Compute Engine

Secure and customizable compute service that lets you create and run virtual machines on Google's infrastructure.

New customers get $300 in free credits to spend on Google Cloud. All customers get a general purpose machine (e2-micro instance) per month for free, not charged against your credits.

- [Predefined machine types:](#) Start running quickly with pre-built and ready-to-go configurations
- [Custom machine types](#): Create VMs with optimal amounts of vCPU and memory, while balancing cost
- [Spot machines:](#) Reduce computing costs by up to 91%.
- [Confidential computing](#): Encrypt your most sensitive data while it's being processed
- [Rightsizing recommendations:](#) Optimize resource utilization with automatic recommendations

KEY FEATURES

Choosing the right virtual machine type

Scale-out workloads (T2D)

T2D offers [the best price-performance](#) compared to general-purpose VMs from any of the leading public cloud vendors. It is the first instance type in the [Tau VM family](#) and comes in predefined shapes, with up to 60vCPUs per VM and 4GB of memory per vCPU. T2D is ideal for scale-out workloads like web servers, containerized microservices, media transcoding, and large scale java applications.

General purpose workloads (E2, N2, N2D, N1)

[E2](#), [N2](#), [N2D](#), and [N1](#) are general-purpose machines offering a good balance of price and performance, and are suitable for a wide variety of common workloads including databases, development and testing environments, web applications, and mobile gaming. They support up to 224 vCPUs and 896 GB of memory.

Ultra-high memory (M2, M1)

[Memory-optimized machines](#) offer the highest memory configurations with up to 12 TB for a single instance. They are well suited to memory-intensive workloads

such as large in-memory databases like SAP HANA, and in-memory data analytics workloads.

Compute-intensive workloads (C2, C2D)

Compute-optimized machines provide the highest performance per core on Compute Engine and are optimized for workloads such as high performance computing (HPC), game servers, and latency-sensitive API serving.

Most demanding applications and workloads (A2)

Accelerator-optimized machines are based on the NVIDIA Ampere A100 Tensor Core GPU. Each A100 GPU offers up to 20x the compute performance compared to the previous generation GPU. These VMs are designed for your most demanding workloads such as machine learning and high performance computing.

All features

VM Manager	VM Manager is a suite of tools that can be used to manage operating systems for large virtual machine (VM) fleets running Windows and Linux on Compute Engine.
Confidential VMs	Confidential VMs are a breakthrough technology that allows you to encrypt data in use—while it's being processed. It is a simple, easy-to-use deployment that doesn't compromise on performance. You can collaborate with anyone, all while preserving the confidentiality of your data.
Live migration for VMs	Compute Engine virtual machines can live-migrate between host systems without rebooting, which keeps your applications running even when host systems require maintenance.
Sole-tenant nodes	Sole-tenant nodes are physical Compute Engine servers dedicated exclusively for your use. Sole-tenant nodes simplify deployment for bring-your-own-license (BYOL) applications. Sole-tenant nodes give you access to the same machine types and VM configuration options as regular compute instances.

Custom machine types	Create a virtual machine with a custom machine type that best fits your workloads. By tailoring a custom machine type to your specific needs, you can realize significant savings.
Predefined machine types	Compute Engine offers predefined virtual machine configurations for every need from small general purpose instances to large memory-optimized instances with up to 11.5 TB of RAM or fast compute-optimized instances with up to 60 vCPUs.
Spot VMs	Affordable compute instances suitable for batch jobs and fault-tolerant workloads. Spot VMs provide significant savings of up to 91%, while still getting the same performance and capabilities as regular VMs.
Instance groups	An instance group is a collection of virtual machines running a single application. It automatically creates and deletes virtual machines to meet the demand, repairs workload from failures, and runs updates.
Persistent disks	Durable, high-performance block storage for your VM instances. You can create persistent disks in HDD or SSD formats. You can also take snapshots and create new persistent disks from that snapshot. If a VM instance is terminated, its persistent disk retains data and can be attached to another instance.
Local SSD	Compute Engine offers always-encrypted local solid-state drive (SSD) block storage. Local SSDs are physically attached to the server that hosts the virtual machine instance for very high input/output operations per second (IOPS) and very low latency compared to persistent disks.
GPU accelerators	GPUs can be added to accelerate computationally intensive workloads like machine learning, simulation, and

	virtual workstation applications. Add or remove GPUs to a VM when your workload changes and pay for GPU resources only while you are using them. Our new A2 VM family is based on the NVIDIA Ampere A100 GPU. You can learn more about the A2 VM family by requesting access to our alpha program.
Global load balancing	Global load-balancing technology helps you distribute incoming requests across pools of instances across multiple regions, so you can achieve maximum performance, throughput, and availability at low cost.
Linux and Windows support	Run your choice of OS, including Debian, CentOS, CoreOS, SUSE, Ubuntu, Red Hat Enterprise Linux, FreeBSD, or Windows Server 2008 R2, 2012 R2, and 2016. You can also use a shared image from the Google Cloud community or bring your own.
Per-second billing	Google bills in second-level increments. You pay only for the compute time that you use.
Commitment savings	With committed-use discounts, you can save up to 57% with no up-front costs or instance-type lock-in.
Container support	Run, manage, and orchestrate Docker containers on Compute Engine VMs with Google Kubernetes Engine.
Reservations	Create reservations for VM instances in a specific zone. Use reservations to ensure that your project has resources for future increases in demand. When you no longer need a reservation, delete the reservation to stop incurring charges for it.
Right-sizing recommenda-tions	Compute Engine provides machine type recommendations to help you optimize the resource utilization of your virtual machine (VM) instances. Use these recommendations to

	resize your instance's machine type to more efficiently use the instance's resources.
OS patch management	With OS patch management, you can apply OS patches across a set of VMs, receive patch compliance data across your environments, and automate installation of OS patches across VMs—all from a centralized location.
Placement Policy	Use Placement Policy to specify the location of your underlying hardware instances. Spread Placement Policy provides higher reliability by placing instances on distinct hardware, reducing the impact of underlying hardware failures. Compact Placement Policy provides lower latency between nodes by placing instances close together within the same network infrastructure.