Week 1, Cloud Application Development

Azamat Serek, PhD, Assist.Prof

What is Cloud Computing?

Understanding the types of cloud computing resources can be time-consuming and costly. Enterprises need to buy physical servers and other infrastructure through procurement processes that can take months, and support the architecture of cloud computing. The acquired systems require a physical space, typically a specialized room with sufficient power and cooling. After configuring and deploying the systems, enterprises need expert personnel to manage them.

This long process is difficult to scale when demand spikes or business expands. Enterprises can acquire more computing resources than needed, ending up with low utilization numbers.

Cloud computing addresses these issues by offering computing resources as scalable, on-demand services. Learn more about Google Cloud, a suite of cloud computing service models offered by Google.

Cloud computing defined

Cloud computing is the on-demand availability of computing resources (such as storage and infrastructure), as services over the internet. It eliminates the need for individuals and businesses to self-manage physical resources themselves, and only pay for what they use.

The main cloud computing service models include infrastructure as a service offers compute and storage services, platform as a service offers a develop-and-deploy environment to build cloud apps, and software as a service delivers apps as services.

Understanding how cloud computing works

Cloud computing service models are based on the concept of sharing on-demand computing resources, software, and information over the internet. Companies or individuals pay to access a virtual pool of shared resources, including compute, storage, and networking services, which are located on remote servers that are owned and managed by service providers.

One of the many advantages of cloud computing is that you only pay for what you use. This allows organizations to scale faster and more efficiently without the burden of having to buy and maintain their own physical data centers and servers.

In simpler terms, cloud computing uses a network (most often, the internet) to connect users to a cloud platform where they request and access rented computing services. A central server handles all the communication between client devices and servers to facilitate the exchange of data. Security and privacy features are common components to keep this information secure and safe.

When adopting cloud computing architecture, there is no one-size-fits-all. What works for another company may not suit you and your business needs. In fact, this flexibility and versatility is one of the hallmarks of cloud, allowing enterprises to quickly adapt to changing markets or metrics.

There are three different cloud computing deployment models: public cloud, private cloud, and hybrid cloud.

Types of cloud computing deployment models

Public cloud

Public clouds are run by third-party cloud service providers. They offer compute, storage, and network resources over the internet, enabling companies to access shared on-demand resources based on their unique requirements and business goals.

Private cloud

Private clouds are built, managed, and owned by a single organization and privately hosted in their own data centers, commonly known as "on-premises" or "on-prem." They provide greater control, security, and management of data while still enabling internal users to benefit from a shared pool of compute, storage, and network resources.

Hybrid cloud

Hybrid clouds combine public and private cloud models, allowing companies to leverage public cloud services and maintain the security and compliance capabilities commonly found in private cloud architectures.

Why cloud

Everything you need to build and scale

Virtual machines
Create and run virtual machines with Compute Engine.

Object storage
Store any amount of data across classes with Cloud Storage.

Data warehouse
Run analytics and democratize insights with BigQuery.

Relational databases
Run fully managed PostgreSQL-compatible database services with AlloyDB

Managed Kubernetes
Manage containerized applications with Google Kubernetes Engine.

Command line tools
Get libraries and tools for Google Cloud services with Cloud SDK.

Content delivery network
Deliver video and web content at global scale with Cloud CDN.

Streaming analytics
Unify stream and batch data processing with Dataflow.

What are the types of cloud computing services?

There are three main types of cloud computing service models that you can select based on the level of control, flexibility, and management your business needs:

Infrastructure as a service (IaaS)

Infrastructure as a service (IaaS) offers on-demand access to IT infrastructure services, including compute, storage, networking, and virtualization. It provides the highest level of control over your IT resources and most closely resembles traditional on-premises IT resources.

Platform as a service (PaaS)

Platform as a service (PaaS) offers all the hardware and software resources needed for cloud application development. With PaaS, companies can focus fully on application development without the burden of managing and maintaining the underlying infrastructure.

Software as a service (SaaS)

Software as a service (SaaS) delivers a full application stack as a service, from underlying infrastructure to maintenance and updates to the app software itself. A SaaS solution is often an end-user application, where both the service and the infrastructure is managed and maintained by the cloud service provider.

What are the benefits of cloud computing?

It's flexible

Due to the architecture of cloud computing, enterprises and their users can access cloud services from anywhere with an internet connection, scaling services up or down as needed.

It's efficient

Enterprises can develop new applications and rapidly get them into production—without worrying about the underlying infrastructure.

It offers strategic value

Because cloud providers stay on top of the latest innovations and offer them as services to customers, enterprises can get more competitive advantages—and a higher return on investment—than if they'd invested in soon-to-be obsolete technologies.

How cloud computing can help your organization

The pace of innovation—and the need for advanced computing to accelerate this growth—makes cloud computing a viable option to advance research and speed up new product development. Cloud computing can give enterprises access to scalable resources and the latest technologies without needing to worry about capital expenditures or limited fixed infrastructure. What is the future of cloud computing? It's expected to become the dominant enterprise IT environment.

If your organization experiences any of the following, you're probably a good candidate for cloud computing:

- High business growth that outpaces infrastructure capabilities
- Low utilization of existing infrastructure resources
- Large volumes of data that are overwhelming your on-premises data storage resources
- Slow response times with on-premises infrastructure
- Delayed product development cycles due to infrastructure constraints
- Cash flow challenges due to high computing infrastructure expenses
- Highly mobile or distributed user population

These scenarios require more than traditional data centers can provide.

Use cases

Cloud computing offers a broad range of possible applications that can benefit organizations. Here are some common use cases:

Infrastructure scaling

Many organizations, including those in retail, have wildly varying needs for compute capacity. Cloud computing easily accommodates these fluctuations.

Disaster recovery

Rather than building more data centers to ensure continuity during disasters, businesses use cloud computing to safely back up their digital assets.

Data storage

Cloud computing helps overloaded data centers by storing large volumes of data, making it more accessible, easing analysis, and making backup easier.

Application development

Cloud computing offers enterprise developers quick access to tools and platforms for building and testing applications, speeding up time to market.

Big data analytics

Cloud computing offers almost unlimited resources to process large volumes of data to speed research and reduce time to insights.

GCP

https://cloud.google.com/gcp/

Advantages

Put your data to work

Bring the simplicity, scale, security, and intelligence of Google's information approach to your organization. Google offers a complete data foundation to unify all workloads and manage the entire data life cycle. The solution is designed to run data anywhere, so you can leverage your data across all clouds, on-premises, and access it in the most popular SaaS apps. This solution is built with and for AI, so you can get the latest tools for machine learning analysis, prompting, tuning, training, and deploying custom foundation models—all connected to your business data.

Explore the Data and AI cloud

900+

Partners and software integrations in our data and AI ecosystem

135

Languages translated in just a few clicks with Translation Hub

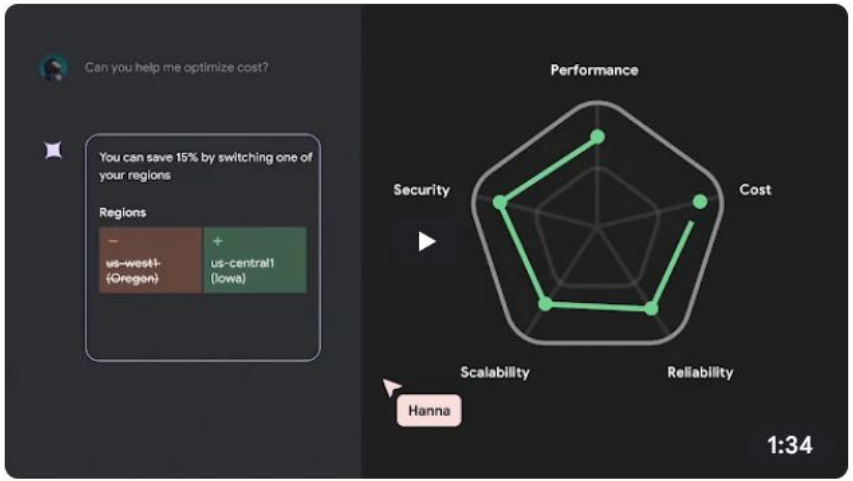
>70%

Of generative AI unicorns are Google Cloud customers

110+

TB of data per second analyzed by BigQuery customers

Advantages



Gemini for Google Cloud: the next frontier in AI-powered developer productivity

Modernize your infrastructure

You're ready for AI, but is your cloud? Google Cloud helps developers build quickly, securely, and cost effectively with the next generation of modern infrastructure designed to meet specific workload and industry needs. Get infrastructure that's optimized for AI, container-based applications, traditional enterprise workloads, and high-performance, distributed workloads—all while helping to cut costs and your carbon footprint.

Explore infrastructure modernization solutions

Advantages

1,800

Organizations get back to business quickly every year post-breach

40K+

Vulnerabilities found by Google in open source software projects

6M+

Sites protected by reCAPTCHA

9B

Files and URLs analyzed in threat observatory platform VirusTotal

Get built-in security

Benefit from the same security capabilities that Google uses to keep more people and organizations safe online than anyone else. We help organizations transform their cybersecurity programs with frontline intelligence from Mandiant to understand the latest cyber attacks; a modern security operations platform for detecting, investigating, and responding to threats; and a secure-by-design, secure-by-default infrastructure platform with controls to help maintain digital sovereignty.

Strengthen security with AI

Use case

https://cloud.google.com/trace/docs/trace-app-latency?hl=en

View latency of app requests

Learn how to collect and view latency data from your applications:

1. Create a Google Kubernetes Engine (GKE) cluster by using the Google Cloud CLI.
2. Download and deploy a sample application to your cluster.
3. Create a trace by sending an HTTP request to the sample application.
4. View the latency information of the trace you created.
5. Clean up.

Before you begin

1. Security constraints defined by your organization might prevent you from completing the following steps. For troubleshooting information, see Develop applications in a constrained Google Cloud environment.
2. In the Google Cloud console, on the project selector page, select or create a Google Cloud project.

   ★ **Note**: If you don't plan to keep the resources that you create in this procedure, create a project instead of selecting an existing project. After you finish these steps, you can delete the project, removing all resources associated with the project.

   Go to project selector
3. Make sure that billing is enabled for your Google Cloud project.
4. Enable the Google Kubernetes Engine and Cloud Trace APIs.

   Enable the APIs

Create a GKE cluster

1. In the toolbar, click **Activate Cloud Shell**, and then perform the following steps in the Cloud Shell.
2. Create a cluster:

```
gcloud container clusters create cloud-trace-demo --zone us-central1-c
```

   The previous command, which takes several minutes to complete, creates a standard cluster with the name `cloud-trace-demo` in the zone `us-central1-c`.

3. Configure `kubectl` to automatically refresh its credentials to use the same identity as the Google Cloud CLI:

```
gcloud container clusters get-credentials cloud-trace-demo --zone us-central1-c
```

4. Verify access to your cluster:

```
kubectl get nodes
```

A sample output of this command is:

```
NAME                                                STATUS   ROLES    AGE   VERSION
gke-cloud-trace-demo-default-pool-063c0416-113s   Ready    <none>   78s   v1.22.12-gke.2300
gke-cloud-trace-demo-default-pool-063c0416-1n27   Ready    <none>   79s   v1.22.12-gke.2300
gke-cloud-trace-demo-default-pool-063c0416-frkd   Ready    <none>   78s   v1.22.12-gke.2300
```

Download and deploy and application

Download and deploy a Python application, which uses the Flask framework and the OpenTelemetry package. The application is described in the About the app section of this page.

In the Cloud Shell, do the following:

1. Clone a Python app from GitHub:

```
git clone https://github.com/GoogleCloudPlatform/python-docs-samples.git
```

2. Run the following command to deploy the sample application:

```
cd python-docs-samples/trace/cloud-trace-demo-app-opentelemetry && ./setup.sh
```

The script `setup.sh` takes several minutes to complete.

The script configures three services using a pre-built image and then waits for all resources to be provisioned. The workloads are named `cloud-trace-demo-a`, `cloud-trace-demo-b`, and `cloud-trace-demo-c`.

A sample output of this command is:

```
deployment.apps/cloud-trace-demo-a is created
service/cloud-trace-demo-a is created
deployment.apps/cloud-trace-demo-b is created
service/cloud-trace-demo-b is created
deployment.apps/cloud-trace-demo-c is created
service/cloud-trace-demo-c is created

Wait for load balancer initialization complete......
Completed.
```

Create trace data

A *trace* describes the time it takes an application to complete a single operation.

To create a trace, in the Cloud Shell, run the following command:

```
curl $(kubectl get svc -o=jsonpath='{.items[?(@.metadata.name=="cloud-trace-demo-a")].status.loadBa
```

The response of the previous command looks like the following:

```
Hello, I am service A
And I am service B
Hello, I am service C
```

You can execute the `curl` command multiple times to generate multiple traces.

View latency data

1. In the Google Cloud console, go to the **Trace explorer** page:

   Go to Trace explorer

   You can also find this page by using the search bar.

   Each trace is represented by a dot on the graph and a row in the table.

   In the following screenshot shows multiple traces:

Exercise (will not be given grade)

https://cloud.google.com/trace/docs/trace-app-latency?hl=en

Write report thereafter

References

1) https://cloud.google.com/
2) https://cloud.google.com/learn/what-is-cloud-computing?hl=ru