searce | SYNOPSYS®

Hackathon-Workshop

searce | SYNOPSYS®

Agenda

1. Introduction to Google Cloud
2. Identity Access management (IAM)
3. Resource management
4. Networking
5. Google Compute Engine - VM
6. Google Cloud Storage - GCS
7. Bigquery
8. Simple Docker application deployment on VM
9. Pub/sub
10. Google Kubernetes Engine - GKE

searce | SYNOPSYS®

Introduction to Google Cloud

Region & Zones Terminology

Region:

- Regions are **independent geographic areas** that consist of zones.
- Locations within regions tend to have round-trip network latencies of under <1ms on the 95th percentile.

Zones:

- Is a deployment area for Google Cloud resources within a region.
- Each Region consists of **3-4 zones**

Note: Google Cloud services and resources can be either zonal, regional, or managed by Google across multiple regions.

searce | SYNOPSYS®

GCP Global Cloud Locations

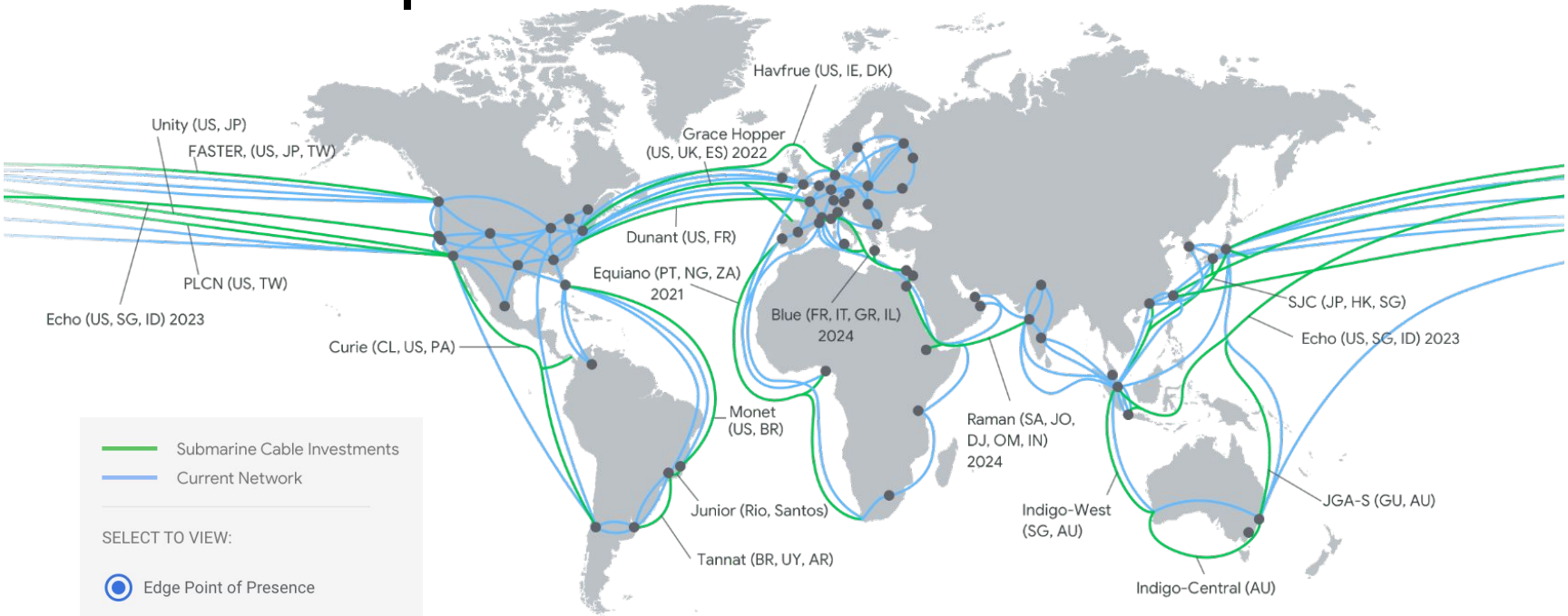
COMING SOON! Google Cloud will continue expanding into the following regions: Doha (Qatar), Turin (Italy), Berlin (Germany), Dammam (Kingdom of Saudi Arabia), Mexico, Malaysia, Thailand, New Zealand, Greece, Norway, Austria and Sweden.

searce | SYNOPSYS®

Region Map

searce | SYNOPSYS®

Network Map



searce | SYNOPSYS®

Identity Access management (IAM)

searce | SYNOPSYS®

Controlling access

Authentication

Cloud Identity

Authorization

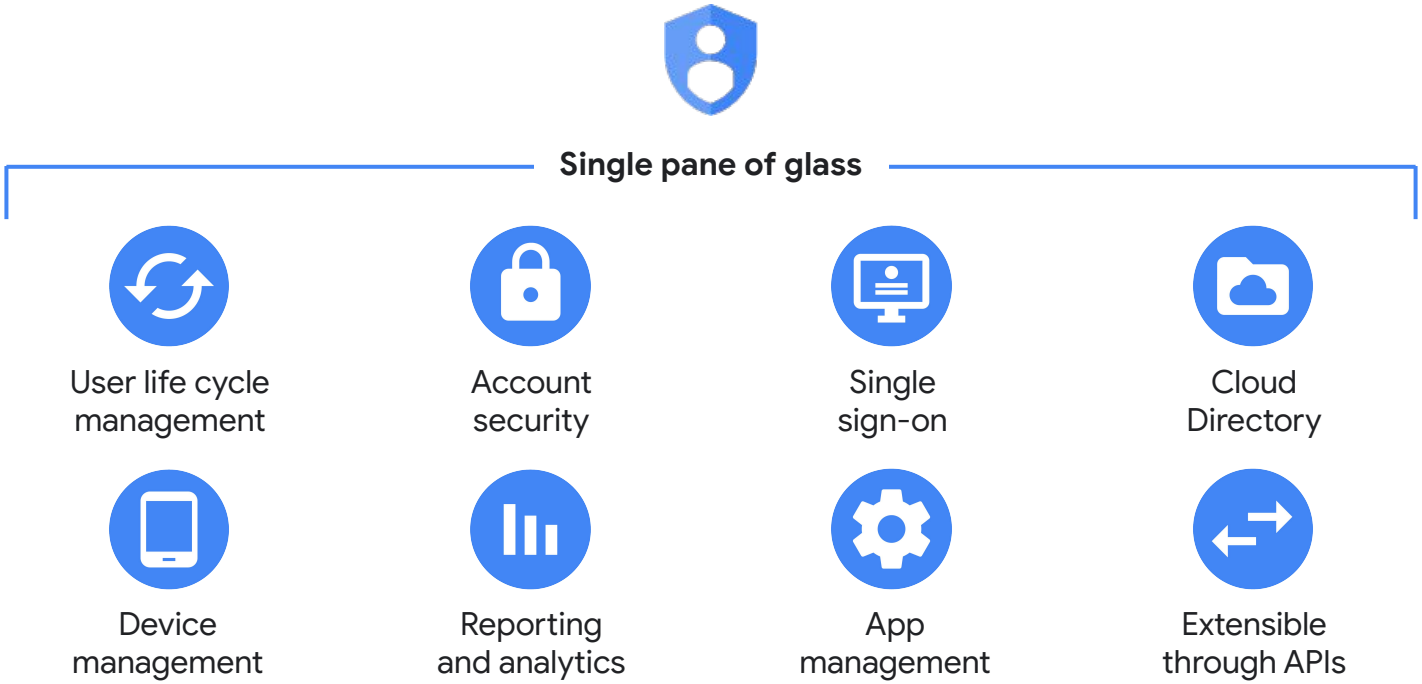
Identity Access Management (IAM)

What is Cloud Identity?

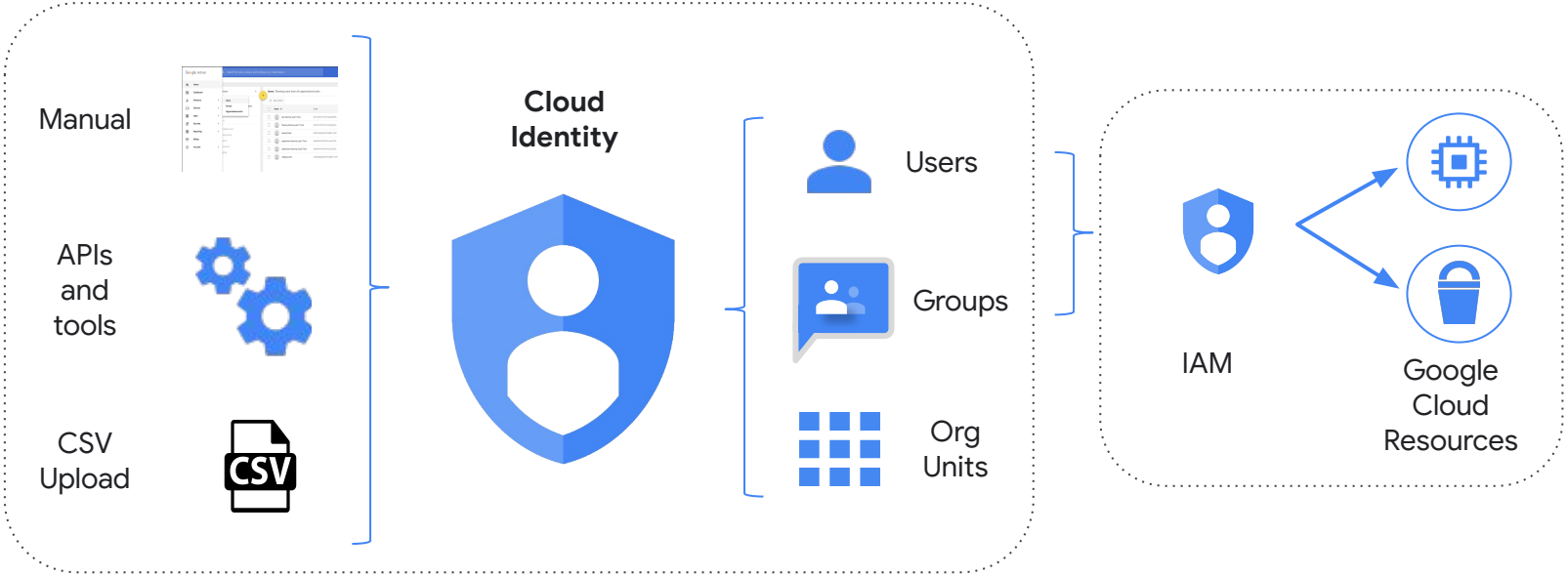
- Cloud Identity is an Identity as a Service (IDaaS) solution that **allows you to centrally manage users and groups** who can access Google Cloud and Google Workspace resources
- It is the same identity service that powers Google Workspace and can also be used as IdP for third-party applications (supports SAML and LDAP applications)

searce | SYNOPSYS®

Cloud Identity provides



Users and groups

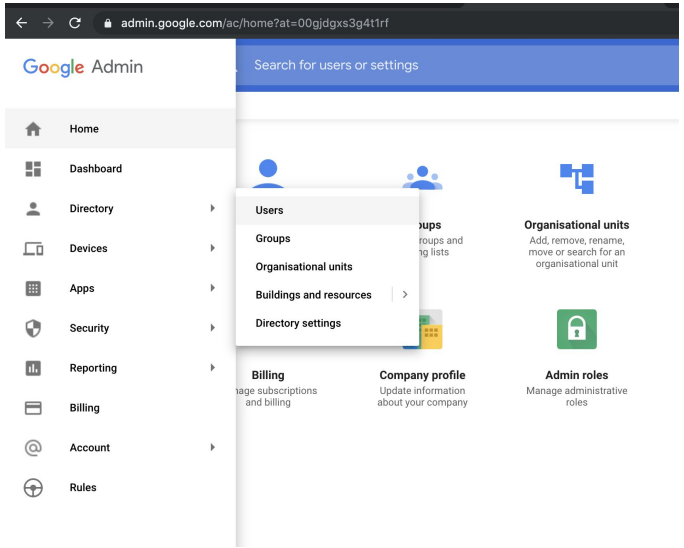


Users and groups created in Cloud Identity are the **Google Identities** that *can* be assigned **IAM roles** in the Google Cloud console.

The **Cloud Identity roles** only manage aspects of Cloud Identity such as user/group management, **and are different from Google Cloud roles,** which manage permissions to cloud resources.

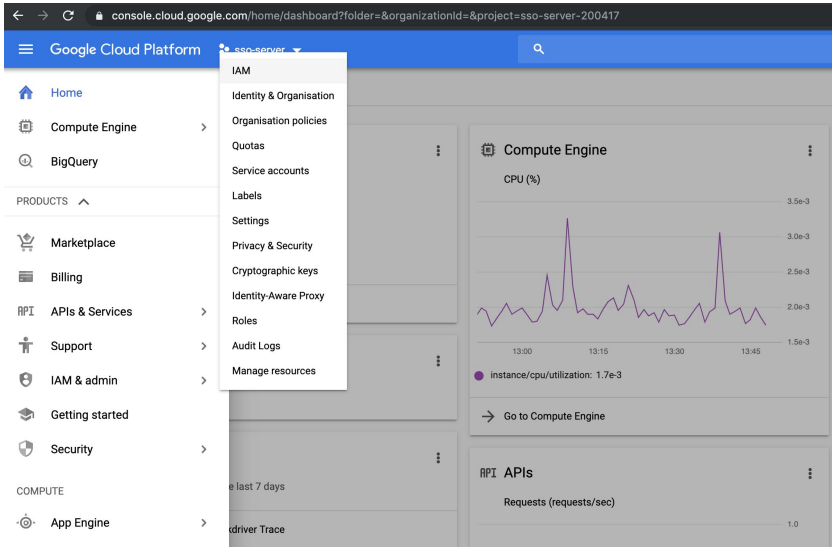
searce | SYNOPSYS®

Two consoles for administration



Cloud Identity (`admin.google.com`)

Managing Users, Groups, and Authentication settings



Google Cloud (`console.cloud.google.com`)

Roles & Authorization for Google Cloud

searce | SYNOPSYS®

Administration and management

Cloud Identity

Admin console, Admin SDK, GAM

- Creating and managing user accounts
- Creating and managing groups
- Assigning access and identity administration roles for users
- Enforcing authentication options for users

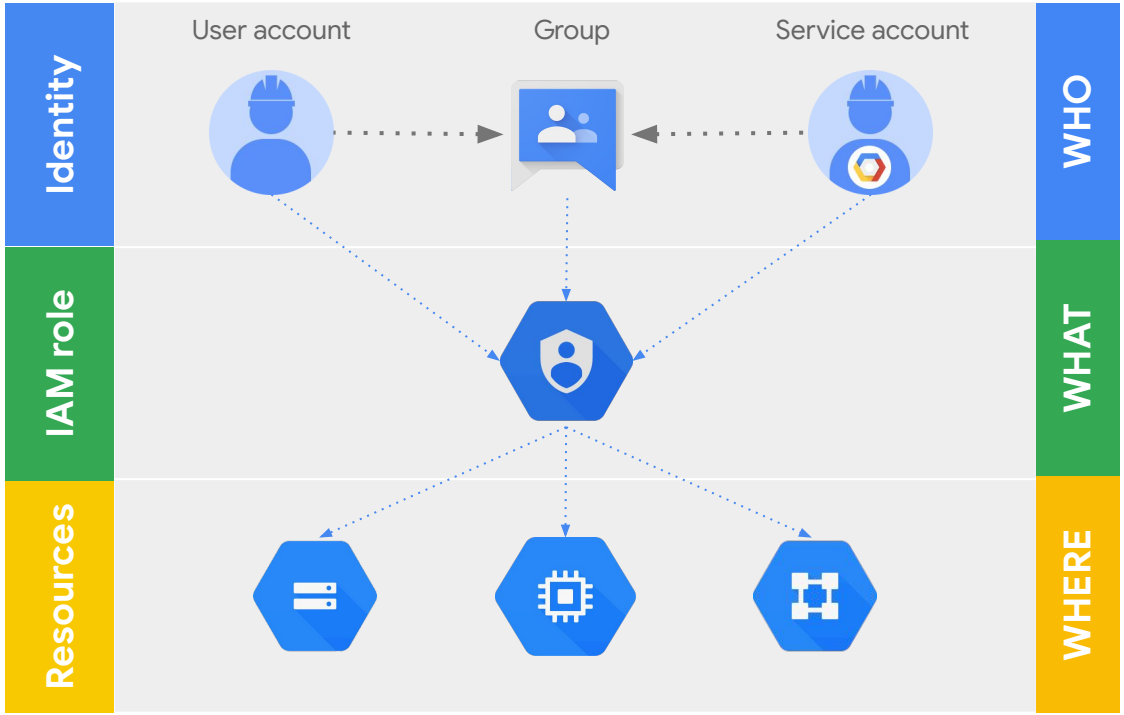
Google Cloud

Console, gcloud CLI, Terraform, API

- Provisioning Cloud Platform resources
- Assigning access and identity management roles for Cloud platform resources to users and groups set in Cloud Identity
- Configuring networking and on-premises integration

searce | SYNOPSYS®

Identity, roles, and resources



searce | SYNOPSYS®

Two types of identity

	User accounts	Service accounts	Groups	Domain
Identity	Human	Robot	Both	Cloud Identity accounts
Authentication	Google password or SSO	Keys	Via user	Via user
Authorization	IAM roles	IAM roles	IAM roles	IAM roles
Management	Admin console	Cloud console	Admin console	Admin console

IAM policy

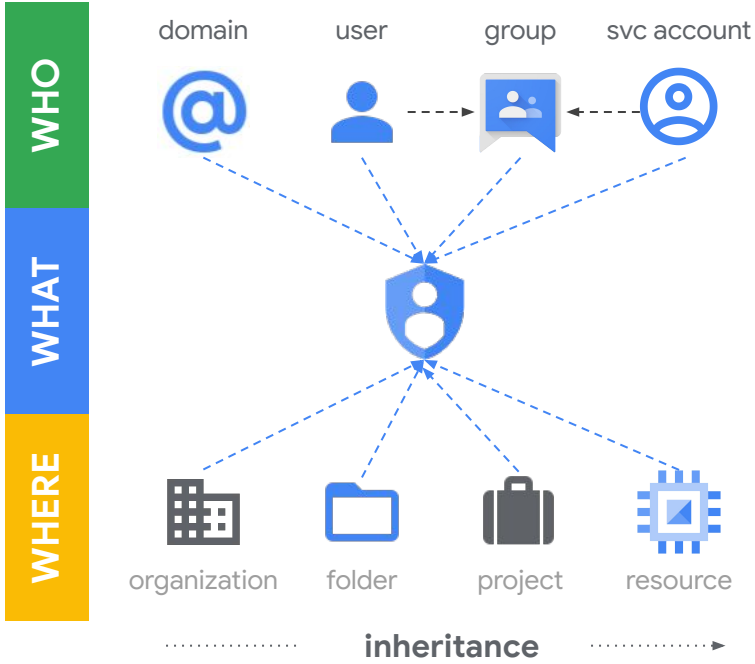
Who can do what and where on Google Cloud.

IAM policies **manage access control for Google Cloud resources**. They are collections of **IAM bindings**, each one "binding" together a principal, a role, and the resource to which the policy is attached.

Binding **principles** can be

- an org domain, granting the role to all org members
- a Google Workspace/Cloud Identity user
- a service account (described later in this deck)
- a Google Workspace/Cloud Identity group

What is commonly thought of as an authorization group is, on Google Cloud, an IAM binding: the union of an identity group and a role, bound to a specific resource or hierarchy node.



searce | SYNOPSYS®

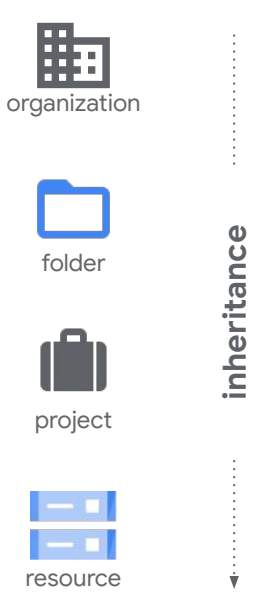
IAM role

Groups a set of related fine-grained permissions.

roles/storage.objectAdmin

- resourcemanager.projects.get
- resourcemanager.projects.list
- storage.objects.create
- storage.objects.delete
- storage.objects.get
- storage.objects.getIamPolicy
- storage.objects.list
- storage.objects.setIamPolicy
- storage.objects.update

can apply to



searce | SYNOPSYS®

IAM role types

Basic, predefined, and custom roles available.

- `roles/browser`
- `roles/viewer`
- `roles/editor`
- `roles/owner`

- `roles/storage.admin`
- `roles/storage.objectAdmin`
- `roles/storage.objectCreator`
- `roles/storage.objectViewer`
- `...`

- `[define your own]`

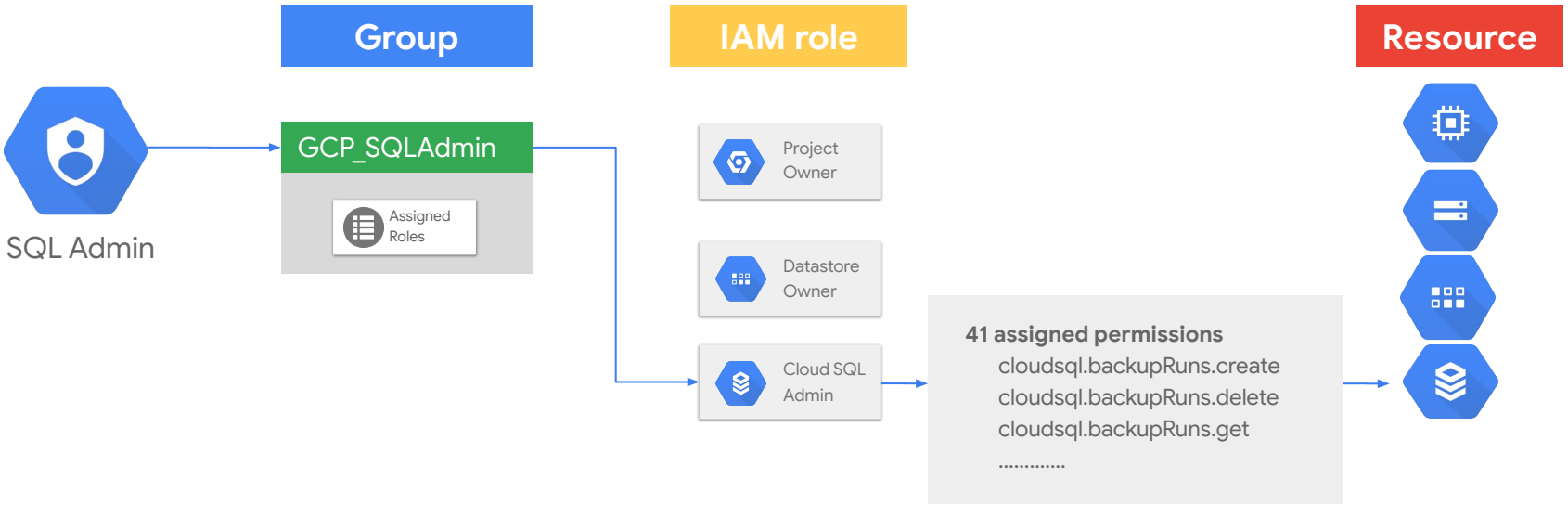
- **Broad permissions scope**, concentric (*for example, owner includes editor permissions)*
- **Easy** to understand and apply
- Often include **unwanted permissions** *(for example, owner has broad IAM grant permissions, editor/viewer have billing report access*

- Narrower **per-service permissions scope**
- Admin roles usually have IAM grant permissions at the service resource level
- More effort but **safer** than primitive roles
- Maps well to the model of "which services users are allowed to use"

- **Arbitrarily defined permissions scope**
- Can be defined at org or project level
- Not all permissions available
- Limit in number of roles and permissions per role
- **Substantial maintenance effort** managing dependencies and updates

searce | SYNOPSYS®

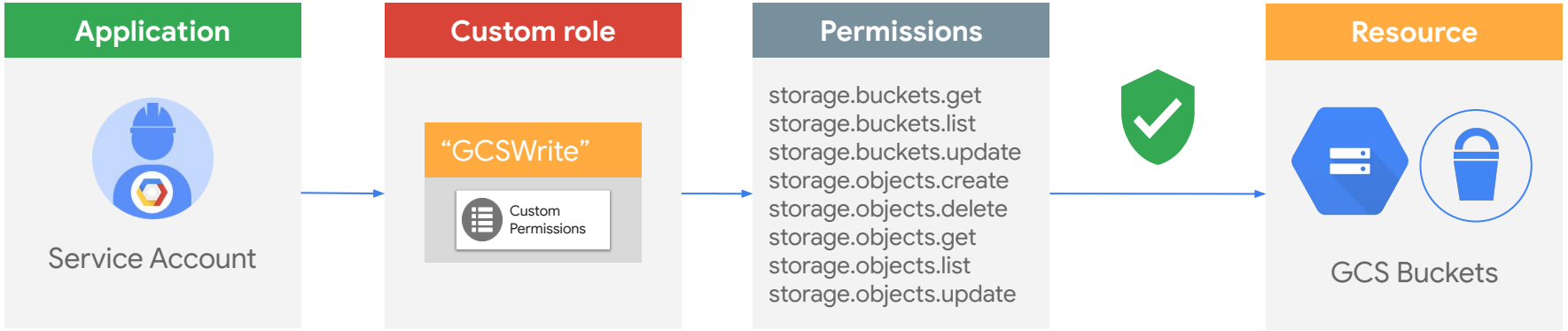
IAM - predefined roles



searce | SYNOPSYS®

IAM - custom roles

Custom IAM roles enable administrators to create roles based on specific permissions associated with a resource to ensure that users and service accounts have the precise level of access they require.

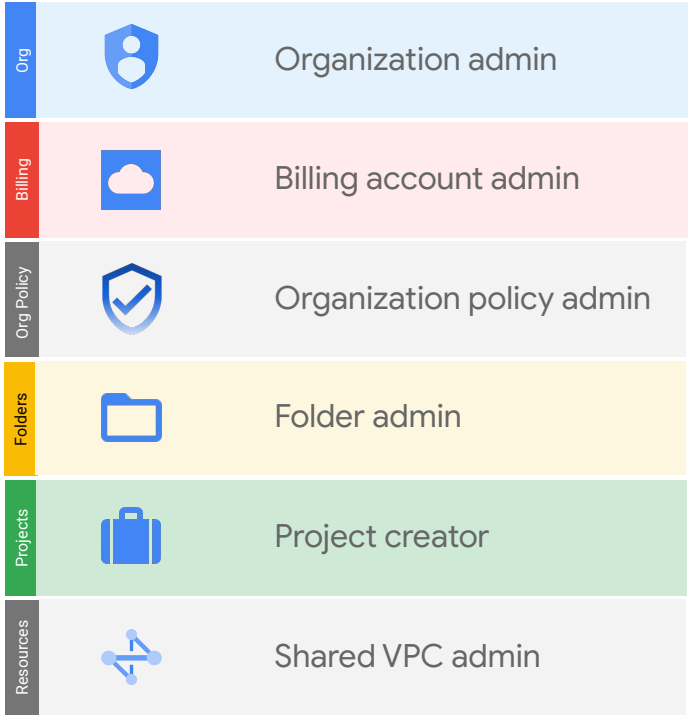


searce | SYNOPSYS®

Typical organization-level roles

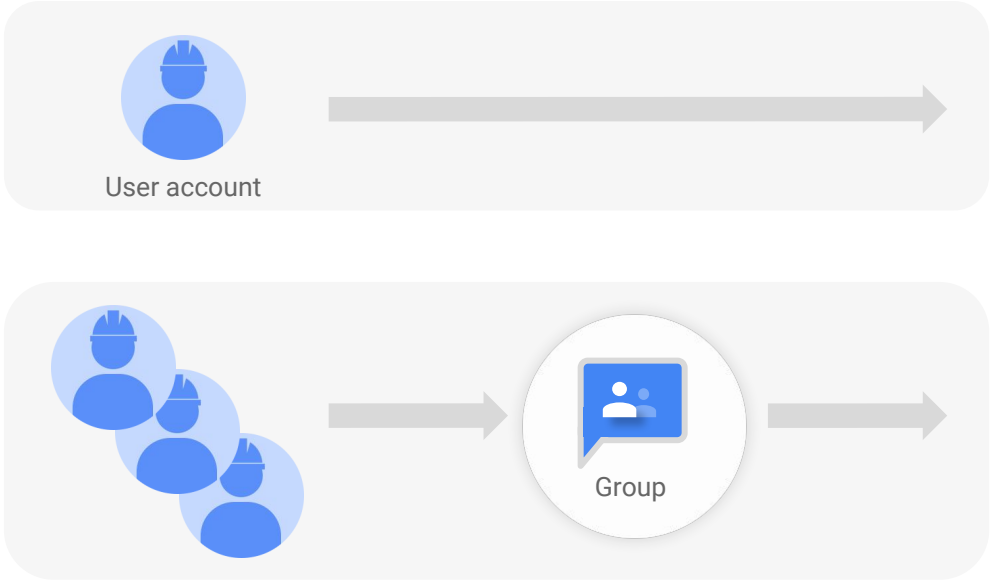
Some of the roles typically assigned at the organization level.

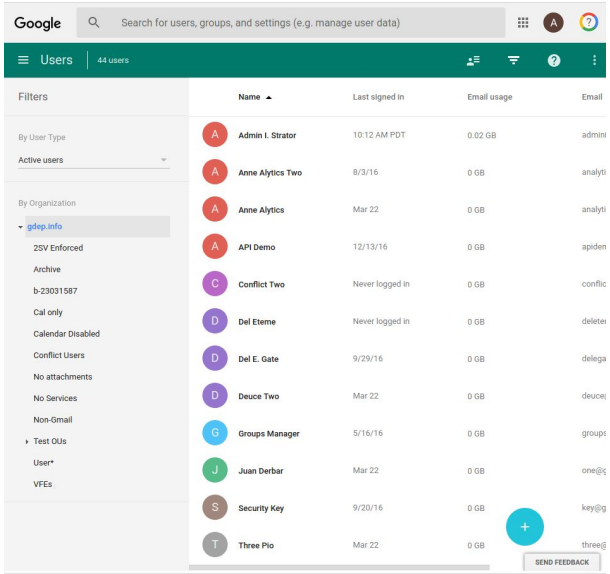
- **Organization admin**
  Set IAM policies at all levels, assigns roles to other users.
- **Billing account admin**
  Full access to manage billing accounts and project billing assignments.
- **Organization policy admin**
  Set or override organization policies.
- **Folder admin**
  Manage folders, set IAM permissions on folders.
- **Project creator**
  Create projects, set IAM permissions on projects.*
- **Shared VPC admin**
  Configure shared VPC at the project level.



searce | SYNOPSYS®

Users and groups





searce | SYNOPSYS®

IAM conditions

Attribute-based access control for IAM policies.

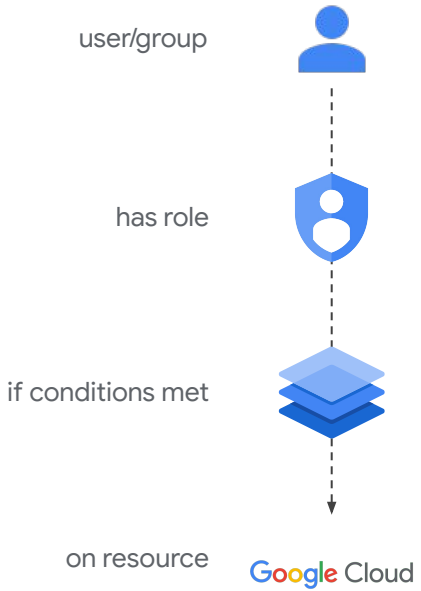
Grant resource access only if configured conditions are met, by restricting role grants based on dynamic attributes.*

Resource attributes, examples:

- Only grant role on specific resource types**
  `resource.type == "compute.googleapis.com/Instance"`
- Or specific sets of resources by name/prefix
  `resource.name.startsWith("projects/_/buckets/myco-team-a-")`

Request attributes, examples:

- Only grant role at specific times
  `request.time.getHours("Europe/Berlin") >= 9 &&`
  `request.time.getHours("Europe/Berlin") <= 17 &&`
- Or specific IP/ports (for IAP tunnels)
  `destination.port == 22`



*** not supported for basic roles*
*** also possible via custom role*

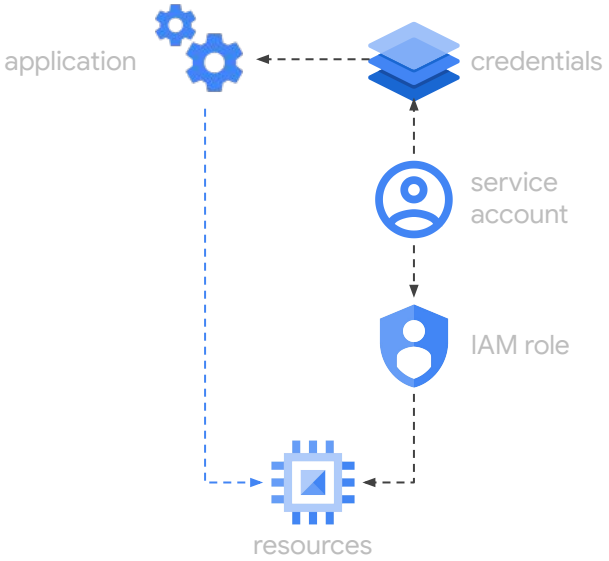
searce | SYNOPSYS®

Service accounts

Special types of accounts used by applications and services.

Non-human access to Google Cloud APIs and services is usually done via service accounts.

- Created and managed **within projects** like most other resources.
- **No associated password**, cannot log in via browser or cookies (no console access).
- Authentication is done via **private/public key pairs**, either Google or customer-managed, or identity federation.
- Included in org for domain-restricted sharing org policy, excluded for roles assigned to org domain.
- Can be **impersonated** by either regular users, or other service accounts (via IAM roles).



searce | SYNOPSYS®

Service account types

Some types of service accounts are built-in to Google Cloud services.

User managed

Created by customers and managed like all other resources.

- Arbitrary number, naming, and usage
- No IAM role assigned by default
- Use via key, VM association, impersonation

- CI/CD service accounts
- automation service accounts
- per-app service accounts

Service default

Created at API activation, used by default when no customer SA is selected.

- Fixed naming convention
- Editor IAM role assigned at creation
- IAM can be limited via organization policy

- Compute Engine default service account for VMs
- AppEngine default service account

Google managed (robots)

Created at API activation, used by Google Cloud services to perform actions on customer resources.

- Fixed naming convention
- Specific IAM roles assigned at creation
- Additional roles may be needed to enable extra functionality (like CMEK)

- Google Kubernetes Engine (GKE) robot account
- Compute Engine robot account
- Cloud Storage robot account

Resource management

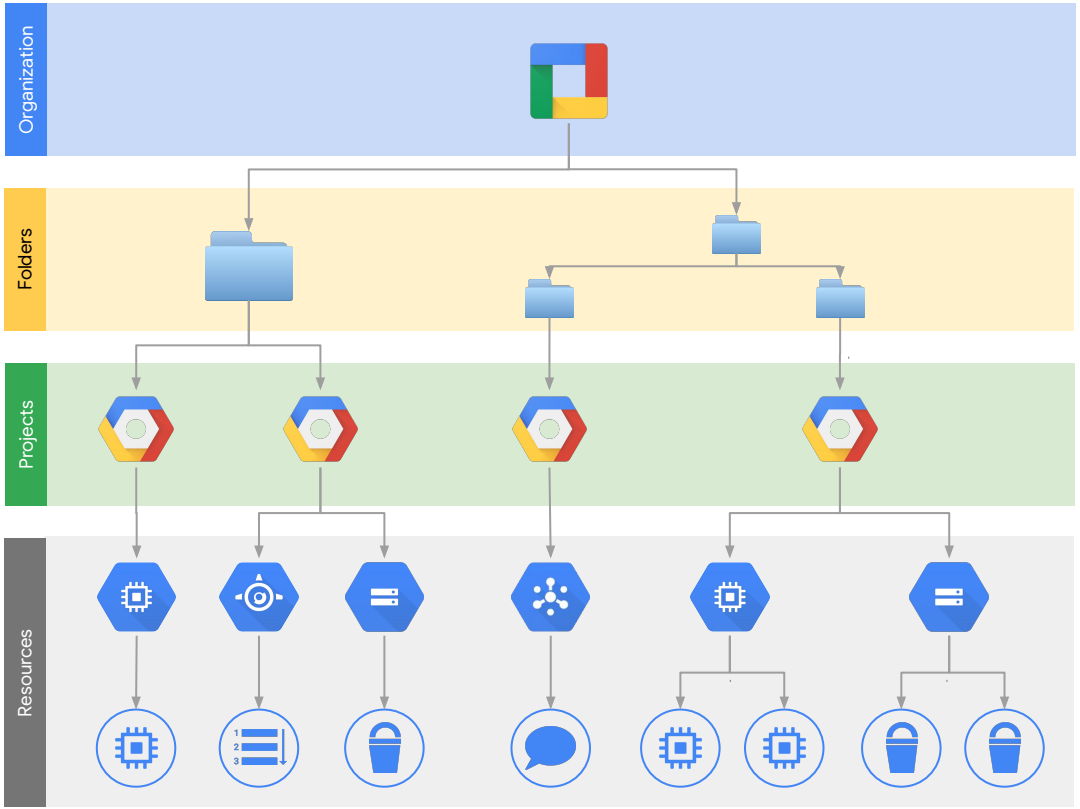
Cloud Resource Manager: Organization hierarchy

searce | SYNOPSYS®

IAM

Top-down inheritance

Additive only



Policies

Top-down inheritance

Allows overrides

What is an organization?

searce | SYNOPSYS®

- The **Organization** resource is the root node in the Google Cloud Platform resource hierarchy and the hierarchical super-node of projects.
- It allows for the following controls to be set across an entire collection of projects:
  - Defining IAM policies
  - Determining the initial folder structure of the resource hierarchy
  - Delegating responsibility over critical components such as networking, billing, resource hierarchy through IAM roles
- An Organization has a single **Cloud Identity directory** containing users and groups associated with the organization.

What are folders?

searce | SYNOPSYS®

- **Folders** are nodes in the Cloud Platform organization hierarchy. A folder can contain projects, folders up to four levels deep, or a combination of both.
- **Folders** are used to...
  - Control access to the resources in the folder through folder-level IAM policies.
  - Enforce constraints on allowed resource configurations through the Organization Policy Service.

What is a project?

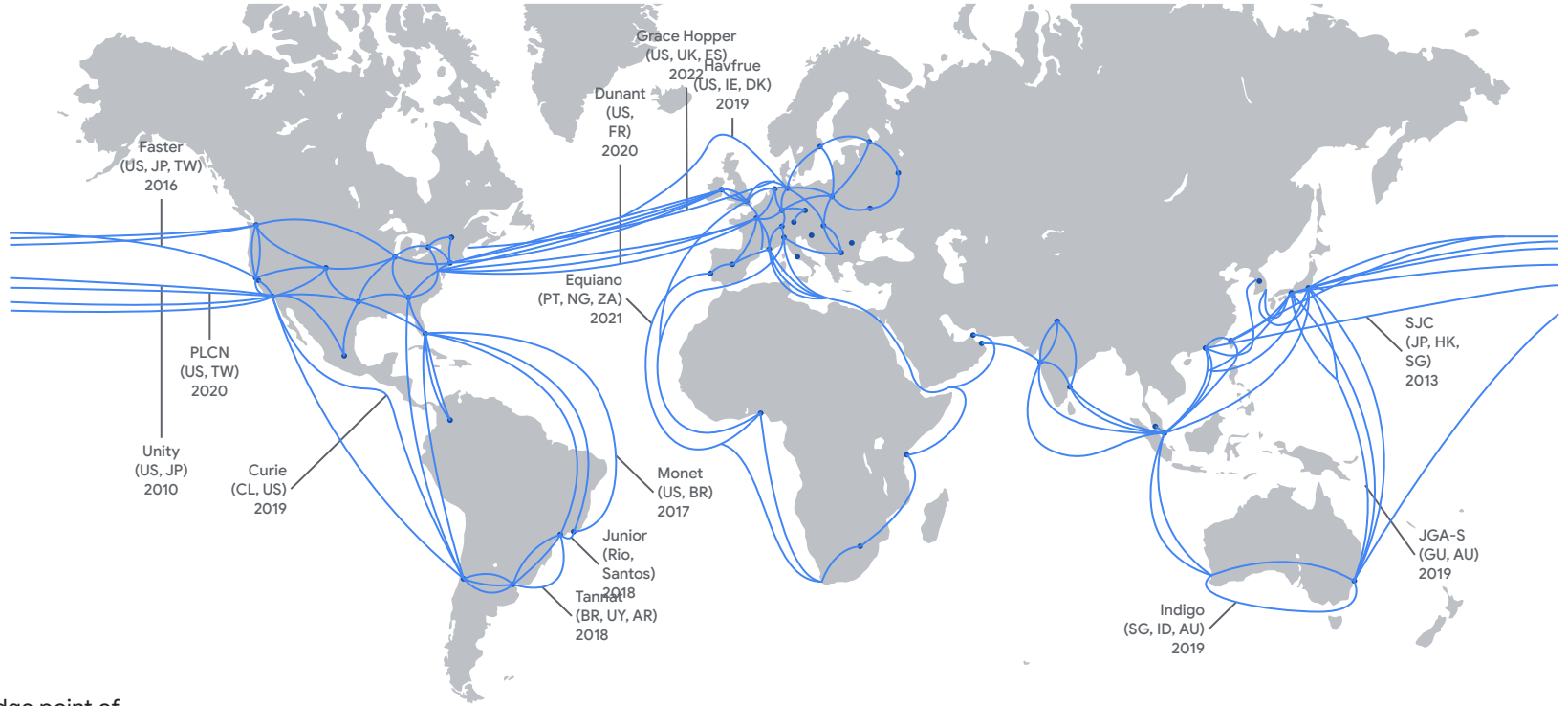
- A resource container
- An IAM enforcement point
- Projects are completely separate from one another
- Resources are part of a project
- Projects can be part of an organization

Networking

searce | SYNOPSYS®

Google's global network infrastructure

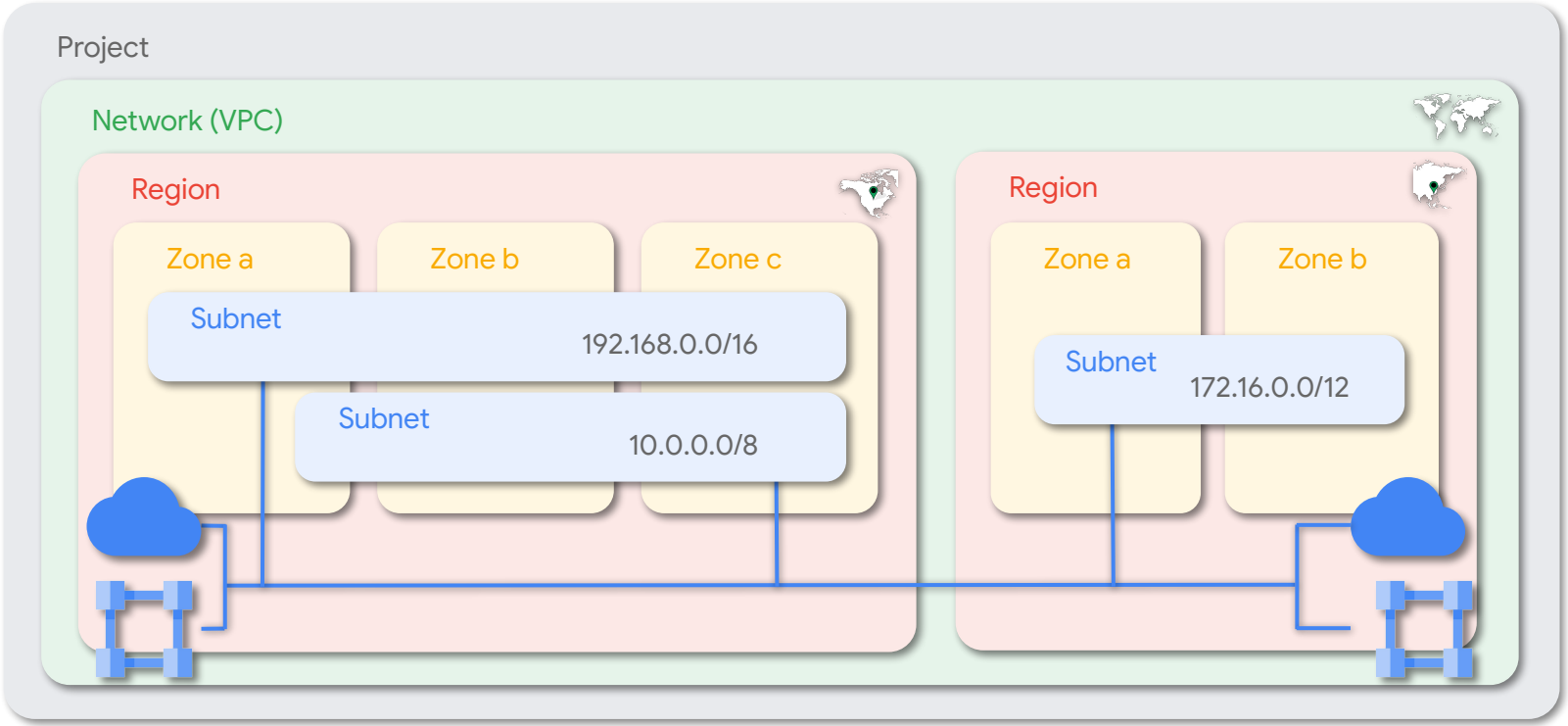
Hundreds of thousands of miles of fiber optic cable connecting all of our data center regions and 140+ points of presence.



• Edge point of presence — Network

searce | SYNOPSYS®

Network concepts



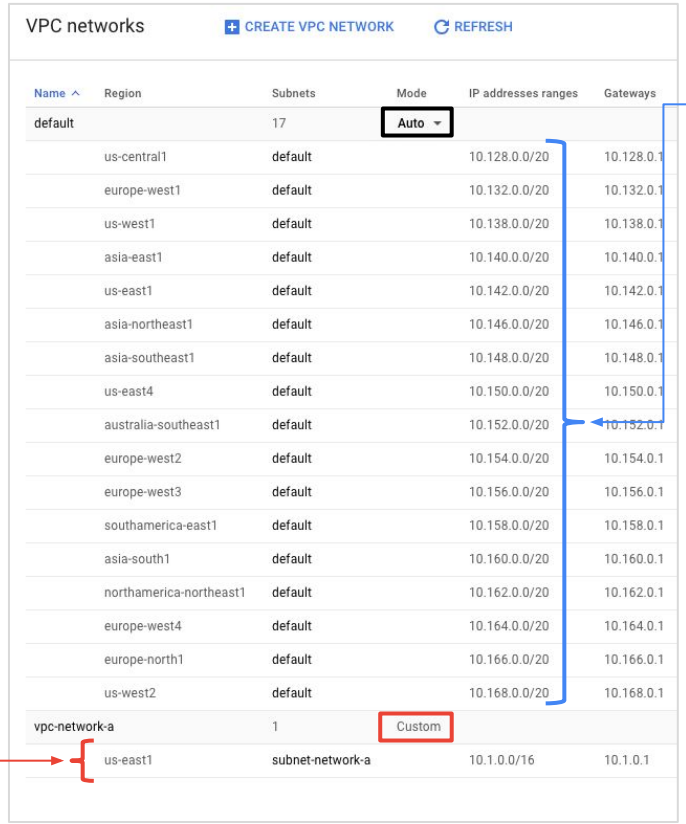
searce | SYNOPSYS®

Subnet creation modes

Best practice

Custom subnet mode

- Network admin **defines subnets and IP ranges**
- No default firewalls rules
- **Expandable** to any RFC-1918 size
- Good for
  - **Production** environments
  - **Preventing CIDR overlap** between environments



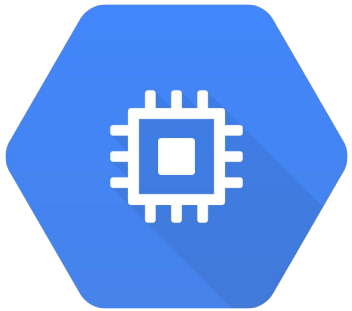
Auto subnet mode

- Default network **when project is created**
- **Default /20** subnetwork per region
- **Expandable** up to /16
- Subnets created as new regions are launched
- Comes with **default FW rules** (for example, TCP 22)
- Good for isolated use cases (PoCs, testing)
- Organization policy to skip default Network creation

Google Compute Engine - VM

Differentiators

searce | SYNOPSYS®



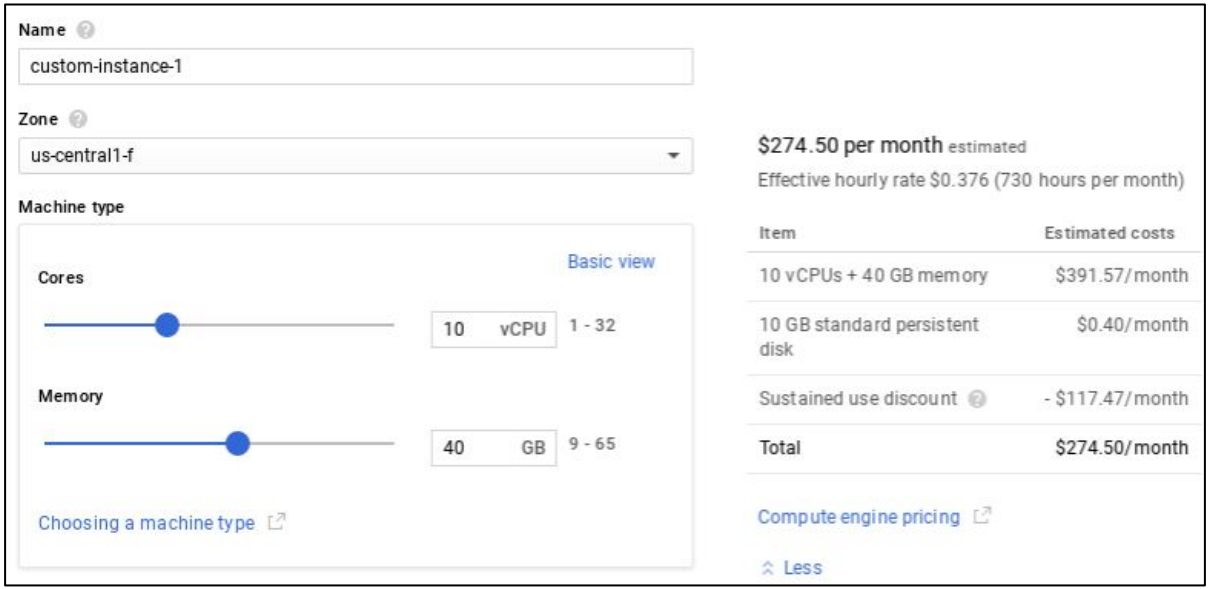
- Consistent Performance
- Best Price per Performance
- Open and Flexible
- Highly Secure
- Globally Interconnected using Google’s Connected Network

Machine Types

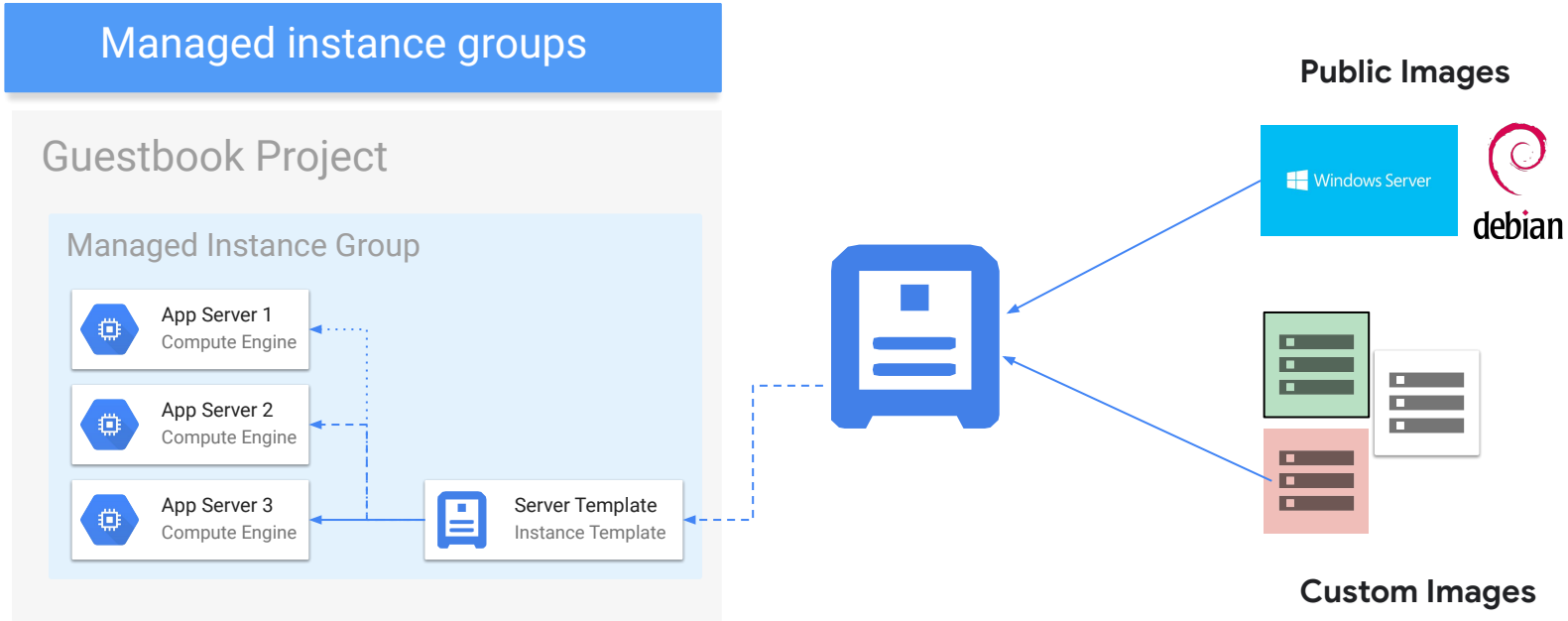
- Selecting a machine type from a machine family determines the resources available to that VM.

	Predefined					Custom	Accelerated Computing
Type	n1-standard	n1-highmen	n1-highcpu	f1-micro	n1-ultramem n1-megamem	Create your own	GPUs
Description	Standard Balanced, good for consistent workload	High ratio of memory to compute	High ratio of compute to memory	Shared Core Burstable, good for changing workloads	Good for in-memory databases	User defines compute and memory, good for custom needs	Good for graphics processing and other GPU uses

GCE Custom Machines

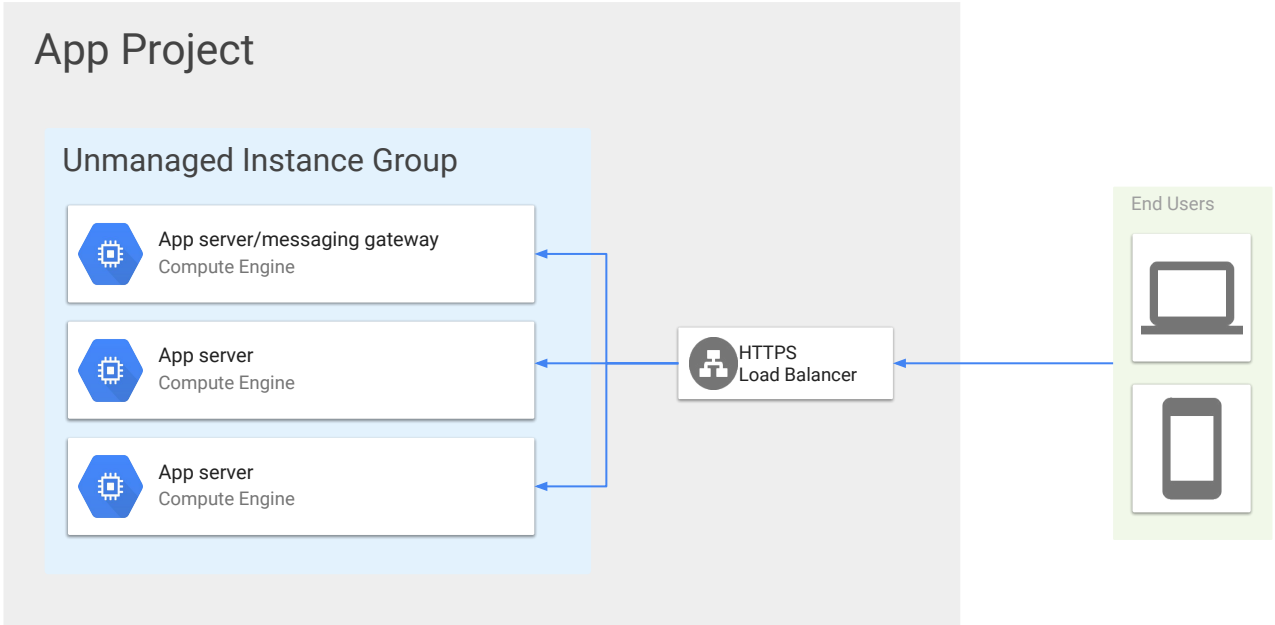


Managed Instance Groups

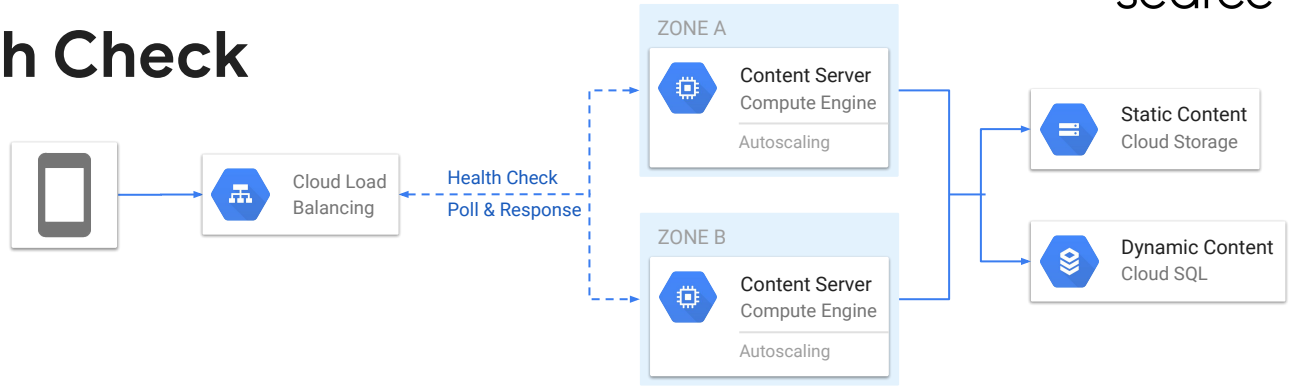


searce | SYNOPSYS®

Unmanaged Instance Groups



Health Check



Health Check Type	HTTP/HTTPS	SSL/TLS
How it works	Probes instance on selected port specified number of times over a configured interval to determine instance health	Simple handshake or request/response
Healthy or unhealthy	Healthy instances return code 200 Failed check marks instance as unhealthy, load balancers stop serving traffic but existing connections are unaffected	Handshake is successful or response is provided Failed check marks instance as unhealthy, load balancers stop serving traffic but existing connections are unaffected
Good for	Use in conjunction with load-balanced managed instance groups to maintain application capacity	

searce | SYNOPSYS®

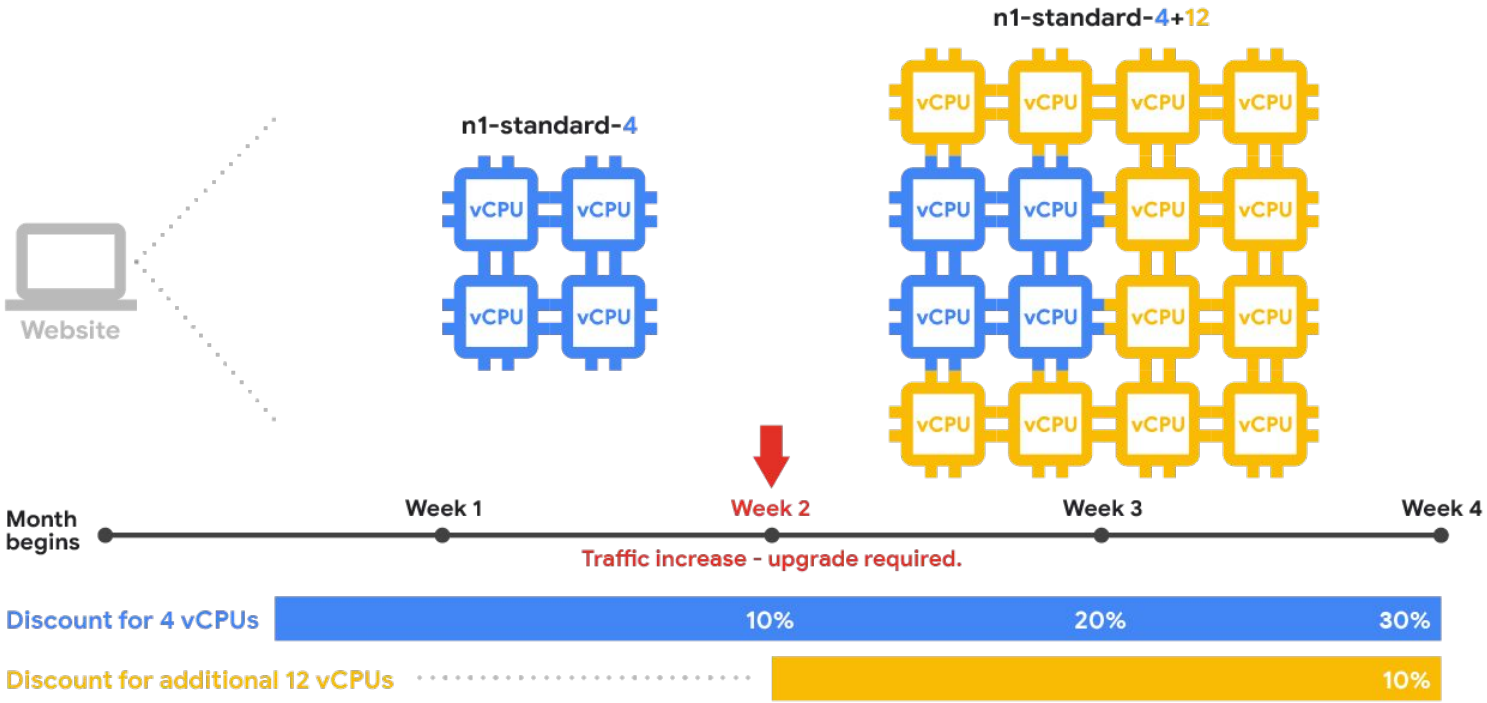
Sustained Use Discount

- Sustained use discounts are automatic discounts for running specific Compute Engine resources a significant portion of the billing month.

Resources	Usage level (% of the month)	% at which incremental is charged	Incremental rate (USD/hour) example: `c2-standard-4` instance
General-purpose N2 and N2D predefined and custom machine types, and Compute-optimized machine types	0%–25%	100% of base rate	$0.2088
	25%–50%	86.78% of base rate	$0.1811
	50%–75%	73.3% of base rate	$0.1530
	75%–100%	60% of base rate	$0.1252

searce | SYNOPSYS®

Sustained Use Discount



searce | SYNOPSYS®

Compute Engine Tips and Tricks

1. Performance for HDD and SSD Persistent Disks increases with volume size up to the per-VM maximums
2. Add auto-scaling to managed instance groups to dynamically increase application capacity
3. Create a library of custom images built on Google's images to use across GCE
4. Consider using Stackdriver to export application logs to enable "statelessness" of managed instance-group machines

searce | SYNOPSYS®

Committed Use Discount

- Compute Engine offers the ability to purchase committed use contracts in return for deeply discounted prices for VM usage. These discounts are referred to as committed use discounts.
- When you purchase a committed use contract, you purchase a certain amount of vCPUs, memory, GPUs, and local SSDs at a discounted price in return for committing to paying for those resources for 1 year or 3 years.
- The discount is up to 57% for most resources like machine types or GPUs.
- The discount is up to 70% for memory-optimized machine types.
- Commitments must be purchased on a per-region basis.

searce | SYNOPSYS®

Spot Virtual Machine

- Spot VMs are available at 60-91% discounts on machine types and GPUs, and significantly lower discounts on local SSDs compared to the on-demand pricing of standard VMs.
- Spot VMs can drastically lower your Compute Engine expenses if your workloads are fault-tolerant and able to endure potential VM preemption.

Preemption Process :

- The VM receives a preemption notification from Compute Engine in the form of an ACPI G2 Soft Off signal. After 30 seconds, if the VM is still running, Compute Engine notifies the operating system with an ACPI G3 Mechanical Off signal.
- The final state of Spot VMs varies depending on specified termination action for each VM:
    - If the termination action is set to STOP or not specified, then Compute Engine stops the VM.
    - If the termination action is set to DELETE, then Compute Engine deletes the VM.

searce | SYNOPSYS®

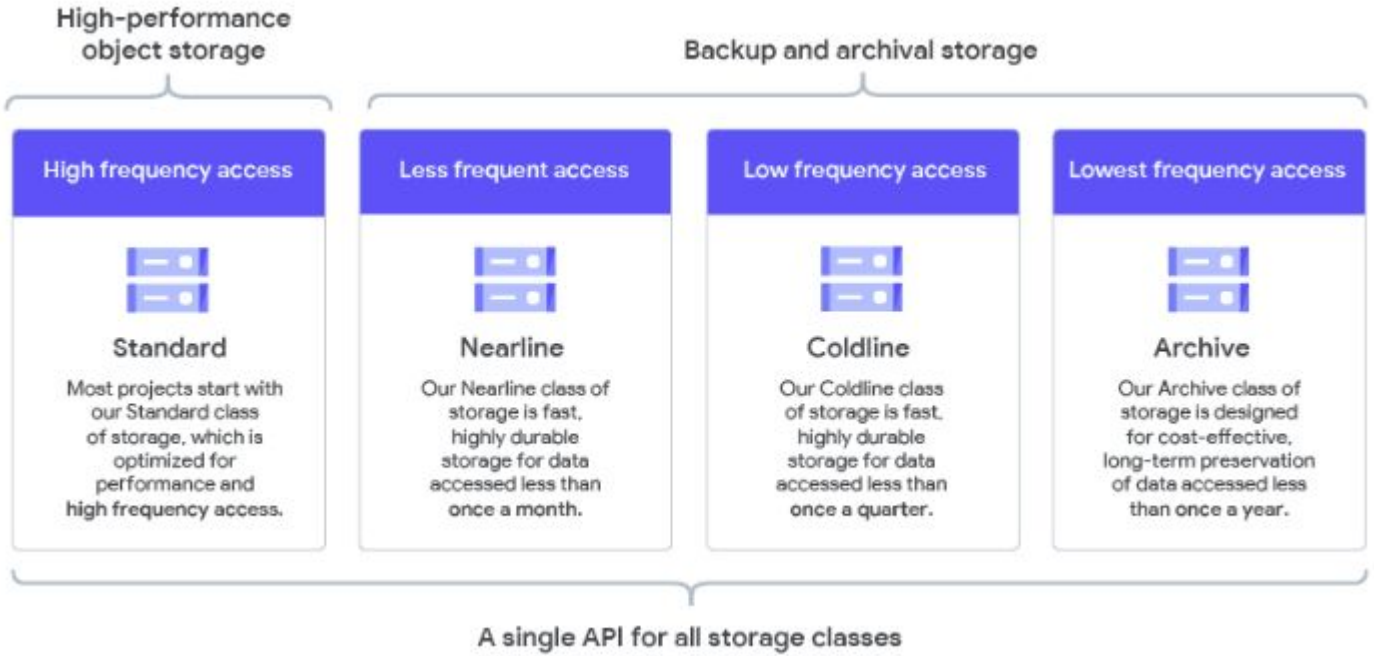
DEMO

COMPUTE LAB : LINK

Google Cloud Storage

searce | SYNOPSYS®

Google Cloud Storage - Types



searce | SYNOPSYS®

Cloud Storage: Access control overview

Permission method	Scope	Access control	Use case
IAM permission	Project, bucket	• Grant access to project's bucket and objects • User must be in IAM	Google Cloud user access
Access control lists (object ACL)	Object	• Grant read or write access to users for objects • Can permit users from outside	User access per object
Signed URLs	Object	• Grant time-limited read or write access to an object • Access to anyone you share URL with	Time limited, temporary access
Signed policy document	Bucket	• Policy control contents that can be uploaded	Session-based uploads (such as web form)

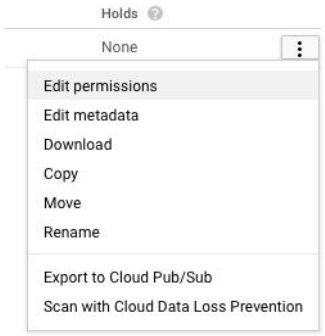
Cloud Storage: Access control models

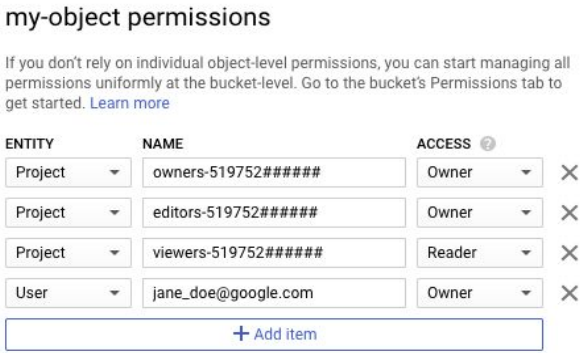
Bucket and object level

- **IAM** for project and bucket level
- **ACLs** for object level
- **Default** configuration

Use when

- Object-level access is absolutely needed





Object-level permissions

Risk of unintended access, including outside the organization

Limited ability to audit access

Cloud Storage: Access control models

Bucket policy only

Best practice

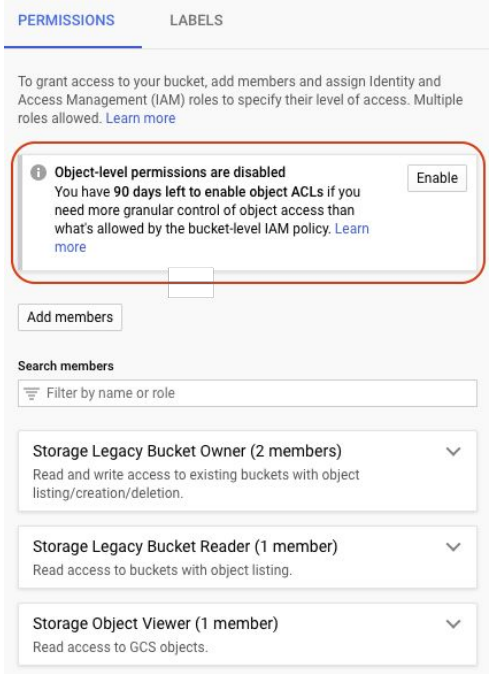
- Project and bucket level only based on **IAM**
- Object level permissions disabled and revoked

Use when

- Object-level access is not needed

Recommended org policies

- Enforce with Bucket Policy
- Complement with Domain Restricted Sharing



Consistent access control

Single permissions systems

Easy to audit access

Object-level access not possible

Export from certain Google Cloud services not supported (like Google Cloud's operations suite, Cloud Audit Logs)

Cloud Storage: Retention lifecycle

Object versioning

- Creates an **archived version of an object** each time the live version of the object is overwritten or deleted.
  - Uniquely identified by a generation number.
  - Retains its ACLs and does not necessarily have the same permissions as the live version of the object.
- Owners can delete object versions

Object lifecycle management

- Set deletion of objects, with versioning enabled:
  - Deleting a live object archives the object
  - Deleting an archived object deletes the object permanently.
- Change the storage class of live and/or archived objects. This action can be applied to both versioned and non-versioned objects.

DEMO

STORAGE LAB : LINK

searce | SYNOPSYS®

Bigquery

searce | SYNOPSYS®

Google BigQuery

- BigQuery is a fully-managed, serverless enterprise data warehouse that enables scalable analysis over petabytes of data and supports querying using ANSI SQL.

- It helps you manage and analyze your data with built-in features like machine learning, geospatial analysis, and business intelligence.

BigQuery Data Transfer Service

The BigQuery Data Transfer Service automates data movement into BigQuery on a scheduled, managed basis. Your analytics team can lay the foundation for a BigQuery data warehouse without writing a single line of code.

You can access the BigQuery Data Transfer Service using the:

- Google Cloud console
- bq command-line tool
- BigQuery Data Transfer Service API

After you configure a data transfer, the BigQuery Data Transfer Service automatically loads data into BigQuery on a regular basis.

You can also initiate data backfills to recover from any outages or gaps. Currently, you cannot use the BigQuery Data Transfer Service to transfer data out of BigQuery.

searce | SYNOPSYS®

DEMO

BIGQUERY LAB : LINK

searce | SYNOPSYS®

Simple docker application deployment on GCE VM

searce | SYNOPSYS®

DEMO

DOCKER APP ON GCE LAB : LINK

searce | SYNOPSYS®

Pub/sub

searce | SYNOPSYS®

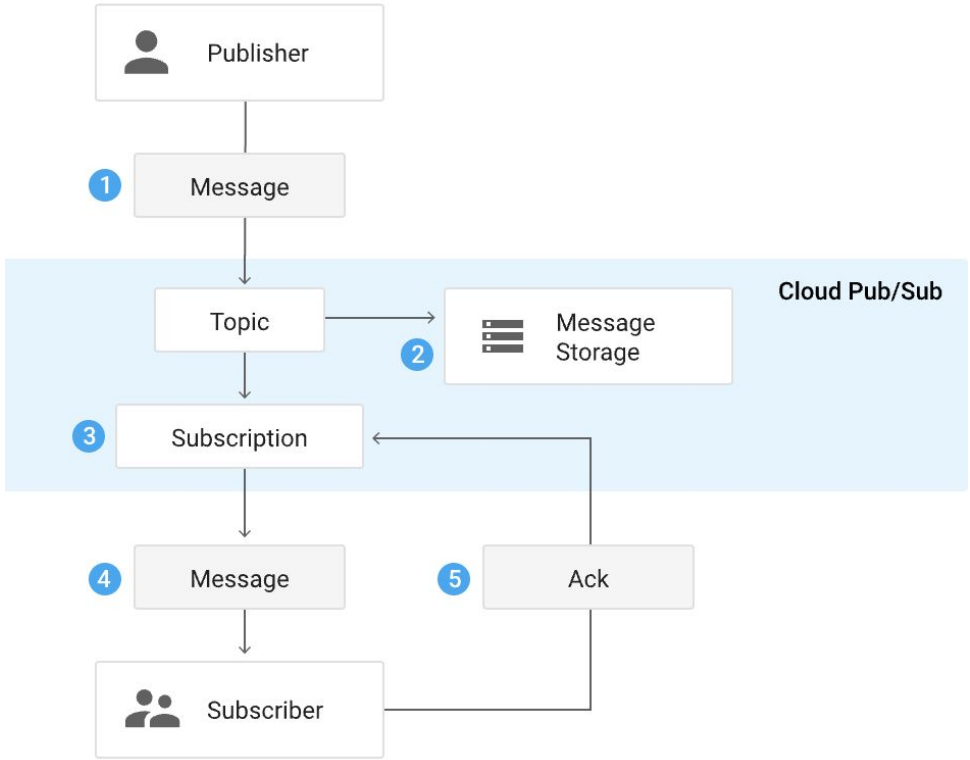
Pub/Sub

- Pub/Sub is a fully-managed real-time messaging service that allows you to send and receive messages between independent applications.

- Pub/Sub allows services to communicate asynchronously, with latencies on the order of 100 milliseconds.

Core concepts

- Topic
- Subscription
- Push and pull
- Publisher
- Subscriber
- Acknowledgment (or "ack")



Types of Pub/Sub services

Pub/Sub consists of two services:

- **Pub/Sub service :** This service is selected by default by users and applications. Offers the highest reliability, largest integration set, and automatic capacity management.

- **Pub/Sub Lite service :** A separate, similar messaging service designed for low cost. Less reliable than Pub/Sub. Provides zonal or regional theme storage.

searce | SYNOPSYS®

Common use case

- Ingestion user interaction and server events.
- Data streaming from applications, services, or IoT devices.
- Refreshing distributed caches.
- Real-time event distribution.

DEMO

MIG AUTOSCALING LAB : LINK

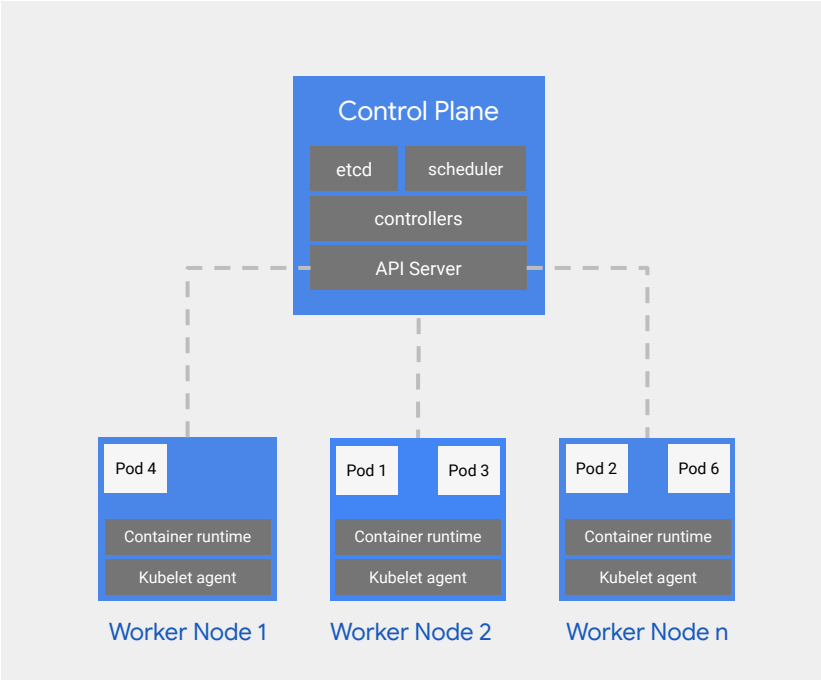
searce | SYNOPSYS®

Google Kubernetes Engine (GKE) Foundations

Kubernetes architecture

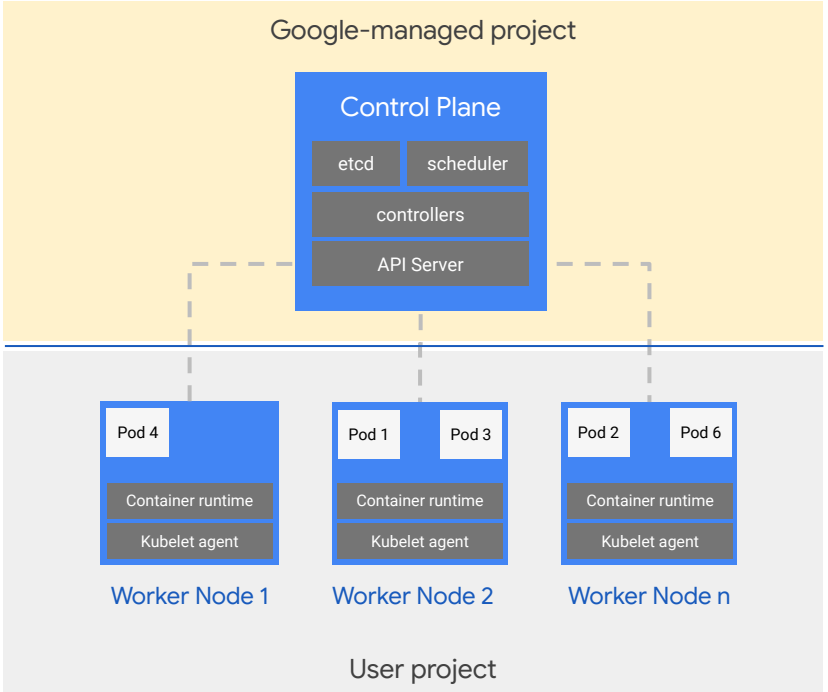
A Kubernetes cluster is composed of several pieces:

- A **control plane**, responsible of handling the overall status of the cluster. Includes components like:
  - etcd database
  - scheduler
  - controllers
  - API server
- Worker **nodes**, responsible for running user workloads and management components (container runtime, kubelet)



Standard GKE on Google Cloud architecture

GKE is managed service. Google takes care of the **control plane creation and maintenance**.



GKE Autopilot flavor

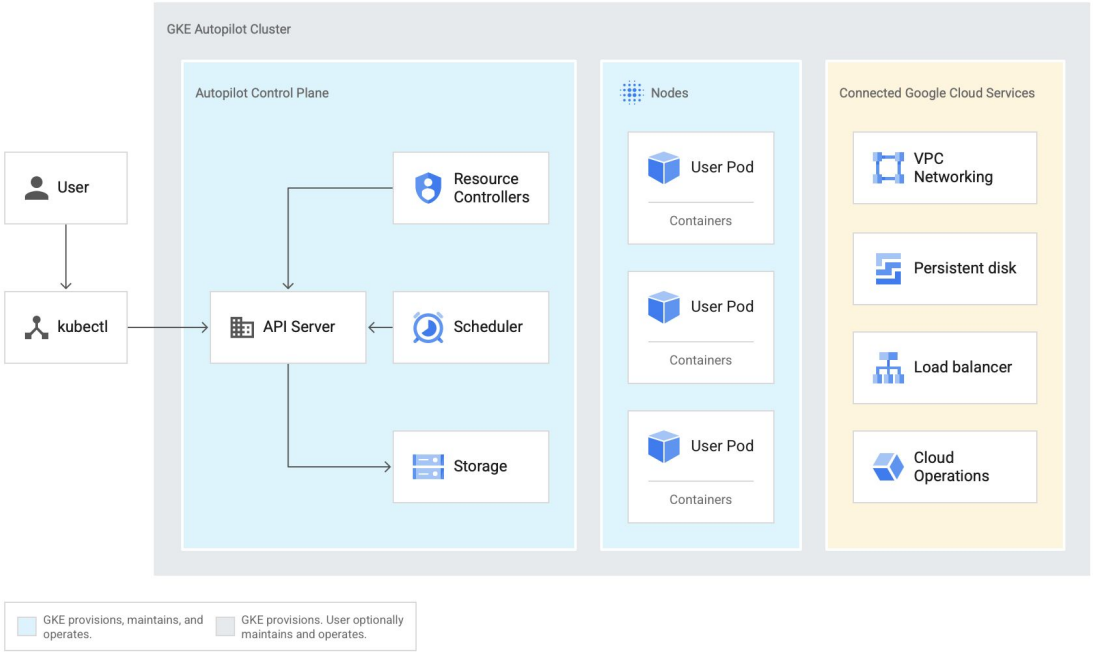
Google manages the entire cluster including cluster nodes

Key Benefits:

- Rich, powerful UI
- SRE monitoring
- Automated repair of apps
- Resource optimized app deployments
- Load balancing and autoscaling of resources
- Global Virtual Private Cloud
- SLA for entire cluster

Limitations:

- Cannot run privileged pods
- Doesn't support Service Mesh (istio/ASM)
- HostPort and hostNetwork not supported



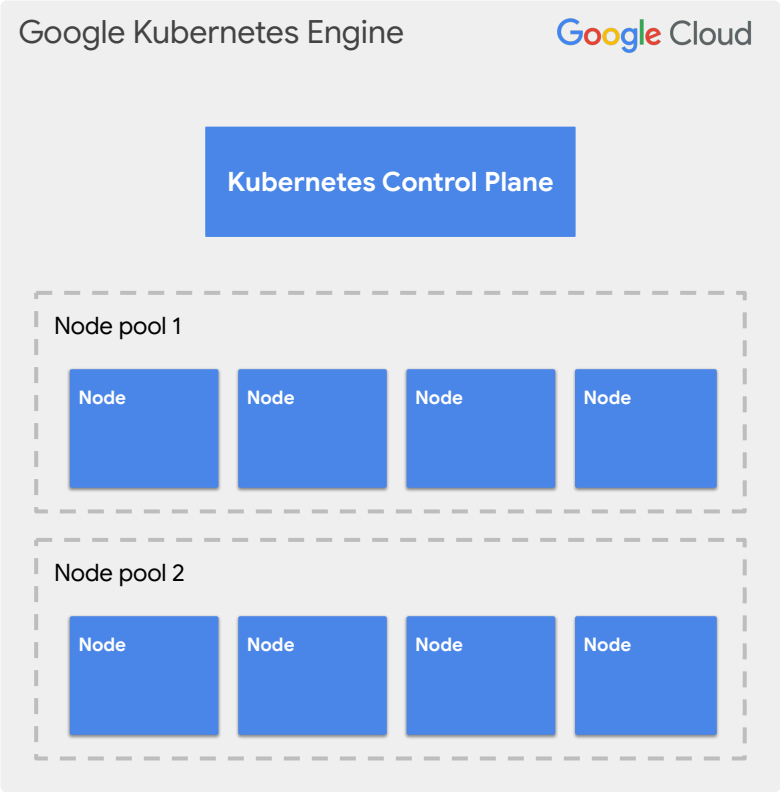
Node pools

A node pool is a **group of nodes that have the same configuration**

Node pools run in the **customer's project**

You can have **one or more node pools** per cluster

Allows you to **mix-and-match machine configurations** (for example, local SSDs, GPU, preemptible VMs, specific node images, larger instance sizes, and more)



Working with node pools

Use a **`nodeSelector`** together with labels to constrain in which nodes a pod can be scheduled. As a convenience, GKE automatically populates the **`gke-nodepool`** label for every node, which contains the name of the node's node pool.

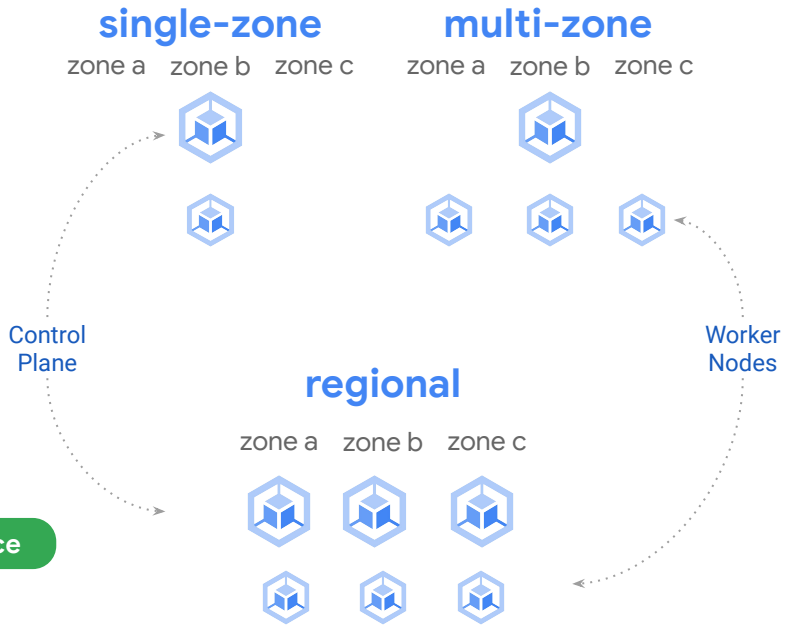
```yaml
apiVersion: v1
kind: Pod
metadata:
  name: nginx
  labels:
    env: test
spec:
  containers:
    - name: nginx
      image: nginx
  nodeSelector:
    cloud.google.com/gke-nodepool: nodepool-gpu
```

Working with node pools (cont.)

To achieve **high availability**, the Kubernetes control plane and its nodes need to be spread across different zones.

- **Zonal clusters** have a single control plane in a single zone.
- **Multi-zonal** clusters have a single replica of the control plane running in a single zone, and has nodes running in multiple zones.
- **Regional clusters** have multiple replicas of the control plane, running in multiple zones within a given region. Nodes can run in multiple zones.

Best practice



GKE cluster types

Routes-based

- Can only use RFC 1918 addresses
- Uses VPC custom routes for pod-to-pod traffic
- Little control over service range

VPC-native

Best practice

- **Default and recommended for new clusters.**
- Can leverage a larger set of private IP ranges.
- Pods get *VPC-native* endpoints, no need for additional custom routes.
- Separate, VPC-routable IP ranges for pods, nodes and services.

IP management in GKE clusters

GKE clusters require up to five network ranges

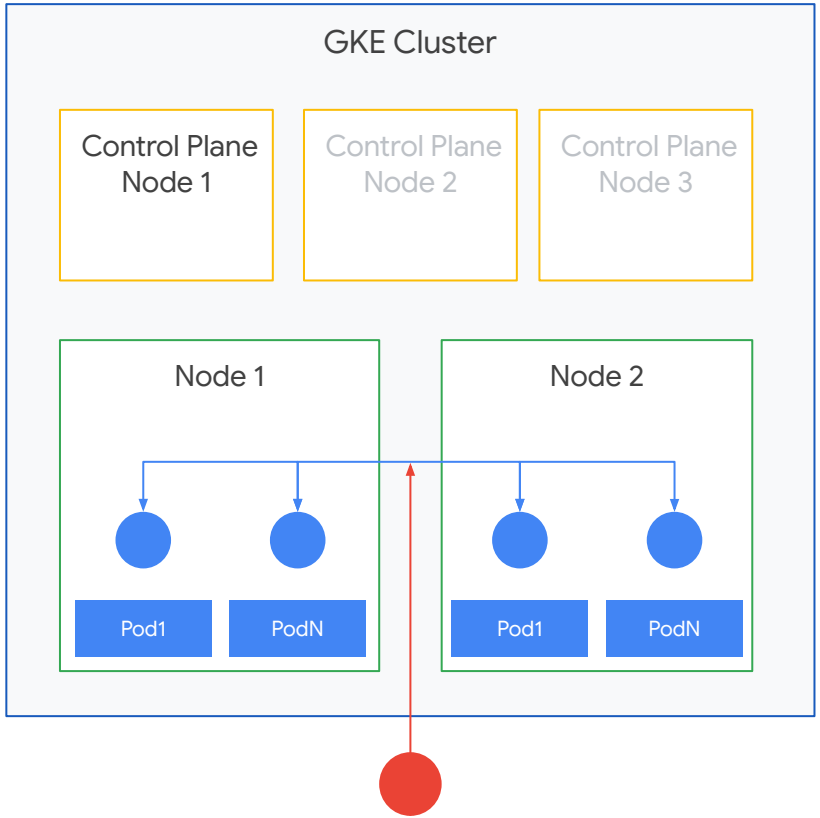
1) **Nodes Range**
   IPs for the VMs backing the cluster

2) **Pods Range**
   IPs for pods created by the cluster. Each node gets a slice (subnet) within this range.

3) **Services Range**
   IPs for Kubernetes services (ClusterIPs)

4) **Control Plane Range**
   IP address range for control plane node (private clusters only)

5) **Internal Load Balancer Ranges**
   (Optional) IP address range for internal load balancers



searce | SYNOPSYS®

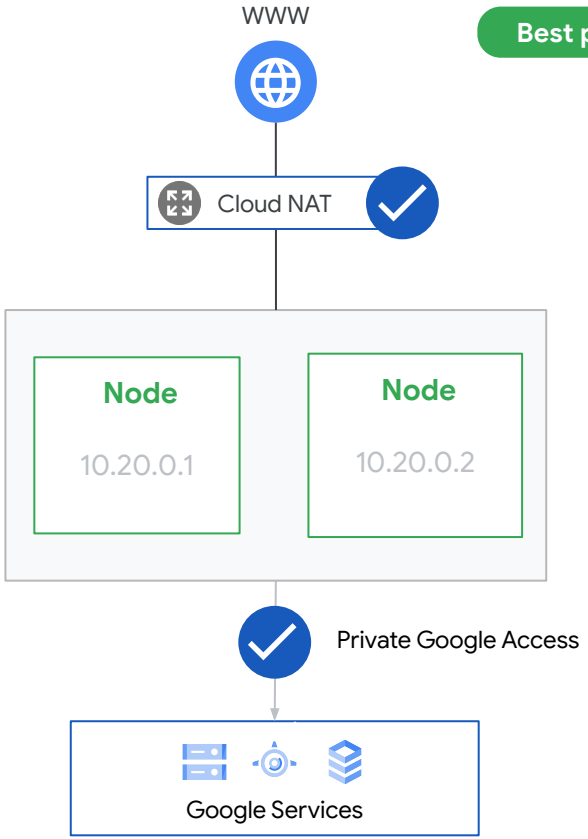
Why do GKE clusters require so many ranges?

- Better integration with Google Cloud's networking stack
- Reserve ranges for pods and services
- **Pod & service IP addresses are natively routable within the cluster's VPC network**
- Better efficiency for routing purposes
- Each resource can scale independently

Private clusters

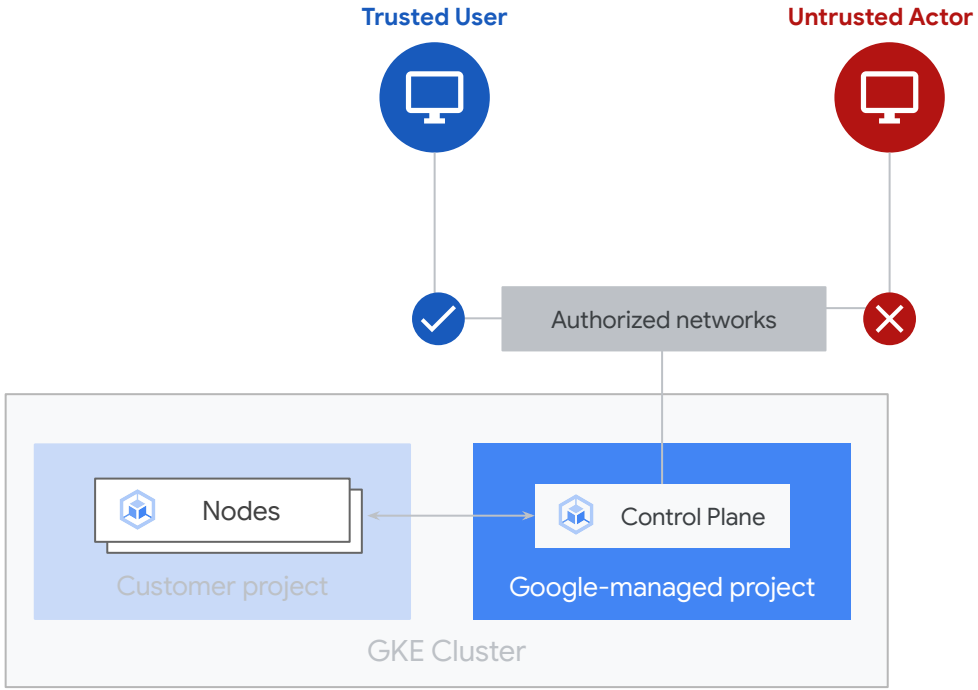
Private clusters **isolate nodes** from having inbound and outbound connectivity to the public internet

- Nodes have only **private IP addresses**
- Nodes use **Private Google Access** to communicate with Google APIs
- Nodes can use Cloud NAT to reach the internet
- **Control Plane** gets an additional private endpoint **for the cluster nodes to talk to the control plane**



Controlling access to the control plane

- Authorized networks **restrict access to the control plane to trusted CIDR ranges.** Mandatory for private clusters.
- By default **nodes and pods ranges are allowed.**
- **Google recommends activating it for all clusters.**
- Control plane access via public IP can be disabled (recommended). This is called Private endpoint.



searce | SYNOPSYS®

DEMO

TRACING LAB : LINK

DEMO

GKE AUTOSCALING LAB : LINK