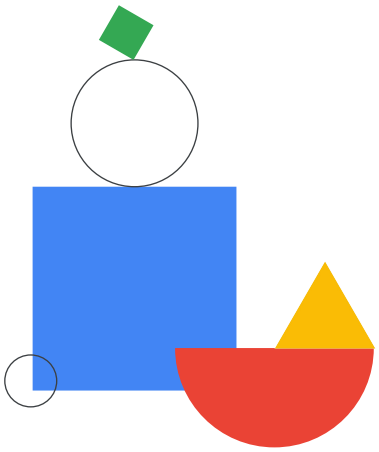
Google Cloud

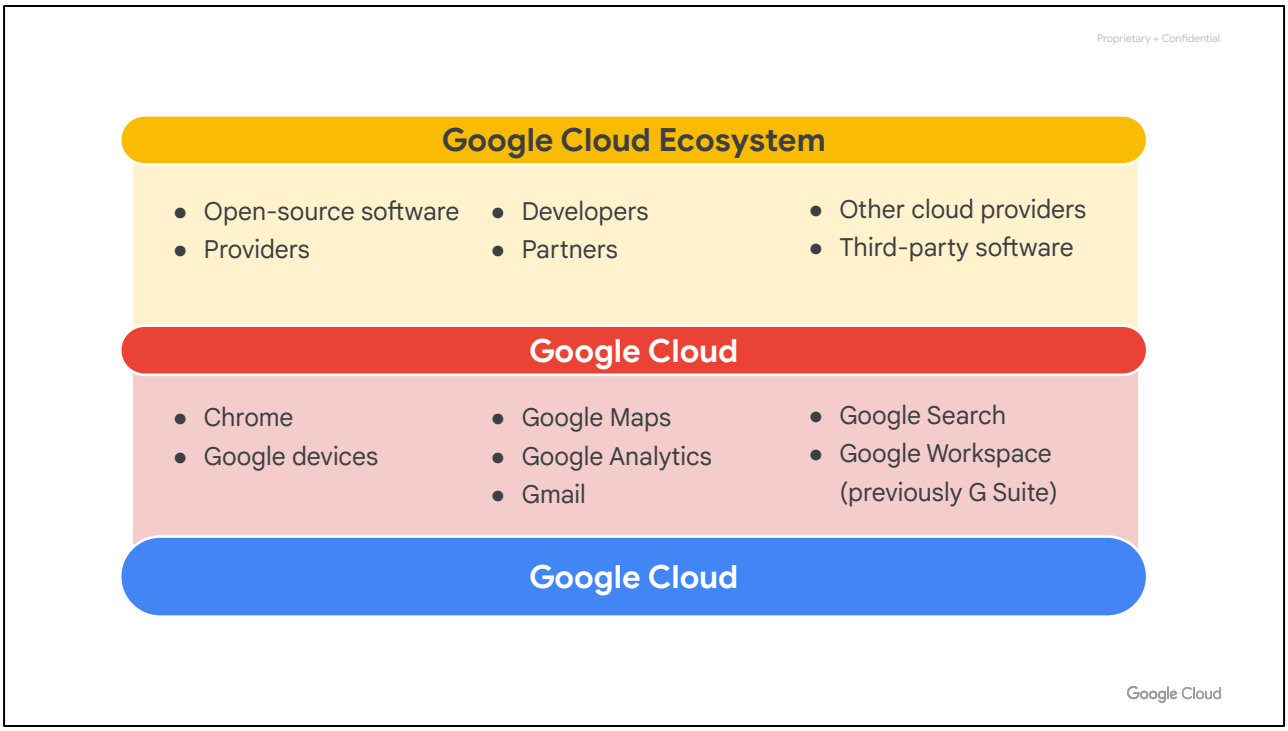
Architecting with Google Compute Engine

Philipp Maier
Mylene Biddle



Hello, I'm Philipp Maier, and I'm Mylene Biddle. We're both course developers at Google and we want to welcome you to Architecting with Compute Engine, a series of 3 courses.

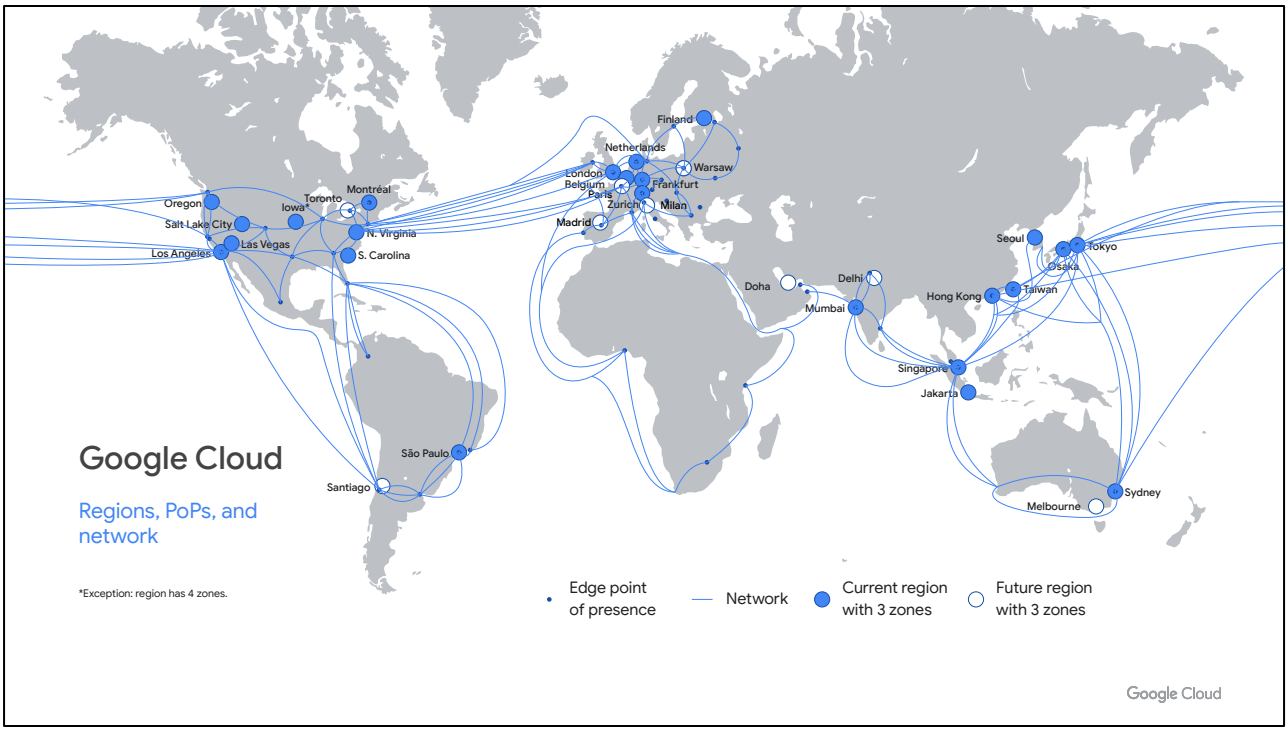
Before we start using all of the different services that Google Cloud offers, let's talk about what Google Cloud is.



When you look at Google Cloud, you'll see that it's actually part of a much larger ecosystem. This ecosystem consists of open-source software, providers, partners, developers, third-party software, and other cloud providers.

Google is actually a very strong supporter of open-source software.

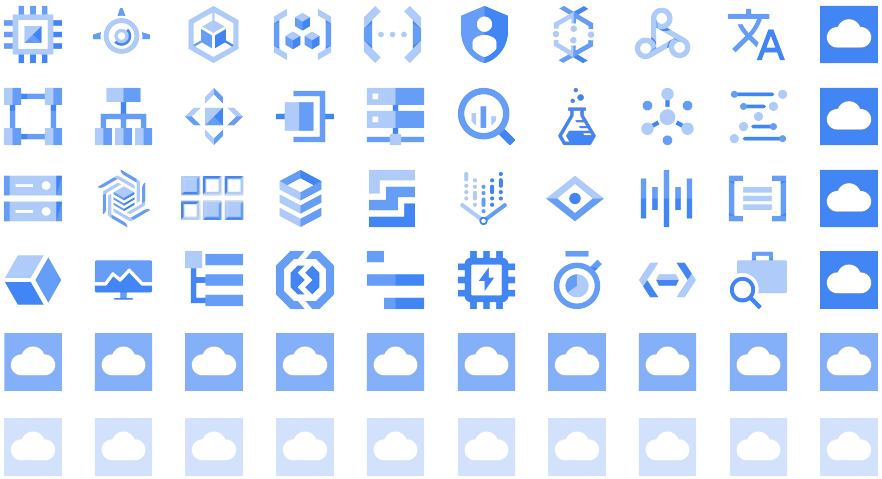
That's right, now Google Cloud consists of Chrome, Google Devices, Google Maps, Gmail, Google Analytics, Google Workspace, Google Search and the Google Cloud platform. Google Cloud itself is a computing solution platform that really encompasses three core features: infrastructure, platform, and software.



This map represents Google Cloud's global infrastructure. As of this recording, Google Cloud's well-provisioned global network connects over 60 zones to over 130 points of presence through a global network of fiber optic cables.

And Google is continuously investing in this network, with new regions, points of presence, and subsea cable investments.

Google Cloud is...



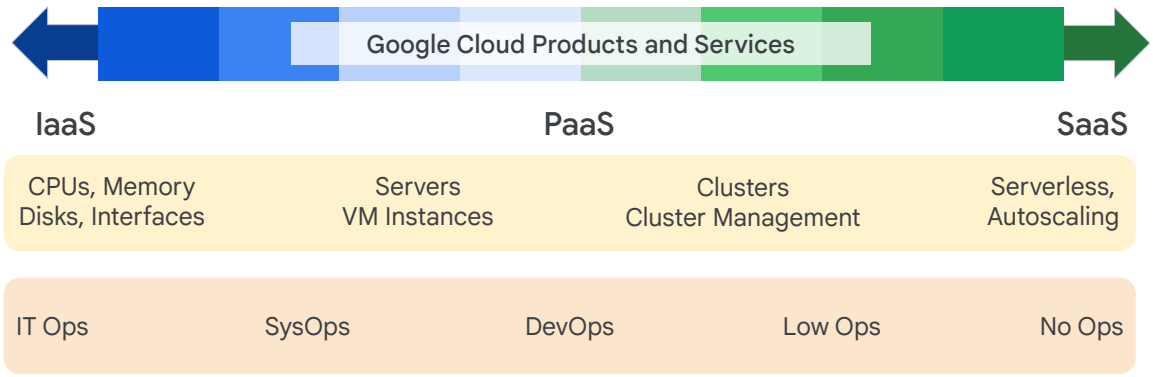
On top of this infrastructure, Google Cloud uses state-of-the-art software-defined networking and distributed systems technologies to host and deliver your services around the world.

These technologies are represented by a suite of cloud-based products and services that is continuously expanding.

Google Cloud is...

Now, it's important to understand that there is usually more than one solution for a task or application in Google Cloud. To better understand this, let's look at a solution continuum.

Solution continuum

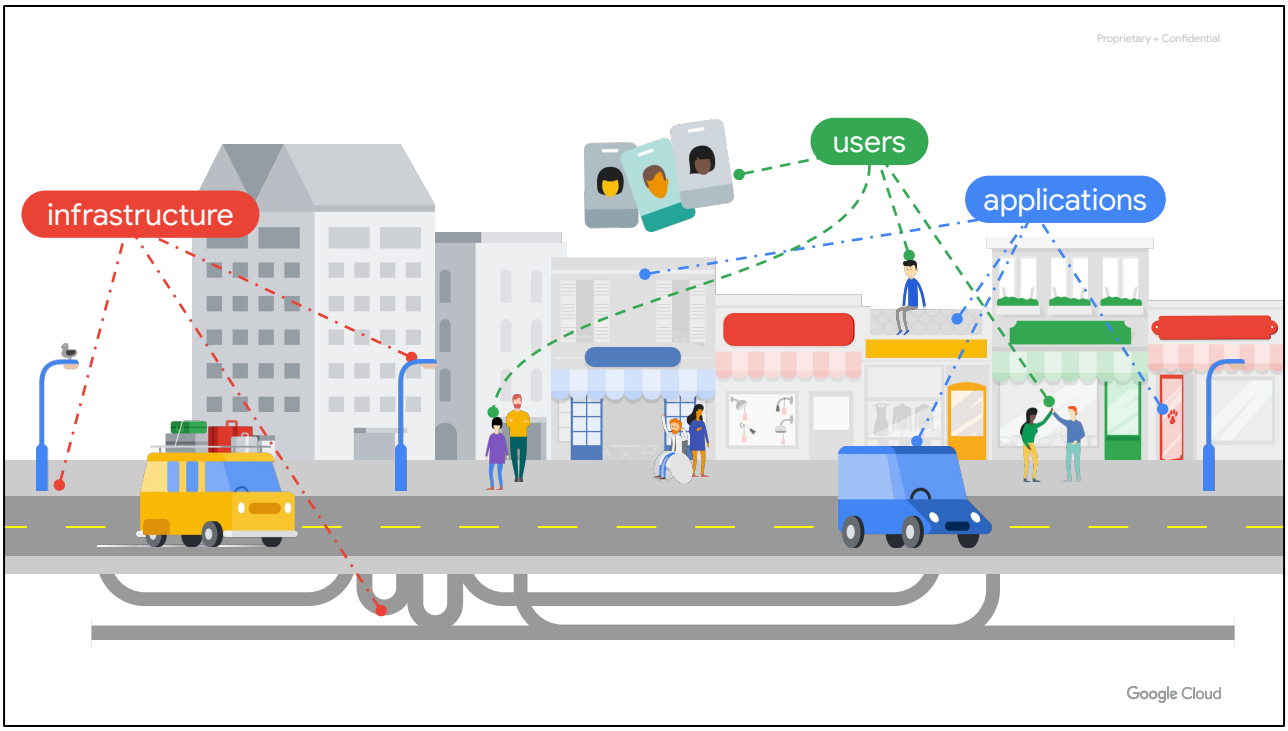


Google Cloud spans from Infrastructure as a Service (or IaaS) to Software as a Service (or SaaS). You really can build applications on Google Cloud for the web or mobile that are global, autoscaling, and assistive, and that provide services where the infrastructure is completely invisible to the user. It is not just that Google has opened the infrastructure that powers applications like Search, Gmail, Google Maps, and Google Workspace. Google has opened all of the services that make these products possible and packaged them for your use.

Alternative solutions are possible. For example, you could start up your own VM in Google Compute Engine, install open source MySQL on it, and run it just like a MySQL database on your own computer in a data center.

Or you could use the Cloud SQL service, which provides a MySQL instance and handles operational work like backups and security patching for you, using the same services Google does to automate backups and patches.

You could even move to a noSQL database that is autoscaling and serverless so that growth no longer requires adding server instances or possibly changing the design to handle the new capacity.



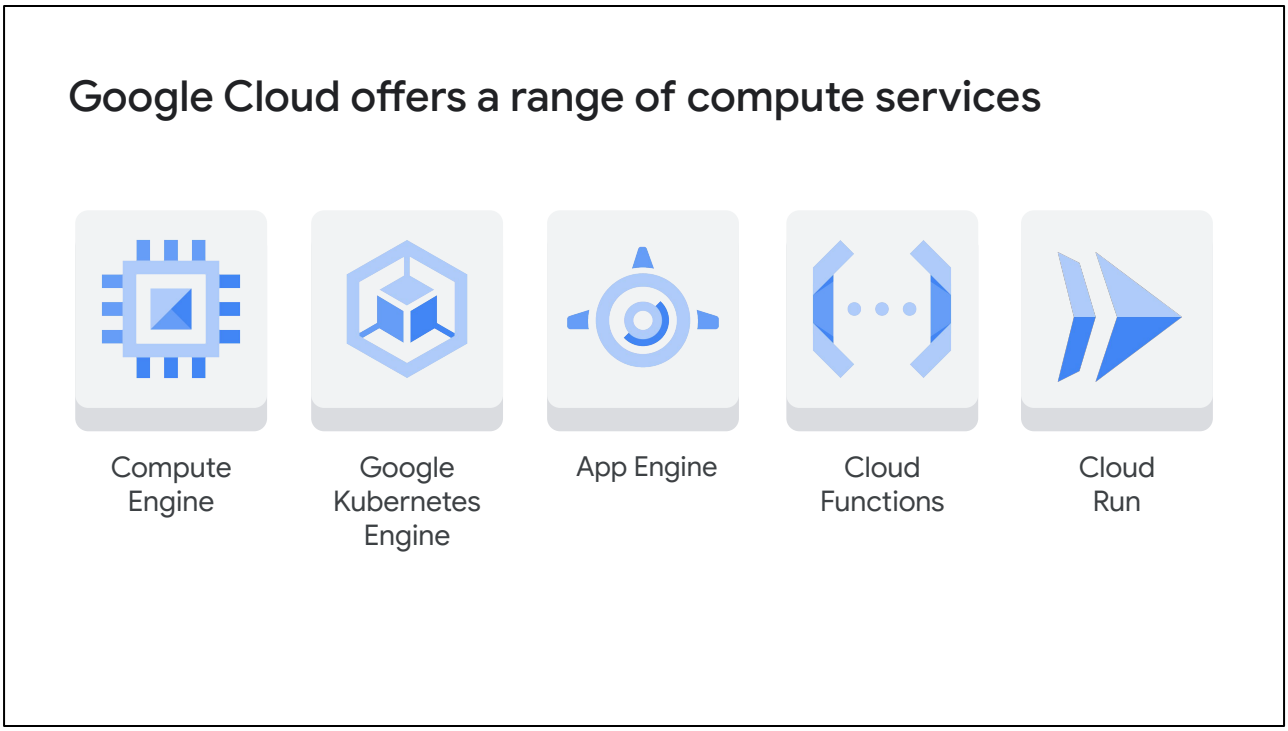
This series of course focuses on the infrastructure. An IT infrastructure is like a "city infrastructure."

The infrastructure is the basic underlying framework of fundamental facilities and systems such as transport, communications, power, water, fuel, and other essential services.

The people in the city are like "users," and the cars and bikes and buildings in the city are like "applications." Everything that goes into creating and supporting those applications for the users is the infrastructure.

The purpose of this course is to explore, as efficiently and clearly as possible, the infrastructure services provided by Google Cloud. You should become familiar enough with the infrastructure services that you will know what the services do and how to use them.

We won’t go into very deep-dive case studies on specific vertical applications, but you'll know enough to put all the building blocks together to build your own solution.



Now, Google Cloud offers a range of compute services. The service that might be most familiar to newcomers is Compute Engine, which lets you run virtual machines on demand in the cloud. It's Google Cloud's Infrastructure-as-a-Service solution. It provides maximum flexibility for people who prefer to manage server instances themselves.

Google Kubernetes Engine lets you run containerized applications on a cloud environment that Google manages for you, under your administrative control. Think of containerization as a way to package code that's designed to be highly portable and to use resources very efficiently, and think of Kubernetes as a way to orchestrate code in containers.

App Engine is Google Cloud's fully managed Platform-as-a-Service framework. That means it's a way to run code in the cloud without having to worry about infrastructure. You just focus on your code, and let Google deal with all the provisioning and resource management. You can learn a lot more about App Engine in the "Developing Applications with Google Cloud" course series.

Cloud Functions is a completely serverless execution environment, or Functions-as-a-Service. It executes your code in response to events, whether those events occur once a day or many times per second. Google scales resources as required, but you only pay for the service while your code runs. The "Developing Applications with Google Cloud" course series also discusses Cloud Functions.

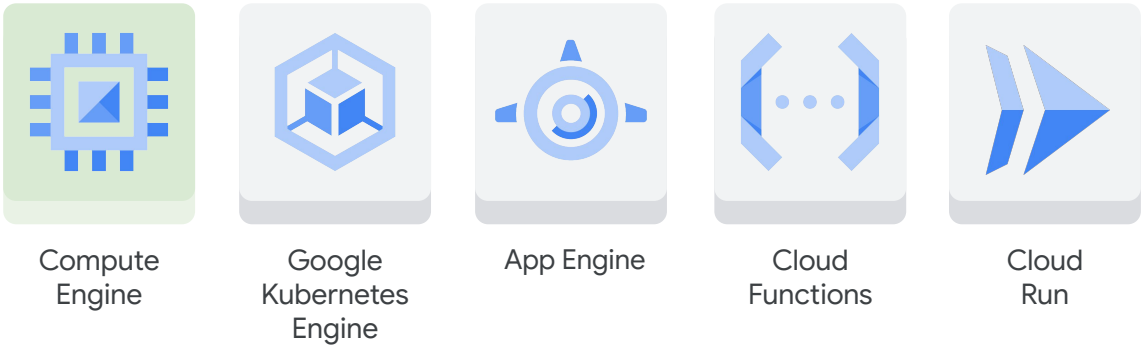
Cloud Run, a managed compute platform that lets you run stateless containers via web requests or Pub/Sub events.

Cloud Run is serverless. That means it removes all infrastructure management tasks so you can focus on developing applications.

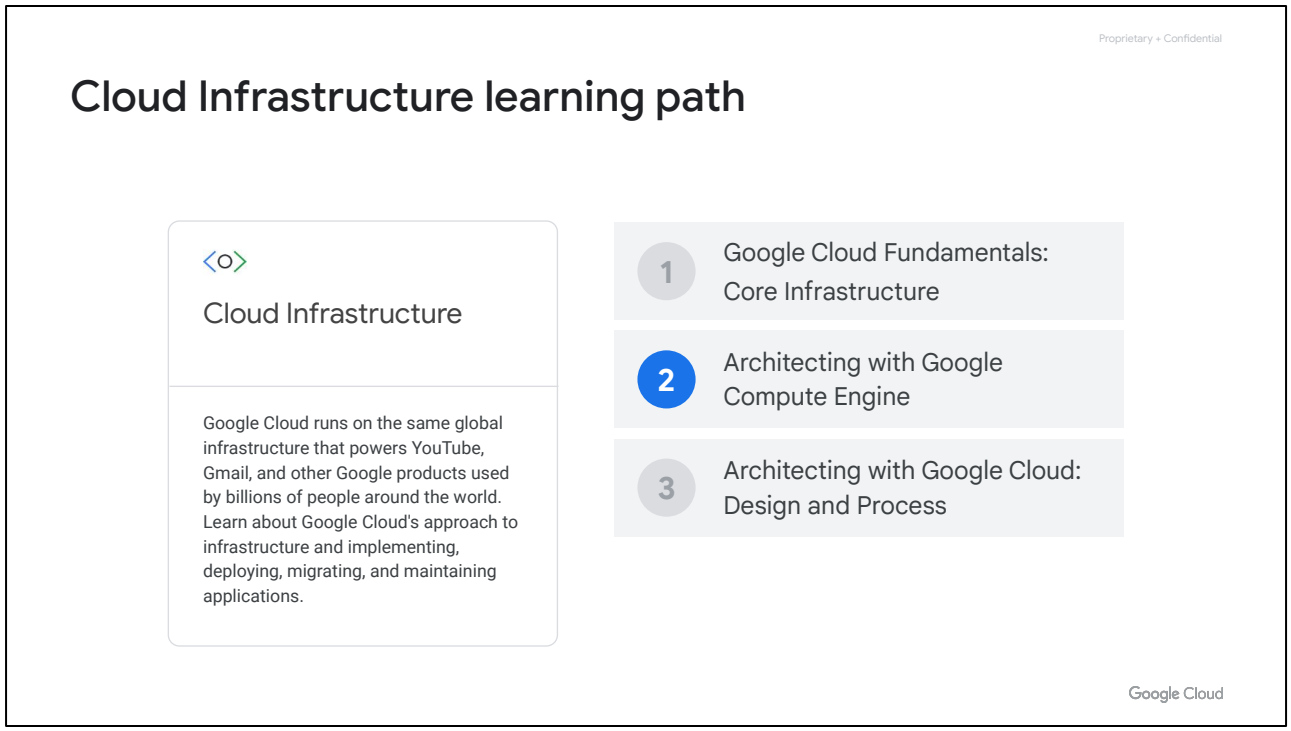
It is built on Knative, an open API and runtime environment built on Kubernetes that gives you freedom to move your workloads across different environments and platforms. It can be fully managed on Google Cloud, on Google Kubernetes Engine, or anywhere Knative runs.

Cloud Run is fast. It can automatically scale up and down from zero almost instantaneously, and it charges you only for the resources you use, calculated down to the nearest 100 milliseconds, so you‘ll never pay for your over-provisioned resources.

Google Cloud offers a range of compute services



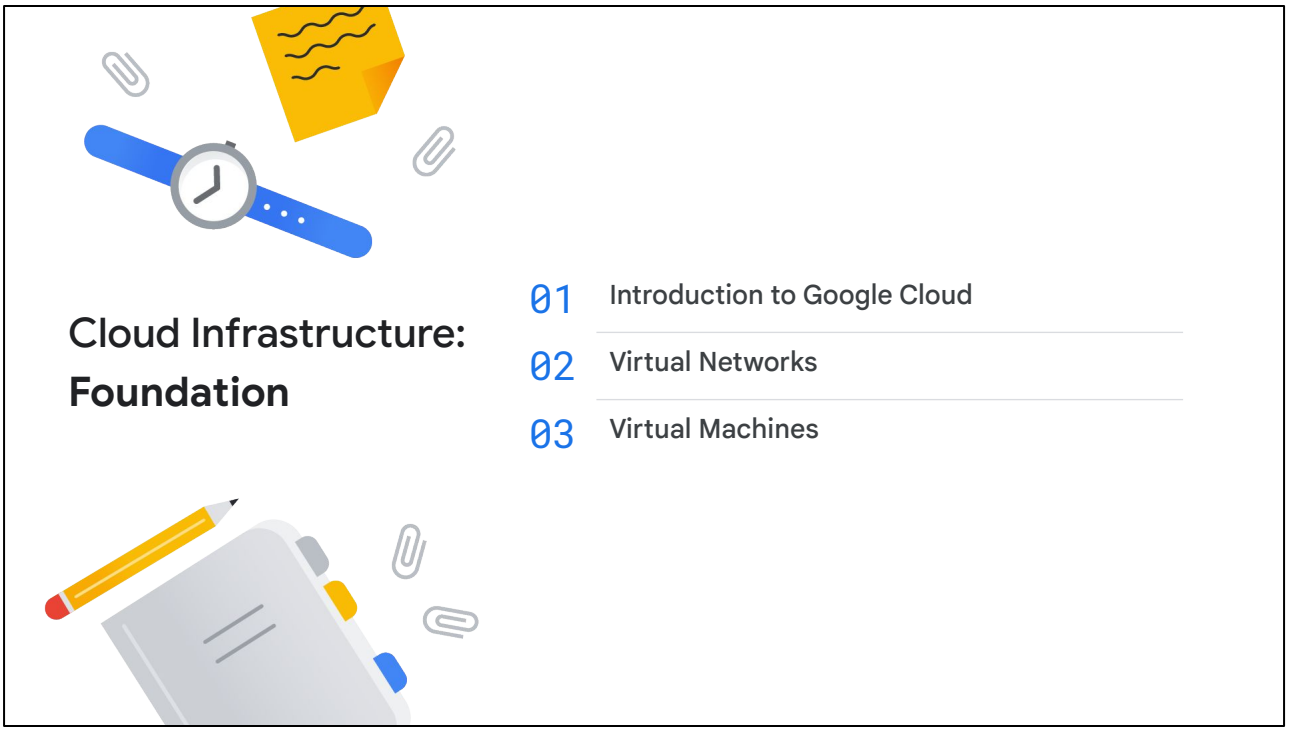
In this series of courses, Compute Engine will be our main focus.



The “Architecting with Google Compute Engine” courses are part of the Cloud Infrastructure learning path. This path is designed for IT professionals who are responsible for implementing, deploying, migrating, and maintaining applications in the cloud.

The prerequisite for these courses is the “Google Cloud Fundamentals: Core Infrastructure” course, which you can find in the links section for this video [https://google.qwiklabs.com/courses/870].

The Architecting with Compute Engine series consists of 3 courses.

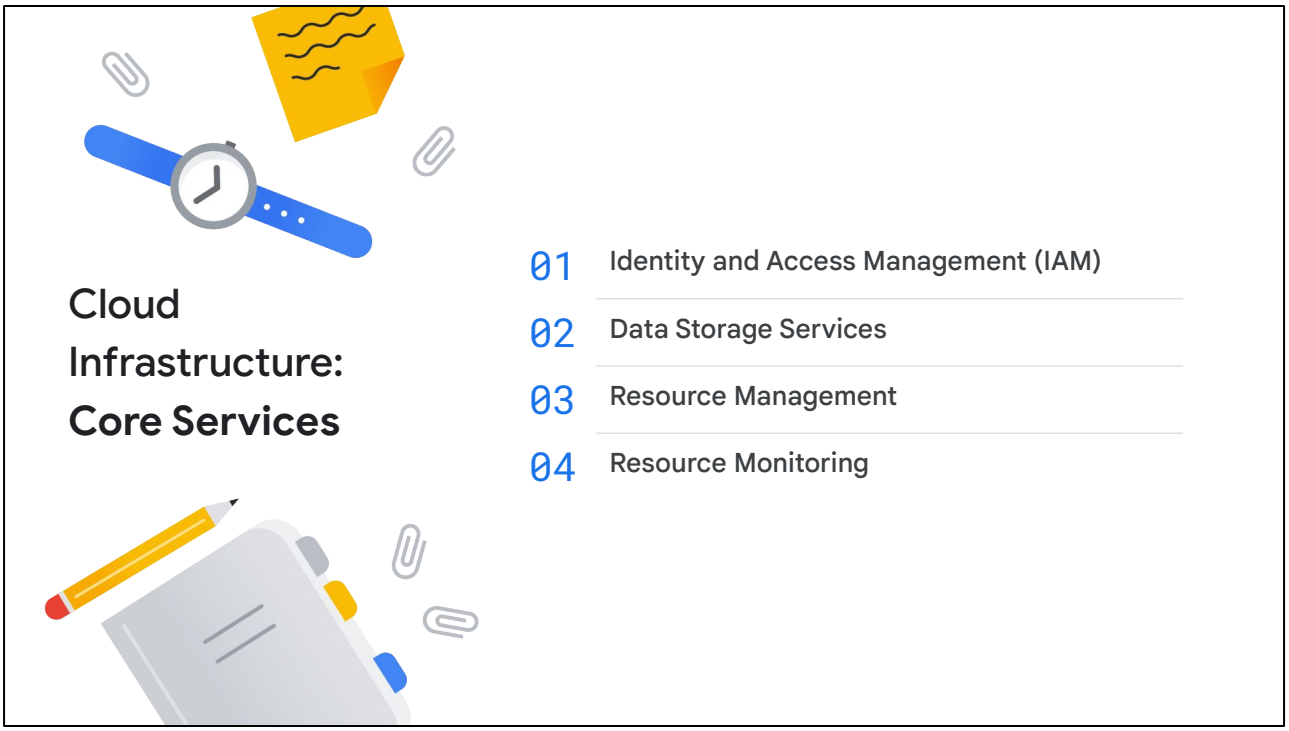


“Essential Cloud Infrastructure: Foundation” is the first course of the Architecting with Compute Engine series.

In that course, we start by introducing you to Google Cloud and how to interact with the Cloud Console and Cloud Shell.

Next we’ll get into virtual networks, and you will create VPC networks and other networking objects.

Then we’ll take a deep dive into virtual machines, and you will create virtual machines using Compute Engine.



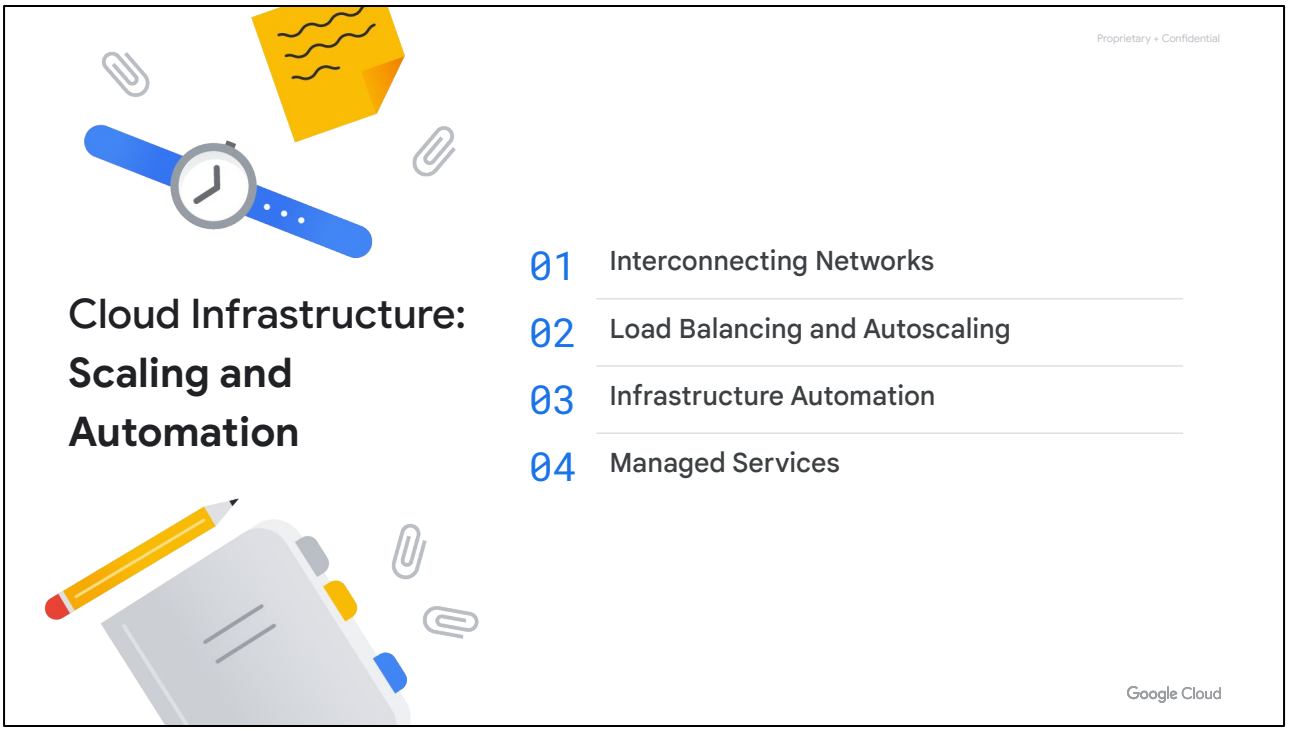
“Essential Cloud Infrastructure: Core Services” is the second course of this series.

In that course, we start by talking about IAM, and you will administer Identity and Access Management for resources.

Next, we’ll cover the different data storage services in Google Cloud, and you will implement some of those services.

Then we’ll go over resource management, where you will manage and examine billing of Google Cloud resources.

Lastly we’ll talk about resource monitoring, and you will monitor Google Cloud resources using Google Cloud’s operations suite.



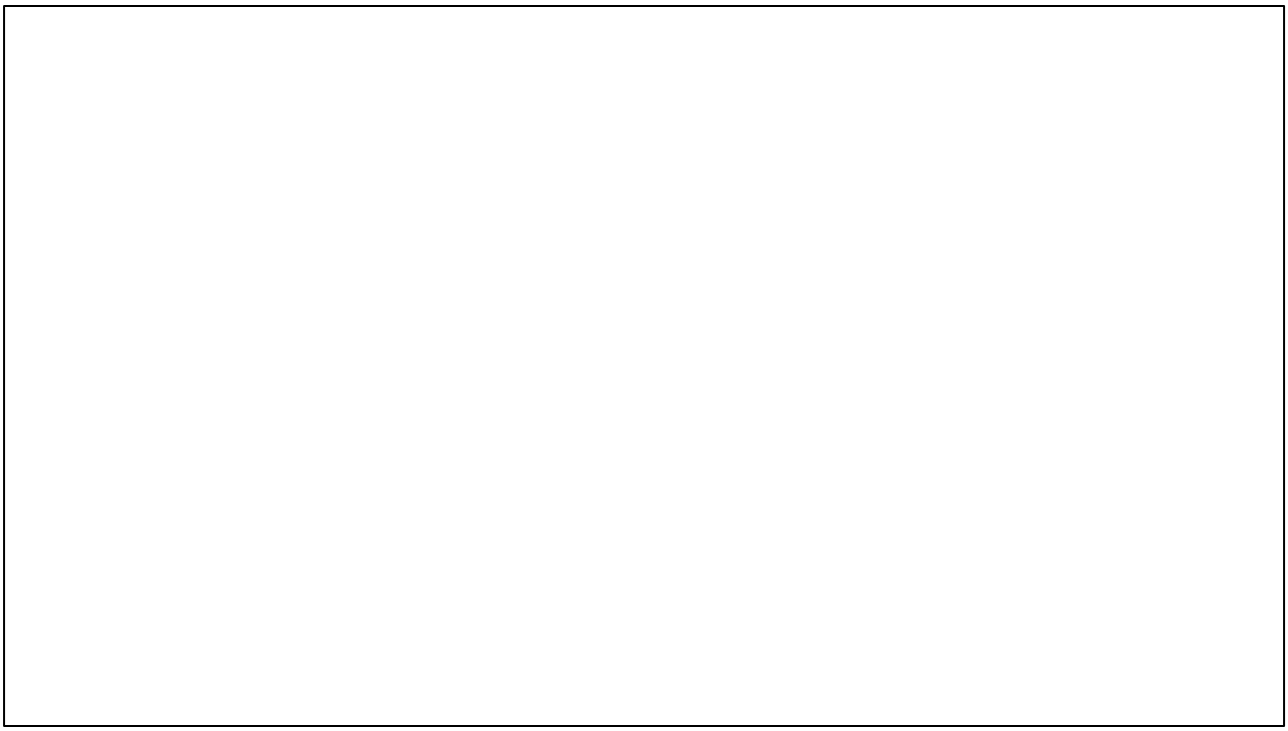
“Elastic Cloud Infrastructure: Scaling and Automation” is the last course of this series.

In that course, we start by going over the different options to interconnect networks to enable you to connect your infrastructure to Google Cloud.

Next, we’ll go over Google Cloud’s load balancing and autoscaling services, which you will get to explore directly.

Then we’ll cover infrastructure automation services like Terraform, so that you can automate the deployment of Google Cloud infrastructure services.

Lastly, we’ll talk about other managed services that you might want to leverage in Google Cloud.



Now, our goal for you is to remember and understand the different Google Cloud services and features and also be able to apply your knowledge, analyze requirements, evaluate different options, and create your own services.

And that’s why these courses include interactive hands-on labs through the Qwiklabs platform. Qwiklabs provisions you with a Google account and credentials, so you can access the Cloud Console for each lab at no cost.