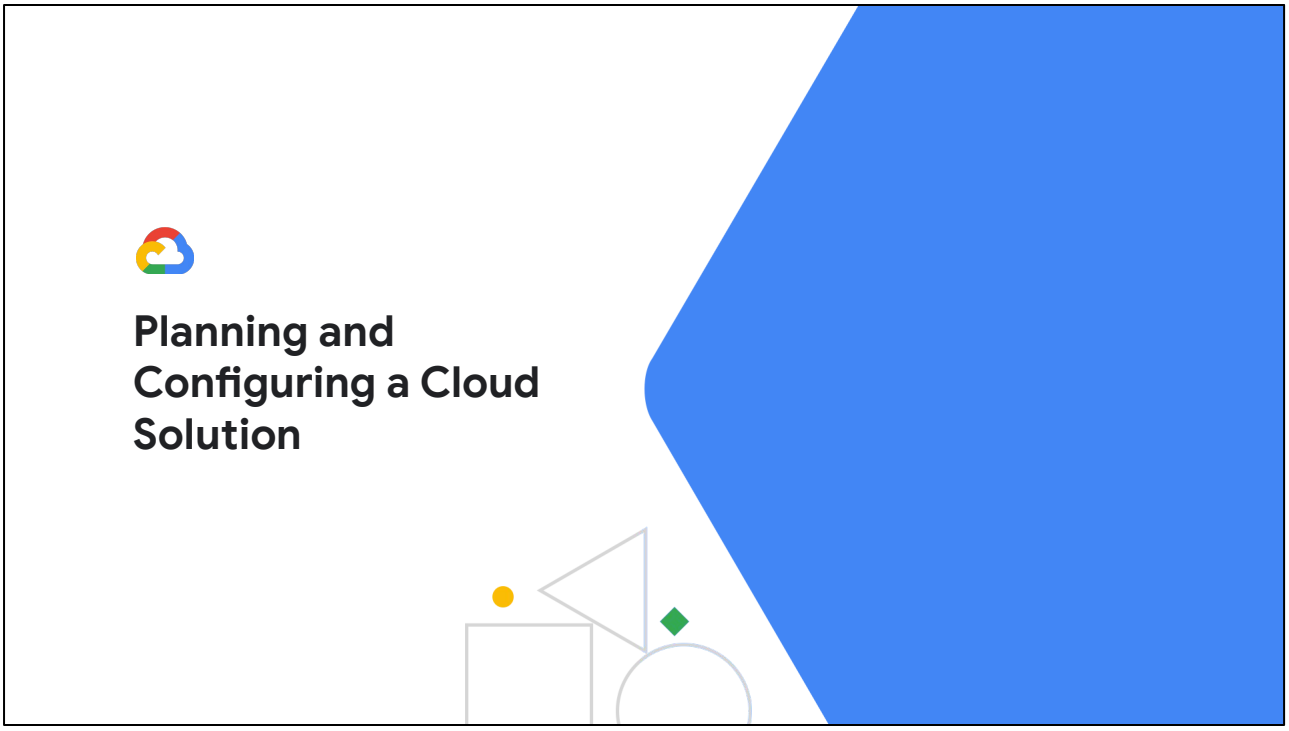

Planning and
Configuring a Cloud
Solution

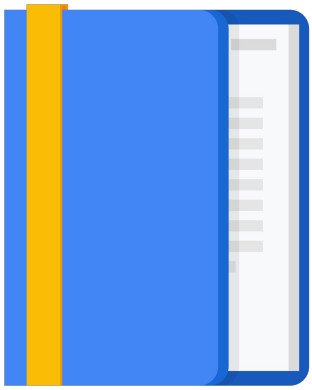
Agenda

Section 2.1 - Planning and estimating using the Pricing Calculator

Section 2.2 - Planning and configuring Compute resources

Section 2.3 - Planning and configuring data storage options

Section 2.4 - Planning and configuring network resources



Google Cloud

Exam Guide: Section 2.1

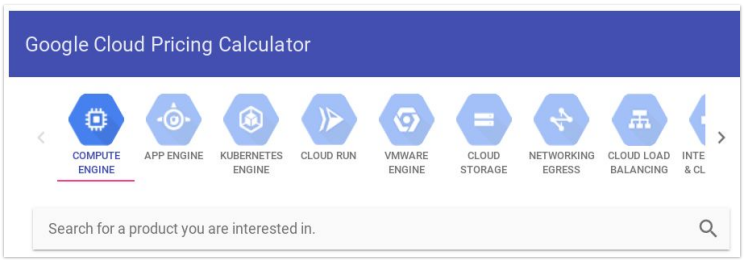
2.1 Planning and estimating Google Cloud product use using the Pricing Calculator.

Google Cloud

One of the first stages in setting up a new project is budgeting.
Because there can be quite a number of variables that go into pricing a particular Cloud product, having a way to pull all of that information together into a unified report - without having to actually configure that product first - is very helpful. This is where Google's Pricing Calculator comes into play.

Google Cloud Pricing Calculator

- Select a product from the scrolling list at the top of the form.
- The variables for that product will then be shown in the form below.
- Fill out the form with your target configuration.
- Submit each section filled out to add it to your overall estimate.



https://cloud.google.com/products/calculator/

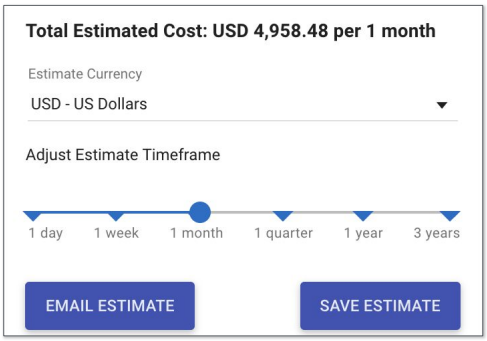
Google Cloud

The pricing calculator is a multi-section form.
Once you have some possible configurations in mind, you can use it to estimate costs for the different products you will be using,

Google Cloud Pricing Calculator

The Pricing Calculator gives you a total estimated cost that you can view in daily, weekly, monthly, quarterly, yearly and 3-year increments.

This cost is only an estimate, however. Total costs may differ depending on how closely your estimated usage matches your actual usage. The estimate is also not a binding contract, it is just a planning tool.



Google Cloud

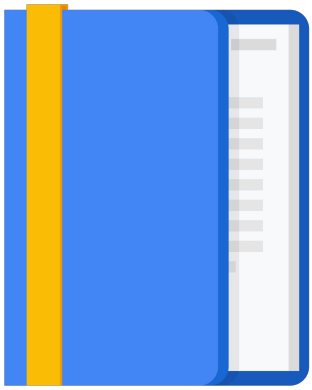
Agenda

Section 2.1 - Planning and estimating using the Pricing Calculator

Section 2.2 - Planning and configuring Compute resources

Section 2.3 - Planning and configuring data storage options

Section 2.4 - Planning and configuring network resources



Google Cloud

Exam Guide: Section 2.2

2.2 Planning and configuring compute resources.

Considerations include:

- Selecting appropriate compute choices for a given workload (e.g., Compute Engine, Google Kubernetes Engine, and App Engine).
- Using preemptible VMs and custom machine types as appropriate.

Exam Guide: Section 2.2

2.2 Planning and configuring compute resources.

Considerations include:

- **Selecting appropriate compute choices for a given workload (e.g., Compute Engine, Google Kubernetes Engine, and App Engine).**
- Using preemptible VMs and custom machine types as appropriate.

Compute options and use cases

Option	Use when you need...	Typical use cases
Google App Engine Flexible, zero-ops platform for building apps.	• To just focus on writing code. • Developer velocity. • To minimize operational overhead.	• Web sites • Apps (of course!) • Gaming back ends • IoT applications
Google Compute Engine Virtual machines running in Google's global data centers.	• Complete control. • Ability to make OS level changes. • To be able to move to the cloud without rewriting your code. • To use custom VM images.	• Any workload requiring a specific OS or configuration. • On-premises software that you want to run in the cloud.
Google Kubernetes Engine Logical infrastructure powered by Kubernetes, the open source container orchestration system.	• No dependencies on a specific OS. • Increased velocity and operability. • To manage containers in production.	• Containerized workloads. • Cloud-native distributed systems. • Hybrid applications.

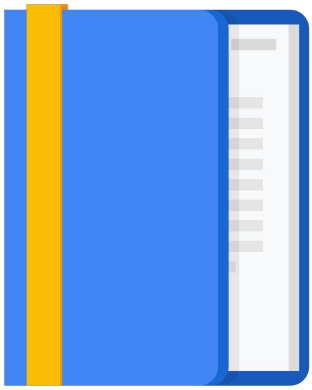
Agenda

Section 2.1 - Planning and estimating using the Pricing Calculator

Section 2.2 - Planning and configuring Compute resources

Section 2.3 - Planning and configuring data storage options

Section 2.4 - Planning and configuring network resources



Google Cloud

Exam Guide: Section 2.3

2.3 Planning and configuring data storage options.

Considerations include:

- Product choice (e.g., Cloud SQL, BigQuery, Cloud Spanner, Cloud Bigtable)
- Choosing storage options (e.g., Standard, Nearline, Coldline, Archive)

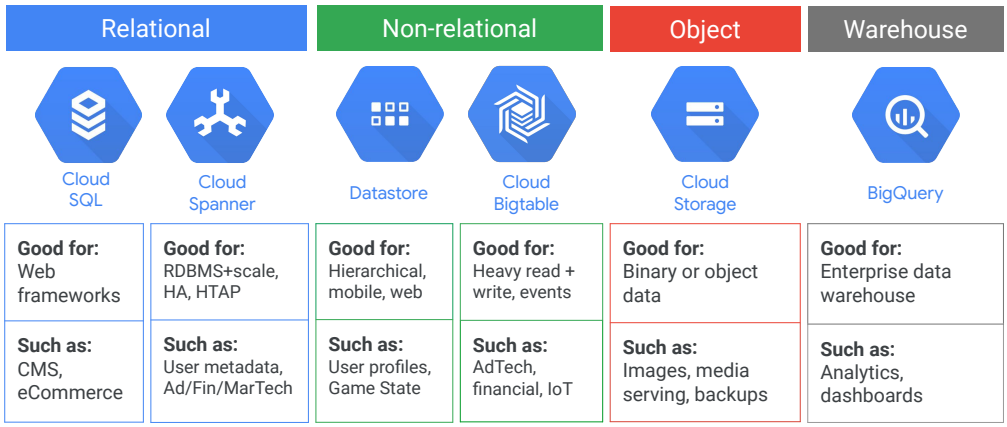
Exam Guide: Section 2.3

2.3 Planning and configuring data storage options.

Considerations include:

- **Product choice (e.g., Cloud SQL, BigQuery, Cloud Spanner, Cloud Bigtable)**
- Choosing storage options (e.g., Standard, Nearline, Coldline, Archive)

Comparing data storage and database options



Data storage options and use cases

Option	Use when you need...	Typical use cases
Cloud SQL	Fully managed MySQL and PostgreSQL database service.	• Web frameworks • Structured data • OLTP workloads
BigQuery	A scalable, fully managed enterprise data warehouse (EDW) with SQL and fast ad-hoc queries.	• OLAP workloads up to petabyte scale • Big data exploration and processing
Cloud Spanner	Mission-critical, relational database service with transactional consistency, global scale, and high availability.	• Adtech • Financial services • Global supply chain • Retail
Cloud Bigtable	A scalable, fully managed NoSQL wide-column database that is suitable for both low-latency single-point lookups and precalculated analytics.	• IoT, finance, adtech • Monitoring • Geospatial datasets • Graphs

Exam Guide: Section 2.3

2.3 Planning and configuring data storage options.

Considerations include:

- Product choice (e.g., Cloud SQL, BigQuery, Cloud Spanner, Cloud Bigtable)
- **Choosing storage options (e.g., Standard, Nearline, Coldline, Archive)**

Choosing among Cloud Storage classes

Storage Class	Minimum duration	Availability SLA	Typical monthly availability	Use cases	Name for APIs and gsutil
Standard Storage	None	Multi-region 99.95% Dual-region 99.95% Region 99.9%	>99.99% availability in multi-regions and dual-regions; 99.99% in regions	Access data frequently ("hot" data) and/or store for brief periods • Serve website content • Stream videos • Interactive workloads • Mobile and gaming apps	STANDARD
Nearline Storage	30 days	Multi-region 99.9% Dual-region 99.9% Region 99.0%	99.95% availability in multi-regions and dual-regions; 99.9% in regions	Read/modify data ≤ once per month • Data backup • Serve long-tail multimedia content	NEARLINE
Coldline Storage	90 days			Read/modify data no more than once a quarter	COLDLINE
Archive Storage	365 days	None		Read/modify data < once a year • Cold data storage • Disaster recovery	ARCHIVE

Google Cloud

Google Cloud Storage has four primary storage classes, with different characteristics, use cases, and prices for your needs.

Standard Storage is best for data that is frequently accessed ("hot" data) and/or stored for only brief periods of time. When used in a region, co-locating your resources maximizes the performance for data-intensive computations and can reduce network charges. When used in a dual-region, you still get optimized performance when accessing Google Cloud products that are located in one of the associated regions, but you also get the improved availability that comes from storing data in geographically separate locations. When used in a multi-region, Standard Storage is appropriate for storing data that is accessed around the world, such as serving website content, streaming videos, executing interactive workloads, or serving data supporting mobile and gaming applications.

Nearline Storage is a low-cost, highly durable storage service for storing infrequently accessed data. Nearline Storage is a better choice than Standard Storage in scenarios where slightly lower availability, a 30-day minimum storage duration, and costs for data access are acceptable trade-offs for lowered at-rest storage costs. Nearline Storage is ideal for data you plan to read or modify on average once per month or less. Nearline Storage is appropriate for data backup, long-tail multimedia content, and data archiving.

Coldline Storage is a very-low-cost, highly durable storage service for storing infrequently accessed data. Coldline Storage is a better choice than Standard Storage

or Nearline Storage in scenarios where slightly lower availability, a 90-day minimum storage duration, and higher costs for data access are acceptable trade-offs for lowered at-rest storage costs. Coldline Storage is ideal for data you plan to read or modify at most once a quarter.

Archive Storage is the lowest-cost, highly durable storage service for data archiving, online backup, and disaster recovery. Archive Storage has higher costs for data access and operations, as well as a 365-day minimum storage duration. Archive Storage is the best choice for data that you plan to access less than once a year. For example, cold data storage, such as data stored for legal or regulatory reasons, and disaster recovery.

For more information, see: https://cloud.google.com/storage/docs/storage-classes

Characteristics applicable to all storage classes

- Unlimited storage with no minimum object size.
- Worldwide accessibility and worldwide storage locations.
- Can be used in any multi-region, dual-region, or region.
- Geo-redundancy if the data is stored in a multi-region or dual-region.
- Low latency (time to first byte typically tens of milliseconds).
- High durability (99.999999999% annual durability).
- A uniform experience with Cloud Storage features, security, tools, and APIs.

Google Cloud

We've discussed the four primary storage classes and differentiated between them in terms of characteristics, availability and use cases. It is worth noting that there are a number of characteristics that apply across all storage classes. These include:

- Unlimited storage with no minimum object size requirement,
- Worldwide accessibility and locations,
- Can be used in any multi-region, dual-region, or region,
- Geo-redundancy if data is stored in a multi-region or dual-region,
- Low latency and high durability, and
- A uniform experience, which extends to security, tools, and APIs.

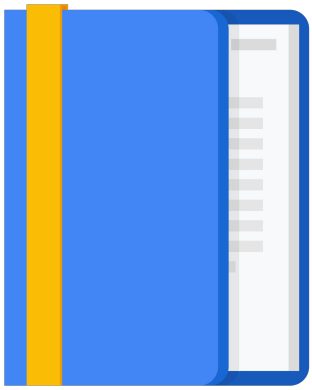
Agenda

Section 2.1 - Planning and estimating using the Pricing Calculator

Section 2.2 - Planning and configuring Compute resources

Section 2.3 - Planning and configuring data storage options

Section 2.4 - Planning and configuring network resources



Google Cloud

Exam Guide: Section 2.4

2.4 Planning and configuring network resources.

Tasks include:

- Differentiating load balancing options.
- Identifying resource locations in a network for availability.
- Configuring Cloud DNS.

Google Cloud

Exam Guide: Section 2.4

2.4 Planning and configuring network resources.

Tasks include:

- **Differentiating load balancing options.**
- Identifying resource locations in a network for availability.
- Configuring Cloud DNS.

Google VPC offers a suite of load-balancing options

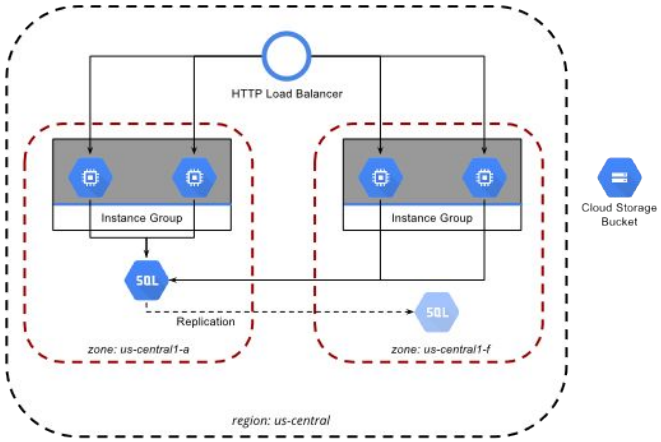
Global HTTP(S)	Global SSL Proxy	Global TCP Proxy	Regional	Regional internal
Layer 7 load balancing based on load.	Layer 4 load balancing of non-HTTPS SSL traffic based on load.	Layer 4 load balancing of non-SSL TCP traffic.	Load balancing of any traffic (TCP, UDP).	Load balancing of traffic inside a 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.

If you need cross-regional load balancing for a Web application, use HTTP(S) load balancing. For Secure Sockets Layer traffic that is not HTTP, use the Global SSL Proxy load balancer. If it's other TCP traffic that does not use Secure Sockets Layer, use the Global TCP Proxy load balancer.

Those two proxy services only work for specific port numbers, and they only work for TCP. If you want to load balance UDP traffic, or traffic on any port number, you can still load balance across a Google Cloud region with the Regional load balancer.

Finally, what all those services have in common is that they're intended for traffic coming into the Google network from the Internet. But what if you want to load balance traffic inside your project, say, between the presentation layer and the business layer of your application? For that, use the Internal load balancer. It accepts traffic on a Google Cloud internal IP address and load balances it across Compute Engine VMs.

Load balancing overview



Google Cloud

Load balancing is when two or more identical servers or server clusters have been created so that if the load becomes too great, or if one or more servers should fail, then the remainder can assist with or take over handling the load. This is one way to create applications and services that are “highly available.”

Load balancing allows multiple servers (or clusters of servers) to function as a single computing resource. Load balancers can also be configured to add, or remove, these servers or server clusters from the system to better meet demand. This is known as “autoscaling.”

Deciding on load balancing options

Cloud load balancer considerations can be divided up as follows:

- Global versus regional load balancing
- External versus internal load balancing
- Traffic type

The slides that follow describe the use cases for different types of load balancers.

Deciding on load balancing options

Load balancer	Traffic type	Global/Regional	External/Internal	External Ports for Load Balancing
HTTP(S)	HTTP or HTTPS	Global	External	HTTP on 80 or 8080; HTTPS on 443
SSL Proxy	TCP with SSL offload	Global	External	25, 43, 110, 143, 195, 443, 465, 587, 700, 993, 995, 1883, and 5222
TCP Proxy	TCP without SSL offload. Does not preserve client IP addresses	Global	External	25, 43, 110, 143, 195, 443, 465, 587, 700, 993, 995, 1883, 5222
Network TCP/UDP	TCP/UDP without SSL offload. Preserves client IP addresses.	Regional	External	Any
Internal TCP/UDP	TCP or UDP	Regional	Internal	Any

Google Cloud

Google Cloud's load balancers can be divided into external and internal load balancers.

External load balancers distribute traffic coming from the internet to your Google Cloud network.

Internal load balancers distribute traffic within your Google Cloud network.

Use global load balancing when your users and instances are globally distributed, your users need access to the same applications and content, and you want to provide access using a single anycast IP address. Global load balancing can also provide IPv6 termination.

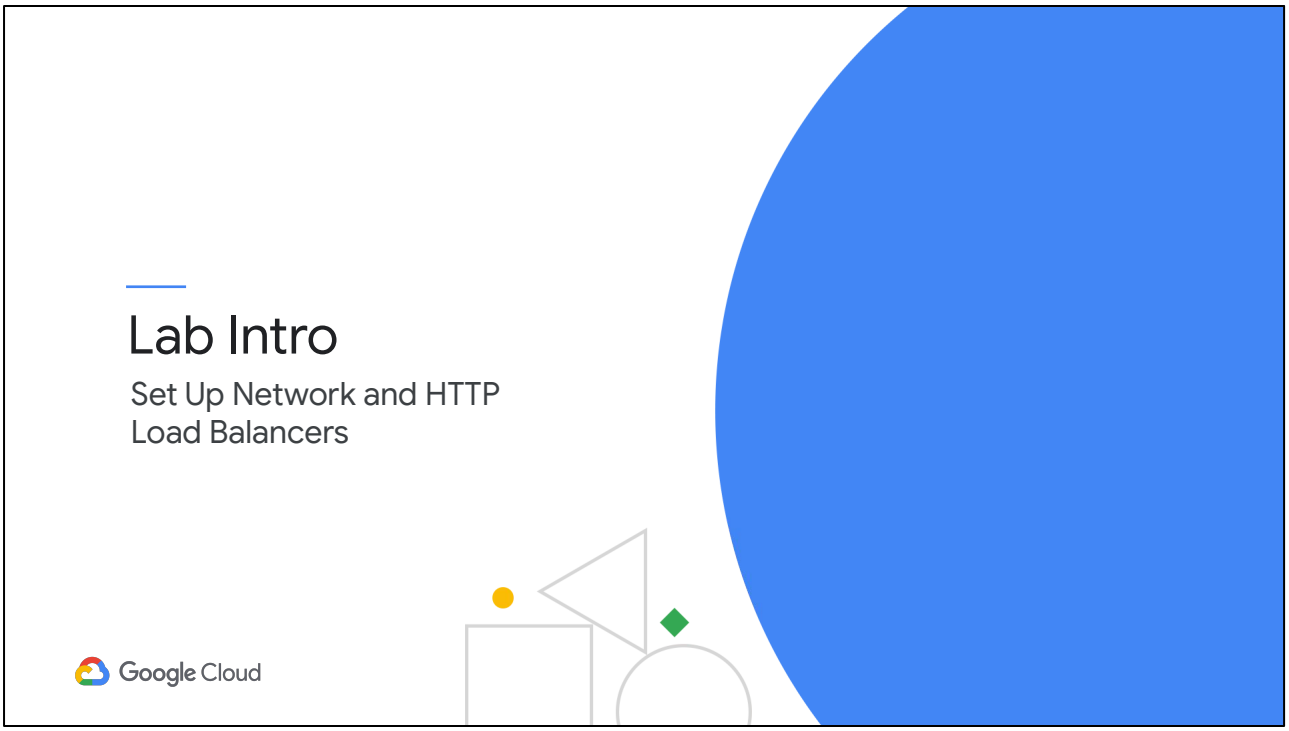
Use regional load balancing when your users and instances are concentrated in one region and you only require IPv4 termination.

Global load balancing requires that you use the Premium Tier of Network Service Tiers. For regional load balancing, you can use Standard Tier.

Traffic type is a deciding factor in choosing a load balancer

The type of traffic you need your load balancer to handle is another factor in determining which load balancer to use.

- HTTP and HTTPS traffic require global, external load balancing.
- TCP traffic can be handled by global, external load balancing; external, regional load balancing; or internal, regional load balancing.
- UDP traffic can be handled by external regional load balancing or internal regional load balancing.



In this hands-on lab, you'll learn the differences between a network load balancer and a HTTP load balancer, and how to set them up for your applications running on Google Compute Engine virtual machines.

This lab is part of the Qwiklabs Cloud Architecture Quest.

Suggested study resources for this section

Google Cloud pricing overview: https://cloud.google.com/pricing/

Google Cloud Pricing Calculator: https://cloud.google.com/products/calculator/

Compute Engine documentation: https://cloud.google.com/compute/docs/

Choosing the right compute option in Google Cloud: https://cloud.google.com/blog/products/gcp/choosing-the-right-compute-option-in-gcp-a-decision-tree

Choosing an application hosting option: https://cloud.google.com/hosting-options

Storage classes: https://cloud.google.com/storage/docs/storage-classes

Cloud Storage products: https://cloud.google.com/products/storage

Load Balancing: https://cloud.google.com/load-balancing/docs/load-balancing-overview


Google Cloud