The GCP Dictionary of Pain

a straightforward guide to thorny cloud terms

A Cloud Guru

The GCP Dictionary of Pain

This idea of helping a team reach a "baseline cloud fluency" is a phrase we hear a lot at A Cloud Guru. You can have two engineers that excel at one kind of cloud operation, but if the rest of the team can't speak their language, nothing gets done. Things slow down—or stop—when people can't communicate.

At A Cloud Guru, our goal is to teach the world to cloud. So we turned to our data to help us understand how we can help bring organizations up to that baseline cloud fluency—or at least in the ballpark. Our goal with this guide is to make sure the next time someone mentions a major issue with Google Cloud Run (the product, not the verb), you won't be left scratching your head.

We analyzed 2.7 million responses to hundreds of questions across multiple areas of cloud expertise. We specifically looked at tough questions, where the correct response rate fell below what you'd probably call a C average (or 60%). From these questions we identified key products, technologies, and topics that were often represented in those questions—and are here to help you get a handle on them.

In this guide, we'll cover some core concepts (like load balancing and availability zones) as well as some of the most common—though clearly challenging!—GCP tools you'll find in a modern cloud stack.

You'll still probably want to do a lot better than just a C average. But while you won't need everyone to be an expert at everything, they still need to speak the same language. Or, at least, understand when someone's telling a cloud-related joke that isn't meant to be taken literally.

Here are the trip-up topics that stump cloud teams the world over: the GCP Dictionary of Pain.

Pain Forecast: Thunderstorms

Cloud & Identity Access Management

[kloud, ī'den(t)ədē, and, akˌses, manijmənt]

AWS *AWS Identity & Access Management* | ***Azure*** *Identity and access management*

WHAT IS IT? Identity and Access Management (or IAM) is your frontline defense for your cloud services. Security will be at the top of mind for basically everyone running a cloud operation—and it should be on top of mind for you as well. Google manages permissions and identity and access management with Cloud IAM and Cloud Identity.

Cloud IAM is a unified resource access management system for both users and services. Identities can come in many forms, such as Google accounts, unmanaged accounts, service accounts, and collections of these items—such as Google Groups and G-Suite domains.

There are three critical elements to IAM.

POLICIES : A set of rules that governs who can do one or more tasks with specific resources. We can decide who can do what on which resources.

ROLES : A collection of permissions assigned to identities or members.

RESOURCES : Made up, well, your resources. It can be your projects, folders, cloud services, or parts of those services like instances or buckets.

Google refers to Cloud Identity as a nice alphabet soup acronym dubbed “identity as a service” (or IDaaS). Cloud Identity started with G-Suite with a separate administrative portal but now integrates with Google Cloud. One of its main benefits is it can securely allow people outside your organization access to specific resources within your organization. It also enables single sign-on (or SSO).

WHY IS IT HARD? Lots of abstraction! Cloud Identity creates what Google calls “folders” for one or more Google Cloud Projects. You might have a folder for your developer team and your live team. You can apply policies to your folders, where all resources within the projects in those folders inherit permissions from those policies.

On top of all this, IAM can quickly become one of your points of failure in managing security. The scope of your IAM policies represents a large part of the surface area for Bad Things To Happen. You might misconfigure the permissions you’re giving a third party for your resources or set policies too broad. Or you might just lose the password altogether. Setting specific, robust IAM policies is one of the best ways to protect your cloud operations from breaches and bad actors.

A Cloud Guru learners missed tough questions related to identity and access management 50.5% of the time. Here are a few sample questions.

> What IAM predefined roles allow the modification of the default cookie expiration time for Google App Engine standard environments? (68.9% correct)

> Google recommends organizing users with the same responsibilities into groups, and then assigning Cloud IAM roles to those groups rather than to individuals. As per Google’s best practices for enterprise customers, which of the following are not recommended groups to create when you begin to implement Cloud IAM? (43.8% correct)

Pain Forecast: Scattered Thunderstorms

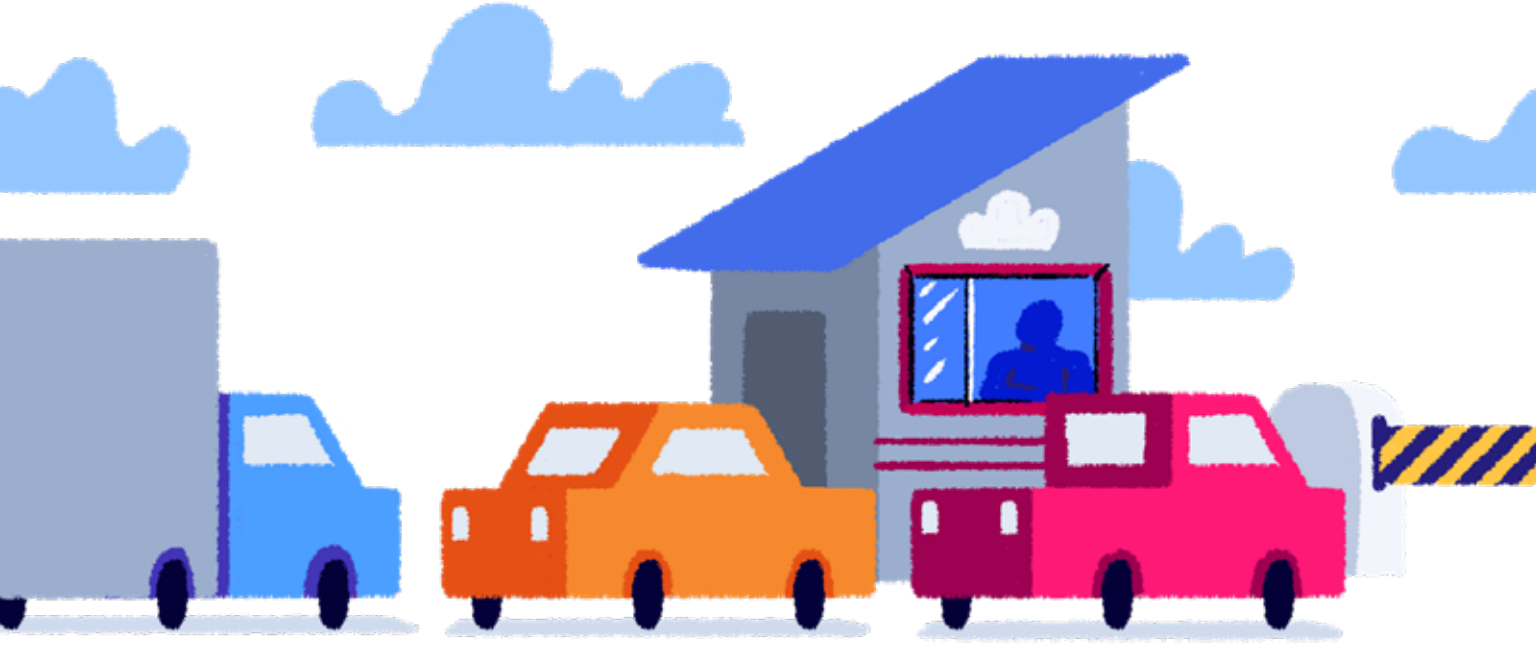
Cloud Load Balancing

[kloud, lōd, 'balənsING]

***AWS** Load Balancers | **Azure** Also load balancers!*

WHAT IS IT? Let's say you have plenty of instances available and a buttload of traffic on the way. How do you figure out how to optimize everything? Figuring out how to route the traffic—and balancing it across all of those instances. If your application is popular enough to use multiple VMs, you'll need to consider using load balancers to ensure that your users have a good experience.

It's a fully managed incoming traffic service responsible for distributing traffic across several VM instances. It is autoscaling, set by policy, CPU utilization, or serving capacity (or all of them). It supports heavy traffic and is instrumental in routing traffic to the closest instances, and can detect and remove unhealthy instances. You can also determine what percentage of failure is unhealthy.



WHY IS IT HARD? Part of the challenge is understanding the type of traffic you need to balance across your instances. There are three main types of load balancing supported by Google Cloud:

GLOBAL EXTERNAL : HTTP/S, SSL, and TCP

REGIONAL EXTERNAL : TCP/UDP within a region

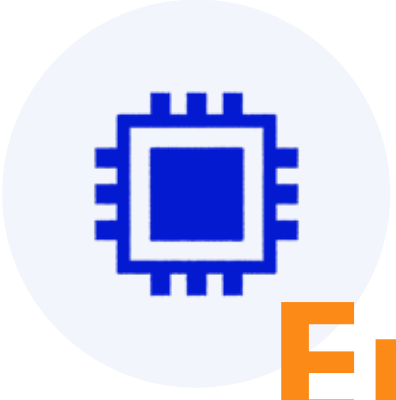
REGIONAL INTERNAL : between groups of instances within a region

A Cloud Guru learners missed tough questions related to load balancers 53% of the time. Here are a few sample questions.

> You just created a new Google Project. After enabling the compute engine API, the project now has a default VPC network created. You need to create a custom VPC network with subnets in the us-east1 and europe-west2 regions of GCP. The project will also need a load balancer that can route traffic to either region based on user location. What services must be configured to allow traffic through the load balancer to the backend services? (58.8% correct)

> Your application development team has updated files that are in use behind a Google Cloud HTTP(S) Load balancer with Content Delivery Network enabled. It has been determined that these files contain sensitive information, and must be removed ASAP from any cache server where they are stored. How would you accomplish this? (30% correct)

Pain Forecast: Mostly Sunny



Google Compute Engine(GCE)

[gōōg(ə)l, kəm'pyōōt, enjən]

***AWS** EC2 | **Azure** Azure VMs*

WHAT IS IT? Google Compute Engine, in short, is where just about everything happens. GCE consists of virtual machines (VMs) that run all of the operations for your applications. Combined with Google Cloud Storage, GCE is your operation running in the cloud without you ever needing to touch a power button. GCE is a type of infrastructure as a service (or IaaS)

Like other providers, GCE abstracts away the complicated processes of managing server hardware. You can select the amount of compute power you need for your operations and pay for as much as you need without having to buy or take care of the actual hardware. Google Cloud offers several machine types that vary in compute power, memory, and cost.

These VMs can run public disk images pre-configured and offered by Google, or private disk images accessible by you and only with you. Google Cloud also configures VMs to work with containers. They also come with virtual private network (or VPC) connectivity—we'll have more on that later.

WHY IS IT HARD? While you can't see it, it's still your computer. And as such, you'll still have to maintain parts of the operation without ever taking a look at the hardware. You'll have to provision the right number of resources, configure your machine types cor-

rectly, and finally make sure you're optimizing for performance and cost.

Fortunately, Google will offer some price estimates right in the dashboard when you're configuring your GCE VM. Google Cloud, in particular, offers four machine types for Google Cloud Engine:

1 : General-purpose machine types that work well for day-to-day needs. You'll see types like E2, N2, N2D, and N1—all of which may work well in different scenarios.

2 : A set of memory-optimized machines when you might have specific requirements around managing memory. These offer more memory per core.

3 : Compute-optimized instances, for when you need to optimize for computing power, even if it may cost more or sacrifice memory performance.

4 : Shared-core machine types, which timeshare a physical core and may be more cost-effective.

There are, of course, other configuration options. You'll have to set your zone (again, more on that later), determine whether to deploy a container image to your VM, select a type of disk image, pick sizes for those disks, and so forth.

A Cloud Guru learners missed tough questions related to GCE and VMs 50.3% of the time. Here are a few sample questions.

> How should you enable a GCE instance to read files from a bucket in the same project? (40.4% correct)

> You believe that the startup script for a newly created GCE instance did not run at all. What steps should you take to help investigate this problem? (41.8% correct)

Pain Forecast: Scattered Thunderstorms

Google App Engine

[gōōg(ə)l, ap, enjən]

WHAT IS IT? App Engine is one of the original four Google Cloud services. App Engine requires less management than its other compute products. That means you're abstracting the hardware away from your work. Suppose your operation falls under one the use cases of Google's four serverless products (we'll see the other three later). In that case, you could save an enormous amount of money—and a lot of headaches—by deploying your application on serverless technology. Google bills you for the time you spend processing, and you can scale up as much as you need.

App Engine is a platform-as-a-service or PaaS. So it requires no manual server setup and provisioning. It provides automatic scaling and load-balancing. It's great for websites, mobile apps, and line of business apps that need to get up and running quickly. Developers often use App Engine for HTTP applications.

It's exceptionally straightforward to get started with, code-centric, and provides automatic scaling regionally. It has easy access to the rest of the platform as a full-fledged product within Google Cloud. App Engine also has access to CloudSQL for relational work and works with Cloud Storage.

WHY IS IT HARD? You can't customize App Engine as you can with Compute Engine.

App Engine excels when working with scalable mobile apps—primarily gaming. Rapid scalability is a crucial factor when creating and handling online games, and App Engine can handle surges and scale up when needed.

App Engine also now comes in two environments: standard and flexible.

Standard is Google's original App Engine product: **1 :** It's more proprietary. **2 :** Source code must be written in specific versions of the supported programming languages. **3 :** The standard environment has the fastest spin-up time—literally milliseconds. **4 :** Good if you're expecting spikes in traffic. **5 :** It's standard so, of course, less expensive.

Flexible is Google's newer App Engine project: **1 :** It's less proprietary and (ironically) more standardized. **2 :** Flexible runs on Docker containers and works with more languages as a result. **3 :** You can access the resources of your Google Cloud project that reside in the Google Compute Engine network (more on that in a moment). **4 :** Flexible App Engine spin-up is slower and cannot scale down to zero, whereas Standard can. **5 :** It isn't recommended over Standard unless what you're working with doesn't work best on Standard.

A Cloud Guru learners missed tough questions related to App Engine 40.6% of the time. Here are a few sample questions.

> You need to view both request and application logs for your Python-based App Engine app. Which of the following options would be best? (56.8% correct)

> You are working together with a contractor from the Acme company and you need to allow App Engine running in one of Acme's GCP projects to write to a Cloud Pub/Sub topic you own. What information do you need to let you enable that access? (60.1% correct)

Pain Forecast: Severe Thunderstorms

Google Cloud Functions & Cloud Run

[gōōg(ə)l, kloud, fəNG(k)SH(ə)n, and, kloud, rən]

AWS *Lambda* | ***Azure*** *Azure Functions*

WHAT IS IT? App Engine isn't Google's only serverless product. Google also has three others: Cloud Functions, Cloud Run, and Cloud Run for Anthos.

Cloud Run operates on event-driven code. Instead of App Engine, something has to trigger them. You can think of Cloud Functions as the glue between the various services, whether they're on Google Cloud or third party. It scales automatically and is extremely low cost—Google bills you for execution time to the nearest 100 ms.

You can use Cloud Run for containerized apps, like those found in Google Kubernetes Engine. Cloud Run also scales on-demand and is suitable for continuous integration/continuous delivery. Unlike App Engine, use can use any language, library, or binary—even something completely custom.

There's also Cloud Run for Anthos, which enables hybrid multi-cloud apps. Cloud Run for Anthos runs on Google Kubernetes Engine. You can use custom domains, integrated logging, and monitoring services.

WHY IS IT HARD? There are a lot of serverless services from Google Cloud! And each is best for different scenarios (including App Engine, which we reviewed earlier). Let's review really quickly:

APP ENGINE : HTTP applications, such as modern web applications, to host front-end, back-end, or both.

CLOUD FUNCTIONS : Event-driven code, like third-party apps and API integrations, such as Twilio and Stripe. It also works well with IoT products for real-time processing, data collection, and processing. And it integrates with a large number of services in Google Cloud, making it work well for real-time processing.

CLOUD RUN : Container-based operations, including mature technology stacks. It's useful for internal company-only applications in addition to regular robust websites. You could also use it to, for example, create invoices monthly created by a cloud scheduler task.

CLOUD RUN FOR ANTHOS : Good for enterprise-grade CI/CD pipelines so you can continuously build new content. It's suitable for integrating on-premise services, executing your applications closest to customers (or "at the edge").

Your operation may work best with one, or multiple, or none whatsoever. So you'll have to determine the best approach to take when choosing whether to go serverless.

A Cloud Guru learners missed tough questions related to serverless operations 48% of the time. Here are a few sample questions.

> You've been asked to evaluate Google Cloud's implementation of its serverless compute service. What are some of the benefits of Cloud Functions you might include in your report? (59% correct)

Pain Forecast: Mostly Cloudy

Google Cloud SQL

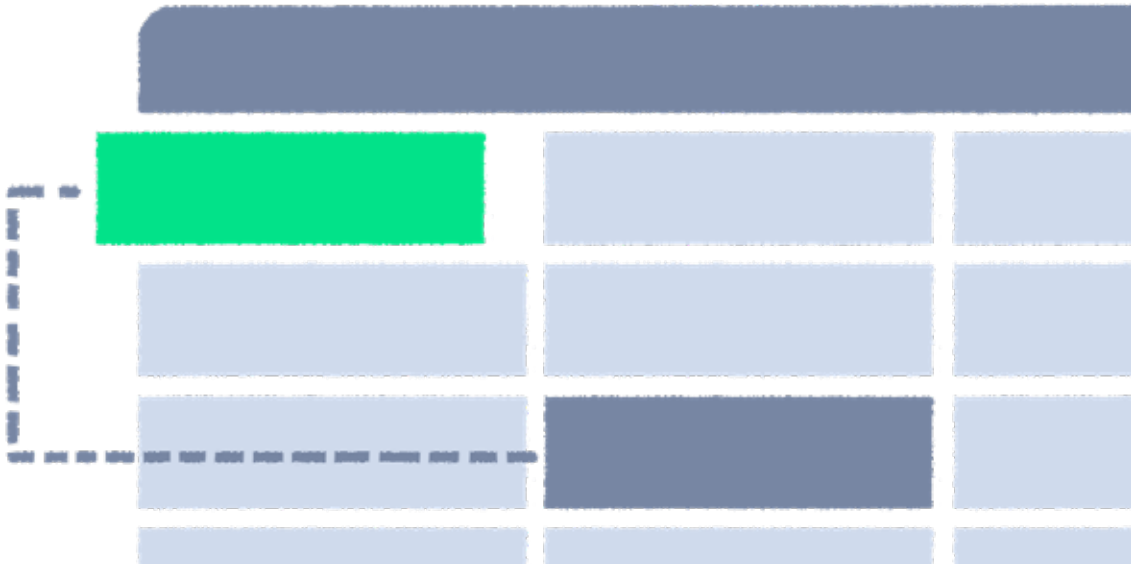
[gōōg(ə)l, kloud, es,kyōō,el]

***AWS** RDS | **Azure** Azure SQL Family*

WHAT IS IT? CloudSQL is Google Cloud's entry in the market of managing data with SQL. It's a relational database service that works across various needs for data-rich companies—ranging anywhere from flight services or ed-tech startups. CloudSQL manages, maintains, and administers PostgreSQL, SQL server, and MySQL (including two generations of MySQL). It just abstracts away the need to deal with hardware, and you can focus on working with the data.

Data is automatically encrypted and has a default firewall. CloudSQL also automates and replicates your backups, providing you with 99.95% uptime. You can easily copy data to another zone or region for high availability, so any time your source database crashes, you can failover to your backup. It's also easily accessible from GCE and App Engine.

Once you have the database instances created, all you have to do is use them. Even the console itself is pretty sparse.



WHY IS IT HARD? Configs! Everything, again, comes down to configuration for your CloudSQL databases. You'll have to add high availability, you'll have to specify a region, and you have to select a zone. And you'll have to set up the times for auto-backups as well as a maintenance schedule—which, if set automatically, may end up creating some service disruptions.

In theory, you could just go with best-practice versions or defaults, but your service probably has unique needs that could justify going deep into the config options.

A Cloud Guru learners missed tough questions related to CloudSQL and relational databases 49.6% of the time. Here are a few sample questions.

> A CloudSQL database is experiencing more read transactions recently due to the increasing need for admin reports by a system. Which of the following can be used to reduce the load issues experienced by the system? (69.6% correct)

Pain Forecast: Partly Cloudy

Google Cloud Storage

[gōōg(ə)l, kloud, stôrij]

Amazon S3 | ***Azure*** *Blog storage*

WHAT IS IT? Cloud storage is the unstructured data service on google cloud based on buckets and objects in those buckets. Those buckets can hold, well, anything that's digital—pictures, files, or entire databases. It has practically infinite size, assuming you have a credit limit high enough to hit it. You only pay for the amount of data currently in that bucket, and you don't have to pre-allocate space in advance.

A project can have one or more buckets assigned to it, and it observes the inherited format for IAM. You have the option of applying an IAM role to an individual bucket. You can think of objects and files (often used interchangeably, even though the technical term is an object).

You can apply an IAM role to an individual bucket. Objects in those buckets inherit those buckets' permissions, and those buckets can inherit permissions from the project. Technically cloud storage is an object storage system, not a file system (which you would use Filestore). It's also not used for block storage that you need for a hard drive. It has different write and performance characteristics.

There's also Cloud Run for Anthos, which enables hybrid multi-cloud apps. Cloud Run for Anthos runs on Google Kubernetes Engine. You can use custom domains, integrated logging, and monitoring services.

WHY IS IT HARD? There are also different types of storage classes, and they are great for various reasons. You can apply a storage class to an entire bucket, or even the objects in those buckets. They all have the same performance, but there's a difference

in geolocation, the SLA, and operations restrictions. Here are the storage classes:

STANDARD REGIONAL : The same as a multi-regional bucket, but the bucket resides in a single region. If you're using services in that region, you may have slightly better performance accessing the data in a regional bucket and have lower costs.

STANDARD MULTI-REGIONAL : This has your data spanning multiple regions in a continental region, suitable for hot data from different geographic regions. It has no retrieval cost and has the highest SLA.

NEARLINE (ALSO IN REGIONAL AND MULTI-REGIONAL) : The cost drops to a much lower per-gigabyte fee, but comes with a retrieval fee. Nearline storage is suitable for regular backups that you plan to access periodically.

COLDLINE (ALSO IN REGIONAL AND MULTI-REGIONAL) : Even cheaper than nearline at a per-gigabyte level, but has an even higher cost per retrieval. It is suitable for data that you don't really intend to touch, but you might want to have it just in case.

You can change the storage class for your data, but you can't change it from multi-regional to regional or vice versa. When you change the storage class of your bucket, it only affects new objects coming into that bucket—you'll have to change the storage class for objects inside the bucket.

A Cloud Guru learners missed tough questions related to cloud storage 49.2% of the time. Here are a few sample questions.

> If you have a GCE instance with no external IP address, and the instance needs to access Google Cloud Storage, what action must be performed to fulfill this requirement? (65.5% correct)

> You are troubleshooting an issue regarding a VM instance's ability to read information from a Google Storage Bucket. While researching the problem you find this connection has never been successful. The VM instance is using a project service account to access the Google Cloud APIs. You decide the problem is that the VM does not have permissions to view the data stored in Cloud Storage. What permissions need to be granted to the service account to allow permissions to view Cloud Storage? (67.6% correct)

Pain Forecast: Scattered Thunderstorms

Instance Groups

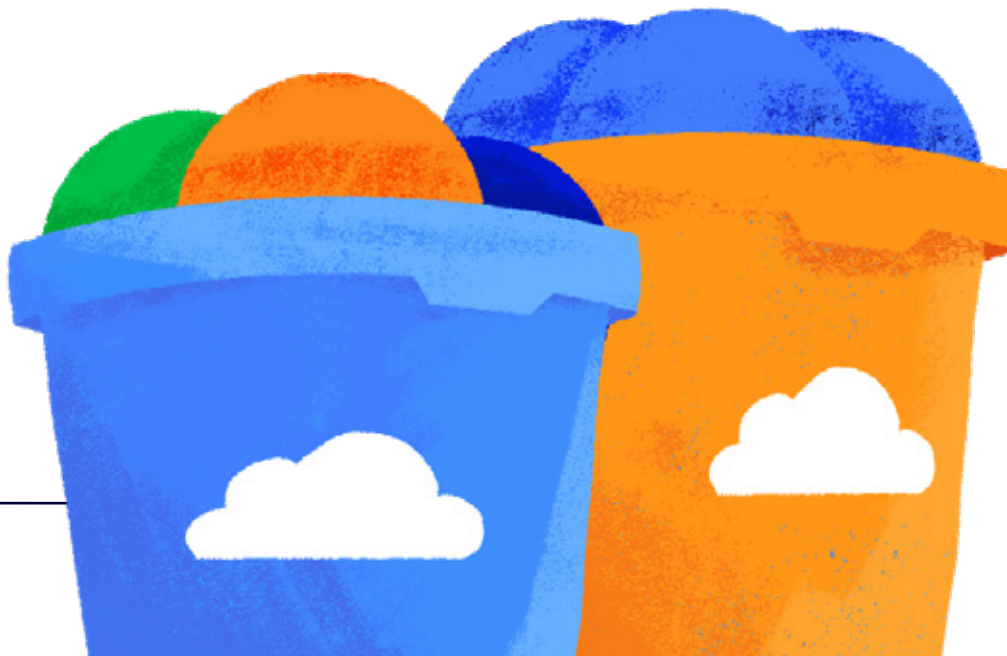
[instəns, groōps]

WHAT IS IT? Things might start to get a little out of hand once you start spinning up dozens, or hundreds, of Google Compute Engine instances. You can get around the complexity here by batching together instances into groups—buckets of instances where you can try to make changes to a bunch of them all at once.

There are two types of instanced groups that Google offers.

UNMANAGED GROUPS : collections of instances with different configurations, which you can use to apply load balancing to existing configurations.

MANAGED GROUPS : collections of instances that are identical that you can manage as a single unit. If you need to make changes to all of them at once, it'll make more sense to run it as an instance group. If there are health problems, the managed instance group will automatically recreate that instance.



WHY IS IT HARD? There are advantages and disadvantages to both. If you want the benefits of managed groups, you'll have to ensure that all the instances are identical. And you'll probably run into plenty of scenarios where you're running multiple identical instances! That makes it friendly and convenient when you want to handle auto-scaling and health checks for your instances.

But you might frequently run into scenarios where you'll use unmanaged instance groups. They're not uniform, so you won't get some of the features of managed groups. You don't get auto-scaling or update support if you're running an unmanaged instance group. But you can still enable some batch processes, so you'll want to group them anyway.

A Cloud Guru learners missed tough questions related to instance groups 46.5% of the time. Here are a few sample questions.

> You have an autoscaled managed instance group that is set to scale based on CPU utilization of 60%. There are currently 3 instances in the instance group. You're connected to one of the instances and notice that the CPU usage is a 70%. However, the instance group isn't starting up another instance. What's the most likely reason? (68.2% correct)

Pain Forecast: Thunderstorms

Snapshots

[snap,SHät]

***AWS** Snapshots | **Azure** Also snapshots!*

WHAT IS IT? The primary backup method in which a compute engine takes a snapshot of an instance or attached disk, creating a backup of that disk. You can take that snapshot in any disaster recovery situation and create a new instance or disk. You can create a snapshot while an instance is running.

You can create a restored snapshot or instance in a completely different zone, and you can create a snapshot from a boot disk or an attached disk.

WHY IS IT HARD? Snapshots operate as incremental backups, which can save you a fair bit of money. A snapshot does not create a new backup of a disk. Instead, it's an incremental backup: the second snapshot copies the changes from the first, the third from the second, and so forth.

You generally want to reduce activity whenever you can. You can create a snapshot when an instance runs, but you want to make sure your disk is consistent and isn't changing too much while the snapshot is in process. You might want to pause operations that write data or unmount the disk completely. You also probably want to schedule a snapshot when there isn't a log of data changing or if there isn't a log going on with your application.

A Cloud Guru learners missed tough questions related to Snapshots 50% of the time. Here are a few sample questions.

> As part of your backup plan, you create regular boot-disk snapshots of Compute Engine instances that are running. You want to be able to restore these snapshots using the fewest possible steps for replacement instances. What should you do? (69.2% correct)

Pain Forecast: Severe Thunderstorms

VPC Networks & Subnets

[vē,pē,sē, net ̩wərks, səb'nets]

WHAT IS IT? The GCP VPC (say that five times fast) is a virtualized network that provides ipv4 connectivity for GCP resources, such as infrastructure as a service and Google Kubernetes Engine. It's the central foundational component of all other networking functions. VPCs create a virtualized global network instance to consume resources in all GCP locations.

It's an implemented software-defined private network on Google Cloud—which means no routers, switches, servers, and tripping over tangled cables. This allows you to customize and scale your services rapidly, and all communication between resources within that network has no exposure to the public internet.

The VPC relies on the Google Global Network for connectivity, divided into GCP regions—or data center locations in various geographic locations. Google also offers premium and standard networking tiers in the VPC. GCP also requires an IP range assigned in a specific region (called a subnet). Regions can also speak to one another via Routes that allow communication between those networks. Google Cloud assigns all this infrastructure based on availability zones.

A Google Project can contain one or more VPC networks, creating configurations within each region. In a VPC, you can consume resources in all regions by creating a subnet within one region. Subnets use the Google global network, building subnets within a zone. We use the VPC to assign a subnet to one, multiple, or all regions. Each IP range provides IP connectivity for our infrastructure in those data centers. And unlike other cloud providers, subnets can span multiple availability zones within a region.

WHY IS IT HARD? While VPCs exist within projects, projects separate users—while VPCs separate systems. You can use shared VPCs that allow you to connect multiple projects to a single VPC network, making it possible for them to communicate with one another.

Let's say you have multiple GCE instances in the same project but across numerous VPCs. The resources in one VPC cannot communicate with the resources in another VPC without VPC network peering. However, resources within one VPC can communicate over a private network within multiple regions.

There are also two tiers of networking: premium and standard. Here are some of the differences:

PREMIUM-TIER DATA goes through Google's network, with the user's traffic entering Google's network at the nearest location and exiting at the global edge. This is sometimes referred to as "cold potato" traffic.

STANDARD-TIER TRAFFIC goes through Google's network through peering, ISP, or transit networks in the region where you've deployed GCP resources. This is sometimes referred to as "hot potato" traffic.

A Cloud Guru learners missed tough questions related to GCP VPCs 42.2% of the time. Here are a few sample questions.

> Your company has begun its migration to the Google Cloud Platform. You have already created a custom VPC network and started migrating workloads to the new platform. During the migration you discovered that a service being migrated relied on an on-premises VPN connection to a third-party data source. The third-party vendor has the same data available in GCP VPC. Which is the best option to access the data? (44.6% correct)

> Shared VPC is a single VPC network that can be shared across multiple projects within an organization in GCP. When utilizing a shared VPC network, a service project must define these network resources within the same service project? (36.1% correct)

A CLOUD GURU

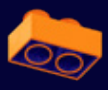
Cloud Smarter. Cloud Better. Learn with A Cloud Guru.

Since 2015, we've helped more than two million engineers learn to cloud with approachable courses taught by industry experts. Scale your teams, get them certified, generate value, and learn a ton of cool stuff along the way.

Accelerate Adoption
Learn the critical skills you need to supercharge your products with top cloud tools and technology.

Speak Cloud
Our scalable, sprint-based program lifts everyone to a common base of knowledge quickly and effectively.

Stay Informed
See the latest trends in the way learners build their cloud knowledge based on our research.



Learn By Doing
Give your team valuable hands-on experience—and let them break everything—without exposing your production environments.

Keep Growing
Provide ongoing, guided learning that takes your employees from novice to guru in specialized cloud career tracks.

Ready to take off? We'd love to help you reach the sky.

Request A Demo