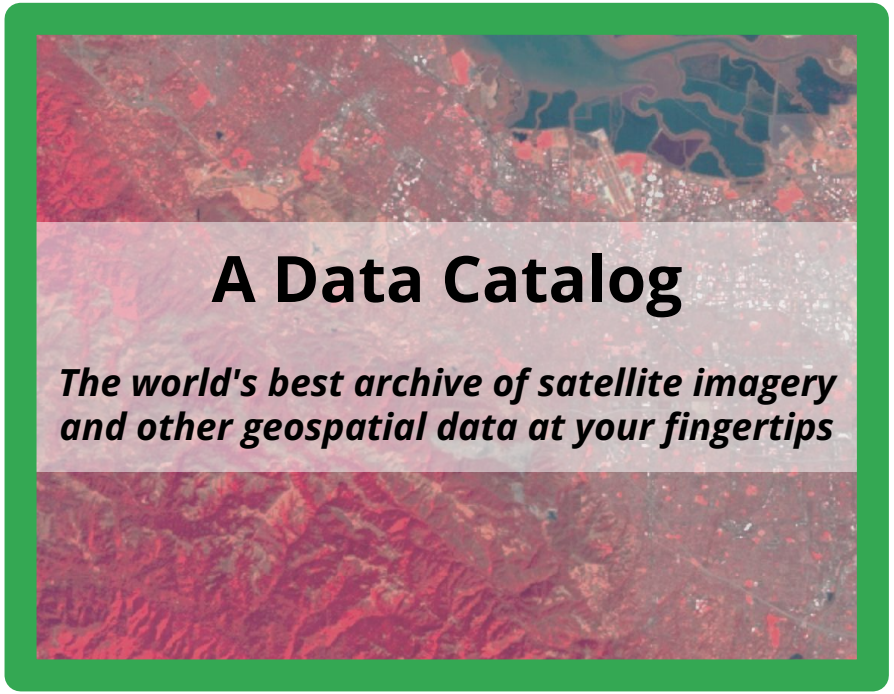
06

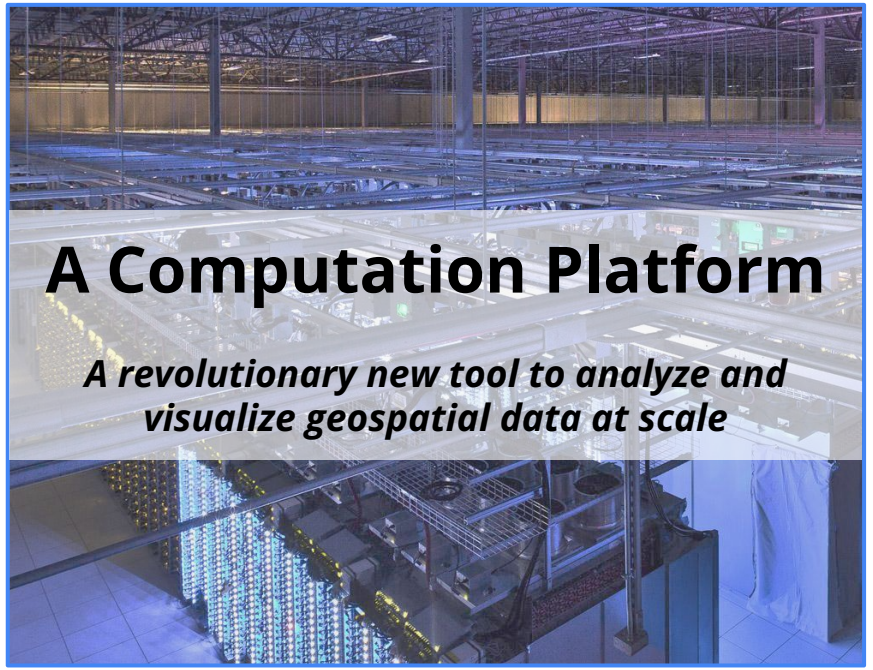
Google Cloud

Google Maps API Platform



What is Earth Engine?





The Earth Engine Data Catalog

Landsat & Sentinel
10-30m, weekly

MODIS
250m daily

Vector Data
WDPA, TIGER, WHC

Terrain & Land Cover

Weather & Climate
NOAA NCEP, OMI, ...

... and upload your own vectors and rasters

800+ public datasets

70+ petabytes of data

100+ datasets added yearly

1+ PB of new data every month

Computation

API

Apps

Data

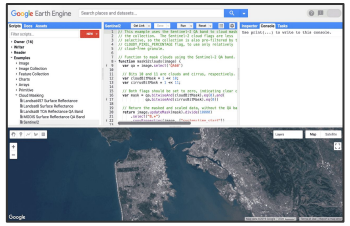
Earth Engine Backend

REST API

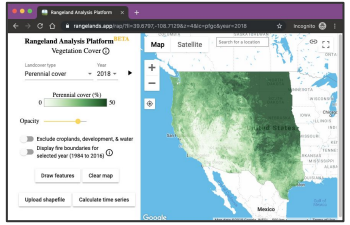
Javascript Client Library

Python Client Library

Earth Engine Web Explorer Code Editor

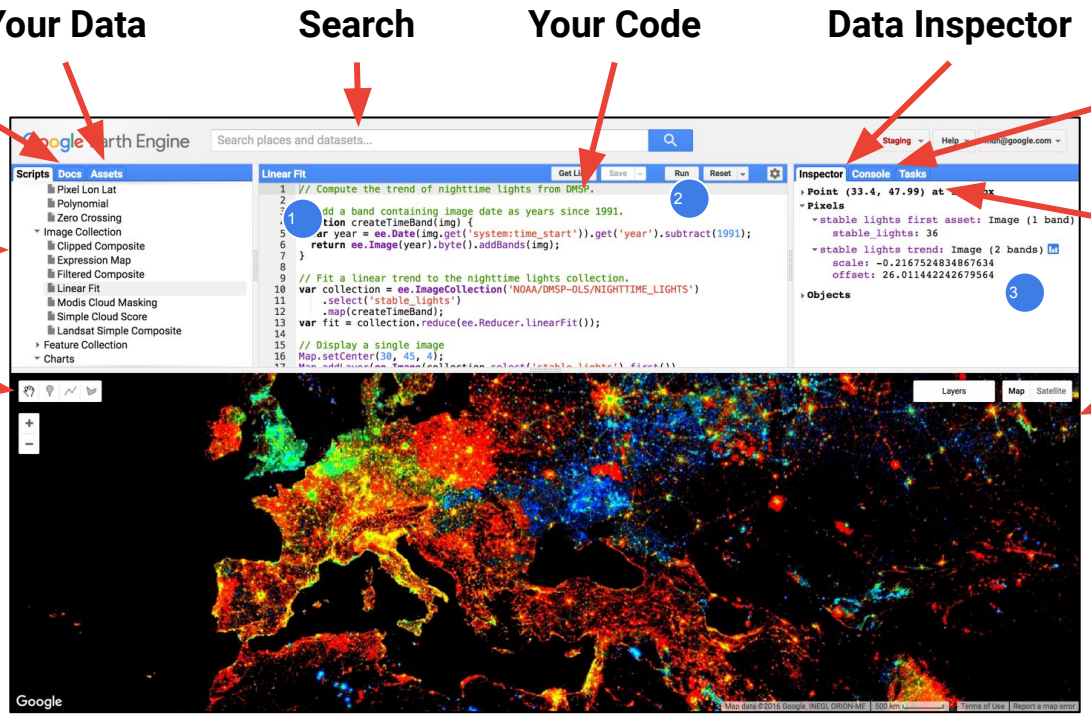


Web Applications



Jupyter-based Development Environment

The Earth Engine Code Editor



In interactive mode, user can:

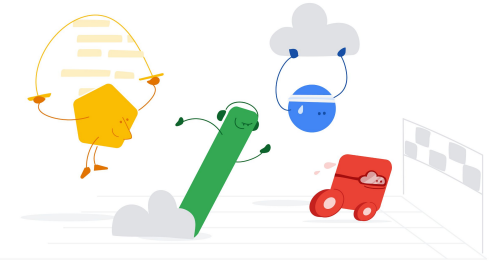
- Edit their script 1
- Run the script 2
- Evaluate the results 3
- Repeat, refine, and perfect
- Share and publish

code.earthengine.google.com

Google Cloud Labs
Cloud Hero Game

Cloud Hero

Cloud Hero is a program designed for developer hands-on engagement. Cloud Hero events bring developers together to **learn** and engage in friendly **competition** against one another in **gamified labs** using their Google Cloud skills. At its core, a Cloud Hero game is made up of hands-on labs (Google Cloud Skills Boost) layered with activity-tracking and scoring.



The objective of Cloud Hero is to deliver hands-on training and upskill in Google Cloud solutions, while having fun along the learning journey!

Signín for account at https://www.cloudskillsboost.google/

- **Sign-in** with Gmail account or if you already have an account or **Join** to create a new one.

https://www.cloudskillsboost.google/

https://www.cloudskillsboost.google/users/sign_in

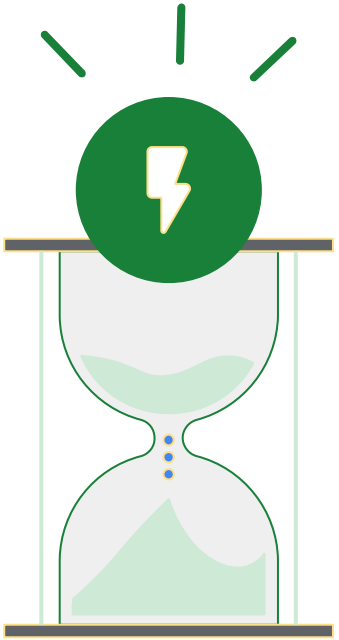
Timing and Scoring

Points are earned by completing the steps in the lab.... and bonus points are earned for speed!

You can redo each of the labs up to 5 times. **Your best score will count!**

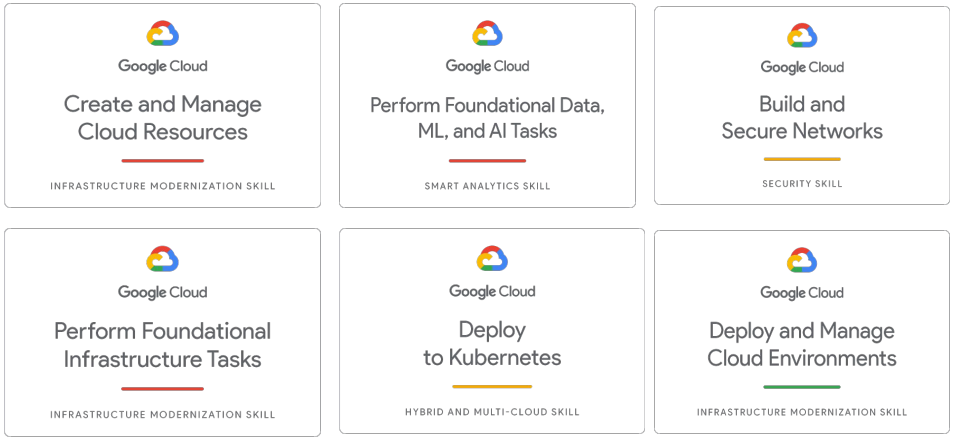
Be sure to complete each lab by clicking **END Lab** to get the maximum points.

Labs usually have checkpoints. To get maximum points, you need to have all the checkpoints green.



Google Cloud Skill Badges

By playing Cloud Hero games, you are on your way to demonstrating your growing Google Cloud-recognized skill set through exclusive digital Google Cloud skill badges. These are real Google Cloud credentials, which can help you grow professionally and expand your Google Cloud console knowledge.



We will have 3 Games this afternoon

Infra Skills

- Navigate through the GCP console and shell
- Understand the IAM permissions
- Create VM and connect to it
- Manage traffic through a load balancer

Data Skills

- Dataprep for easy data cleaning
- Dataproc to submit Spark jobs
- Dataflow for managing a steam pipeline
- Natural Language API

Vertex AI Skills:

- BigQuery ML
- AutoML Vision to train and label images
- Vertex AI workbench for managed Jupyter Notebooks with Tensorflow Enterprise

Cloud Hero Infra II Skills Game

Continue your progress on the game.
Here is the link:
https://www.cloudskillsboost.google/games/4082
The access code is
ch4-na-universitat-219

Access to the lab:

You will be prompted to log into Cloud Skills Boost. If you did not already have a Cloud Skills Boost account, you'll need to create one. Creating an account is always free.

1. Click "**Join this game**" and enter your access code: **ch4-na-universitat-219**
2. Start taking labs to score points and watch your name rise on the leaderboard!

Cloud Hero Infra II Skills Game

1. **A Tour of Google Cloud Hands-on Labs**
   - Navigate through the console and projects, understand IAM roles, enable APIs&services
2. **Compute Engine Qwik Start-Windows**
   - Create a Windows VM using GCP console, setup password and connect through RDP
3. **Getting Started with Cloud Shell and gcloud**
   - Activate cloud shell in GCP console, set project, regions&zones.
   - Create VM using cloud shell and list VMs. List and analyze firewall rules
   - Connect to VM and install nginx server
   - Update firewall list to allow connection to nginx server through port 80
   - View and analyze logs
4. **Setup Network and HTTP load balancers**
   - Set project and zone configuration
   - Create three Compute Engine VM instances and install Apache on them, then add a firewall rule that allows HTTP traffic to reach the instances
   - Configure load balancing service
   - Send traffic to your instance
   - Create an HTTP load balancer and a managed instance group
     - Test sending traffic to your instances through the load balancer

Cloud Hero Data Skills Game

Continue your progress on the game.
Here is the link:
https://www.cloudskillsboost.google/games/4083
The access code is
ch4-na-universitat-220

Access to the lab:

You will be prompted to log into Cloud Skills Boost. If you did not already have a Cloud Skills Boost account, you'll need to create one. Creating an account is always free.

1. Click "**Join this game**" and enter your access code: **ch4-na-universitat-220**
2. Start taking labs to score points and watch your name rise on the leaderboard!

Cloud Hero Data Skills Game (I)

Dataprep Qwik Start

- Create a Cloud Storage bucket (uncheck enforce public access prevention) to allow external IP access
- Initialize Cloud Dataprep for your user lab
- Create a new flow
- Import datasets and modify them
- Test and run some transformations

Dataflow Qwik Start

- Create a dataset and table in BigQuery and Create a storage bucket (select cloud shell or GCP console)
- Create a dataflow pipeline and Run job (Pub/Sub topic to BigQuery). Wait until job finishes
- View the data written to BigQuery and execute queries

Dataproc Qwik Start

- Create dataproc cluster
- Submit Job
- Review Job output
- Update the cluster and resubmit job

Cloud Hero Data Skills Game (II)

Cloud Natural Language Qwik Start

- Create an API key and set the GOOGLE_APPLICATION_CREDENTIALS environment variable
- Make an entity analysis request (from a Compute Engine already created. SSH into the VM to launch the request)
-

Cloud Hero Vertex AI Game

Continue your progress on the game.
Here is the link:
https://www.cloudskillsboost.google/games/4084

The access code is
ch4-na-universitat-221

Access to the lab:

You will be prompted to log into Cloud Skills Boost. If you did not already have a Cloud Skills Boost account, you'll need to create one. Creating an account is always free.

1. Click "**Join this game**" and enter your access code: **ch4-na-universitat-221**
2. Start taking labs to score points and watch your name rise on the leaderboard!

Cloud Hero Vertex Game (I)

Getting Started with BigQuery Machine Learning

- Create a dataset in BigQuery
- Create a model to predict whether a visitor will make a transaction (model creation will take approximately 2 minutes or less)
- Evaluate the model using different data (changing time frame)
  - Predict purchases by country
  - Predict purchases by user

Vertex AI Qwikstart

- Enable Google Cloud services
- Create Vertex AI custom service account for Vertex Tensorboard integration
  - Create service account and grant access to Cloud Storage, BigQuery and Vertex AI
- Create a Vertex AI Notebook with Tensorflow Enterprise 2 (Wait while notebook is provisioned)
  - Launch JupyterLab notebook
  - Clone the lab repository
  - Install lab dependencies
- Open lab instance (**lab_exercise.ipynb**)
  - Run each cell and analyze results

Cloud Hero Vertex Game (II)

Identify Damaged cars with AutoML Vision

- Upload training images to Cloud Storage
  - Create a Cloud Storage bucket
  - Upload car images to Cloud Storage bucket
  - Review if images have been correctly uploaded through GCP console
- Create a dataset
  - Upload CSV with taggs for images
  - Enable Vertex AI APIs and create dataset
  - Connect dataset to training images and upload images (it can take some minutes)
- Inspect images:
  - Analyze images and tags
- Train the model
  - Train model and wait until model is trained (it can take some time but you can continue with next session as there is a model already deployed)
- Request a prediction from a hosted model in Cloud Run
  - Get URL of the model and run a prediction creating a new request
  - Call Cloud Run endpoint with payload.json created
  - Check prediction result

Cloud Hero Vertex Game (III)

Deploy a BigQuery ML Customer Churn Classifier to Vertex AI for Online Predictions

- Launch JupyterLab notebook
- Clone the lab repository
- Install lab dependencies
- Run each cell and analyze results
-

Cloud Hero Tips

- Use an incógnito window to open the Google Cloud Console
- Check your progress to earn points
- Don´t forget to Submit “Finish your lab” to get credits
- Read the instructions carefully and use the same names, zones, regions... to get correct results
- Check the remaining time
- Don’t follow just the instructions. Investigate and try with GCP console and GCP shell

ASK ME!!!

Access to Games summary

Infra II Skills Game

- https://www.cloudskillsboost.google/games/4082
- Access Code: **ch4-na-universitat-219**

Data Skills Game

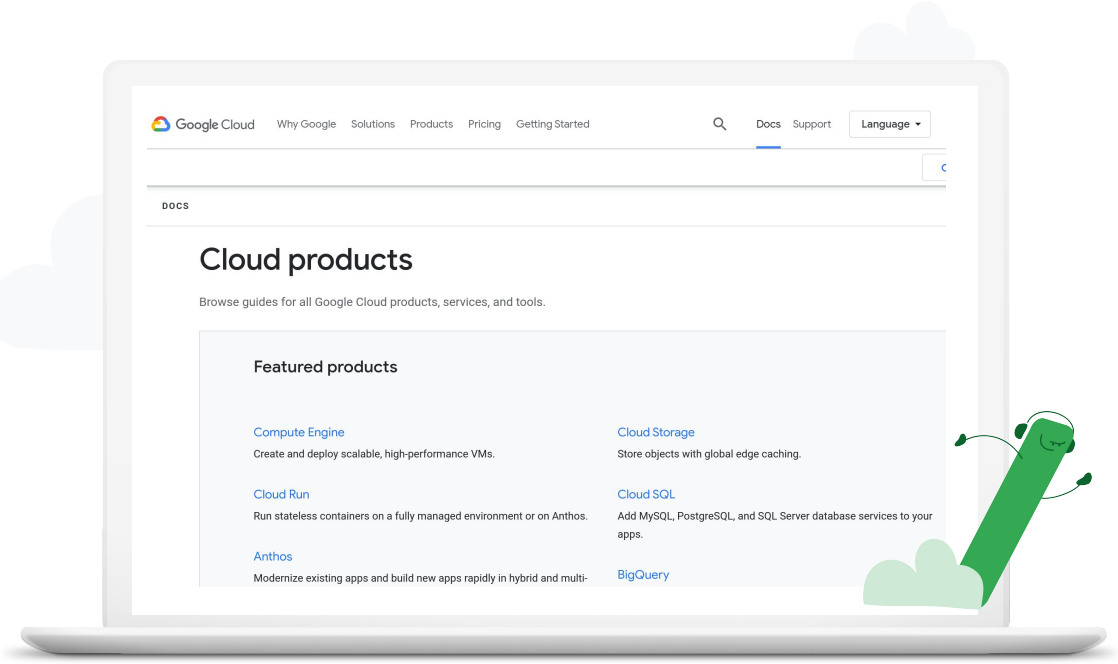
- https://www.cloudskillsboost.google/games/4083
- Access Code: **ch4-na-universitat-220**

Vertex AI Skills Game

- https://www.cloudskillsboost.google/games/4084
- Access Code: **ch4-na-universitat-221**

Additional Resources

Google it! Review support documentation at cloud.google.com/docs



Thank you!

Google Cloud