Free Trial Restrictions

- No more than 8 vCPUs (total simultaneous)
- No GPUs (video card chips)
- No TPUs (custom chips for TensorFlow)
- No Quota increases
- No cryptomining allowed
- No SLAs
- No premium OS licenses (e.g. Windows)
- No Cloud Launcher products with extra usage fees

or Cloud Launcher products with extra usage fees.

Free-Tier GCP Accounts

"Always Free"

A CLOUD GURU

- "Always Free usage does not count against your free trial credits."
- Last beyond end of free trial
- Full details at https://cloud.google.com/free/docs/always-free-usage-limits

but I'm gonna go through just a few highlights here.

Free-Tier GCP Accounts

“Always Free” Storage Highlights

A CLOUD GURU

- *Storage averaged over the month*
- 5 GB of Regional Cloud Storage, including some operations
- 1 GB of Cloud Datastore storage, including some operations
- 10 GB of BigQuery storage, with 1 TB/month of query processing
- 30 GB HDD storage on GCE and AE
- 5 GB snapshot storage on GCE and AE
- 5 GB of StackDriver logs with 7 day retention

and five gigs of StackDriver logs

“Always Free” Compute Highlights

A CLOUD GURU

- 24h/day of f1-micro runtime, *in most US regions, only*
- 28h/day of App Engine runtime, *in North America*
- 2M/month of Cloud Functions invocations (with runtime/size limits)

You can also run two million Cloud Functions every month

“Always Free” Networking Highlights

A CLOUD GURU

- *Egress to China and Australia not free!*
- 1 GB/month of App Engine data egress
- 1 GB/month of Compute Engine data egress
- 5 GB/month of egress by Cloud Function invocations
- 5 GB/month of egress from Cloud Storage based in North America
- 10 GB/month of Cloud PubSub messages

that goes through it each month for free.

“Always Free” Other Highlights

A CLOUD GURU

- 120 build-minutes/day of Google Cloud Container Builder
- 60 minutes/month of Google Cloud Speech API recognition from audio/video
- 1,000 units/month of Cloud Vision API calls
- 5,000 units/month Google Cloud Natural Language API
- Google Cloud Shell with 5 GB of persistent disk storage quota
- 1 GB of Google Cloud Source Repositories private hosting

If not, please feel free to move on to the next lecture,

Recap

- Export must be set up per billing account
- Resources should be placed into appropriate projects
- Resources should be tagged with labels
- Billing export is not real-time
  - Delay is hours

This means that there are definitely some interesting things

Billing IAM

A CLOUD GURU

> **Role:** Billing Account User
> **Purpose:** Link projects to billing accounts.
> **Level:** Organization or billing account.
> **Use Case:** This role has very restricted permissions, so you can grant it broadly, typically in combination with Project Creator. These two roles allow a user to create new projects linked to the billing account on which the role is granted.
>
> —GCP Docs

So we'll see how this plays out in the lab.

Highlights

A CLOUD GURU

- Web browser access
  - No need for local terminal
    - Chromebook 👍
    - No PuTTY! 🎉
  - Automatic SSH key management
- 5 GB of persistent storage
- Easy-access to preinstalled tools
  - gcloud, bq, kubectl, docker, npm/node, pip/python, ruby, vim, emacs, bash, etc.
- Pre-authorized and always up-to-date
- Web preview of web app running on local port

Three Core Components

Network	Moving
Compute	Processing
Storage	Remembering

and storage services.

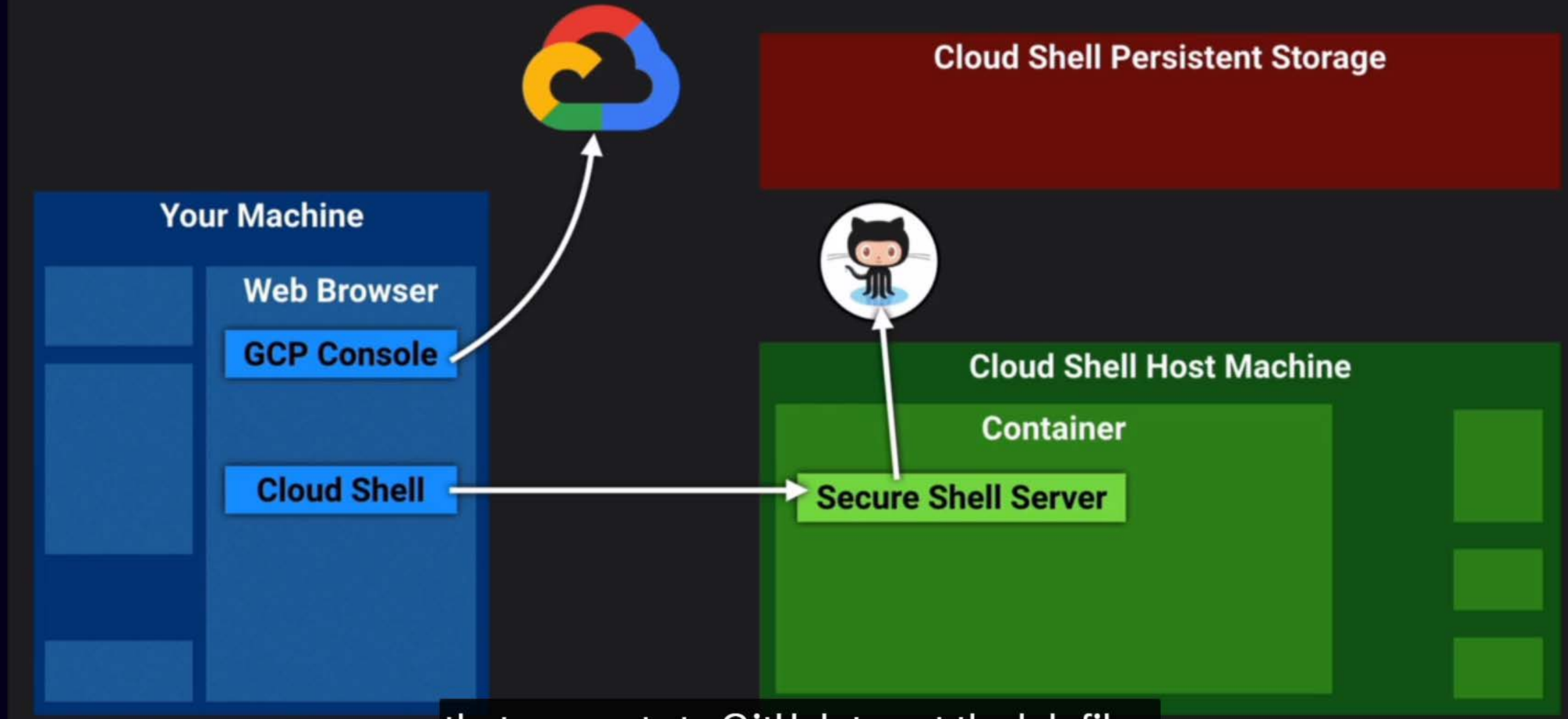
Mental Models

- A simplified representation of reality, which is...
- Used by your mind to anticipate events or draw conclusions
- Systems combine
  - Build larger systems out of smaller ones (abstractions)
  - Zooming in and out

in the Cert Prep Guide Two,

Data Flows

Cloud Shell Lab Data Flow

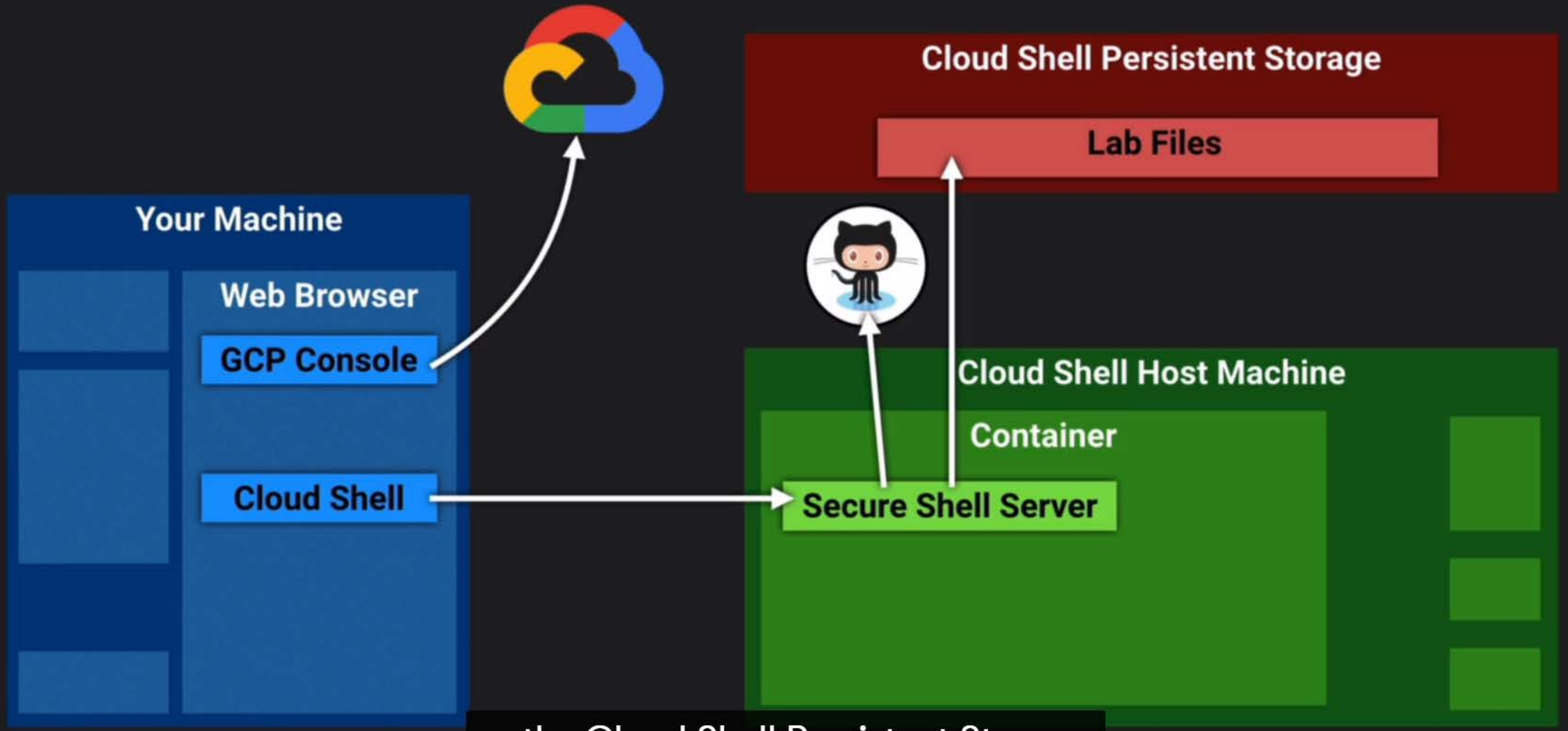


that connects to GitHub to get the lab files,

Data Flows

Cloud Shell Lab Data Flow

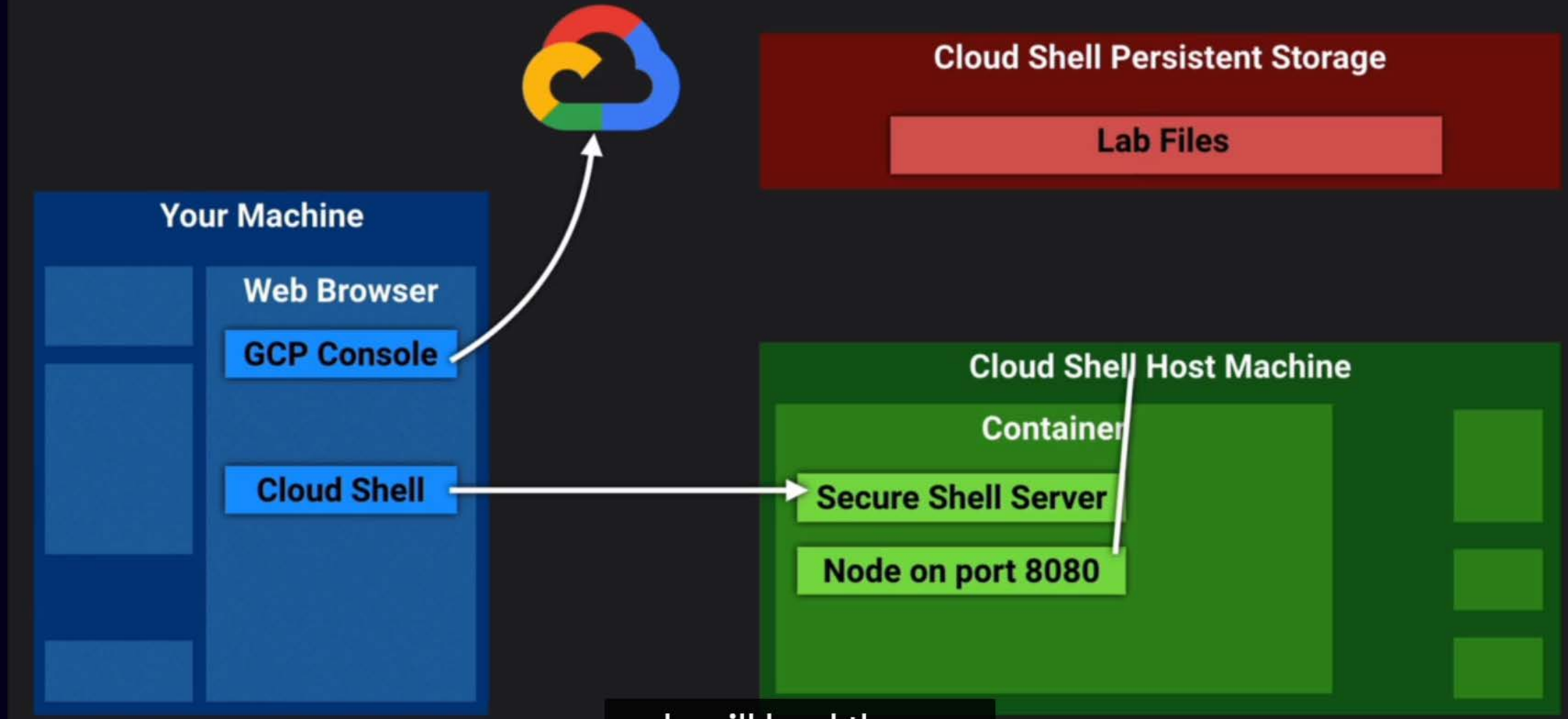
A CLOUD GURU



on the Cloud Shell Persistent Storage.

Data Flows

Cloud Shell Lab Data Flow

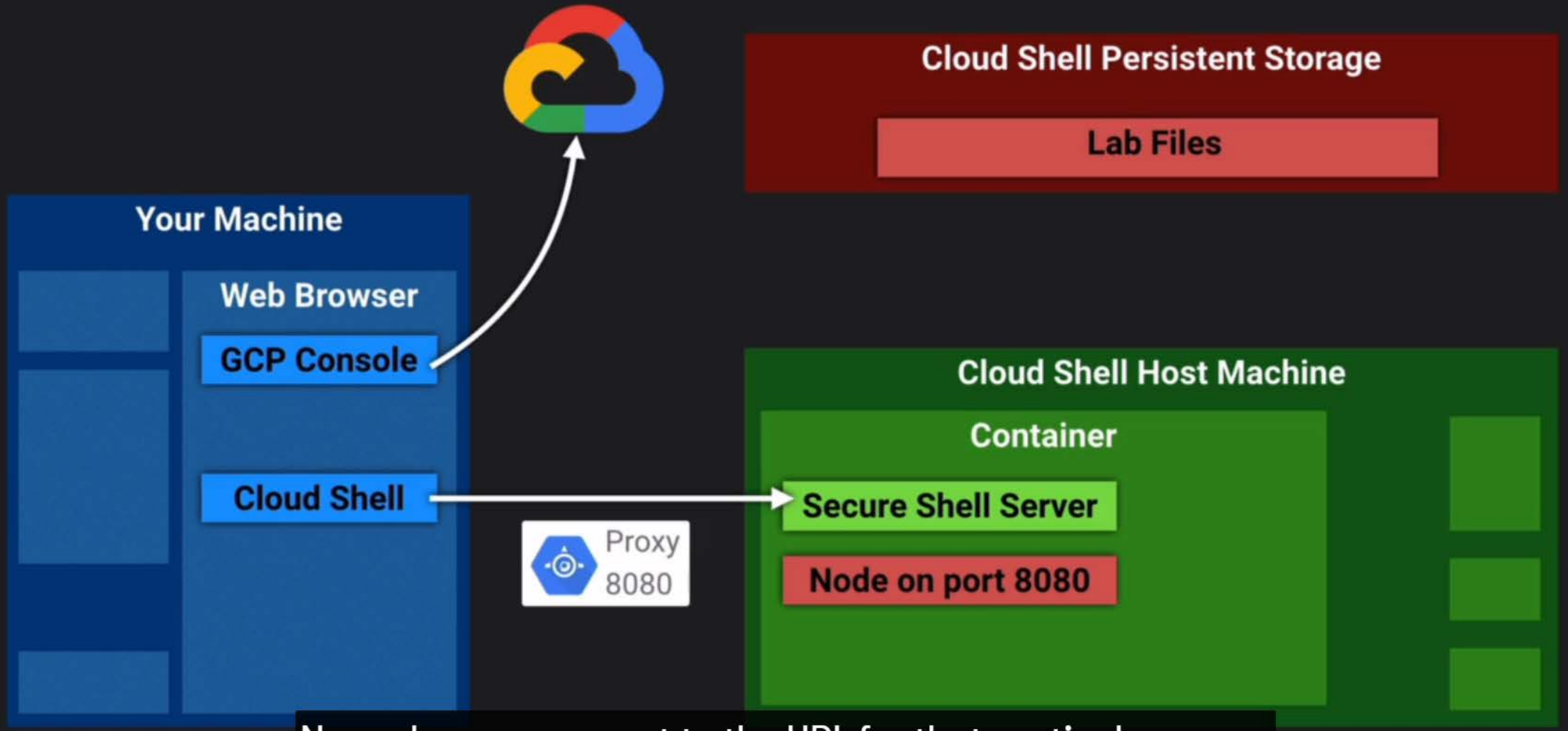


node will load them up,

Data Flows

Cloud Shell Lab Data Flow

A CLOUD GURU

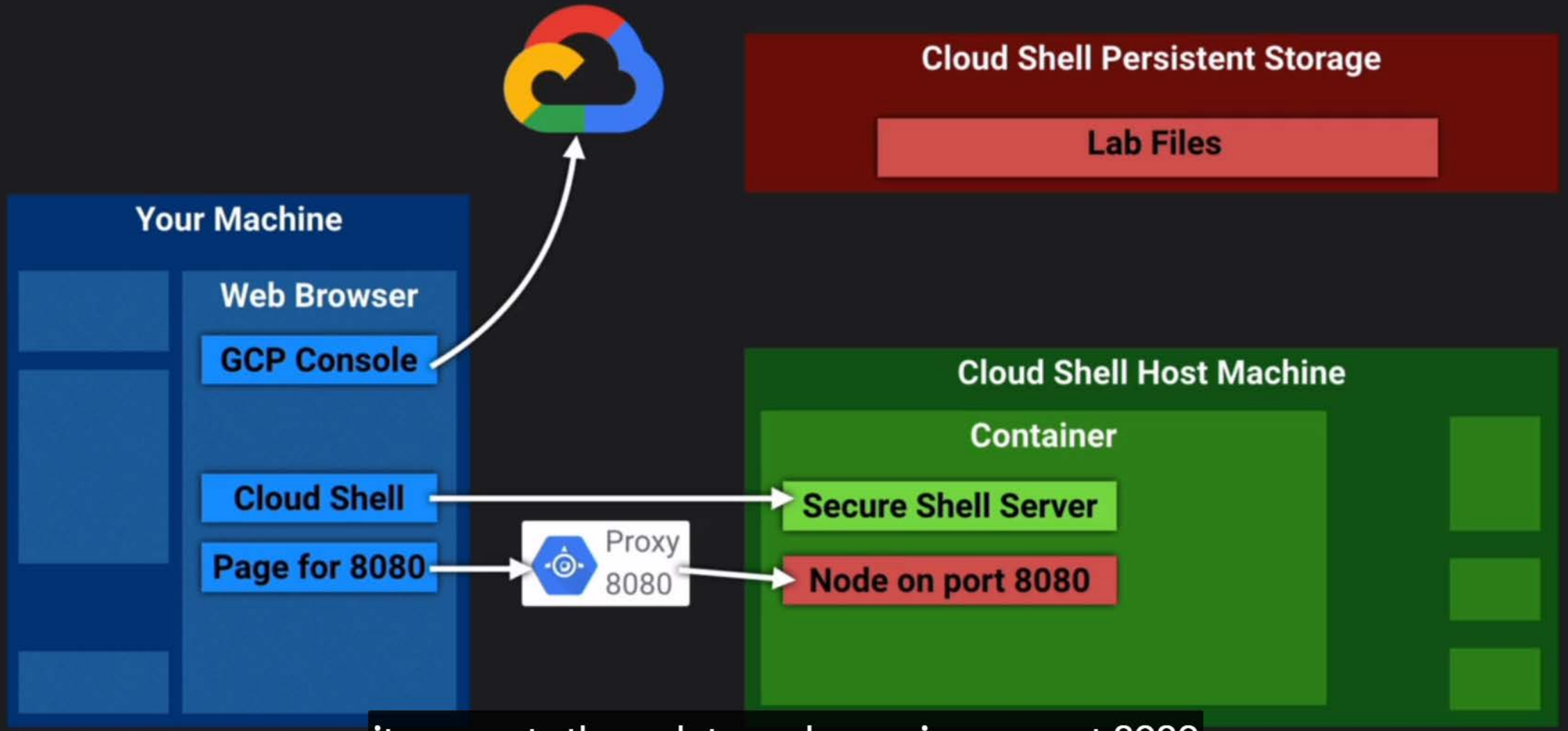


Now when we connect to the URL for that particular proxy,

Data Flows

Cloud Shell Lab Data Flow

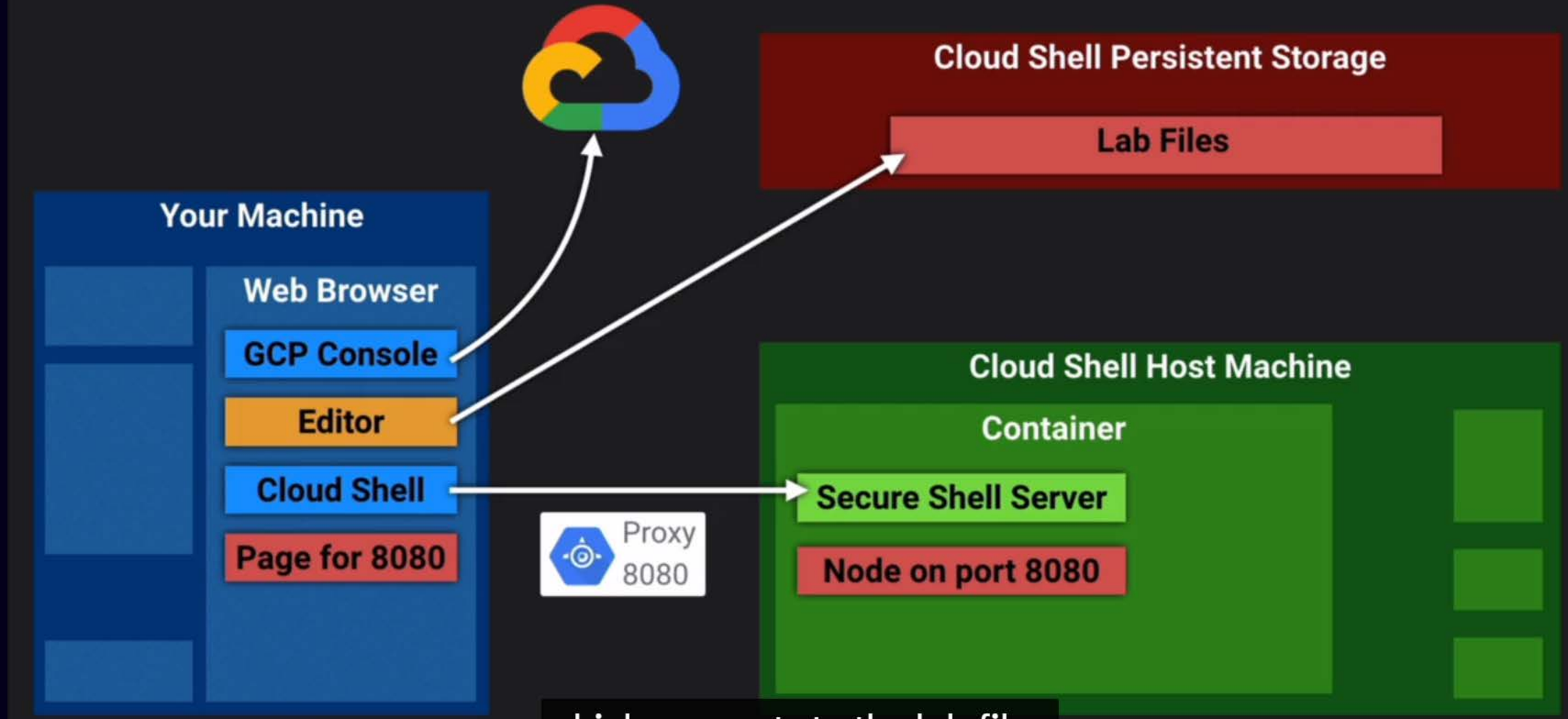
A CLOUD GURU



it connects through to node running on port 8080

Data Flows

Cloud Shell Lab Data Flow

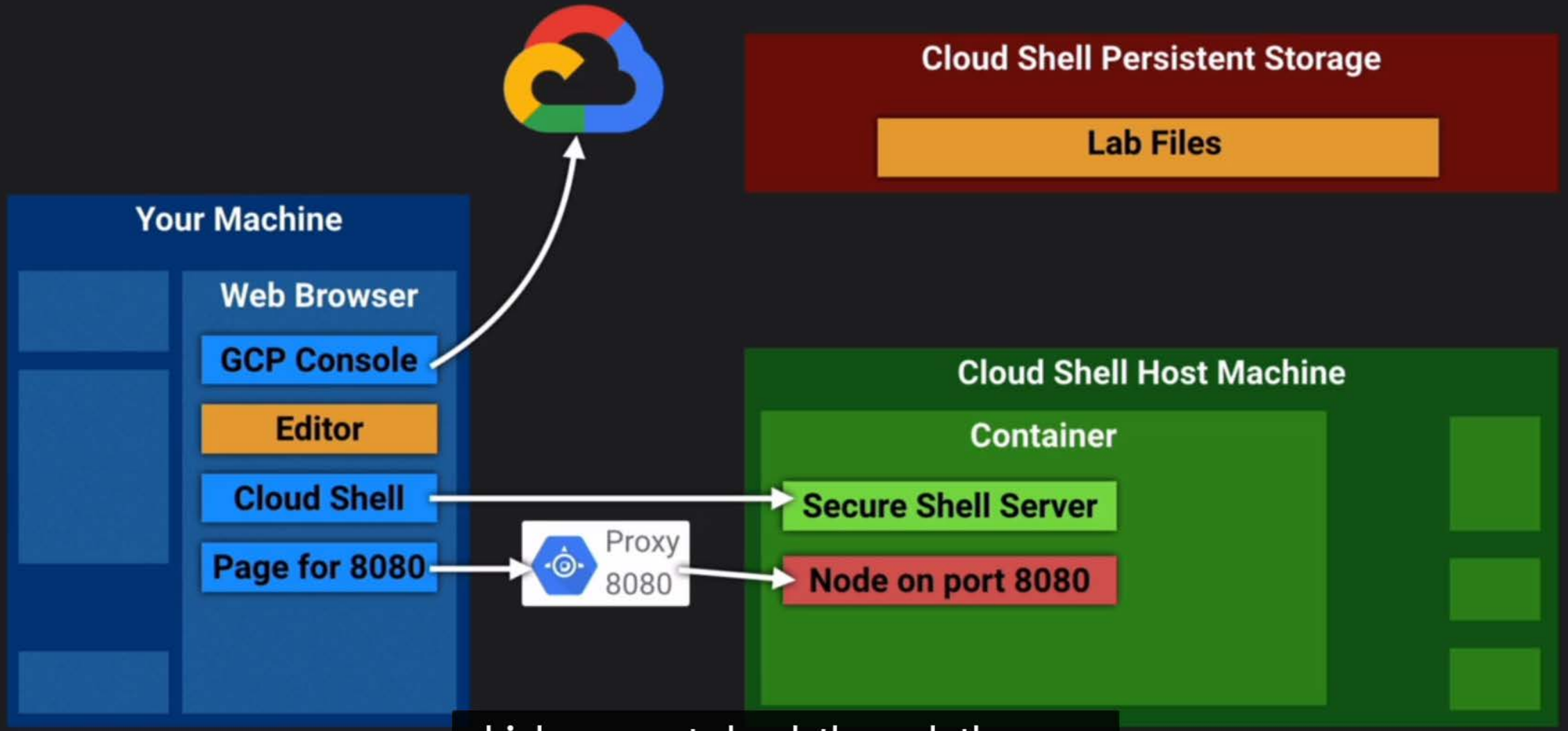


which connects to the lab files,

Data Flows

Cloud Shell Lab Data Flow

A CLOUD GURU

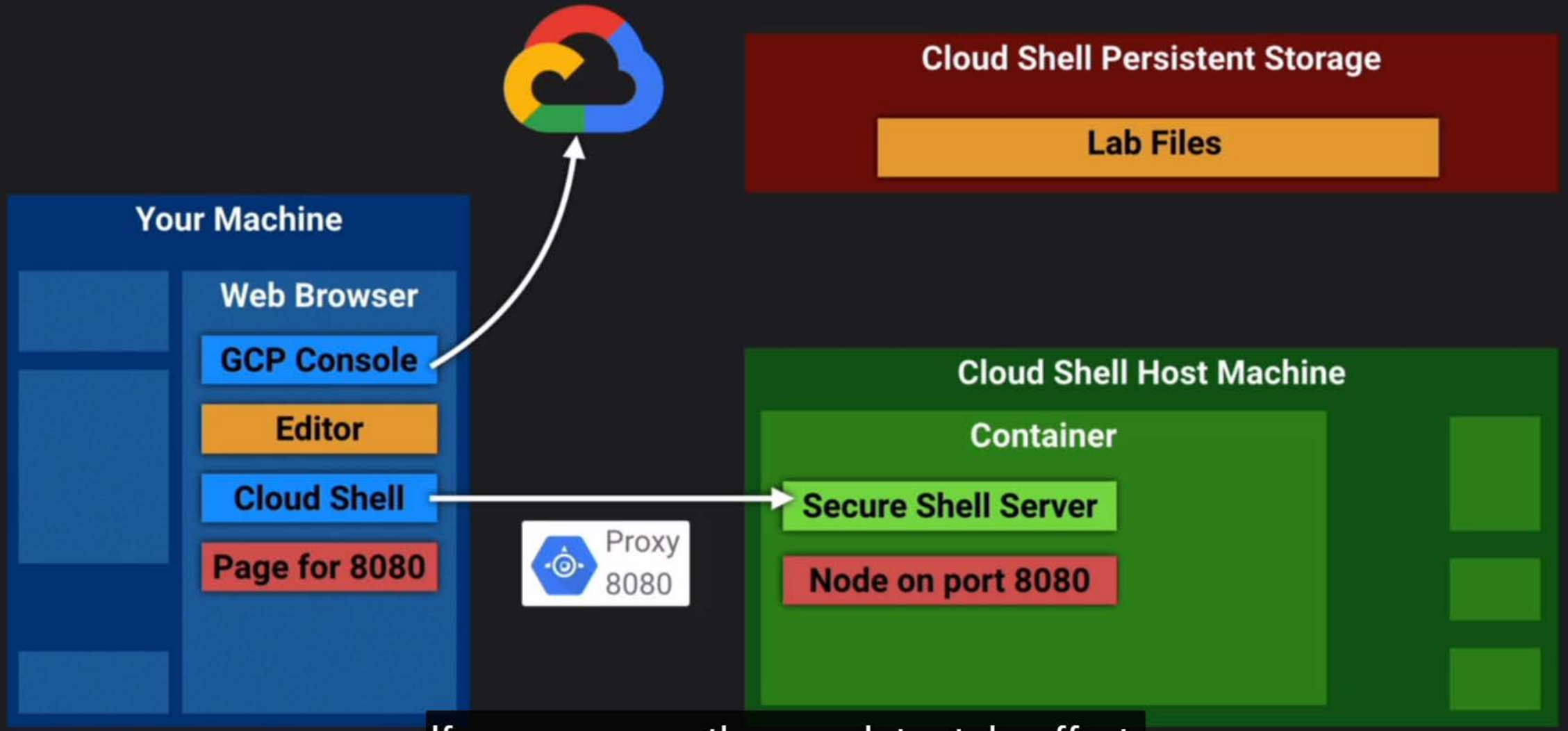


which connects back through the proxy

Data Flows

Cloud Shell Lab Data Flow

A CLOUD GURU

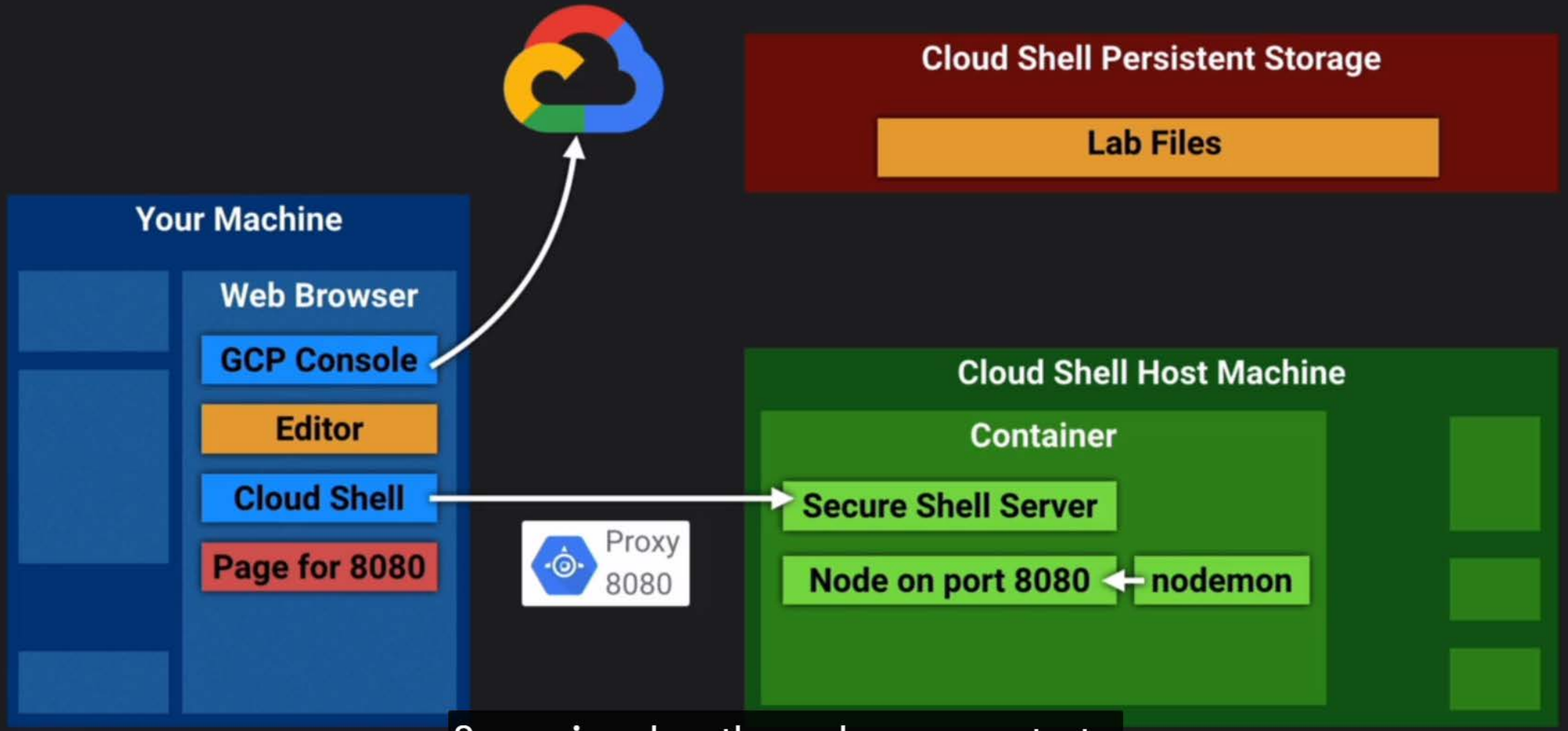


If we wanna see those updates take effect,

Data Flows

Cloud Shell Lab Data Flow

A CLOUD GURU

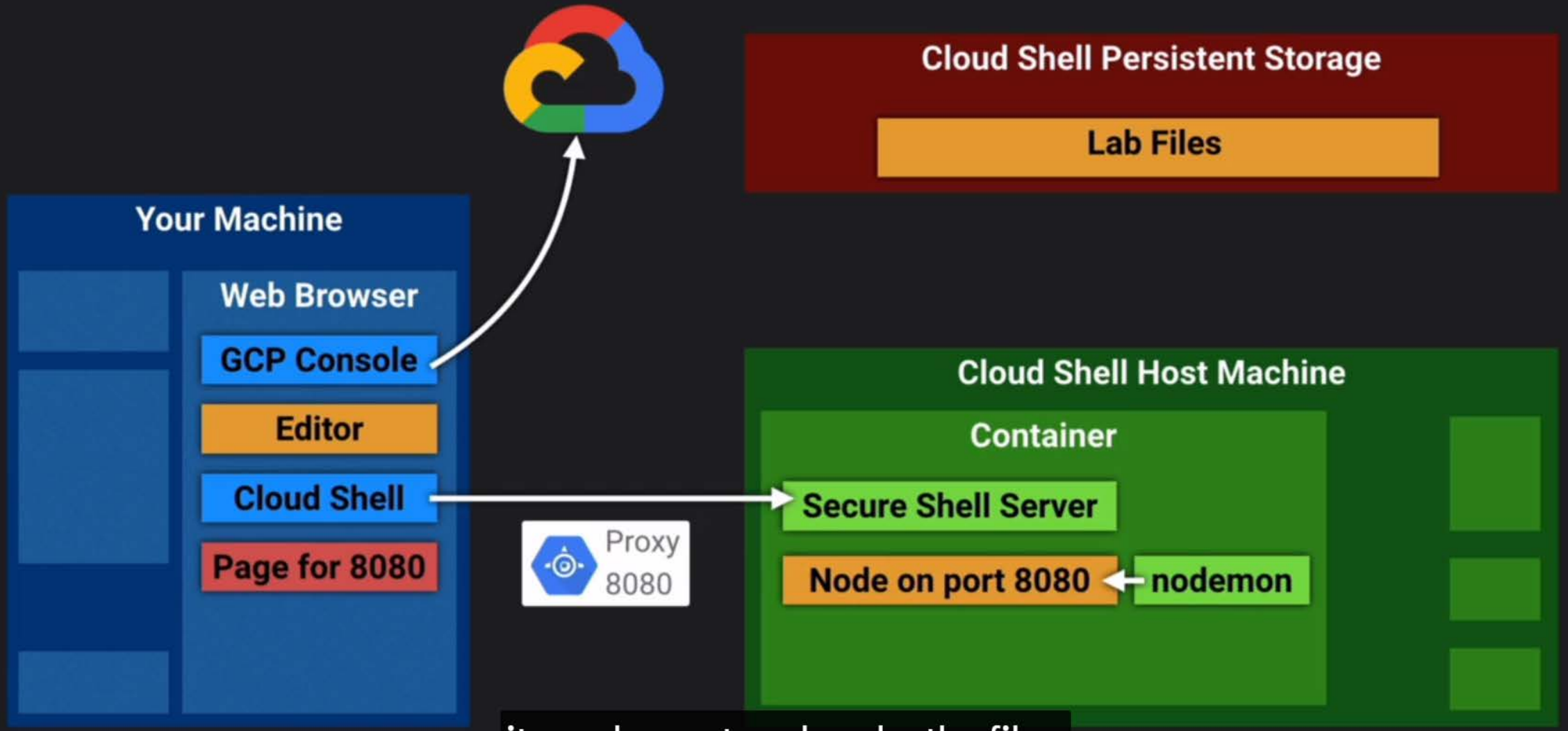


So again, when the node process starts,

Data Flows

Cloud Shell Lab Data Flow

A CLOUD GURU

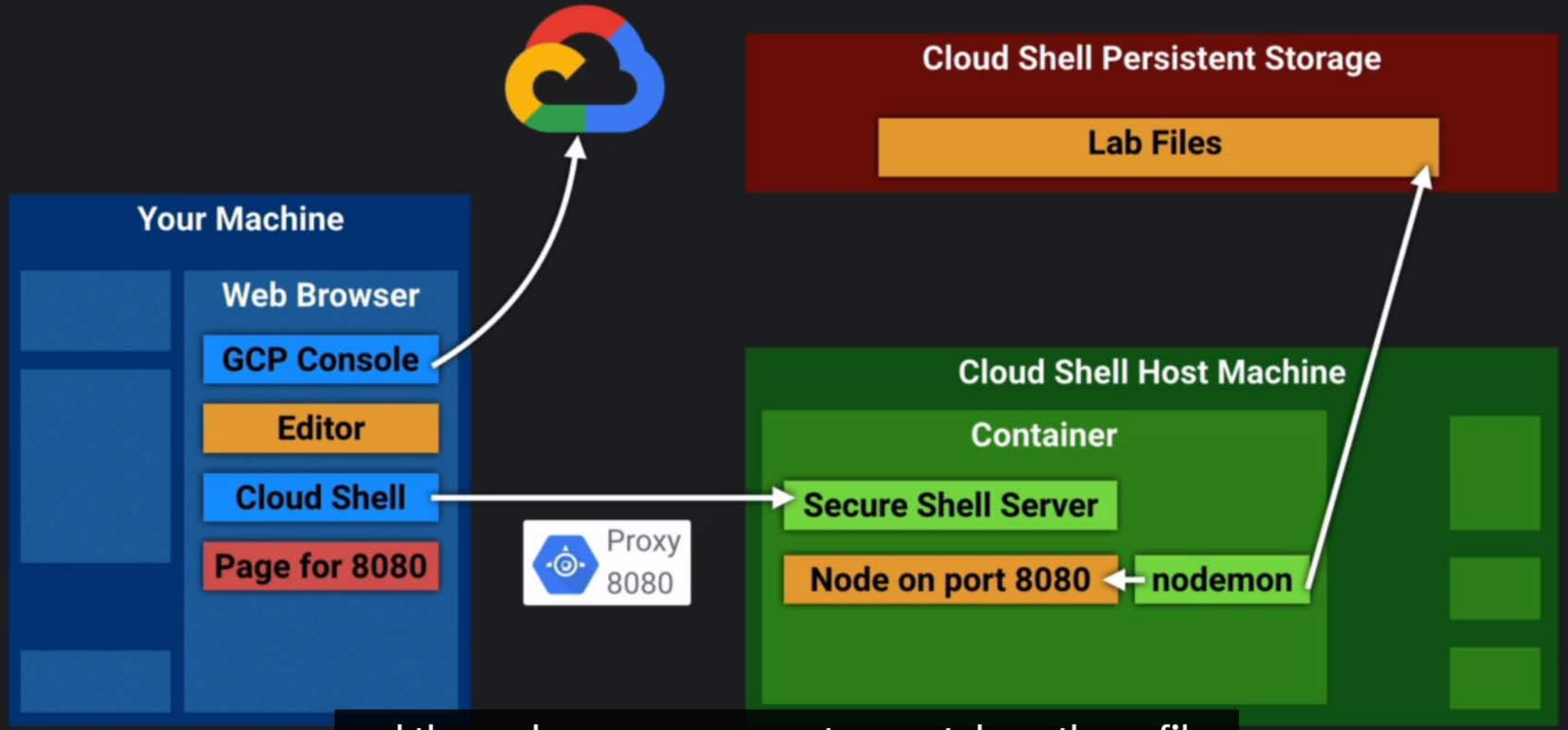


it reaches out and grabs the files,

Data Flows

Cloud Shell Lab Data Flow

A CLOUD GURU

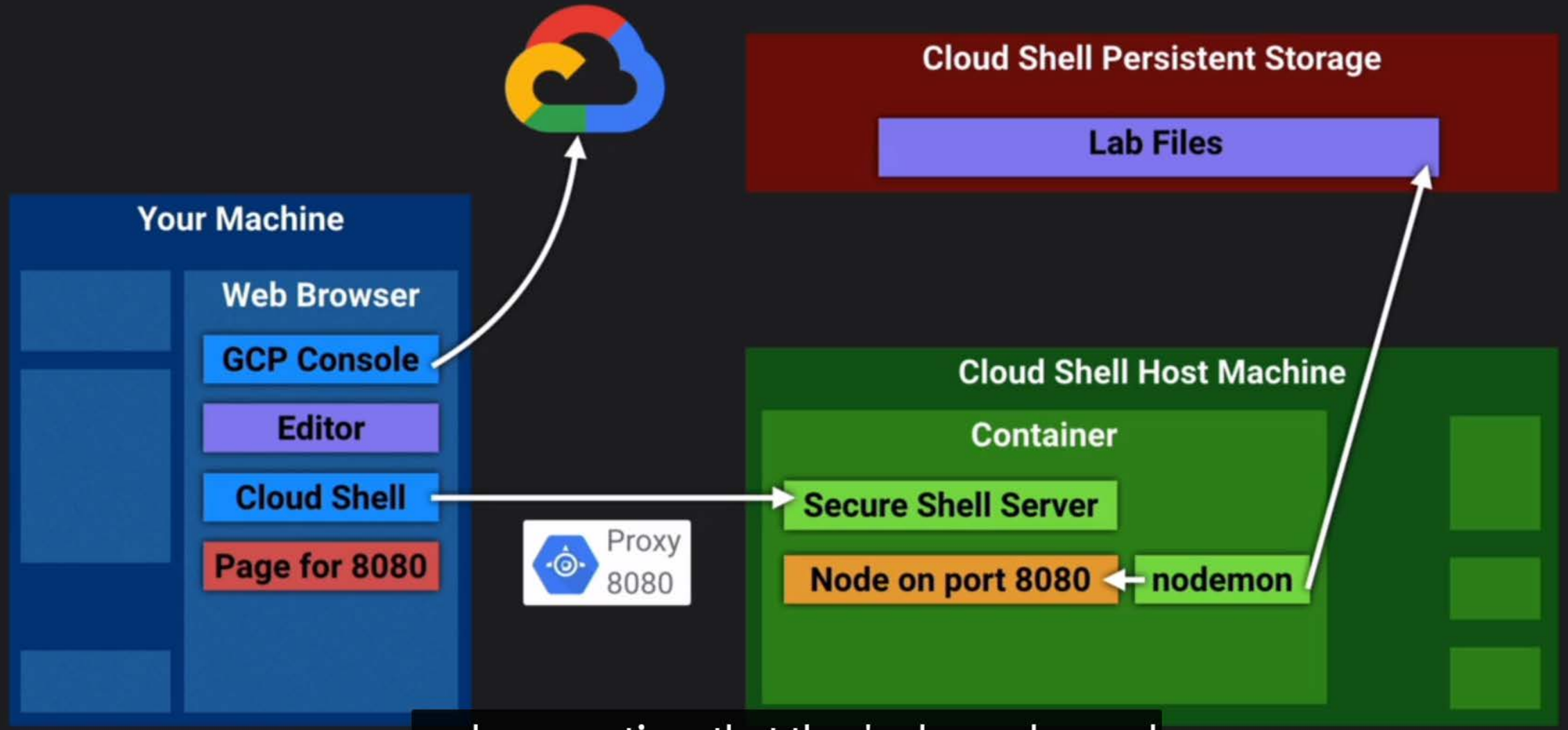


and the nodemon process puts a watch on those files.

Data Flows

Cloud Shell Lab Data Flow

A CLOUD GURU

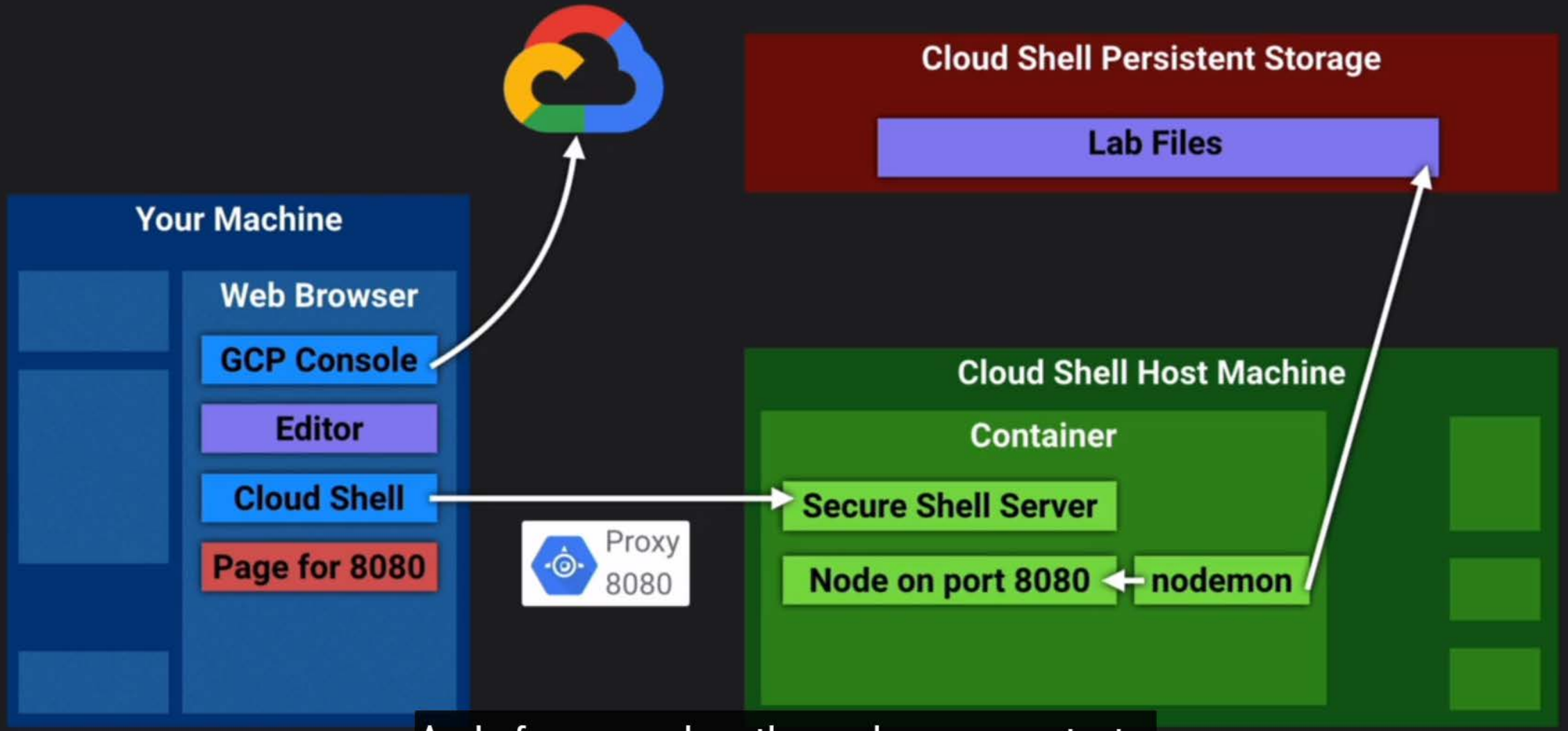


nodemon notices that they've been changed

Data Flows

Cloud Shell Lab Data Flow

A CLOUD GURU

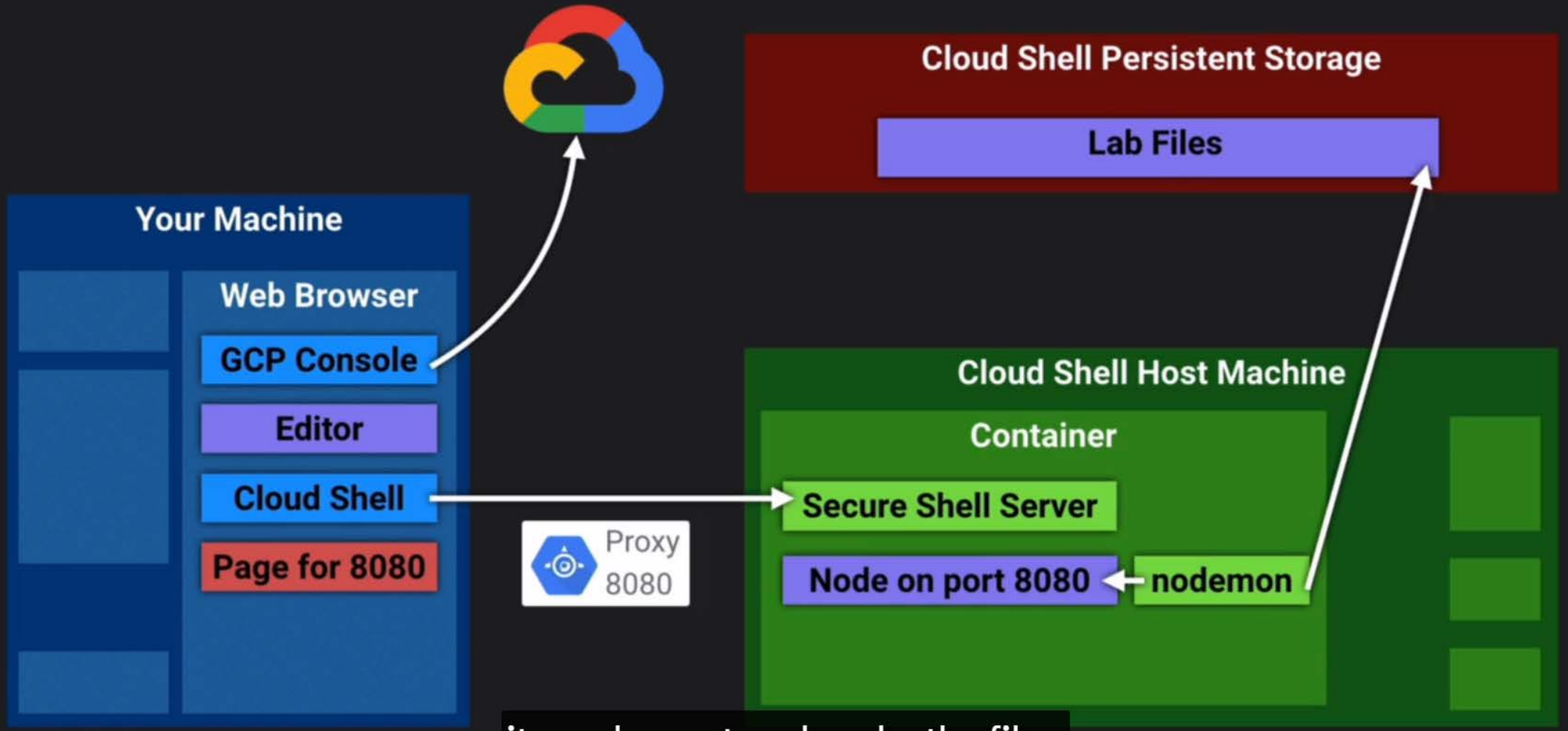


And of course when the node process starts,

Data Flows

Cloud Shell Lab Data Flow

A CLOUD GURU



it reaches out and grabs the files.

Key Takeaways

- Data flows are the foundation of *every* system
- Moving, Processing, Remembering
  - Not *just* Network, Compute, Storage
- Build mental models
  - Helps you make predictions
- Identify and think through data flows
  - Highlights potential issues
- Requirements and options not always clear
  - Especially in the real world
- Critical skills for both real world and exam questions

So the skills we practiced here are critical

Rundown on gcloud

Basic Syntax

A CLOUD GURU

```
gcloud <global flags> <service/product> <group/area> <command> <flags> <parameters>
```

- Always drills down (from left to right)
- Examples:
  - `gcloud --project myprojid compute instances list`
  - `gcloud --project=myprojid compute instances list`
  - `gcloud compute instances create myvm`
  - `gcloud services list --available`
  - `gsutil ls`

against the Google Cloud Storage service,

Basic Syntax

A CLOUD GURU

```
gcloud <global flags> <service/product> <group/area> <command> <flags> <parameters>
```

- Always drills down (from left to right)
- Examples:
  - `gcloud --project myprojid compute instances list`
  - `gcloud --project=myprojid compute instances list`
  - `gcloud compute instances create myvm`
  - `gcloud services list --available`
  - `gsutil ls`
  - `gsutil mb -l northamerica-northeast1 gs://storage-lab-cli`
  - `gsutil label set bucketlabels.json gs://storage-lab-cli/`

For this example gsutil also has you specify the area

Global Flags

- `--help`
- `-h`
- `--project <ProjectID>`
- `--account <Account>`
- `--filter`
  - Not always available, but often better than using grep
- `--format`
  - Can choose JSON, YAML, CSV, etc.
  - Can pipe ("|") JSON to "jq" command for further processing
- `--quiet` (or `-q`)

This could definitely be helpful if you have a script

Config Properties

A CLOUD GURU

- Values entered once and used by any command that needs them
- Can be overridden on a specific command with corresponding flag
- Used very often for account, project, region, and zone
  - Set “core/account” or “account” to replace “--account”
  - Set “core/project” or “project” to replace “--project”
  - Set “compute/region” to replace “--region”
  - Set “compute/zone” to replace “--zone”
- Set with `gcloud config set <property> <value>`
- Check with `gcloud config get-value <property>`
- Clear with `gcloud config unset <property>`

Configurations

- Can maintain groups of settings and switch between them
- Most useful when using multiple projects
- Interactive workflow to set common properties in a config with `gcloud init`
- List all properties in a configuration with `gcloud config list`
- List all configurations with `gcloud config configurations list`
  - IS_ACTIVE column shows which one is currently being used
  - Other columns list account, project, region, zone, and the name of the config
- Make new config with `gcloud config configurations create ITS_NAME`
- Start using config with `gcloud config configurations activate ITS_NAME`
  - Or use for just one command with `--configuration=ITS_NAME`

Structured Breakdown of Questions

A CLOUD GURU

Understand	• Determine the key question—the “kicker” • Figure out what everything means—question and responses
Eliminate	• Get rid of responses that have fake info or other errors • Get rid of responses that conflict with the key question
Evaluate	• Think through all the tradeoffs for remaining responses • Consider both stated and implied dimensions
Choose	• Pick exactly the right number • Select the best options or eliminate the worst ones
Validate	• Make sure your responses *answer* the key question • Make sure your responses don’t conflict with any details

What is “proper” data flow? (CIA)

A CLOUD GURU

- You cannot view data you shouldn’t — **Confidentiality**
- You cannot change data you shouldn’t — **Integrity**
- You *can* access data you *should* — **Availability**

But let's take a look at these three things

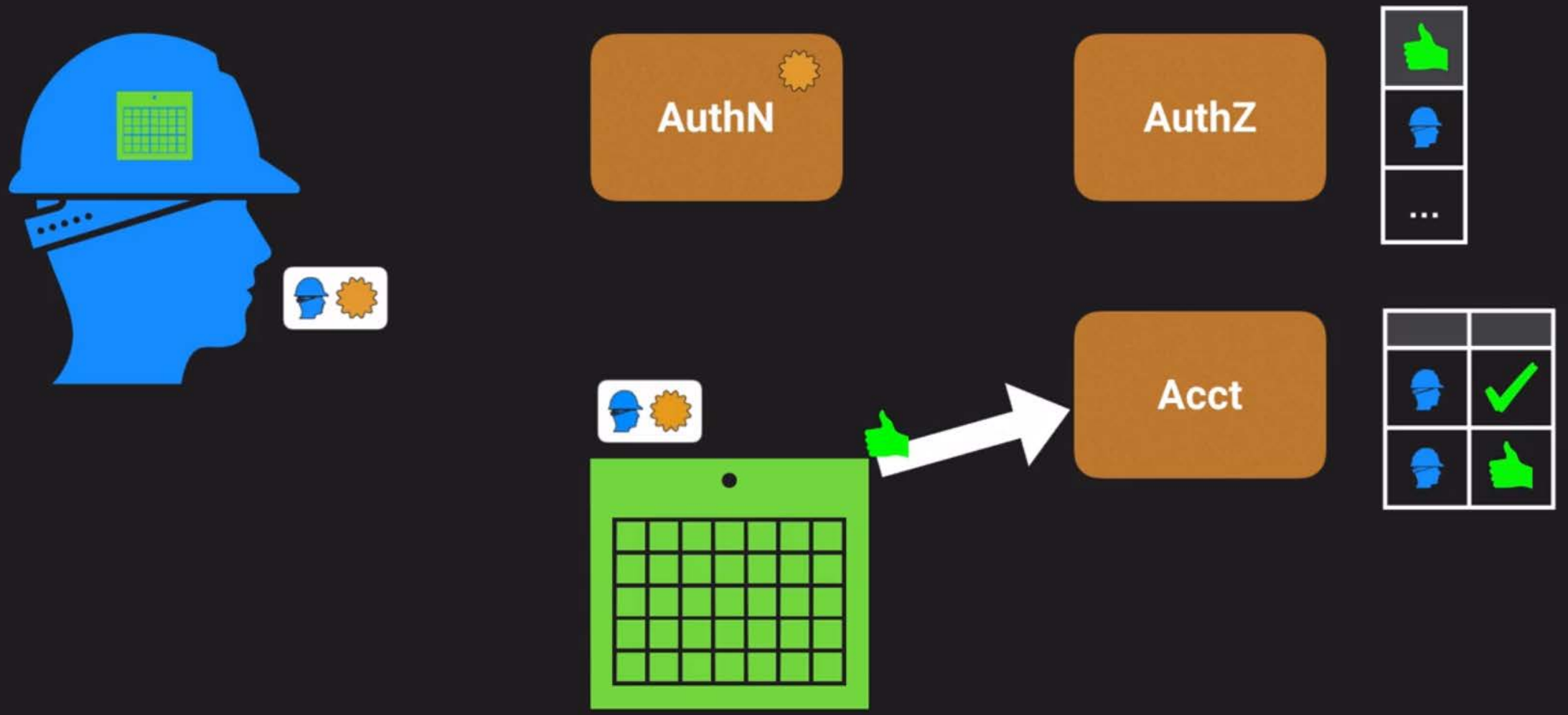
How do we control data flow? (AAA)

A CLOUD GURU

- Authentication - Who are you?
- Authorization - What are you allowed to do?
- Accounting - What did you do?

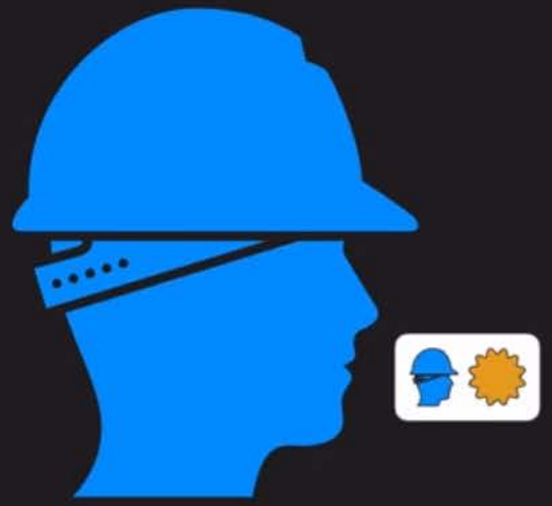
but it can also include viewing information.

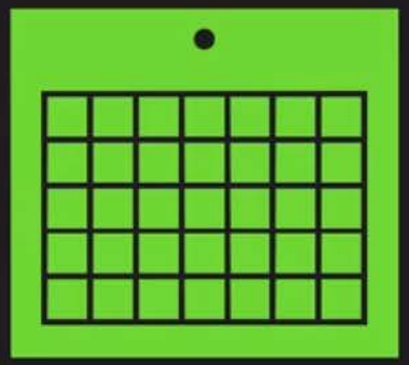
AAA Data Flow



and this action is then logged in the accounting system too.

AAA Data Flow





AuthN

AuthZ

Acct

and records that rejected access in the accounting system.

What enables security in GCP?

A CLOUD GURU

- Security Products
- Security Features
- Security Mindset
  - Includes Availability Mindset

The availability mindset is something

Key Security Products/Features — AuthN

A CLOUD GURU

- Identity
  - Humans in G Suite, Cloud Identity
  - Applications & services use Service Accounts
- Identity hierarchy
  - Google Groups
- Can use Google Cloud Directory Sync (GCDS) to pull from LDAP (no push)

you can use Google Cloud Directory Sync

Key Security Products/Features — AuthZ

A CLOUD GURU

- Identity hierarchy (Google Groups)
- Resource hierarchy (Organization, Folders, Projects)
- Identity and Access Management (IAM)
  - Permissions
  - Roles
  - Bindings
- GCS ACLs
- Billing management
- Networking structure & restrictions

can definitely be set up to help support security,

Key Security Products/Features — Acct

A CLOUD GURU

- Audit / Activity Logs (provided by Stackdriver)
- Billing export
  - To BigQuery
  - To file (in GCS bucket)
    - Can be JSON or CSV
- GCS Object Lifecycle Management

of sensitive material on the correct schedule.

Permissions

- A Permission allows you to perform a certain action
- Each one follows the form `Service.Resource.Verb`
- Usually correspond to REST API methods
- Examples:
  - `pubsub.subscriptions.consume`
  - `pubsub.topics.publish`

and pubsub dot topics dot publish.

IAM Breakdown

Roles

A CLOUD GURU

- A Role is a collection of Permissions to use or manage GCP resources
- Primitive Roles — Project-level and often too broad
  - Viewer is read-only
  - Editor can view and change things
  - Owner can also control access & billing
- Predefined Roles — Give granular access to specific GCP resources
  - E.g.: `roles/bigquery.dataEditor`, `roles/pubsub.subscriber`

and the permission to consume from a subscription

IAM Breakdown

Roles

A CLOUD GURU

- A Role is a collection of Permissions to use or manage GCP resources
- Primitive Roles — Project-level and often too broad
  - Viewer is read-only
  - Editor can view and change things
  - Owner can also control access & billing
- Predefined Roles — Give granular access to specific GCP resources
  - E.g.: `roles/bigquery.dataEditor`, `roles/pubsub.subscriber`
  - Read through the list of roles for each product! Think about why each exists.
- Custom Role — Project- or Org-level collection you define of granular permissions

or organization level and you can put whatever Permissions

Predefined Role Examples — App Engine

A CLOUD GURU

Role Name	Role Title	Description
roles/appengine.appAdmin	App Engine Admin	Read/Write/Modify access to all application configuration and settings.
roles/appengine.serviceAdmin	App Engine Service Admin	Read-only access to all application configuration and settings. Write access to module-level and version-level settings. Cannot deploy a new version.
roles/appengine.deployer	App Engine Deployer	Read-only access to all application configuration and settings. Write access only to create a new version; cannot modify existing versions other than deleting versions that are not receiving traffic.
roles/appengine.appViewer	App Engine Viewer	Read-only access to all application configuration and settings.
roles/appengine.codeViewer	App Engine Code Viewer	Read-only access to all application configuration, settings, and deployed source code.

Members

A CLOUD GURU

- A Member is some Google-known identity
- Each Member is identified by a unique email address
- Can be:
  - **`user:`** Specific Google account
    - G Suite, Cloud Identity, Gmail, or validated email
  - **`serviceAccount:`** Service account for apps/services
  - **`group:`** Google group of users and service accounts
  - **`domain:`** Whole domain managed by G Suite or Cloud Identity
  - **`allAuthenticatedUsers`** — *Any* Google account or service account
  - **`allUsers`** — Anyone on the Internet (Public)

Groups

A CLOUD GURU

- "A Google group is a named collection of Google accounts and service accounts."
- "Every group has a unique email address that is associated with the group."
- You never act *as* the group
  - But membership in a group can grant capabilities to individuals
- Use them for everything!
- Can be used for owner when within an organization
- *Can* nest groups in an organization
  - Example: one group for each department, all those in group for all staff

Policies

A CLOUD GURU

- A Policy binds Members to Roles for some scope of Resources
- Answers: Who can do what to which thing(s)?
- Attached to some level in the Resource Hierarchy
  - Organization, Folder, Project, Resource
- Roles and Members listed in policy, but Resources identified by attachment
- Always additive (“Allow”) and never subtractive (no “Deny”)
  - "Child policies cannot restrict access granted at a higher level.”

Policies (cont.)

- One policy per resource
- Max 1500 member bindings per policy
  - Ridiculously high max
  - Anywhere close and “You’re doing it wrong!”
  - Use groups, instead!!!
    - Not kidding!
    - Don’t forget this
- You should use groups!
- Usually takes less than 60s to apply changes (both granting and revoking)
- “[M]ay take up to 7 minutes for… changes to fully propagate across the system”

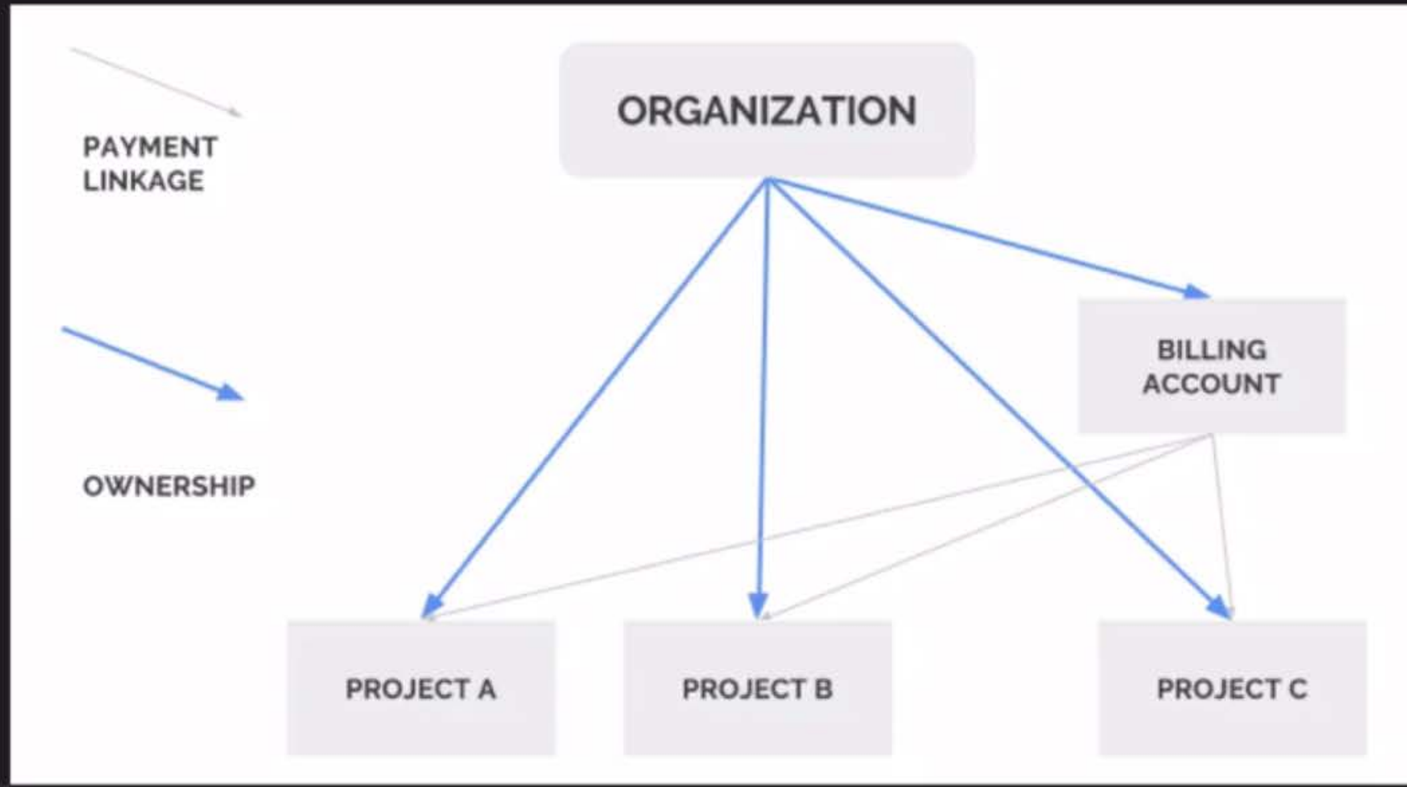
Managing Policy Bindings

- Can use `get-iam-policy`, edit the JSON/YAML, and `set-iam-policy` back
  - But don't! Instead, prefer...
- `gcloud [GROUP] add-iam-policy-binding [RESOURCE-NAME] --role [ROLE-ID-TO-GRANT] --member user:[USER-EMAIL]`
- `gcloud [GROUP] remove-iam-policy-binding [RESOURCE-NAME] --role [ROLE-ID-TO-REVOKE] --member user:[USER-EMAIL]`
- Atomic operations are better because changes:
  - Are simpler, less work, and less error-prone (than editing JSON/YAML)
  - *Avoid race conditions, so can happen simultaneously*
- `gcloud beta compute instances add-iam-policy-binding myhappyvm --role roles/compute.instanceAdmin --member user:me@example.com`

Billing Accounts

A CLOUD GURU

- A Billing Account represents some way to pay for GCP service usage
- Type of Resource that lives outside of Projects
- Can belong to an Organization (i.e. be owned by it)
  - Inherits Org-level IAM policies
- Can be linked to projects
  - But does not own them
    - No impact on project IAM



Billing IAM Roles

Role	Purpose	Scope
Billing Account Creator	Create new self-serve billing accounts.	Org
Billing Account Administrator	Manage billing accounts (but not create them).	Billing Account
Billing Account User	Link projects to billing accounts.	Billing Account
Billing Account Viewer	View billing account cost information and transactions.	Billing Account
Project Billing Manager	Link/unlink the project to/from a billing account.	Project

Monthly Invoiced Billing

A CLOUD GURU

- Get billed monthly and pay by invoice due date
- Can pay via check or wire transfer
- Can increase project and quota limits
- Billing administrator of org's current billing account contacts Cloud Billing Support
  - To determine eligibility
  - To apply to switch to monthly invoicing
- Eligibility depends on
  - account age
  - typical monthly spend
  - country

what country you're in.

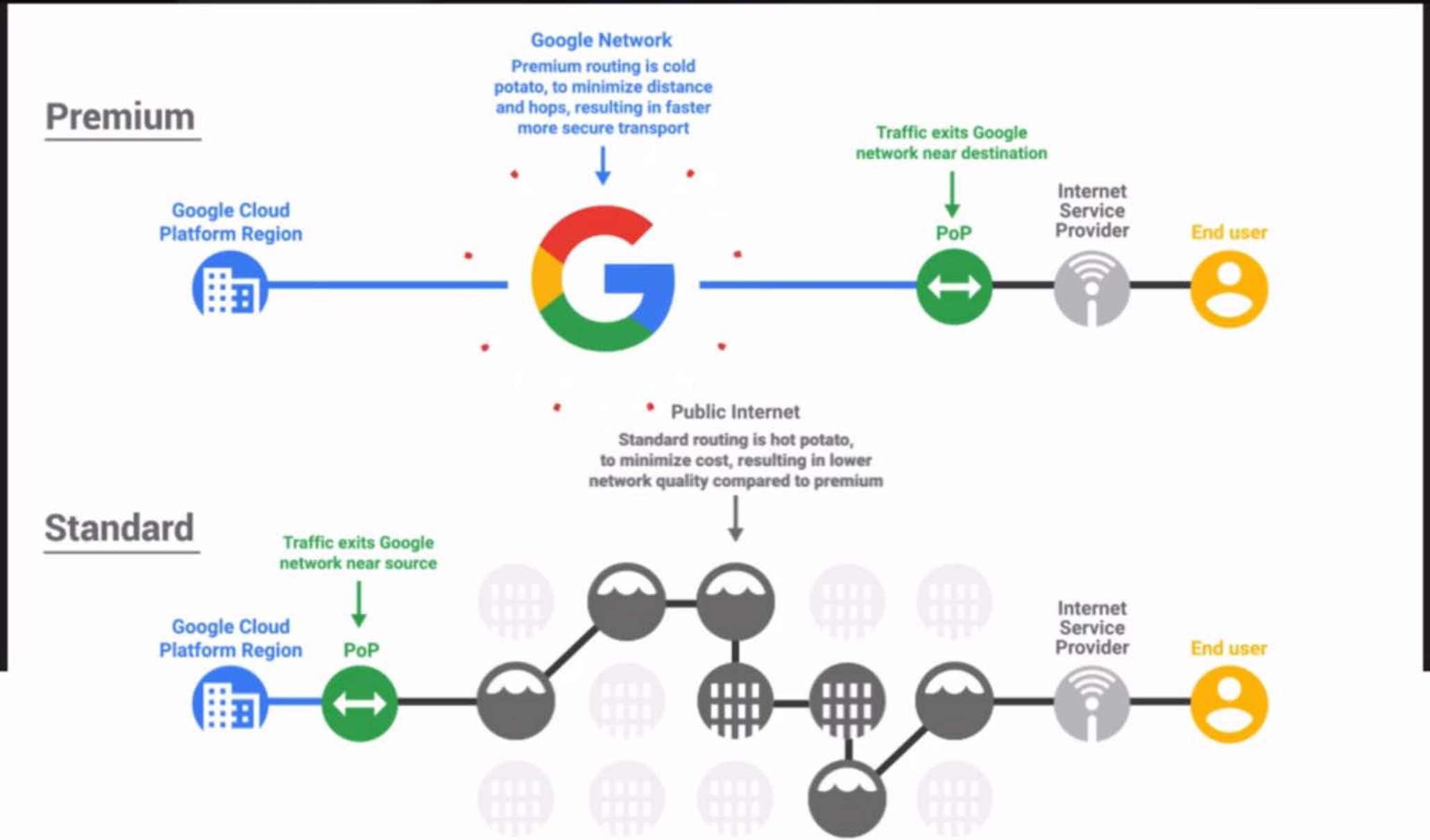
What is it?

A CLOUD GURU

- Routing is about deciding where data should go next
- Like a direction marker on a hiking trail
  - Fork in the trail:
    - Go that way to get to the parking lot
    - Go that way to get to the waterfall
    - Go that way to get to the lake
    - Go that way to get to the peak
  - At the peak:
    - Go that way (to get down)
- Many local decisions — No full map or path

local decisions about how to move data from

Getting data to Google's network



Getting data to the right resource

A CLOUD GURU

- **Latency reduction**
  - Use servers physically close to clients

Cross-Region Load Balancing
(with Global Anycast IPs)

- **Load balancing**
  - Separate from auto-scaling

Cloud Load Balancer
(all types; internal and external)

- **System design**
  - Different servers may handle different parts of the system
  - Especially when using microservices (instead of a monolith)

and also load balancers at both layer four or layer seven

Getting data to the right resource

A CLOUD GURU

- **Latency reduction** — **Cross-Region Load Balancing**
  - Use servers physically close to clients — (with Global Anycast IPs)
- **Load balancing** — **Cloud Load Balancer**
  - Separate from auto-scaling — (all types; internal and external)
- **System design** — **HTTP(S) Load Balancer**
  - Different servers may handle different parts of the system — (with URL Map)
  - Especially when using microservices (instead of a monolith)

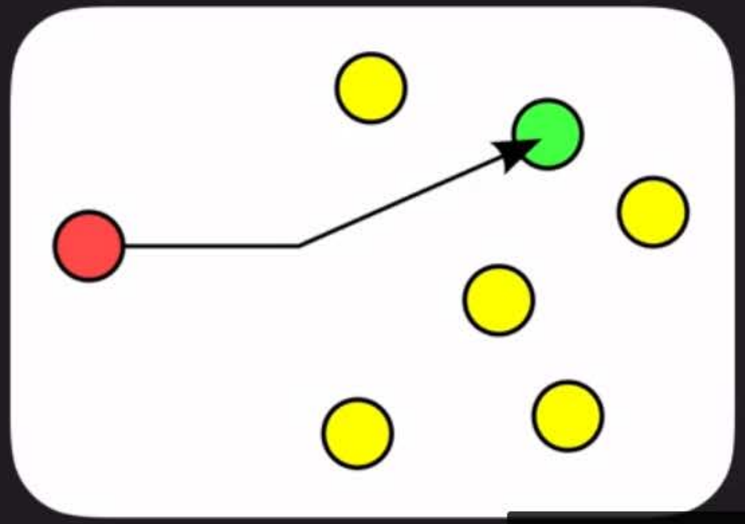
This also we'll take a closer look at in just a minute.

Unicast vs Anycast

A CLOUD GURU

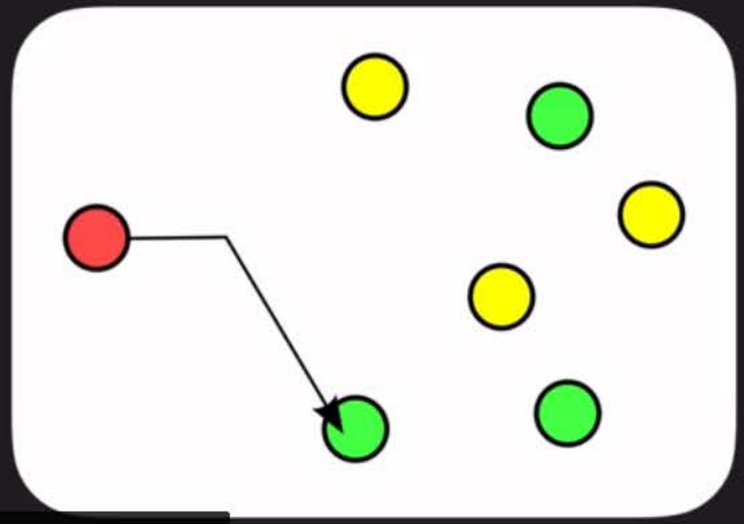
Unicast

There is only one **uni**que device
in the world
that can handle this;
send it there.



Anycast

There are multiple devices
that could handle this;
send it to **any** one—
but ideally the closest.



So it's quite special that Google offers

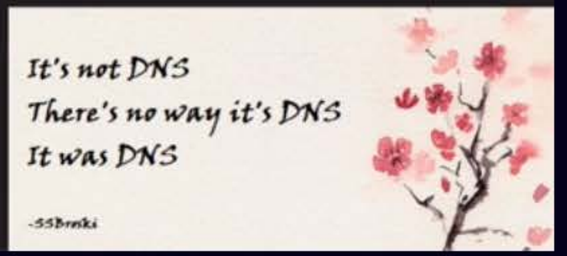
Layer 4 vs. Layer 7

- TCP (of TCP/IP) is usually called Layer 4 (L4)
  - It works solely with IP addresses
- HTTP and HTTPS work at Layer 7 (L7)
  - These know about URLs and paths
- Each layer is built on the one below it
- Therefore:
  - To route based on URL paths, routing needs to understand L7
  - L4 cannot route based on the URL paths defined in L7

So let's say that you have a bunch of packages,

What about DNS?

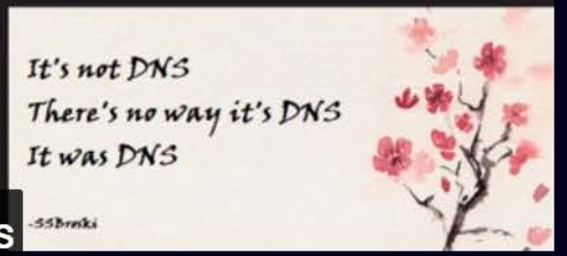
- Name resolution (via the Domain Name System) can be the first step in routing
- But that comes with a number of problems:
  - Layer 4 — Cannot route L4 based on L7's URL paths
  - Chunky — DNS queries often cached and reused for huge client sets
  - Sticky — DNS lookup "locks on" and refreshing per request has high cost
    - Extra latency because each request includes another round-trip!
    - More money for additional DNS request processing
  - Not Robust — Relies on the client always doing the right thing
    - Spoiler: They don't



And here's a spoiler, they don't.

What about DNS?

- Name resolution (via the Domain Name System) can be the first step in routing
- But that comes with a number of problems:
  - Layer 4 — Cannot route L4 based on L7's URL paths
  - Chunky — DNS queries often cached and reused for huge client sets
  - Sticky — DNS lookup "locks on" and refreshing per request has high cost
    - Extra latency because each request includes another round-trip!
    - More money for additional DNS request processing
  - Not Robust — Relies on the client always doing the right thing
    - Spoiler: They don't



There have been many documented cases of clients

What about DNS?

A CLOUD GURU

- Name resolution (via the Domain Name System) can be the first step in routing
- But that comes with a number of problems:
  - Layer 4 — Cannot route L4 based on L7's URL paths
  - Chunky — DNS queries often cached and reused for huge client sets
  - Sticky — DNS lookup "locks on" and refreshing per request has high cost
    - Extra latency because each request includes another round-trip!
    - More money for additional DNS request processing
  - Not Robust — Relies on the client always doing the right thing
    - Spoiler: They don't
- Premium tier "cold potato" routing with global anycast IPs avoids these problems

Getting data from one resource to another

A CLOUD GURU

- VPC (global) is Virtual Private Cloud — Your private SDN space in GCP
  - Not just resource-to-resource — Also manages the doors to outside & peers
- Subnets (regional) create logical spaces to contain resources
  - All Subnets can reach all others — globally, without any need for VPNs
- Routes (global) define “next hop” for traffic based on destination IP
  - Routes are global and apply by Instance-level Tags, not by Subnet
  - No route to the internet gateway means no such data can flow
- Firewall Rules (global) further filter data flow that would otherwise route
  - All Firewall Rules are global and apply by Instance-level Tags or Service Acct.
  - Default Firewall Rules are restrictive inbound and permissive outbound

Compute Engine (GCE)

Zonal Regional Multi-Regional Global

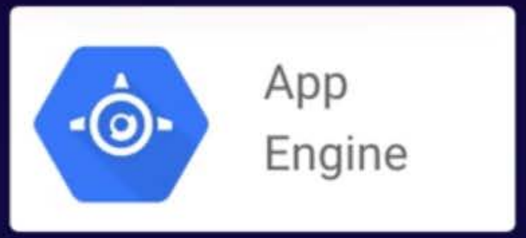
- Fast-booting Virtual Machines (VMs) you can rent, on demand
- Infrastructure as a Service (IaaS)
- Pick set machine type—standard, highmem, highcpu—or custom CPU/RAM
- Pay by the second (60 second min.) for CPUs, RAM
- Automatically cheaper if you keep running it (“sustained use discount”)
- Even cheaper for “preemptible” or long-term use commitment in a region
- Can add GPUs and paid OSes for extra cost
- Live Migration: Google seamlessly moves instance across hosts, as needed

Kubernetes Engine (GKE)

Zonal Regional Multi-Regional Global

- Managed Kubernetes cluster for running Docker containers (with autoscaling)
- Used to be called “Google Container Engine” (but still GKE) until Nov, 2017
- Kubernetes DNS on by default for service discovery
- No IAM integration (unlike AWS’s ECS)
- Integrates with Persistent Disk for storage
- Pay for underlying GCE instances
  - Production cluster should have 3+ nodes
- No GKE management fee, no matter how many nodes in cluster, as of Nov 2017

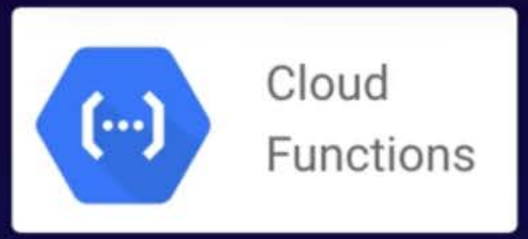
App Engine (GAE)



Zonal Regional Multi-Regional Global

- Platform as a Service (PaaS) that takes your code and runs it
- Much more than just compute — Integrates storage, queues, NoSQL, ...
- Flex mode ("App Engine Flex") can run any container & access VPC
- Auto-scales based on load
  - Standard (non-Flex) mode can turn off last instance when no traffic
- Effectively pay for underlying GCE instances and other services

Cloud Functions (GCF)



Zonal Regional Multi-Regional Global

- Runs code in response to an event — Node.js, Python, Java, Go
- Functions as a Service (FaaS), “Serverless”
- Pay for CPU and RAM assigned to function, per 100ms (min. 100ms)
- Each function automatically gets an HTTP endpoint
- Can be triggered by GCS objects, Pub/Sub messages, etc.
- Massively scalable (horizontally) — Runs many copies when needed
- Often used for chatbots, message processors, IoT, automation, etc.

Local SSD

Zonal Regional Multi-Regional Global

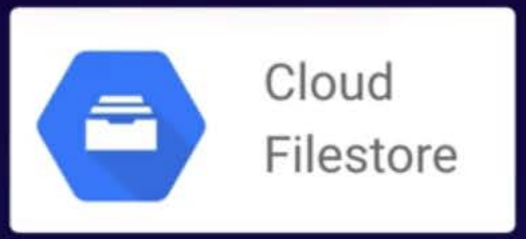
- Very fast 375GB solid state drives physically attached to the server
- Can stripe across eight of them (3TB) for *even* better performance
- DATA WILL BE LOST whenever the instance shuts down
  - But can survive a Live Migration
- Like all data at rest, always encrypted
- Pay by GB-month provisioned (prorated, as always)

Persistent Disk (PD)

Zonal Regional Multi-Regional Global

- Flexible, *block-based* network-attached storage; boot disk for every GCE instance
- Perf scales with volume size; max way below Local SSD, but still plenty fast
- Persistent disks *persist*, and are replicated (zone or region) for durability
- Can resize while in use (up to 64TB), but will need file system update within VM
- Snapshots (and machine images) add even more capability and flexibility
  - "Magical": Pay for incremental ($ and time), but use/delete like full backups
- *Not* file-based NAS, but can mount to multiple instances if *all* are read-only
- Pay for GB/mo provisioned depending on perf. class; plus snapshot GB/mo used

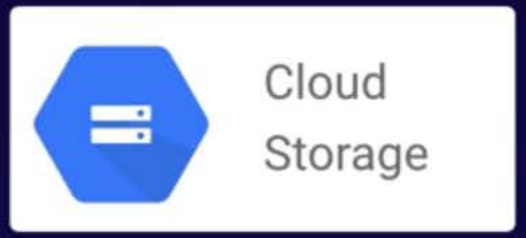
Cloud Filestore



Zonal Regional Multi-Regional Global

- Fully-managed file-based storage
- “Predictably fast performance for your file-based workloads”
- Accessible to GCE and GKE through your VPC, via NFSv3 protocol
- Primary use case is application migration to cloud (“lift and shift”)
- Fully manages file *serving*, but *not* backups
- Pay for provisioned TBs in “Standard” (slow) or “Premium” (fast) mode
- Minimum provisioned capacity of 1TB (Standard) or 2.5TB (Premium)

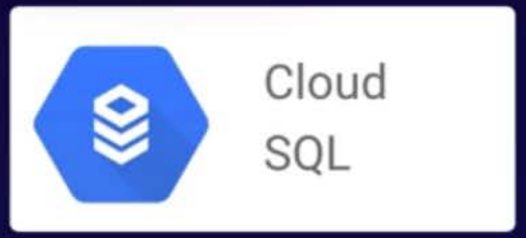
Cloud Storage (GCS)



Zonal Regional Multi-Regional Global

- Infinitely scalable, fully-managed, versioned, and highly-durable object storage
- Designed for 99.999999999% (“eleven nines”) durability
- Strongly consistent (even for overwrite PUTs and DELETEs)
- Integrated site hosting and CDN functionality
- Lifecycle transitions across classes: Multi-Regional, Regional, Nearline, Coldline
  - Diffs in cost & availability (99.95%, 99.9%, 99%, 99%), not latency (no thaw delay)
- All classes have same API, so can use gsutil and gcsfuse (but beware)
- Pay for data operations & GB-months stored by class
- Nearline/Coldline: Also pay for GBs retrieved—plus early deletion fee, if < 30/90 days

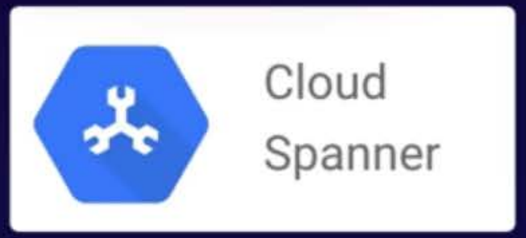
Cloud SQL



Zonal Regional Multi-Regional Global

- Fully-managed and reliable MySQL and PostgreSQL databases
- Supports automatic replication, backup, failover, etc.
- Scaling is manual (both vertically and horizontally)
- Effectively pay for underlying GCE instances and PDs
  - Plus some baked-in service fees

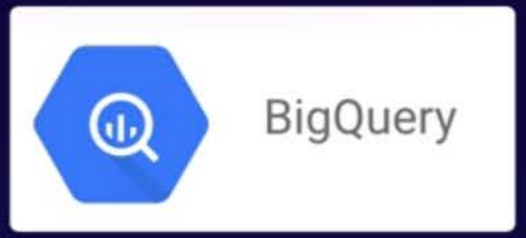
Cloud Spanner



Zonal Regional Multi-Regional Global

- "The first horizontally scalable, strongly consistent, relational database service"
  - "From 1 to hundreds or thousands of nodes"
  - "A minimum of 3 nodes is recommended for production environments."
- Chooses Consistency and Partition-Tolerance (CP of CAP theorem)
- But still *high* Availability: SLA has 99.999% SLO (five nines) for multi-region
  - Nothing is actually 100%, really
  - *Not* based on fail-over
- Pay for provisioned node time (by region/multi-region) plus used storage-time

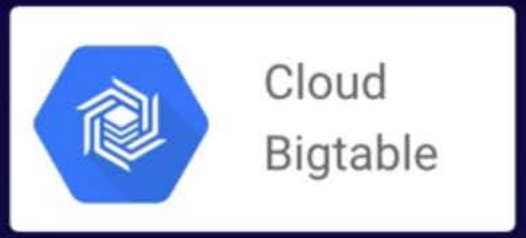
BigQuery (BQ)



Zonal Regional Multi-Regional Global

- Serverless column-store data warehouse for analytics using SQL
- Scales internally, so it “can scan TB in seconds and PB in minutes”
- Pay for GBs actually considered (scanned) during queries
  - Attempts to reuse cached results, which are free
- Pay for data stored (GB-months)
  - Relatively inexpensive
  - Even cheaper when table not modified for 90 days (reading still fine)
- Pay for GBs added via streaming inserts

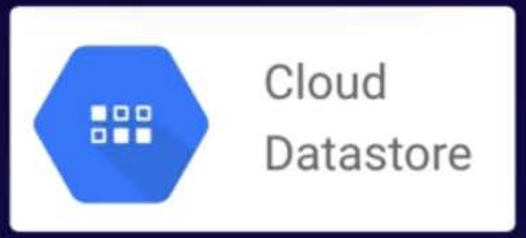
Cloud Bigtable



Zonal Regional Multi-Regional Global

- Low latency & high throughput NoSQL DB for large operational & analytical apps
- Supports open-source HBase API
- Integrates with Hadoop, Dataflow, Dataproc
- Scales seamlessly and unlimitedly
  - Storage autoscales
  - Processing nodes must be scaled manually
- Pay for processing node hours
- Pay for GB-hours used for storage (cheap HDD or fast SSD)

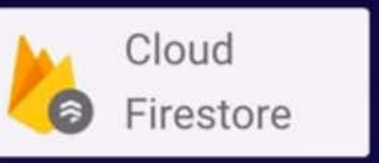
Cloud Datastore



Zonal Regional Multi-Regional Global

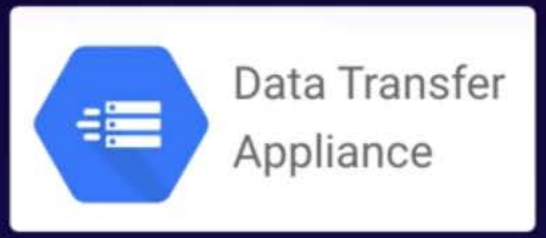
- Managed & autoscaled NoSQL DB with indexes, queries, and ACID trans. support
- NoSQL, so queries can get complicated
  - No joins or aggregates and must line up with indexes
  - NOT, OR, and NOT EQUALS (<>, !=) operations not natively supported
- Automatic “built-in” indexes for simple filtering and sorting (ASC, DESC)
- Manual “composite” indexes for more complicated, but beware them “exploding”
- Pay for GB-months of storage used (including indexes)

Firebase Realtime DB & Cloud Firestore



- NoSQL document stores with ~real-time client updates via managed websockets
- Firebase DB is single (potentially huge) JSON doc, located only in central US
- Cloud Firestore has collections, documents, and contained data
- Free tier (Spark), flat tier (Flame), or usage-based pricing (Blaze)
  - Realtime DB: Pay more for GB/month stored and GB downloaded
  - Firestore: Pay for operations and much less for storage and transfer

Data Transfer Appliance



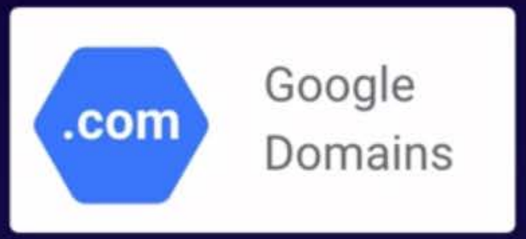
- Rackable, high-capacity storage server to physically ship data to GCS
- Ingest only; not a way to avoid egress charges
- 100TB or 480TB versions
- 480TB/week is faster than a saturated 6Gbps link

Storage Transfer Service

Zonal Regional Multi-Regional Global

- Copies objects for you, so you don't need to set up a machine to do it
- Destination is always GCS bucket
- Source can be S3, HTTP/HTTPS endpoint, or another GCS bucket
- One-time or scheduled recurring transfers
- Free to use, but you pay for its actions

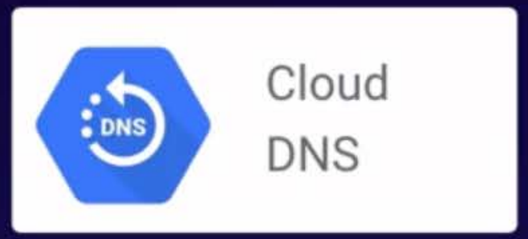
Google Domains



Zonal Regional Multi-Regional Global

- Google's registrar for domain names
- Private Whois records
- Built-in DNS or custom nameservers
- Supports DNSSEC
- Email forwarding with automatic setup of SPF and DKIM (for built-in DNS)

Cloud DNS



Zonal Regional Multi-Regional Global

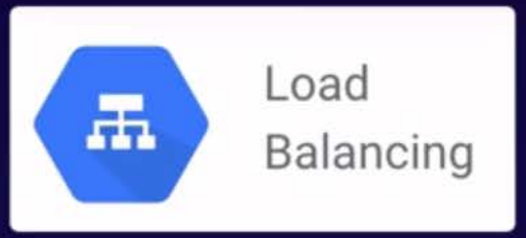
- Scalable, reliable, & managed authoritative Domain Name System (DNS) service
- 100% uptime guarantee
- Public and private managed zones
- Low latency globally
- Supports DNSSEC
- Manage via UI, CLI, or API
- Pay fixed fee per managed zone to store and distribute DNS records
- Pay for DNS lookups (i.e. usage)

Static IP Addresses

Zonal Regional Multi-Regional Global

- Reserve static IP addresses in projects and assign them to resources
- Regional IPs used for GCE instances & Network Load Balancers
- Global IPs used for global load balancers:
  - HTTP(S), SSL proxy, and TCP proxy
  - “Anycast IP” simplifies DNS
- Pay for reserved IPs that are not in use, to discourage wasting them

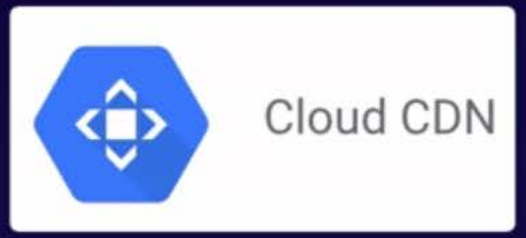
Cloud Load Balancing (CLB)



Zonal Regional Multi-Regional Global

- High-perf, scalable traffic distribution integrated with autoscaling & Cloud CDN
- SDN naturally handles spikes without any prewarming; no instances or devices
- Regional Network Load Balancer: health checks, round robin, session affinity
  - Forwarding rules based on IP, protocol (e.g. TCP, UDP), and (optionally) port
- Global load balancers w/ multi-region failover for HTTP(S), SSL proxy, & TCP proxy
  - Prioritize low-latency connection to region near user, then gently fail over in bits
  - Reacts quickly (unlike DNS) to changes in users, traffic, network, health, etc.
- Pay by making ingress traffic billable (cheaper than egress) plus hourly per rule

Cloud CDN



Zonal Regional Multi-Regional Global

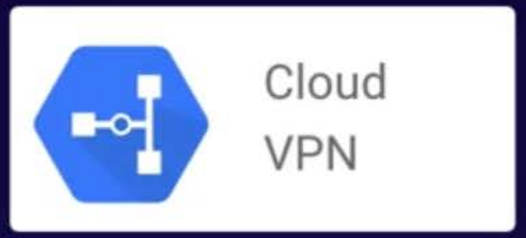
- Low-latency content delivery based on HTTP(S) CLB & integrated w/ GCE & GCS
- Supports HTTP/2 and HTTPS, but no custom origins (GCP only)
- Simple checkbox on HTTP(S) Load Balancer config turns this on
- On cache miss, pay origin→POP “cache fill” egress charges (cheaper for in-region)
- Always pay POP→client egress charges, depending on location
- Pay for HTTP(S) request volume
- Pay per cache invalidation request (not per resource invalidated)
- Origin costs (e.g. CLB, GCS) can be much lower because cache hits reduce load

Virtual Private Cloud (VPC)

Zonal Regional Multi-Regional Global

- Global IPv4 unicast Software-Defined Network (SDN) for GCP resources
- Automatic mode is easy; custom mode gives control
- Configure subnets (each with a private IP range), routes, firewalls, VPNs, BGP, etc.
- VPC is global and subnets are regional (not zonal!)
- Can be shared across multiple projects in same org and peered with other VPCs
- Can enable private (internal IP) access to some GCP services (e.g. BQ, GCS)
- Free to configure VPC (container)
- Pay to use certain services (e.g. VPN) and for network egress

Cloud Virtual Private Network (VPN)



Zonal Regional Multi-Regional Global

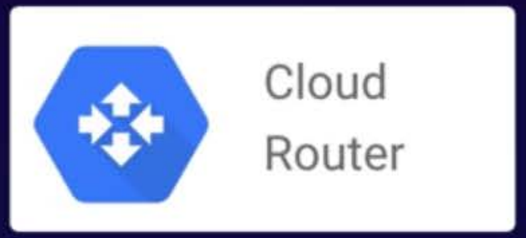
- IPsec VPN to connect to VPC via public internet for low-volume data connections
- For persistent, static connections between gateways (i.e. not for a dynamic client)
  - Peer VPN gateway must have static (unchanging) IP
- Encrypted link to VPC (as opposed to Dedicated Interconnect), into one subnet
- Supports both static and dynamic routing
- 99.9% availability SLA
- Pay per tunnel-hour
- Normal traffic charges apply

Dedicated Interconnect

Zonal Regional Multi-Regional Global

- Direct physical link between VPC and on-prem for high-volume data connections
- VLAN attachment is private connection to VPC in one region; no public GCP APIs
  - Region chosen from those supported by particular Interconnect Location
- Links are private but not encrypted; can layer your own encryption
- Redundant connections in different locations recommended for critical apps
  - Redundancy achieves 99.99% availability; otherwise 99.9% SLA
- Pay fee per 10 Gbps link, plus (relatively small) fee per VLAN attachment
- Pay reduced egress rates from VPC through Dedicated Interconnect

Cloud Router



Zonal Regional Multi-Regional Global

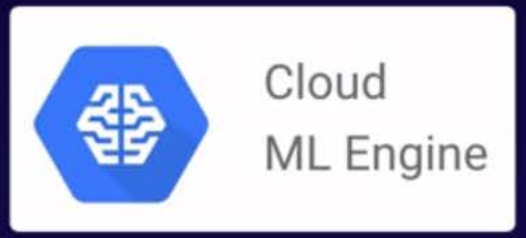
- Dynamic routing (BGP) for hybrid networks linking GCP VPCs to external networks
- Works with Cloud VPN and Dedicated Interconnect
- Automatically learns subnets in VPC and announces them to on-prem network
- Without Cloud Router you must manage static routes for VPN
  - Changing the IP addresses on either side of VPN requires recreating it
- Free to set up
- Pay for usual VPC egress

CDN Interconnect

Zonal Regional Multi-Regional Global

- Direct, low-latency connectivity to certain CDN providers, with cheaper egress
- For external CDNs, not Google's Cloud CDN service
  - Supports Akamai, Cloudflare, Fastly, and more
- Works for both pull and push cache fills
  - Because it's for all traffic with that CDN
- Contact CDN provider to set up for GCP project and which regions
- Free to enable, then pay less for the egress you configured

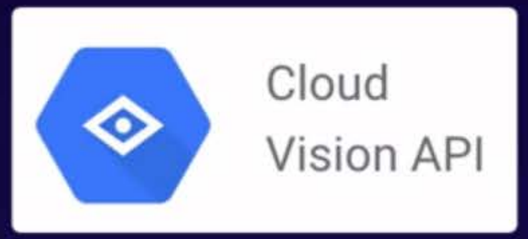
Cloud Machine Learning (ML) Engine



Zonal Regional Multi-Regional Global

- Massively scalable managed service for training ML models & making predictions
- Enables apps/devs to use TensorFlow on datasets of any size; endless use cases
- Integrates: GCS/BQ (storage), Cloud Datalab (dev), Cloud Dataflow (preprocessing)
- Supports online & batch predictions, prioritizing latency (online) & job time (batch)
- Or download models & make predictions anywhere: desktop, mobile, own servers
- HyperTune automatically tunes model hyperparameters to avoid manual tweaking
- Training: Pay per hour depending on chosen cluster capabilities (ML training units)
- Prediction: Pay per provisioned node-hour plus by prediction request volume made

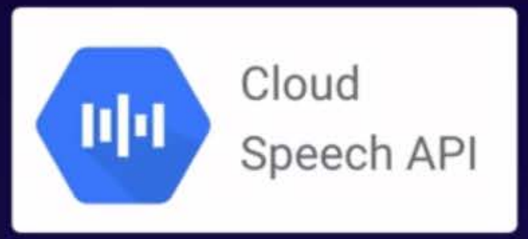
Cloud Vision API



Zonal Regional Multi-Regional Global

- Classifies images into categories, detects objects/faces, & finds/reads printed text
- Pre-trained ML model to analyze images and discover their contents
- Classifies into thousands of categories (e.g., "sailboat", "lion", "Eiffel Tower")
- Upload images or point to ones stored in GCS
- Pay per image, based on detection features requested
  - Higher price for OCR of full documents and finding similar images on the web
  - Some features are prices together: Labels + SafeSearch, ImgProps + Cropping
  - Other features priced individually: Text, Faces, Landmarks, Logos

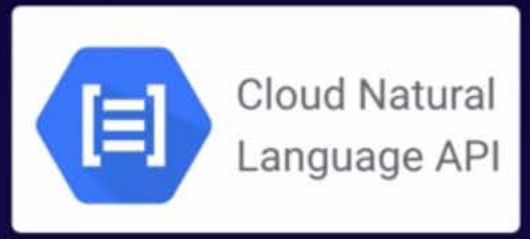
Cloud Speech API



Zonal Regional Multi-Regional Global

- Automatic Speech Recognition (ASR) to turn spoken word audio files into text
- Pre-trained ML model for recognizing speech in 110+ languages/variants
- Accepts pre-recorded or real-time audio, & can stream results back in real-time
- Enables voice command-and-control and transcribing user microphone dictations
- Handles noisy source audio
- Optionally filters inappropriate content in some languages
- Accepts contextual hints: words and names that will likely be spoken
- Pay per 15 seconds of audio processed

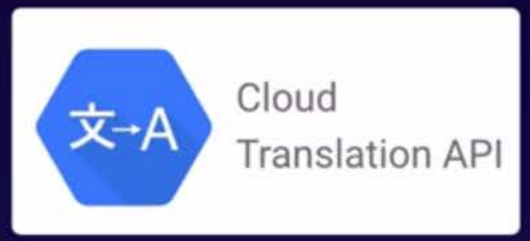
Cloud Natural Language API



Zonal Regional Multi-Regional Global

- Analyzes text for sentiment, intent, & content classification, and extracts info
- Pre-trained ML model for understanding what text means, so you can act on it
- Excellent with Speech API (audio), Vision API (OCR), & Translation API (or built-ins)
- Syntax analysis extracts tokens/sentences, parts of speech & dependency trees
- Entity analysis finds people, places, things, etc., labels them, & links to Wikipedia
- Analysis for sentiment (overall) and entity sentiment detect +/- feelings & strength
- Content classification puts each document into one of 700+ predefined categories
- Charged per request of 1000 characters, depending on analysis types requested

Cloud Translation API



Zonal Regional Multi-Regional Global

- Translate text among 100+ languages; optionally auto-detects source language
- Pre-trained ML model for recognizing and translating semantics, not just syntax
- Can let people support multi-regional clients in non-native languages, 2-way
- Combine with Speech, Vision, & Natural Language APIs for powerful workflows
- Send plain text or HTML and receive translation in kind
- Pay per character processed for translation
- Also pay per character for language auto-detection

Dialogflow

Zonal Regional Multi-Regional Global

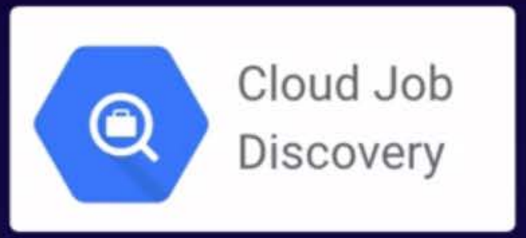
- Build conversational interfaces for websites, mobile apps, messaging, IoT devices
- Pre-trained ML model and service for accepting, parsing, lexing input & responding
- Enables useful chatbots and other natural user interactions with your custom code
- Train it to identify custom entity types by providing a small dataset of examples
- Or choose from 30+ pre-built agents (e.g. car, currency, dates) as starting template
- Supports many different languages and platforms/devices
- Free plan has unlimited text interactions and capped voice interactions
- Paid plan is unlimited but charges per request: more for voice, less for text

Cloud Video Intelligence API

Zonal Regional Multi-Regional Global

- Annotates videos in GCS (or directly uploaded) with info about what they contain
- Pre-trained ML model for video scene analysis and subject identification
- Enables you to search a video catalog the same way you search text documents
- “Specify a region where processing will take place (for regulatory compliance)”
- *Label Detection*: Detect entities within the video, such as "dog", "flower" or "car"
- *Shot Change Detection*: Detect scene changes within the video
- *SafeSearch Detection*: Detect adult content within the video
- Pay per minute of video processed, depending on requested detection modes

Cloud Job Discovery



Zonal Regional Multi-Regional Global

- Helps career sites, company job boards, etc. to improve engagement & conversion
- Pre-trained ML model to help job seekers search job posting databases
- `Most job sites rely on keyword search to retrieve content which often omits relevant jobs and overwhelms the job seeker with irrelevant jobs. For example, a keyword search with any spelling error returns 0 results, and a keyword search for “dental assistant” returns any "assistant" role that offers dental benefits.’
- Integrates with many job/hiring systems
- Lots of features, such as commute distance and recognizing abbreviations/jargon
- “Show me jobs with a 30 minute commute on public transportation from my home”

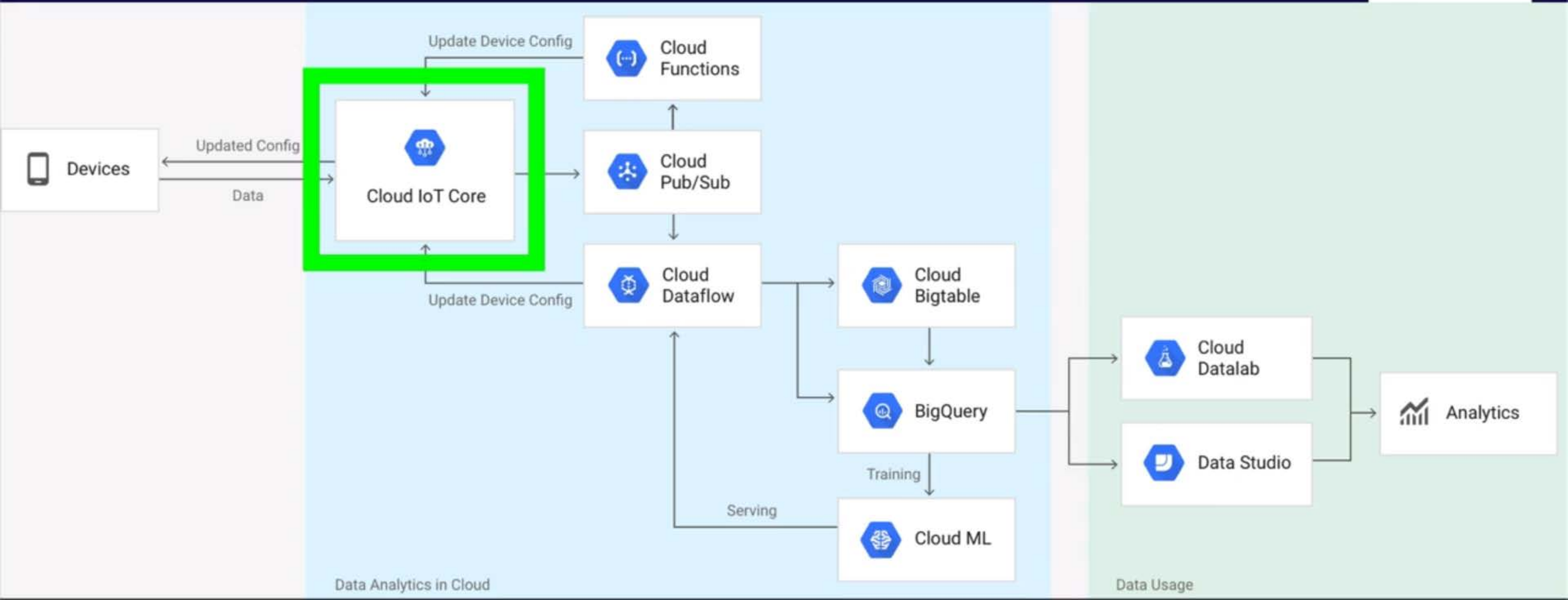
BIG DATA & IOT

Cloud Internet of Things (IoT) Core

A CLOUD GURU

Cloud IoT Core

Zonal Regional Multi-Regional Global

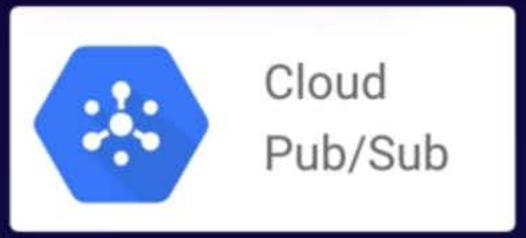


Cloud Internet of Things (IoT) Core

Zonal Regional Multi-Regional Global

- Fully-managed service to connect, manage, and ingest data from devices globally
- Device Manager handles device identity, authentication, config, & control
- Protocol Bridge publishes incoming telemetry to Cloud Pub/Sub for processing
- Connect securely using IoT industry-standard MQTT or HTTPS protocols
- CA signed certificates can be used to verify device ownership on first connect
- Two-way device communication enables configuration & firmware updates
- Device shadows enable querying & making control changes while devices offline
- Pay per MB of data exchanged with devices; no per-device charge

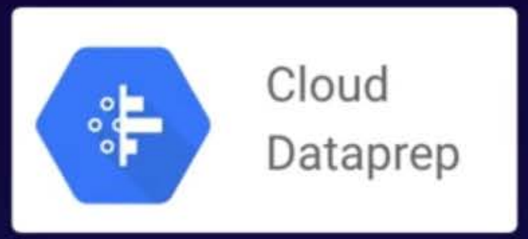
Cloud Pub/Sub



Zonal Regional Multi-Regional Global

- Infinitely-scalable at-least-once messaging for ingestion, decoupling, etc.
- "Global by default: Publish... and consume from anywhere, with consistent latency."
- Messages can be up to 10MB and undelivered ones stored for 7 days—but no DLQ
- Push mode delivers to HTTPS endpoint & succeeds on HTTP success status code
  - "Slow-start" algorithm ramps up on success and backs off & retries, on failures
- Pull mode delivers messages to requesting clients and waits for ACK to delete
  - Lets clients set rate of consumption, and supports batching and long-polling
- Pay for data volume
  - Min 1KB per publish/push/pull request (not by message)

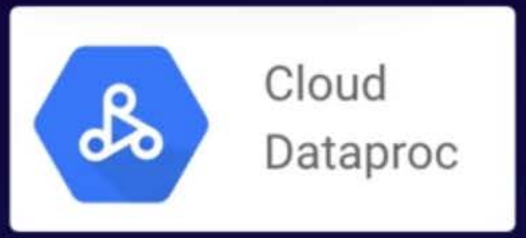
Cloud Dataprep



Zonal Regional Multi-Regional Global

- Visually explore, clean, and prepare data for analysis without running servers
- "Data Wrangling" (i.e. "ad-hoc ETL") for business analysts, not IT pros
  - Who might otherwise spend 80% of their time cleaning data
- Managed version of Trifacta Wrangler—and managed by Trifacta, not Google
- Source data from GCS, BQ, or file upload—formatted in CSV, JSON, or relational
- Automatically detects schemas, datatypes, possible joins, and various anomalies
- Pay for underlying Dataflow job, plus management overhead charge
- Pay for other accessed services (e.g. GCS, BQ)

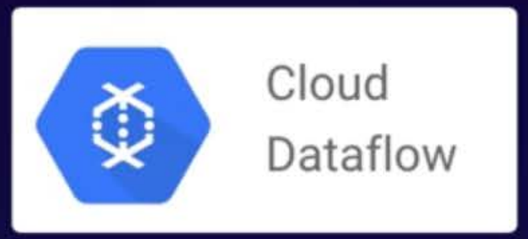
Cloud Dataproc



Zonal Regional Multi-Regional Global

- Batch MapReduce processing via configurable, managed Spark & Hadoop clusters
- Handles being told to scale (adding or removing nodes) even while running jobs
- Integrated with Cloud Storage, BigQuery, Bigtable, and some Stackdriver services
- "Image versioning" switches between versions of Spark, Hadoop, & other tools
- Pay directly for underlying GCE servers used in the cluster—optionally preemptible
- Pay a Cloud Dataproc management fee per vCPU-hour in the cluster
- Best for moving *existing* Spark/Hadoop setups to GCP
    - Prefer Cloud Dataflow for new data processing pipelines — "Go with the flow"

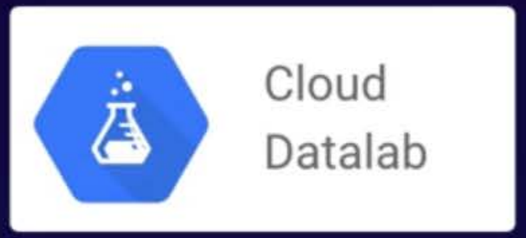
Cloud Dataflow



Zonal Regional Multi-Regional Global

- Smartly-autoscaled & fully-managed batch or stream MapReduce-like processing
- Released as open-source Apache Beam
- Autoscales & dynamically redistributes lagging work, mid-job, to optimize run time
- Integrated with Cloud Pub/Sub, Datastore, BQ, Bigtable, Cloud ML, Stackdriver, etc.
- Dataflow Shuffle service for batch offloads Shuffle ops from workers for big gains
- Effectively pay for underlying worker GCE via consolidated charges
  - Pay per second for vCPUs, RAM GBs PD/PD-SSD (more for streaming)
- Dataflow Shuffle charged for time per GB used

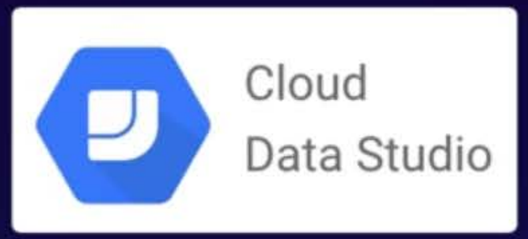
Cloud Datalab



Zonal Regional Multi-Regional Global

- Interactive tool for data exploration, analysis, visualization and machine learning
- Uses Jupyter Notebook
  - "[A]n open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more."
- Supports iterative development of data analysis algorithms in Python/SQL/~JS
- Pay for GCE/GAE instance hosting and storing (on PD) your notebooks
- Pay for any other resources accessed (e.g. BigQuery)

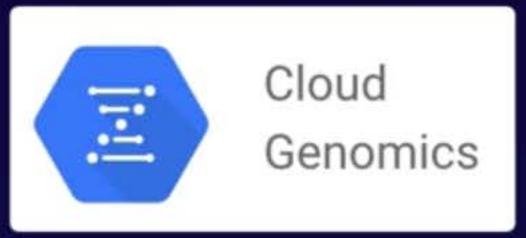
Cloud Data Studio



Zonal Regional Multi-Regional Global

- Big Data Visualization tool for dashboards and reporting
- Meaningful data stories/presentations enable better business decision making
- Data sources include BigQuery, Cloud SQL, other MySQL, Google Sheets, Google Analytics, Analytics 360, AdWords, DoubleClick, & YouTube channels
- Visualizations include time series, bar charts, pie charts, tables, heat maps, geo maps, scorecards, scatter charts, bullet charts, & area charts
- Templates for quick start; customization options for impactful finish
- Familiar G Suite sharing and real-time collaboration
- Pay only for services accessed

Cloud Genomics



Zonal Regional Multi-Regional Global

- Store and process genomes and related experiments
- Query complete genomic information of large research projects in seconds
- Process many genomes and experiments in parallel
- Open industry standards (e.g. from Global Alliance for Genomics and Health)
- Supports “Requester Pays” sharing

Big Data Lifecycle

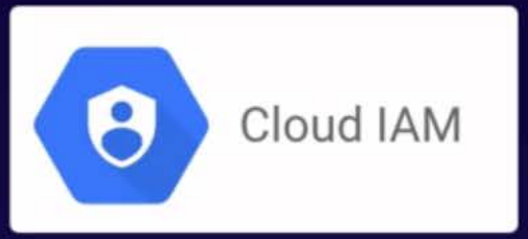
Ingest	Store	Process & Analyze	Explore & Visualize
App Engine	Cloud Storage	Cloud Dataflow	Cloud Datalab
Compute Engine	Cloud SQL	Cloud Dataproc	Google Data Studio
Container Engine	Cloud Datastore	BigQuery	Google Sheets
Cloud Pub/Sub	Cloud Bigtable	Cloud ML	
Stackdriver Logging	BigQuery	Cloud Vision API	
Cloud Transfer Service		Cloud Speech API	
		Translate API	
		Cloud Natural Lang API	

Roles

Zonal Regional Multi-Regional Global

- Roles are collections of Permissions to use or manage GCP resources
- Permissions allow you to perform certain actions: `Service.Resource.Verb`
- Primitive Roles: Owner, Editor, Viewer
  - Viewer is read-only; Editor can change things; Owner can control access & billing
  - Pre-date IAM service, may still be useful (e.g. dev/test envs), but often too broad
- Predefined Roles: Give granular access to specific GCP resources (IAM)
  - E.g.: `roles/bigquery.dataEditor`, `roles/pubsub.subscriber`
- Custom Roles: Project- or Org-level collections you define of granular permissions

Cloud Identity and Access Management (IAM)



Zonal Regional Multi-Regional Global

- Controls access to GCP resources: authorization, not really authentication/identity
- Member is user, group, domain, service account, or the public (e.g. “allUsers”)
  - Individual Google account, Google group (👍), G Suite / Cloud Identity domain
  - Service account (👍) belongs to application/instance, not individual end user
  - Every identity has a unique e-mail address, including service accounts
- Policies bind Members to Roles at a hierarchy level: Org, Folder, Project, Resource
  - Answer: *Who* can do *what* to *which thing(s)*?

IDENTITY & ACCESS

Cloud Iden

A CLOUD GURU

Cl

- Contr
- Memb
  - Inc
  - Se
  - Ev
- Polici
  - An

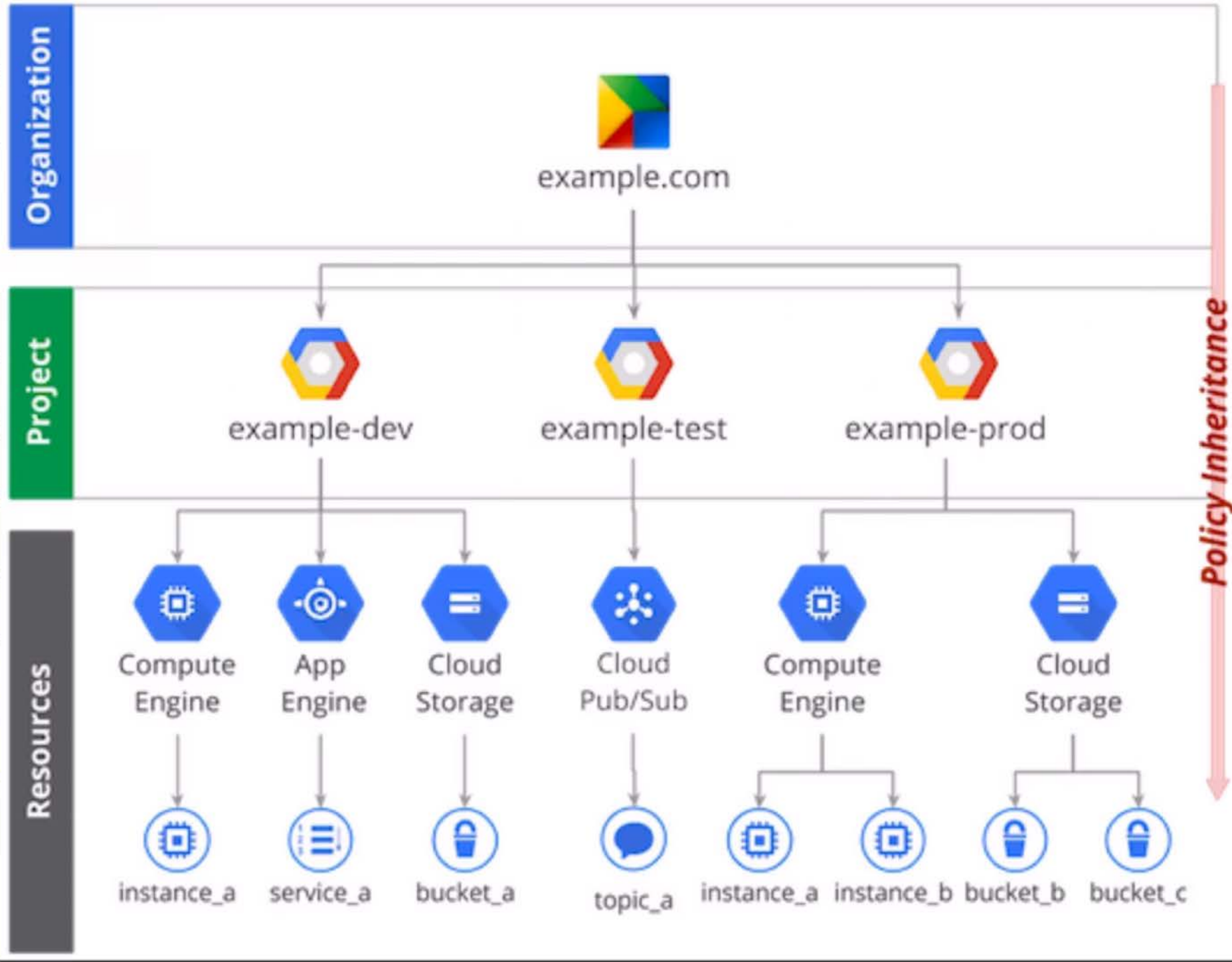
n/identity

ers")

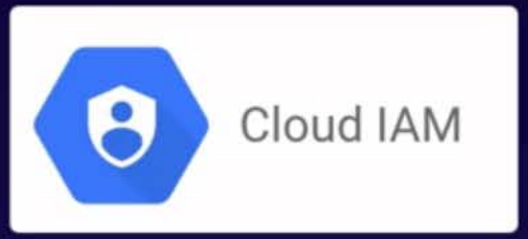
lomain

d user

esource



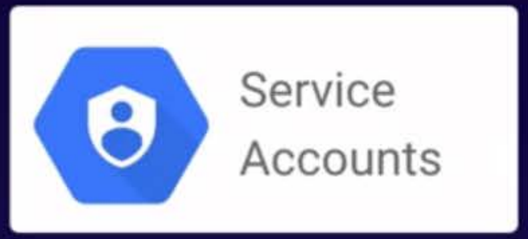
Cloud Identity and Access Management (IAM)



Zonal Regional Multi-Regional Global

- Controls access to GCP resources: authorization, not really authentication/identity
- Member is user, group, domain, service account, or the public (e.g. “allUsers”)
  - Individual Google account, Google group (👍), G Suite / Cloud Identity domain
  - Service account (👍) belongs to application/instance, not individual end user
  - Every identity has a unique e-mail address, including service accounts
- Policies bind Members to Roles at a hierarchy level: Org, Folder, Project, Resource
  - Answer: *Who* can do *what* to *which thing(s)*?
- IAM is free; pay for authorized GCP service usage

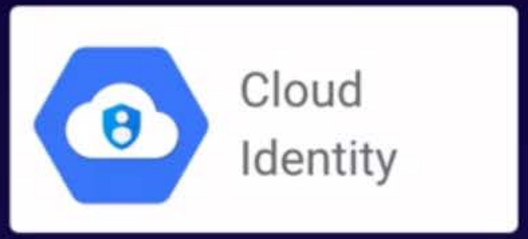
Service Accounts



Zonal Regional Multi-Regional Global

- Special type of Google account that represents an application, not an end user
- Can be “assumed” by applications or individual users (when so authorized)
- “Important: For almost all cases, whether you are developing locally or in a production application, you should use service accounts, rather than user accounts or API keys.”
- Consider resources and permissions required by application; use least privilege
- Can generate and download private keys (user-managed keys), for non-GCP, but...
- Cloud-Platform-managed keys (👍) are better, for GCP (i.e. GCF, GAE, GCE, and GKE)
  - No direct downloading: Google manages private keys & rotates them once a day

Cloud Identity



Zonal Regional Multi-Regional Global

- Identity as a Service (IDaaS, not DaaS) to provision and manage users and groups
- Free Google Accounts for non-G-Suite users, tied to a verified domain
- Centrally manage all users in Google Admin console; supports compliance
- 2-Step verification (2SV/MFA) and enforcement (👍), including security keys
- Sync from Active Directory and LDAP directories via Google Cloud Directory Sync
- Identities work with other Google services (e.g. Chrome)
- Identities can be used to SSO with other apps via OIDC, SAML, OAuth2
- Cloud Identity is free; pay for authorized GCP service usage

Security Key Enforcement

Zonal Regional Multi-Regional Global

- USB or Bluetooth 2-step verification device that prevents phishing
- Not like just getting a code via email or text message...
- Device also verifies the target service
- Eliminates man-in-the-middle (MITM) attacks against GCP credentials

IDENTITY & ACCESS

Cloud Resource

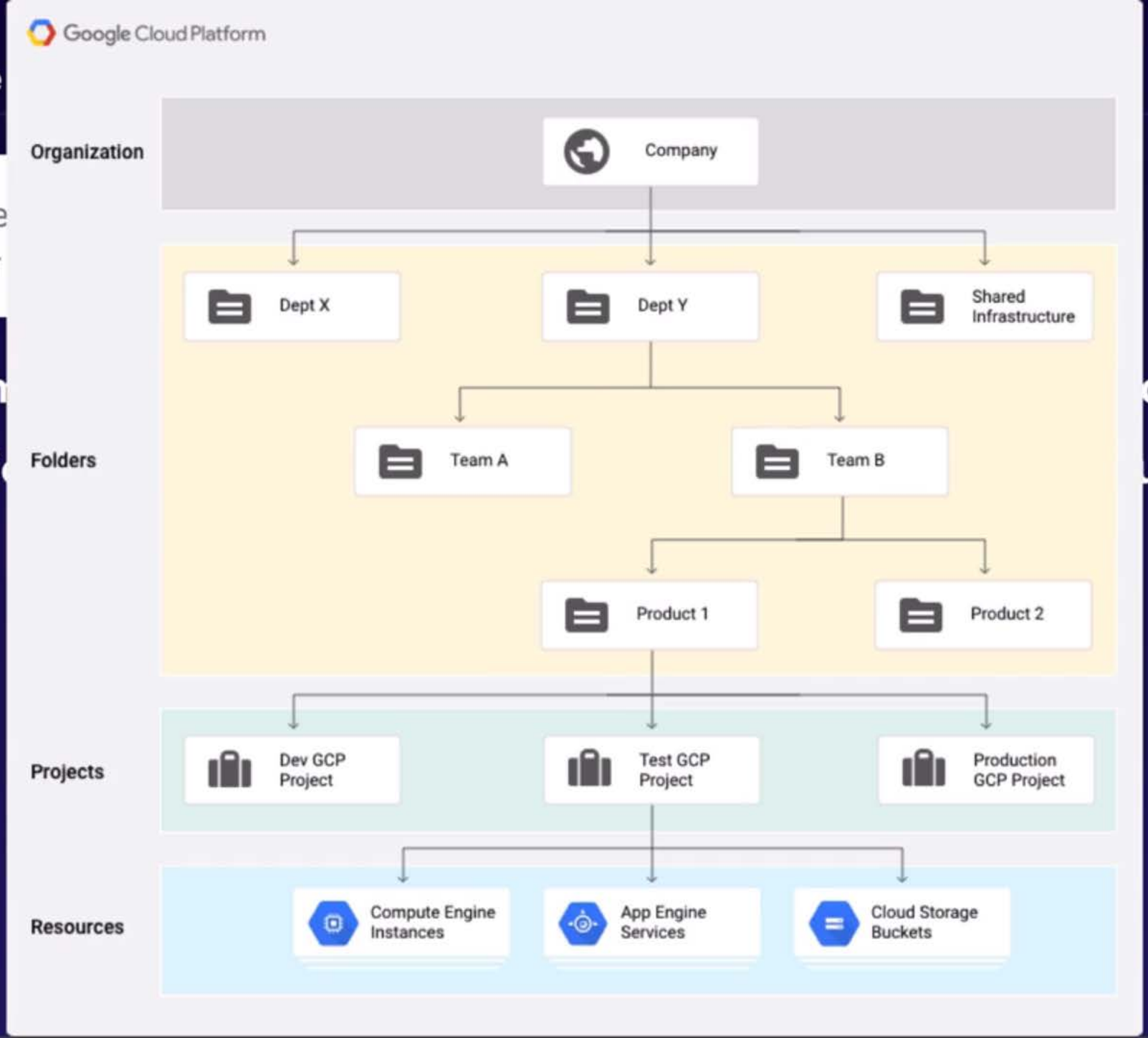
A CLOUD GURU

Resource Manager

- Centrally m
- Organizati

der hierarchy

usiness needs



Cloud Resource Manager

Zonal Regional Multi-Regional Global

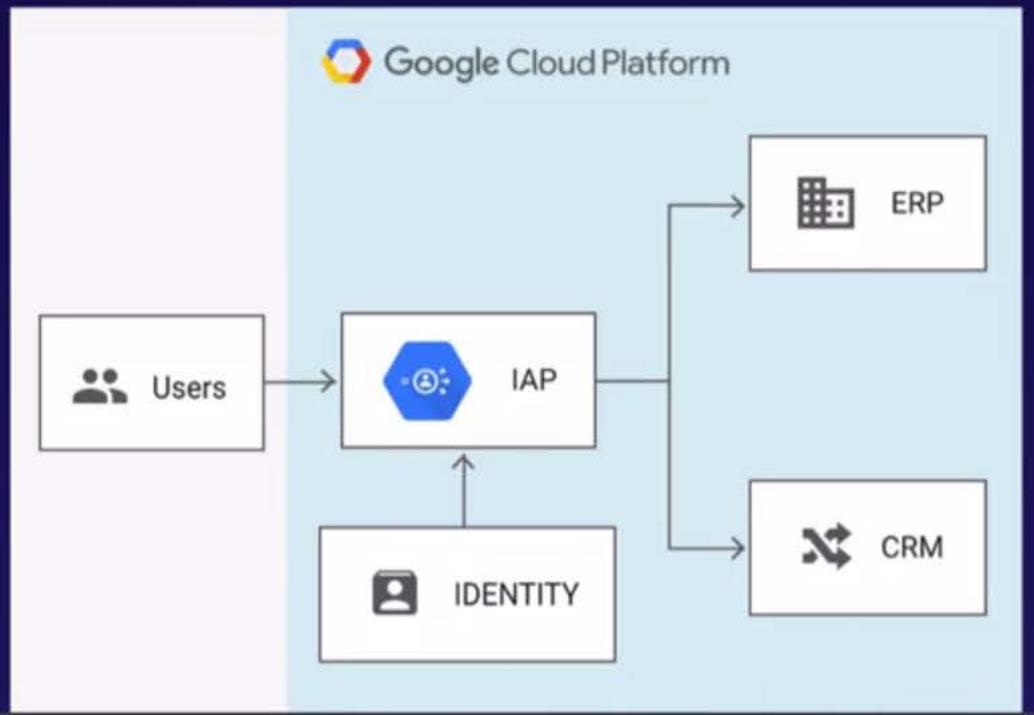
- Centrally manage & secure organization's projects with custom folder hierarchy
- Organization resource is root node in hierarchy; folders per your business needs
- Tied 1:1 to a Cloud Identity / G Suite domain, then owns all newly-created projects
  - Without this organization, specific identities (people) must own GCP projects
- "Recycle bin" allows undeleting projects
- Define custom IAM roles at org level
- Apply IAM policies at organization, folder, or project levels
- No charge for this service

Cloud Identity-Aware Proxy (IAP)

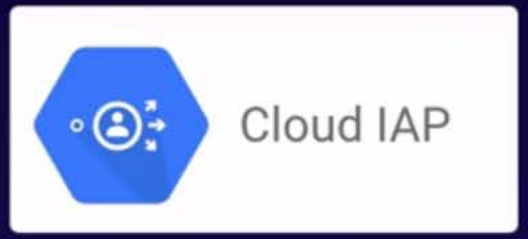
Cloud IAP

Zonal Regional Multi-Regional Global

- Guards apps running on GCP via identity verification, not VPN access
- Based on CLB & IAM, and only passes authed requests through



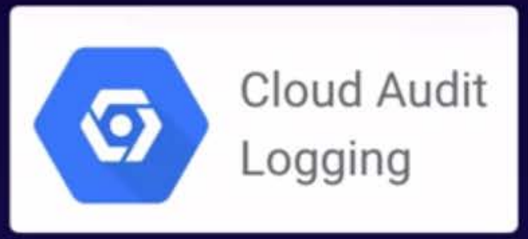
Cloud Identity-Aware Proxy (IAP)



Zonal Regional Multi-Regional Global

- Guards apps running on GCP via identity verification, not VPN access
- Based on CLB & IAM, and only passes authed requests through
- Grant access to any IAM identites, incl. groups & service accounts
- Relatively straightforward to set up
- Pay for load balancing / protocol forwarding rules and traffic

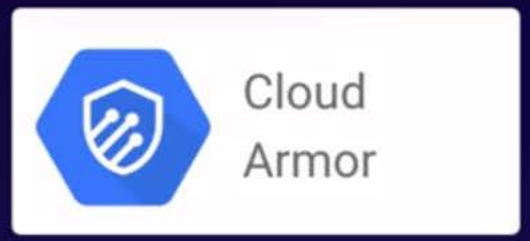
Cloud Audit Logging



Zonal Regional Multi-Regional Global

- Answers the questions "Who did what, where, and when?" within GCP projects
- Maintains non-tamperable audit logs for each project and organization:
  - Admin Activity and System Events (400 day retention)
  - Access Transparency (400 day retention)
    - Shows actions by Google support staff
  - Data Access (30 day retention)
    - For GCP-visible services (e.g. Can't see into MySQL DB on GCE)
- Data Access logs priced through Stackdriver Logging; rest are free

Cloud Armor



Zonal Regional Multi-Regional Global

- Edge-level protection from DDoS & other attacks on global HTTP(S) LB
- Offload work: Blocked attacks never reach your systems
- Monitor: Detailed request-level logs available in Stackdriver Logging
- Manage IPs with CIDR-based allow/block lists (aka whitelist/blacklist)
- More intelligent rules forthcoming (e.g. XSS, SQLi, geo-based, custom)
- Preview effect of changes before making them live
- Pay per policy and rule configured, plus for incoming request volume

Cloud Security Scanner

Zonal Regional Multi-Regional Global

- Free but limited GAE app vulnerability scanner with “very low false positive rates”
- “After you set up a scan, Cloud Security Scanner automatically crawls your application, following all links within the scope of your starting URLs, and attempts to exercise as many user inputs and event handlers as possible.”
- Can identify:
  - Cross-site-scripting (XSS)
  - Flash injection
  - Mixed content (HTTP in HTTPS)
  - Outdated/insecure libraries

Cloud Data Loss Prevention API (DLP)

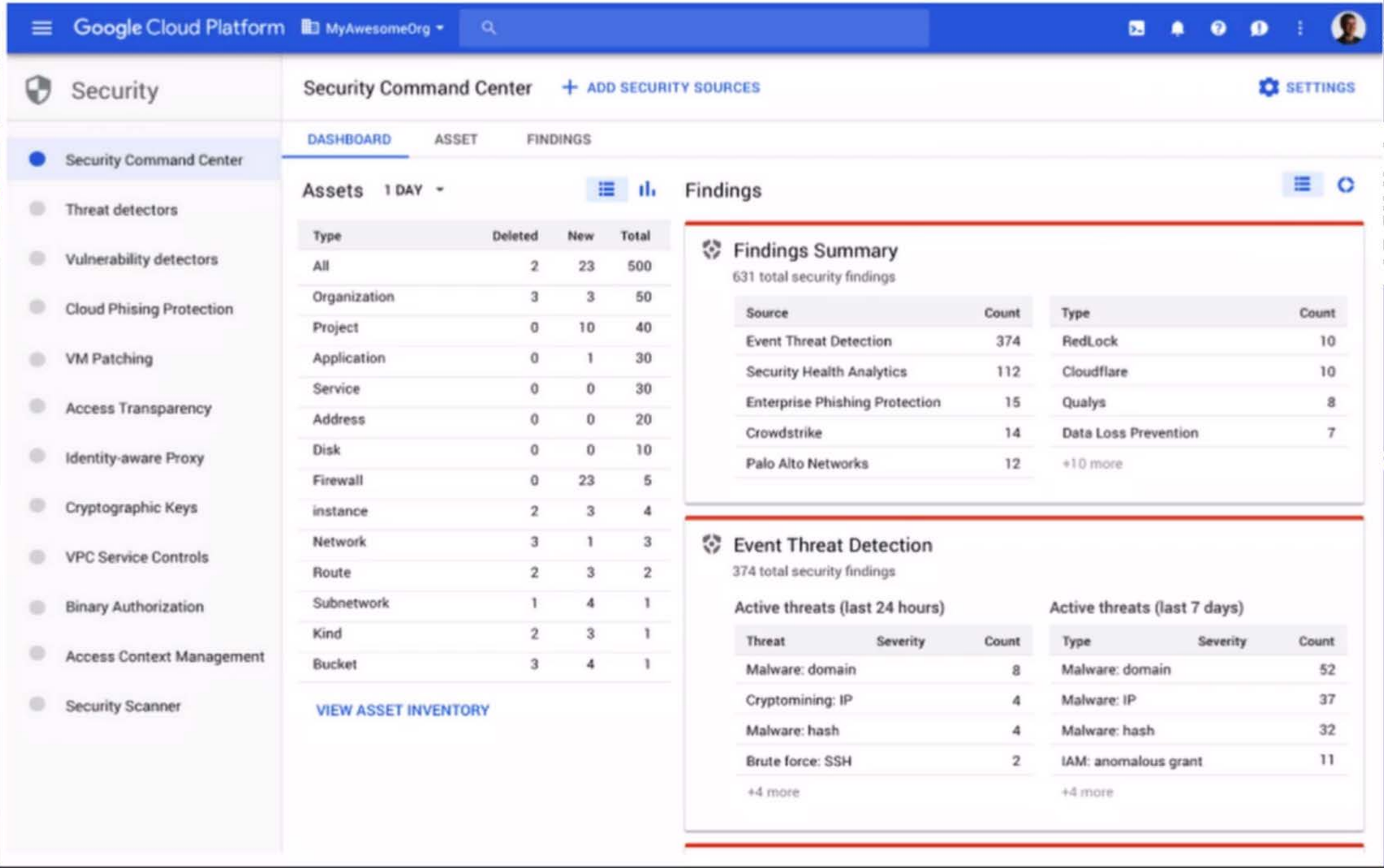
Zonal Regional Multi-Regional Global

- Finds and optionally redacts sensitive info in unstructured data streams
- Helps you minimize what you collect, expose, or copy to other systems
- 50+ sensitive data detectors, including: credit card numbers, names, social security numbers, passport numbers, driver's license numbers (US and some other jurisdictions), phone numbers, and other personally identifiable information (PII)
- Data can be sent directly, or API can be pointed at GCS, BQ, or Cloud DataStore
- Can scan both text and images
- Pay for amount of data processed (per GB)—and gets cheaper when large volume
    - Pricing for storage now very simple (June 2019), but for streaming is still a mess

Event Threat Detection (ETD)

Zonal Regional Multi-Regional Global

- Automatically scans your Stackdriver logs for suspicious activity
- Uses industry-leading threat intelligence, including Google Safe Browsing
- Quickly detects many possible threats, including:
  - Malware, cryptomining, outgoing DDoS attacks, port scanning, brute-force SSH
  - Also: Unauthorized access to GCP resources via abusive IAM access
- Can export parsed logs to BigQuery for forensic analysis
- Integrates with SIEMs like Google's Cloud SCC or via Cloud Pub/Sub
- No charge for ETD, but charged for its usage of other GCP services (like SD Logging)

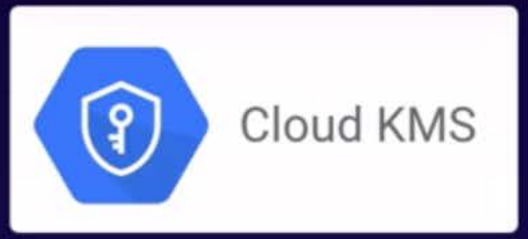

Google Cloud Platform
MyAwesomeOrg
Security
Security Command Center
ADD SECURITY SOURCES
SETTINGS
Security Command Center
Threat detectors
Vulnerability detectors
Cloud Phising Protection
VM Patching
Access Transparency
Identity-aware Proxy
Cryptographic Keys
VPC Service Controls
Binary Authorization
Access Context Management
Security Scanner
DASHBOARD
ASSET
FINDINGS
Assets
1 DAY
Type Deleted New Total
All 2 23 500
Organization 3 3 50
Project 0 10 40
Application 0 1 30
Service 0 0 30
Address 0 0 20
Disk 0 0 10
Firewall 0 23 5
instance 2 3 4
Network 3 1 3
Route 2 3 2
Subnetwork 1 4 1
Kind 2 3 1
Bucket 3 4 1
VIEW ASSET INVENTORY
Findings
Findings Summary
631 total security findings
Source Count
Event Threat Detection 374
Security Health Analytics 112
Enterprise Phishing Protection 15
Crowdstrike 14
Palo Alto Networks 12
Type Count
RedLock 10
Cloudflare 10
Qualys 8
Data Loss Prevention 7
+10 more
Event Threat Detection
374 total security findings
Active threats (last 24 hours)
Threat Severity Count
Malware: domain 8
Cryptomining: IP 4
Malware: hash 4
Brute force: SSH 2
+4 more
Active threats (last 7 days)
Type Severity Count
Malware: domain 52
Malware: IP 37
Malware: hash 32
IAM: anomalous grant 11
+4 more

Cloud Security Command Center (SCC)

Zonal Regional Multi-Regional Global

- "Comprehensive security management and data risk platform for GCP"
- Security Information and Event Management (SIEM) software
- "Helps you prevent, detect, & respond to threats from a single pane of glass"
- Use "Security Marks" (aka "marks") to group, track, and manage resources
- Integrate ETD, Cloud Scanner, DLP, & many external security finding sources
- Can alert to humans & systems; Can export data to external SIEM
- Free! But charged for services used (e.g. DLP API, if configured)
- Could also be charged for excessive uploads of external findings

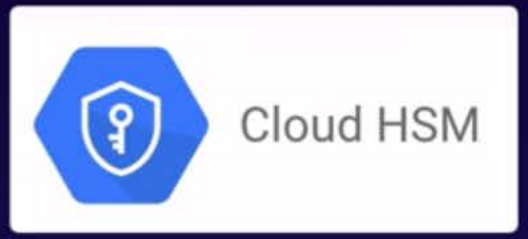
Cloud Key Management Service (KMS)



Zonal Regional Multi-Regional Global

- Low-latency service to manage and use cryptographic keys
- Supports symmetric (e.g. AES) and asymmetric (e.g. RSA, EC) algorithms
- Move secrets out of code (and the like) and into the environment, in a secure way
- Integrated with IAM & Cloud Audit Logging to authorize & track key usage
- Rotate keys used for new encryption either automatically or on demand
  - Still keeps old active key versions, to allow decrypting
- Key deletion has 24 hour delay, "to prevent accidental or malicious data loss”
- Pay for active key versions stored over time
- Pay for key use operations (i.e. encrypt/decrypt; admin operations are free)

Cloud Hardware Security Module (HSM)



Zonal Regional Multi-Regional Global

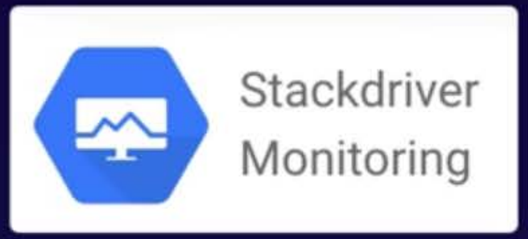
- Cloud KMS keys managed by FIPS 140-2 Level 3 certified HSMs
- Device hosts encryption keys and performs cryptographic operations
- Enables you to meet compliance that mandates hardware environment
- Fully integrated with Cloud KMS
  - Same API, features, IAM integration
- Priced like Cloud KMS: Active key versions stored & key operations
  - But some key types more expensive: RSA, EC, Long AES

Google Stackdriver

Zonal Regional Multi-Regional Global

- Family of services for monitoring, logging, & diagnosing apps on GCP/AWS/hybrid
- Service integrations add lots of value—among Stackdriver and with GCP
- One Stackdriver account can track multiple:
  - GCP projects
  - AWS accounts
  - Other resources
- Simple usage-based pricing
  - No longer previous system of tiers, allotments, and overages

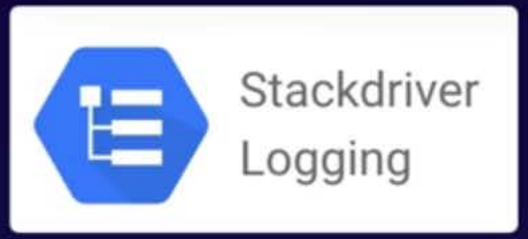
Stackdriver Monitoring



Zonal Regional Multi-Regional Global

- Gives visibility into perf, uptime, & overall health of cloud apps (based on collectd)
- Includes built-in/custom metrics, dashboards, global uptime monitoring, & alerts
- Follow the trail: Links from alerts to dashboards/charts to logs to traces
- Cross-cloud: GCP, of course, but monitoring agent also supports AWS
- Alerting policy config includes multi-condition rules & resource organization
- Alert via email, GCP Mobile App, SMS, Slack, PagerDuty, AWS SNS, webhook, etc.
- Automatic GCP/Anthos metrics always free
- Pay for API calls & per MB for custom or AWS metrics

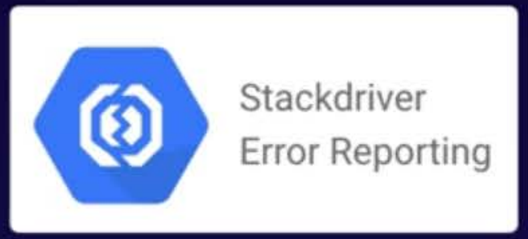
Stackdriver Logging



Zonal Regional Multi-Regional Global

- Store, search, analyze, monitor, and alert on log data & events (based on Fluentd)
- Collection built into some GCP, AWS support with agent, or custom send via API
- Debug issues via integration with Stackdriver Monitoring, Trace & Error Reporting
- Create real-time metrics from log data, then alert or chart them on dashboards
- Send real-time log data to BigQuery for advanced analytics and SQL-like querying
- Powerful interface to browse, search, and slice log data
- Export log data to GCS to cost-effectively store log archives
- Pay per GB ingested & stored for one month, but first 50GB/project free
- Cloud Audit Logging / Access Transparency logs also free

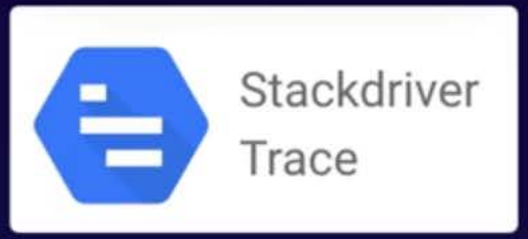
Stackdriver Error Reporting



Zonal Regional Multi-Regional Global

- Counts, analyzes, aggregates, & tracks crashes in helpful centralized interface
- Smartly aggregates errors into meaningful groups tailored to language/framework
- Instantly alerts when a new app error cannot be grouped with existing ones
- Link directly from notifications to error details:
  - Time chart, occurrences, affected user count, first/last seen dates, cleaned stack
- Exception stack trace parser knows Java, Python, JavaScript, Ruby, C#, PHP, & Go
- Jump from stack frames to source to start debugging
- No direct charge; pay for source data in Stackdriver Logging

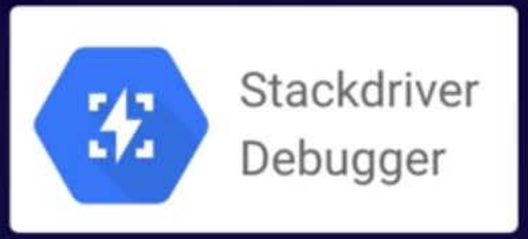
Stackdriver Trace



Zonal Regional Multi-Regional Global

- Tracks and displays call tree & timings across distributed systems, to debug perf
- Automatically captures traces from Google App Engine
- Trace API and SDKs for Java, Node.js, Ruby, and Go capture traces from anywhere
- Zipkin collector allows Zipkin tracers to submit data to Stackdriver Trace
- View aggregate app latency info or dig into individual traces to debug problems
- Generate reports on demand and get daily auto reports per traced app
- Detects app latency shift (degradation) over time by evaluating perf reports
- Pay for ingesting and retrieving trace spans

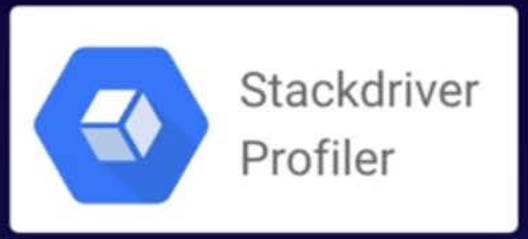
Stackdriver Debugger



Zonal Regional Multi-Regional Global

- Grabs program state (callstack, variables, expressions) in live deploys; low impact
- Logpoints repeat for up to 24h; fuller snapshots run once but can be conditional
- Source view supports Cloud Source Repository, Github, Bitbucket, local, & upload
- Java and Python supported on GCE, GKE, and GAE (Standard and Flex)
- Node.js and Ruby supported on GCE, GKE, and GAE Flex; Go only on GCE and GKE
- Automatically enabled for Google App Engine apps; agents available for others
- Share debugging sessions with others (just send URL)
- Free to use

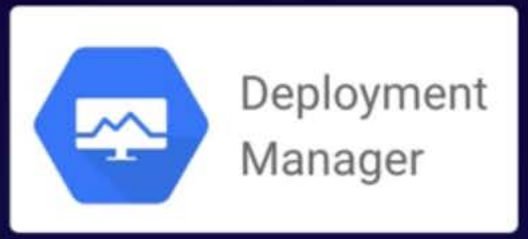
Stackdriver Profiler



Zonal Regional Multi-Regional Global

- Continuous CPU and memory profiling to improve perf & reduce cost
- Low overhead (Typical: 0.5%; Max: 5%)—so use it in prod, too!
- Supports Go, Java, Node.js, and Python (3.2+)
- Agent-based
- Saves profiles for 30 days
- Can download profiles for longer-term storage
- Free to use

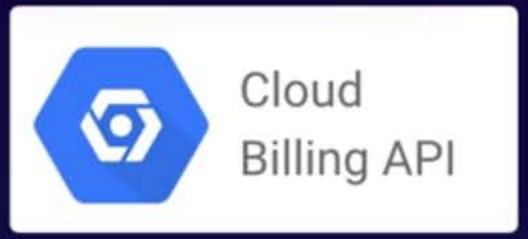
Cloud Deployment Manager



Zonal Regional Multi-Regional Global

- Create/manage resources via declarative templates: “Infrastructure as Code”
- Declarative allows automatic parallelization
- Templates written in YAML, Python, or Jinja2
- Supports input and output parameters, with JSON schema
- Create and update of deployments both support preview
- Free service; just pay for resources involved in deployments

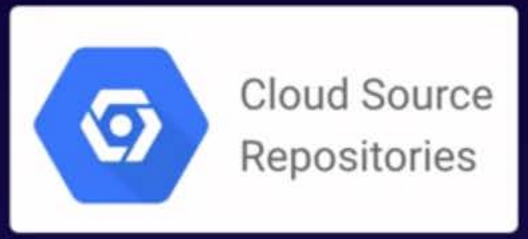
Cloud Billing API



Zonal Regional Multi-Regional Global

- Programmatically manage billing for GCP projects and get GCP pricing
- Billing config
  - List billing accounts; get details and associated projects for each
  - Enable (associate), disable (disassociate), or change project's billing account
- Pricing
  - List billable SKUs; get public pricing (including tiers) for each
  - Get SKU metadata like regional availability
- Export of current bill to GCS or BQ is possible—but configured via console, not API

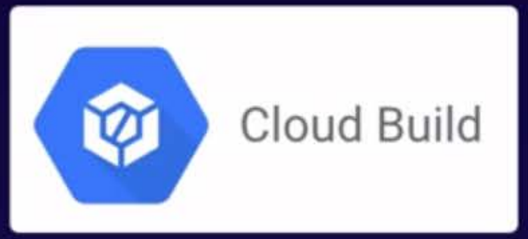
Cloud Source Repositories



Zonal Regional Multi-Regional Global

- Hosted private Git repositories, with integrations to GCP and other hosted repos
- Supports standard Git functionality
- No enhanced workflow support like pull requests
- Can set up automatic sync from GitHub or Bitbucket
- Natural integration with Stackdriver debugger for live-debugging deployed apps
- Pay per project-user active each month (not prorated)
- Pay per GB-month of data storage (prorated)
- Pay per GB of data egress

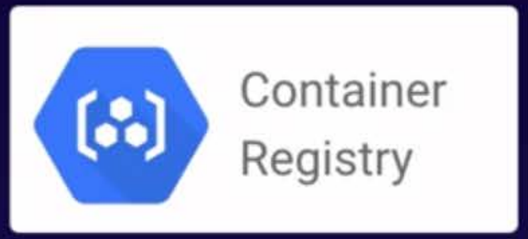
Cloud Build



Zonal Regional Multi-Regional Global

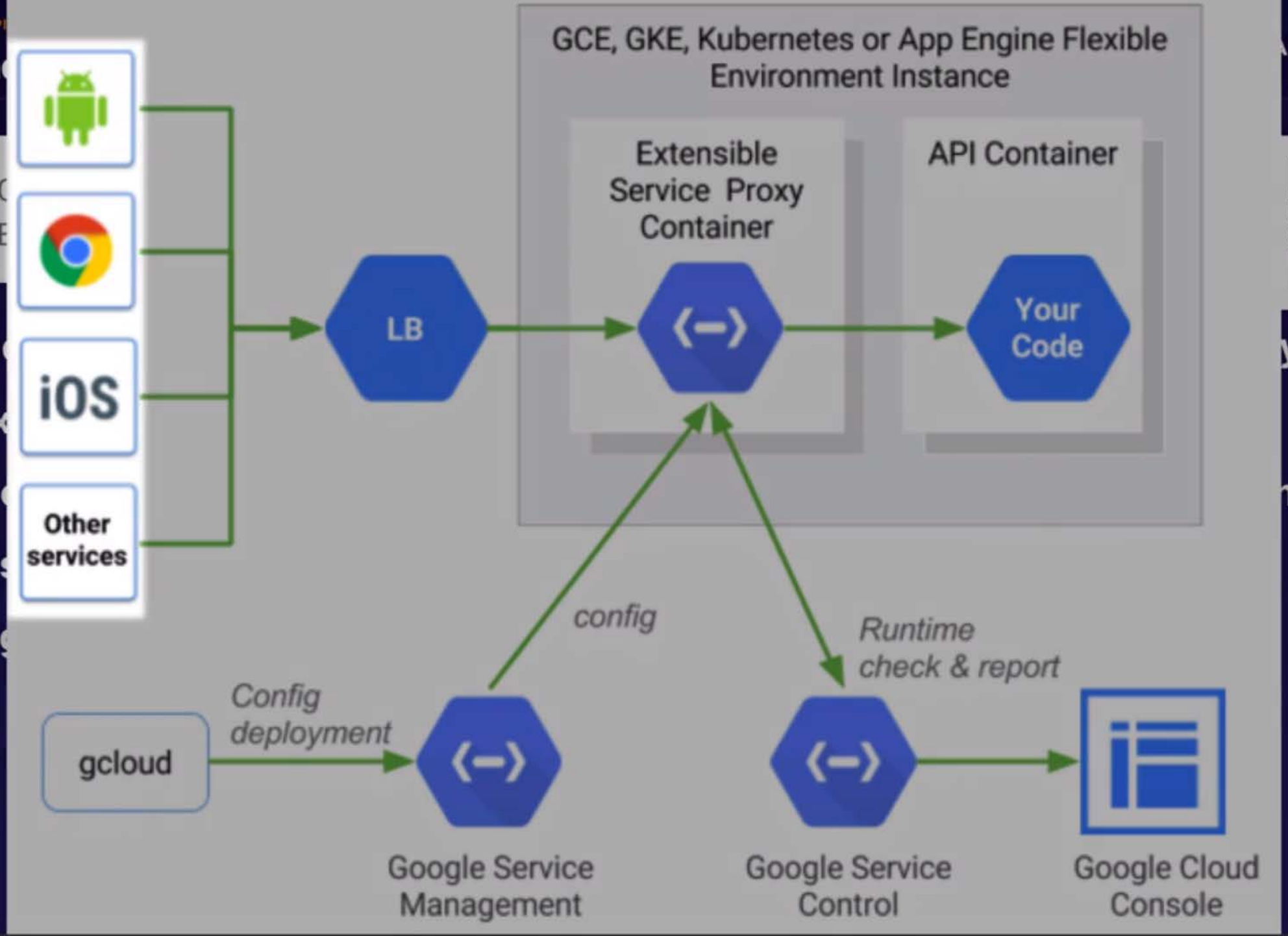
- Continuously takes source code and builds, tests, and deploys it — CI/CD service
- Trigger from Cloud Source Repository (by branch, tag, or commit) or zip in GCS
  - Can trigger from GitHub and Bitbucket via Cloud Source Repositories RepoSync
- Runs many builds in parallel (currently 10 at a time)
- Dockerfile: super-simple build+push—plus scans for package vulnerabilities
- JSON/YAML file: flexible & parallel steps
- Push to GCR & export artifacts to GCS—or anywhere your build steps write
- Maintains build logs and build history
- Pay per minute of build time—but free tier is 120 minutes per day

Container Registry (GCR)

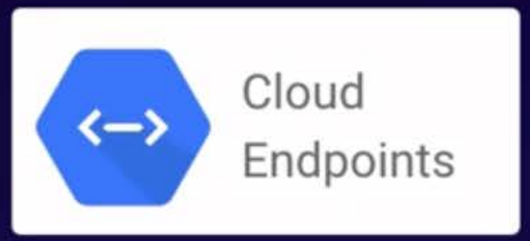


Zonal Regional Multi-Regional Global

- Fast, private Docker image storage (based on GCS) with Docker V2 Registry API
- Creates & manages a multi-regional GCS bucket, then translates GCR calls to GCS
- IAM integration simplifies builds and deployments within GCP
- Quick deploys because of GCP networking to GCS
- Directly compatible with standard Docker CLI; native Docker Login support
- UX integrated with Cloud Build & Stackdriver Logs
- UI to manage tags and search for images
- Pay directly for storage and egress of underlying GCS (no overhead)


GCE, GKE, Kubernetes or App Engine Flexible Environment Instance
Extensible Service Proxy Container
API Container
LB
Your Code
iOS
Other services
gcloud
Config deployment
config
Runtime check & report
Google Service Management
Google Service Control
Google Cloud Console

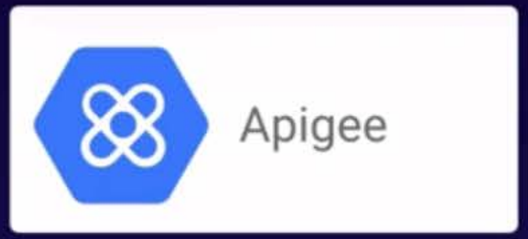
Cloud Endpoints



Zonal Regional Multi-Regional Global

- Handles authorization, monitoring, logging, & API keys for APIs backed by GCP
- Proxy instances are distributed and hook into Cloud Load Balancer
- Super-fast Extensible Service Proxy (ESP) container based on nginx: <1 ms / call
- Uses JWTs and integrates with Firebase, Auth0, & Google Auth
- Integrates with Stackdriver Logging and Stackdriver Trace
- Extensible Service Proxy (ESP) can transcode HTTP/JSON to gRPC
  - But API needs to be resource-oriented (i.e. RESTful)
- Pay per call to your API

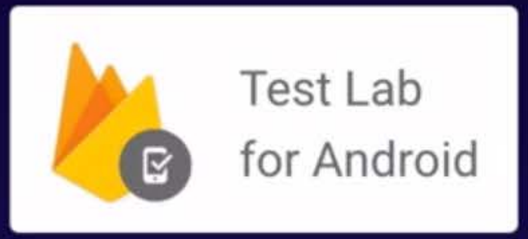
Apigee API Platform



Zonal Regional Multi-Regional Global

- Full-featured & enterprise-scale API management platform for whole API lifecycle
- Transform calls between different protocols: SOAP, REST, XML, binary, custom
- Authenticate via OAuth/SAML/LDAP; authorize via Role-Based Access Control
- Throttle traffic with quotas, manage API versions, etc.
- Apigee Sense identifies and alerts administrators to suspicious API behaviors
- Apigee API Monetization supports various revenue models / rate pans
- Team and Business tiers are flat monthly rate with API call quotas & feature sets
- “Enterprise” tier and special feature pricing are “Contact Sales”

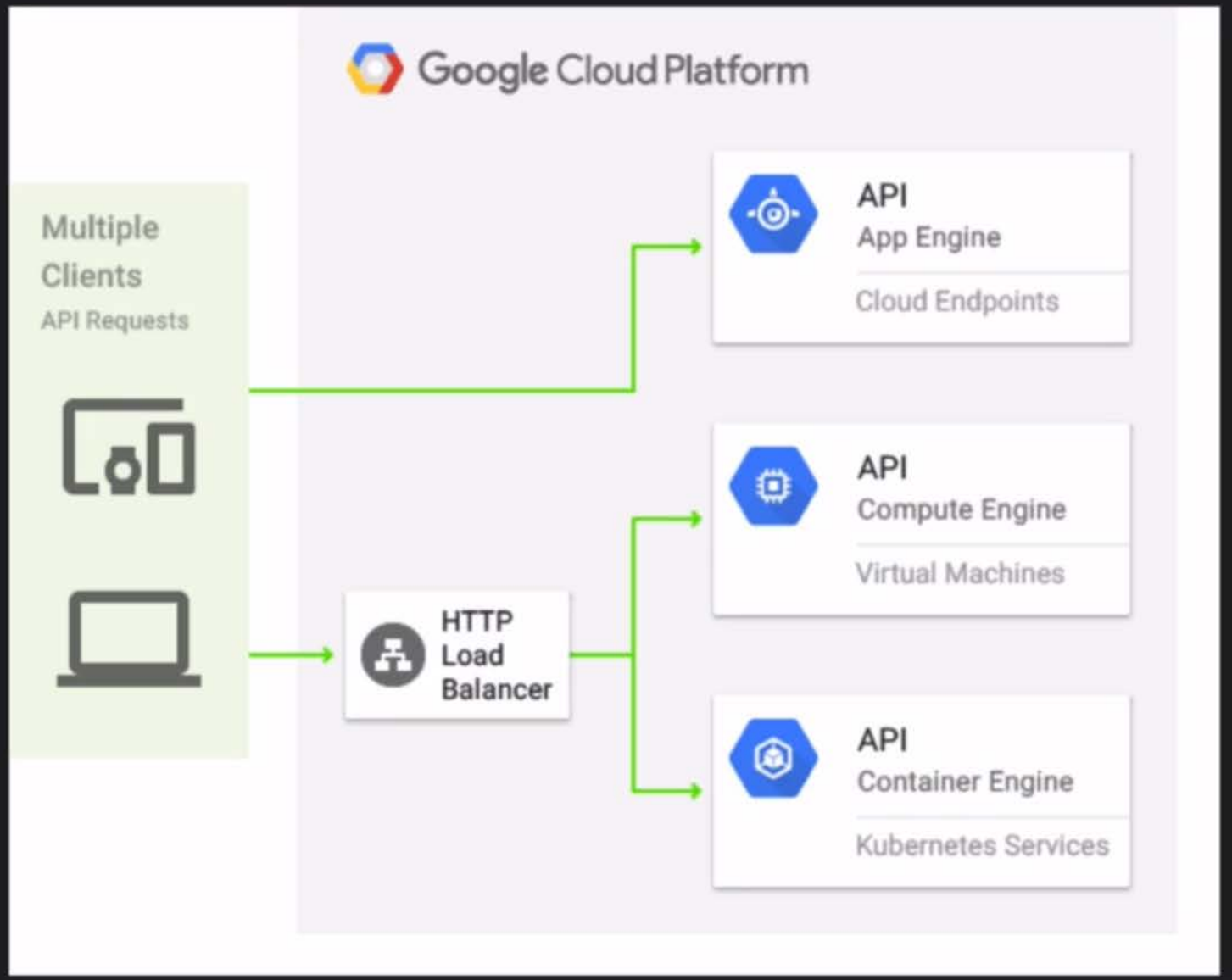
Test Lab for Android



Zonal Regional Multi-Regional Global

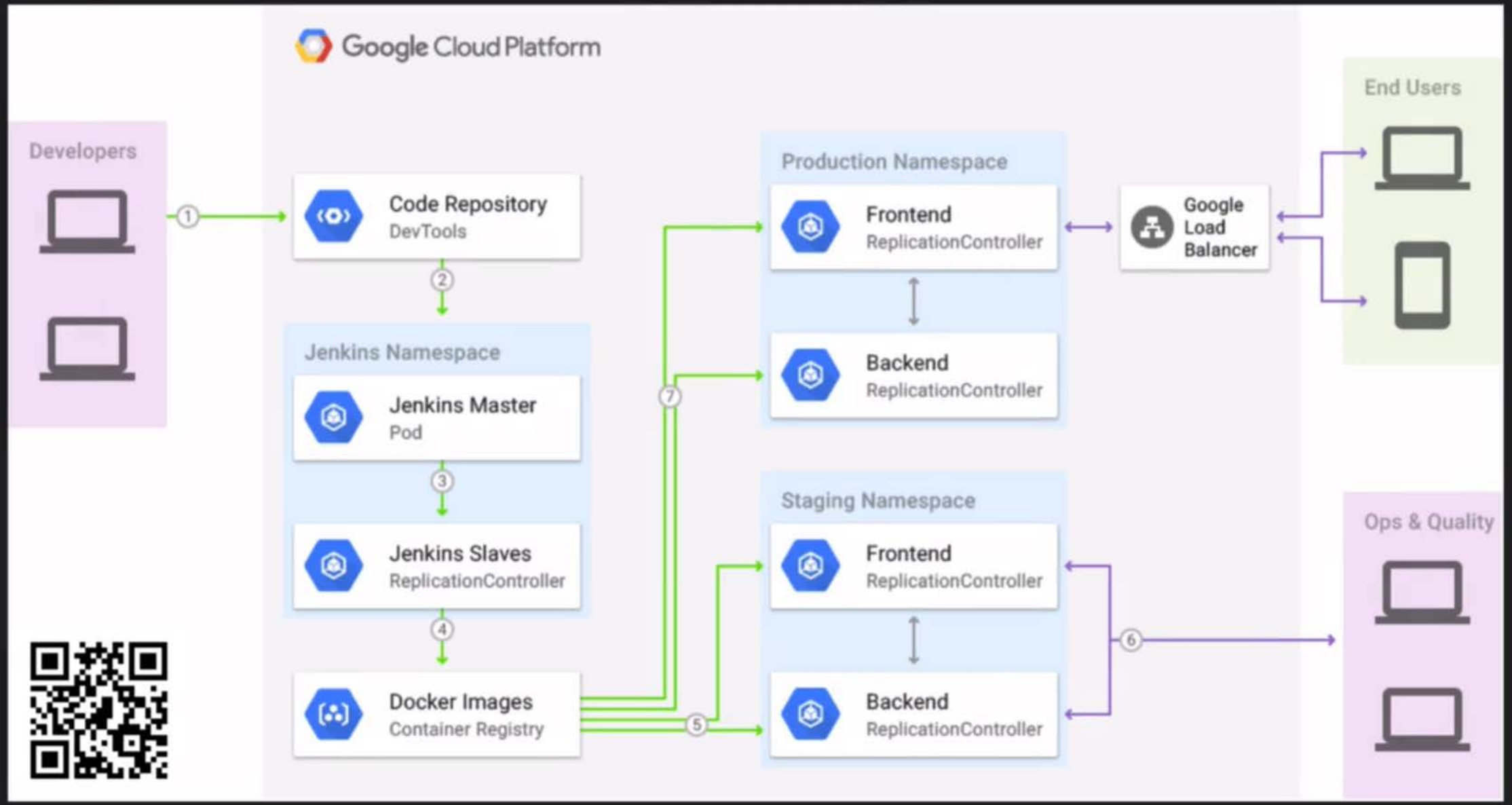
- Cloud infrastructure for running test matrix across variety of real Android devices
- Production-grade devices flashed with Android version and locale you specify
- Robo test captures log files, saves annotated screenshots & video to show steps
  - Default completely automatic but still deterministic, so can show regressions
  - Can record custom script
- Can also run Espresso and UI Automator 2.0 instrumentation tests
- Firebase Spark and Flame plans have daily allotment of physical and virtual tests
- Blaze (PAYG) plan charges per device-hour—much less for virtual devices

API Hosting

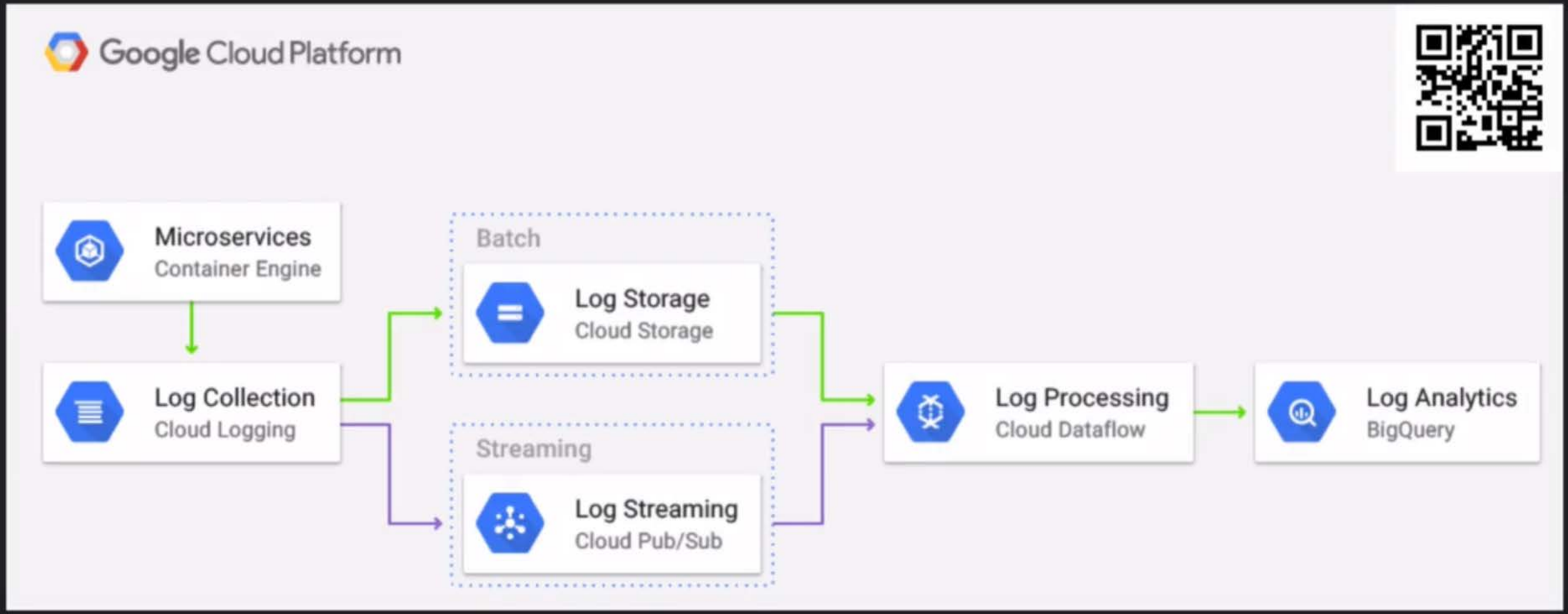

Google Cloud Platform
Multiple Clients
API Requests
HTTP Load Balancer
API
App Engine
Cloud Endpoints
API
Compute Engine
Virtual Machines
API
Container Engine
Kubernetes Services

Jenkins on Kubernetes

A CLOUD GURU



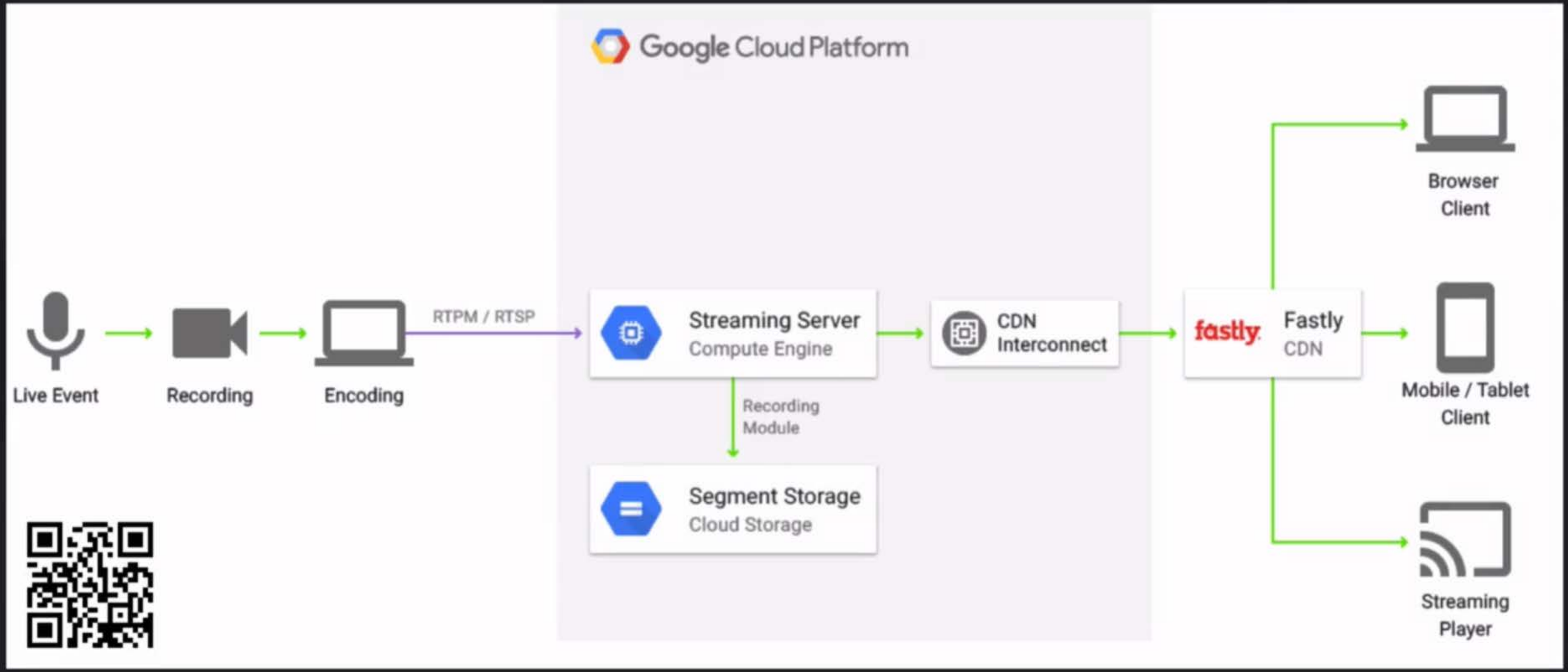
Log Processing



Putting It All Together

Live Streaming

A CLOUD GURU



Shopping Cart Analysis

A CLOUD GURU