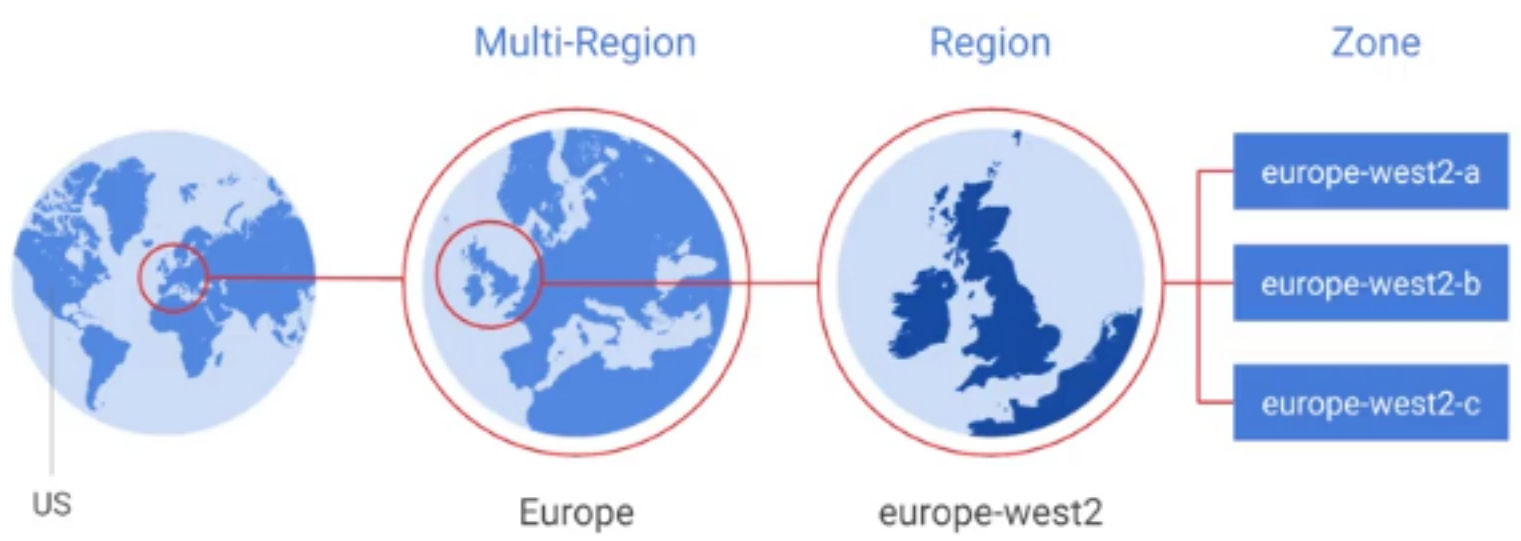
Here's how GCP is organized. Let's start at the finest grain level, the Zone, shown here on the right. A zone is a deployment area for Google Cloud Platform Resources.

For example, when you launch a virtual machine in GCP using Compute Engine, which we'll discuss later, it runs in a zone you specify. Although people think of a zone as being like a GCP Data Center, that's not strictly accurate because a zone doesn't always correspond to a single physical building.

You can still visualize the zone that

Google Cloud Platform is organized into regions and zones



You can still visualize the zone that way, though. Zones are grouped into regions, independent geographic areas, and you can choose what regions your GCP resources are in. All the zones within a region have fast network connectivity among them.

Locations within regions usually have round trip network latencies of under five milliseconds. Think of a zone as a single failure domain within a region. As part of building a fault tolerant application, you can spread their resources across multiple zones in a region.

Environmental responsibility

Google's data centers were the first to achieve ISO 14001 certification, which is a standard that maps out a frame-work for improving resource efficiency and reducing waste. This is Google's data center in Hamina, Finland, one of the most advanced and efficient data centers in the Google fleet.

Its cooling system uses seawater from the bay of Finland to reduce energy use. It's the first of its kind anywhere in the world. Google is one of the world's largest corporate purchasers of wind and solar energy.

Google has been a hundred percent

Its cooling system uses seawater from the bay of Finland to reduce energy use. It's the first of its kind anywhere in the world. Google is one of the world's largest corporate purchasers of wind and solar energy.

Google has been a hundred percent carbon neutral since 2007, and will shortly reach a hundred percent renewable energy sources for its data centers. Just like its customers, Google is trying to do the right things for the planet.

GCP customers have environmental goals of their own, and running their

Google offers customer-friendly pricing

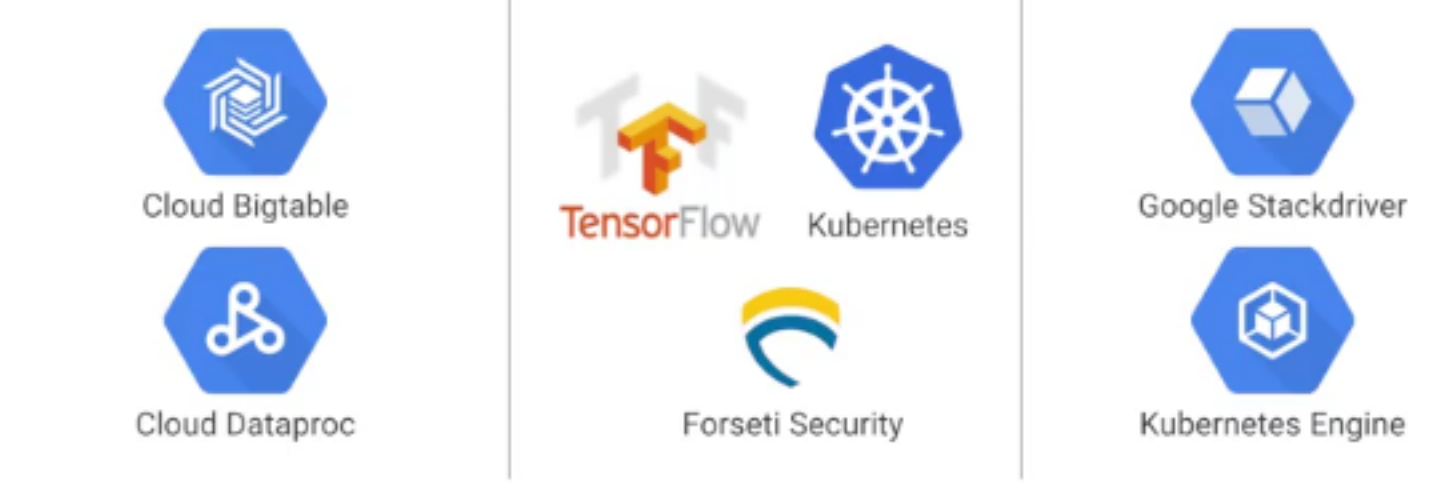
Billing in sub-hour increments	Discounts for sustained use	Discounts for committed use	Discounts for preemptible use	Custom VM instance types
For compute, data processing and other services	Automatically applied to virtual machine use over 25% of a month	Pay less for steady, long-term workloads	Pay less for interruptible workloads	Pay only for the resources you need for your application

Google was the first major Cloud provider to deliver per second billing for its Infrastructure as a Service Compute offering, Google Compute Engine. Fine-grain billing is a big cost savings for workloads that are bursty, which is a lot of them.

Many of the best-known GCP services billed by the second, including Compute Engine and Kubernetes Engine and you'll learn about them and others in this course.

Compute Engine offers automatically applied sustained use discounts which are automatic discounts that

Open APIs

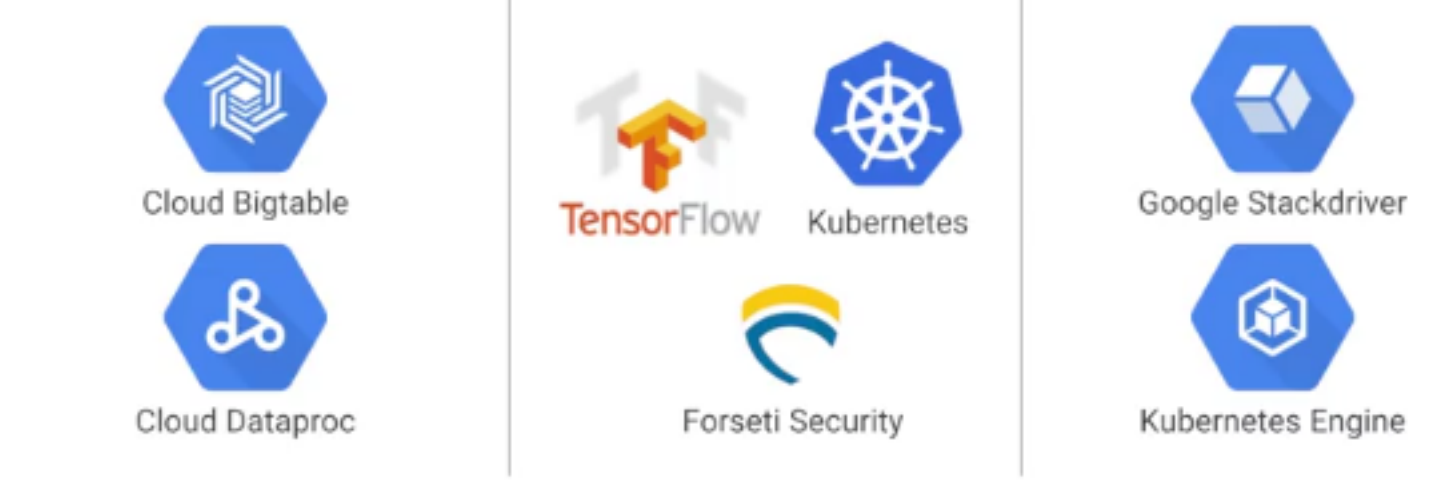


Here are some examples of how Google helps its customers avoid feeling locked in. GCP services are compatible with open source products. For example, take Cloud Bigtable, a database we'll discuss later.

Bigtable uses the interface of the open source database Apache HBase, which gives customers the benefit of code portability. Another example, Cloud Dataproc offers the open source big data environment Hadoop, as a managed service.

Google publishes key elements of technology using open source licenses to create ecosystems that

Open APIs



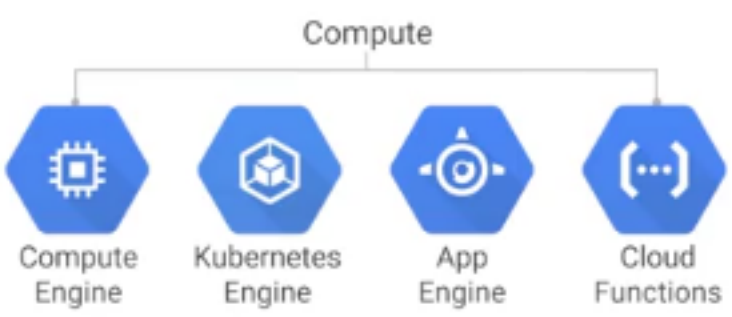
Bigtable uses the interface of the open source database Apache HBase, which gives customers the benefit of code portability. Another example, Cloud Dataproc offers the open source big data environment Hadoop, as a managed service.

Google publishes key elements of technology using open source licenses to create ecosystems that provide customers with options other than Google. For example, TensorFlow, an open source software library for machine learning developed inside Google, is at the heart of a strong open source ecosystem.

Many GCP technologies provide

Why choose Go...loud Platform

Review: Google Cloud Platform offers a range of compute services



Google Cloud Platform's products and services can be broadly categorized as compute, storage, big data, machine learning, networking and operations and tools. This course considers each of the compute services and discusses why customers might choose each.

The course will examine each of Google Cloud Platform storage services, how it works and when customers use it. To learn more about these services, you can participate in the training courses in Google Cloud's Data Analyst learning track.

This course also examines the

Why choose Go...loud Platform

and discusses why customers might choose each.

The course will examine each of Google Cloud Platform storage services, how it works and when customers use it. To learn more about these services, you can participate in the training courses in Google Cloud's Data Analyst learning track.

This course also examines the function and purpose of Google Cloud Platform's big data and machine learning services. More details about these services are also available in the training courses in Google Cloud's Data Analyst learning track.

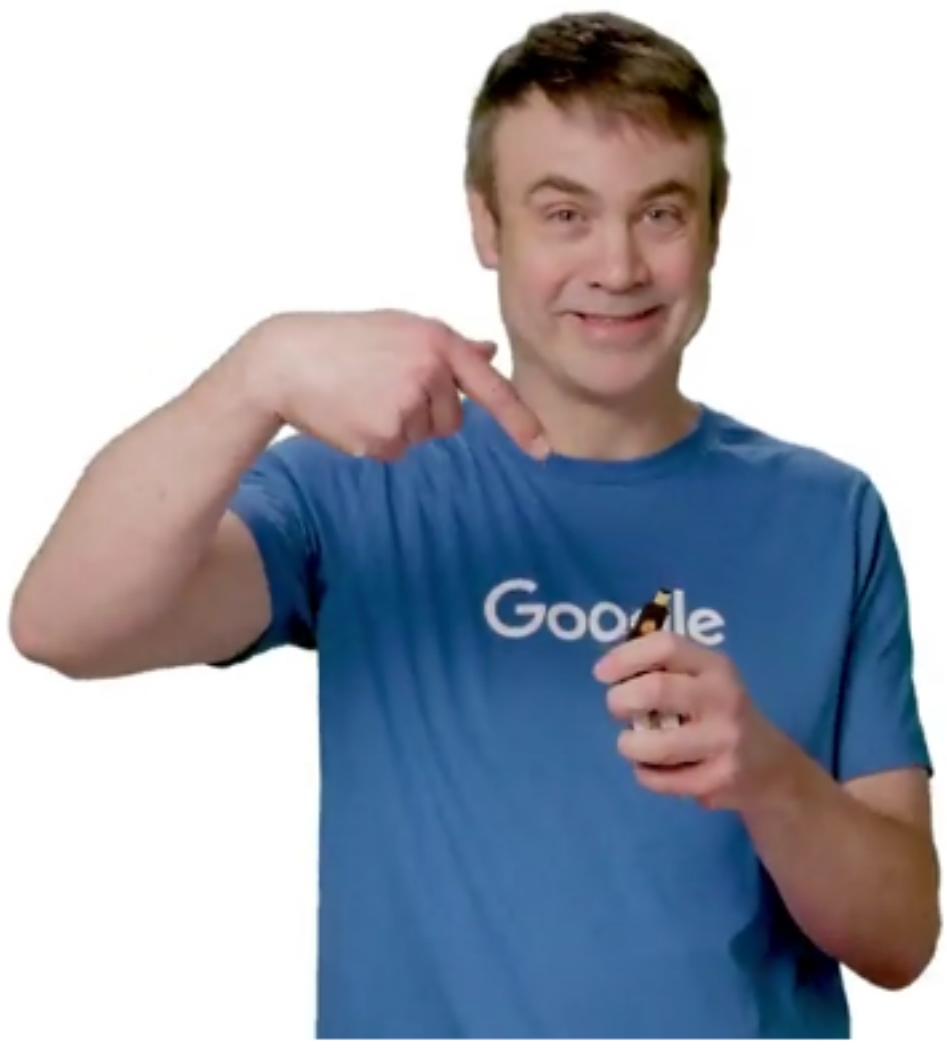
Security is designed into Google's technical infrastructure

Layer	Notable security measures (among others)
Operational security	Intrusion detection systems; techniques to reduce insider risk; employee U2F use; software development practices
Internet communication	Google Front End; designed-in Denial of Service protection
Storage services	Encryption at rest
User identity	Central identity service with support for U2F
Service deployment	Encryption of inter-service communication
Hardware infrastructure	Hardware design and provenance; secure boot stack; premises security

Google also designs custom chips, including a hardware security chip called Titan that's currently being deployed on both servers and peripherals. Google server machines use cryptographic signatures to make sure they are booting the correct software.

Google designs and builds its own data centers which incorporate multiple layers of physical security protections. Access to these data centers is limited to only a very small fraction of Google employees, not including me.

Google's infrastructure provides cryptographic privacy and integrity for



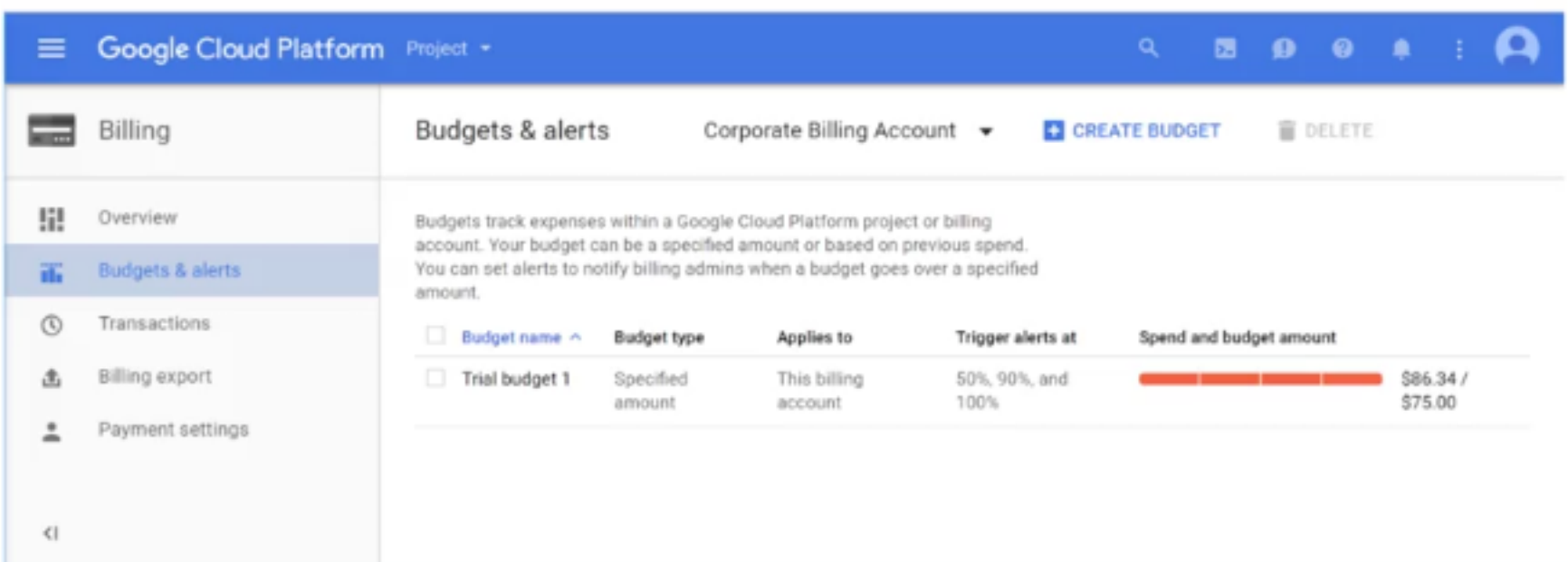
Users can also use second factors when signing in, including devices based on the universal second factor U2F open standard. Here's mine.

Most applications at Google access physical storage indirectly via storage services and encryption is built into those services.

Google also enables hardware encryption support in hard drives and SSDs. That's how Google achieves encryption at rest of customer data.

Google services that want to make themselves available on the Internet register themselves with an infrastructure service called the Google Front End, which checks incoming network

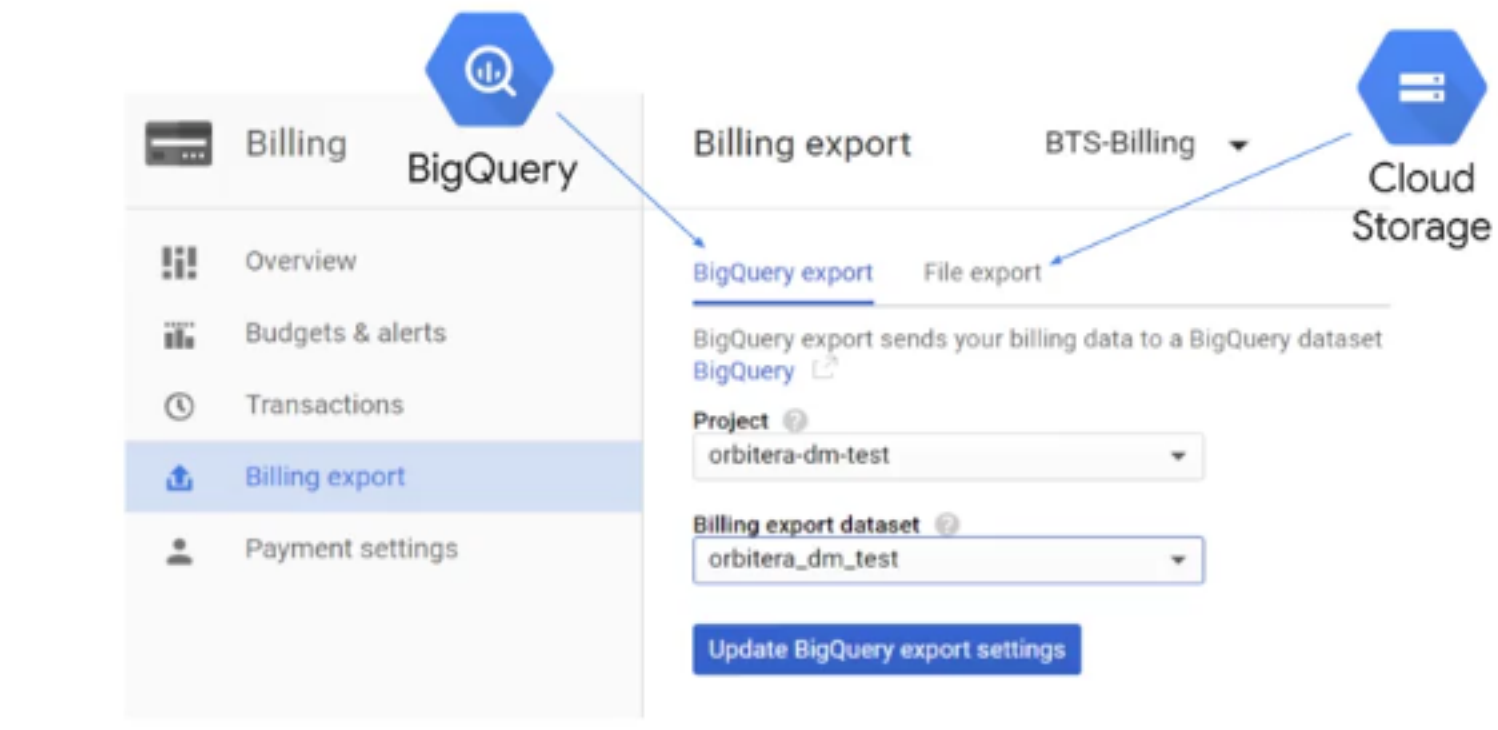
Budgets and alerts keep your billing under control



look at budgets and alerts first. You can define budgets either per billing account or per GCP project.

A budget can be a fixed limit or you can tie it to another metric. For example, a percentage of the previous month spend. To be notified when costs approach your budget limit, create an alert.

For example, with a budget limit of $20,000 and an alert set at 90 percent, you'll receive a notification alert when your expenses reach $18,000. Alerts are generally set at 50 percent, 90 percent, and 100 percent. But you can customize that.

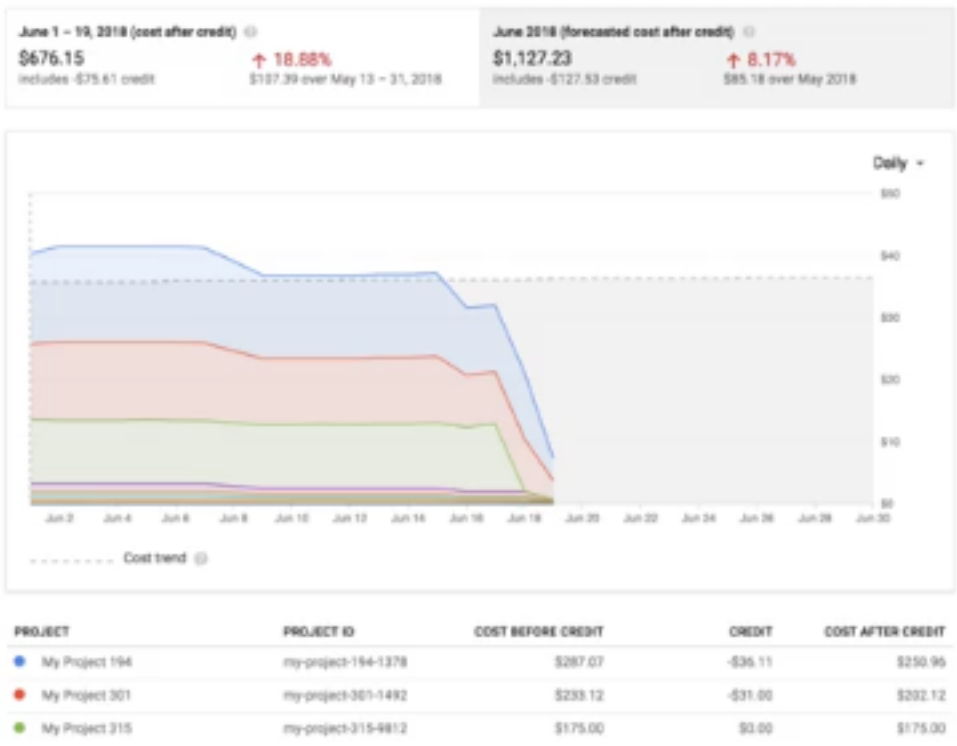


your expenses reach $18,000. Alerts are generally set at 50 percent, 90 percent, and 100 percent. But you can customize that.

Billing export lets you store detailed billing information in places where it's easy to retrieve for more detailed analysis, such as a BigQuery dataset or a Cloud storage bucket. Reports is a visual tool in the GCP console that allows you to monitor your expenditure.

GCP also implements quotas, which protect both account owners and the GCP community as a whole. Quotas are designed to prevent the over-consumption of resources.

Reports is a visual tool to monitor expenditure



or a Cloud storage bucket. Reports is a visual tool in the GCP console that allows you to monitor your expenditure.

GCP also implements quotas, which protect both account owners and the GCP community as a whole. Quotas are designed to prevent the over-consumption of resources, whether because of error or malicious attack. There are two types of quotas: rate quotas and allocation quotas.

Both get applied at the level of the GCP project. Rate quotas reset after a specific time. For example, by default, the Kubernetes Engine service sets a

Quotas are helpful limits

Rate quota
GKE API: 1,000 requests per 100 seconds

Allocation quota
5 networks per project

Many quotas are changeable

GCP also implements quotas, which protect both account owners and the GCP community as a whole. Quotas are designed to prevent the over-consumption of resources, whether because of error or malicious attack. There are two types of quotas: rate quotas and allocation quotas.

Both get applied at the level of the GCP project. Rate quotas reset after a specific time. For example, by default, the Kubernetes Engine service sets a quota of a 1000 calls to its API from each GCP project every 100 seconds. After that 100 seconds, the limit is reset.

Quotas are helpful limits

Rate quota
GKE API: 1,000 requests per 100 seconds

Allocation quota
5 networks per project

Many quotas are changeable

each GCP project every 100 seconds. After that 100 seconds, the limit is reset.

Allocation quotas, on the other hand, govern the number of resources you can have in your projects. For example, by default, each GCP project has a quota allowing it no more than five Virtual Private Cloud networks.

Although projects all start with the same quotas, you can change some of them by requesting an increase from Google Cloud support.

reset.

Allocation quotas, on the other hand, govern the number of resources you can have in your projects. For example, by default, each GCP project has a quota allowing it no more than five Virtual Private Cloud networks.

Although projects all start with the same quotas, you can change some of them by requesting an increase from Google Cloud support.

Module introduction

Cloud security requires collaboration

- Google is responsible for managing its infrastructure security.
- You are responsible for securing your data.
- Google helps you with best practices, templates, products, and solutions.

And he's still very, very sorry. GCP customers use IM to implement least privilege, and it makes everybody happier. There are four ways to interact with GCP's management layer: through the web-based console, through the SDK and its command-line tools, through the APIs, and through a mobile app.

In this class, you'll mostly use the console and the command-line tools. When you build an application on your on-premises infrastructure, you're responsible for the entire stack security.

From the physical security of the hardware, and the premises in which they're housed, through the encryption

Resource hierarchy levels define trust boundaries

- Group your resources according to your organization structure.
- Levels of the hierarchy provide trust boundaries and resource isolation.

the GCP resource hierarchy from the bottom up. All the resources you use, whether they're virtual machines, cloud storage buckets, tables and big query or anything else in GCP are organized into projects. Optionally, these projects may be organized into folders.

Folders can contain other folders. All the folders and projects used by your organization can be brought together under an organization node. Projects, folders and organization nodes are all places where the policies can be defined.

Some GCP resources let you put poli-

The Google Cl...urce hierarchy

All GCP services you use are associated with a project

- Track resource and quota usage
- Enable billing
- Manage permissions and credentials
- Enable services and APIs

Google services.

Each project is a separate compartment and each resource belongs to exactly one. Projects can have different owners and users - they're built separately and they're managed separately. Each GCP project has a name and a project ID that you assign.

The project ID is a permanent, unchangeable identifier and it has to be unique across GCP. You use project IDs in several contexts to tell GCP which project you want to work with. On the other hand, project names are for your convenience and you can assign them.

The Google Cl...urce hierarchy

Projects have three identifying attributes

Project ID	Globally unique	Chosen by you	Immutable
Project name	Need not be unique	Chosen by you	Mutable
Project number	Globally unique	Assigned by GCP	Immutable

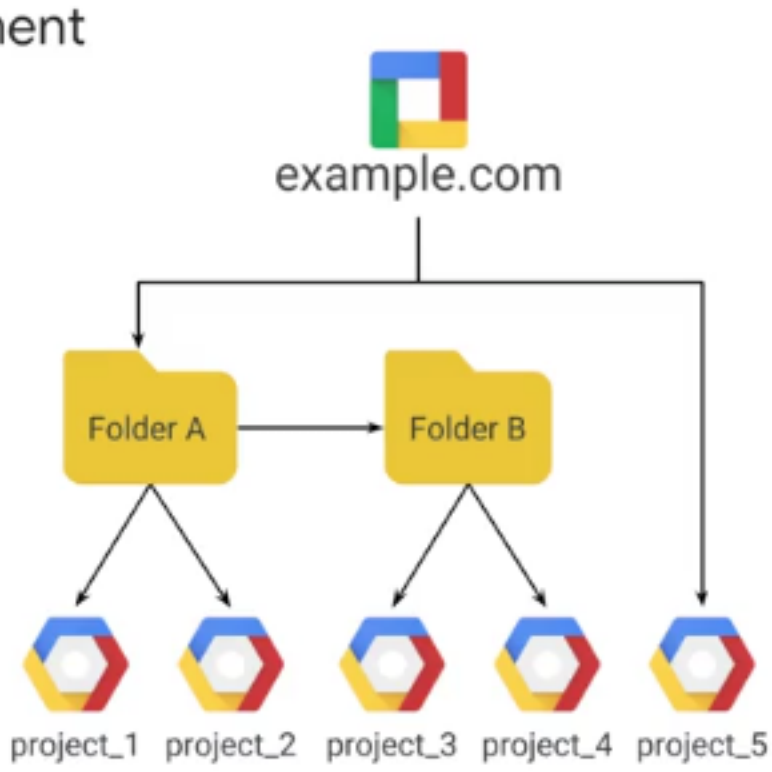
The project ID is a permanent, unchangeable identifier and it has to be unique across GCP. You use project IDs in several contexts to tell GCP which project you want to work with. On the other hand, project names are for your convenience and you can assign them.

GCP also assigns each of your projects a unique project number and you'll see a display to you in various contexts. But using it is mostly outside the scope of this course. In general, project IDs are made to be human readable strings and you'll use them frequently to refer to projects.

You can organize projects into folders

Folders offer flexible management

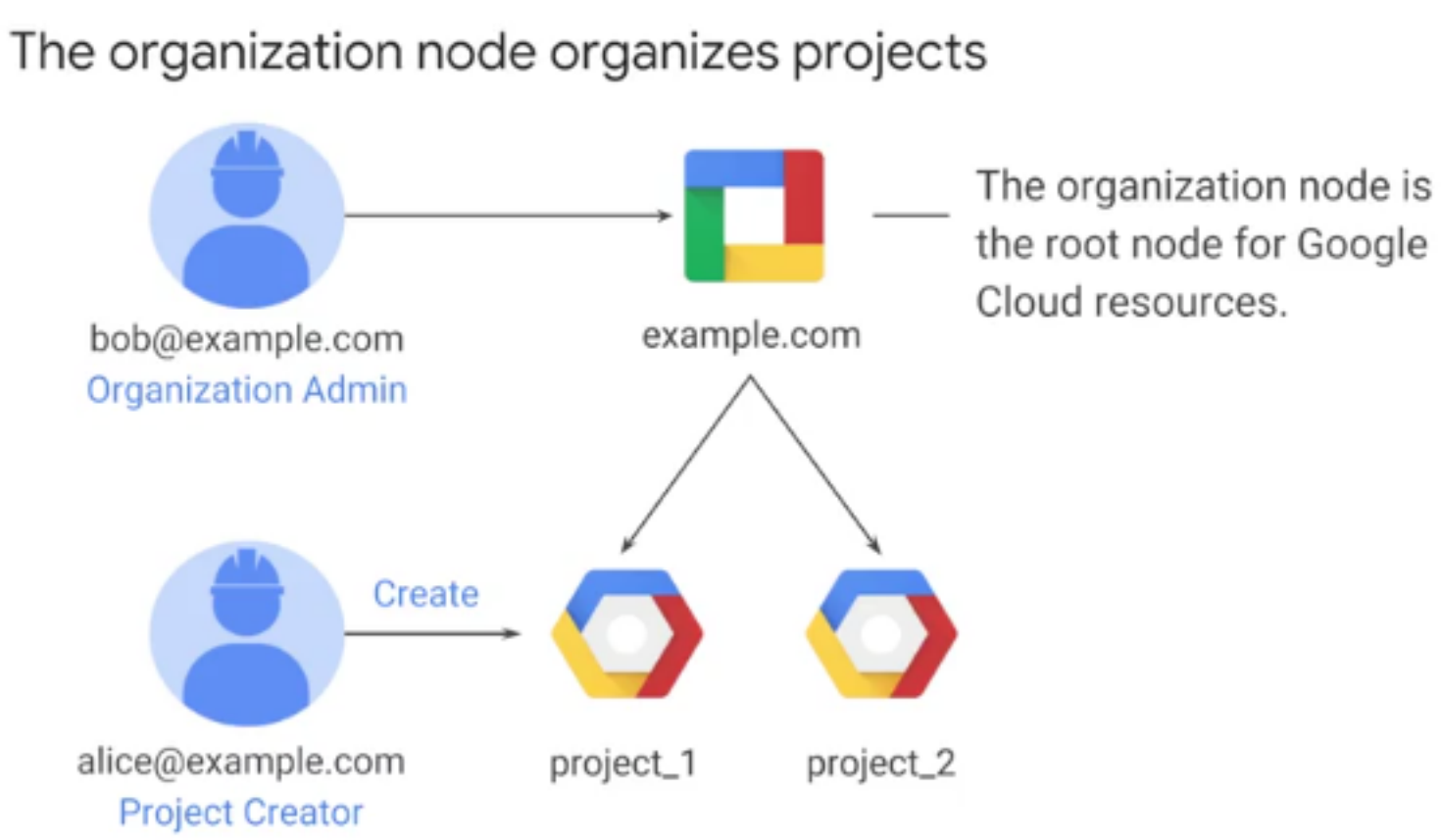
- Folders group projects under an organization.
- Folders can contain projects, other folders, or both.
- Use folders to assign policies.



You can organize projects into folders, although you don't have to. They're a tool at your disposal to make your life easier. For example, you can use folders to represent different departments, teams, applications or environments in your organization.

Folders let teams have the ability to delegate administrative rights, so they can work independently. The resources in a folder inherit IAM policies from the folder. So, if project three and four are administered by the same team by design, you can put IAM policies into folder B instead.

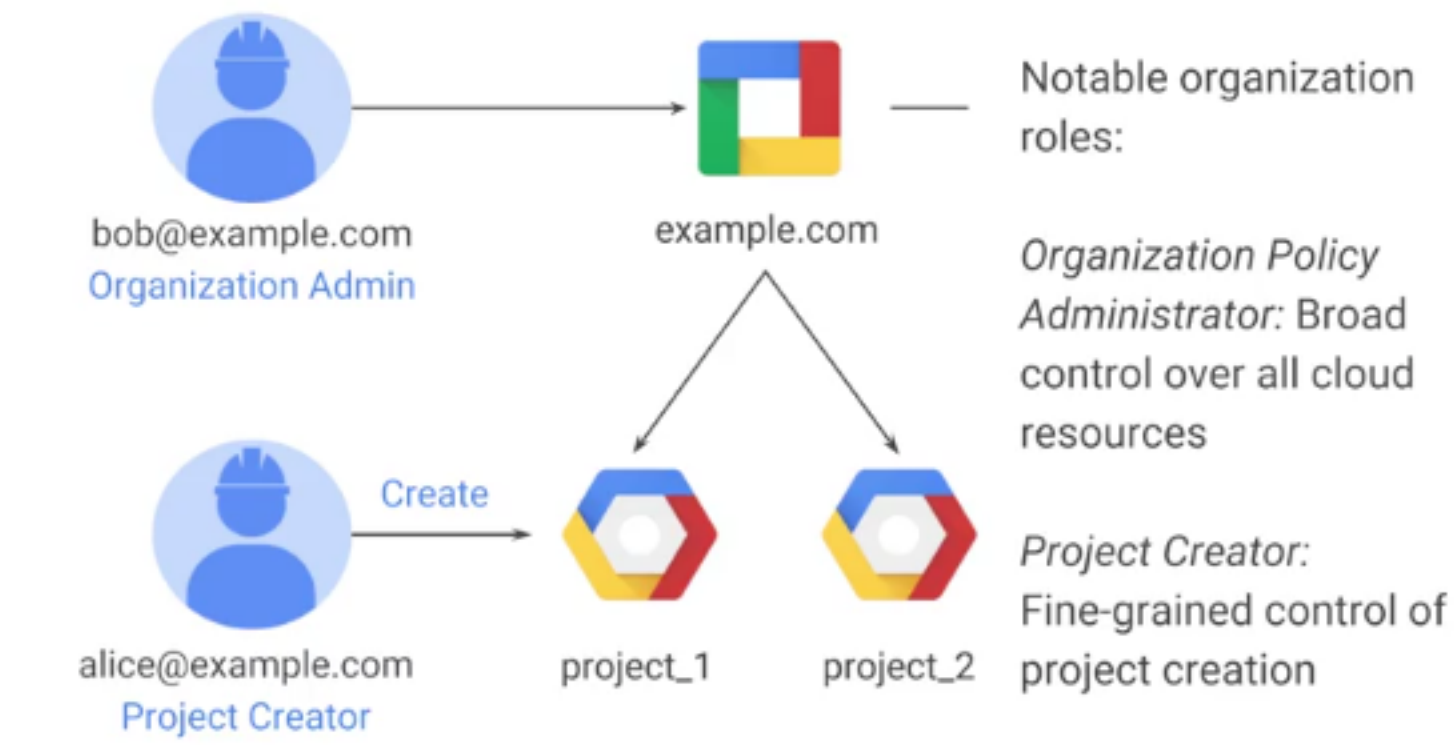
Doing it the other way, putting



You probably want to organize all the projects in your company into a single structure. Most companies want the ability to have centralized visibility on how resources are being used and to apply policy centrally. That's what the organization node is for. It's the top of the hierarchy.

There are some special roles associated with it. For example, you can designate an organization policy administrator so that only people with privilege can change policies. You can also assign a project creator role, which is a great way to control who can spend money.

So how do you get an organization



There are some special roles associated with it. For example, you can designate an organization policy administrator so that only people with privilege can change policies. You can also assign a project creator role, which is a great way to control who can spend money.

So how do you get an organization node? In part the answer depends on whether your company is also a G Suite customer. If you have a G Suite domain, GCP projects will automatically belong to your organization node. Otherwise, you can use Google Cloud Identity to create one. Here's a tip.

So how do you get an organization node? In part the answer depends on whether your company is also a G Suite customer. If you have a G Suite domain, GCP projects will automatically belong to your organization node. Otherwise, you can use Google Cloud Identity to create one. Here's a tip.

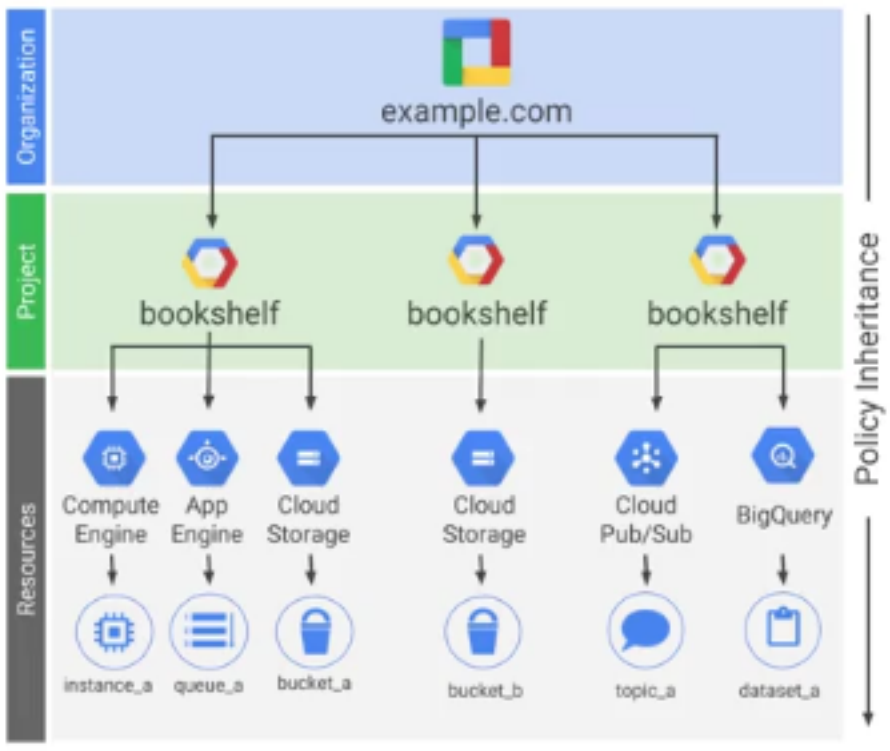
When you get a new organization node, it lets anyone in the domain create projects and billing accounts just as they could before. That's to avoid surprises and disruption.

But it'd be a great first step with a new organization node to decide who on

The Google Cl...urce hierarchy

An example IAM resource hierarchy

- A policy is set on a resource.
  - Each policy contains a set of roles and role members.
- Resources inherit policies from parent.
  - Resource policies are a union of parent and resource.
- A less restrictive parent policy overrides a more restrictive resource policy.



Resources inherit the policies of their parent resource.

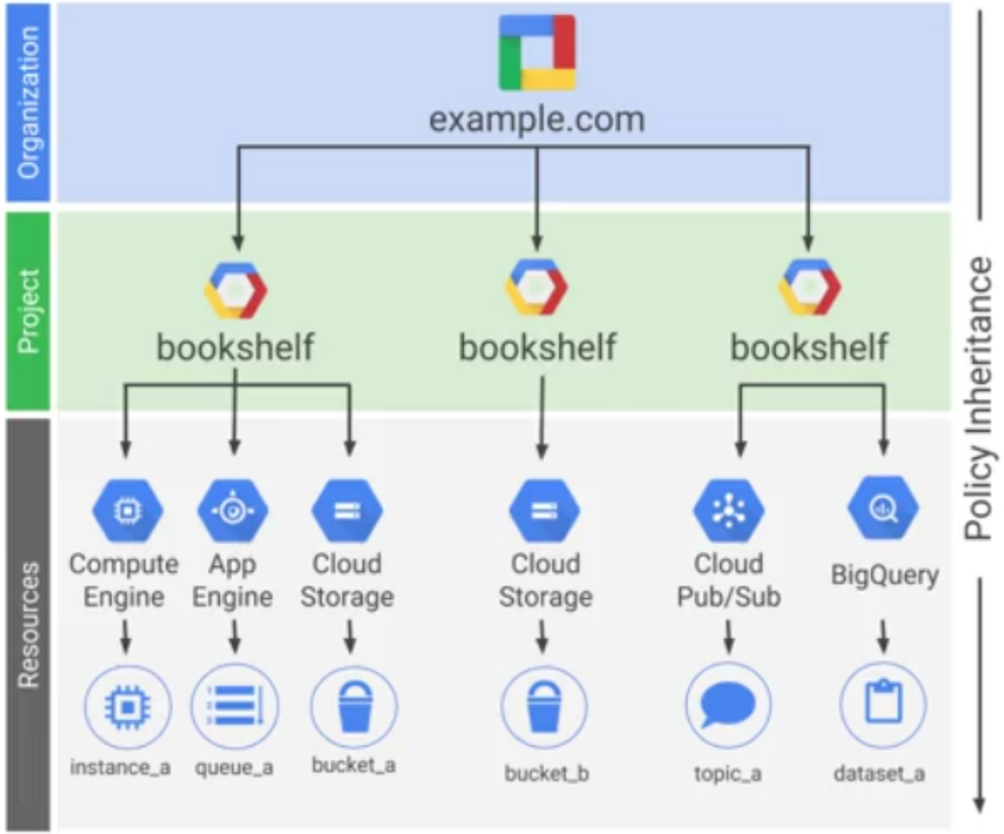
For instance, if you set a policy at the organization level, it is automatically inherited by all its children projects. And this inheritance is transitive, which means that all the resources in those projects inherit the policy too. There's one important rule to keep in mind.

The policies implemented at a higher level in this hierarchy can't take away access that's granted at a lower level.

For example, suppose that a policy applied on the bookshelf project gives user Pat the right to modify a cloud

An example IAM resource hierarchy

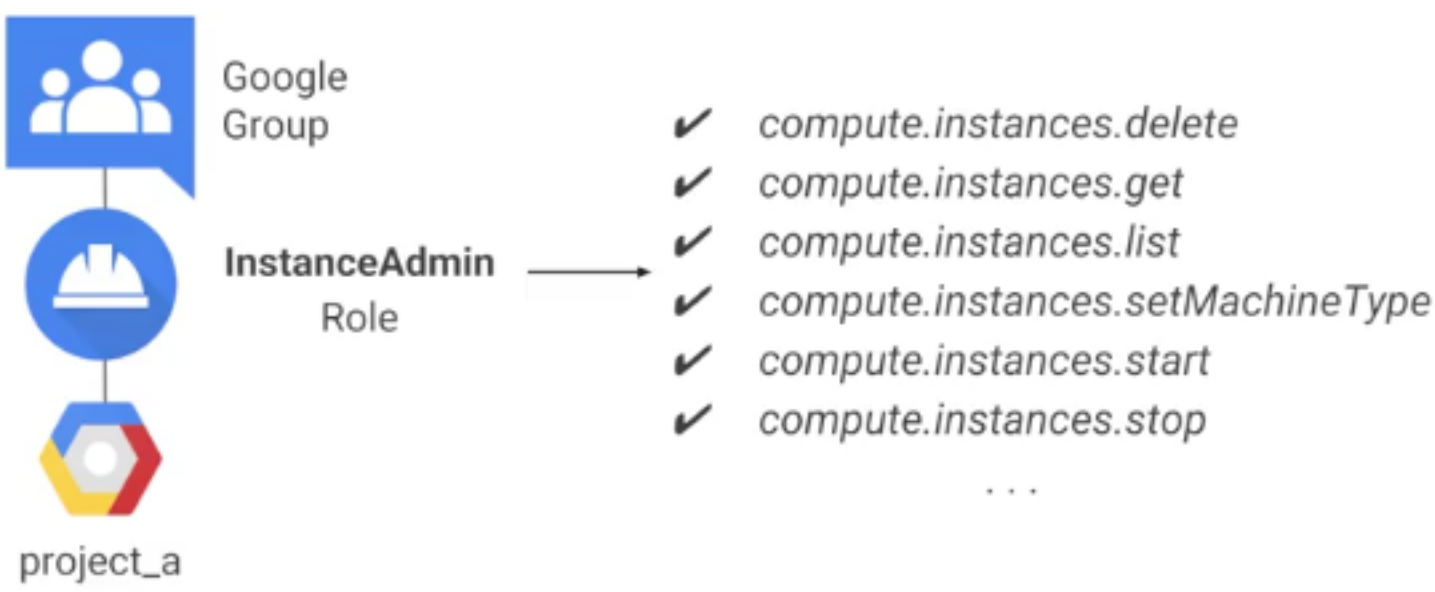
- A policy is set on a resource.
  - Each policy contains a set of roles and role members.
- Resources inherit policies from parent.
  - Resource policies are a union of parent and resource.
- A less restrictive parent policy overrides a more restrictive resource policy.



but a policy at the organization level says that Pat can only

IAM roles

IAM predefined roles offer more fine-grained permissions on particular services



Compute Engines InstanceAdmin Role lets whoever has that role perform a certain set of actions on virtual machines. The actions are: listing them, reading and changing their configurations, and starting and stopping them. And which virtual machines? Well, that depends on where the roles apply.

In this example, all the users of a certain Google Group have the role, and they have it on all the virtual machines in project_a. If you need something even finer-grained, there are custom roles.

A lot of companies have a

Service Accounts and IAM

- Service accounts authenticate using keys.
  - Google manages keys for Compute Engine and App Engine.
- You can assign a predefined or custom IAM role to the service account.

to authenticate your VM to cloud storage. Service accounts are named with an email address. But instead of passwords, they use cryptographic keys to access resources. In this simple example, a service account has been granted Compute Engine's InstanceAdmin Role.

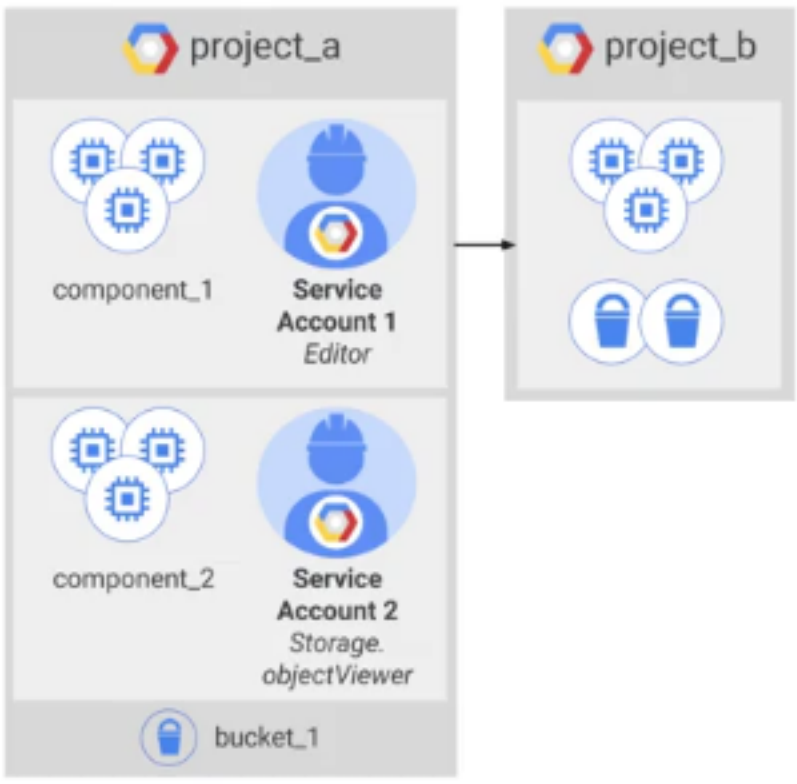
This would allow an application running in a VM with that service account to create, modify, and delete other VMs. Incidentally, service accounts need to be managed, too.

For example, maybe Alice needs to manage what can be act as a given

IAM roles

Example: Service Accounts and IAM

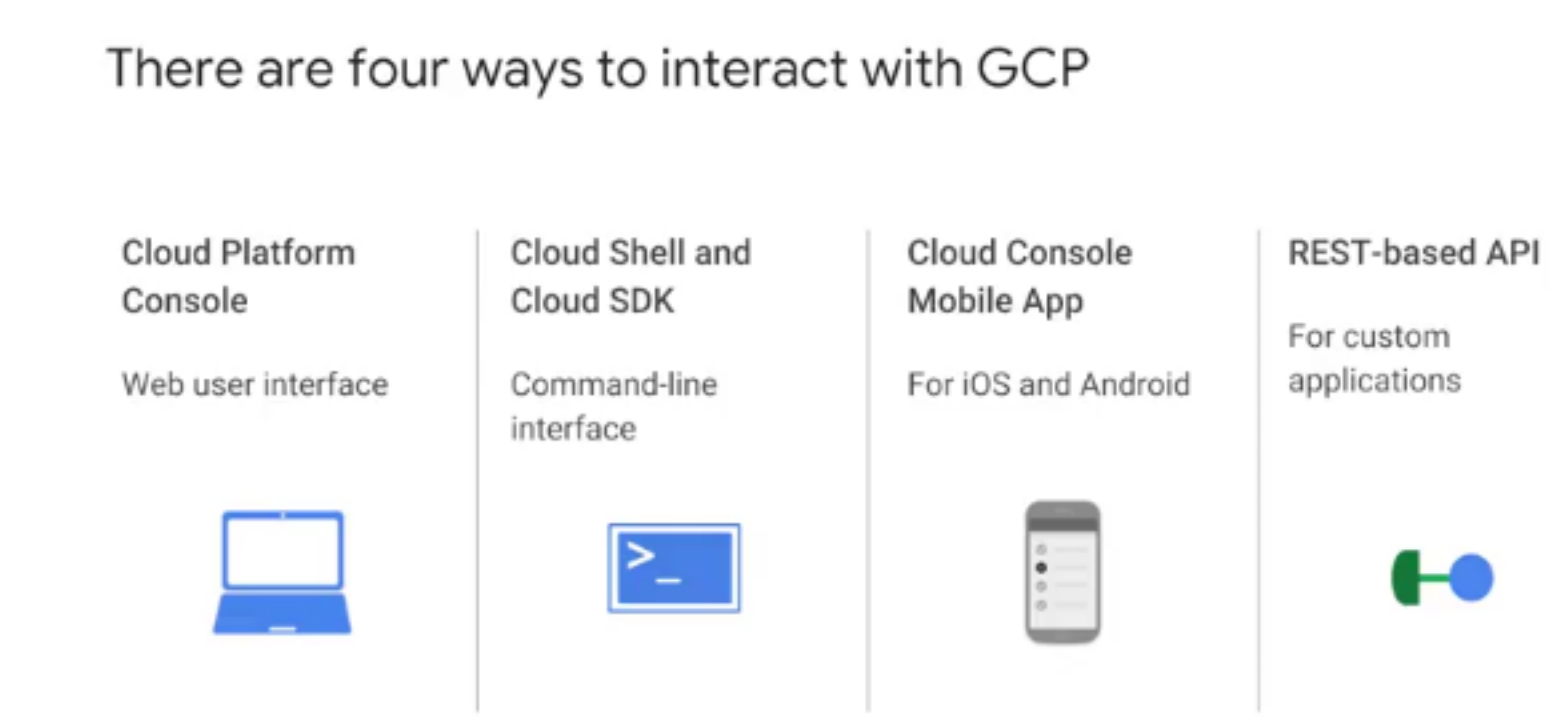
- VMs running component_1 are granted **Editor** access to project_b using Service Account 1.
- VMs running component_2 are granted **objectViewer** access to bucket_1 using Service Account 2.
- Service account permissions can be changed without recreating VMs.



having to recreate the VMs. Here's a more complex example. Say you have an application that's implemented across a group of Compute Engine virtual machines.

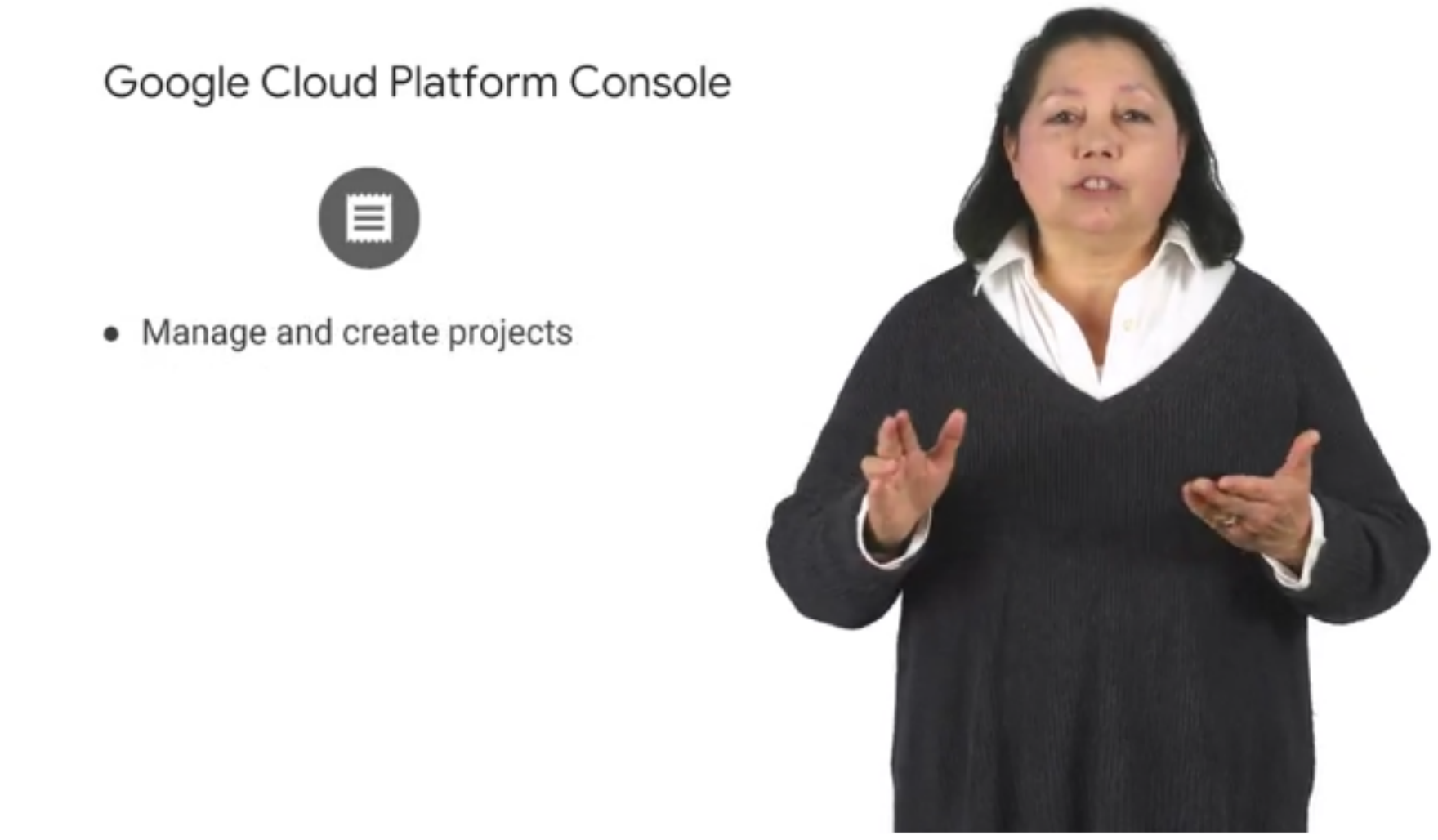
One component of your application needs to have an editor role on another project, but another component doesn't. So you would create two different service accounts, one for each subgroup of virtual machines. Only the first service account has privilege on the other project.

That reduces the potential impact of a miscoded application or a compromised virtual machine.



There are four ways you can interact with Google Cloud Platform, and we'll talk about each in turn: the Console, the SDK and Cloud Shell, the Mobile App and the APIs. The GCP Console is a web-based administrative interface. If you build an application in GCP, you'll use it.

Although, the end users of your application won't. It lets you view and manage all your projects and all the resources they use. It also lets you enable, disable and explore the APIs of GCP services. And it gives you access to Cloud Shell.



Although, the end users of your application won't. It lets you view and manage all your projects and all the resources they use. It also lets you enable, disable and explore the APIs of GCP services. And it gives you access to Cloud Shell.

That's a command-line interface to GCP that's easily accessed from your browser. From Cloud Shell, you can use the tools provided by the Google Cloud Software Development kit SDK, without having to first install them somewhere. What's the Software Development Kit? We'll talk about that

Google Cloud Platform Console

- Access to Google Cloud Platform APIs

You're using Co

Google Cloud Platform Console

- Offers access to Cloud Shell
  - A temporary virtual machine with Google Cloud SDK preinstalled

Google Cloud SDK

- Includes command-line tools for Cloud Platform products and services
  - gcloud, gsutil (Cloud Storage), bq (BigQuery)

The Google Cloud SDK is a set of tools that you can use to

Use APIs Explorer to help you write your code

- The APIs Explorer is an interactive tool that lets you easily try Google APIs using a browser.
- With the APIs Explorer, you can:
  - Browse quickly through available APIs and versions.
  - See methods available for each API and what parameters they support along with inline documentation.
  - Execute requests for any method and see responses in real time.
  - Easily make authenticated and authorized API calls.

It lets you see what APIs are available and in what versions.

Cloud Console Mobile App

- Manage virtual machines and database instances
- Manage apps in Google App Engine
- Manage your billing
- Visualize your projects with a customizable dashboard

Finally, one more tool that's of interest to everyone, not just developers.

Cloud Launcher gives quick access to solutions

- A solution marketplace containing pre-packaged, ready-to-deploy solutions
  - Some offered by Google
  - Others by third-party vendors

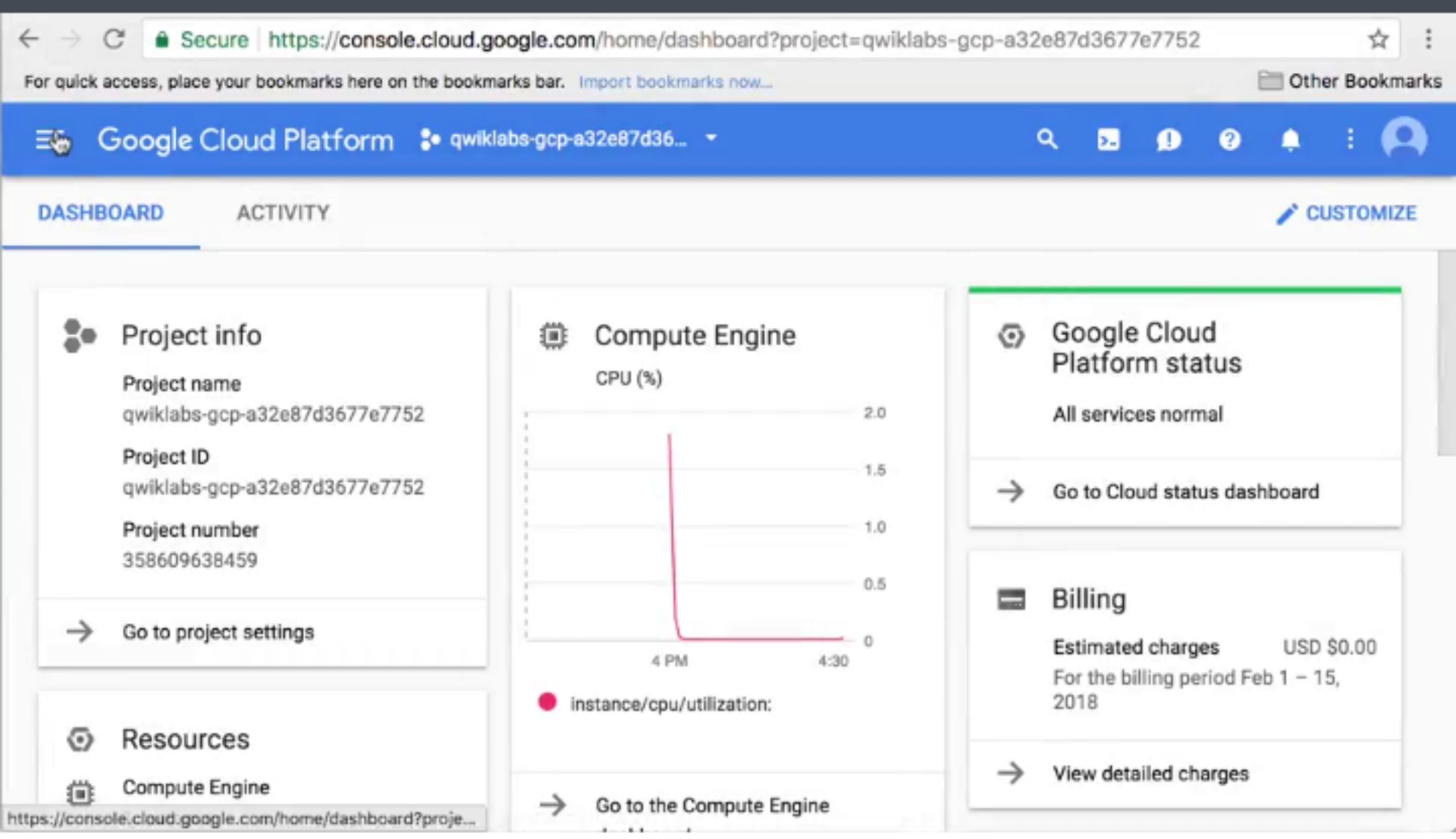
It's a tool for quickly deploying

Some Cloud Launcher images charge users fees, particularly those published by third parties with commercially licensed software. But they all show you estimates of their monthly charges before you launch them. Be aware that these estimates are just that, estimates.

In particular, they don't attempt to estimate networking costs since those will vary based on how you use the applications. A second note of caution. GCP updates the base images for these software packages to fix critical issues and vulnerabilities.

Some Cloud Launcher images charge users fees, particularly those published by third parties with commercially licensed software. But they all show you estimates of their monthly charges before you launch them. Be aware that these estimates are just that, estimates.

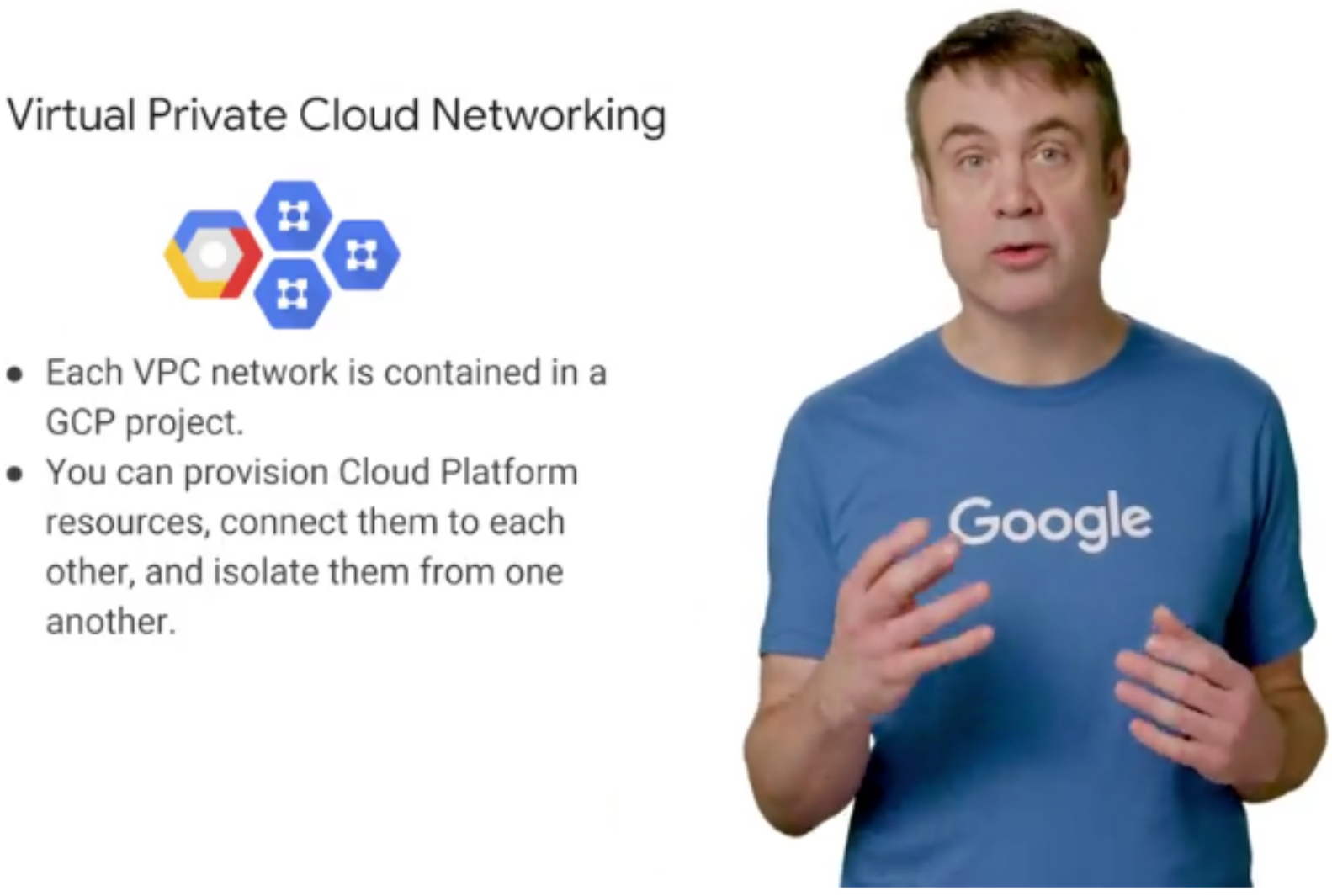
In particular, they don't attempt to estimate networking costs since those will vary based on how you use the applications. A second note of caution. GCP updates the base images for these software packages to fix critical issues and vulnerabilities.



In this demonstration, I'll use Cloud Launcher to deploy a solution on Google Cloud platform. The solution I've chosen is a LAMP stack. LAMP stands for Linux, Apache, MySQL, PHP. It's an easy environment for developing web applications.

I'll use Cloud Launcher to deploy that Stack into a Compute Engine Instance. In the GCP Console's Products and Services menu, I click Cloud Launcher. In the Search Bar, I type LAMP. LAMP Stacks are environments for web development. Notice that estimated costs are provided.

Now I click Launch On Compute Engine. I leave the deployment name

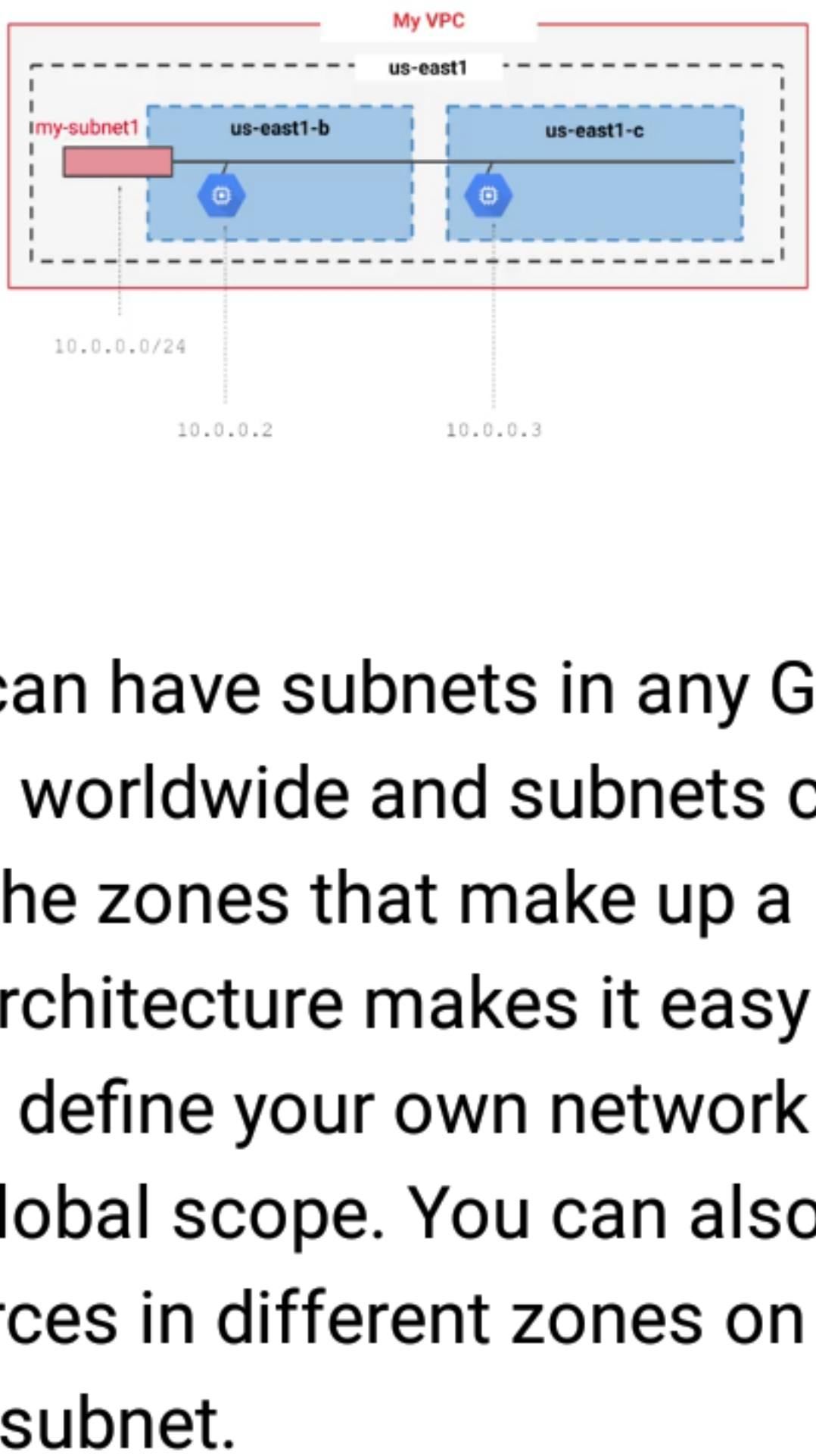


The way a lot of people get started with GCP is to define their own Virtual Private Cloud inside their first GCP project, or they can simply choose the default VPC and get started with that. Regardless, your VPC networks connect your Google Cloud platform resources to each other and to the internet.

You can segment your networks, use firewall rules to restrict access to instances, and create static routes to forward traffic to specific destinations. Here's something that surprises a lot of people who are new to GCP. The Virtual Private Cloud networks that you define have global scope.

Virtual Private...(VPC) Network

Google Cloud VPC networks are global; subnets are regional



They can have subnets in any GCP region worldwide and subnets can span the zones that make up a region. This architecture makes it easy for you to define your own network layout with global scope. You can also have resources in different zones on the same subnet.

You can dynamically increase the size of a subnet in a custom network by expanding the range of IP addresses allocated to it. Doing that doesn't affect already configured VMs. In this example, your VPC has one network. So far, it has one subnet defined in GCP us-east1 region.

Compute Engine

Compute Engine offers managed virtual machines

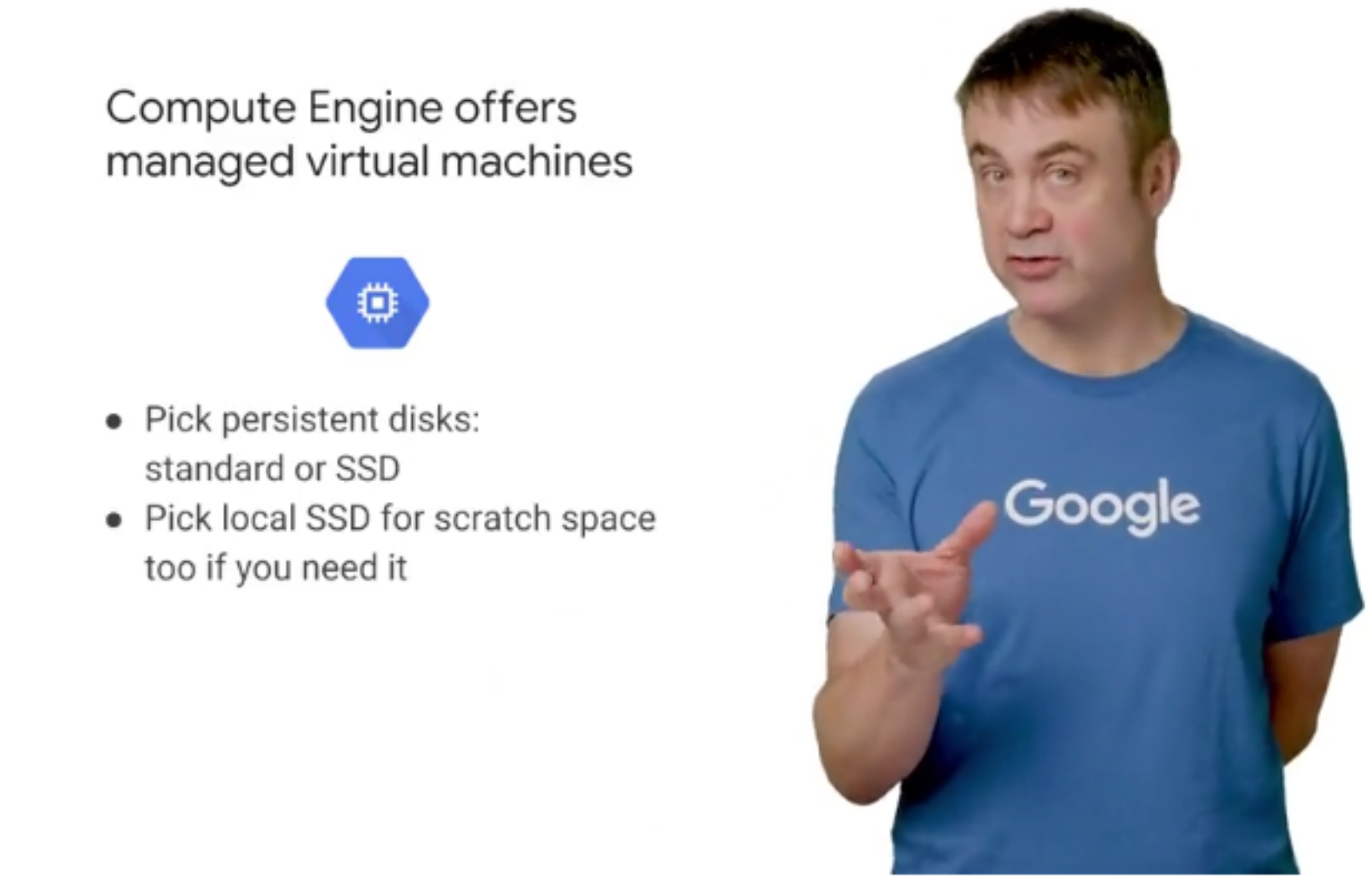
- Pick memory and CPU: use predefined types, or make a custom VM
- Pick GPUs if you need them

When you create a VM, pick a machine type which determines how much memory and how many virtual CPUs it has. These types range from very small to very large indeed. If you can't find a predefined type that meets your needs perfectly, you can make a custom VM.

Speaking of processing power, if you have workloads like machine learning and data processing that can take advantage of GPUs, many GCP zones have GPU's available for you. Just like physical computers need disks, so do VM. You can choose two kinds of persistent storage; standard or SSD.

If your application needs

Compute Engine

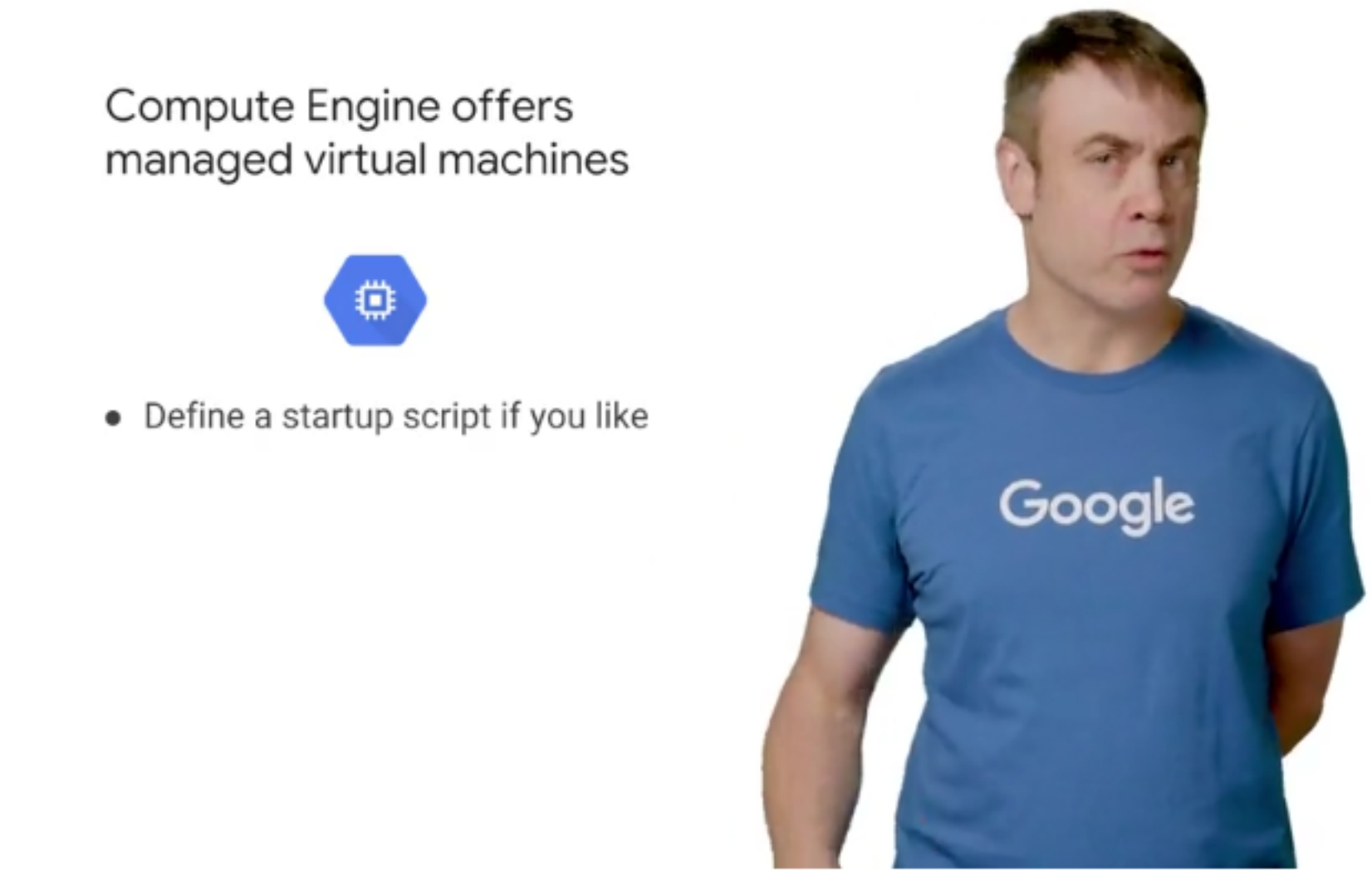


If your application needs high-performance scratch space, you can attach a local SSD, but be sure to store data of permanent value somewhere else because local SSDs content doesn't last past when the VM terminates. That's why the other kinds are called persistent disks.

Anyway, most people start off with standard persistent disks and that's the default. You'll also choose a boot image. GCP offers lots of versions of Linux and Windows ready to go and you can import your own images too.

Lots of GCP customers want their VMs to always come up with certain configurations like installing software

Compute Engine



Lots of GCP customers want their VMs to always come up with certain configurations like installing software packages on first boot. It's very common to pass GCP VM startup scripts that do just that. You can also pass in other kinds of metadata too.

Once your VMs are running, it's easy to take a durable snapshot of their disks. You can keep these as backups or use them when you need to migrate a VM to another region.

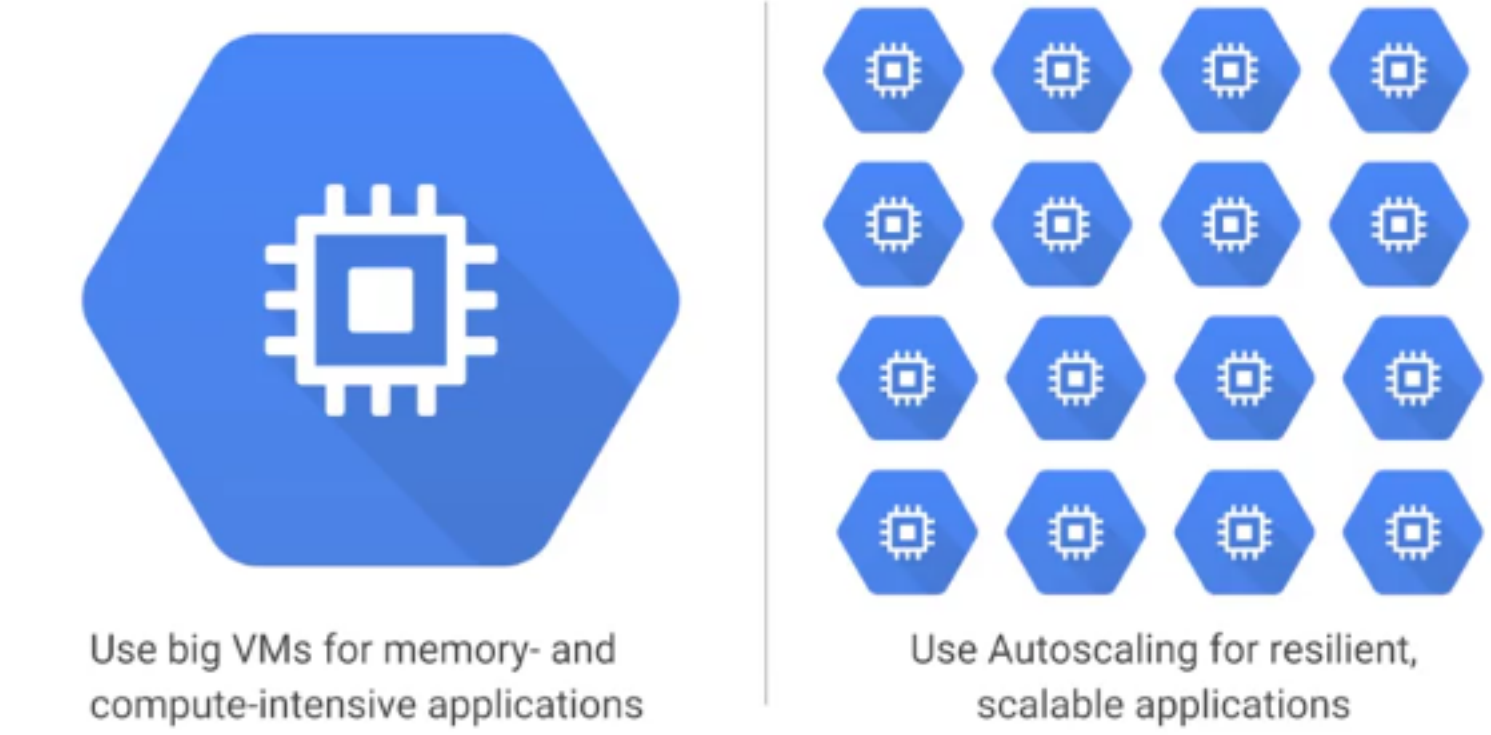
Suppose you have a workload that no human being is sitting around waiting to finish, say a batch job analyzing large dataset, you can save money by choosing preemptible VMs to run

Compute Engine

large dataset, you can save money by choosing preemptible VMs to run the job. A preemptible VM is different from an ordinary Compute Engine VM in only one respect.

You've given compute engine permission to terminate it if it's resources are needed elsewhere. You can save a lot of money with preemptible VMs, although be sure to make your job able to be stopped and restarted.

You can choose the machine properties of your instances such as the number of virtual CPUs and the amount of memory by using a set of predefined machine types or by creating your own custom machine

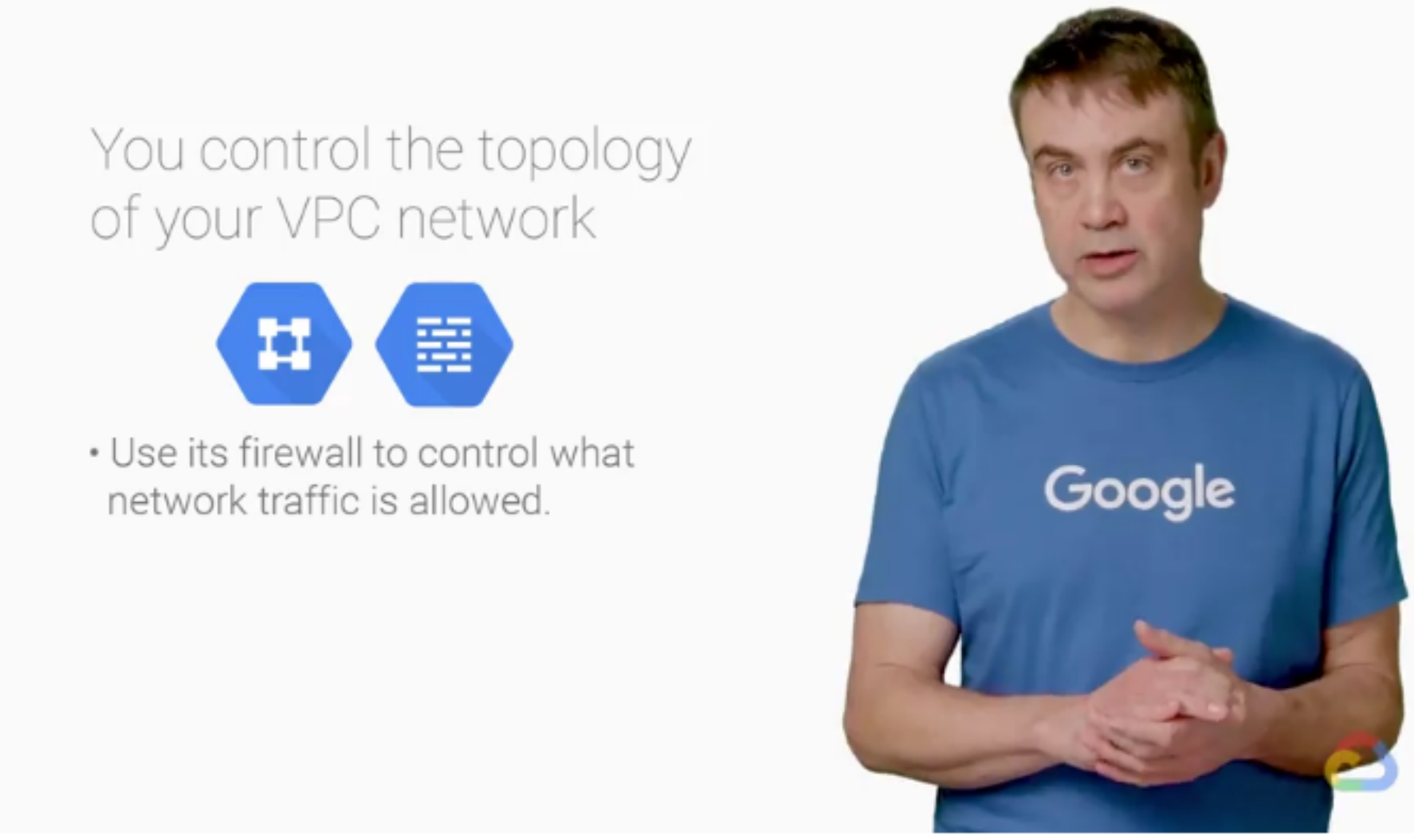


types. I mentioned a bit ago that you can make very large VMs in compute engine.

At the time this video was produced, the maximum number of virtual CPUs and the VM was 96 and the maximum memory size was in beta at 624 gigabytes. Check the GCP website to see where these maximums are today.

These huge VMs are great for work-loads like in-memory databases and CPU intensive analytics, but most GCP customers start off with scaling out not scaling up. Compute Engine has a feature called auto scaling that lets you add and take away VMs from your application based on load metrics.

Important VPC capabilities

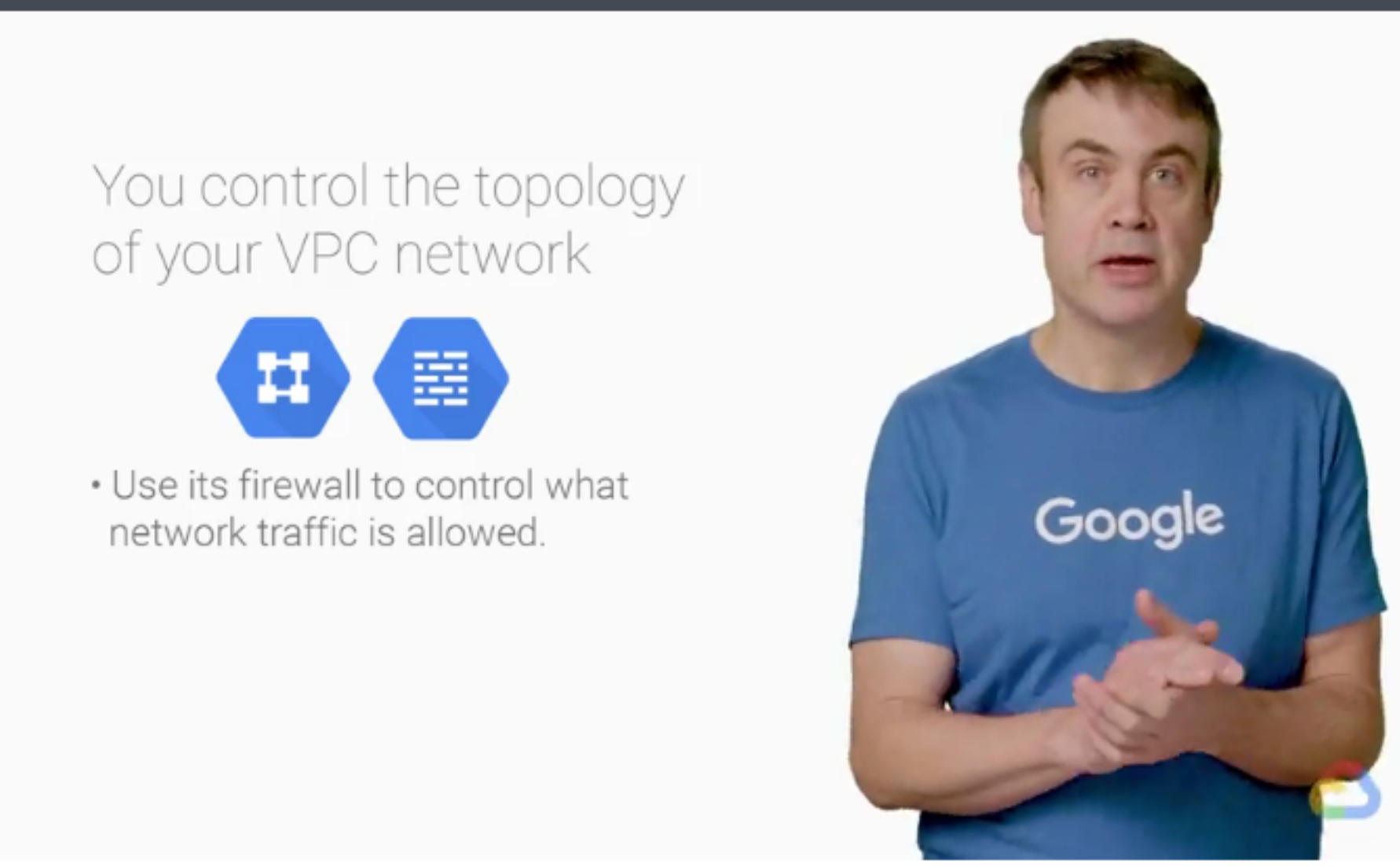


Much like physical networks, VPCs have routing tables. These are used to forward traffic from one instance to another instance within the same network. Even across sub-networks and even between GCP zones without requiring an external IP address.

VPCs routing tables are built in, you don't have to provision or manage a router. Another thing you don't have to provision or manage for GCP, a firewall instance. VPCs give you a global distributed firewall. You can control to restrict access to instances, both incoming and outgoing traffic.

You can define firewall rules in terms

Important VPC capabilities

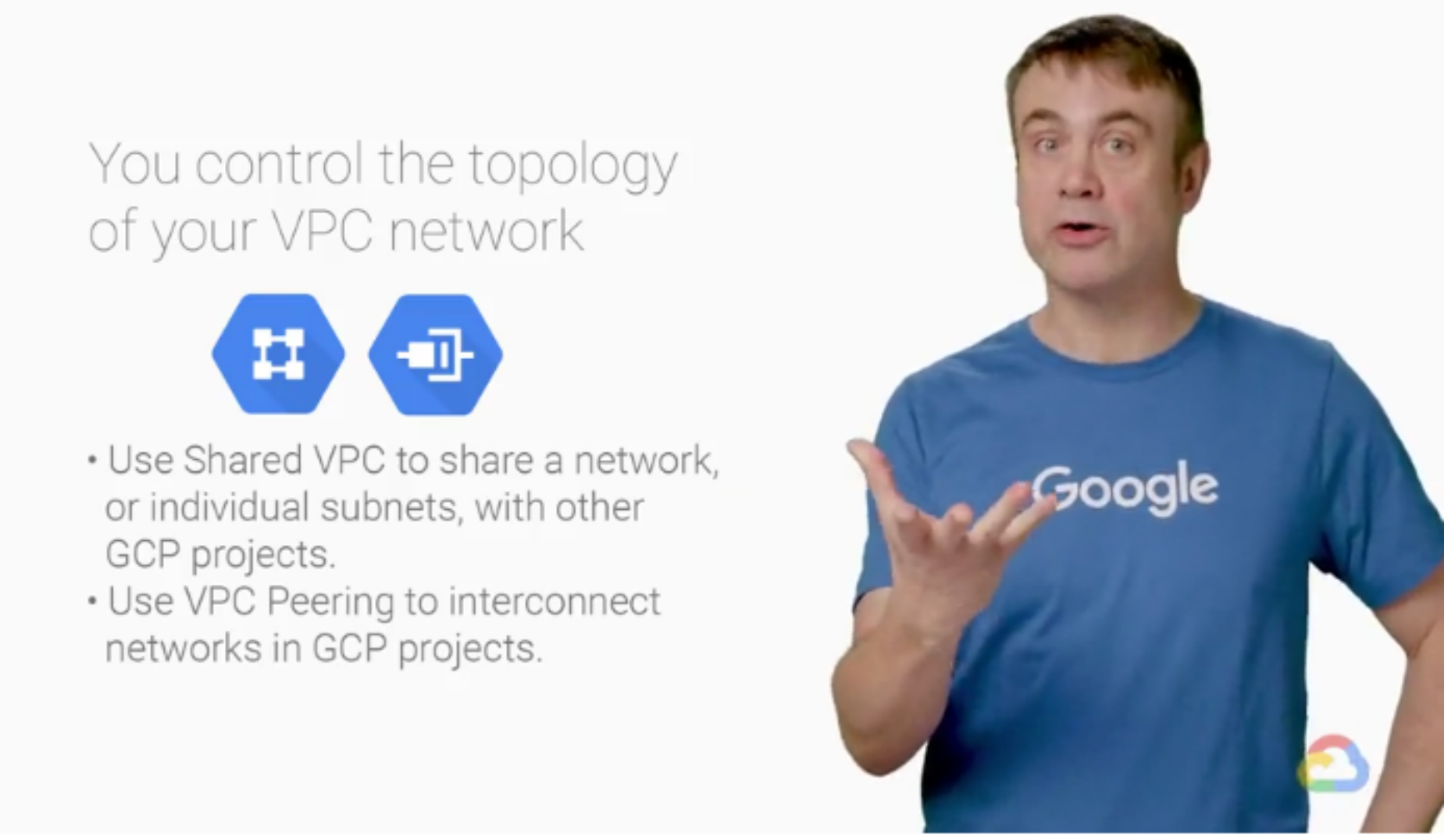


For example, you can tag all your web servers with say, "web," and write a firewall rule saying that traffic on ports 80 or 443 is allowed into all VMs with the "web" tag, no matter what their IP address happens to be. Remember, I mentioned that VPCs belong to GCP projects.

But what if your company has several GCP projects and the VPCs need to talk to each other? Don't worry, that's totally doable and manageable. If you simply want to establish a peering relationship between two VPCs so that they can exchange traffic, that's what VPC Peering does.

On the other hand, if you want to use

Important VPC capabilities



simply want to establish a peering relationship between two VPCs so that they can exchange traffic, that's what VPC Peering does.

On the other hand, if you want to use the full power of IAM to control who and what in one project can interact with a VPC in another, that's what Shared VPC is for. A few slides back, we talked about how virtual machines can auto-scale to respond to changing load.

But how do your customers get to your application when it might be provided by four VMs one moment

Important VPC capabilities

With global Cloud Load Balancing, your application presents a single front-end to the world

- Users get a single, global anycast IP address.
- Traffic goes over the Google backbone from the closest point-of-presence to the user.
- Backends are selected based on load.
- Only healthy backends receive traffic.
- No pre-warming is required.

Shared VPC is for. A few slides back, we talked about how virtual machines can auto-scale to respond to changing load.

But how do your customers get to your application when it might be provided by four VMs one moment and 40 VMs at another? Cloud Load Balancing is the answer. Cloud Load Balancing is a fully distributed, software-defined managed service for all your traffic.

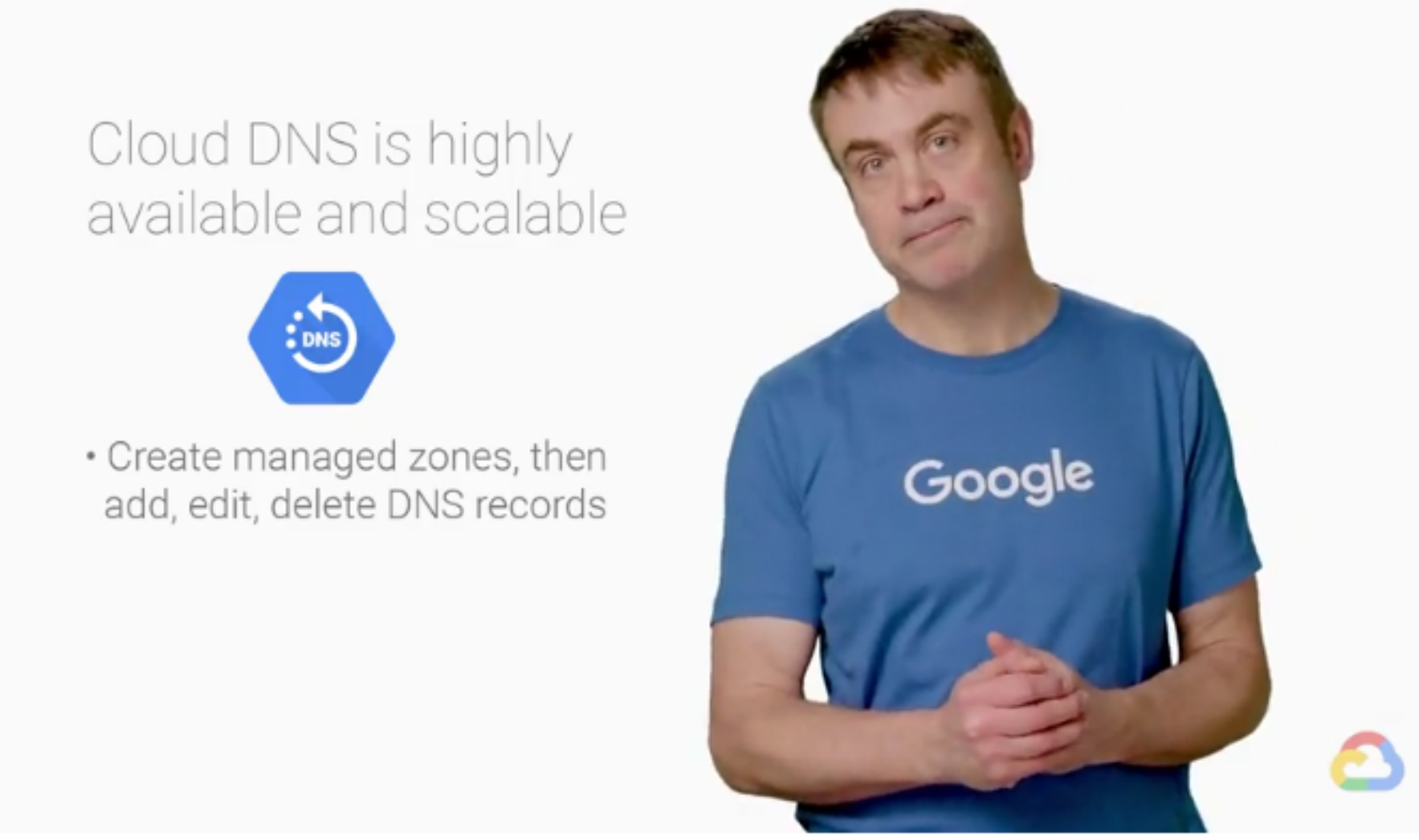
And because the load balancers don't run in VMs you have to manage, you don't have to worry about scaling or managing them. You can put Cloud Load Balancing in front of all your

Google VPC offers a suite of load-balancing options

Global HTTP(S)	Global SSL Proxy	Global TCP Proxy	Regional	Regional internal
Layer 7 load balancing based on load	Layer 4 load balancing of non-HTTPS SSL traffic based on load	Layer 4 load balancing of non-SSL TCP traffic	Load balancing of any traffic (TCP, UDP)	Load balancing of traffic inside a VPC
Can route different URLs to different back ends	Supported on specific port numbers	Supported on specific port numbers	Supported on any port number	Use for the internal tiers of multi-tier applications

For Secure Sockets Layer traffic that is not HTTP,

Important VPC capabilities



One of the most famous Google services that people don't pay for is 8.8.8.8, which provides a public domain name service to the world. DNS is what translates internet host names to addresses. And as you would imagine, Google has a highly developed DNS infrastructure.

It makes 8.8.8.8 available so that everybody can take advantage of it. But what about the internet host names and addresses of applications you build in GCP? GCP offers Cloud DNS to help the world find them. It's a managed DNS service running on the same infrastructure as Google.

It has low latency and high availability

Important VPC capabilities

interface or the API. Google has a global system of edge caches. You can use this system to accelerate content delivery in your application using Google Cloud CDN.

Your customers will experience lower network latency. The origins of your content will experience reduced load and you can save money too. Once you've set up HTTPS load balancing, simply enable Cloud CDN with a single checkbox. There are lots of other CDNs out there of course.

What if you're already using one? Chances are, your CDN is a part of GCPs, CDN interconnect partner

Google Cloud Platform offers many interconnect options

VPN

Secure multi-Gbps connection over VPN tunnels

Direct Peering

Private connection between you and Google for your hybrid cloud workloads

Carrier Peering

Connection through the largest partner network of service providers

Dedicated Interconnect

Connect N X 10G transport circuits for private cloud traffic to Google Cloud at Google POPs

to use it. Lots of GCP customers want to interconnect their other networks to their Google VPCs, such as on-premises networks or their networks in other clouds.

There are many good choices. Many customers start with a Virtual Private Network connection over the internet using the IPSEC protocol. To make that dynamic, they use a GCP feature called Cloud Router.

Cloud Router lets your other networks and your Google VPC exchange route information over the VPN using the Border Gateway Protocol. For instance, if you add a new subnet to

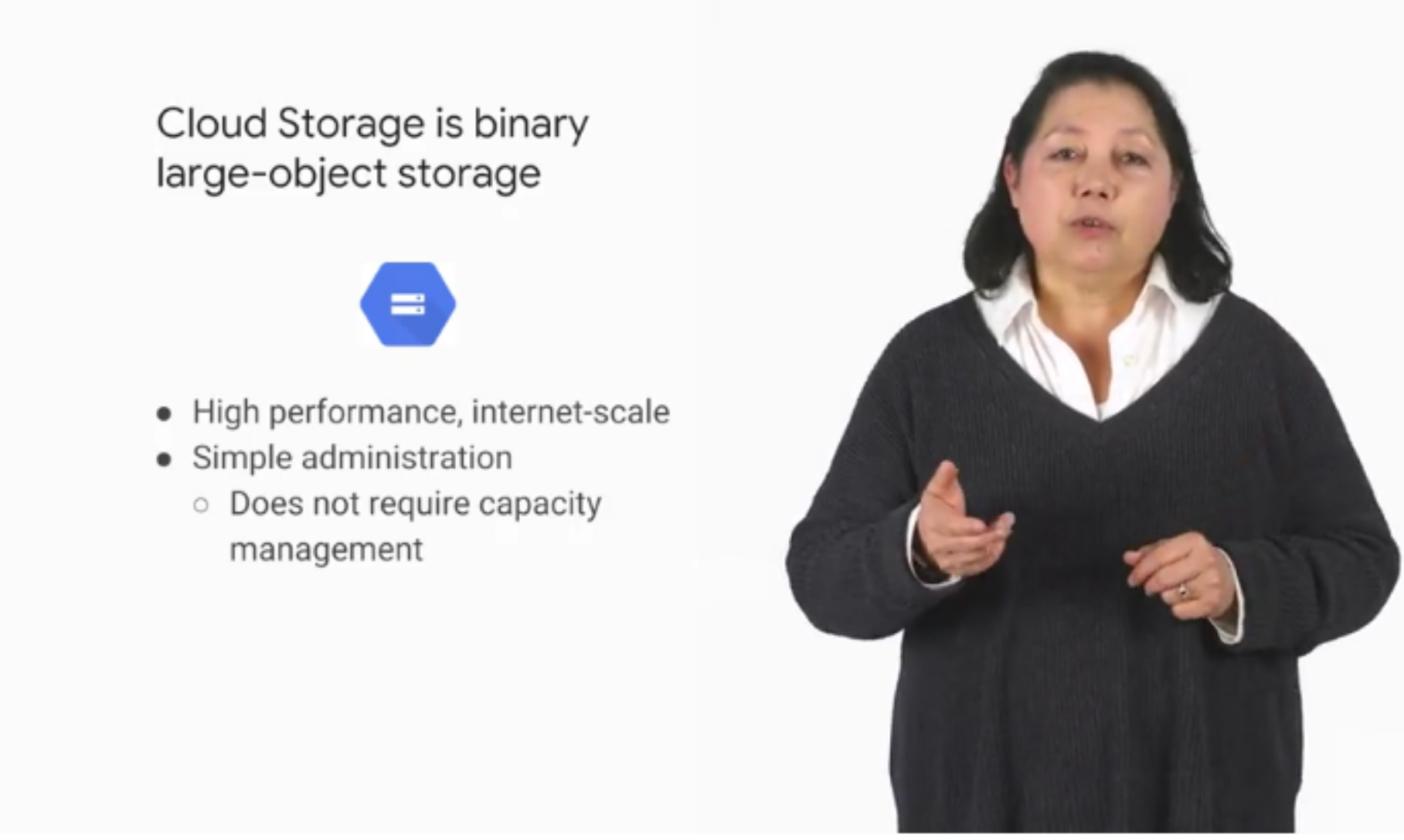
2600 CP. Different applications and workloads required different storage database solutions.

You already know that you can store data on your VM's persistent disk. Google Cloud Platform has other storage options to meet your needs for structured, unstructured, transactional and relational data.

In this module, I'll tell you about the core storage options: Cloud Storage, Cloud SQL, Cloud Spanner, Cloud Data Store and Google Big Table. Depending on your application, you might want to use one or several of these services to get the job done.

In this module, I'll tell you about the core storage options: Cloud Storage, Cloud SQL, Cloud Spanner, Cloud Data Store and Google Big Table. Depending on your application, you might want to use one or several of these services to get the job done.

Cloud Storage



Let's start with Google Cloud Storage. What's object storage? It's not the same as file storage, in which you manage your data as a hierarchy of folders. It's not the same as block storage, in which your operating system manages your data as chunks of disk.

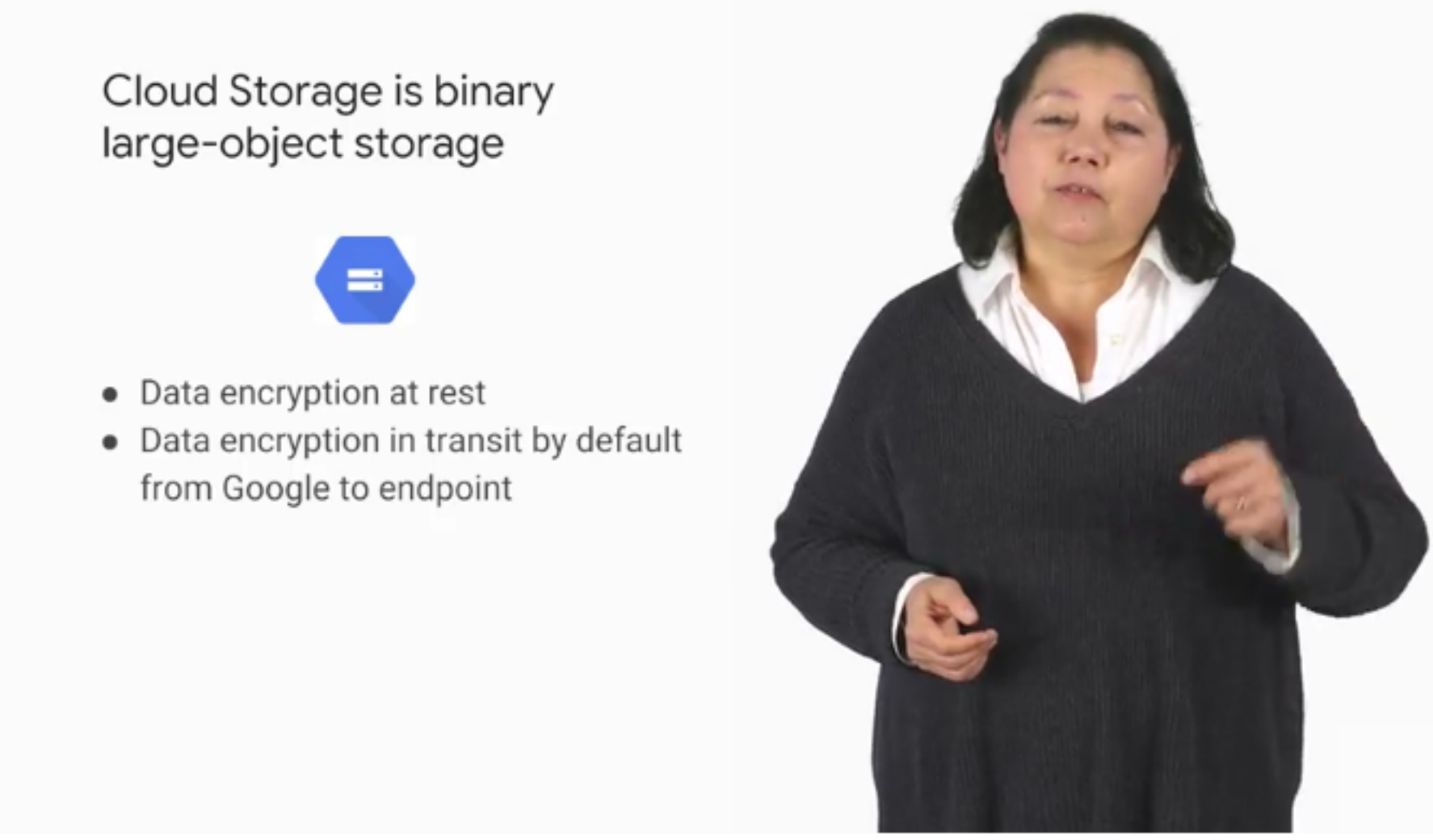
Instead, object storage means you save to your storage here, you keep this arbitrary bunch of bytes I give you and the storage lets you address it with a unique key. That's it. Often these unique keys are in the form of URLs which means object storage interacts nicely with Web technologies.

Cloud Storage

Cloud Storage works just like that, except better. It's a fully managed scalable service. That means that you don't need to provision capacity ahead of time. Just make objects and the service stores them with high durability and high availability.

You can use Cloud Storage for lots of things: serving website content, storing data for archival and disaster recovery, or distributing large data objects to your end users via Direct Download. Cloud Storage is not a file system because each of your objects in Cloud Storage has a URL.

Each feels like a file in a lot of ways



Cloud Storage always encrypts your data on the server side before it is written to disk and you don't pay extra for that. Also by default, data in-transit is encrypted using HTTPS.

Speaking of transferring data, there are services you can use to get large amounts of data into Cloud Storage conveniently. We'll discuss them later in this module. Once they are in Cloud Storage, you can move them onwards to other GCP storage services.

Just as I discussed, your Cloud Storage files are organized into buckets. When you create a bucket, you give it a globally unique name. You specify a

Just as I discussed, your Cloud Storage files are organized into buckets. When you create a bucket, you give it a globally unique name. You specify a geographic location where the bucket and its contents are stored and you choose a default storage class.

Pick a location that minimizes latency for your users. In other words, if most of your users are in Europe, you probably want to pick a European location. Speaking of your users, there are several ways to control access to your objects and buckets. For most purposes, Cloud IAM is sufficient.

Roles are inherited from project to bucket to object. If you need finer

Your Cloud Storage files are organized into buckets

Bucket attributes	Bucket contents
Globally unique name	Files (in a flat namespace)
Storage class	
Location (region or multi-region)	
IAM policies or Access Control Lists	Access Control Lists
Object versioning setting	
Object lifecycle management rules	

Roles are inherited from project to bucket to object. If you need finer control, you can create access control lists - ACLs - that offer finer control. ACLs define who has access to your buckets and objects as well as what level of access they have.

Each ACL consists of two pieces of information, a scope which defines who can perform the specified actions, for example, a specific user or group of users and a permission which defines what actions can be performed. For example, read or write.

Remember I mentioned that Cloud Storage objects are immutable. You

Remember I mentioned that Cloud Storage objects are immutable. You can turn on object versioning on your buckets if you want. If you do, Cloud Storage keeps a history of modifications. That is, it overrides or deletes all of the objects in the bucket.

You can list the archived versions of an object, restore an object to an older state or permanently delete a version as needed. If you don't turn on object versioning, new always overrides old. What if versioning sounds good to you but you're worried about junk accumulating?

Cloud Storage also offers lifecycle

an object, restore an object to an older state or permanently delete a version as needed. If you don't turn on object versioning, new always overrides old. What if versioning sounds good to you but you're worried about junk accumulating?

Cloud Storage also offers lifecycle management policies. For example, you could tell Cloud Storage to delete objects older than 365 days.

Or you could tell it to delete objects created before January 1, 2013 or keep only the three most recent versions of each object in a bucket that has versioning enabled.

Cloud Storage interactions

Choosing among Cloud Storage classes

	Multi-regional	Regional	Nearline	Coldline
Intended for data that is...	Most frequently accessed	Accessed frequently within a region	Accessed less than once a month	Accessed less than once a year
Availability SLA	99.95%	99.90%	99.00%	99.00%
Access APIs	***Consistent APIs***			
Access time	***Millisecond access***			
Storage price	Price per GB stored per month			
Retrieval price	Total price per GB transferred			
Use cases	Content storage and delivery	In-region analytics, transcoding	Long-tail content, backups	Archiving, disaster recovery

Cloud Storage lets you choose among four different types of storage classes: Regional, Multi-regional, Nearline, and Coldline. Here's how to think about them. Multi-regional and Regional are high-performance object storage, whereas Nearline and Coldline are backup and archival storage.

That's why I placed that heavy dividing line between these two groups. All of the storage classes are accessed in comparable ways using the cloud storage API and they all offer millisecond access times. Now, let's talk about how they differ.

Regional storage lets you store

Choosing among Cloud Storage classes

	Multi-regional	Regional	Nearline	Coldline
Intended for data that is...	Most frequently accessed	Accessed frequently within a region	Accessed less than once a month	Accessed less than once a year
Availability SLA	99.95%	99.90%	99.00%	99.00%
Access APIs	***Consistent APIs***			
Access time	***Millisecond access***			
Storage price	Price per GB stored per month			
Retrieval price	Total price per GB transferred			
Use cases	Content storage and delivery	In-region analytics, transcoding	Long-tail content, backups	Archiving, disaster recovery

Regional storage lets you store your data in a specific GCP region: US Central one, Europe West one or Asia East one. It's cheaper than Multi-regional storage but it offers less redundancy. Multi-regional storage on the other hand, cost a bit more but it's Geo-redundant.

That means you pick a broad geographical location like the United States, the European Union, or Asia and cloud storage stores your data in at least two geographic locations separated by at least 160 kilometers. Multi-regional storage is appropriate for storing frequently accessed data.

Cloud Storage interactions

Choosing among Cloud Storage classes

	Multi-regional	Regional	Nearline	Coldline
Intended for data that is...	Most frequently accessed	Accessed frequently within a region	Accessed less than once a month	Accessed less than once a year
Availability SLA	99.95%	99.90%	99.00%	99.00%
Access APIs	***Consistent APIs***			
Access time	***Millisecond access***			
Storage price	Price per GB stored per month			
Retrieval price	Total price per GB transferred			
Use cases	Content storage and delivery	In-region analytics, transcoding	Long-tail content, backups	Archiving, disaster recovery

For example, website content, interactive workloads, or data that's part of mobile and gaming applications. People use regional storage on the other hand, to store data close to their Compute Engine, virtual machines, or their Kubernetes engine clusters.

That gives better performance for data-intensive computations. Nearline storage is a low-cost, highly durable service for storing infrequently accessed data.

The storage class is a better choice than Multi-regional storage or Regional storage in scenarios where you plan to read or modify your data

Cloud Storage interactions

Choosing among Cloud Storage classes

	Multi-regional	Regional	Nearline	Coldline
Intended for data that is...	Most frequently accessed	Accessed frequently within a region	Accessed less than once a month	Accessed less than once a year
Availability SLA	99.95%	99.90%	99.00%	99.00%
Access APIs	***Consistent APIs***			
Access time	***Millisecond access***			
Storage price	Price per GB stored per month			
Retrieval price	Total price per GB transferred			
Use cases	Content storage and delivery	In-region analytics, transcoding	Long-tail content, backups	Archiving, disaster recovery

The storage class is a better choice than Multi-regional storage or Regional storage in scenarios where you plan to read or modify your data once a month or less on average.

For example, if you want to continuously add files to cloud storage and plan to access those files once a month for analysis, Nearline storage is a great choice. Coldline storage is a very low cost, highly durable service for data archiving, online backup, and disaster recovery.

Coldline storage is the best choice for data that you plan to access -at most - once a year. This is due to its slightly lower availability, 90-day

Choosing among Cloud Storage classes

	Multi-regional	Regional	Nearline	Coldline
Intended for data that is...	Most frequently accessed	Accessed frequently within a region	Accessed less than once a month	Accessed less than once a year
Availability SLA	99.95%	99.90%	99.00%	99.00%
Access APIs	***Consistent APIs***			
Access time	***Millisecond access***			
Storage price	Price per GB stored per month			
Retrieval price	Total price per GB transferred			
Use cases	Content storage and delivery	In-region analytics, transcoding	Long-tail content, backups	Archiving, disaster recovery

most - once a year. This is due to its slightly lower availability, 90-day minimum storage duration, costs for data access, and higher per operation costs.

For example, if you want to archive data or have access to it in case of a disaster recovery event. Availability of these storage classes varies with Multi-regional having the highest availability of 99.95 percent followed by Regional with 99.9 percent and Nearline and Coldline with 99 percent.

As for pricing, all storage classes in-cur a cost per gigabyte of data stored per month, with Multi-regional having the highest storage price and Coldline

Cloud Storage interactions

Choosing among Cloud Storage classes

	Multi-regional	Regional	Nearline	Coldline
Intended for data that is...	Most frequently accessed	Accessed frequently within a region	Accessed less than once a month	Accessed less than once a year
Availability SLA	99.95%	99.90%	99.00%	99.00%
Access APIs	***Consistent APIs***			
Access time	***Millisecond access***			
Storage price	Price per GB stored per month			
Retrieval price	Total price per GB transferred			
Use cases	Content storage and delivery	In-region analytics, transcoding	Long-tail content, backups	Archiving, disaster recovery

As for pricing, all storage classes incur a cost per gigabyte of data stored per month, with Multi-regional having the highest storage price and Coldline the lowest storage price. Egress and data transfer charges may also apply.

In addition to those charges, Nearline storage also incurs an access fee per gigabyte of data read and Coldline storage incurs a higher fee per gigabyte of data read. Regardless of which storage class you choose, there are several ways to bring data into cloud storage.

Many customers simply use gsutil which is the cloud storage command

Cloud Storage interactions

There are several ways to bring data into Cloud Storage

Online transfer

Self-managed copies using command-line tools or drag-and-drop

Storage Transfer Service

Scheduled, managed batch transfers

Transfer Appliance Beta

Rackable appliances to securely ship your data

Many customers simply use gsutil which is the cloud storage command from this cloud SDK. You can also move data in with a drag and drop in the GCP console, if you use the Google Chrome browser. But what if you have to upload terabytes or even petabytes of data?

Google Cloud platform offers the online storage transfer service and the offline transfer appliance to help. The storage transfer service lets you schedule and manage batch transfers to cloud storage from another cloud provider from a different cloud storage region or from an HTTPS endpoint.

Cloud Storage interactions

There are several ways to bring data into Cloud Storage

Online transfer

Self-managed copies using command-line tools or drag-and-drop

Storage Transfer Service

Scheduled, managed batch transfers

Transfer Appliance Beta

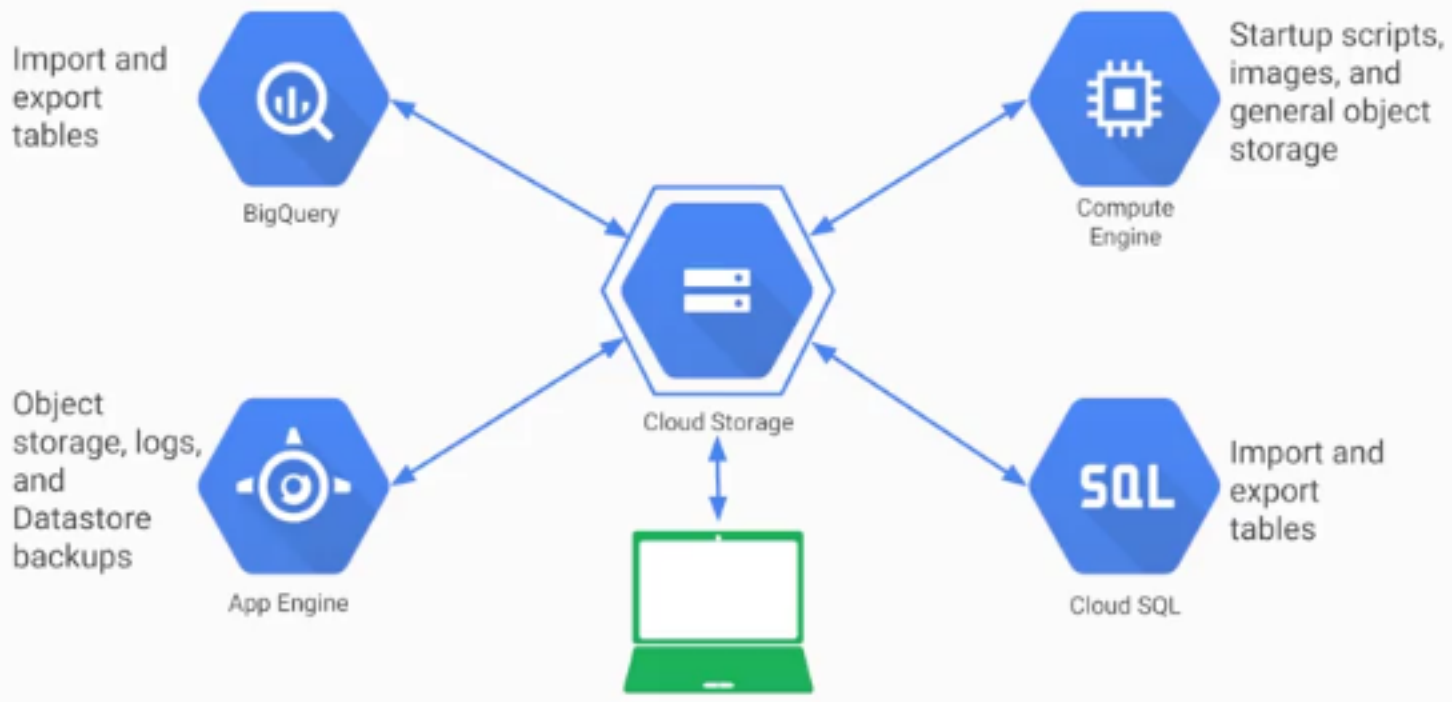
Rackable appliances to securely ship your data

Google Cloud platform offers the online storage transfer service and the offline transfer appliance to help. The storage transfer service lets you schedule and manage batch transfers to cloud storage from another cloud provider from a different cloud storage region or from an HTTPS endpoint.

The transfer appliance is a rackable, high-capacity storage server that you lease from Google Cloud. You simply connect it to your network, load it with data, and then ship it to an upload facility where the data is uploaded to cloud storage.

This service enables you to securely

Cloud Storage works with other GCP services

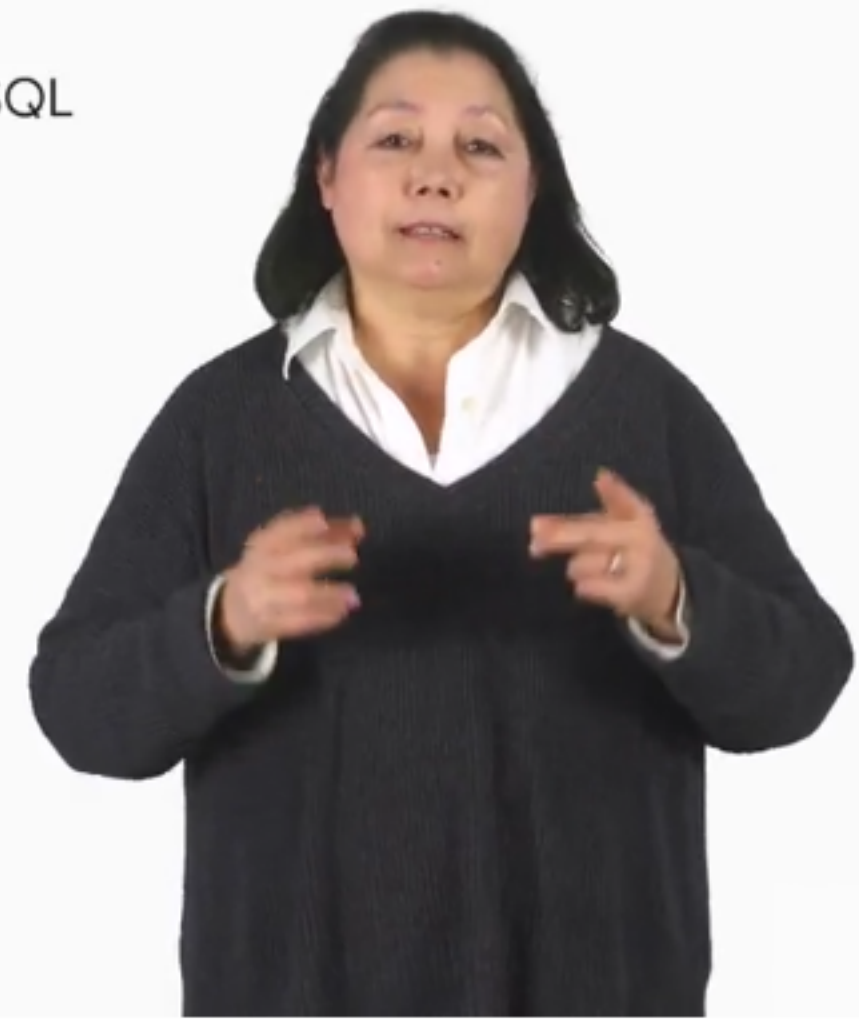


There are other ways of getting your data into cloud storage as this storage option is tightly integrated with many of the Google cloud platform products and services. For example, you can import and export tables from and to BigQuery as well as Cloud SQL.

You can also store App Engine logs, cloud data store backups, and objects used by App Engine applications like images. Cloud storage can also store instant startup scripts, Compute Engine images, and objects used by Compute Engine applications.

In short, cloud storage is often the ingestion point for data being moved into the cloud and is frequently the

Google Cloud Bigtable



Cloud Bigtable is Google's NoSQL, big data database service. What is NoSQL mean? Well, this isn't a database course, so I'll give you a very informal picture.

Think first of a relational database as offering you tables in which every row has the same set of columns, and the database engine enforces that rule and other rules you specify for each table. That's called the database schema.

An enforce schema is a big help for some applications and a huge pain for others. Some applications call for a much more flexible approach. For example, a NoSQL schema. In other

a much more flexible approach. For example, a NoSQL schema. In other words, for these applications not all the rows might need to have the same columns.

And in fact, the database might be designed to take advantage of that by sparsely populating the rows. That's part of what makes a NoSQL database what it is. Which brings us to Bigtable.

Your databases in Bigtable are sparsely populated tables that can scale to billions of rows and thousands of columns allowing you to store petabytes of data. GCP fully manages the surface, so you don't

Google Cloud Bigtable

Cloud Bigtable is managed NoSQL

- Fully managed NoSQL, wide-column database service for terabyte applications

thousands of columns allowing you to store petabytes of data. GCP fully manages the surface, so you don't have to configure and tune it. It's ideal for data that has a single lookup key.

Some applications developers think of Bigtable as a persistent hash table. Cloud Bigtable is ideal for storing large amounts of data with very low latency.

It supports high throughput, both read and write, so it's a great choice for both operational and analytical applications including Internet of Things, user analytics and financial data analysis.

Google Cloud Bigtable

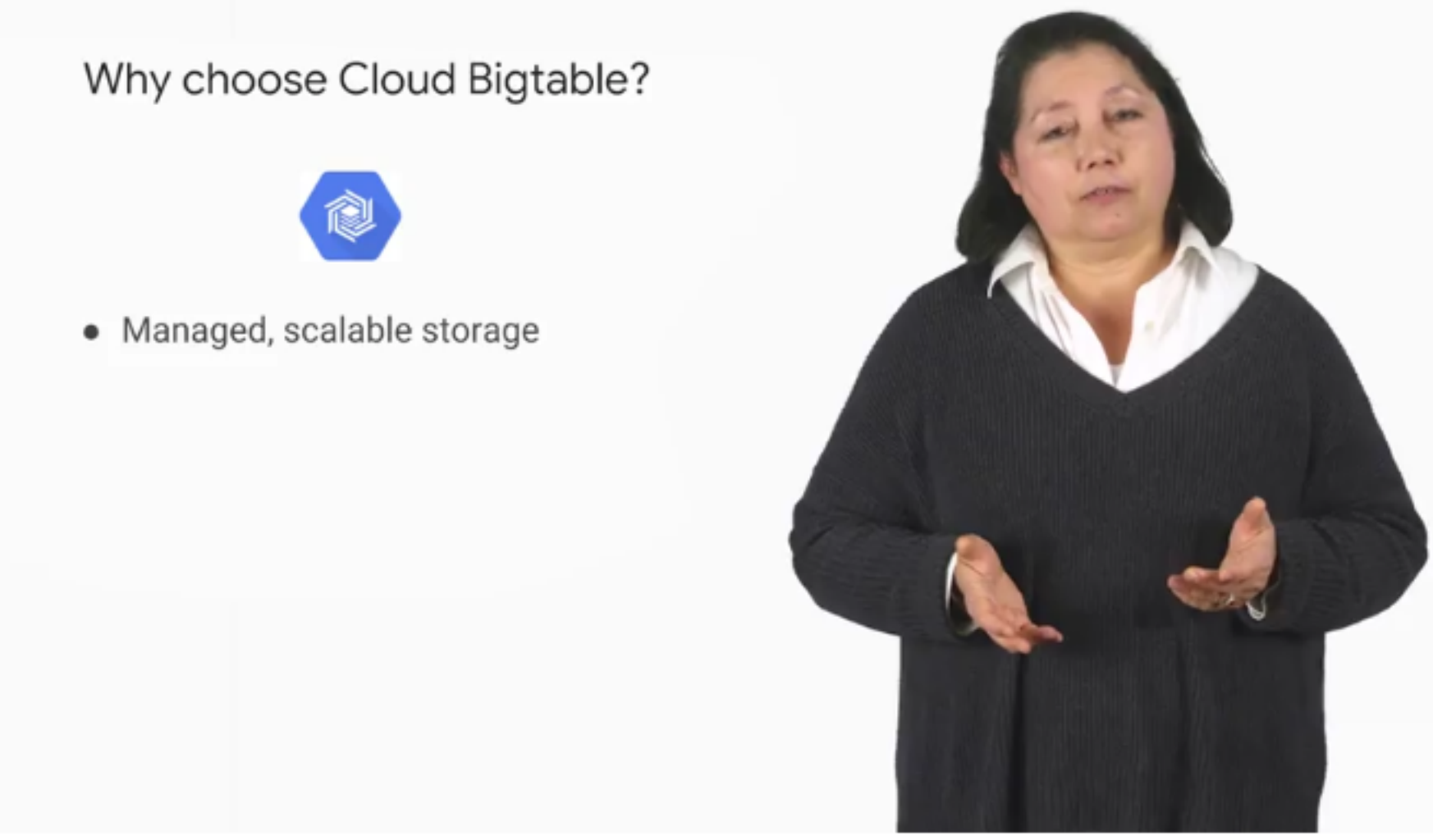
Cloud Bigtable is managed NoSQL

- Accessed using HBase API
- Native compatibility with big data, Hadoop ecosystems

Cloud Bigtable is offered through the same open source API as HBase, which is the native database for the Apache Hadoop project. I'll talk more about Hadoop later in the course. Anyway, having the same API enables portability of applications between HBase and Bigtable.

Given that you could manage your own Apache HBase installation, you might ask yourself, why should I choose Bigtable? Here are a few reasons why you might. First, scalability.

If you manage your own Hbase installation, scaling past a certain rate of queries per second is going to be



If you manage your own Hbase installation, scaling past a certain rate of queries per second is going to be tough, but with Bigtable you can just increase your machine count which doesn't even require downtime.

Also, Cloud Bigtable handles administration tasks like upgrades and restarts transparently. All data in Cloud Bigtable is encrypted in both in-flight and at rest. You can even use IAM permissions to control who has access to Bigtable data. One last reference point.

Bigtable is actually the same database that powers many of Google's core services including

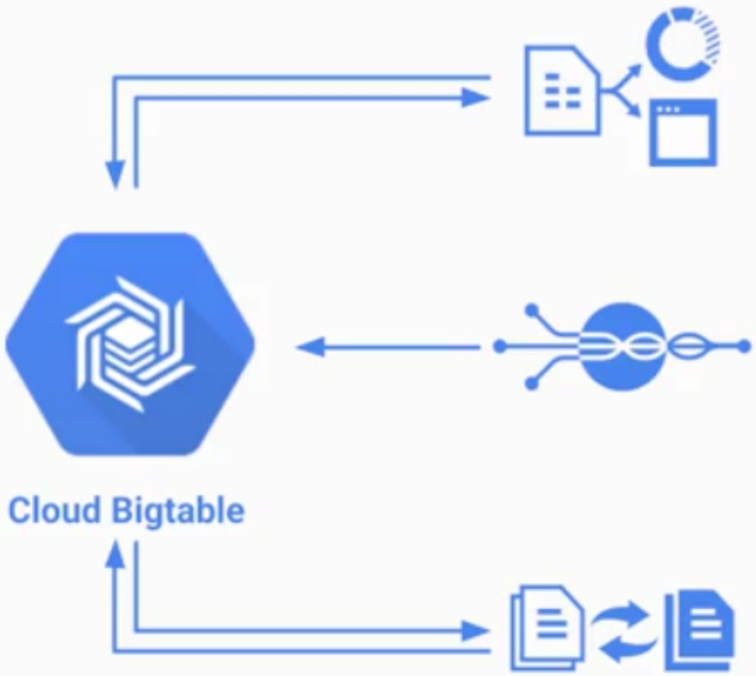
Google Cloud Bigtable

Bigtable is actually the same database that powers many of Google's core services including search, analytics, maps and Gmail. As Cloud Bigtable is part of the GCP ecosystem, it can interact with other GCP services and third-party clients.

From an application API perspective, data can be read from and written to Cloud Bigtable through a data service layer like Managed VMs, the HBase rest server or a Java server using the HBase client. Typically, this will be to serve data to applications, dashboards and data services.

Data can also be streamed in

Bigtable Access Patterns



Application API

Data can be read from and written to Cloud Bigtable through a data service layer like Managed VMs, the HBase REST Server, or a Java Server using the HBase client. Typically this will be to serve data to applications, dashboards, and data services.

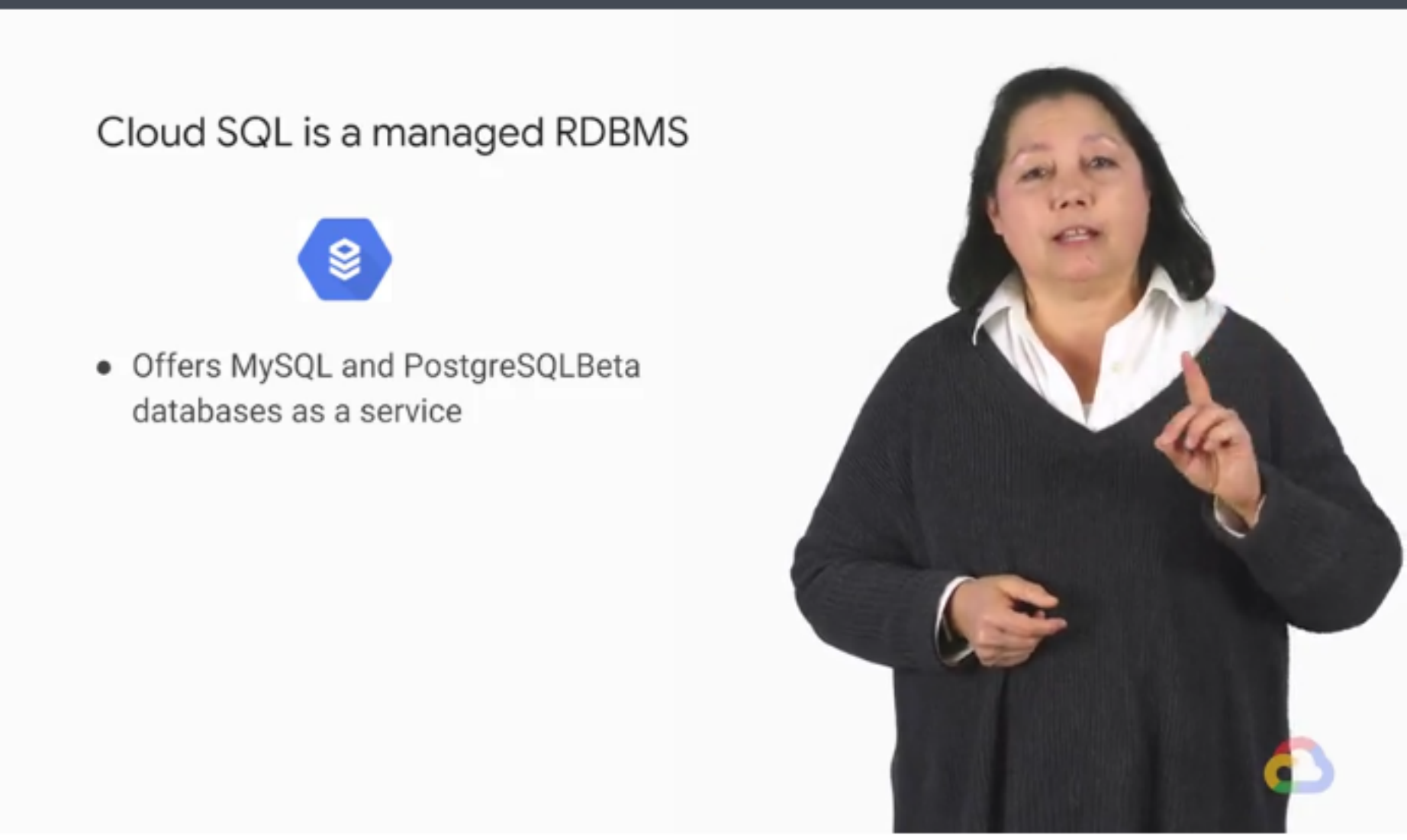
Streaming

Data can be streamed in (written event by event) through a variety of popular stream processing frameworks like Cloud Dataflow Streaming, Spark Streaming, and Storm.

Batch Processing

Data can be read from and written to Cloud Bigtable through batch processes like Hadoop MapReduce, Dataflow, or Spark. Often, summarized or newly calculated data is written back to Cloud Bigtable or to a downstream database.

From an application API perspective,



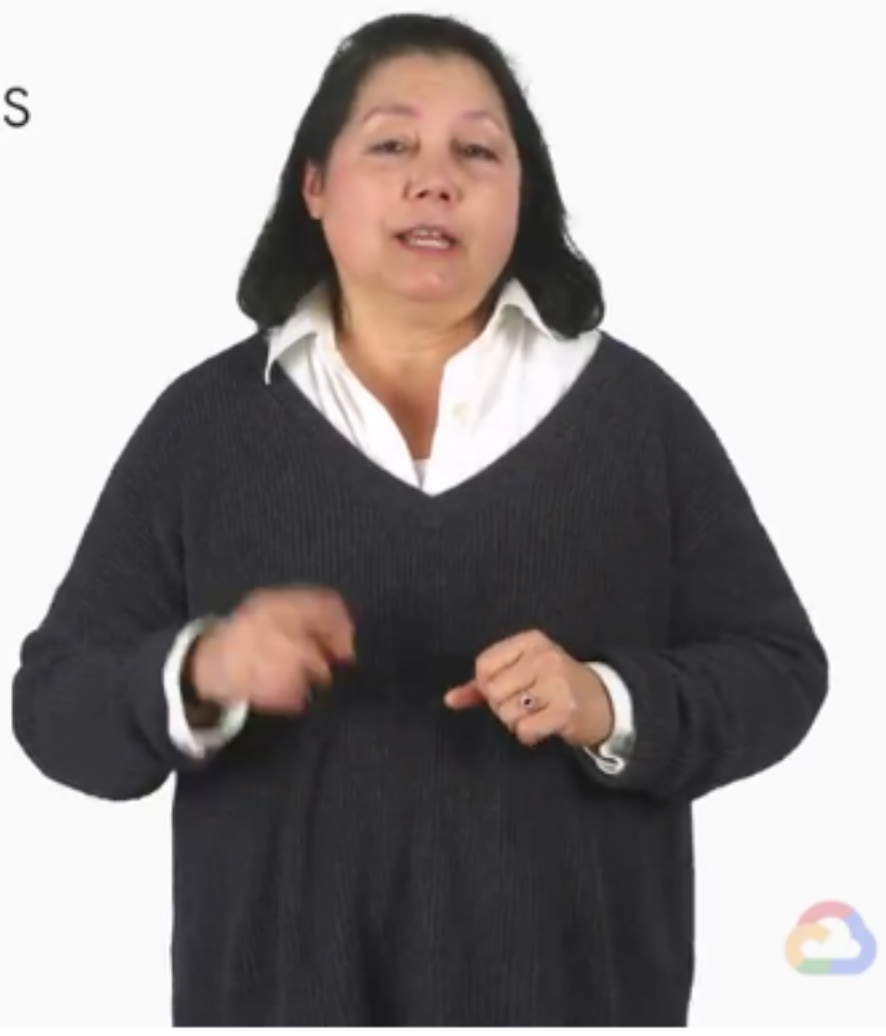
A moment ago, I discussed NoSQL databases. Now, let's turn our attention to relational database services. Remember, these services use a database schema to help your application keep your data consistent and correct.

Another feature of relational database services that helps with the same goal - transactions. Your application can designate a group of database changes as all or nothing. Either they all get made, or none do.

Without database transactions, your online bank wouldn't be able to offer you the ability to move money from one account to another. What if, after

Cloud SQL is a managed RDBMS

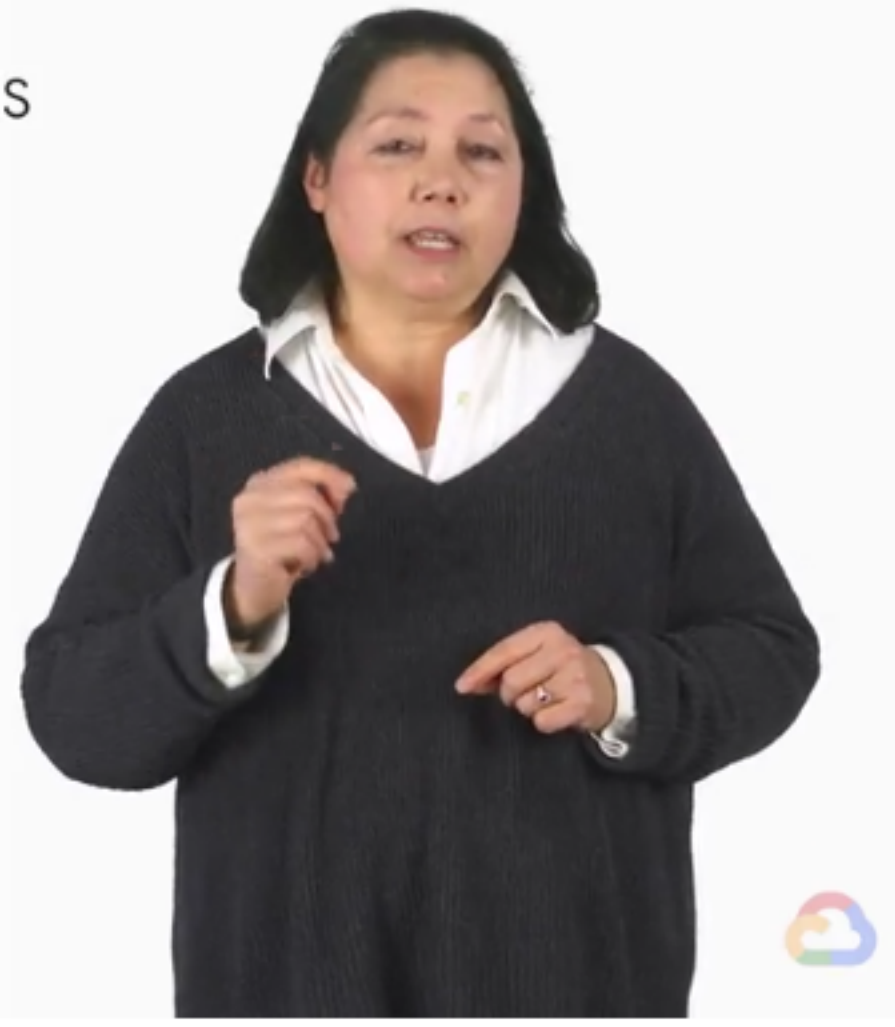
- Offers MySQL and PostgreSQLBeta databases as a service



It offers you your choice of the MySQL or PostgreSQL database engines as a fully managed service. Cloud SQL offers both MySQL and PostgreSQL databases that are capable of handling terabytes of storage. As of this recording, Cloud SQL for PostgreSQL is in beta.

So, check the website for details of its status. Of course, you could always run your own database server inside a Compute Engine virtual machine, which a lot of GCP customers do. But there are some benefits of using the Cloud SQL managed service instead.

First, Cloud SQL provides several



there are some benefits of using the Cloud SQL managed service instead.

First, Cloud SQL provides several replica services like read, failover, and external replicas. This means that if an outage occurs, Cloud SQL can replicate data between multiple zones with automatic failover. Cloud SQL also helps you backup your data with either on-demand or scheduled backups.

It can also scale both vertically by changing the machine type, and horizontally via read replicas.

From a security perspective, Cloud SQL instances include network firewalls, and customer data is encrypted

Cloud SQL is a managed RDBMS

Cloud SQL can be used with App Engine using standard drivers.

You can configure a Cloud SQL instance to follow an App Engine application.

Compute Engine instances can be authorized to access Cloud SQL instances using an external IP address.

Cloud SQL instances can be configured with a preferred zone.

Cloud SQL can be used with external applications and clients.

Standard tools can be used to administer databases.

External read replicas can be configured.

Cloud SQL instances and configure

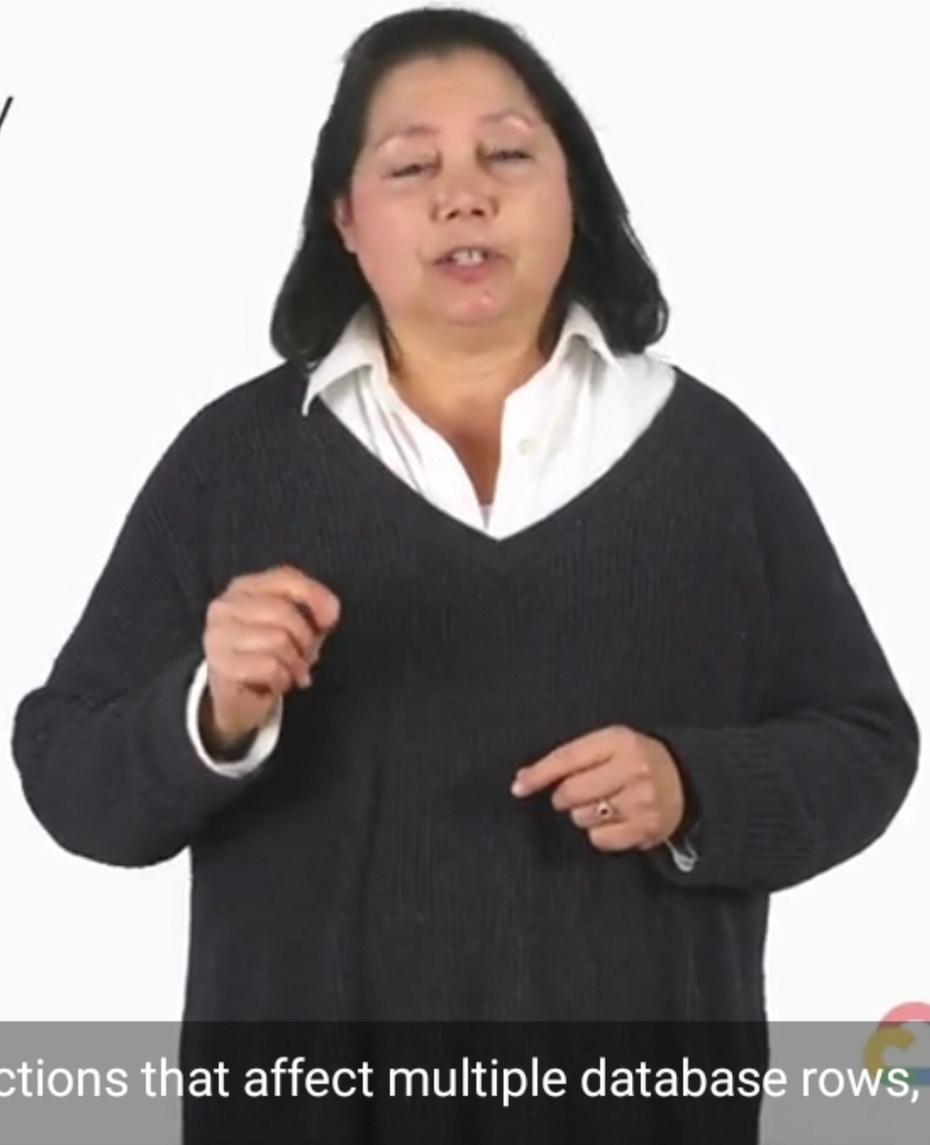
standard MySQL drivers. If Cloud SQL does not fit your requirements be-cause you need horizontal scaleability, consider using Cloud Spanner.

It offers transactional consistency at a global scale, schemas, SQL, and automatic synchronous replication for high availability. And, it can provide pedabytes of capacity.

Consider using Cloud Spanner if you have outgrown any relational database, or sharding your databases for throughput high performance, need transactional consistency, global data and strong consistency, or just want to consolidate your database.

Cloud Datastore is a horizontally scalable NoSQL DB

- Designed for application backends
- Supports transactions
- Includes a free daily quota



Unlike Cloud Bigtable, it also offers transactions that affect multiple database rows,

As you would expect from a fully-managed service, Cloud Datastore automatically handles sharding and replication, providing you with a highly available and durable database that scales automatically to handle load.

Unlike Cloud Bigtable, it also offers transactions that affect multiple database rows, and it lets you do SQL-like queries. To get you started, Cloud Datastore has a free daily quota that provides storage, reads, writes, deletes and small operations at no charge.

Comparing storage options: technical details

	Cloud Datastore	Bigtable	Cloud Storage	Cloud SQL	Cloud Spanner	BigQuery
Type	NoSQL document	NoSQL wide column	Blobstore	Relational SQL for OLTP	Relational SQL for OLTP	Relational SQL for OLAP
Transactions	Yes	Single-row	No	Yes	Yes	No
Complex queries	No	No	No	Yes	Yes	Yes
Capacity	Terabytes+	Petabytes+	Petabytes+	Terabytes	Petabytes	Petabytes+
Unit size	1 MB/entity	~10 MB/cell ~100 MB/row	5 TB/object	Determined by DB engine	10,240 MiB/ row	10 MB/row

let's compare them to help you choose the right service for your application or workflow.

Note - Cloud Datastore actually stores *structured* objects.

CONTINUE

Comparing storage options: technical details

	Cloud Datastore	Cloud Bigtable	Cloud Storage	Cloud SQL	Cloud Spanner	BigQuery
Type	NoSQL document	NoSQL wide column	Blobstore	Relational SQL for OLTP	Relational SQL for OLTP	Relational SQL for OLAP
Best for	Semi-structured application data, durable key-value data	"Flat" data, Heavy read/write, events, analytical data	Structured and unstructured binary or object data	Web frameworks, existing applications	Large-scale database applications (> ~2 TB)	Interactive querying, offline analytics
Use cases	Getting started, App Engine applications	AdTech, Financial and IoT data	Images, large media files, backups	User credentials, customer orders	Whenever high I/O, global consistency is needed	Data warehousing

Considering the technical differentiators of the different storage services,