Google Certified Professional - Cloud Architect - Part 2

Welcome to Google Certified Professional - Cloud Architect - Part 2

Course Introduction

What is this course about?

Part 2 of 3 for preparing for the Google Cloud Architect exam

Emphasis on working with wide assortment of GCP services

Point of view of an 'architect' building cloud infrastructure

Pre-requisites

GCP Part 1 course, or equivalent knowledge

Assumption that you know the 'basics'

What this course covers

Role of the cloud architect

How to properly manage access to your GCP resources

Lots of command line interaction

How to monitor your GCP environment for problems

How to connect Google Cloud as an extension of your private network

How to manage Software Defined Networks on GCP

Deep exploration of Compute Engine's services and features

Learn about automating our workloads to work smarter, not harder

Finally, couple real life business scenarios that you'd see in practice

Let's get started!

Google Certified Professional - Cloud Architect - Part 2

Purpose of this Course

What is the point of this Part 2 course?

The GCP Architect exam is a **professional** level exam

Very wide range of skillsets tested

Course is in three parts to start with beginner fundamentals, and move to advanced understanding

Theme of each course

Part 1 – Beginner's Introduction

- Purposely kept at a beginner's level to understand the basics

Part 2 – Become expert craftsmen with GCP's tools – build 'houses'

- Taking off the 'training wheels', hands on with tools

Part 3 – Big picture topics – build 'skyscrapers'

Objectives throughout:

Prepare for the exam focused topics and questions

Become experts in GCP to use in real life scenarios

Google Certified Professional - Cloud Architect - Part 2

Role of the Cloud Architect

What is a Google Cloud Platform Architect?

A Google Certified Professional - Cloud Architect enables organizations to leverage Google Cloud technologies. Through an understanding of cloud architecture and Google technology, this individual can design, develop, and manage robust, secure, scalable, highly available, and dynamic solutions to drive business objectives.

What does this mean in practice?

What does a traditional architect do?

An **architect** is someone who plans, designs, and reviews the construction of buildings.

A poorly designed building does not withstand the test of time, and may collapse.

An architect must know the tools to build a structure with, and how to best use them

A cloud architect has the same role

A GCP Architect plans, designs, and builds the **infrastructure** for an organization to host their workload on GCP

Must be skilled in using the tools available

Infrastructure must stand the test of time

Must also know when to use the best tool for a particular business requirement

- Several methods to fulfill a task, some are better than others

How is a cloud architect role different?

Traditional architects are constrained by physical resources

However, cloud technology allows infrastructure to scale, automate, and expand with no limits

- Worldwide reach – Google deploys and manages thousands of resources every day

A skilled architect is able to plan to scale

The subjects of **scalability** and **automation** will be a frequent theme throughout this course

As we go through this course...

Adopt the mindset of scalability and automation

We're going to start small, but work toward thinking at a larger scale

Think like an architect! You are designing and building a structure that many will benefit from!

Google Certified Professional - Cloud Architect - Part 2

The Importance of Hands on Practice

How do you build experience with GCP?

PRACTICE PRACTICE PRACTICE

Emphasis on hands on approach

Do not be afraid to experiment and break something

Free Trial and Always Free give generous allotment

Create, Experiment, Break, Delete, Repeat

Google Certified Professional - Cloud Architect - Part 2

Management Services

In this section...

Cloud Resource Manager (Quotas, IAM, Billing)

Cloud IAM (Identity and Access Management)

Stackdriver

To put another way:

How is our stuff organized, and what are we paying for?

Who has access to our stuff?

What is happening to our resources?

Role of Cloud Architect (why this matters)

While not as fun as **creating** infrastructure, **managing** that infrastructure is just as important

If not properly managed:

- Unauthorized access
- Runaway/hidden costs
- Project resources halted due to improper billing setup
- Application/service errors without us knowing about it or knowing where to fix

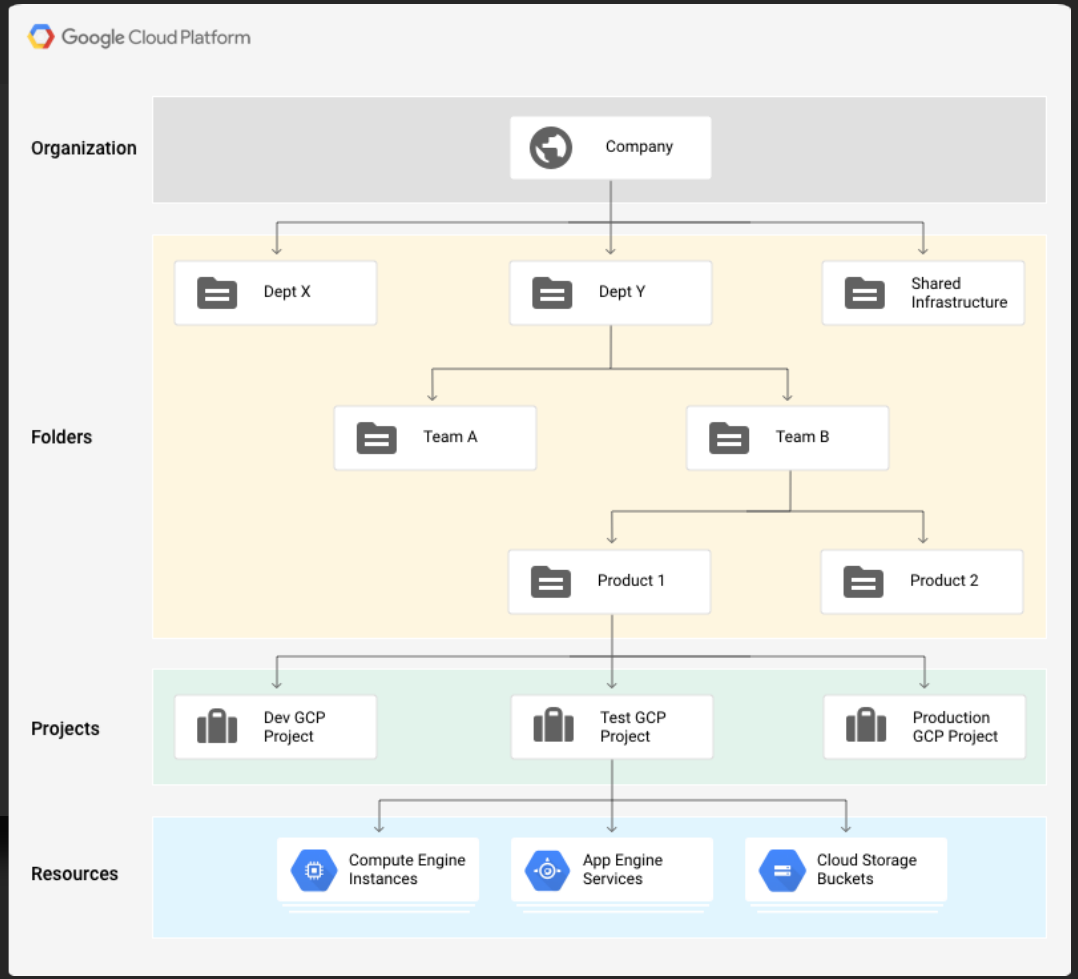
Management services concepts will be on the exam

Google Certified Professional - Cloud Architect - Part 2

Organization Node and Folders

GCP Organization Structure

- Organization
- Folders
- Projects
- Resources



What is the Organization node?

Root node for all GCP resources (in an organization)

- Example: linuxacademy.com, professionalwireless.net

Personal gmail accounts do not have an organization (or folders)

For G Suite or Cloud Identity domains

Control of account can be given to different people in the organization

Organization admin – useful for auditing

Organization owner – reserved for G Suite super admin

- Best practice is to have more than one org owner

Folders

Second node in Cloud Platform Resource Hierarchy

Group projects under organization in a hierarchy

- example: group projects on per-department basis

Group resources that share common IAM policies

Roles granted to folder apply to resources inside. Great for grouping permissions.

Be careful when moving resources. Roles granted at folder level, not resource level will be lost!

Google Certified Professional - Cloud Architect - Part 2

Quotas

What are quotas?

Caps on resources you can create

- example: 48 total CPU's per region, 5 static IP's per project

Prevent unexpected spikes in usage

Generally one of three types:

- Resources per project
- API rate limit requests per project
- Per region

Free trial has additional quotas in place – trial should be primary for testing/evaluating

Why do we need quotas?

Ability to instantly create infinite resources can lead to massive costs

Protection from unexpected spikes in resource usage

Prevent runaway consumption due to error or malicious intent

Prevent unexpected spikes in billing

Review sizing considerations

- "Do you **really** need a single 64 core VM?"

Increasing quota caps

Most quotas are soft caps – can be raised by request

Support ticket or self service form

Quotas can be viewed in console

Best practice – proactively request increase for anticipated demand

Google Certified Professional - Cloud Architect - Part 2

Labels

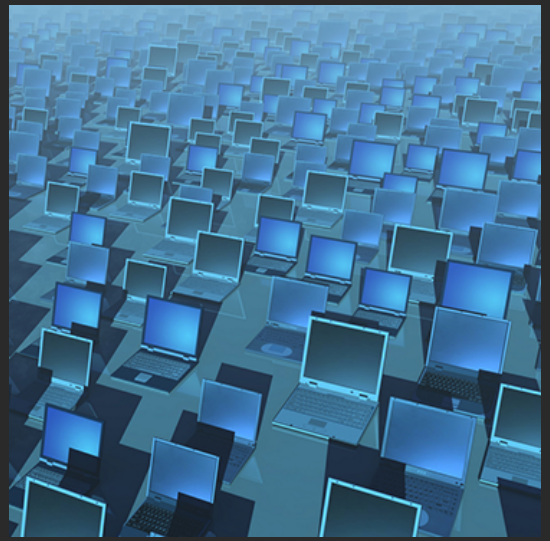
Why labels?

Methods of organization and segregation

- Project
- Folders

However, even further levels of organization may be necessary

- What if we're tracking thousands of VM's, disks, etc



What are labels?

Tool for organizing GCP resources

A virtual sticky note

Almost any resource can be labeled

- Projects, VM's, snapshots, images

Set in console, gcloud, or API

No set rules on how to label – depends on your organization needs

Up to 64 labels per resource

How it works

Key:value pair

- Key – unique identifier
- Value – identified data or pointer to data location

Popular in lookup tables and configuration files

Key cannot be empty, but Value can

Can be applied to multiple items at once

Examples

Environment - env:prod/env:test

Owner or point of contact - owner:matt, contact:devops

Team or cost center – team:research, team:marketing

App component – component:backend, component:frontend

Resource state – state:readyfordeletion, state:inuse

Labels vs. tags

Labels:

- Can be applied across all of GCP
- Organization purposes
- Does not affect resource operation

Tags:

- Only for network/VPC resources
- Affects resource operation (e.g. firewall rule application, network route)

Google Certified Professional - Cloud Architect - Part 2

Cloud IAM Recap and Overview

To Recap from Part 1

IAM determines who – can do what – on which resource

Who = members

What = roles

Resource = GCP services

Members

- People – Google account, Google group, G Suite domain, Cloud Identity domain
- Applications – service account

Roles

- Collection of permissions in form of ***<service>.<resource>.<verb>.*** E.g.: compute.instances.delete
- Permissions not assigned directly to user but bundled in roles
- NEW – custom roles (currently in Alpha)

Primitive and Predefined (Curated) Roles

- Primitive – broad, assigned at project level (Viewer, Editor, Owner)
- Predefined – granular, assigned at resource level (e.g. Compute Engine, Storage)

IAM Policy

- Collections of statements that define access
- Full list of roles granted to a member or a resource

Policy Hierarchy

- Parent policy overrules more restrictive child policy
- Child policy inherit parent role
- Can use a restrictive parent role, and grant more access at child role level

Resources

- Organization, folders (new topic), projects, GCP services

ALL of these are "IAM Objects"

Building upon Previous IAM Lesson

Organization node

Folders

Service accounts in depth

IAM best practices

Hands-on and exercises

Google Certified Professional - Cloud Architect - Part 2

Service Accounts

What Is a Service Account?

Special type of Google account not attached to a user

Identity attached to a VM or application

Resources don't need **end user authentication**

Identity shown by email address

- example: 126692436861-compute@developer.gserviceaccount.com

Authenticate between services

More stringent, better logging capabilities

Types of Service Accounts

Google-managed

- Represent different Google services and are automatically granted IAM roles
- [PROJECT_NUMBER]@cloudservices.gserviceaccount.com
- Generally invisible to end user

User-managed

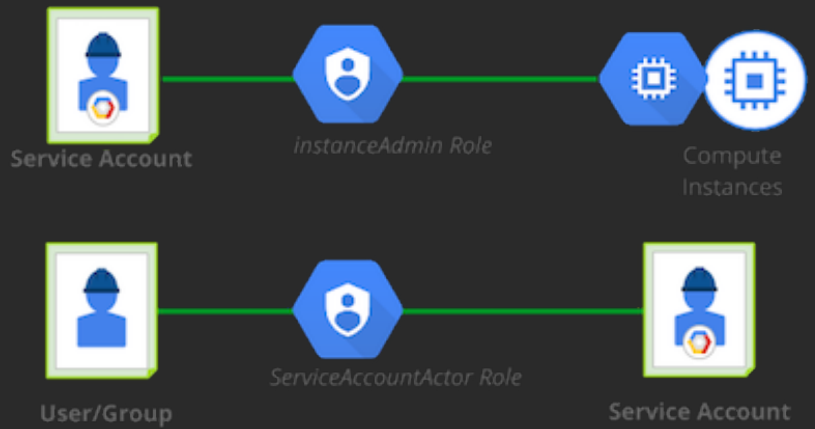
- Created for/by you, based on enabled APIs in project
- [PROJECT-NUMBER]-compute@developer.gserviceaccount.com
- [PROJECT-ID]@appspot.gserviceaccount.com
- Both automatically created and user-created

Both a Member and a Resource

Service accounts can be both a member and a resource

Service accounts are granted permissions to a resource

Users are granted **serviceAccountActor** Role to a Service account



Service Account Keys

Service account access managed by **account keys**

- Think of it as the service account 'password'

Default service account keys are managed by Google and can't be accessed/edited

Custom service accounts can use custom keys, which you store/manage

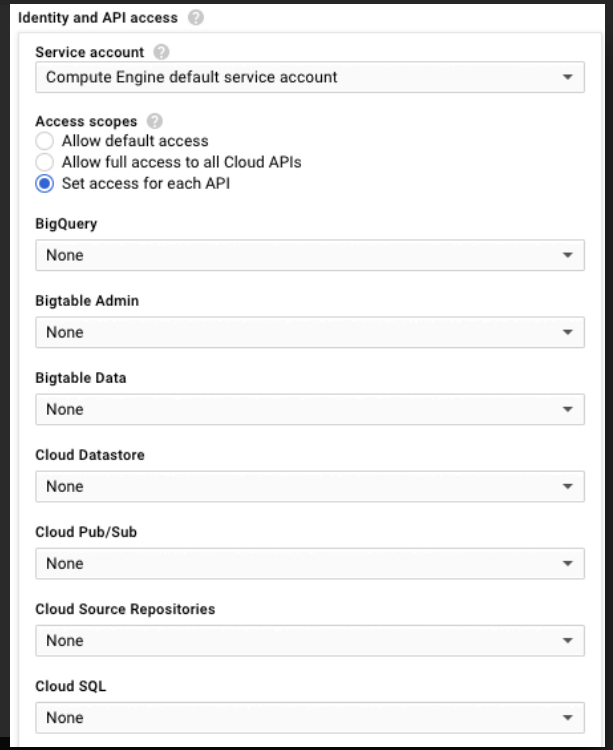
- Google maintains public copy for verification, but the public/private key pair is yours to manage
- If you lose your private copy of the key, Google cannot retrieve it!

Scopes

Legacy method of granting permissions for service account for an individual instance

Combine with IAM roles with service accounts to grant per-instance permissions to other GCP resources via the instance

IAM roles and scopes determine service account permission for that instance



To summarize:

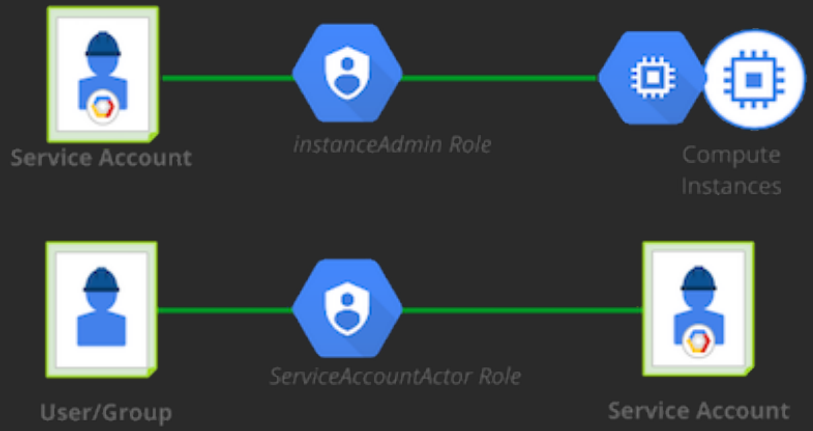
Service accounts grant application/VMs access

Users are granted access to act as a service account...

- ...and service accounts are granted access to resources.

Service accounts use keys for access

Service accounts granted access based on both scopes and IAM



Time for Some Hands On!

Google Certified Professional - Cloud Architect - Part 2

Cloud IAM Best Practices

Principal of Least Privilege

Members should only have just enough permissions to perform their job role

Use predefined roles over primitive roles

Treat each app component as a separate trust boundary

- Create separate service account for each service

Remember that child policies cannot overrule parent

Grant roles at smallest scope necessary

- e.g. Compute Instance Admin vs. Compute Admin

Restrict service account access

Restrict who can create and manage service accounts (Service Account Admin)

Careful with Owner roles (Editor might be better)

- Owner can change IAM policy
- Audit tracking

Service Accounts

Rotate service account keys (user managed)

- IMPORTANT – don't delete service in use by running instances

Don't check in service account keys to source code or leave in downloads directory!

Name service keys to reflect use and permissions

Auditing

Use Cloud Audit logs to **regularly** audit IAM policy changes

Export audit logs to Cloud Storage for long term retention

Restrict log access with cloud logging roles

Other Best Practices

When possible, use groups

Separate production and development environments

Google Certified Professional - Cloud Architect - Part 2

Billing

Why is this important?

The bigger you scale, the greater in number your resources

Need to be informed of costs and identify runaway costs

Billing + Cloud IAM

Billing roles defined in Cloud IAM

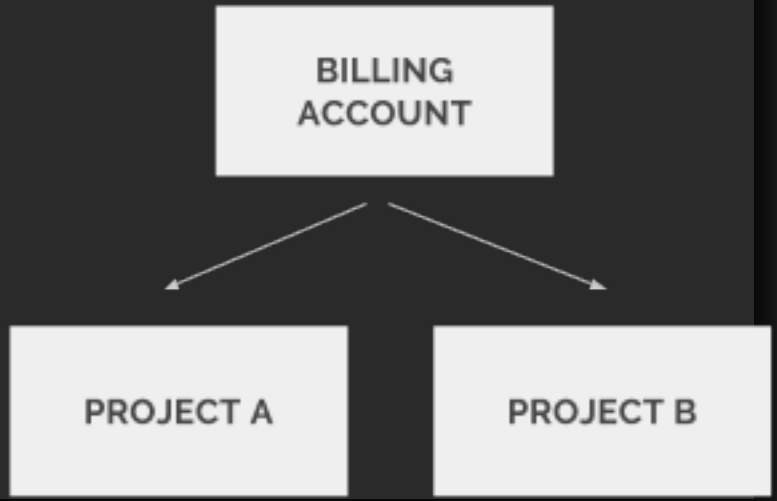
- Billing Account Creator, Billing Account Administrator, Billing Account User, Billing Account Viewer, Project Billing Manager

Like IAM, billing resources are hierarchical

Organization is top of hierarchy

Billing accounts are linked to Projects

- Requires Billing Account User



Viewing billing info

"What am I paying for?"

View in web console

Export to Cloud Storage and Big Query

- Exports can be forwarded to financial department/auditors

Set budgets and alerts

- Does not halt billing/resource usage, but alerts at thresholds

Google Certified Professional - Cloud Architect - Part 2

Stackdriver Overview

Answers the question:

HEY, WHAT'S GOING ON?

What is Stackdriver?

Suite of tools for monitoring, logging, and tracking diagnostics for your applications

Recent acquisition – previously exclusive to AWS

Native monitoring of **both** GCP **and** AWS

Dynamically discover all GCP resources

- Install Stackdriver client on VMs for even greater levels of monitoring

Will be tested on the Architect exam!

Five Different Products

Monitoring

- Monitor metrics, health checks, dashboards and alerts

Logging

- Audit of activity

Error Reporting

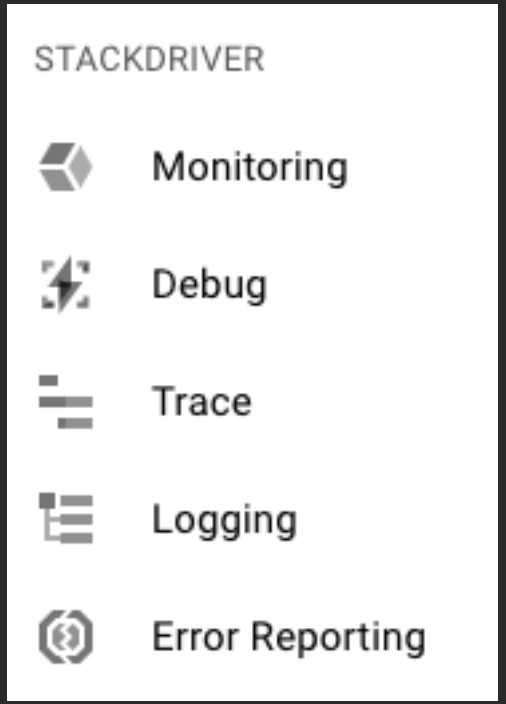
- Identify and understand application errors

Trace

- App Engine – find bottlenecks

Debugger

- Find/fix code errors in production



Benefits of Stackdriver

Multi-cloud monitoring – native GCP and AWS

Identify trends and prevent problems before they occur

Centralized logging for all of GCP/AWS

- One-stop shop

Better signal-to-noise ratio

- More relevant alerts

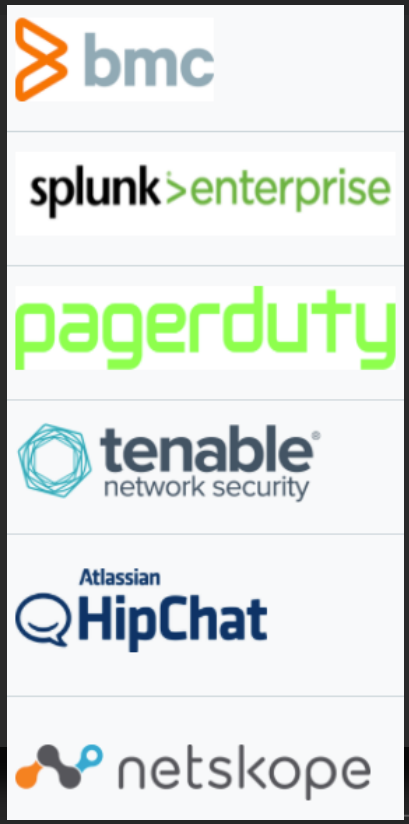
Find and fix problems faster

Third Party Integrations

Site Reliability Engineer (SRE) vendors

Centralized logging integrates with third party products

BMC, Splunk, Hipchat, PagerDuty, Netskope



Best Practices

Create single project for Stackdriver monitoring

'Single pane of glass' – monitors all GCP/AWS resources across projects

Determine monitoring needs in advance

IAM controls – separate Stackdriver accounts for data and control isolation

Google Certified Professional - Cloud Architect - Part 2

Stackdriver Logging

What Is It?

Repository for log data and events – GCP and AWS

- Single repository for data from multiple sources

Store, search, analyze, monitor, and alert

Collect platform, system, and application logs (with agent)

Tight integration with Stackdriver Monitoring

Real time and batch monitoring

Export logs to other sources for long term storage/analysis

Basic Concepts and Terms

Associated by project

- Logs viewer only shows logs for one project

Log entry – records status or event

- Includes **log name** (e.g. 'syslog', 'compute.googleapis.com/activity')

Logs – named collection of log entries

- Only exist if there are log entries

Retention period

- Depends on admin/data access, basic or premium tiers

Audit Log Types

"Who did what, where, and when?"

Admin Activity

- Change, create, modify configuration or metadata
- Example: new Compute Engine instance created, IAM roles change
- Requires IAM role Logging/Logs Viewer or Project Viewer (or higher)
- Always enabled, no charge

Data Access

- Create, modify, or read **user-provided** data
- Requires IAM role Logging Private Logs Viewer or Project Owner
- Disabled by default, charge depending on quantity produced

Retention

Depends on log type, subscription tier

Admin Activity logs – 400 days – all tiers

Data Access logs – 7 days (Basic), 30 days (Premium)

Non-audit logs - 7 days (Basic), 30 days (Premium)

Allotment:

Basic – 50GB per project

Premium – 50GB per project + 14.25MB per chargeable resource per hour

Overage charge - $0.50 per GB per project

- Only largest customers are likely to exceed free allotment

Exporting Logging Data

Long term storage (Cloud Storage), big data analysis (BigQuery), stream to other sources (Pub/Sub)

The basics:

Requires a project and destination service

Create filter – select log entries to export

Choose destination – Cloud Storage, BigQuery, Pub/Sub

Filter and destination held in a sink – direct what entries to copy to which destination

Only new entries will be exported after sink creation

Best Practices

Search for specific values for faster searches

- Log entry name, resource type, labels

Use advanced filters – cut out 'white noise'

Use advanced viewing interface

Google Certified Professional - Cloud Architect - Part 2

Stackdriver Monitoring

What Is It?

Full-stack monitoring, powered by Google

- What is up? What is down? What is overloaded?

Native monitoring of GCP, AWS, and third-party applications

Monitor system and application metrics

Interacts with Stackdriver Logging

Easy to view insights with dashboards and alerts

Pricing

Basic and Premium tier

Separate from GCP account status

Only applies to Monitoring – other Stackdriver products are 'baked in'

New account gets Premium features for 30 days before downgraded to Basic account –

- No auto charge when trial ends

Basic vs. Premium Comparison

Feature	Basic Tier	Premium Tier
Price	Free	$8.00[1] per chargeable resource[2] per month (prorated hourly)
Supported clouds	GCP only	GCP and AWS
Logs allotment[3]	50 GB per project per month	50 GB per project per month plus 14.25 MB per chargeable resource[2] per hour
Logs retention: Admin activity audit logs Data access audit logs Non-audit logs	 400 days 7 days 7 days	 400 days 30 days 30 days
User-defined metric allotment[4]	None	500 time series per chargeable resource[2], and 250 metric types per project
Metric data retention	6 weeks	6 weeks
Alerting policies	Some limitations[5]	No limitations
Stackdriver VM agents	Logging agent only	Logging agent and Monitoring agent
Stackdriver Error Reporting, Stackdriver Debugger, and Stackdriver Trace	For applications on GCP	For applications on GCP and AWS (if supported)
Cloud Console Mobile App	Included	Included

Stackdriver Agent

Software installed on VMs

Recommended but not required

Agentless, can still get CPU, disk/network traffic, and uptime info

Agent accesses additional resource and application service info

Requires Premium tier

Monitors many third party applications

Apache web server	Kafka	RabbitMQ
Cassandra	Memcached	Redis
CouchDB	MongoDB	Riak
Elasticsearch	MySQL	Tomcat
HBase	Nginx	Varnish
JVM Monitoring	PostgreSQL	ZooKeeper

Google Certified Professional - Cloud Architect - Part 2

Stackdriver Trace, Error Reporting, and Debugger

Error Reporting

Real-time error monitoring and alerting in your application

Quickly understand errors

Automatic and real-time analysis

- Alerts and dashboards

Built into App Engine – automatically enabled

In Beta for GAE Flexible, GCE, GKE, EC2 (AWS), and Cloud Functions

GCE, GKE, EC2 require Stackdriver logging agent to be installed

Java, Python, JavaScript, Ruby, C#, PHP, and Go

Trace

Find performance bottlenecks – latency

Collect data from Google App Engine (GAE), Google HTTP load balancers, or apps with Stackdriver Trace SDK

Integrated into App Engine Standard – automatically enabled

Available for GCE, GKE, and GAE (Flexible)

- Requires enabling Stackdriver Trace API or SDK (depending on library)

Can be installed on non-GCP resources

Trace Helps Answer Questions

How long does it take my application to handle a given request?

Why is it taking my application so long to handle a request?

Why do some of my requests take longer than others?

What is the overall latency of requests to my application?

Has latency for my application increased or decreased over time?

What can I do to reduce application latency?

Debugger

Debug application

Inspect application state without stopping or slowing app

Does not require adding log statements

Automatically enabled in GAE Standard

Available in GAE Flexible, GCE, and GKE with additional configuration

Java, Python, Go, Node.js

Can be installed on non-GCP resources

Google Certified Professional - Cloud Architect - Part 2

Google Cloud Storage Review/Concepts

What is Google Cloud Storage?

Storage for unstructured data

- Anything that is not database/tables and rows
- Pictures, videos, files, scripts, etc

Virtually limitless size

Pay per use, not allocation (elastic)

Primary unit is a bucket

- Access managed via IAM
- Can be arranged in file/folder format for organizational purposes

Objects go inside of the bucket

- Objects = files
- Note – folders are also considered objects
- Inherit bucket permissions and storage class

Storage classes

Regional	Single region Hot data in single geographical location No retrieval costs
Multi-Regional	Span multiple regions Hot data from different geographical locations No retrieval costs
Nearline	Regional or multi-regional Archive data accessed once per month Lower monthly cost with some retrieval costs
Coldline	Regional or multi-regional Archive data accessed once per year Even lower monthly costs with higher retrieval costs

Changing storage classes

Nearline/Coldline can be regional or multi-regional

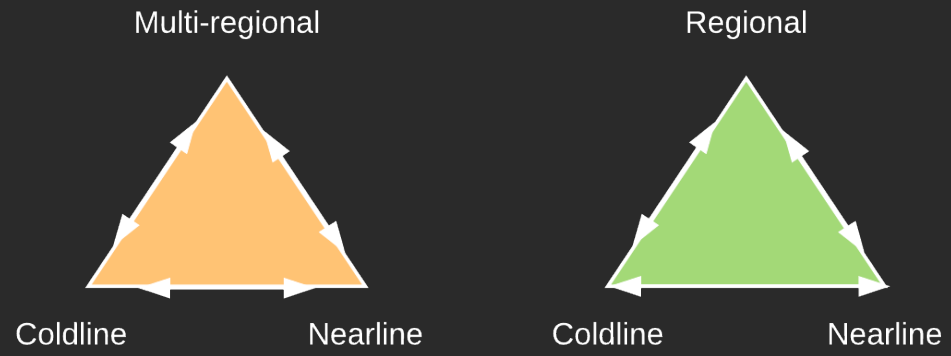
Cannot change from multi-regional to regional (and vice versa)

Changing class only affects new objects

- Previous objects are previous class (change with gsutil)

Objects can be moved to another bucket

- Same class – can use web console
- Different class – must use gsutil



Google Certified Professional - Cloud Architect - Part 2

Cloud Storage Security Concepts

Security Is REALLY Important!

When taken lightly, you can end up on the news... and not in a good way

There's a Hole in 1,951 Amazon S3 Buckets

Corporate tech giant leaves secret data exposed to public internet

Top level executives among those responsible in data leakage in Asean

How Private Data Became Public on Amazon's Cloud

Amazon S3 storage buckets set to 'public' are ripe for data-plundering

Among 12,328 randomly selected Amazon Simple Storage Service buckets, researchers found nearly 2,000 containing freely accessible data

Security Is in Your Control!

Cloud storage access is as 'locked down' or 'wide open' as you choose

Access Management Principles

Two methods: IAM and ACL

ACL = Access Control Lists

Some overlap between the two

Security via IAM

Same IAM principles as the rest of GCP

Hierarchical in nature

Granted at project, resource, or individual bucket (but not objects)

Possible to grant access to manage bucket but not view/read objects inside

IAM Roles

Primitive project level roles (Owner/Editor/Viewer)

Standard Storage Roles (work independently from ACLs)

- Storage Admin – applied to either project or individual bucket
- Storage Object Admin
- Storage Object Viewer
- Storage Object Creator

Legacy roles – work in tandem with ACL permissions

- Storage Legacy Bucket Owner
- Storage Legacy Bucket Reader
- Storage Legacy Bucket Writer
- Storage Legacy Object Owner
- Storage Legacy Object Reader

Access Control Lists (ACLs)

Define who has access to buckets/objects and what level of access

Can be applied to bucket or individual objects

Objects inherit ACL from default bucket ACL

- But can also be independent

For non-legacy IAM roles, no overlap

Sounds like a lot of overlap between IAM and ACL... because there is

Should You Use ACLs?

Best practice – use IAM over ACL whenever possible

- IAM gives enterprise-grade control across all of GCP
- IAM leaves an audit trail for access
- When in doubt, use IAM

However, use ACL to grant access to an object without granting access to bucket

- More fine-grained control

Signed URLs - Timed Access to Object Data

Set a timer on access to a bucket or object

Useful for temporarily giving access without need of signing in with Google account

Gives user read, write, or delete access for limited time

Anyone with the URL can access within the time period

Best Practices

NEVER share credentials!

Remove application access when no longer needed

Don't make bucket names a target – e.g. 'gs://confidential-customer-info'

Use groups over individual users for IAM when possible

Check default object ACL before adding objects

Make sure publically readable data is intended!

- Once 'out', you can't bring it back in

Though possible, publically writable buckets are a bad idea

Use signed URLs for secure access w/o need of Google account

Google Certified Professional - Cloud Architect - Part 2

Object Versioning and Lifecycle Management

Object Versioning

Retrieve objects that are deleted or overwritten

Applied to bucket level

Disabled by default

When enabled, deleted and overwritten objects are archived instead of deleted

Object keeps same name but paired with unique identifier number

Actions possible with versioning enabled:

- List archived versions
- Restore live version to an older state
- Permanently delete older version

If versioning is disabled, existing versions remain but new ones not created

Versioning considerations

No default limit on versions – can increase bucket size (and cost) greatly

- Lifecycle Management helps with this

Archive versions retain own ACL, which may not be same as live version

Versioning properties

Generation – update when object content overwritten (delete, overwrite file)

Metageneration – metadata generation/change

No relationship between the two

Object Lifecycle Management

Sets Time to Live (TTL) on an object

- Archive/Delete older versions
- Downgrade storage classes

Applied to bucket level

Often paired to object versioning, but not required

Examples:

- Downgrade the storage class of objects older than 365 days to Coldline Storage
- Delete objects created before January 1, 2017
- Keep only the 3 most recent versions of each object in a bucket with versioning enabled

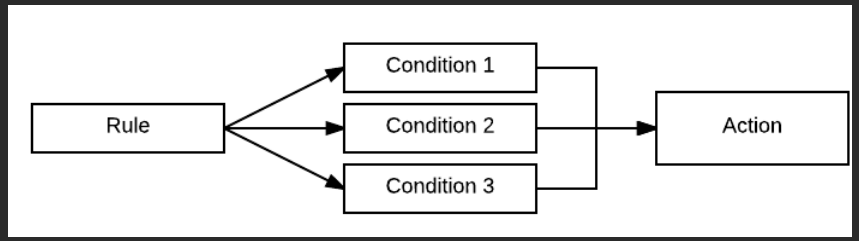
Implemented with combination of Rules, Conditions, and Actions

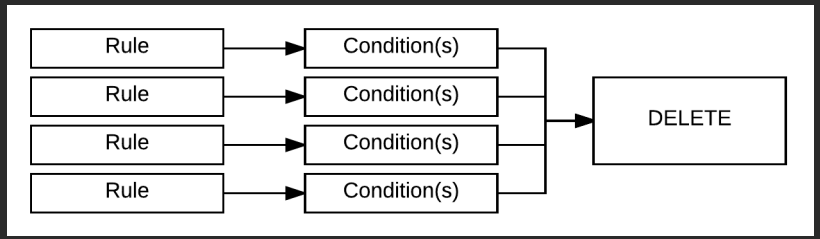
Rules

Specify set of conditions in order to take action

If multiple conditions in a rule, **all** conditions must be met before action taken

However, if multiple rules with same action, **any** met rule executes action





Conditions

Criteria to meet before taking action

Age

- Age in days (TTL)
- Still valid if archived via versioning

CreatedBefore

- Objected created before midnight of specified data (UTC)

IsLive

- Live or archived version of object

MatchesStorageClass

- Condition matches specified storage class

NumberOfNewerVersions

- Condition met when number of newer versions than target is reached

Actions

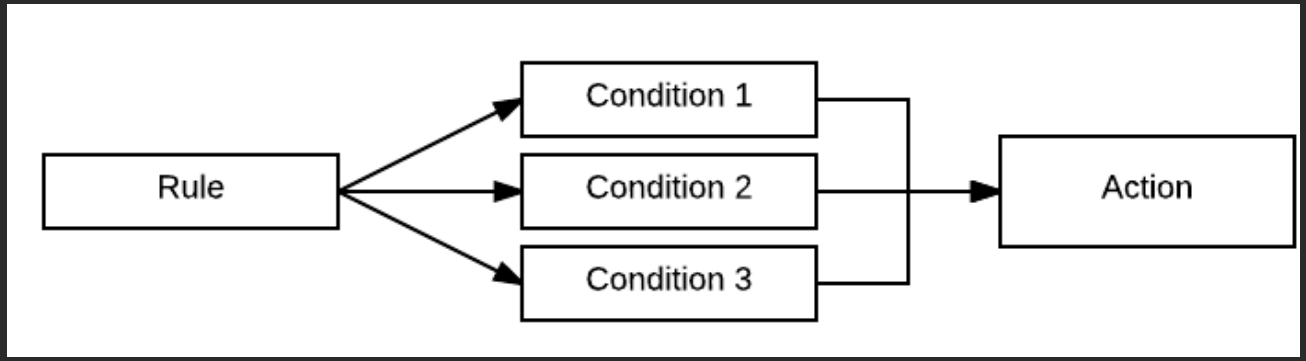
Delete

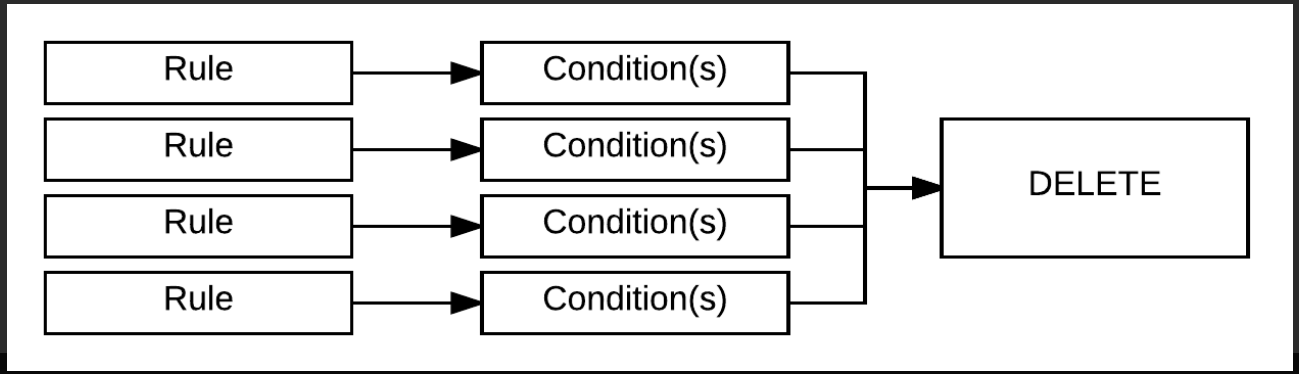
- Versioned or non-versioned
- Delete live (versioned) objects creates archived object
- Deleting archived object deletes permanently
- Deleted objects cannot be recovered!
    - Test rules on development data

SetStorageClass

- Change storage class of affected object
- Remember, cannot go from Regional to Multi-Regional, and vice versa
- Regional Standard → Regional Nearline/Coldline
- Multi-Regional Standard → Multi-Regional Nearline/Coldline

Rules, conditions, and actions (and/or requirements)






Google Certified Professional
- Cloud Architect - Part 2
The Power of the Network

Google's Network Is AWESOME

Biggest distinction compared to other platforms

Worldwide PRIVATE network

- All regions are on one massive private/internal network
- Communication between regions and on-premises (direct peering/carrier interconnect) NEVER touches public internet!

As a result – networking is handled differently than others

Software-Defined Networking (SDN)

Traditional network/data center – manage network hardware

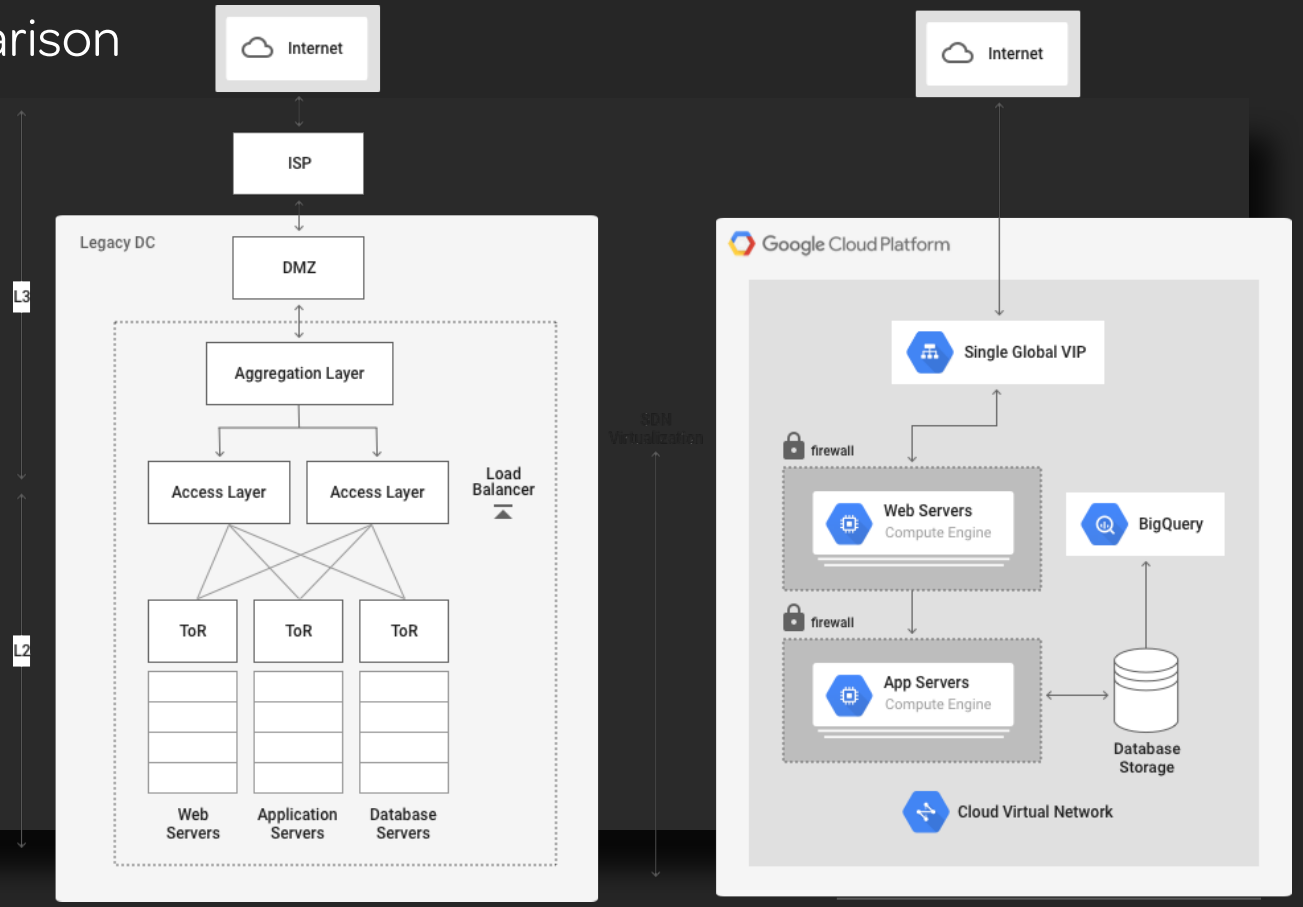
- Switches, routers, load balancers, firewalls, storage devices etc.
- Detailed device configurations, monitor network software, high management overhead

Software-Defined Networking – everything is virtualized

- Removes overhead
- Rapidly customize and scale services
- High throughput
- Global availability
- Seamless upgrades

Note: Traditional networking concepts still apply, such as subnet, routes, firewall rules, DNS, etc.

Hardware/SDN comparison



How Does This Affect Networking on GCP?

Single global/cross-region VPCs

- No managing multiple private networks for global availability

Global internal DNS/load balancing/firewalls/routes

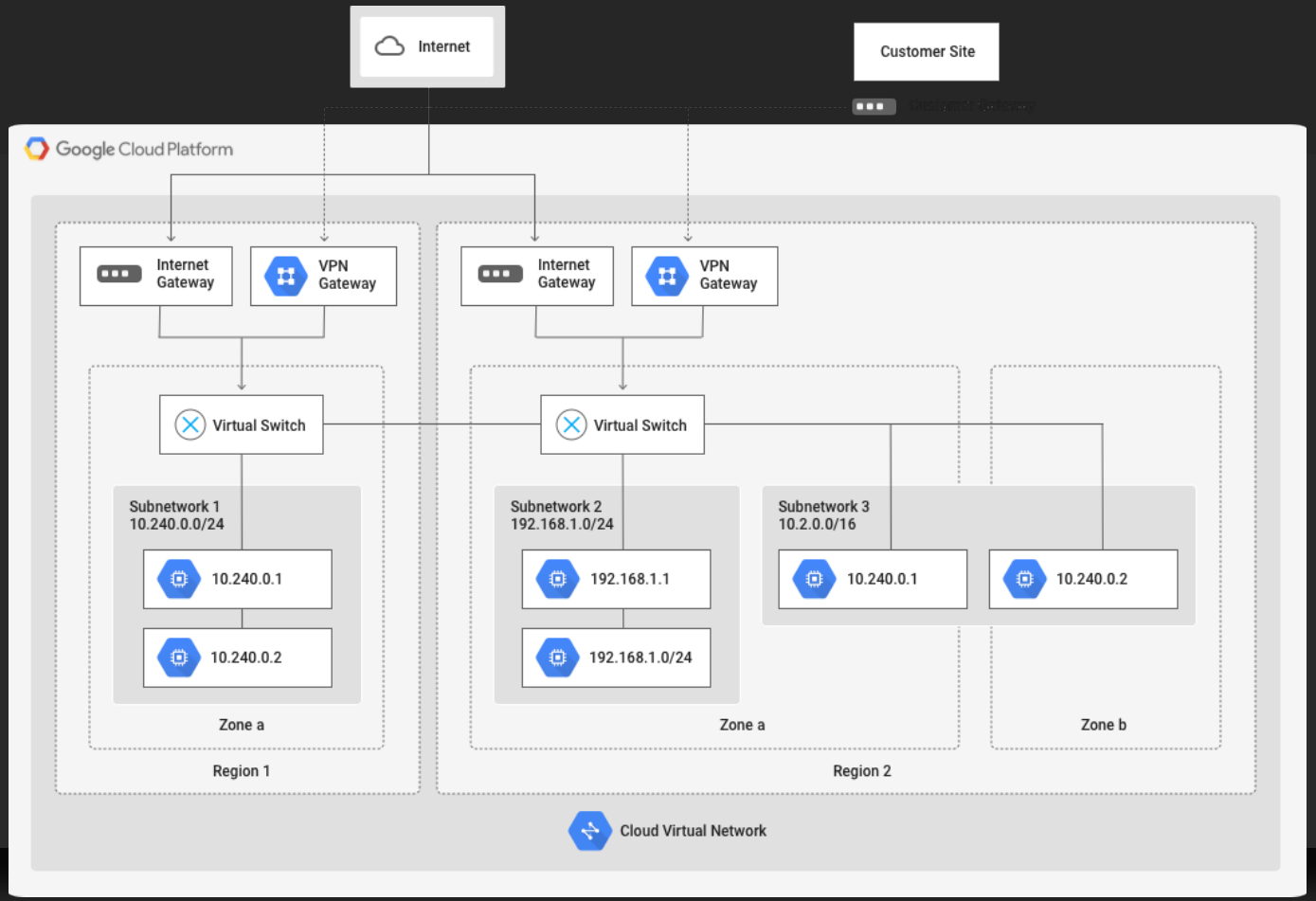
- Separate resources with tags

Global public DNS

Rapid scaling with global load balancers (Layer 7/HTTP)

Subnets within VPC group resources by region/zone

IP ranges between subnets are dynamically expandable


Internet
Customer Site
Google Cloud Platform
Internet Gateway
VPN Gateway
Internet Gateway
VPN Gateway
Virtual Switch
Virtual Switch
Subnetwork 1
10.240.0.0/24
10.240.0.1
10.240.0.2
Zone a
Region 1
Subnetwork 2
192.168.1.0/24
192.168.1.1
192.168.1.0/24
Zone a
Subnetwork 3
10.2.0.0/16
10.240.0.1
10.240.0.2
Zone b
Region 2
Cloud Virtual Network

Extend Google Private Network to On-Premises

VPN

Cloud Interconnect

Direct Peering


Google Certified Professional
- Cloud Architect - Part 2
Connecting your Network to Google

Connection options

Extend your network to Google's network

Global term - Cloud Interconnect

Dedicated Interconnect

Peering

Cloud VPN

Dedicated Interconnect

Physically connect on-premises network to GCP VPC via Google edge location

- Traffic does not touch public Internet
- VPC internal IP addresses directly accessible from company network

Enterprise grade connection to GCP

Useful for

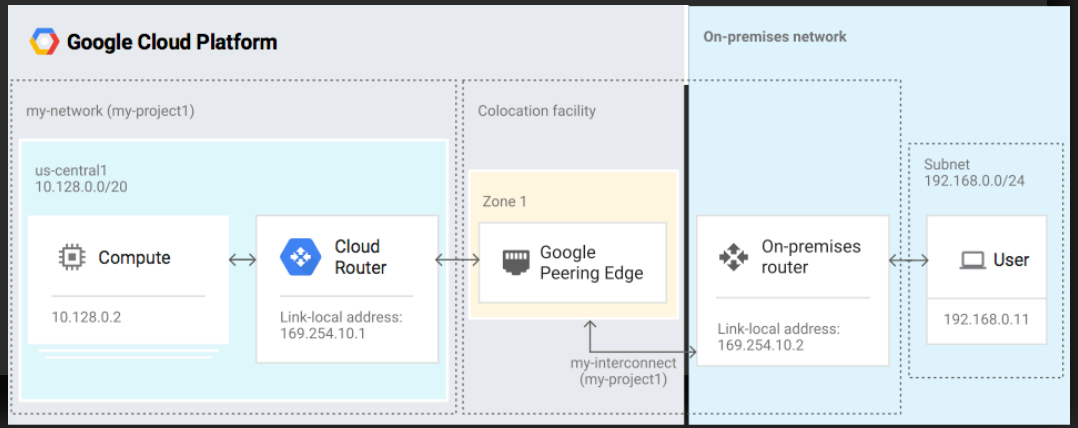
- Hybrid environments – extend corporate data center IP space into Google Cloud
- High bandwidth traffic (e.g. transfer large datasets)

Must be at supported peering location

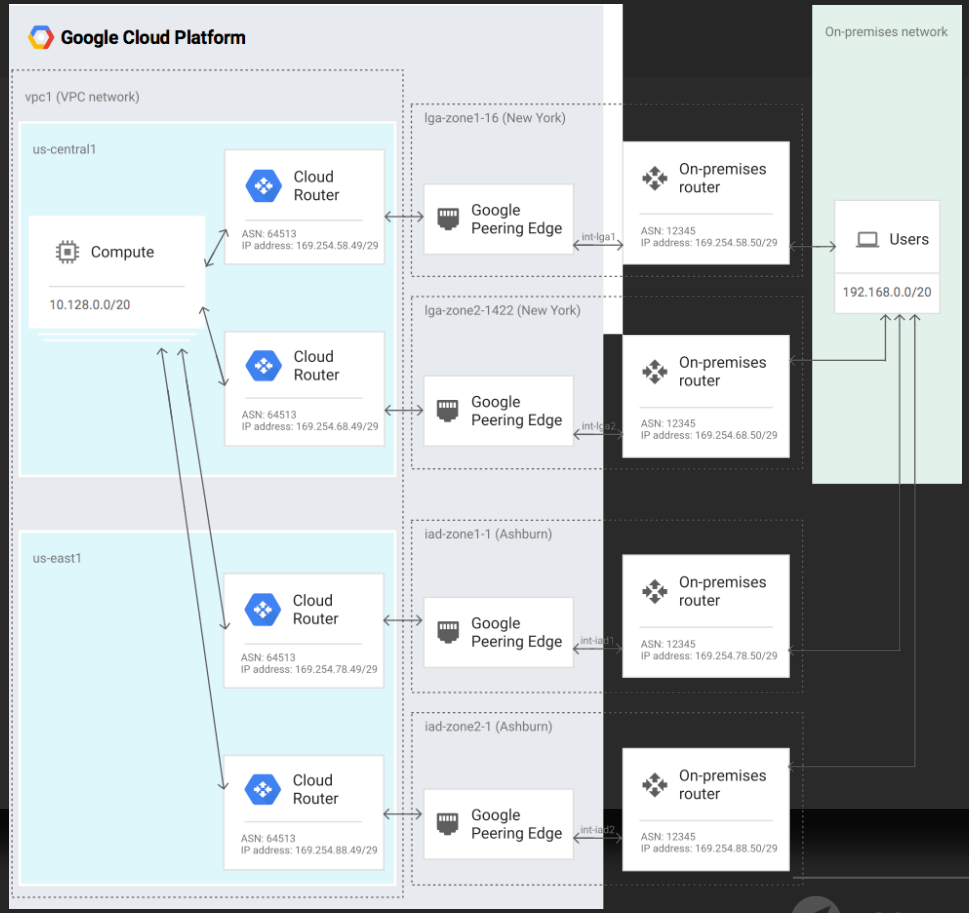
Can be direct with Google or through carrier

$1700 per 10Gbps link, up to 80Gbps total

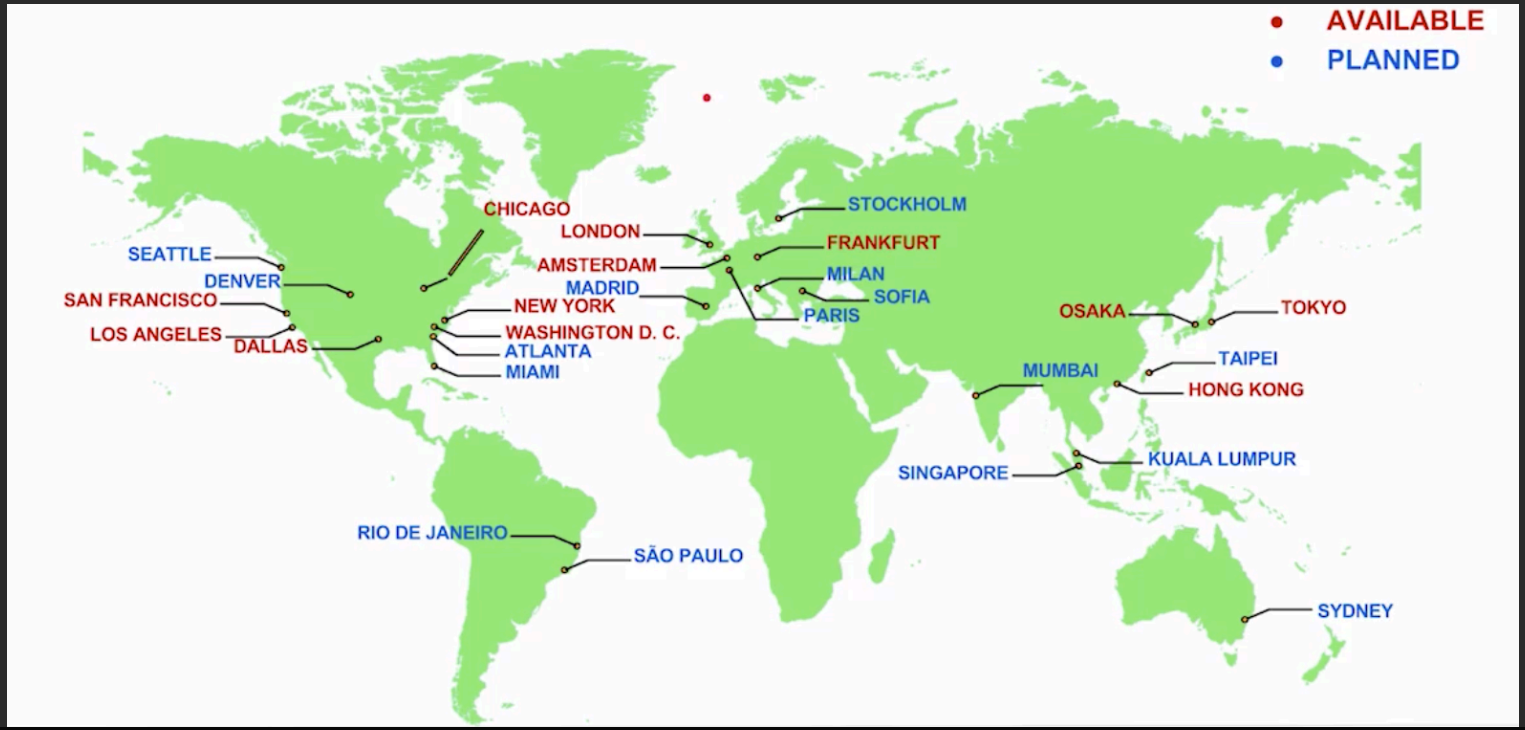
Reduced egress fees



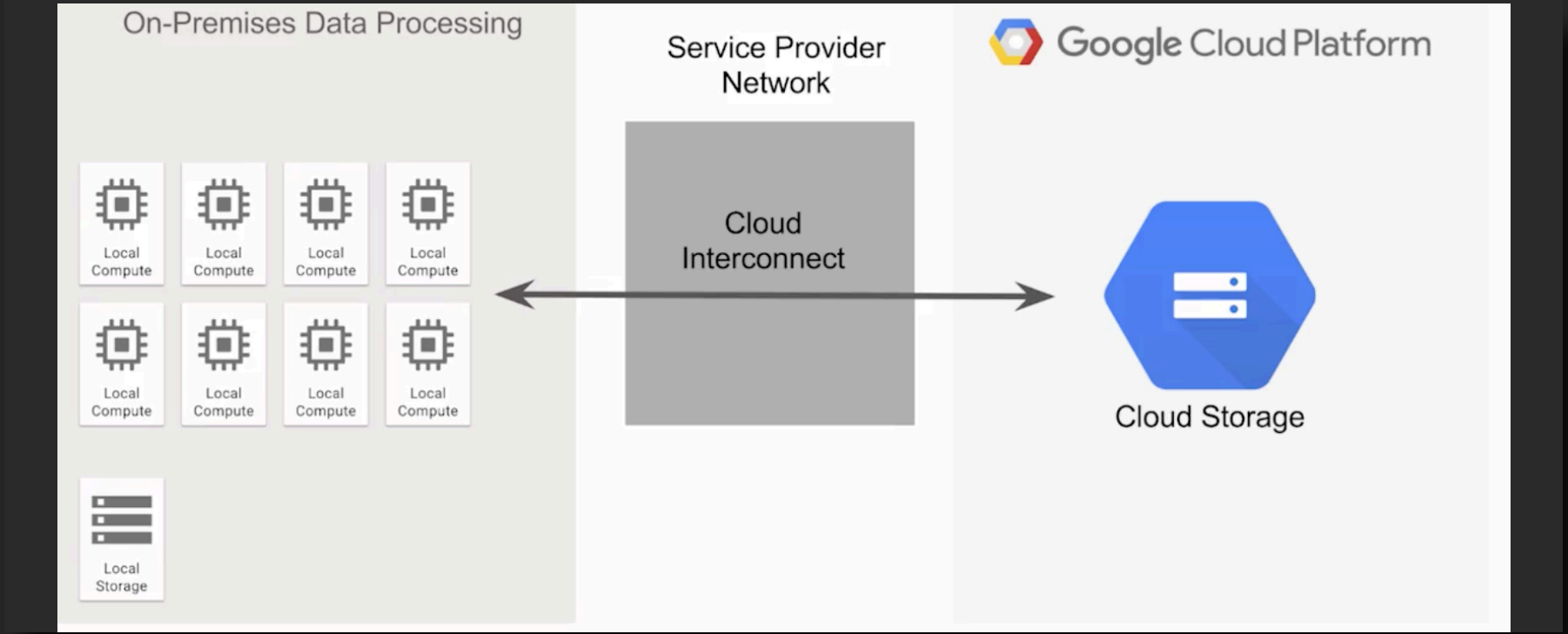
Multiple Interconnects for high availability



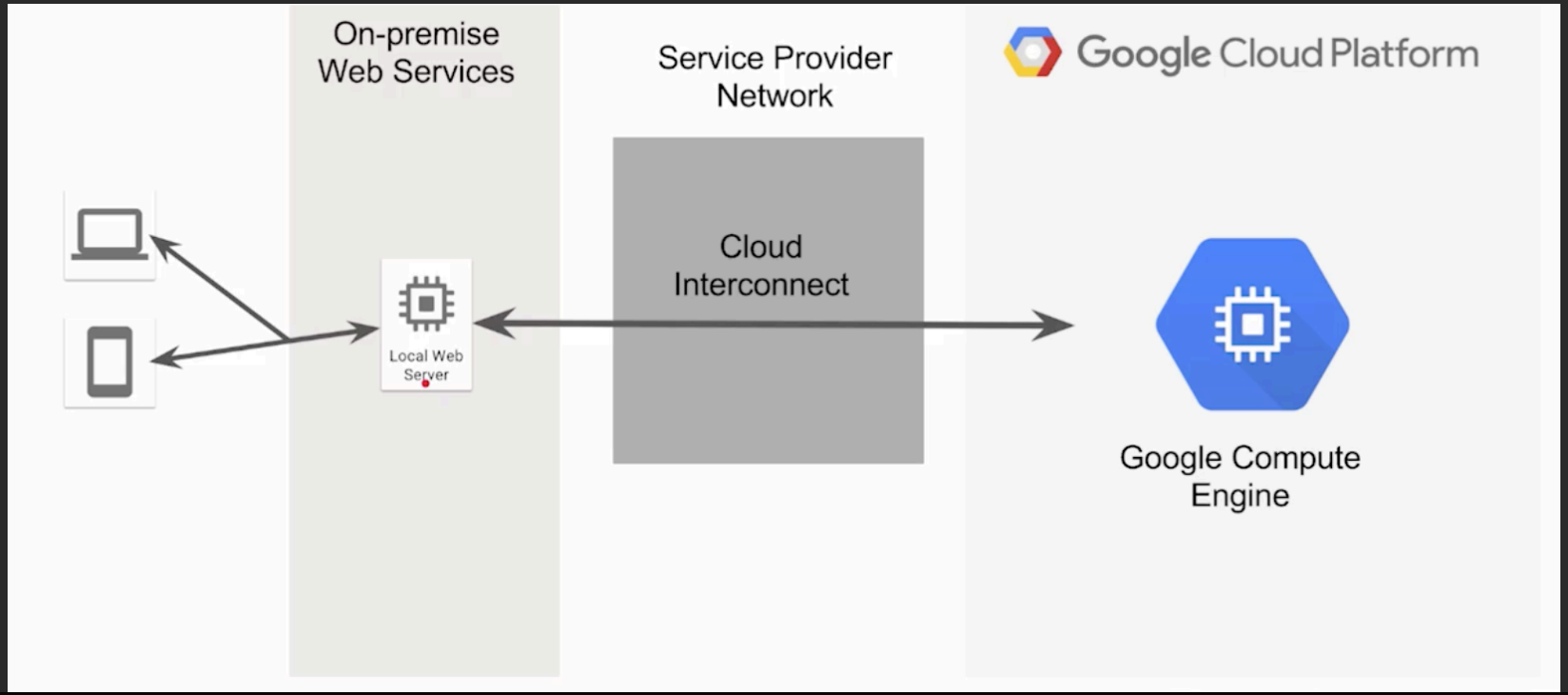
Direct Interconnect locations



Use cases – heavy processing



Use cases – low latency needs



Peering

Connect business directly to Google

70+ location in 33 countries for Direct Peering

Not GCP specific, but exchanging Internet traffic with Google

Exchange Border Gateway Patrol (BGP) routes

Direct and Carrier Peering

Does not connect to external Internet

Useful for connecting directly to Google (not just GCP)

Also save on egress fees

10Gbps per link (direct), variable for carrier

Cloud VPN

Site to site VPN connection over IPSec

Connect internal network to GCP over encrypted tunnel over public Internet

Up to 1.5 Gbps per tunnel

Can use multiple tunnels for increased performance

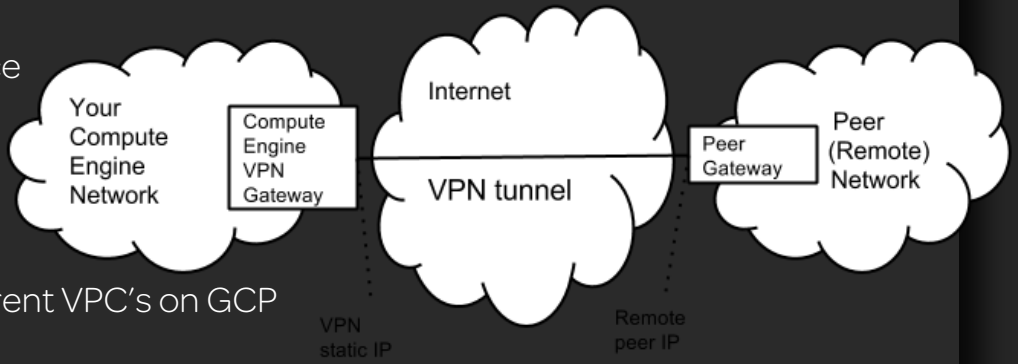
- example: transfer legacy resources to GCP

Static and dynamic routes (using Cloud Router)

Supports IKEv1 and IKEv2 using shared secret

Connect on-premises to GCP or connect two different VPC's on GCP

No site to client option available



Breakdowns

CONNECTION	ACCESS TYPE	CAPACITY	COST	OTHER CONSIDERATIONS
Dedicated Interconnect				
Dedicated, direct connection to VPC networks	Internal IP addresses in RFC 1918 address space	10 Gbps for each link	Reduced egress costs, fee for each link and VLAN	Requires you to have a connection in a Google supported colocation facility, either directly or through a carrier
IPsec VPN tunnel				
Encrypted tunnel to VPC networks through the public Internet	Internal IP addresses in RFC 1918 address space	1.5-3 Gbps for each tunnel	Egress is billed the same as general network pricing, fee for each tunnel	Requires a VPN device on your on-premises network

CONNECTION	ACCESS TYPE	CAPACITY	COST	OTHER CONSIDERATIONS
Direct Peering				
Dedicated, direct connection to Google's network	Public IP addresses	10 Gbps for each link	Settlement free peering, reduced cost for egress	Requires you to have a connection in a colocation facility, either directly or through a carrier provided wave service
Carrier Peering				
Peering through service provider to Google's public network (list of partners)	Public IP Addresses	Varies based on partner offering	Cost based on partner offering, reduced cost for egress	Requirements vary by partner

Decision tree

Google Certified Professional - Cloud Architect - Part 2

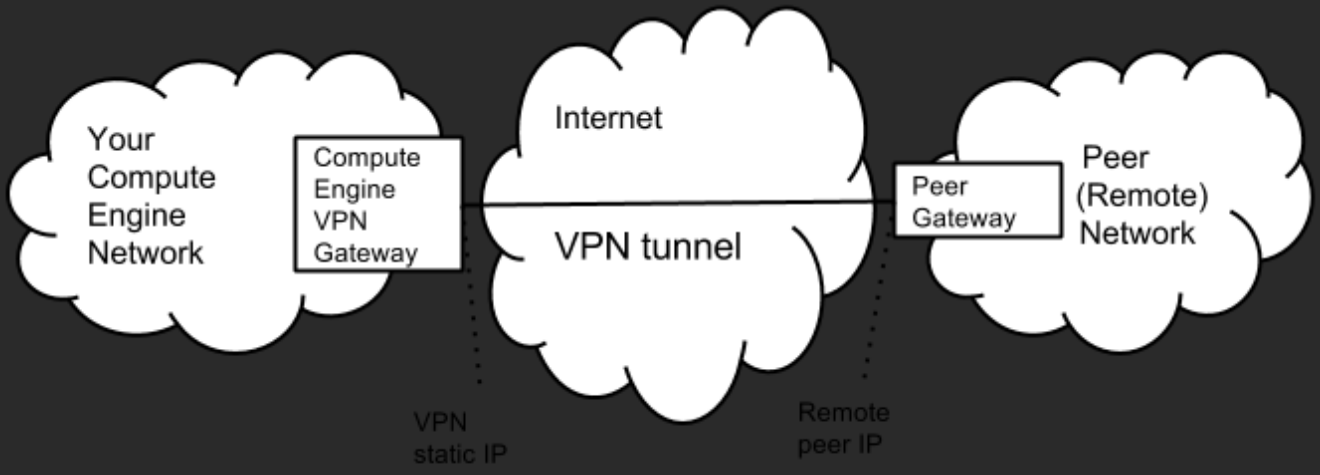
Cloud VPN

What is Cloud VPN?

Connect on-premises network to GCP Virtual Private Cloud (VPC)

IPsec connection over VPN over public Internet

Traffic encrypted by one gateway, then decrypted by other gateway



Cloud VPN traits

99.9% SLA

Site-to-site VPN only, no site-to-client (road warrior)

Up to 1.5Gbps per tunnel, can have multiple tunnels for increased performance

Static and dynamic routes (with Cloud Router)

Supports IKEv1 and IKEv2 using shared secret

Use cases

Connect to on-premises network – act as extension of your own network

- Example: network monitor server on GCE instance monitoring on-premises resources

Connect two different VPC networks on GCP

Requirements

VPN gateway on both ends (peer)

- Non-GCP = on-premises VPN server/router

Peer gateway must have static IP address

- If behind firewall, configure to pass ESP and IKE traffic

Non-conflicting CIDR range/subnet with rest of network

Cloud Router

Not required for VPN router, but makes things much easier

Static vs. dynamic routing

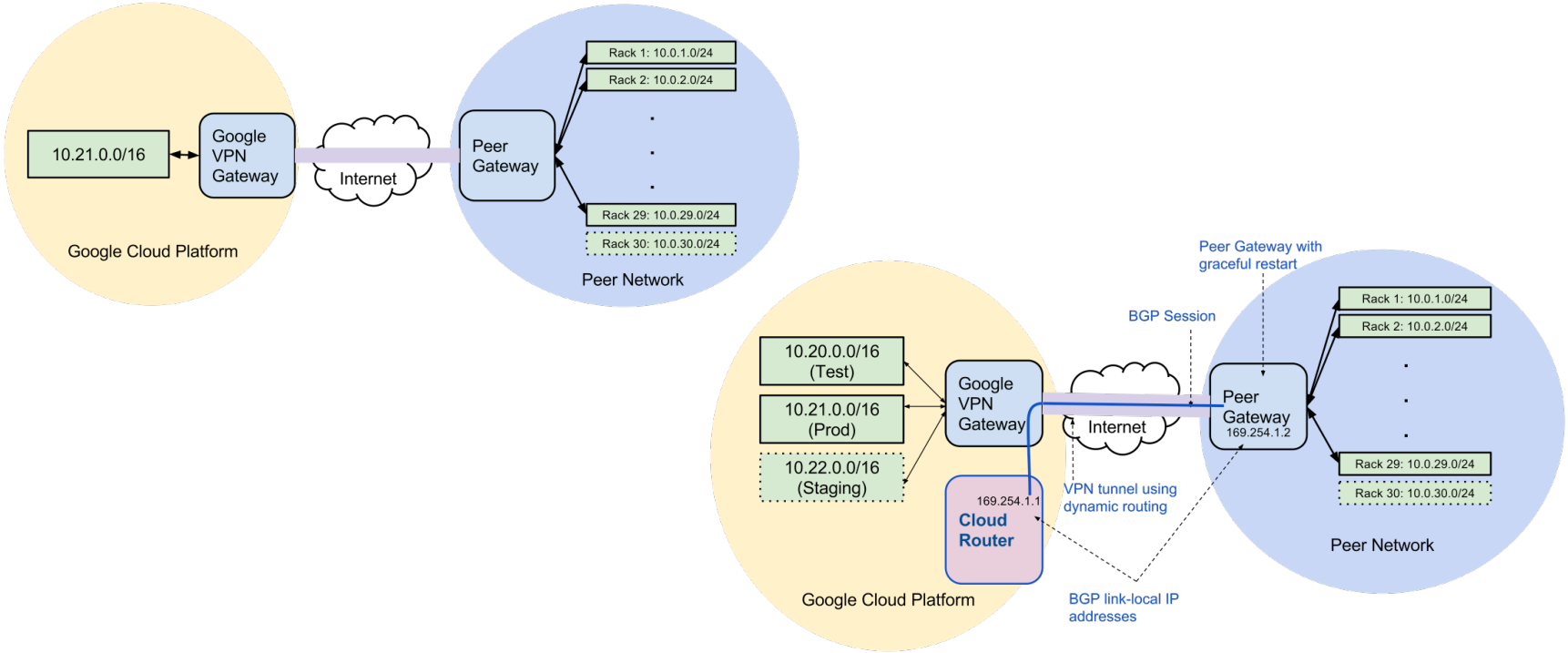
Static – create routing table for all existing and new routes. Can't re-route traffic if link fails

Dynamic – networks automatically discovery topology changes via BGP

- Can re-route traffic if link fails

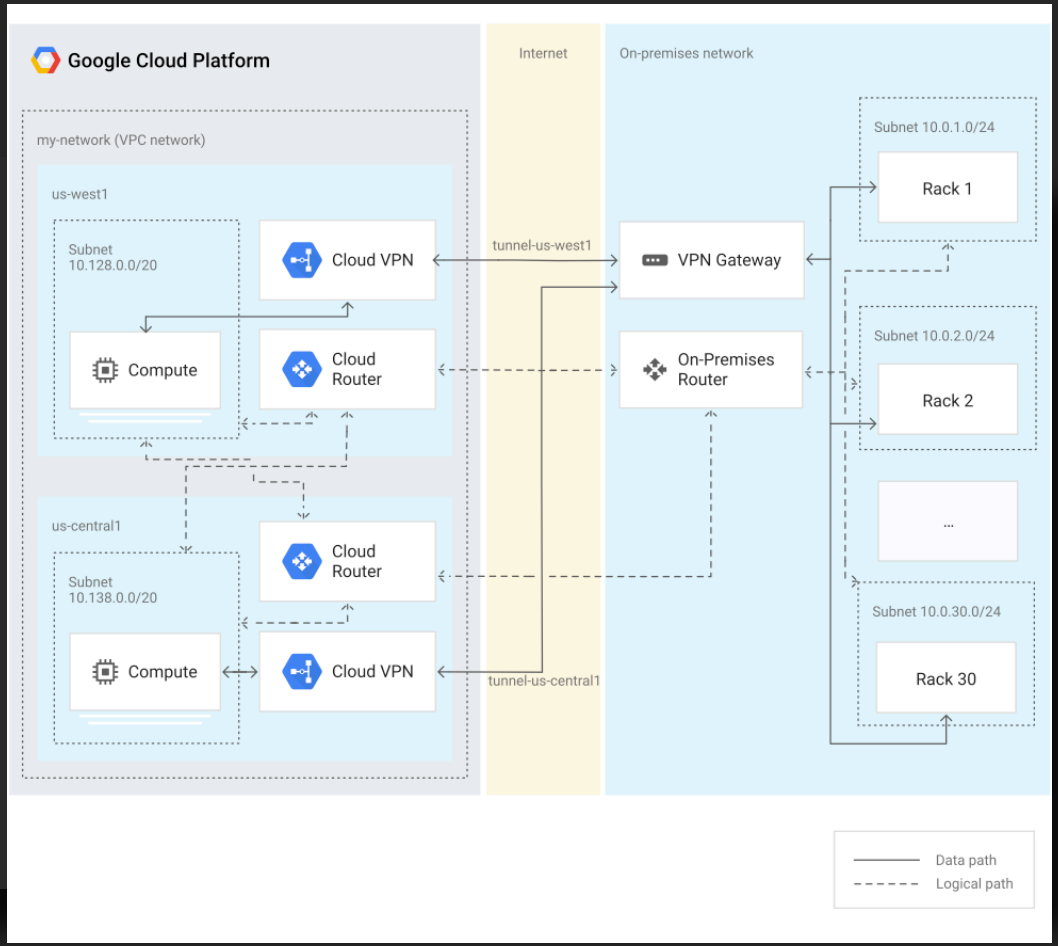
i.e. "the easy way"

Dynamic vs static



Multiple tunnels

Dynamic redundancy and rerouting



Google Certified Professional - Cloud Architect - Part 2

Virtual Private Cloud (VPC) Concepts

What Is VPC?

Software-Defined Network (SDN)

- Virtual version of traditional physical networks

Terms 'VPC' and 'network' are interchangeable

Central foundation of all other networking functions on GCP

Create subnetworks from single VPC

- Subnets are region-bound. Can have single subnet span multiple zones.

Global communications space, private communication among resources

Project based, but can share between projects with Shared VPC

Traditional networking concepts apply

- Firewalls, routes, load balancing, DNS, etc

Hybrid networking with on-premises networks with interconnect options

Quotas and Limits

Limit of 7000 virtual machines

- Cannot be raised via Quotas console
- If need more, create a new VPC network or call customer sales engineer
- No limit per subnet, just across entire VPC

IPV4 unicast traffic only

- No broadcast/multicast
- No IPV6 internally, but global load balancer IP's and traditional App Engine do support

Most other quotas can be increased by request

Network Tags

Primary method of segmenting network traffic access

Apply to firewall and network routes

Individual instances are tagged

Google Certified Professional - Cloud Architect - Part 2

Firewalls

Firewall basics

Single firewall for entire VPC

- Rules apply to entire VPC, connections allowed/denied at instance level

Manage both inbound (ingress) and outbound (egress) traffic

Implied 'deny all' ingress

Implied 'allow all' egress

Rules manage both external access and also access between internal resources

Firewall components:

- Directions – ingress/egress
- Source/destination
- Protocol/port
- Action – allow/deny
- Priority – order rules evaluated – first matching rule applied

Conditions for determining access

Source/target

Protocols

Ports

Tags on instances

The above can be optionally combined for granular access

Google Certified Professional - Cloud Architect - Part 2

Routes

Routing concepts

Software based – not limited by hardware

Routes traffic leaving VM's

Maps IP range to a destination – tells VPC network where to send packets destined for an IP address

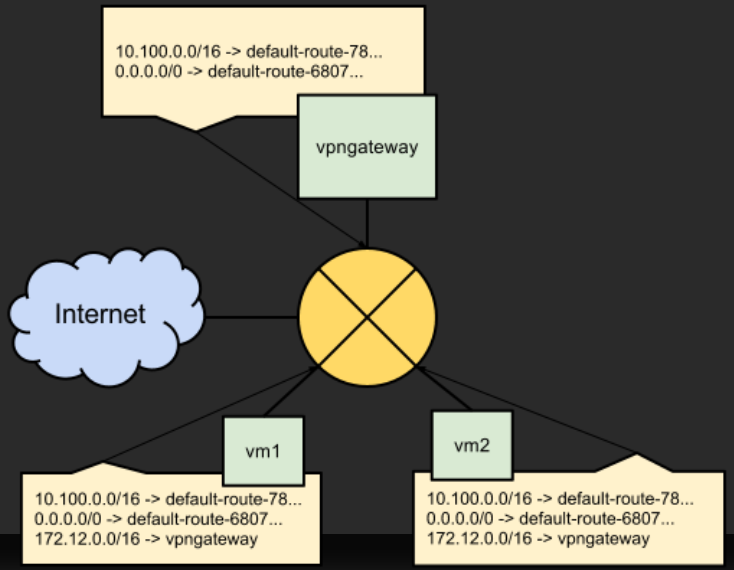
Default routes created for each subnet

Special cases for advanced routing setups

- Many-to-one NAT
- Proxy server

Routes can be network wide or by network tags

Routes + firewall rules combine to determine traffic access



Google Certified Professional - Cloud Architect - Part 2

Shared VPC

Overview

By default, VPC is tied to a single project

Need may exist to share network resources across projects

Shared VPC shares VPC across projects within an organization

i.e. Cross-Project Networking – also its former name

Concepts and terminology

Host project – project hosting the shared VPC

Service project – project with permission to shared VPC

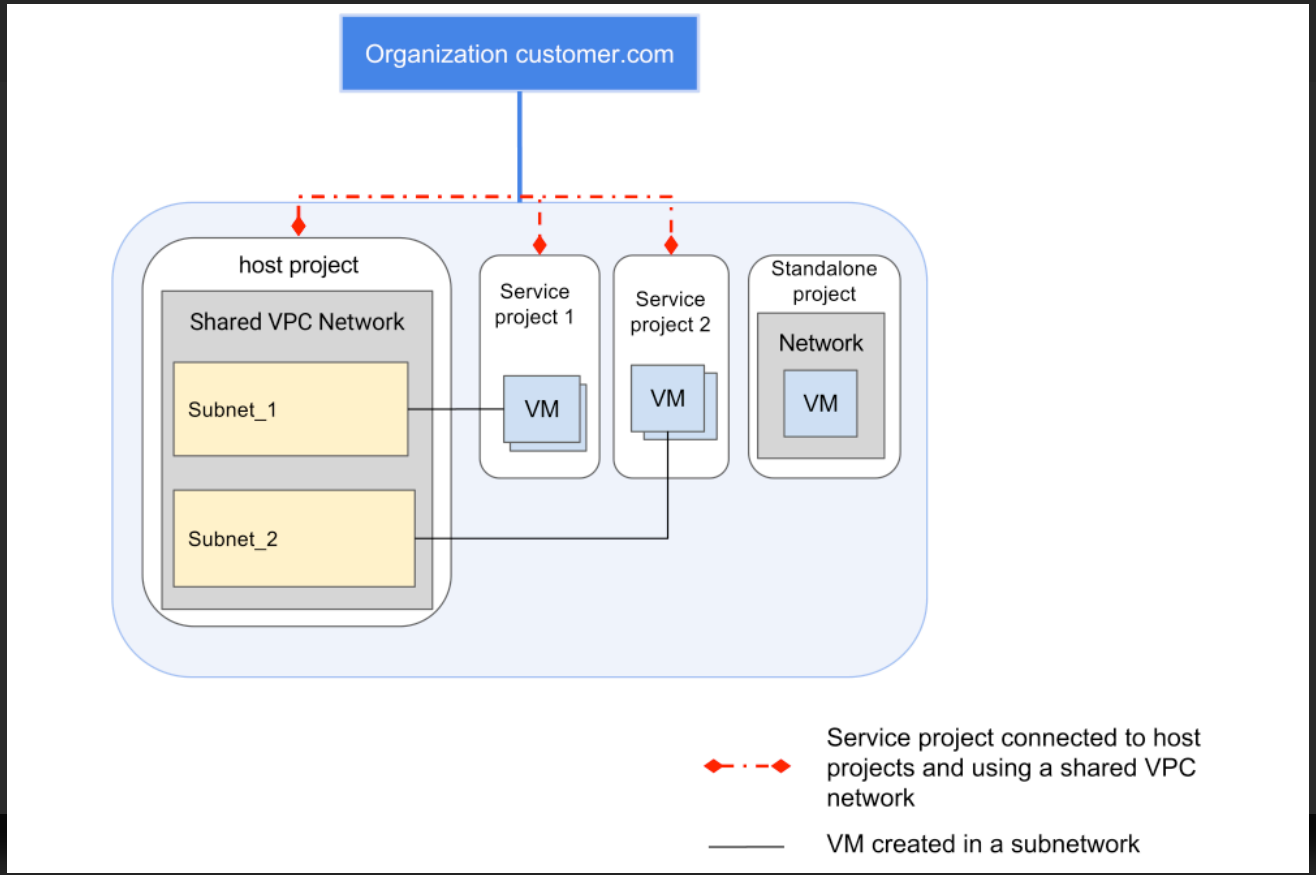
- Shared VPC projects can be controlled by different departments
- Ownership of resources in shared VPC maintained by project

Standalone project – project not using shared VPC

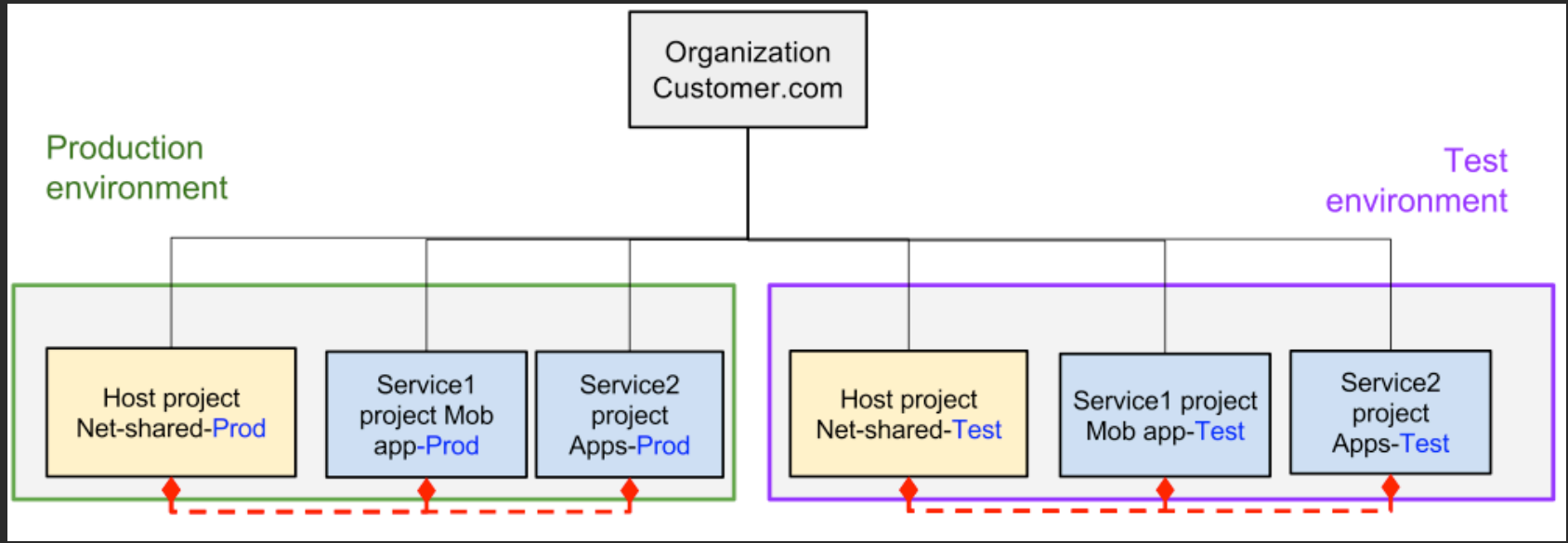
Shared VPC admin – IAM role for administrator of shared VPC

Service project admin – project admin of shared VPC service project

What it looks like



Separation of environments



Considerations

Only within single cloud organization

Service project can only link to single host project

Project cannot be both host and service project

Existing projects can use shared VPC, but existing instances cannot

Reserved (static) IP addresses tied to project that reserved it

Resources tied to shared VPC

Instances

Instance template

Instance groups

Forwarding rules for internal load balancing

Use cases

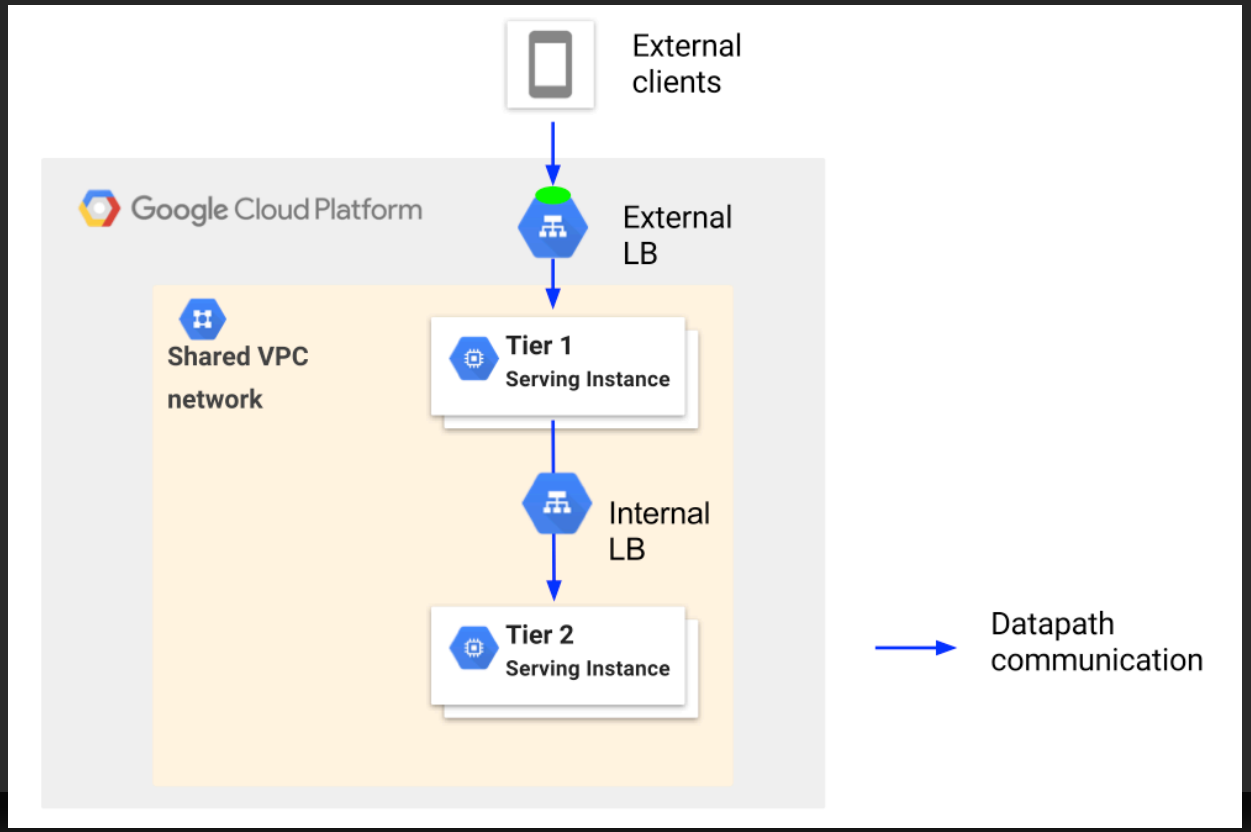
Separation of projects for access control/billing, but need access to same VPC environment

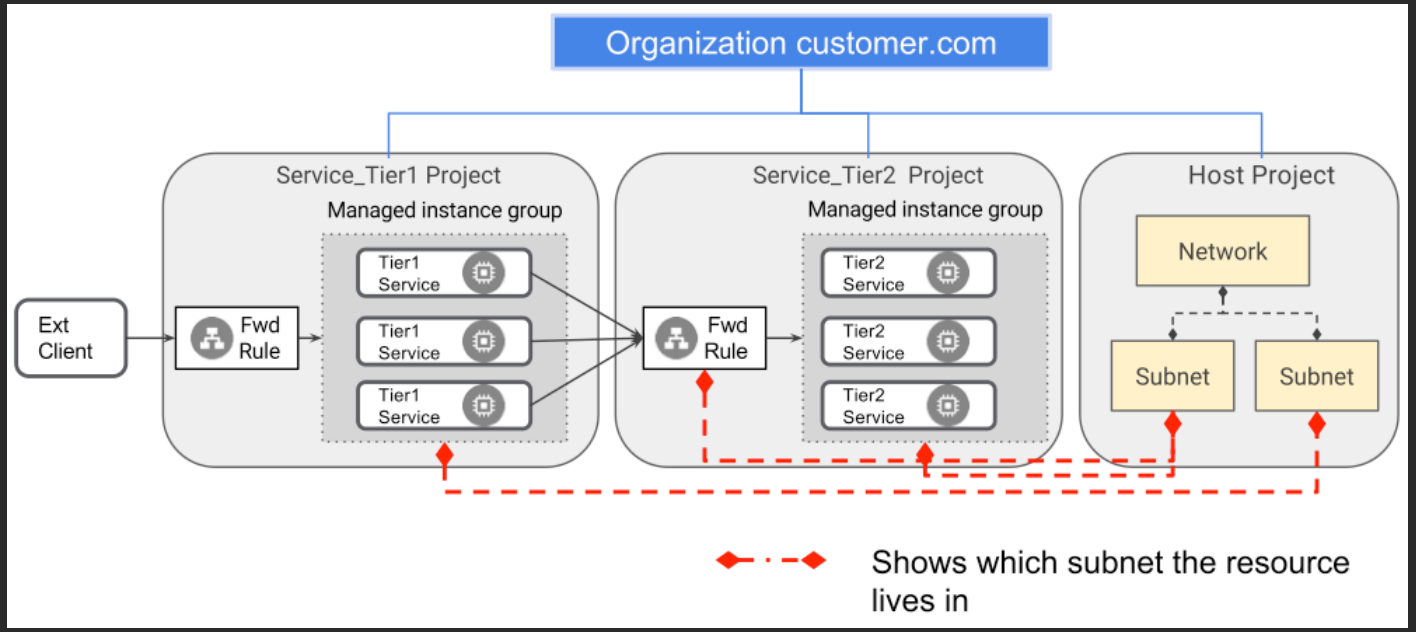
- Example: separate testing and production environments

Two tier web service

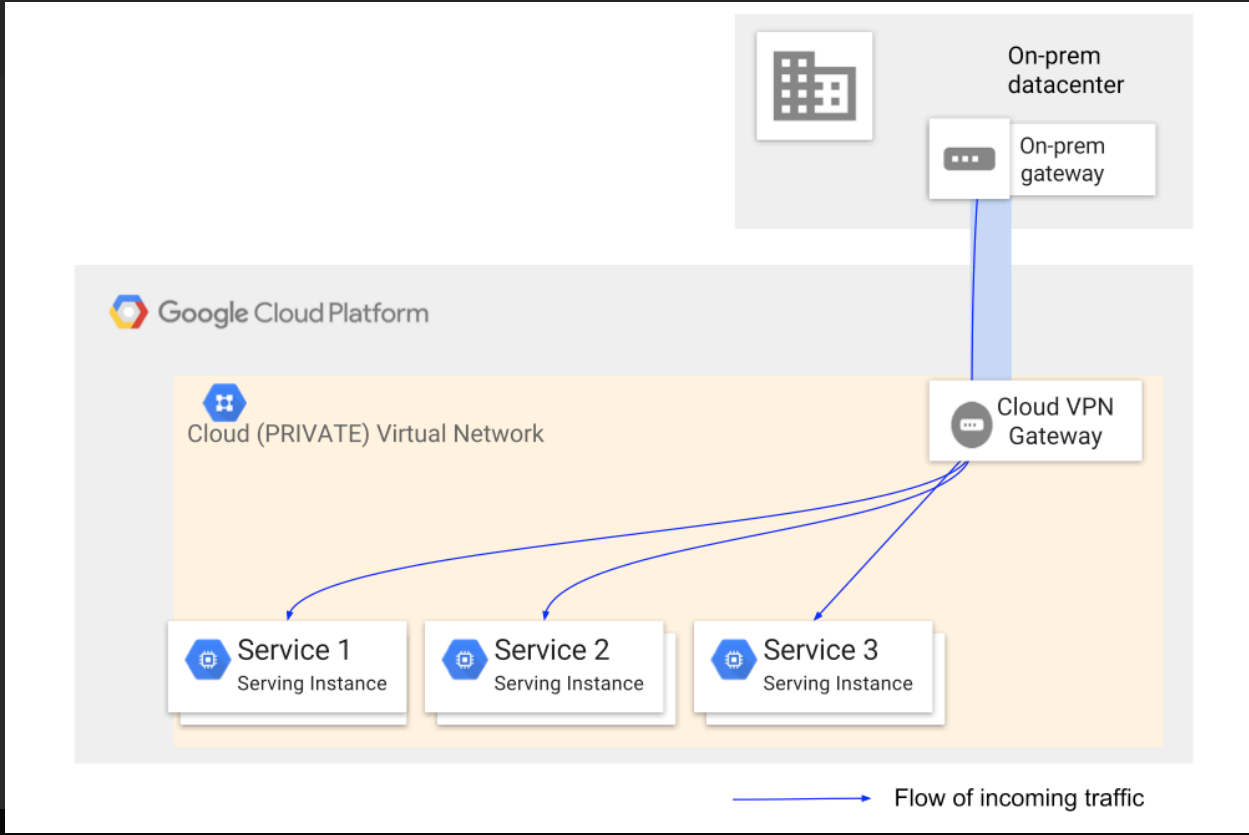
Hybrid cloud scenarios

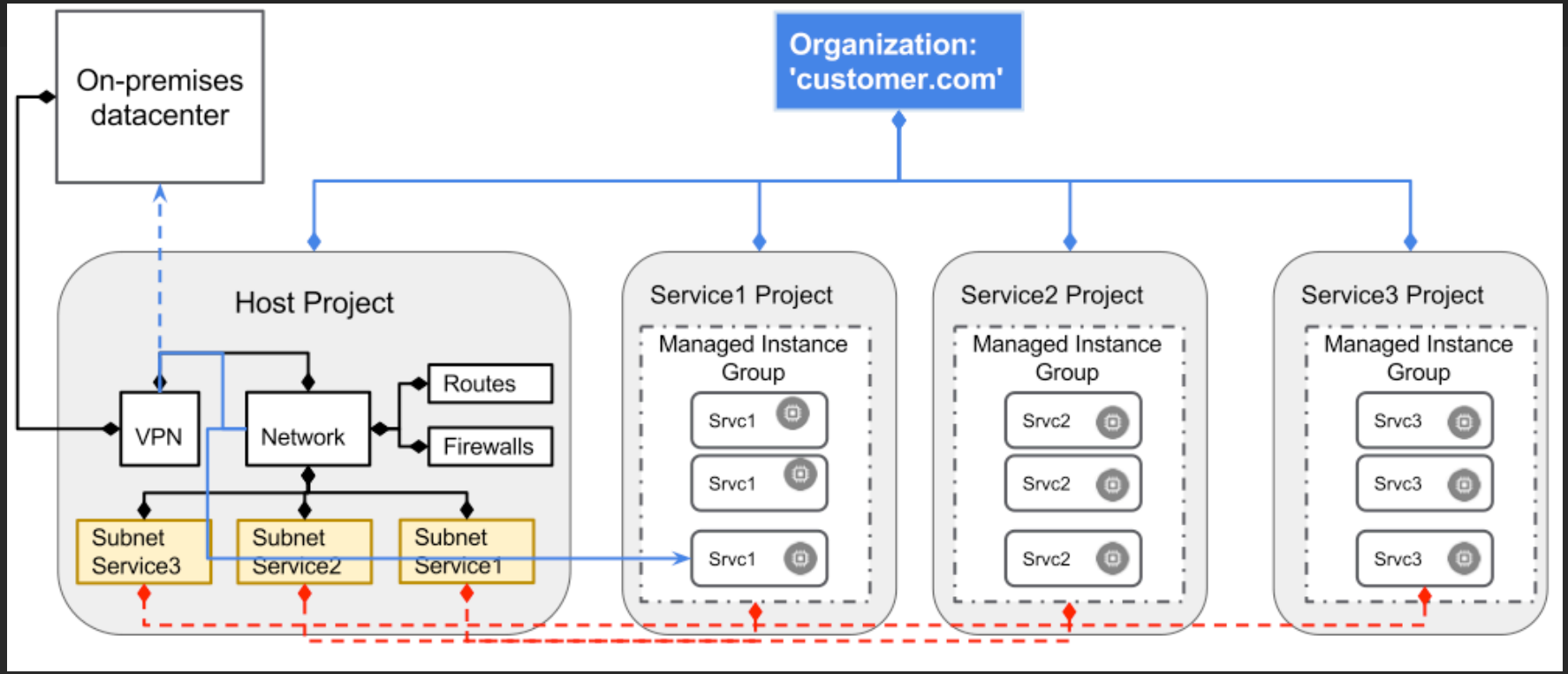
Two tier web service




Organization customer.com
Service_Tier1 Project
Managed instance group
Tier1 Service
Tier1 Service
Tier1 Service
Service_Tier2 Project
Managed instance group
Tier2 Service
Tier2 Service
Tier2 Service
Host Project
Network
Subnet
Subnet
Ext Client
Fwd Rule
Fwd Rule
Shows which subnet the resource lives in

Hybrid cloud




On-premises datacenter
Organization: 'customer.com'
Host Project
VPN
Network
Routes
Firewalls
Subnet Service3
Subnet Service2
Subnet Service1
Service1 Project
Managed Instance Group
Srvc1
Srvc1
Srvc1
Service2 Project
Managed Instance Group
Srvc2
Srvc2
Srvc2
Service3 Project
Managed Instance Group
Srvc3
Srvc3
Srvc3

IAM Roles

Organization Admin

Shared VPC Admin - compute.xpnAdmin

- Organization level role
- Configure shared VPC
- Associate service projects with host projects
- Grant Network User role

Network User – compute.networkUser

- Project level role
- Create resources to use shared VPC
- Discover shared VPC assets
- Requires project admin role (Project Owner, Editor, Compute Engine Admin)

Google Certified Professional - Cloud Architect - Part 2

Compute Engine Deep Dive

Why are we focusing on this so much?

VM's are the heart and soul of GCP

GCE, GKE, and GAE all run on VM's

Substantial portion of exam will be VM focused

Start with single VM's

Move up to 'force multipliers' – the true 'magic' of cloud computing

Automation

Auto-scaling

Managed instance groups

Load balancers

In this deep dive

Assume you know the basics

Custom images

Disk manipulation

Metadata

Startup/shutdown scripts

Snapshots

Persistent disks

gcloud commands

Prepare you to work with ‘force multiplier’ concepts

Google Certified Professional - Cloud Architect - Part 2

Disks

DISK OPTIONS

ALL INSTANCES HAVE A **SINGLE ROOT DISK** FOR THE **OS**

1. **PERSISTENT DISK**
   - **Most common,** default option
   - **NOT directly attached**
   - **Standard** and SSD variety
2. **LOCAL SSD**
   - **Directly attached** to your VM
3. **CLOUD STORAGE BUCKETS**
   - **Niche** option
   - **High collaboration**, 'infinite' space

① PERSISTENT DISK

② LOCAL SSD

③ CLOUD STORAGE BUCKETS

DURABLE + RELIABLE	DEFAULT option
DATA DISTRIBUTED	SEVERAL physical disks redundancy + performance
NOT PHYSICALLY ATTACHED	INDEPENDENT from the VM instance CAN DETACH /move disks PRESERVE DATA after deleting instances MODULAR – resize, move, attach additional disks
PERFORMANCE SCALES WITH SIZE	
EASE OF USE	No RAID configuration necessary Partitioning disk not encouraged – just resize or add additional disks
SSD OPTION AVAILABLE	
ENCRYPTED	EITHER WITH GOOGLE provided keys OR bring your own

① PERSISTENT DISK

- DURABLE + RELIABLE
- DATA DISTRIBUTED
- NOT PHYSICALLY ATTACHED
- PERFORMANCE SCALES WITH SIZE
- EASE OF USE
- SSD OPTION AVAILABLE
- ENCRYPTED

② LOCAL SSD

- TRADE PERFORMANCE DURABILITY + RELIABILITY
- PHYSICALLY ATTACHED TO VM
- CANNOT BE BOOT DEVICE
- HIGHEST PERFORMANCE
- MUST CREATE ON INSTANCE CREATION
- CAN ATTACH BOTH LOCAL SSD + PERSISTENT DISK
- ENCRYPTED
- BEST PRACTICE
- 375 GB IN SIZE

Not automatically replicated
All data lost if instance terminated
Still migrate-able

BUT with caveats

GOOGLE SUPPLIED - cannot use your own

FAST SCRATCH DISK
– replicate workload across multiple instances

NON-CONFIGURABLE – can attach up to 8

③ CLOUD STORAGE BUCKETS

① PERSISTENT DISK

- DURABLE + RELIABLE
- DATA DISTRIBUTED
- NOT PHYSICALLY ATTACHED
- PERFORMANCE SCALES WITH SIZE
- EASE OF USE
- SSD OPTION AVAILABLE
- ENCRYPTED

② LOCAL SSD

- TRADE PERFORMANCE DURABILITY + RELIABILITY
- PHYSICALLY ATTACHED TO VM
- CANNOT BE BOOT DEVICE
- HI
- MUST CREATE ON INSTANCE CREATION
- CAN AT
- SSD + PERSISTENT DISK
- ENCRYPTED
- BEST PRACTICE
- 375 GB IN SIZE

③ CLOUD STORAGE BUCKETS

- **MOST FLEXIBLE, SCALABLE, AND DURABLE STORAGE OPTION**
- **NOT A ROOT DISK**
- **GLOBAL ACCESIBILITY VS. ZONE FOR OTHER DISKS**
- **LOWER PERFORMANCE**
- **ENCRYPTED**

INSTANCES IN MULTIPLE REGIONS/ZONES
can write to same bucket

THAN other disk options

BOTH Google supplied + use your own

STORAGE OPTIONS OVERVIEW

	STANDARD PERSISTENT DISK	SSD PERSISTENT DISK	LOCAL SSDs	CLOUD STORAGE BUCKETS
Storage Type	EFFICIENT + reliable block storage	FAST + reliable block storage	HIGH PERFORMANCE local block storage	AFFORDABLE object storage
Max. Space Per Instance	64 TB	64 TB	3 TB	ALMOST INFINITE
Scope Of Access	Zone	Zone	Instance	Global
Data Redundancy	Yes	Yes	No	Yes
Encryption At Rest	Yes	Yes	Yes	Yes
Custom Encryption Keys	Yes	Yes	No	Yes
Machine Type Support	ALL Machine types	ALL Machine types	MOST Machine types	ALL Machine types
Zone Availability	ALL Zones	ALL Zones	ALL Zones	ALL Zones

PERFORMANCE

	STANDARD PERSISTENT DISK	SSD PERSISTENT DISK	LOCAL SSD (SCSI)	LOCAL SSD (NVMe)
	Maximum Sustained IOPS			
Read IOPS per GB	0.75	30	266.7	453.3
Write IOPS per GB	1.5	30	186.7	240
Read IOPS per Instance	3,000	15,000–40,000*	400,000	680,000
Write IOPS per Instance	15,000	15,000–30,000*	280,000	360,000
	Maximum Sustained Throughput (MB/s)			
Read throughput per GB	0.12	0.48	1.04	1.77
Write throughput per GB	0.12	0.48	0.73	0.94
Read throughput per instance	180	240 – 800*	1,560	2,650
Write throughput per instance	120	240 – 400*	1,090	1,400

PRICING

TYPE	PRICE(PER GB / MONTH)
Standard Persistent Disk	$0.040
SSD Persistent Disk	$0.170
Local SSD (Min. 375 GB disk)	$0.080
Snapshot Storage	$0.026

Google Certified Professional - Cloud Architect - Part 2

Custom Images

Differences between images and snapshots

Images

- Purpose - create new instances, configure instance templates
- Recommended to shut down instances before creating new image
- Access across projects

Snapshots

- Purpose - periodic incremental backup of existing disk/instance
- Can create while running
- Access only from within same project

Create from several sources

Persistent disk

Another image in same project

Imaged shared from another project

Compressed image from Cloud Storage

Managing custom images

Image families simplifies image versioning

- Useful for instance templates and scripts

Groups related images together

- Roll forward and back between image versions
- Family always points to newest non-deprecated version

Deprecating images

As custom images are continually updated, need to retire older versions

Transition users away from older unsupported versions in manageable way

Deprecation states

- Deprecated – still works but gives warning
- Obsolete – new users cannot use it - error if attempt to use, existing links still work
- Deleted – all users cannot use it
- Active – mark deprecated image as active again (command line only)

Sharing and moving images

Share across projects

Requires Compute Engine Image User role to host project

- Example: User in Project A wants to use images from Project B
- User in Project A must have Compute Engine Image User role granted for project B
- Role grants access to all images in project

For managed instance groups, Project A service account must be granted role to Project B

Export image to Cloud Storage

Ideal for sharing with projects without host project access

Export image as a tar.gz to Cloud Storage

Linux only, not available for Windows

Sharing with Image User Role is preferable

Google Certified Professional - Cloud Architect - Part 2

Snapshots

What is it?

Simply put: backups

Periodic backups via a point in time snapshot of disk

Can create while instance is running

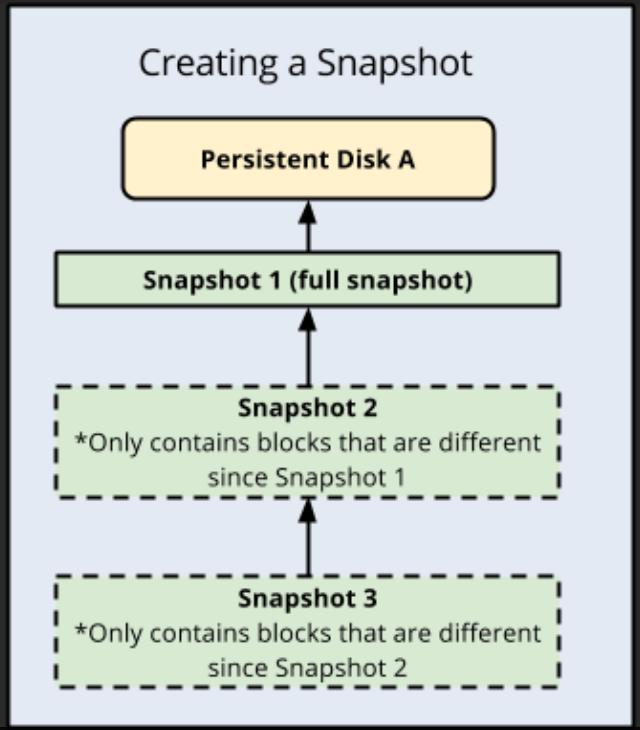
Cannot share across projects

Can create instance copies in new zones

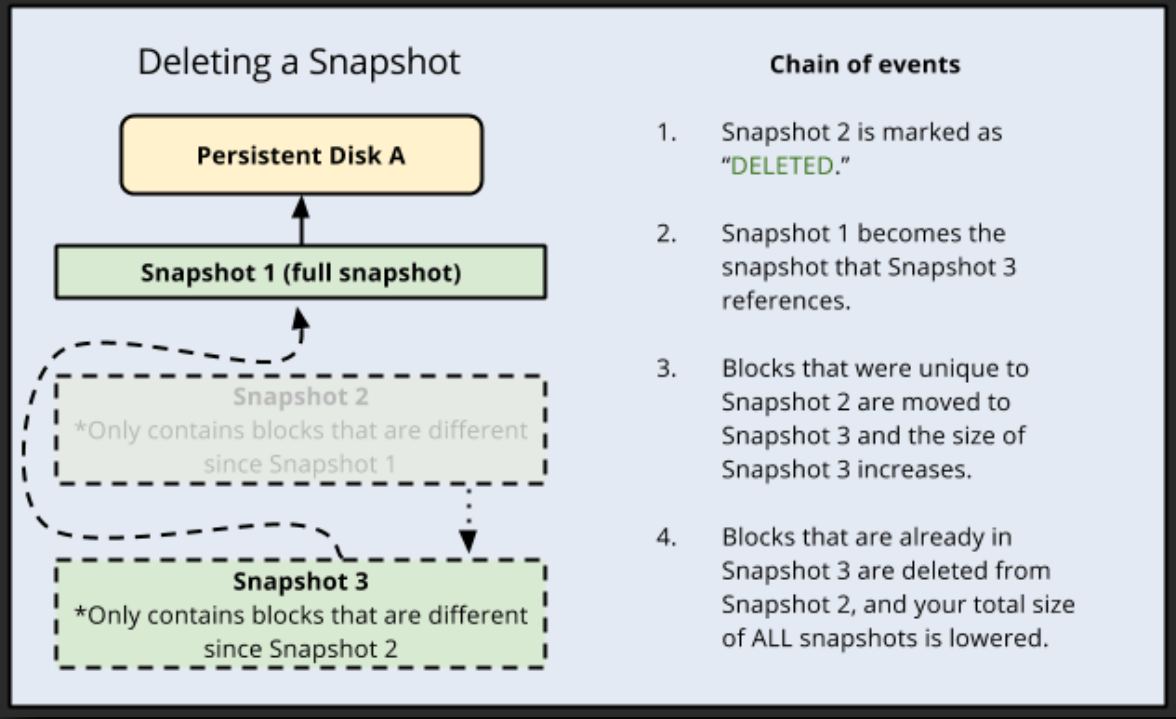
Can snapshot boot disk or attached disks

How it works

Incremental backups



Deleting snapshots



Best practices

Prepare disk for best consistency

Pause applications/processes that write data, then flush disk buffers

If possible, unmount disk completely

For Windows, use VSS Snapshots

Use ext4 for Linux

Take only one snapshot at a time per disk

Schedule during off-peak hours

Use multiple persistent disks for large data volume

Run fstrim before snapshot (Linux) to clean up space

Google Certified Professional - Cloud Architect - Part 2

Startup and Shutdown Scripts

Why is this important?

Automation! Automation! Automation!

Eases management of large number of VM's

Automate software installation, updates, services, and much more

Easily and programmatically customize VM's

Key component in instance groups and scaling capabilities

Considerations

Always run as root/administrator

Can run whatever script types OS recognizes (bash, Python, .bat files)

Compute Engine will run the script verbatim, regardless of type

Input methods

Direct input

Paste in the script field in instance properties

Link to script on Google Cloud Storage

Using metadata server URL

Very useful for large scale automation

Must have access to bucket/object

Shutdown scripts

Best-effort basis

Run during shutdown period. May shut down before script completes.

Great for managed instance group/autoscaler

Example: copy processed data to Cloud Storage, back up logs

Good to pair with preemptible instances

Metadata server

Built into GCP

Manage configuration and environment variables programmatically

Default and custom values

Key/value pair

Google Certified Professional - Cloud Architect - Part 2

Force Multipliers – Elastic Computing

So far...

Focus on individual VM's

Take concepts, and apply to managing many VM's

Force multiplier – Scalable, automatic

Why is automation important?

Repeatable and documented

Easily re-deploy resources if needed

Scalable

Grow and shrink as needed based on demand

Necessary for large architecture

Reduce complexity

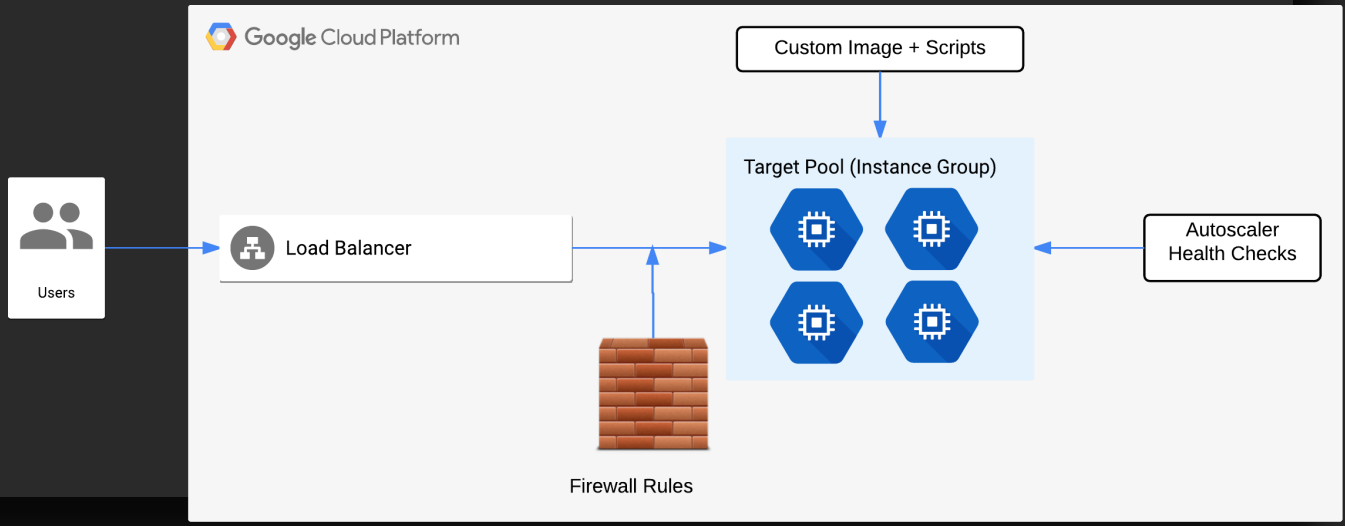
Putting it all together

Use previous concepts

Firewall rules, scripts, custom images, etc

Combine with scalable components

Load Balancers, Instance Groups, Autoscaler



For this section

Three components – one purpose

Load Balancer

Instance Groups

Autoscaling

Go over concepts on each one individually, then combine them

Google Certified Professional - Cloud Architect - Part 2

Load Balancers

What is a load balancer?

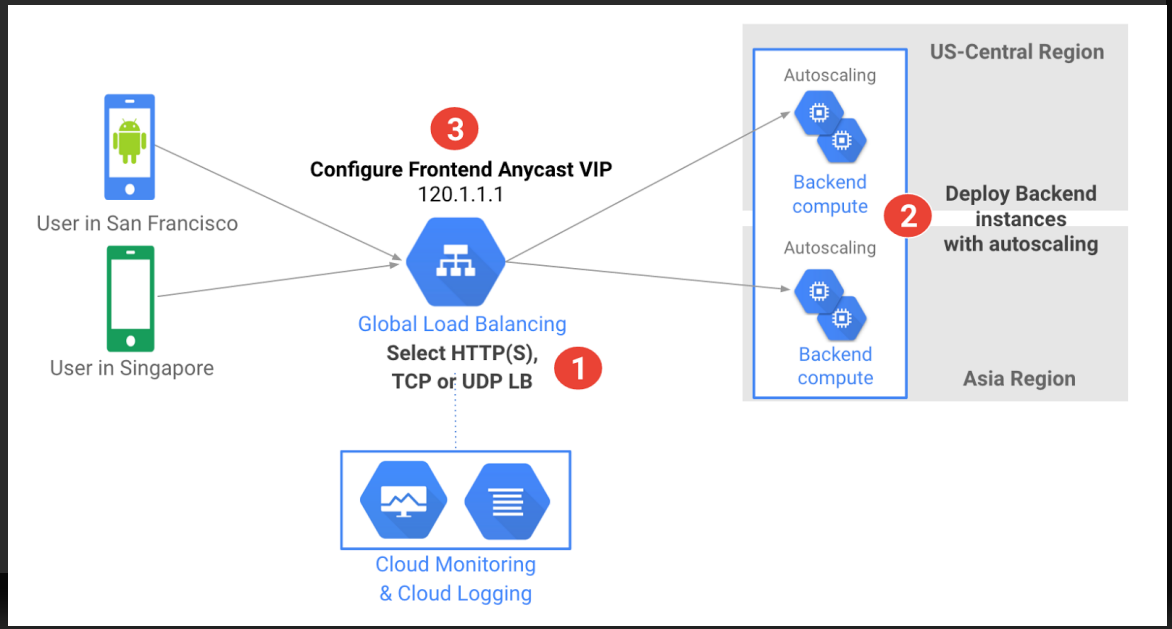
Distributes (balances) user network requests among a pool of instances

Single frontend point of access – multiple backend targets to server traffic

Software Defined – not physical

Global or regional in scope

Traffic subject to firewall rules



Load Balancer types

Global External Load Balancer

HTTP(S) Load Balancer

SSL/TCP Proxy

Regional External Load Balancer

Network Load Balancer (TCP/UDP)

Regional Internal Load Balancer

Internal Load Balancing

HTTP(S) Load Balancer

Manages HTTP(S) requests

Global scope

Distribute traffic to closest region

IPv4 and IPv6 – IPv6 terminated at LB then proxy by IPv4 to backend

Distribute traffic by location or content requested

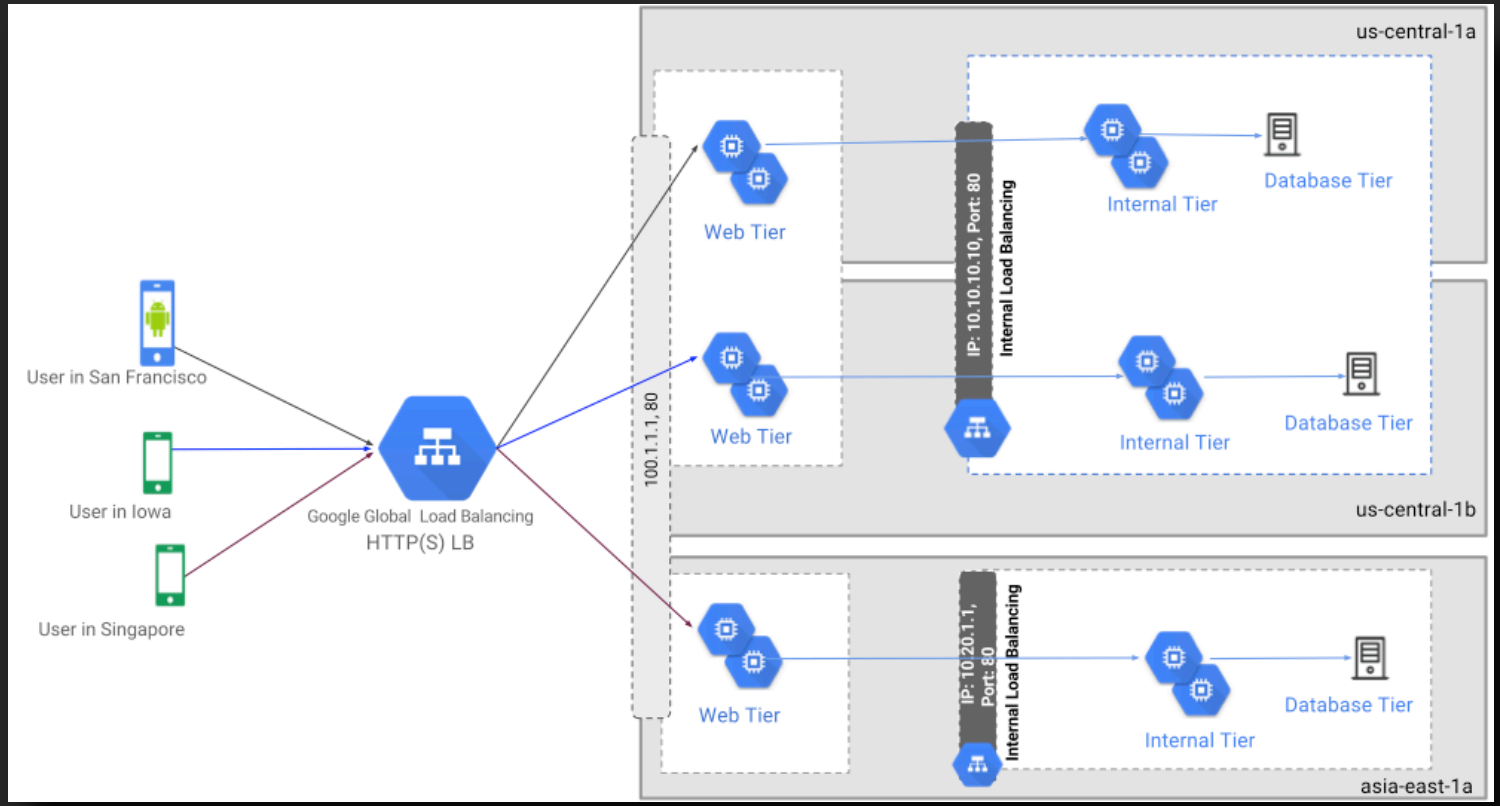
Forwarding rule – forwards traffic to target pool by matched criteria (location, content)

Forwards to target pool

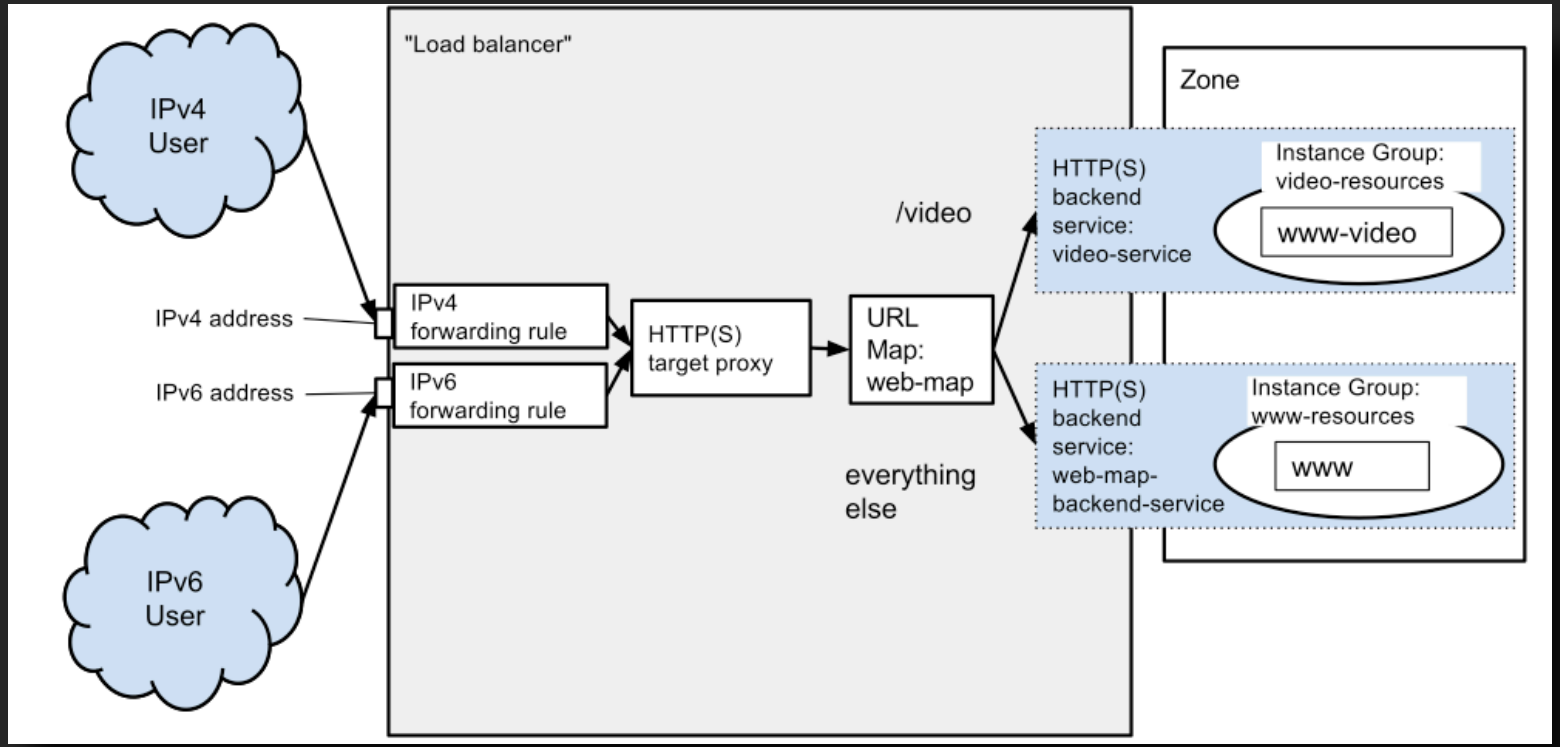
Paired with instance group for backend

Native support for websocket protocol

Location-based



Content-based



Network (External) Load Balancer

Regional External

non-HTTP(S) protocols

Balance requests by IP protocol data (address, port, protocol type)

How it works:

Forwarding Rules – matched criteria = address, protocol, port range

Target Pool – group of VM's (usually an instance group)

Network (Internal) Load Balancer

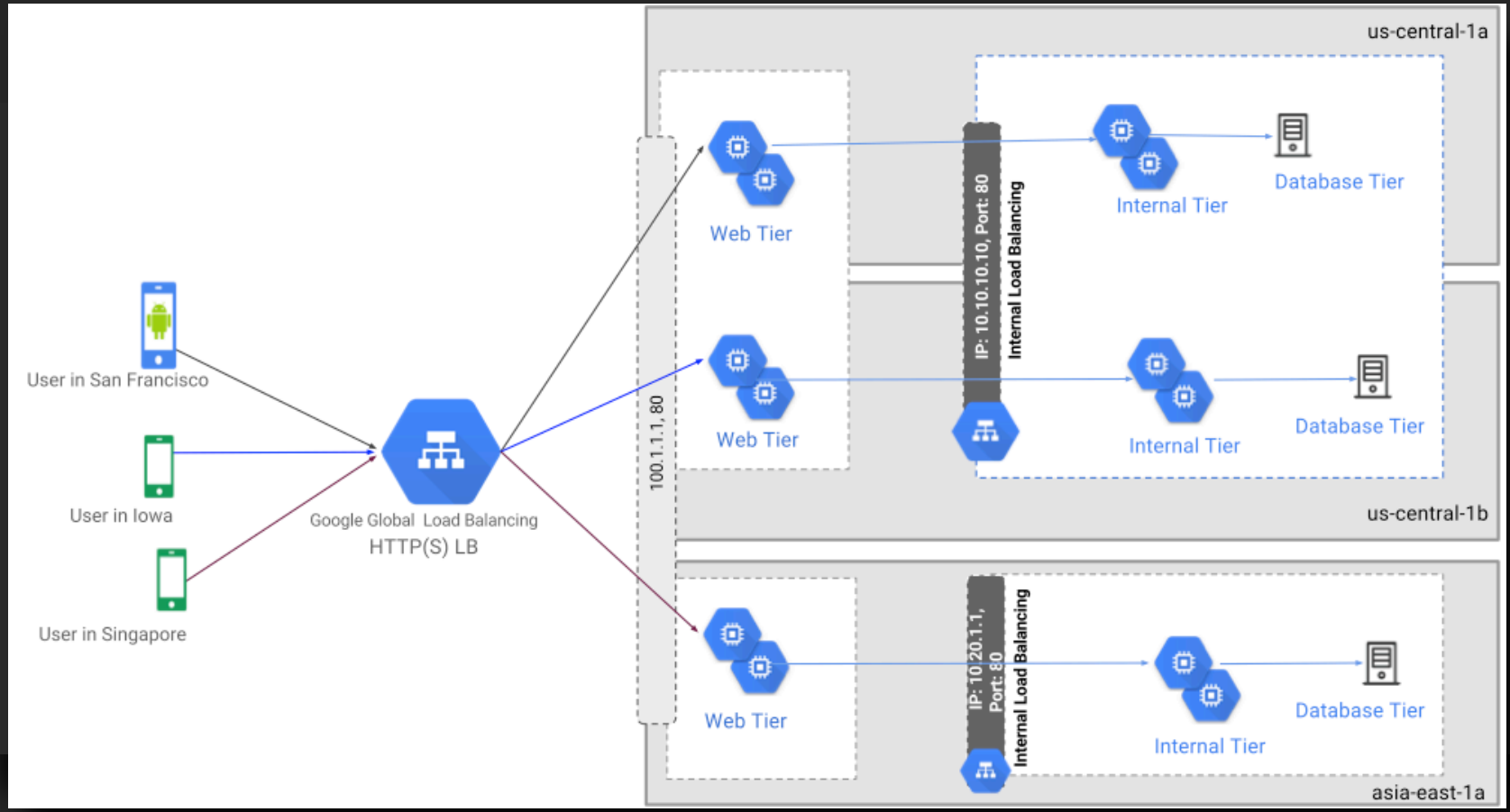
Regional Internal

Private load balancer within same VPC

Same option as Network Load Balancer, with option for internal only

Often used with multi-tier application – nested LB's

Affects Cloud Router dynamic routing


us-central-1a
Web Tier
IP: 10.10.10.10, Port: 80
Internal Load Balancing
Internal Tier
Database Tier
User in San Francisco
User in Iowa
User in Singapore
Google Global Load Balancing
HTTP(S) LB
100.1.1.1, 80
Web Tier
Internal Tier
Database Tier
us-central-1b
Web Tier
IP: 10.20.1.1, Port: 80
Internal Load Balancing
Internal Tier
Database Tier
asia-east-1a

Google Certified Professional - Cloud Architect - Part 2

Instance Groups and Autoscaling

What is it?

Group of instances

Manage as a group, not one at a time

Managed and unmanaged varieties

Managed preferred, and what we'll cover

Features

Automatically scale

Work with load balancers

Health checks – auto-healing groups

How it works

Require Instance Template

Defines group configuration

- machine type, zone, image, scripts

Re-usable for multiple group configurations

Global – not region bound

- Can specify zonal resources (i.e. read only disk), which effectively binds it

From template – create managed instance group

Networking

Subject to firewall rules for allowed traffic

Essential for load balancers

Load balancers must be assigned to a backend – target pool or instance group

HTTP Load balancers must use instance group

Load balancer contains one backend service

Backend service links to one or more backends

Backend links to one instance group

Backend service knows which backends to use – directs traffic

Health checks

Auto-healing

If an instance fails or service fails – delete and recreate identical instance

Managed instance groups only

Updating managed instance groups

Managed Instance Group Updater (Beta)

Update entire group – not just individual machines

Deploy new versions of software

Control pace of update rollout

- Rollout happens automatically
- Can do partial rollouts for canary testing

Deploy inside existing managed instance group

Autoscaling

Automatically scales instance group

Managed instance group only

Automatically creates more instances when demand increases

Automatically remove instances when demand decreases

Set by autoscaling policy

Set metric and threshold

- Average CPU, HTTP requests, Pub/Sub queue workload, Stackdriver Monitoring metrics
- When average of available instances meets threshold, more instances created

Set minimum and maximum instance count

Google Certified Professional - Cloud Architect - Part 2

Cloud Deployment Manager

What is Cloud Deployment Manager?

Infrastructure deployment service

Automates creation/management of GCP resources

Create and manage resources with configuration files and templates

Why is it important?

As infrastructure grows in size and complexity, so does the chance of human error

Standardized and repeatable

Create resources over and over with repeatable results

Highly structured templates and configuration

Document infrastructure in easy to understand format

Used by Cloud Launcher to create easy, one-click deployments

How it works

Deploy with command line only

Infrastructure as code

Calls on API resources

Configuration file – YAML format

Lists each resource to create and its properties

Contains resources section followed by list of resources

Resource components

- Name – user-defined string to identify (my-deployment-project)
- Type – type of resource to deploy (compute.v1.instance, compute.v1.disk)
- Properties – resource parameters (zone: us-central1, boot: true)

Resource Type
appengine.v1.version
appengine.v1beta4.version
appengine.v1beta5.version
bigquery.v2.dataset
bigquery.v2.table
bigtableadmin.v2.instance
bigtableadmin.v2.instance.table
cloudfunctions.v1beta2.function
cloudresourcemanager.v1.project
clouduseraccounts.beta.group
clouduseraccounts.beta.user
compute.beta.address
compute.beta.autoscaler
compute.beta.backendBucket
compute.beta.backendService
compute.beta.disk

Templates

Configuration file can contain templates

Separate configurations into smaller chunks

Update and re-use

Python or Jinja2 format

Advantages:

Easier to manage and maintain

Reusable

Keep consistent definitions in one place

Manifest

Read only output of final configuration

Includes configuration YAML, imported templates, expanded resource list

When troubleshooting, consult the manifest

Exam perspective

Mostly high level understanding needed

Not tested on python/jinja knowledge

May need to troubleshoot manifest/configuration file

```
resources:
- name: {{ env["deployment"] }}-app-ig
  type: compute.beta.instanceGroupManager
  properties:
    baseInstanceName: {{ env["deployment"] }}-app
    zone: us-central1-f
    instanceTemplate: $(ref.{{ env["deployment"]
}}-app-template.selfLink)
    targetSize: 3
- name: {{ env["deployment"] }}-app-disk
  type: compute.v1.disk
  properties:
    zone: us-central1-f
    sizeGb: 20
- name: {{ env["deployment"] }}-app-template
  type: compute.v1.instanceTemplate
  properties:
      properties:
        machineType: n1-standard-1
        networkInterfaces:
        - network: default
          subnetwork: default
          accessConfigs:
          - name: External NAT
            type: ONE_TO_ONE_NAT
        disks:
        - autoDelete: true
          boot: true
          deviceName: app-vm-tmpl-boot-disk
          initializeParams:
            diskSizeGb: 10
            diskType: pd-standard
            sourceImage:
projects/ubuntu-os-cloud/global/images/family/ubuntu-1604-lts
          type: PERSISTENT
        - deviceName: {{ env["deployment"] }}-app-disk
          type: PERSISTENT
          source: $(ref.{{ env["deployment"] }}-app-disk.selfLink)
```

Google Certified Professional - Cloud Architect - Part 2

GKE/GAE Exam Perspective

Focus of Cloud Architect Exam

Infrastructure

Build it

Manage it

Best practices

GKE/GAE are Different

Managed infrastructure

With GKE/GAE infrastructure is built and maintained for you

More developer/code focused

From an Exam Perspective...

Higher level understanding of GKE/GAE

When to choose one of them over other options

Desire for managed services

However, there are some mechanics that may be tested

Google Certified Professional - Cloud Architect - Part 2

GCP Container Resources

Why this matters

Exam will test high level understanding of GCP container resources

Container Builder – “build it”

Container Registry – “store it”

Google Kubernetes (Container) Engine – “run it”

Container Builder – “build it”

Create Docker container images from source code

Pull code from multiple locations:

Google Cloud Storage

Google Cloud Source Repositories

GitHub

BitBucket

Created images automatically stored in Container Registry

Can deploy to GKE, GCE, or GAE (Flexible)

Or any other service that runs Docker containers

Container Registry – “store it”

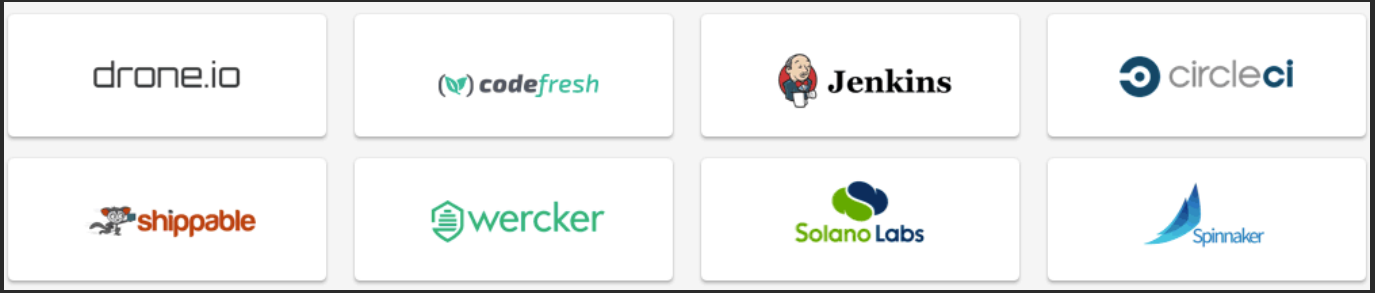
Private Docker repository

Integrate with GCP and external container services

Supports CI/CD model

Push images to the registry

Pull images from registry



Google Kubernetes Engine – “run it”

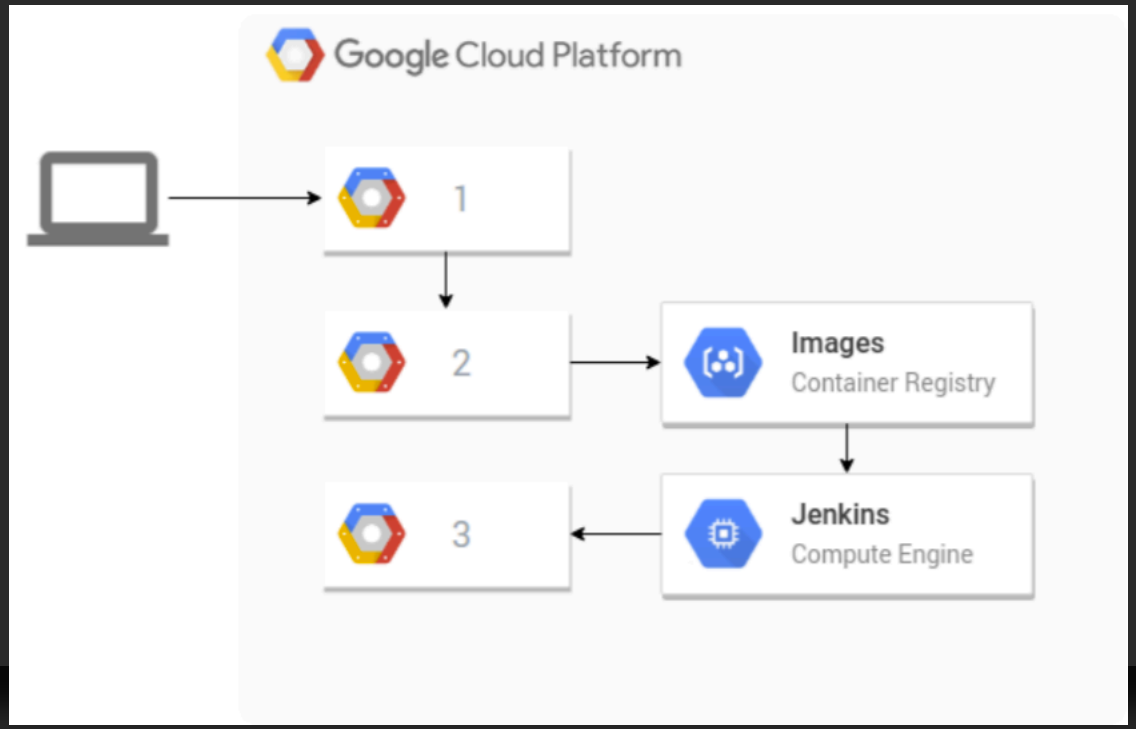
Managed Kubernetes orchestration service

“Kubernetes the easy way”

Run Kubernetes in mixed environment

Exam perspective

Which resources fit where in a data flow pipeline



Google Certified Professional - Cloud Architect - Part 2

App Engine Resources and Management Tools

Cloud Source Repository

Private Git repository hosted on GCP

Collaborative development

Use with Stackdriver to view debug info alongside your code

Connect to GitHub/Bitbucket

Source code browser

May be part of data flow questions

GAE Management

Cloud Shell local environment

Preview app in local environment without deploying

Versions + Split Traffic

Roll out updates slowly

Firewall rules act differently

Default allow all

Control access from IP ranges

Cannot filter by traffic type

Block malicious IP's/DDOS

Best practices for app deployment

Break app into microservices

Roll out updates slowly with Split Traffic

"Canary update"

Use green-blue deployment model

Two parallel production environments, only one is live

Google Certified Professional - Cloud Architect - Part 2

Course Wrap Up and Next Steps

Part 2 overview

Focus on working with GCP services – building blocks

Emphasis on working with mechanics

Exam will have a number of 'low-level' questions

And also high level, conceptual questions

Next Steps - Part 3

Focus on high level concepts

Business use cases

Infrastructure design

Transfer from legacy datacenter

Security and compliance

Data lifecycle

And much more!

Releasing in mid-first quarter 2018