



Compute & Containers

GCP Professional Cloud Architect | Compute Engine, GKE, Serverless & Choosing Compute

Compute Engine Advanced



Compute Abstraction Layers

Managed Instance Groups (MIGs)	Stateless: autoscaling based on CPU/LB/custom metrics, autohealing via health checks, rolling updates with canary. Stateful: preserve disk, IP, metadata across recreations.
Sole-tenant nodes	Dedicated physical servers. Use for: BYOL (Windows Server, Oracle), compliance requiring physical isolation, workloads needing no co-tenancy.
Preemptible vs Spot VMs	Both up to 91% discount. Spot VMs are the current version (no 24h max). Use for batch processing, rendering, CI/CD, Dataproc.
Custom machine types	Tailor vCPU and memory independently. Up to 624 GB memory. Extended memory option for memory-heavy workloads.
Confidential VMs	Hardware-based memory encryption (AMD SEV). Data stays encrypted in use. For highly sensitive workloads.

Placement Policies

- Spread placement: distribute VMs across hosts for HA (anti-affinity)
- Compact placement: co-locate VMs for low latency (HPC, tightly coupled workloads)
- Live migration: VMs automatically move during host maintenance (default, no downtime)

High Control + Stateful

- Compute Engine
- Sole-tenant nodes
- Bare Metal

High Control + Containerized

- GKE Standard
- GKE + GPU/TPU

Low Ops + Stateless

- Cloud Run
- Cloud Functions

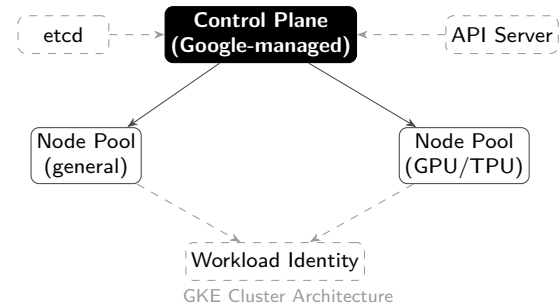
Low Ops + Managed

- GKE Autopilot
- App Engine

Compute Selection Matrix

GKE Standard	GKE Autopilot
Full node control	No node management
Custom machine types	Google manages nodes
Node auto-provisioning	Pod-level SLA
Pay per node	Pay per pod

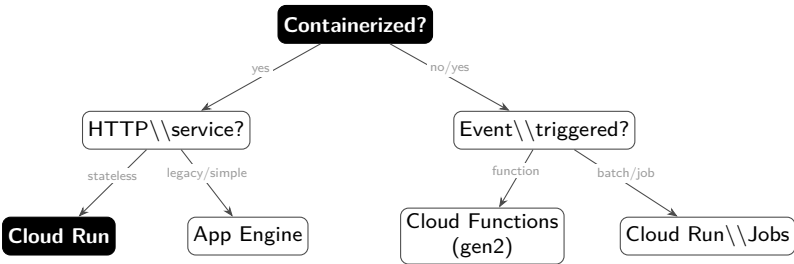
GKE Architecture Deep Dive



Feature	Standard	Autopilot
Node management	You manage	Google manages
Pricing	Per node VM	Per pod CPU/memory/GPU
Security	You harden	Pre-hardened by default
Node access	SSH, privileged pods	No SSH, restricted
GPU/TPU	Full control	Supported (limited configs)
Best for	Custom needs, GPU	Most production workloads

Workload Identity	Map K8s ServiceAccount to GCP ServiceAccount. Preferred auth for GKE workloads — no key files needed.
Node Auto-Provisioning (NAP)	Automatically creates/deletes node pools based on pending pod requirements. Handles heterogeneous workloads.
GKE Sandbox (gVisor)	Extra isolation for untrusted workloads. Kernel-level sandboxing using gVisor runtime.
Binary Authorization	Enforce deploy-time policy: only container images signed by trusted authorities can run on GKE.
Private clusters	Nodes have internal IPs only. Control plane accessible via authorized networks or Private Service Connect.

Serverless Decision Matrix



Choosing Serverless Compute

Cloud Run	Run any container. Scales to zero. Traffic splitting for canary. Up to 8 vCPU, 32 GB RAM. Supports gRPC, WebSockets, HTTP/2.
Cloud Run Jobs	Run containers to completion (no serving). For migrations, data processing, scheduled tasks. Array jobs for parallelism.
Cloud Functions gen2	Built on Cloud Run + Eventarc. Longer timeout (60 min), larger instances, concurrent requests, traffic splitting.
App Engine Standard	Sandboxed runtimes. Fast scale-up, scales to zero. Good for simple web apps. Limited runtime versions.
App Engine Flexible	Custom Docker containers on VMs. SSH access, any runtime. Does NOT scale to zero. Good for WebSockets, background tasks.

8 vCPU	60 min	10,000
Cloud Run Max	Gen2 Functions Timeout	Cloud Run Job Tasks

Choosing the Right Compute

Requirement	Best Choice	Why
Full VM control, BYOL	Compute Engine	IaaS, custom images, sole-tenant
Container orchestration at scale	GKE Standard	Full K8s, GPU/TPU, custom node pools
Managed containers, low ops	GKE Autopilot or Cloud Run	No node management
HTTP microservices	Cloud Run	Scales to zero, any language/container
Event-driven functions	Cloud Functions gen2	Eventarc triggers, quick deploy
Simple web apps	App Engine Standard	Scales to zero, minimal config
HPC, rendering	Compute Engine + MIG	Spot VMs, compact placement, GPUs
Batch processing	Cloud Run Jobs or Batch	Serverless or managed HPC batch

Official resources

⚠ Check the current exam guide

Google Cloud is updating its exams to the products announced at Google Cloud Next '26 (Gemini Enterprise Agent Platform, the data and analytics stack), per the certification hub checked 2026-10-05: read the current exam guide before these sheets.



Professional Cloud Architect exam guide
[services.google...english.pdf](#)
domains and topics



Professional Cloud Architect...
[cloud.google...-architect](#)
format, price, sample questions link, renewal



Professional Cloud Architect learning...
[skills.google/paths/12](#)
Google Skills (ex Cloud Skills Boost)



gcloud CLI cheat sheet
[cloud.google...cheatsheet](#)
the commands on one page

SERVICE	docs	architecture	pricing	
Compute Engine	docs	---	Pricing	2/3
Google Kubernetes Engine	docs	architecture	Pricing	3/3
Cloud Run	docs	---	Pricing	2/3
Architecture Framework	docs	---	---	1/3
Architecture Center	docs	---	---	1/3
gcloud CLI	docs	architecture	---	2/3
per kind	6/6	2/6	3/6	

11 of 18 documentation links (61% coverage) · greyed: no official page given