

Welcome to Google Cloud

We're thrilled to have you join the Google Cloud Startups ecosystem.

Your account team is here to champion your success and help you navigate the resources and support available.



Onboarding guide content

REQUIRED

1. **Set up your Google Cloud Organization**
2. **Setup IAM permissions**
3. **Purchase Support**
4. **Enable Security Command Center**

5. **Set up Budgets and Alerts**
6. **Visualize your spending with Looker Studio**

RECOMMENDED

INFORMATIONAL

7. **Google for Startups program journey**
8. **Request free on-demand training licenses**
9. **Access the Gemini APIs**
10. **Use AI Hypercomputers**
11. **File a support case**
12. **Next steps**

Set up your Google Cloud Organization



GCP | Set up your organization

You can create a Google Cloud organization by purchasing a Google Workspace subscription or by creating an organization directly with Cloud Identity (without Google Workspace):

Google Workspace

Sign up for [Google Workspace](#).

Google for Startups program provides a free 1YR Google Workspace Business Plus subscription for new Google Workspace customers

[Compare Google Workspace editions](#)

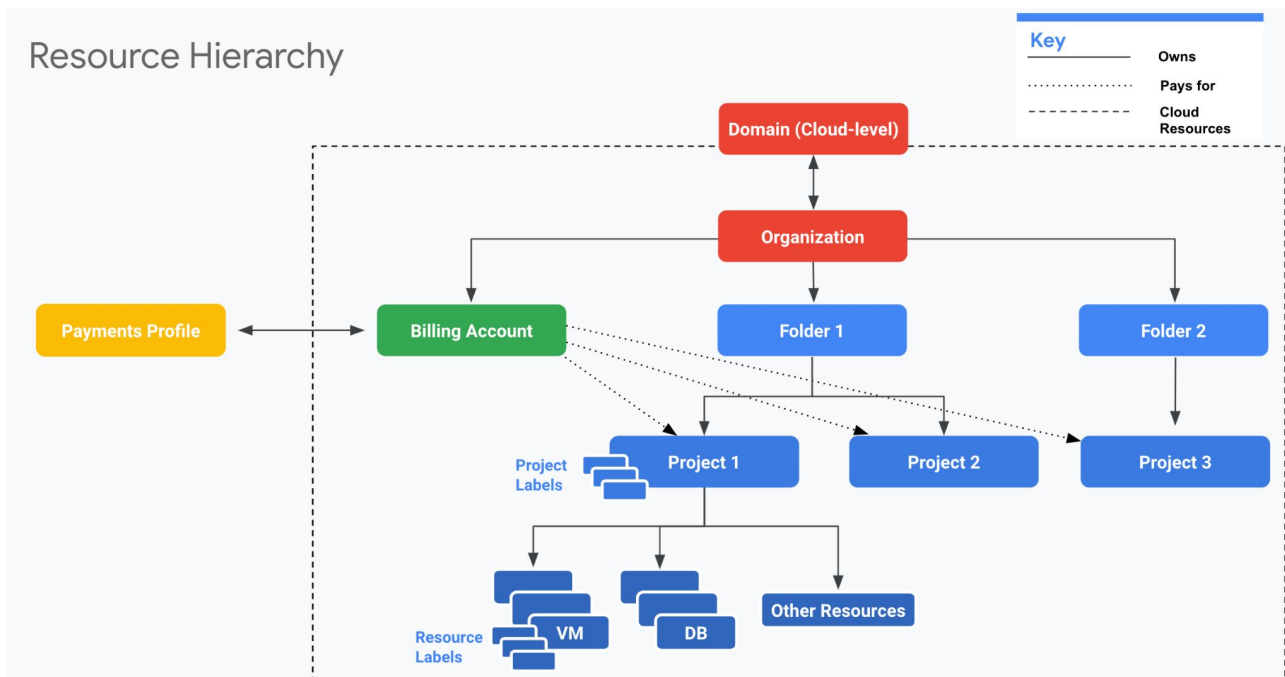


Sign up for [Cloud Identity](#).

Cost Effective
Free tier provides 50 user licenses by default

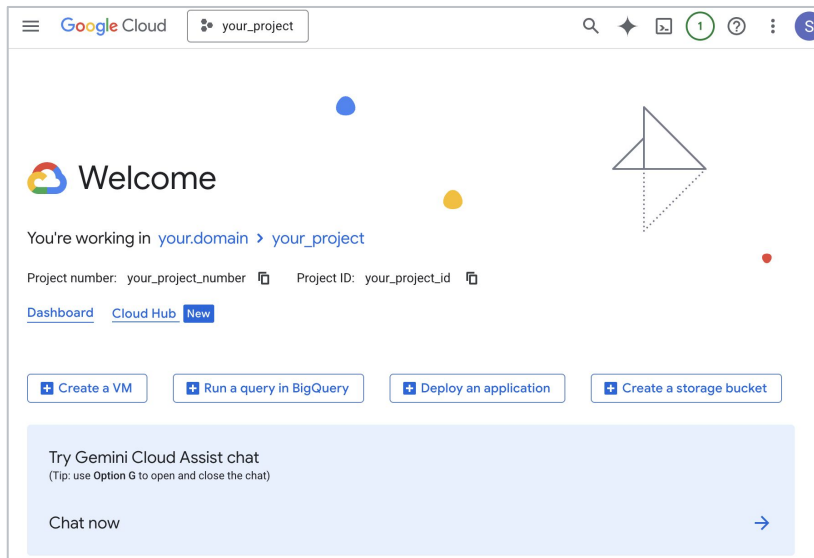
GCP | Setup additional resources

Please follow our [guided flow](#) to set up additional resources in your organization



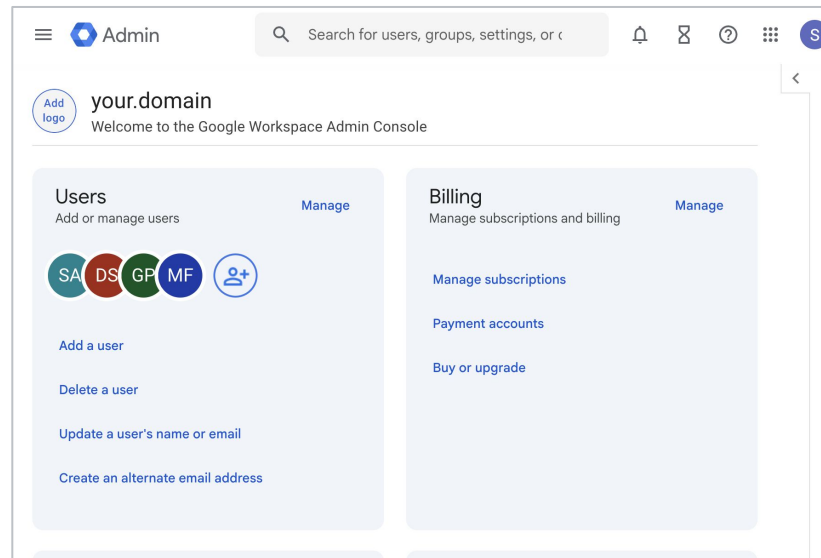
GCP | Consoles

console.cloud.google.com



Allows you to implement granular access controls via Identity and Access Management (IAM), enable APIs and services, create and manage infrastructure and much more.

admin.google.com



Allows you to manage people, groups, access controls, and Google Workspace domains

Setup your Google Cloud Organization



Please use our [Signup Issues Toolkit](#) to:

1. Recover your Google Admin (admin.google.com) account
2. Fix the “Domain already in use” issue



Additional tasks

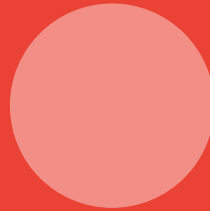
1. [Add an additional Super Admin to your organization](#)
2. [Migrating projects between organization resources](#)
3. [Enable, disable, or change billing for a project](#)
4. [Creating and managing organization policies](#)



Best practices:

1. [Super administrator account best practices](#)
2. [Security best practices for administrator accounts](#)

Setup IAM permissions



IAM | Get to the organization view

The resource selector is at the top left side of your screen

The diagram illustrates the process of selecting a domain or organization in the Google Cloud IAM resource selector. It shows two states of the resource selector: the initial state and the state after selection.

Initial State: The resource selector shows the Google Cloud logo and a dropdown menu with the text "project_name".

Resource Selector Modal: A modal titled "Select a resource" is shown. It includes a dropdown menu with "google.com", a search bar with the text "Search projects and folders", and tabs for "Recent", "Starred", and "All". The "Recent" tab is active. The modal displays a table of resources:

Name	Type	ID
✓ project_name ?	Project	project_id
google.com ?	Organization	organization_id

The "Organization" row is highlighted with a red box. Below the modal, the text "select your domain/organization" is displayed.

Final State: The resource selector shows the Google Cloud logo and a dropdown menu with the text "google.com".

IAM | Add recommended roles

Google Cloud your_domain Search (/) for resources, docs, products, and more

IAM & Admin / IAM

Overview Preview

IAM

PAM

Principal Access Boun...

Organizations

Identity & Organization

Policy Troubleshooter

Policy Analyzer

Organization Policies

Service Accounts

Workload Identity Fede...

Workforce Identity Fed...

Labels

Tags

Settings

Privacy & Security

IAM

Allow Deny Recommendations history

Permissions for organization "your_domain"

These permissions affect this organization and all of its resources. [Learn more](#)

[View by principals](#) View by roles

[Grant access](#) [Remove access](#)

Filter Enter property name or value

Type	Principal	Name	Role
<input type="checkbox"/>	admin@your_domain	Super Admin	Organization Administrator

Please add all 5 recommended IAM roles to complete self-onboarding

Edit access to "your_domain"

Principal [?](#) Organization

admin@your_domain your_domain

Assign roles

Roles are composed of sets of permissions and determine what the principal can do with this resource. [Learn more](#)

- Role: **Organization Administrator** [IAM condition \(optional\) ?](#) [+ Add IAM condition](#)
Access to manage IAM policies and view organization policies for organizations, folders, and projects.
- Role: **Billing Account Administrator** [IAM condition \(optional\) ?](#) [+ Add IAM condition](#)
Authorized to see and manage all aspects of billing accounts.
- Role: **Support Account Administrator** [IAM condition \(optional\) ?](#) [+ Add IAM condition](#)
Allows management of a support account without giving access to support cases.
- Role: **Tech Support Editor** [IAM condition \(optional\) ?](#) [+ Add IAM condition](#)
Full read-write access to technical support cases (applicable for GCP Customer Care and Maps support).
- Role: **Security Center Admin** [IAM condition \(optional\) ?](#) [+ Add IAM condition](#)
Admin(super user) access to security center.

[+ Add another role](#)

[Save](#) [Test changes](#) [Cancel](#)

Purchase Support



Your account team **does not** engage in troubleshooting.

Basic Support (default) **does not** give you access to our support team.

Having a paid support subscription, will ensure that your issues in production are resolved in a timely manner.

Google for Startups - Scale Tier or above already provides dedicated credits to purchase Enhanced Support

Support | Customer Care Portfolio

If you received \$12k in support credits, we **REQUIRE** you to purchase **Enhanced Support**

You **WILL NOT** be able to file support cases if you are in the Basic Support tier (Free / default)

	Basic Support	Standard Support	Enhanced Support	Premium Support
Documentation	✓	✓	✓	✓
Community Support	✓	✓	✓	✓
Multi-channel billing support	✓	✓	✓	✓
Unlimited access to support ⓘ		✓	✓	✓
Multi-channel technical support		✓	✓	✓
Available languages		English	English, Japanese, Mandarin, Chinese, and Korean	English, Japanese, Mandarin, Chinese, and Korean
Service times		8/5 response for high-impact issues	24/7 response for high & critical-impact issues	24/7 response for high & critical-impact issues
Priority issues/cases		4-hour response time	1-hour response time	15-minute initial response time
Active Assist Recommendations API ⓘ	ⓘ	✓	✓	✓
Issue/case escalations ⓘ	ⓘ		✓	✓

Standard Support is recommended if you did not receive support credits

Support | Navigate to the Support page

console.cloud.google.com/support

Google Cloud Search (/) for resources, docs, products, and more

Support Overview

- Overview**
- Cases
- Billing support
- Phone support
- Community support
- Settings

i This information applies only to support for [Google Cloud Platform products](#). You can find support options for other products, such as [Maps](#) and [Apps APIs](#), in their own developer documentation site. Go to [Products](#) for a complete list.

Google Cloud incidents

i **No project selected**
Please select a project to view incidents

→ [Go to Personalized Service Health](#)

i Support information

Your current Customer Care service: Standard

Get responses within four business hours for P2 cases

Resource: Check out actionable recommendations to keep your cloud optimized

→ **View Customer Care services**

→ [Manage IAM permissions](#)

→ [Get started with Standard Support](#)

? [Get help](#)

Support | Purchase a support subscription

Support


[← Customer Care services](#)

Overview

Cases

Billing support

Phone support

Community support

Settings

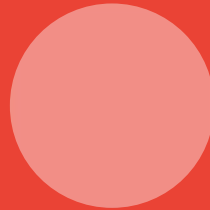
Sign up for Customer Care

Get the most from Google Cloud with Customer Care, choose the technical support service that's right for your organization, and further tailor it to fit your needs with additional Value-Add Services.

	Basic Support	Standard Support	Enhanced Support	Premium Support
	\$0	Minimum spend of \$29.00 OR 3% of monthly Cloud charges	Minimum spend of \$100.00 OR 10% of monthly Cloud charges for the first \$0-\$10K 7% of monthly Cloud charges from \$10K-\$80K 5% of monthly Cloud charges from \$80K-\$250K 3% of monthly Cloud charges over \$250K Charges will be at least the minimum charge of \$29.00 or the result of the calculation, whichever is higher. Calculate cost	Minimum spend of \$15,000.00 OR 10% of monthly charges for the first \$0-\$150K 7% of monthly charges from \$150K-\$500K 5% of monthly charges from \$500K-\$1M 3% of monthly charges over \$1M Charges will be at least the minimum charge of \$15,000.00 or the result of the calculation, whichever is higher. Calculate cost
		BUY NOW	BUY NOW	CONTACT SALES

If "BUY NOW" is grayed out, review your IAM permissions again.

Enable Security Command Center



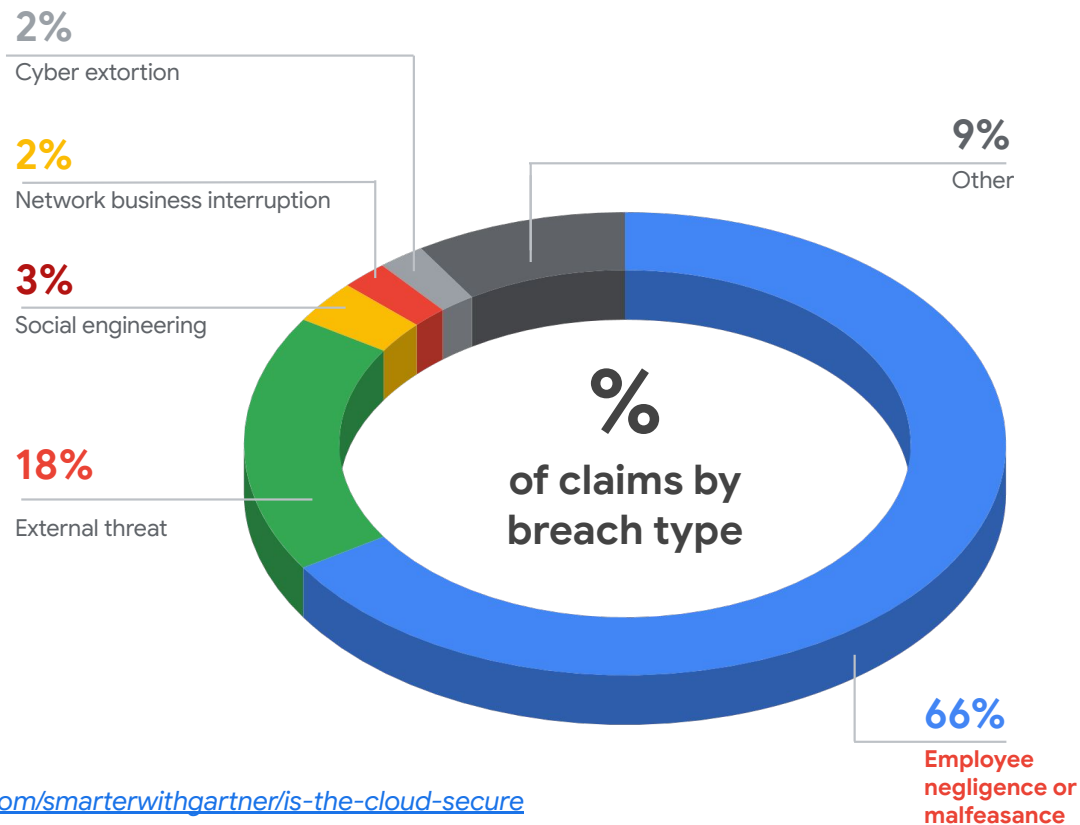
SCC | Security is a shared responsibility.

You're the expert in knowing the security and regulatory requirements for your business, and knowing the requirements for protecting your confidential data and resources. When you run your workloads on Google Cloud, you must identify the security controls that you need to configure in Google Cloud to help protect your confidential data and each workload. To decide which security controls to implement, you must consider the following factors:

- Your regulatory compliance obligations
- Your organization's security standards and risk management plan
- Security requirements of your customers and your vendors

If you do not have a security engineer, we highly recommend that you purchase **SCC Premium to reduce misconfigurations and understand your security posture.**

SCC | Misconfigurations are the largest cause of breaches



To learn more: gartner.com/smarterwithgartner/is-the-cloud-secure

SCC | Some misconfiguration surfaced by SCC Premium

Storage



- Publicly exposed buckets
- Use of legacy bucket ACLs

Networking



- Overly permissive firewall rules
- Use of default and/or legacy networks
- Subnetworks that do not use private access to Google APIs

Logging/ Monitoring



- Monitoring disabled
- Storage buckets with logging disabled
- Stackdriver monitoring for Kubernetes clusters not enabled
- VPC Flow logs disabled

VM Instances



- IP forwarding enabled
- SSH and access misconfigurations

GKE Clusters



- Private cluster disabled
- Network policy disabled
- Master authorized network disabled
- IP alias disabled
- Legacy authorization enabled

Compliance



- Monitoring against the CIS GCP Foundation benchmark
- Certified for CIS Benchmarks 1.0 and 1.1, with more certifications to come
- PCI, ISO, and NIST support

SCC | Navigate to the SCC page

The screenshot displays the Google Cloud Security Command Center (SCC) interface. At the top, the Google Cloud logo and the domain 'your_domain' are visible. The left sidebar contains navigation options: Security, Security Command Center, Risk Overview (highlighted), Threats, Vulnerabilities, Compliance, and Assets. A search bar at the top right of the main content area is active, showing 'Search Results' for 'Risk Overview Security'. The search results are displayed in a white box with a red border around the 'Risk Overview Security' entry. Below the search results, it indicates 'Showing resource results for your_domain only'. The main content area features a 'Get Started Today' section with two steps: '1 Select Premium Tier for the best security' and '2 Configure detectors and services for the right'. Below this, there are five main sections: Visibility (Asset Inventory, Asset tracking), Risk Management (Platform Misconfigurations, OS, Application, Web app vulnerabilities), Threat Detection (Malicious activity in your Cloud Platform, Malicious activity in your compute, Malicious activity in your subdomains), and Compliance Reporting (Best practice: CIS 1.3, NIST 800-53, Industry Standards, PCI DSS v1.2, ISO 27001). A red play button icon is overlaid on the Threat Detection section.

SCC | Navigate to the SCC page

Make sure you are still in the Organization view

Security Command Center

Making Google Cloud the Safest Place for your Applications and Data

Get Started Today

- 1 Select Premium Tier for the best security
- 2 Configure detectors and services for the right security coverage
- 3 Start using SCC results to improve your Google Cloud security

Get Security Command Center

Brief introduction to Security Command Center

Watch on YouTube

SCC | Select SCC Premium

Google Cloud your_domain

Security Command Center

Search

Security Get Security Command Center

Security Command Center

- Risk Overview
- Threats
- Vulnerabilities
- Compliance
- Assets
- Findings
- Sources
- Access Insights
- Posture Management
- Detections and Controls
 - Google SecOps
 - reCAPTCHA
 - Model Armor
 - Web Security Scanner
 - Cyber Insurance Hub
 - Binary Authorization
 - Advisory Notifications
 - Access Approval
 - Managed Microsoft AD
- Data Protection
 - Sensitive Data Protection
 - Data Loss Prevention

- Select a tier
- Select services
- Grant roles
- Complete setup

Get started

	Enterprise	Premium	Standard
	<p>Includes cloud-native application protection capabilities for your organization, such as asset discovery, enhanced threat detection, case management, and more for your Google Cloud or multicloud estate.</p> <p>Contact us</p>	<p>Includes enhanced vulnerability scanning and detection of misconfigurations. PLUS advanced Google Cloud-native Threat detection, Compliance reporting for regulatory and industry mandates, and more.</p> <p>✓ Selected</p>	<p>Includes vulnerability scanning for certain application security risks and detection of basic misconfigurations that could weaken your security posture.</p> <p>Select</p>
Pricing	Based on the usage of certain Google Cloud and AWS services. View pricing details	Based on the usage of certain Google Cloud services. View pricing details	Free
What you get	<ul style="list-style-type: none"> ✓ External asset discovery with Mandiant Attack Surface Management ✓ Google Security Operations with enhanced threat detection and case management ✓ Enhanced compliance reporting ✓ Multicloud findings 	<ul style="list-style-type: none"> ✓ Enhanced vulnerability scanning ✓ Misconfiguration identification ✓ Threat detection ✓ Compliance reporting ✓ Asset query capability 	<ul style="list-style-type: none"> ✓ Application vulnerability scanning ✓ Basic misconfiguration identification
Learn more	Security Command Center Enterprise	Security Command Center Premium	Security Command Center Standard

Select Premium, click "Next"

Next

Cancel

SCC | Enable premium services

Google Cloud | your_domain | Security Command Center | Search

Security | Get Security Command Center | Feedback

Security Command Center

- Risk Overview
- Threats
- Vulnerabilities
- Compliance
- Assets
- Findings
- Sources
- Access Insights
- Posture Management

Detections and Controls

- Google SecOps
- reCAPTCHA
- Model Armor
- Web Security Scanner
- Cyber Insurance Hub
- Binary Authorization
- Advisory Notifications
- Access Approval
- Managed Microsoft AD
- Marketplace
- Release Notes

Get Security Command Center

- Select a tier
- Select services
- Grant roles
- Complete setup

Container Threat Detection Preview

Available for Premium or Enterprise

Use kernel-level instrumentation to identify potential compromise of GKE containers, including suspicious binaries. [Learn more about Container Threat Detection](#)

Container Threat Detection deploys a DaemonSet on your clusters. Ensure your clusters have sufficient resources available to run these DaemonSets. [Learn more](#)

Virtual Machine Threat Detection

Available for Premium or Enterprise

Analyze Compute Engine instances to identify threats, including cryptomining abuse. [Learn more about Virtual Machine Threat Detection](#)

Supported cloud providers [?](#)

Google Cloud, AWS (Preview)

Cloud Run Threat Detection Preview

Available for Premium or Enterprise

Analyze Cloud Run workloads to identify threats, including malicious binary, reverse shell and others. [Learn more](#)

Vulnerability Assessment

Available for Premium or Enterprise

Scan your Compute Engine instances for common vulnerabilities and exposures (CVE).

Supported cloud providers [?](#)

Google Cloud (Preview), AWS

Notebook Security Scanner Preview

Available for Premium or Enterprise

Scan your Colab Enterprise notebooks for package vulnerabilities in the integrated open-source Python packages. [Learn more](#)

Next Back Cancel

Enable Cloud Run Threat Detection if necessary

Click "Next"

SCC | Grant Permissions

Google Cloud your_domain Security Command Center

Security Security Command Center

Security Command Center

- Risk Overview
- Threats
- Vulnerabilities
- Compliance
- Assets
- Findings
- Sources
- Access Insights
- Posture Management

Detections and Controls

Get Security Command Center

Feedback

Grant Permissions

Security Command Center created a service account that doesn't have Cloud IAM permissions. The account must be granted the required IAM roles in order to scan resources for vulnerabilities, store findings, and detect threats.

Choose how to grant roles

Grant roles automatically

Grant roles manually

Service account details

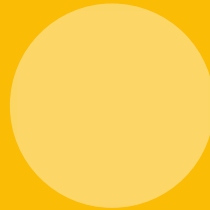
Review the service accounts that were created and the required roles.

Grant Roles

Test account Back Cancel

Click "Grant Roles"
Then click "Next"

Set up Budgets and Alerts



Budget Alert | Billing budgets and alerts

- A budget can be set for a project to send alerts
- Budgets may be set for a **billing account** or for a **project**. After the monthly budget is set, custom thresholds for alerts may be set (for example, 50%, 75%, 100%)
- **Reaching a budget** or threshold has **no resource restriction implications**. All GCP resources continue to function normally.
- You can use Cloud Billing budget notifications with third-party or homegrown cost-management solutions, as well as Google Cloud services.
- **We do not process refunds for runaway bills**, so please set up a budget alert to avoid them.

Budgets & alerts Billing-Reports-Demo [CREATE BUDGET](#) [DELETE](#)

Budgets track expenses within a Google Cloud Platform project or billing account. Your budget can be a specified amount or based on previous spend. You can set alerts to notify billing admins and users when a budget goes over a specified amount.

Setting a budget does not cap resource or API consumption. [Learn more.](#)

<input type="checkbox"/>	Budget name ↑	Budget type	Applies to	Trigger alerts at	Spend and budget amount
<input type="checkbox"/>	amazing-application-budget	Specified am...	Project "Phills Amazing Application"	90%, 100%, and 1...	\$520.85 / Excludes -
<input type="checkbox"/>	Bigquery + AutoML budget - exceed with more queries	Specified am...	Products: "BigQuery" and "Cloud AutoML"	90% and 100%	\$28.90 / No credits
<input type="checkbox"/>	bla	Specified am...	This billing account	50% and 90%	\$1,156.1 / Includes -
<input type="checkbox"/>	Compute Engine Budget - easy to exceed	Specified am...	Product "Compute Engine"	100%	\$1,043.6 / Excludes -
<input type="checkbox"/>	First Demo Budget	Specified am...	Project "Phills Amazing Application"	50%, 90%, and 10...	\$520.85 / Excludes -
<input type="checkbox"/>	Many Products Budget	Specified am...	120 products	50%, 90%, and 10...	\$1,156.1 / Includes -
<input type="checkbox"/>	Many Projects + Products	Specified am...	5 projects and 120 products	50%, 90%, 100%, ...	\$605.56 / Excludes -

Budget Alert | Navigating the console

The screenshot shows the Google Cloud console interface. A search bar at the top right contains the text "budgets & alerts". Below the search bar, a dropdown menu displays search results. The first result, "Budgets and alerts", is highlighted with a red box. The search results list includes:

- Budgets and alerts** (Billing)
- Examples of automated cost control responses | Cloud Billing (This document has examples and step-by-step instructions on...)
- Billing Budget API Quickstart (Node.js) (Interactive Tutorial)
- InfluxDB Cloud (Official Version) (InfluxData, the creators of InfluxDB)
- Unravel Platform (Unravel Data)
- Customize budget alert email recipients | Cloud Billing (To customize who receives the budget alert emails for a specific...)
- Using the Cloud Billing Budget API (Learn how to send a few simple requests to the Cloud Billing...)
- Cloud Billing Budget API (Google Enterprise API)
- Get started with the Cloud Billing Budget API (A budget enables you to track your actual Google Cloud spend...)
- Spot by NetApp (NetApp, Inc.)

The background shows the "Billing / Budgets and alerts / Create budget" page. The left sidebar contains navigation options: Overview, Cost management (Reports, Cost table, Cost breakdown, Budgets & alerts, Billing export, Anomalies), and Cost optimisation (FinOps hub, Committed-use discou..., CUD analysis, Pricing, Cost estimation, Release notes). The bottom of the page shows a "Finish" button and a "Cancel" button. The URL at the bottom is <https://cloud.google.com/billing/docs/how-to/budgets-notification-recipients>.

Budget Alert | Create

The screenshot shows the Google Cloud Billing console interface. At the top, there's a navigation bar with the Google Cloud logo, a search bar containing the word 'budget', and several utility icons. Below the navigation bar, the page title is 'Billing / Budgets and alerts'. The left sidebar contains a list of navigation items: Overview, Cost management (Reports, Cost table, Cost breakdown), Budgets & alerts (highlighted), Billing export, Anomalies, Cost optimisation (FinOps hub, Committed-use discou..., CUD analysis, Pricing, Cost estimation), and Release notes. The main content area is titled 'Budgets & alerts' and includes a '+ Create budget' button and a 'Delete' icon. A 'Learn' link is also present. The central content area features a 'Billing' card with the following text: 'Budgets & alerts', 'Avoid surprises on your bill by creating budgets to monitor all your Google Cloud charges in one place. After you've set a budget, you can create budget alerts to email billing admins and users when charges exceed a certain amount.', and a 'Create budget' button highlighted with a red border. A 'Show debug panel' link is visible on the right side of the main content area.

Budget Alert | Define settings

Define the scope:

- Time range
- Projects, services, labels

Set the budget amount:

- Specified amount or last period's spend

Configure alerts:

- Thresholds (% , amount)
- Trigger on actual or forecasted spend

Manage notifications:

- Email to billing admins, project owners, or custom recipients
- Programmatic notifications with Pub/Sub

Google Cloud | budget

Billing / Budgets and alerts / Edit budget

Billing account: Argolis Billing (alexhoang)

Edit budget

Amount

Set a monthly budget amount

Budget type: Specified amount

A fixed amount that your spend will be compared against.

Target amount: \$ 10000

Actions

Set alert threshold rules

Send email alert notifications after the actual or forecasted spend exceeds a percentage of the budget or a specified amount. [Learn more](#)

Percentage of budget	Amount	Trigger on
50 %	\$ 5000	Actual
90 %	\$ 9000	Actual
100 %	\$ 10000	Actual

[+ Add threshold](#)

Manage notifications

Email alerts to billing admins and users

Email alerts to project owners

Link Monitoring email notification channels to this budget
Select a project and a maximum of 5 Monitoring email notification channels.

Connect a Pub/Sub topic to this budget
Select a project and Pub/Sub topic. Anyone who can view this budget will also be able to view the project ID and the topic name. It may not be possible to add a Pub/Sub topic if it belongs to an organisation that has [domain-restricted sharing](#) enabled.

Cost trend
1 February 2024 – 28 February 2025

\$10k

\$5k

\$0

Feb Mar Apr May Jun Jul Aug Sept Oct Nov Dec Jan Feb

• Total cost

[View report](#)

Budget Alert | Status

Google Cloud

budget

Billing / Budgets and alerts

Billing account: Argolis Billing (alexhoang)

Budgets & alerts [+ Create budget](#) [Delete](#)

Budgets track expenses within a Google Cloud project or billing account. You can set alerts to notify billing admins and users when your costs exceed a budget. Your costs are usually recorded within 24 hours. Set your budget to a lower amount to account for the time to report your costs.

Setting a budget does not cap resource or API consumption. [Learn more.](#)

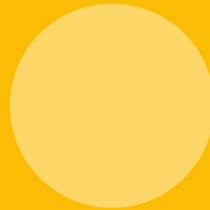
Filter: Enter property name or value

Budget name	Budget period	Budget type	Applies to	Trigger alerts at	Spend and budget amount
<input type="checkbox"/> test	Monthly	Specified amount	This billing account	50%, 90% and 100%	<div style="width: 100%; height: 10px; background-color: #ccc;"></div> \$0.00/\$10,000.00 No credits used

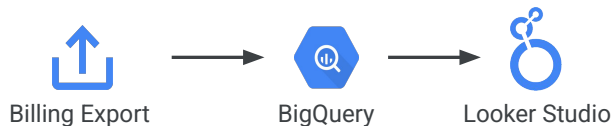
- View a list of your budgets and their status.
- Monitor spending with the progress bar.
- Easily modify or delete existing budgets.

You can [set up programmatic notifications](#) to [disable billing usage based on those notifications](#)

Visualize your spending with Looker Studio



Billing | Visualize your costs with Looker Studio

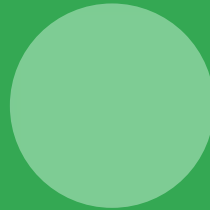


Billing data can be exported to Bigquery and other analytics tools to craft custom dashboards that dive deeper into your cloud usage and costs.

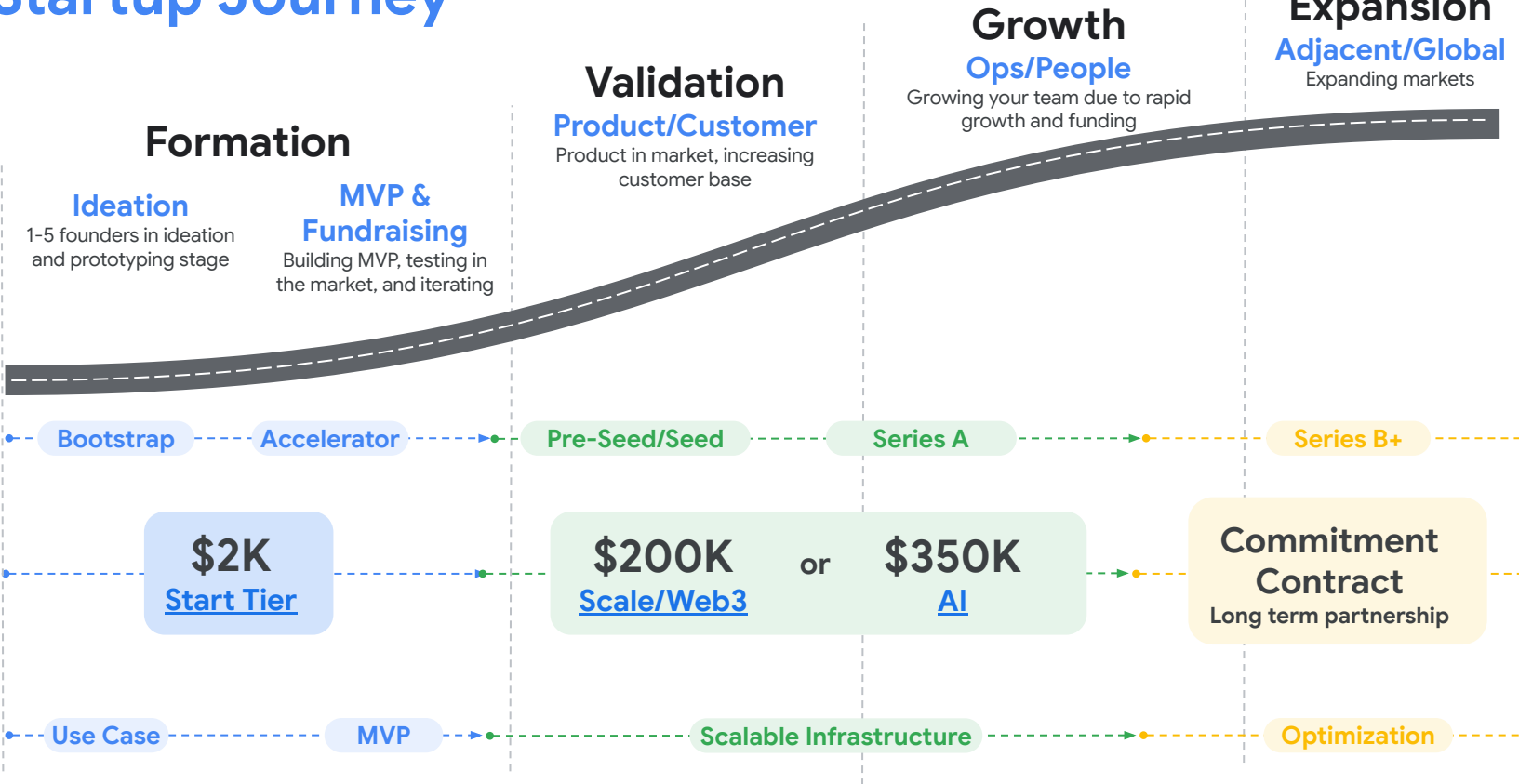
Template: bit.ly/billboard-template



Google for Startups program journey

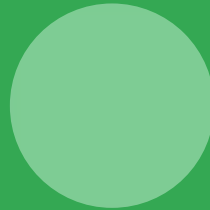


Startup Journey



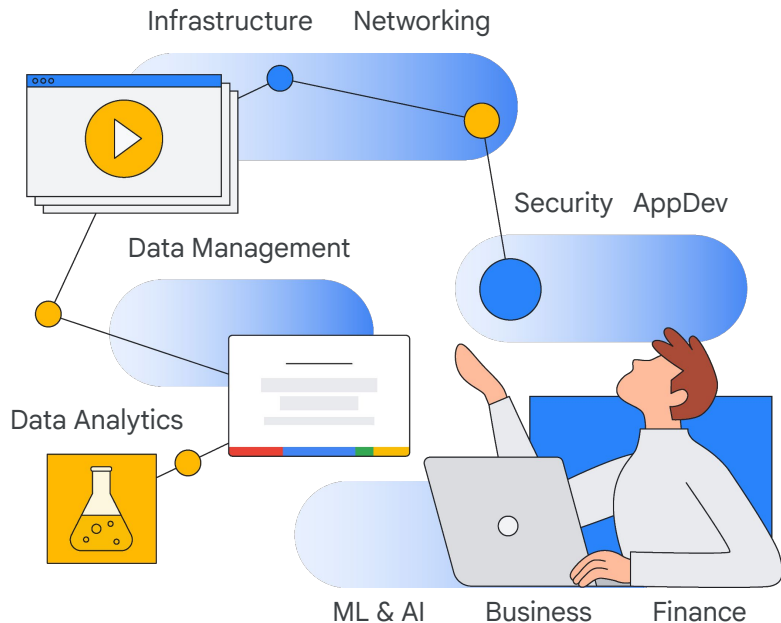
Learn more and apply at cloud.google.com/startup

Request free
on-demand training
licenses



Google Cloud Skills Boost for Organizations

Drive scalable learning for your team with intuitive administrative tools at no cost.



Guided hands-on labs

Build confidence and skills by practicing in a controlled environment

Intuitive administrative tools

Track your team's progress through administrative tools that make it easy to manage your team's learning and key skilling metrics

Skills validation

Measure and recognize cloud capabilities with Google Cloud Credentials

Content by Google Cloud

Content is designed, delivered, and maintained by Google Cloud

No-cost

Up to 500 users at no cost per organization. (Previously \$399/user USD). Additional users available upon request.

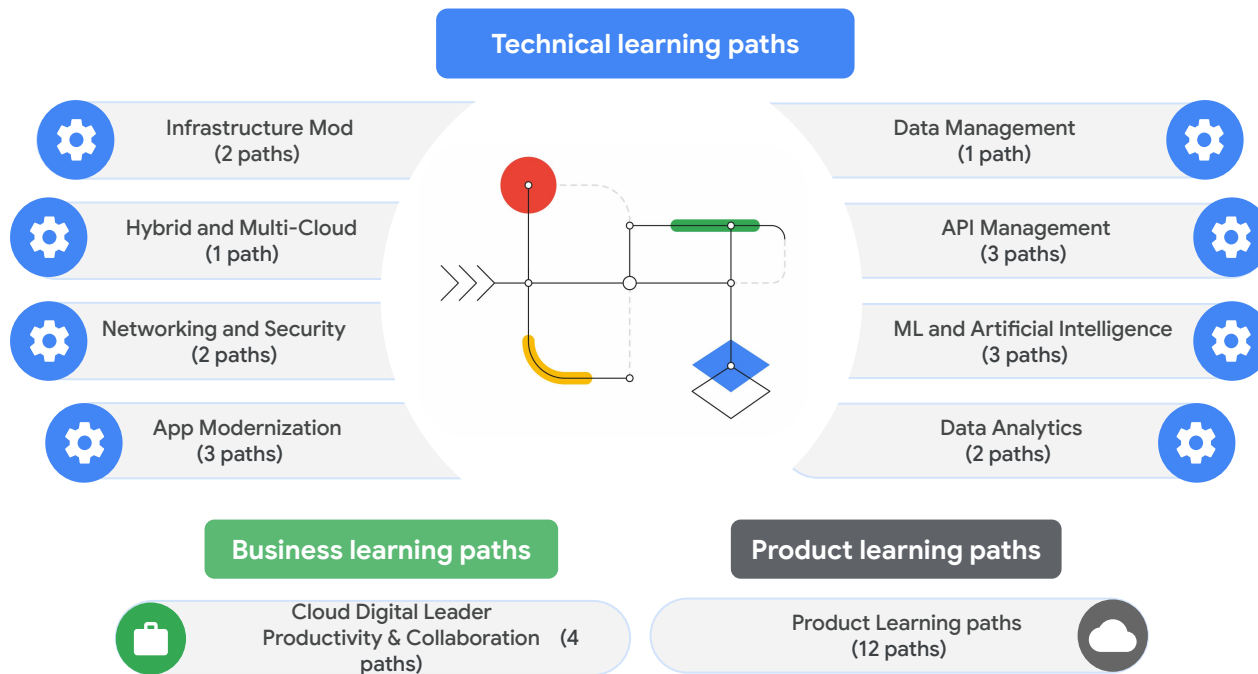
Curated catalog of content

Content created for a range of technical proficiency levels and addresses different learning styles - on-demand videos and courses, self-paced labs, and skill badges

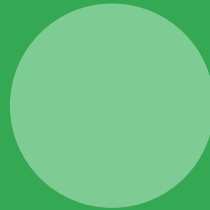
If you are interested, email your account team letting them know the amount of licenses that your team needs.

Learning Paths Cloud Skills Boost

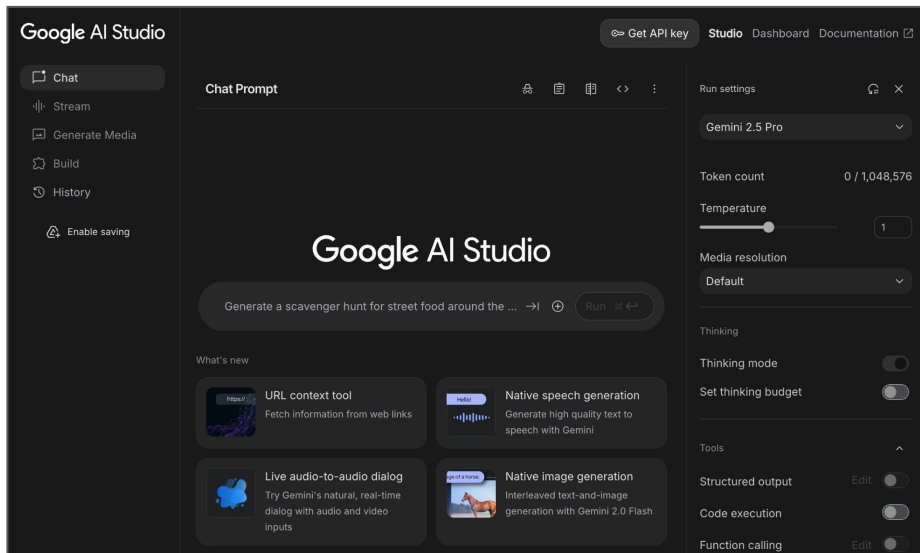
Our [learning paths](#) are curated collections of on-demand content for all business & technical proficiency levels



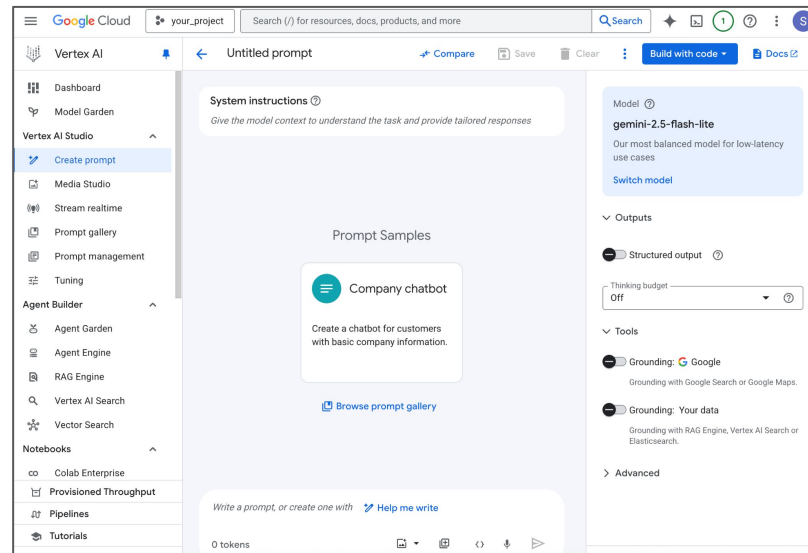
Access the Gemini APIs



Where can I access the Gemini APIs?



[Google AI Studio](#)



[Vertex AI](#) (Recommended)

Use case

Individual developers and students testing

Production Workloads. Enterprise grade platform.

Quota limits

[Spent based rate limits](#)

No RPM or TPM limits from day 1 due to [Dynamic Shared Quota](#)

Multi-region capacity

No supported

[Vertex AI Global Endpoint](#) leverages multi-region capacity which can help you reduce [429 errors](#) in production

Reserved capacity

No supported

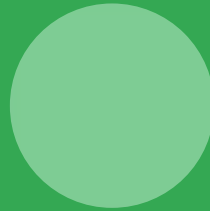
[Provisioned Throughput](#)

Quickstart

[Google AI Studio quickstart](#)

[Gemini API in Vertex AI quickstart](#)

Use AI Hypercomputers

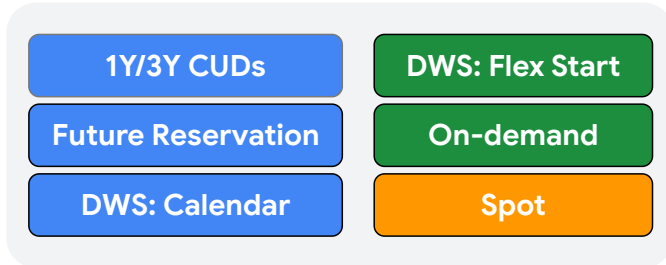


AI Hypercomputers in Google Cloud

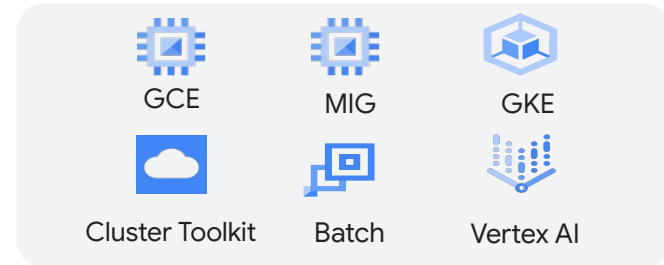
Proprietary + Confidential

AI Hypercomputer is a supercomputing system that is optimized to support your artificial intelligence (AI) and machine learning (ML) workloads. AI Hypercomputer uses best practices and systems-level designs to boost efficiency and productivity across AI pre-training, tuning, and serving.

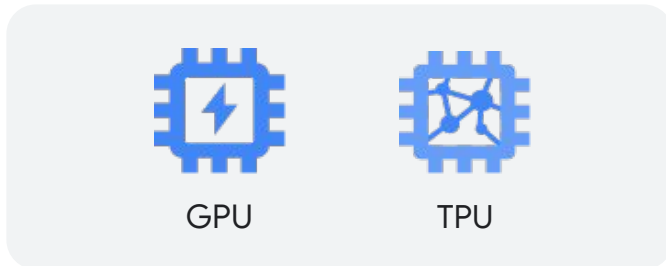
To get started, you need to follow the next steps:



1- Choose your consumption model



2- Select your deployment method



3- Select your accelerator



4- Get your GPU quota limit increased

Consumption options

Provisioning Options

Reserved capacity

Shared capacity

Preemptible

1Y/3Y CUDs

Future Reservation

DWS: Calendar

DWS: Flex Start

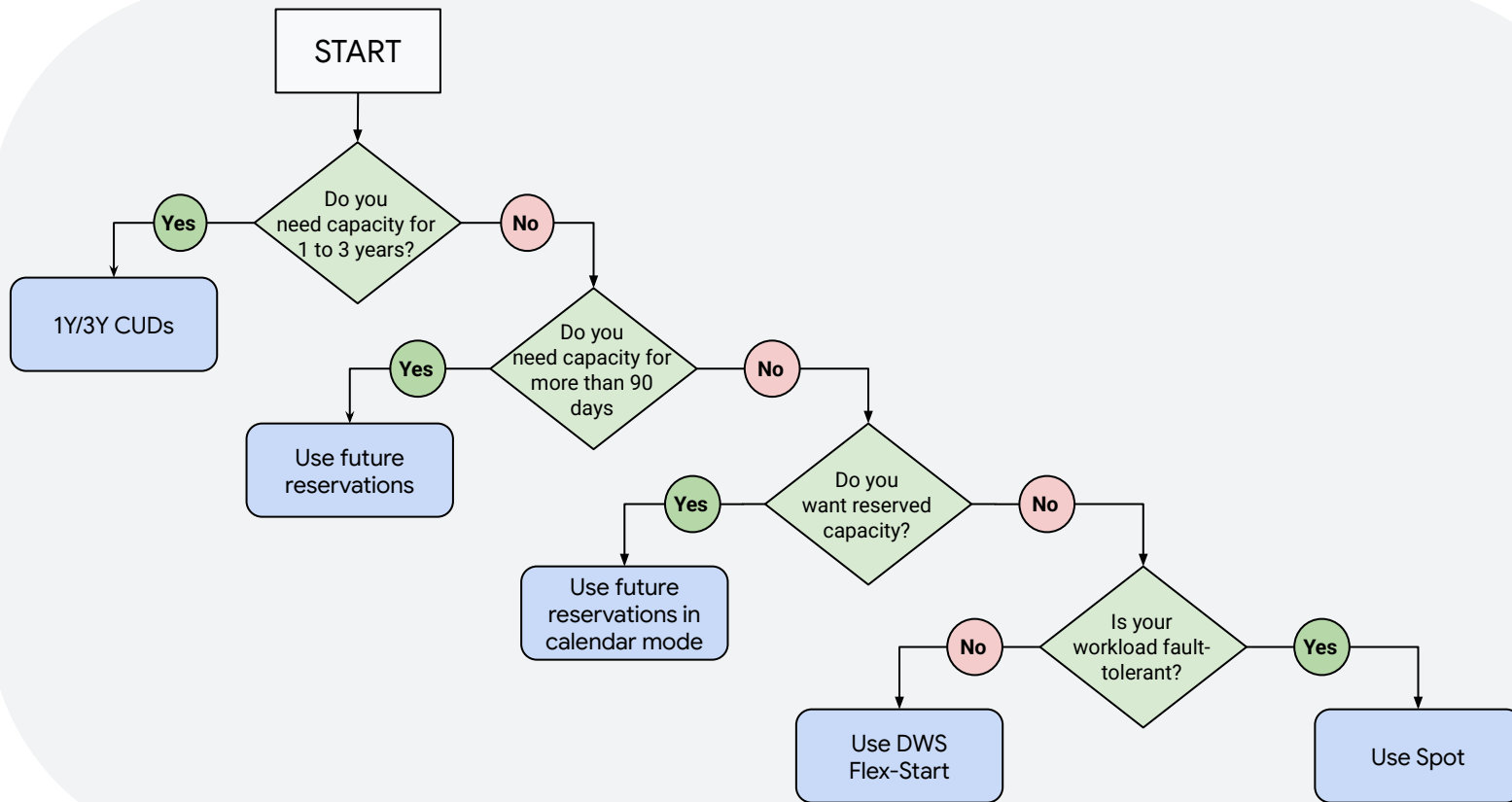
On-demand

Spot

	Reserved capacity	Reserved capacity	Reserved capacity	Shared capacity	Shared capacity	Preemptible
	1Y/3Y CUDs	Future Reservation	DWS: Calendar	DWS: Flex Start	On-demand	Spot
Analogy	Renting an apartment	Subletting an apartment	Booking a hotel	Waiting in line	Window shopping	“While supplies last”
Learn more	Resource-based committed use discounts	About Future reservations	About future Reservation in Calendar Mode	About DWS: Flex-Start	About Standard provisioning	Preemptible VM instances
Start Time	Next day 12 AM PT	Future date	Min 24 hours in advance	As soon as available	Very limited	As soon as available
EndTime	Fixed. Cannot cancel	Cannot cancel but can be extended	Fixed. Cannot cancel or extend	Cancel any time	Stop any time	Stopped by Google at any time
Duration	1 yr, 3 yr Can be extended	Min: 24 hours Max: no limit	Min: 1 day Max: 90 days	Min: 1 min Max: 7 days	Min: 1 min Max: no limit	Min: 1 min Max: no limit
Pricing	Discounted (based on duration)	On demand	Variable (up to 53% off from On demand)	Variable (up to 53% off from On demand)	On demand	Spot pricing
Deployment	Purchase commitments with attached reservations	Create future reservation request	Create a future reservation request in calendar mode	<ol style="list-style-type: none"> 1- GCE: MIG Resize Request 2- GKE: small batch workload & large-scale workload 3- Slurm: Cluster Toolkit 4- Vertex AI: Training & Inference 5- Batch: jobs with GPUs 	You can request on-demand VMs directly from Compute Engine or the deployment method of your choice.	You can request spot VMs directly from Compute Engine or the deployment method of your choice.

Google Cloud

How to choose between the consumption options?



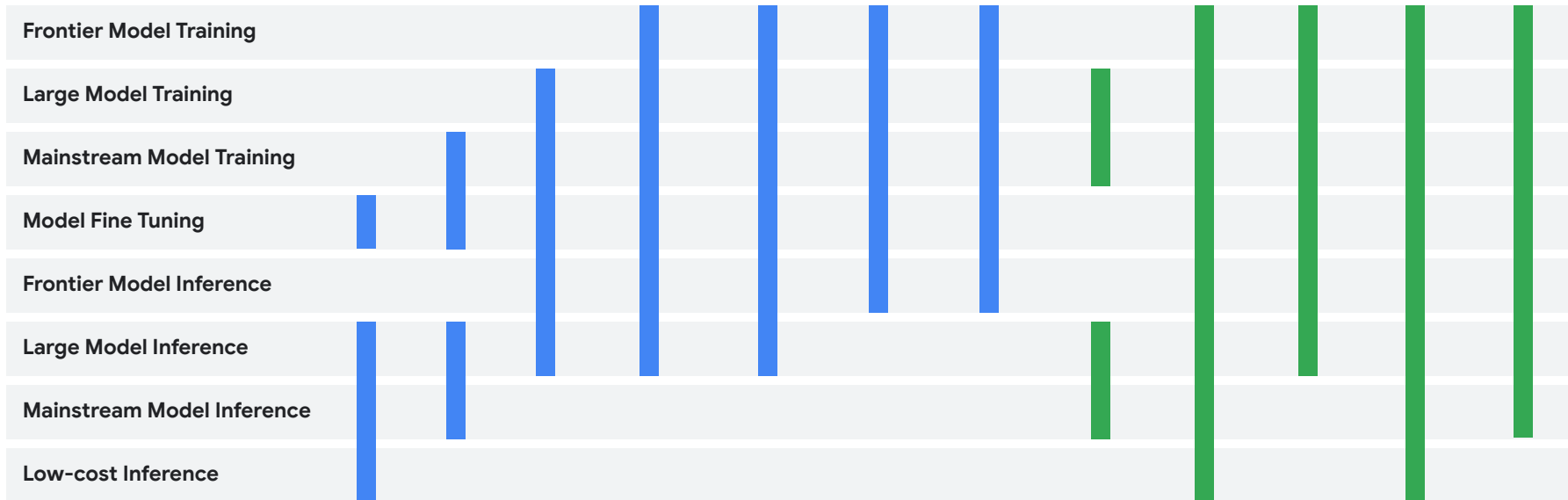
Supported Accelerators

● GPU

Click on the resource names to review their specs

● TPU

[G2 \(L4\)](#) [A2 \(A100\)](#) [A3 \(H100\)](#) [A3 Mega \(H100\)](#) [A3 Ultra \(H200\)](#) [A4 \(B200\)](#) [A4X \(GB200\)](#) [TPU v4](#) [TPU v5e](#) [TPU v5p](#) [Trillium \(TPU v6e\)](#) [Ironwood \(7th Gen\)](#)



In which regions are those resources supported?

- **Compute Engine**
 - [GPU regions and zones](#)
 - [TPU regions and zones](#)
- **Vertex AI:**
 - [GPU and TPU regions](#)

How to request higher quota limits?

- **Compute Engine:** [Request a quota adjustment](#)
- **Cloud Run:** [Request a quota increase](#)
- **Vertex AI:** Next page →



How to increase GPU quota limits for Vertex AI?

Vertex AI has GPU quota limits separate from Compute Engine. Please follow the instructions below to get GPU quota limits specifically for Vertex AI:

1. Visit [Cloud Console > APIs & Services > Vertex AI > Quotas](#)
2. Use the **Filter** to search for the GPU type you need. Here are some examples for naming conventions:
 - a. Custom model **training** Nvidia H200 GPUs per region [[Vertex AI Training](#)]
 - b. Custom model **servicing** Nvidia H200 GPUs per region [[Vertex AI Online and Batch Inference](#)]
 - c. Custom model training **preemptible** Nvidia H200 GPUs per region [[Vertex AI Training + DWS-Flex](#)]
 - d. Custom model serving **preemptible** Nvidia H200 GPUs per region [[Vertex AI Inference + DWS-Flex](#)]
3. Select the **checkbox** next to the quota that you want to update and click **Edit quota**

Name	Type	Dimensions (e.g. location)	Value	Current usage percentage	Current usage	Adjustable	
<input checked="" type="checkbox"/> Custom model training Nvidia H200 GPUs per region	Quota	region : africa-south1	0	0%	0	Yes	Edit quota
<input type="checkbox"/> Custom model training Nvidia H200 GPUs per region	Quota	region : asia-east1	0	0%	0	Yes	
<input type="checkbox"/> Custom model training Nvidia H200 GPUs per region	Quota	region : asia-east2	0	0%	0	Yes	

4. Input the **desired GPU quota limit** and provide a workload description for us to review. Click Next.
5. Provide us a **contact information** to keep you updated on the request progress. **Submit request**.

6. **To expedite your request**, let your Account Team know that you requested GPUs for Vertex AI and send them the project-ids to locate your requests.

^ Edit quota

Vertex AI API

Quota: Custom model training Nvidia H200 GPUs per region

Dimensions: region : africa-south1

Current value: 0

Enter a new quota value. A value above 0 will require approval from your service provider. ?

New value *
16

Request description *
A description of your workload

Your description will be sent to your service provider and is used to evaluate your request. It's useful to include the intent of the quota usage, future growth plans, region or zone spread, and any additional requirements or dependencies.

Done

Next

Contact details

These details will be sent to the approvers while reviewing quota change request.

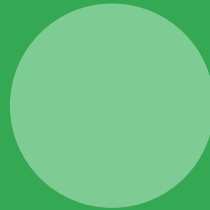
Name *
Your Name

Email *
username@your.domain

Phone

Submit request Back

File a
support case



Support | Navigate to the support cases page

console.cloud.google.com/support/cases

Google Cloud your_domain

Support Cases + GET HELP

Overview

Cases

Billing support

Phone support

Community support

Settings

+ Support cases will be listed here.

TECHNICAL CASES KNOWN ISSUES (10)

These support cases are linked to the "your_domain" organization or any of the projects and resources it contains.

To view cases for a specific project or resource, select a project.

Filter Filter cases

Title	Priority	Status	Project	Case Number	Updated
-------	----------	--------	---------	-------------	---------

Resource: [Create and manage support cases](#)

Support | Fill in the form and submit it

Google Cloud your_domain Search (/) for resources, docs, products, and more Search

Support

Overview

Cases

Billing support

Phone support

Community support

Settings

Help and support

1 Quick details 2 Resources 3 Detailed description

Select your product

Product

Select a product

How would you describe your issue?

Describe your issue

0 / 140

Observed error message

Observed error message

Priority

Priority selection

Next

The following table defines support case priorities. For more information, see [Best practices for working with Support](#)

Priority definition	Example situations
P1: Critical Impact—Service Unusable in Production	<p>The application or infrastructure is unusable in production, having a significant rate of user-facing errors.</p> <p>Business impact is critical (for example, revenue loss or potential data integrity issue).</p> <p>No workaround is available that can be quickly implemented (less than 30 minutes).</p> <p>Affected Google Cloud component or feature is marked as General Availability.</p> <p>Immediate attention from Google is required to resolve the problem.</p>
P2: High Impact—Service Use Severely Impaired	<p>The infrastructure is degraded in production, having a noticeable rate of user-facing errors or difficulties in spinning up a new production system.</p> <p>Business impact is moderate (for example, danger of revenue loss or productivity decrease).</p> <p>A workaround to mitigate critical business impact is available and can be quickly implemented.</p> <p>Affected Google Cloud component or feature is marked as General Availability.</p> <p>Fast response from Google is required.</p>
P3: Medium Impact—Service Use Partially Impaired	<p>The issue is limited in scope and/or severity. The issue has no user-visible impact.</p> <p>Business impact is low (for example, inconvenience or minor business processes affected).</p> <p>Case requires more in-depth investigation and troubleshooting and less frequent interactions.</p>
P4: Low Impact—Service Fully Usable	<p>Little to no business or technical impact.</p> <p>Recommended for consultative tickets where in-depth analysis, troubleshooting or consultancy are preferred to more frequent communications.</p>

Next steps



Google Cloud

Next steps

01.
Let your **Customer Engineer** know about any capacity needs, blockers, or technical questions you may have.
02.
Connect with your **FSR** for any questions about the Google for Startups program.
03.
Keep your **account team** (FSR/CE) in the loop for any launch or milestone updates.

Thank you

Join us as we keep pushing boundaries at
Next 26 on **April 22–24, 2026**
at Mandalay Bay Convention Center in Las Vegas
cloud.withgoogle.com/next

