Cognizant

Google Cloud Platform

Introduction to Google Compute Services

GCE | GAE | GKE

May 6, 2020

Agenda - Master Class

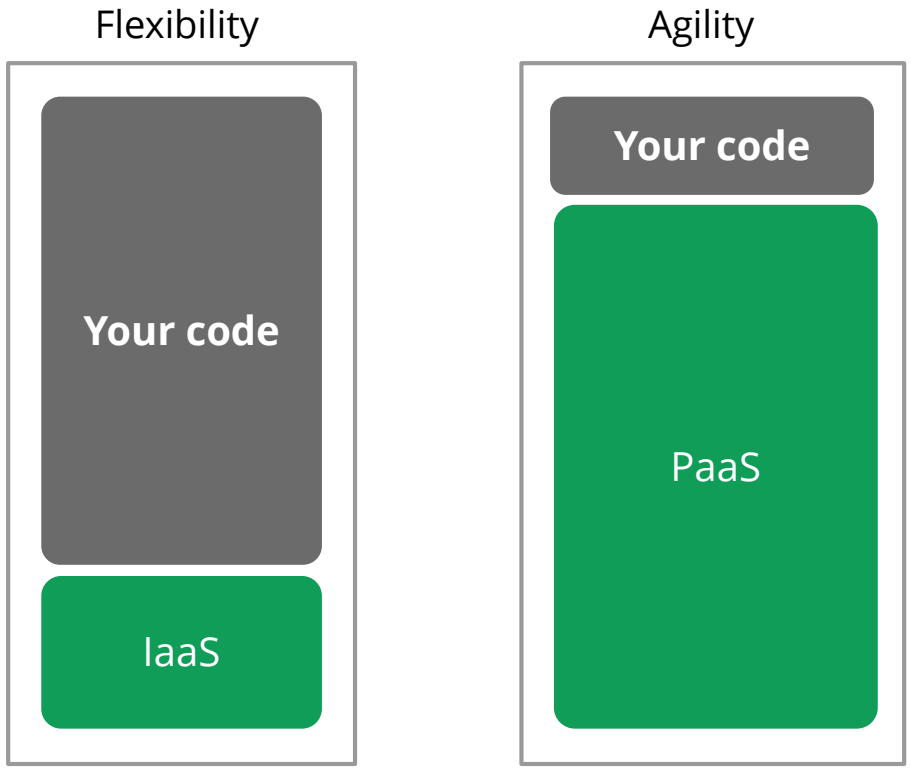
S.No.	Topic (Master Class)	Date	Day
1	Introduction to Google Cloud Platform	4-May	Mon
2	***Introduction to Google Compute Services - GCE | GAE | GKE***	6-May	Wed
3	Introduction to Google Storage - GCS | Bigtable | Big Query | Datastore	8-May	Fri
4	Introduction to Google Networking	11-May	Mon
5	Introduction to GCP Monitoring Services	13-May	Wed
6	**DEMO-I (2 Hours)**	15-May	Fri
7	Introduction to GCP Security Services	18-May	Mon
8	Introduction to Google Data & Serverless Services	20-May	Wed
9	Introduction to GCP DevOps Services	22-May	Fri
10	Introduction to Google API Services	26-May	Tue
11	Introduction to Google Anthos	27-May	Wed
12	**DEMO-II (2 Hours)**	29-May	Fri

Cloud journey…

Cognizant

Yesterday's cloud

Where we were | Where we are | Where we're going



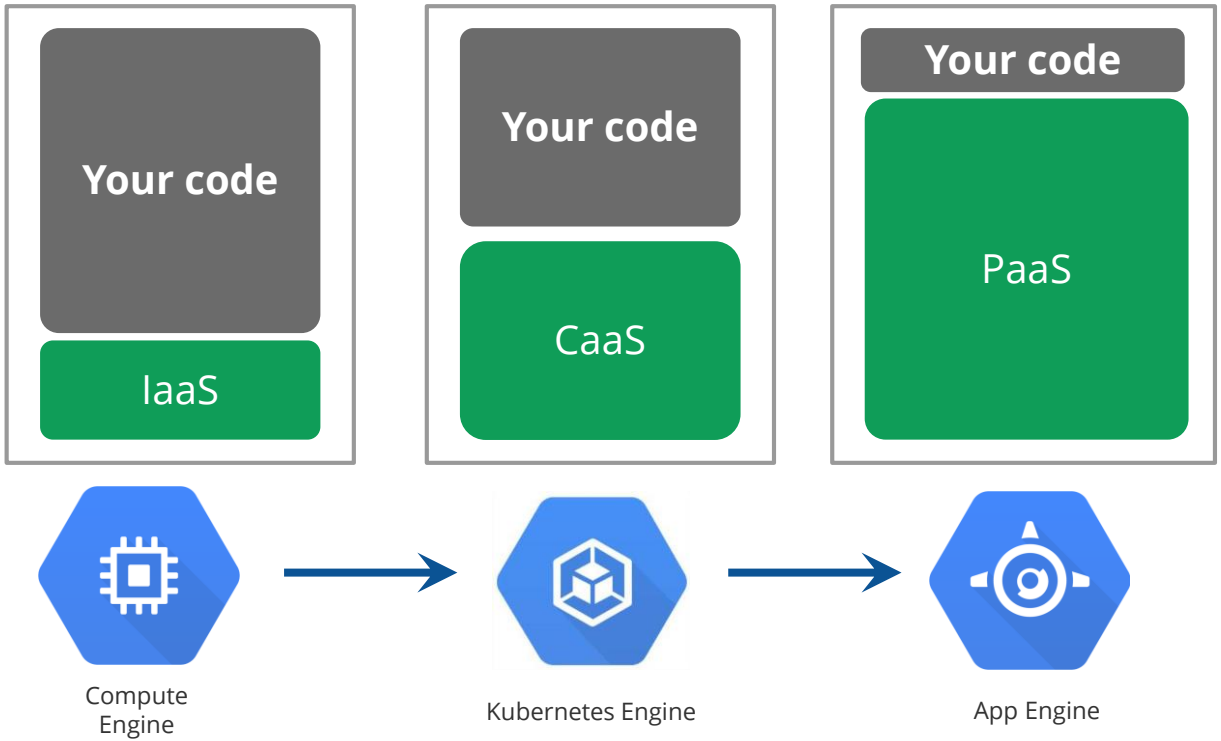
IaaS: Infrastructure as a Service
PaaS: Platform as a Service

Source: Google

Cognizant

Today's cloud

Where we were | Where we are | Where we're going



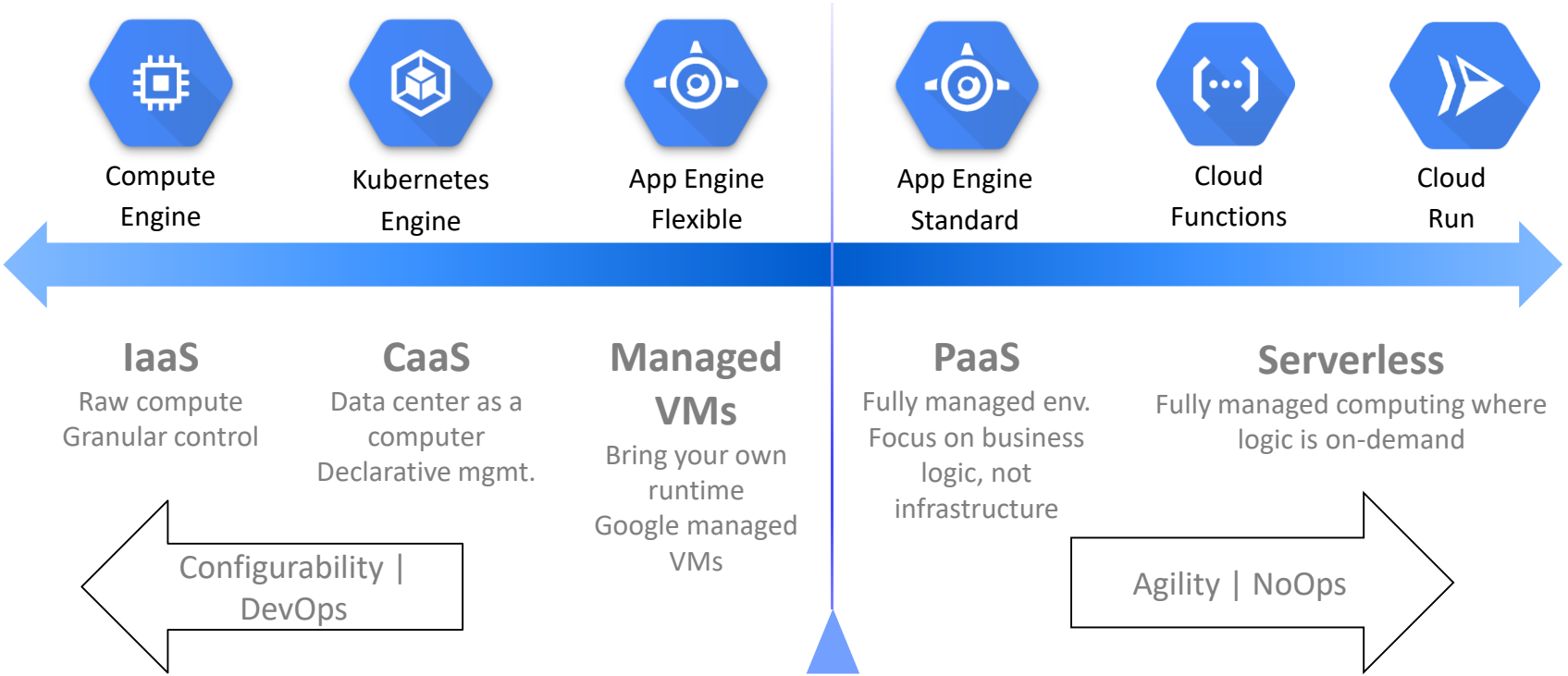
IaaS: Infrastructure as a Service
CaaS: Container as a Service
PaaS: Platform as a Service

Source: Google

Cognizant

Going forward

Where we were | Where we are | Where we're going



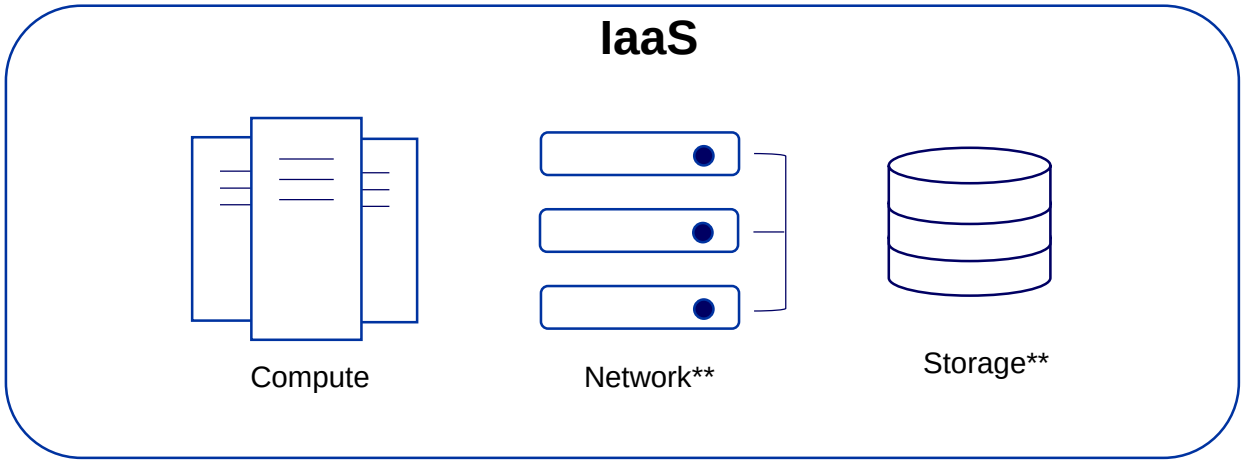
Source: Google

Cognizant

Google Compute Engine (GCE)

Google Compute Engine (GCE)

- Infrastructure as a Service (IaaS) offering of GCP
- It's a zonal resource and available in all regions*



*Some machine types, like N2D, are available in limited regions and being released to other regions gradually.

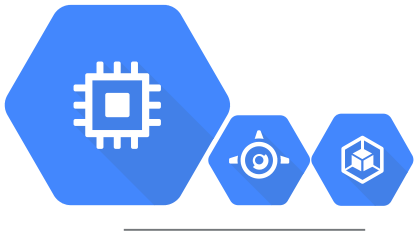
** Network and Storage to be discussed in future sessions

Competitive Products:
AWS : Elastic Compute Cloud (EC2)
Azure : Azure VM

Why use Google Compute Engine (GCE)?

You have an infrastructure-centric view of the world

- **You need complete control** over the virtual-machine infrastructure
- **You need to make kernel-level changes**, such as providing your own network or graphic drivers, to squeeze out the last drop of performance
- You need to **run a software package that can't easily be containerized** or you have existing VM images to move to the cloud



Compute Engine

Virtual machines with industry-leading price/performance

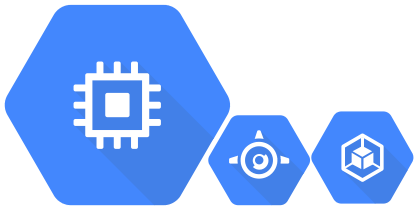
Google Compute Engine (GCE) - Machine Types

- **General Purpose**
- **Compute Optimized**
- **Memory Optimized**
- **GPU Accelerator** e.g. NVIDIA Tesla K80 | P4 | T4 | V100
- **Preemptible**

Machine Type	N1 Generation	N2 Generation
High CPU	0.9GB / vcpu	1 GB / vcpu
High Memory	6.5GB / vcpu	8GB / vcpu

N2D machine types run on the 2nd generation AMD EPYC Rome processor
Support up to 224 vCPUs and 896 GB of memory but do not support GPU

N1 Ultra Memory : 160 vcpu | 3.75 TB



Compute Engine

Virtual machines with industry-leading price/performance

Read More

Cognizant

Recommendation for Machine Types

E2 General purpose	N1, N2, N2D General purpose	M1, M2 Memory-Optimized	C2 Compute-Optimized
Day-to-day computing at a lower cost	Balanced price/performance across a wide range of VM shapes	Ultra high-memory workloads	Ultra high performance for compute-intensive workloads
• Web serving • App serving • Back office applications • Small-medium databases • Microservices • Virtual desktops • Development environments	• Web serving • App serving • Back office applications • Medium-large databases • Cache • Media/streaming	• Large in-memory databases like SAP HANA • In-memory analytics	• HPC • Electronic Design Automation (EDA) • Gaming • Single-threaded applications

Read More

Cognizant

GCE Best Practices

Cost optimization

- Purchase Commitments
- Automate cost optimizations
- Use pre-emptible VMs
- Use auto-scaling (e.g. Horizontal scaling)
- Apply Compute Engine rightsizing recommendations

Region selection

- Region-specific limitations
- Government regulations
- User latency by region
- Latency requirements of your app
- Balance between low latency and simplicity
- Cost variation across regions

Google App Engine (GAE)

Google App Engine (GAE)

PaaS

Google's feature rich PaaS (Platform as a Service) offering

Availability

App Engine is a regional resource available in 18* out of 23 regions

Environment Options

App Engine Standard
App Engine Flexible

Competitive Products

AWS : Elastic Beanstalk (EBS)
Azure : Azure App Service

* *As of May 2020, subject to change*

Cognizant

Google App Engine (GAE)

Focus on your code

Let Google worry about database administration, server configuration, sharding & load balancing.

Deploy at Google scale

You can scale up to 7 billion requests per day and automatically scale down when traffic subsides.

Powerful built-in services

Managed services, such as Task Queues, Memcache and the Users API, let you build any application.

Familiar development tools

Use the tools you know, including Eclipse, IntellIJ, Maven, Git, Jenkins, PyCharm & more.

Popular languages & frameworks

Write applications in some of the most popular programming languages, use existing frameworks and integrate with other familiar technologies.

Multiple storage options

Choose the storage option you need: a traditional MySQL database using Cloud SQL, a schemaless NoSQL datastore, or object storage using Cloud Storage.

App Engine Standard or Flexible?

When to choose the standard environment

Application instances run in a sandbox, using the runtime environment of a supported language listed below.

Applications that need to deal with rapid scaling.

The standard environment is optimal for applications with the following characteristics:

- Source code is written in **specific versions of the supported programming languages**:
  - Python 2.7, Python 3.7
  - Java 8, Java 11
  - Node.js 8, Node.js 10
  - PHP 5.5, PHP 7.2, and PHP 7.3
  - Ruby 2.5 (beta)
  - Go 1.11, Go 1.12 and Go 1.13 (beta)
- Intended to **run for free or at very low cost**, where you pay only for what you need and when you need it. For example, your application can scale to 0 instances when there is no traffic.
- Experiences **sudden and extreme spikes of traffic** which require immediate scaling.

When to choose the flexible environment

Application instances run within Docker containers on Compute Engine virtual machines (VM).

Applications that receive consistent traffic, experience regular traffic fluctuations, or meet the parameters for scaling up and down gradually.

The flexible environment is optimal for applications with the following characteristics:

- Source code that is written in a version of any of the supported programming languages:
  Python, **Java**, **Node.js**, **Go**, **Ruby**, **PHP**, or **.NET**
- Runs in a Docker container that includes a custom runtime or source code written in **other programming languages**.
- Uses or depends on frameworks that include **native code**.
- Accesses the resources or services of your Google Cloud project that reside in the **Compute Engine network**.

Read more

What workloads are ideal?

App Engine's benefits make it ideally suited for building

Mobile backends, especially social and casual games

Software as a Service (SaaS) applications that can disrupt stagnant industries

Internal IT apps that improve productivity and revenue (think Googelplex)

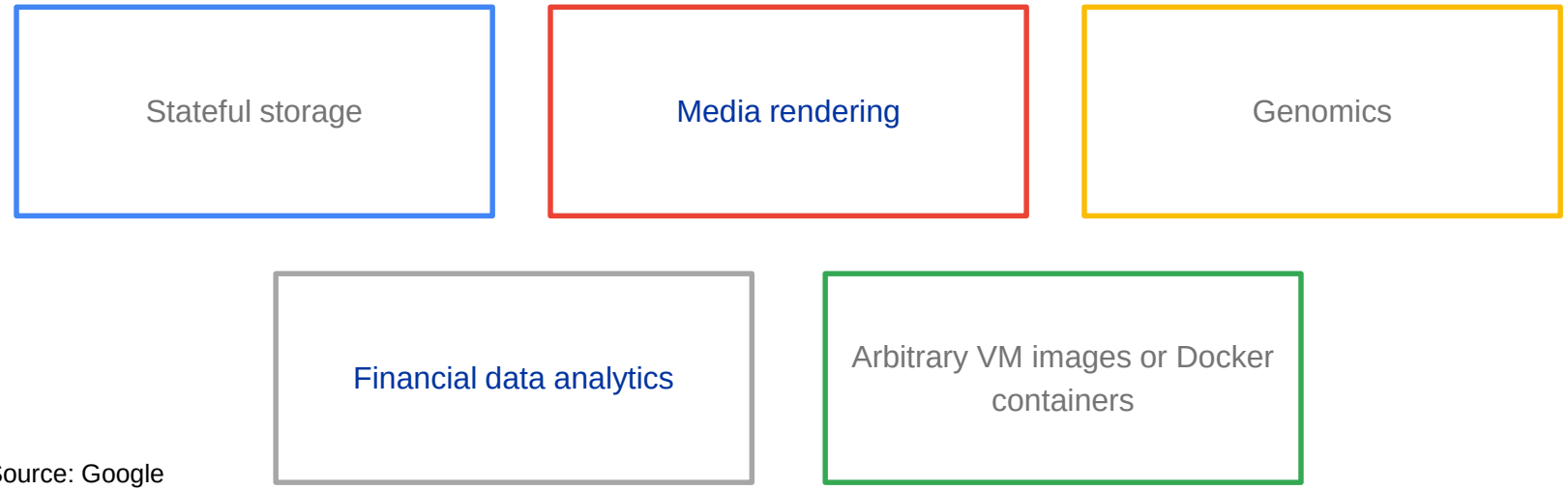
Internet of Things (IoT) front end and backend workloads.

Any web frontend (are you running Tomcat or nginx? stop.)

Source: Google

What workloads aren’t ideal?

Lower-level infrastructure is a better bet here



Source: Google

Cognizant

Google Kubernetes Engine (GKE)

Cognizant

What are containers?

- Just like shipping containers, a software container make it easier for you to package, manage, and ship your code.
- You write software applications that run in a **container.** The container provides the operating system (OS) you need to run your application.
  - This can save a lot of time and cost compared to running servers or virtual machines.

Source: Google

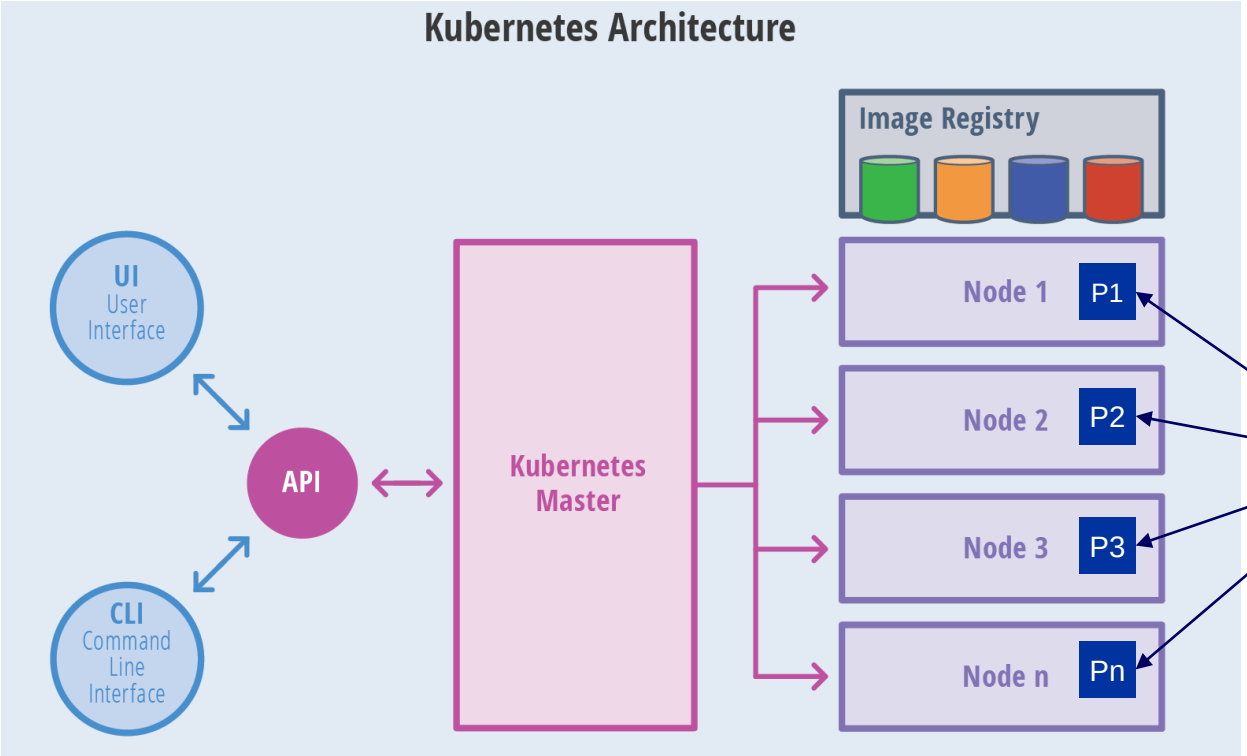
What are containers? Contd.

- Docker is the tool that puts your application and everything it needs in the container.
- Once your application is in a container, you can move it anywhere that will run Docker containers—any laptop, server, or cloud provider.
  - This portability makes code easier to produce, manage, troubleshoot, and update, which creates opportunities for Google partners to sell services.
- As a service provider, containers makes it easy for you to develop code that can be **ported** to your customer and back.

Source: Google

Cognizant

Kubernetes Architecture



Source: TheNewStack

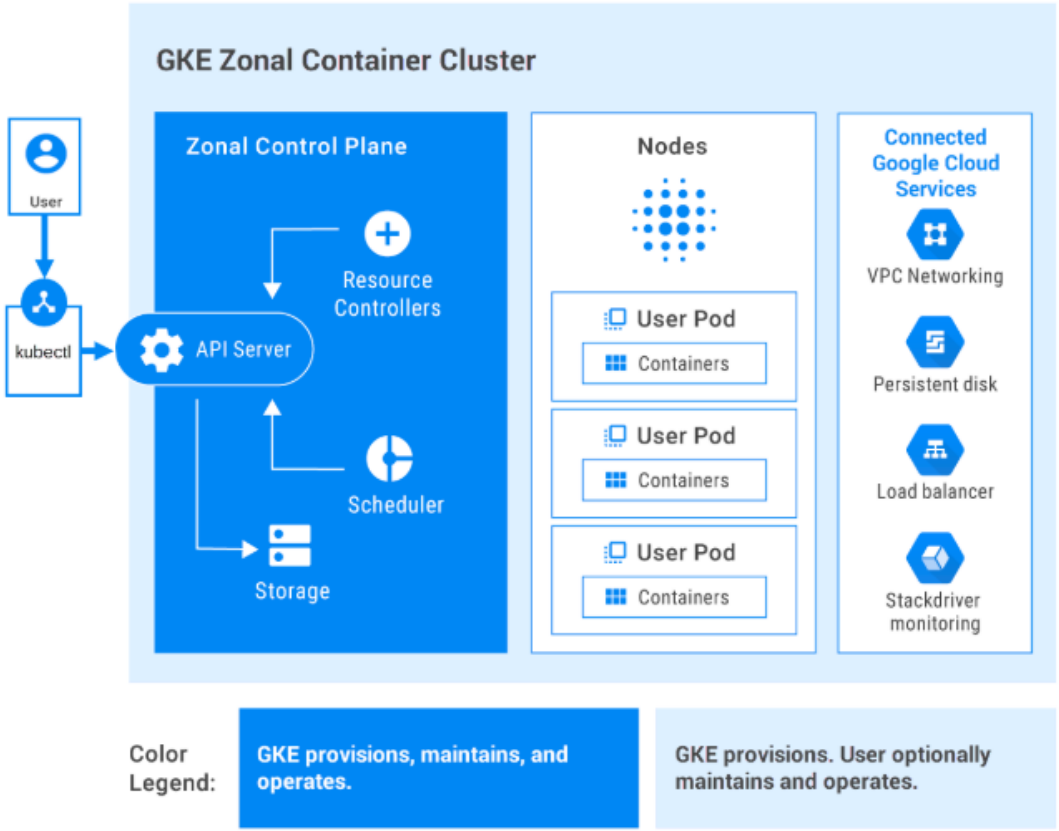
Cognizant

Google Kubernetes Engine (GKE)

- Kubernetes based container orchestration and cluster management system
- Managed service with auto scaling and multi-cluster support. It’s a regional resource and available in all 23 regions
- Start quickly with single-click clusters
- High-availability control plane including multi-zonal and regional clusters.
- Auto-repair, auto-upgrade and release channels
- Secure by default, image vulnerability scanning and data encryption

Cognizant

GKE - Architecture



Cloud Repository

Container Builder

Container Registry

Cognizant

References

- Google Compute Engine - https://cloud.google.com/compute
- Google App Engine - https://cloud.google.com/appengine/
- Google Kubernetes Engine - https://cloud.google.com/kubernetes-engine
- Containers best Practices - https://cloud.google.com/solutions/best-practices-for-operating-containers, https://cloud.google.com/solutions/best-practices-for-building-containers

Thank You

Arun Kumar
475634
Arun.kumar311b80@cognizant.com

Cognizant

Appendix

Cognizant

GKE vs EKS vs AKS

	Google GKE	Amazon EKS	Microsoft AKS
Currently supported Kubernetes version(s)	1.16 (rapid channel - beta), 1.15, 1.14, 1.13(default)	1.14 (default), 1.13, 1.12	1.17 (preview), 1.16(preview), 1.15, 1.14(default), 1.13
Original GA date	Aug-15	Jun-18	Jun-18
Master upgrade process	Automatically upgraded during cluster maintenance window; can be user-initiated	User initiated; user must also manually update the system services that run on nodes (e.g., kube-proxy, coredns, AWS VPC CNI)	User initiated
Node upgrade process	Automatically upgraded (default; can be turned off) during cluster maintenance window; can be user-initiated; GKE drains and replaces nodes	Unmanaged node groups: all user-initiated and managed Managed node groups: User-initiated; EKS will drain and replace nodes	User initiated; AKS will drain and replace nodes
Container runtime	Docker (default), containerd	Docker	Moby Docker
Master/control plane high availability options	Zonal/multi-zonal clusters: single control plane Regional clusters: control plane replicas in multiple zones	Control plane is deployed across multiple Availability Zones	Azure docs do not state redundancy measures for control plane
Master (control plane) SLA	Zonal clusters: 99.5% Regional clusters: 99.95%	99.90%	99.50%

GKE vs EKS vs AKS

	Google GKE	Amazon EKS	Microsoft AKS
Master (control plane) SLA	Zonal clusters: 99.5% Regional clusters: 99.95%	99.90%	99.50%
SLA financially-backed	No	Yes	No
Pricing	Standard costs of GCE machines and other resources	$0.10/hour (USD) per cluster + standard costs of EC2 instances and other resources	Standard costs of node VMs and other resources
GPU support	Yes (NVIDIA); user must install device plugin in cluster	Yes (NVIDIA); user must install device plugin in cluster	Yes (NVIDIA); user must install device plugin in cluster
Master component log collection	Optional, on by default; Sent to Stackdriver	Optional, off by default; Sent to AWS CloudWatch	Optional, off by default; Sent to Azure Monitor
Container performance metrics	Optional, on by default; Sent to Stackdriver	Optional, off by default; Sent to AWS CloudWatch Container Insights	Optional, off by default; Sent to Azure Monitor (preview)
Node health monitoring	Node auto-repair enabled by default	No Kubernetes-aware support; if node instance fails, the AWS autoscaling group of the node pool will replace it	None

Comparison - GAE Standard vs GAE Flexible

Feature	Standard environment	Flexible environment
Automatic in-place security patches	Yes	Yes (excludes container image runtime)
Access to App Engine APIs & Services such as NDB, Users API, Memcache, Images API and others.	• Yes for Python 2.7, PHP 5.5, and Java 8 • No for Java 11, Node.js, Python 3.7, PHP 7.2, PHP 7.3, Ruby 2.5 (beta), Go 1.11, Go 1.12, and Go 1.13 (beta)	No
WebSockets	No Java 8, Python 2, and PHP 5 provide a proprietary Sockets API, but the API is not available in newer standard runtimes.	Yes
Supports installing third-party binaries	• Yes for Java 8, Java 11 Node.js, Python 3.7, PHP 7.2, PHP 7.3, Ruby 2.5 (beta), Go 1.11, Go 1.12, and Go 1.13 (beta). • No for Python 2.7 and PHP 5.5.	Yes
Location	North America, Asia Pacific, or Europe	North America, Asia Pacific, or Europe
Pricing	Based on instance hours	Based on usage of vCPU, memory, and persistent disks