Building High-Performance, Scalable VM Environments with Google Compute Engine

University of Michigan
Agenda

Over the next hour, we’ll be diving deep into how you can make Compute Engine work for you.
Demo

- Let’s try to set up a massively scalable, world-wide, auto-scaling web site in 15 minutes
Compute Engine Functionality
Google Compute Engine is...

Compute Engine delivers virtual machines running in Google’s innovative data centers and worldwide fiber network.

Compute Engine’s tooling and workflow support enable scaling from single instances to global, load-balanced cloud computing.
# Virtual Machines

## Machine types
Instances range from 1-64 CPUs, and 0.6-416GB RAM. You can choose from these configurations:

- Standard
- High-CPU
- High-memory
- Shared CPU
- Custom

## Disk options
GCE offers standard and SSD persistent disks for your instances, with these features:

- High availability
- Durability
- Encryption
- Performance optimization
- Snapshots

## Billing
Prices range from $0.008/hr.-$4.032/hr. Pricing innovations include:

- Per-minute billing
- Sustained-use discounts
- Preemptible instances
Boot Disk Images

Public images
Compute Engine offers many preconfigured public images that have compatible Linux and Windows operating systems. For example:

- Debian
- CentOS
- CoreOS
- Ubuntu
- Red Hat Enterprise Linux (premium)
- Windows Server (premium)
- SQL Server on Windows Server (premium)

Custom images
A custom image is a boot disk image that you own and control access to. You build custom images using one of the following processes:

- Customize a public image
- Use the Google VM Migration Service
- Manually import RAW image file from AWS or other environments
Instance Groups

You can create and manage groups of virtual machine (VM) instances so that you don’t have to individually control each instance in your project.

- Easy resizing
- Self-healing
- Autoscaling
- Load-balancing
- Zonal or regional
Networks

Compute Engine offers a configurable and flexible networking system. You can manage your Compute Engine network by configuring network, firewall, and instance settings.

- Auto/custom routes
- Firewall rules
- VPNs
- Load Balancing
- Cloud DNS
Administration

All management tasks for Compute Engine work through the GCP APIs. There are a variety of Interfaces that systems operations professionals can use to manage infrastructure.

- Cloud Console
- gcloud, gsutil, etc.
- 3rd-party utilities
- Custom developed management apps
“What could you do with Compute Engine?”
Compute Engine reduces render farm load during periods of peak production

- Consumes processing power of up to 15,000 Intel cores at peak rendering times
- Faster rendering time means visual designers can get results and make tweaks more quickly

$300,000+ saved due to eliminating idle cores during production “quiet times”

“By adding Compute Engine to our workflow and allowing our in-house capacity to focus on the studio work, everyone’s project gets computing time – and the creative team can get as imaginative as they want to, with fast views of new iterations.”
Where to Look for Info About Compute Engine
Start with search

The easiest way to find information on Compute Engine is to search using the following phrasing:

- gcp <topic-of-interest>
- gcloud compute <resource>
Browse the docs

Documentation for each GCP product follow a common layout that’s easy to navigate once you know it. Often, searching will lead you to a How-to Guide, and you can easily jump to main page.
Resources

Product Documentation
https://cloud.google.com/compute/

Try it yourself!
https://goo.gl/SS9im3 (Google Cloud Compute Tutorial)
Thanks