



Introduction au FinOps avec Google Cloud

Julien Landuré



CTO

GDG & DevFest Nantes Organizer
GDE Cloud



jlandure.dev
@jlandure

Back to the future

Le Cloud ?

Elasticité rapide

Accès libre aux capacités de calcul

Mutualisation des ressources

Accès via des mécanismes réseaux standards

Mesure fine des usages

5 stratégies de déploiement

Rehost
Replatform
Refactor
Rebuild
Replace



 **Public Cloud & Move to Cloud 5 stratégies possibles !**

[medium.zenika.com](https://medium.com/zenika/public-cloud-et-move-to-cloud-5-strategies-possibles-1234567890) • 5 min read

Compute Spectrum



IaaS

PaaS

SaaS

Servers
VM instances

Clusters
Cluster management

Serverless, autoscaling



Compute
Engine



Kubernetes
Engine



App Engine



Cloud Run

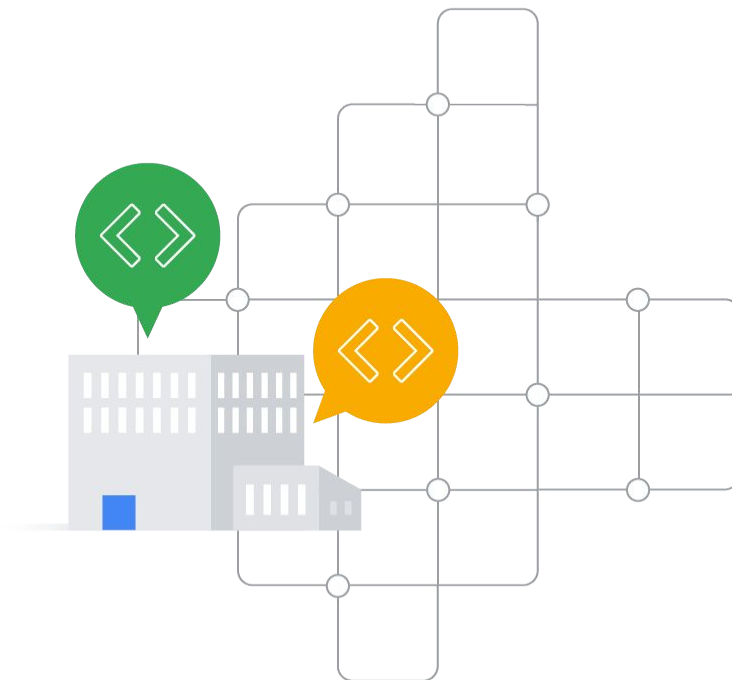


Cloud
Functions

Payment models

IaaS

Pay for what
they allocate

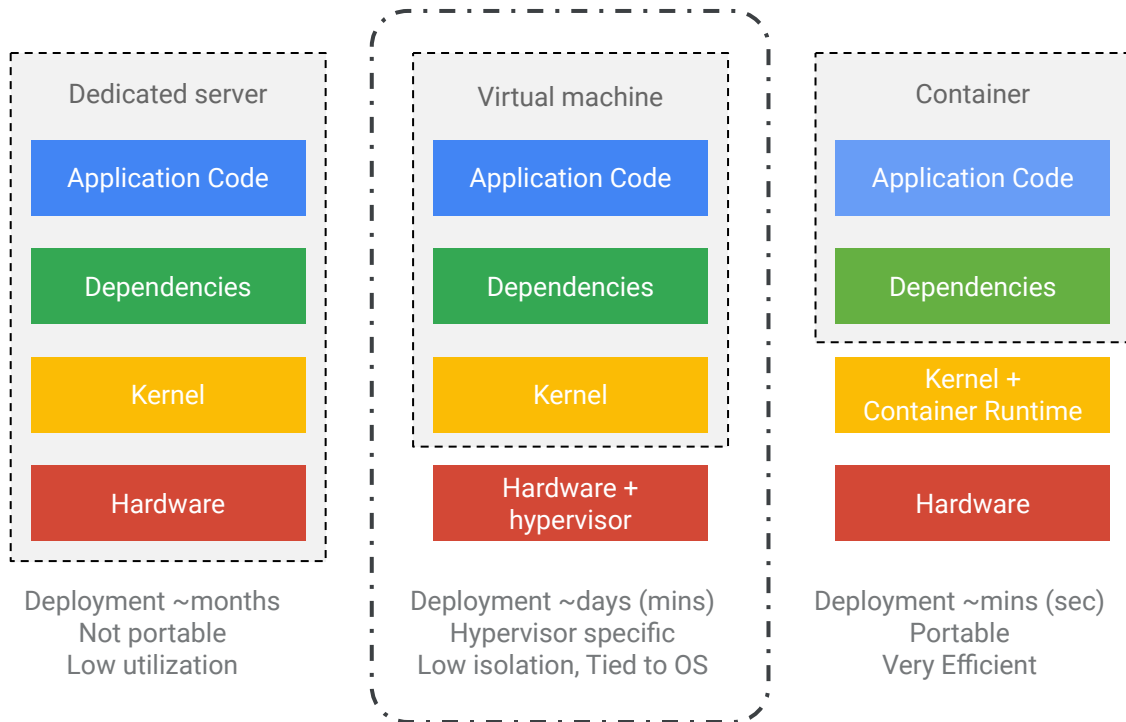


PaaS

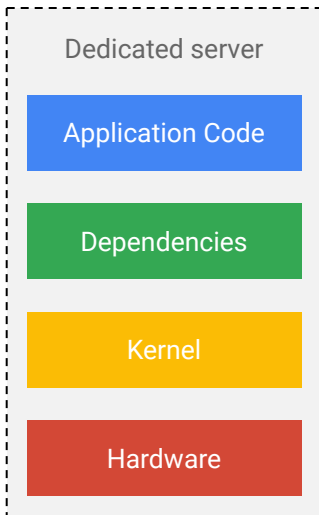
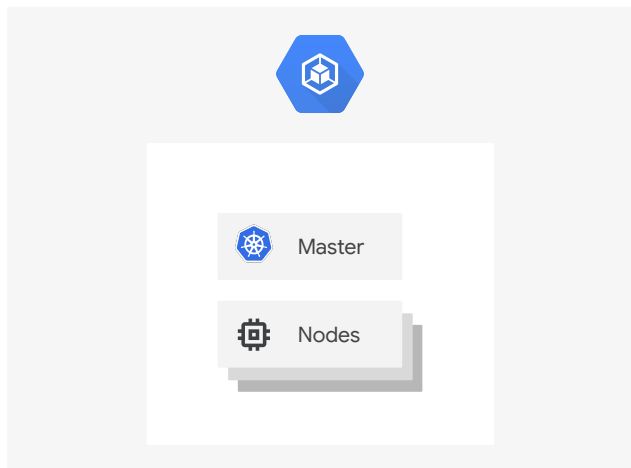
Pay for what
they use

Ce qui existe en standard...
“Pay-as-you-go” model

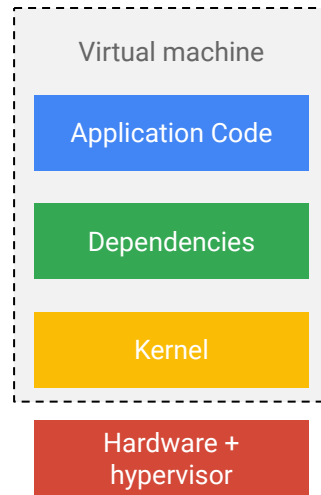
VM World



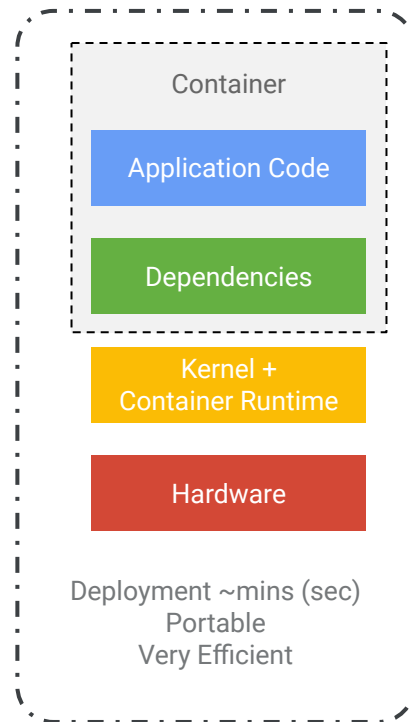
Container World (Kubernetes)



Deployment ~months
Not portable
Low utilization

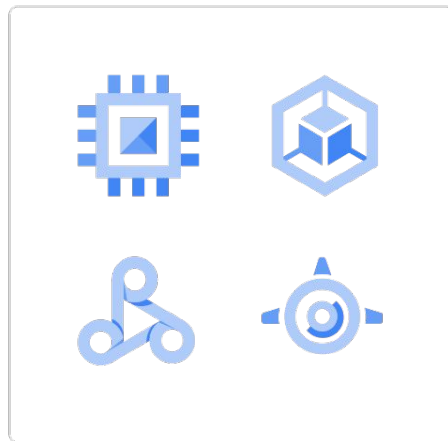
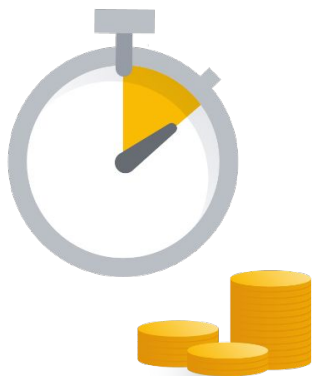


Deployment ~days (mins)
Hypervisor specific
Low isolation, Tied to OS



Deployment ~mins (sec)
Portable
Very Efficient

Google Compute products are billed per-second



What is IaaS / Compute Engine?

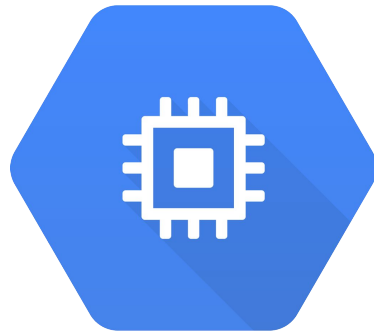
Per VM max:

- 416 vCPUs, 12 TB of RAM, GPUs, TPUs
- Independently configurable, including disk!

Start quickly, ~20 seconds to user code running

Pre-built images, or create your own

- Debian, CentOS, CoreOS, Ubuntu, RHEL, SUSE
- Windows Server Datacenter: 2008, 2012 R2, 2016 , 2019



Disk

Persistent Disk

- Network based
- Configure size (to 64 TB)
- Disk or SSD
- Independent of VM
 - Global Snapshots
 - Multi-VM RO mount
- Larger disks ==
more throughput

Local SSD

- Lowest latency
- 375 GB per disk (up to 8)
- Tied to VM lifecycle



Add

☐ A

☐ A

Manag

Delet

☒ D

Encry

Aut

Addit

Name

Cre

Le

You v

Cre

Equiv

Create a disk

Name ?

disk-1

Description (Optional)

Local SSD scratch disk (maximum 8)

Max disk size: 375 GB

SSD persistent disk

Max disk size: 65,536 GB

Standard persistent disk

Max disk size: 65,536 GB

Size (GB) ? (Optional)

Estimated performance ?

Operation Type

Read

Write

Sustained random IOPS limit

Sustained throughput limit (MB/s)

GCE

Machine Types

Workload type					
General-purpose workloads			Optimized workloads		
Cost-optimized	Balanced	Scale-out optimized	Memory-optimized	Compute-optimized	Accelerator-optimized
E2	N2, N2D, N1	Tau T2D	M2, M1	C2, C2D	A2
Day-to-day computing at a lower cost	Balanced price/performance across a wide range of VM shapes	Best performance/cost for scale-out workloads	Ultra high-memory workloads	Ultra high performance for compute-intensive workloads	Optimized for high performance computing workloads
• Web serving • App serving • Back office apps • Small-medium databases • Microservices • Virtual desktops • Development environments	• Web serving • App serving • Back office apps • Medium-large databases • Cache • Media/streaming	• Scale-out workloads • Web serving • Containerized microservices • Media transcoding • Large-scale Java applications	• Medium-large in-memory databases such as SAP HANA • In-memory databases and in-memory analytics • Microsoft SQL Server and similar databases	• Compute-bound workloads • High-performance web serving • Gaming (AAA game servers) • Ad serving • High-performance computing (HPC) • Media transcoding • AI/ML	• CUDA-enabled ML training and inference • HPC • Massive parallelized computation

Custom Machine feature

Config your own custom machine, with average **19% savings**

- Flexible configurations
- Create a machine type with 1 vCPU and up to 96 vCPUs
- Extended Memory enables higher Mem:vCPU ratios
- Sizing recommendation

Machine type

Basic view

Cores

8 vCPU 1 - 32

Memory

30 GB 7.2 - 52

[Choosing a machine type](#)



```
// CREATE INSTANCE WITH 6 vCPUs and 5 GB MEMORY
```

```
gcloud compute instances create my-vm --custom-cpu 6  
--custom-memory 5
```

#1 Rightsizing

Preemptible / Spot VMs

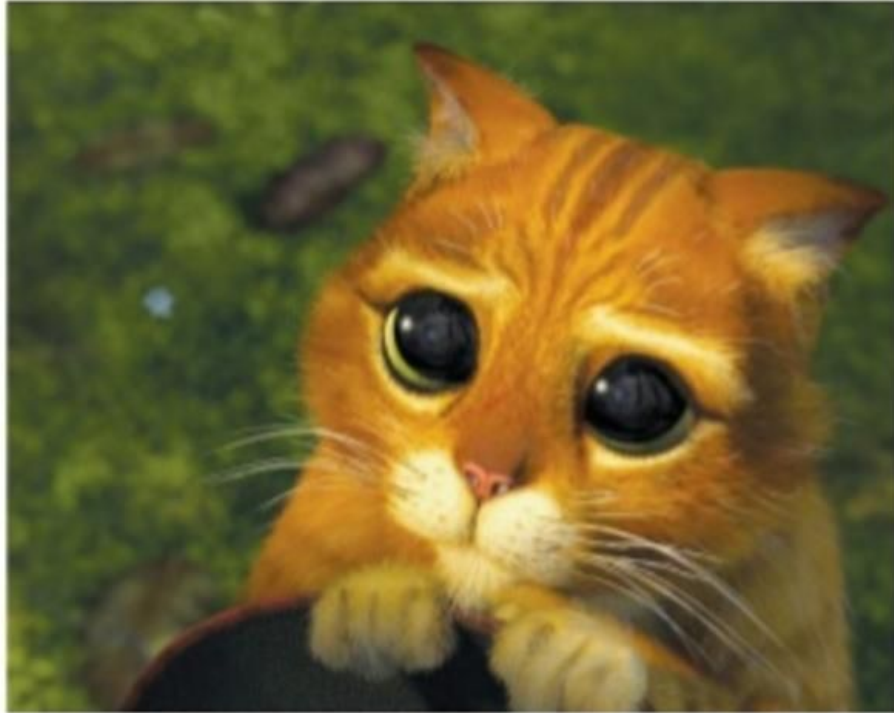
Short lived, low cost virtual machines.

- Can run up to 24 hours
- Might be preempted at any time with 30 seconds notification
- Much lower price (up to 80% cheaper) than regular instances
- Offers the same machine types and options as regular VM instances
- Best for running non-urgent fault-tolerant jobs or short-lived tasks, like genomics, media transcoding, etc.



#2 Spot VM

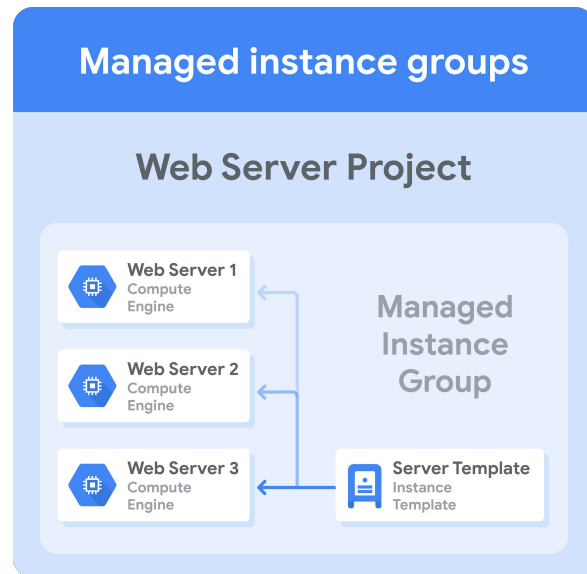
Pet VS Cattle



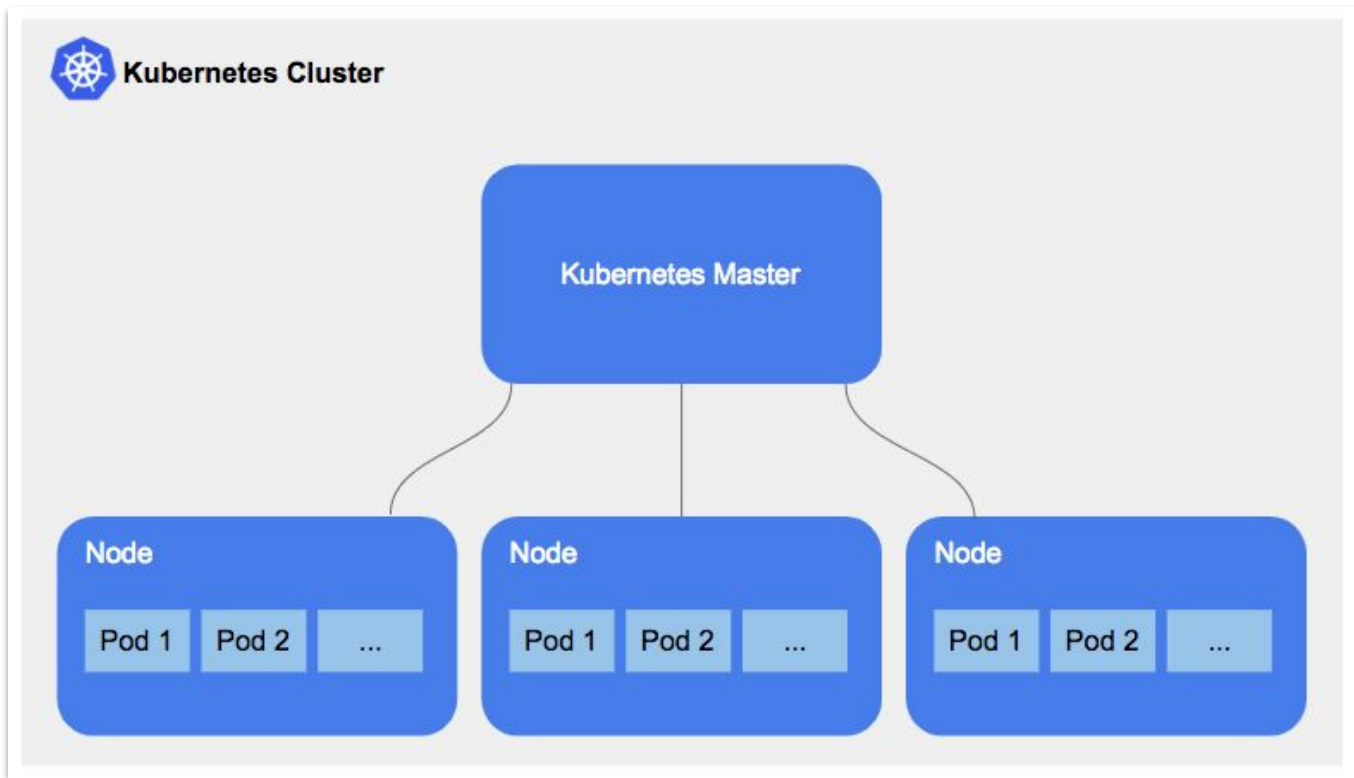
Managed instance groups (MIGs)

A **collection of identical VM instances** that you control as a single entity, using an instance template.

- Instance group can be resized
- Manager ensures all instances are RUNNING
- Typically used with **autoscaler**
- Can be single zone or regional
- Supports **autohealing**



Kubernetes (GKE)

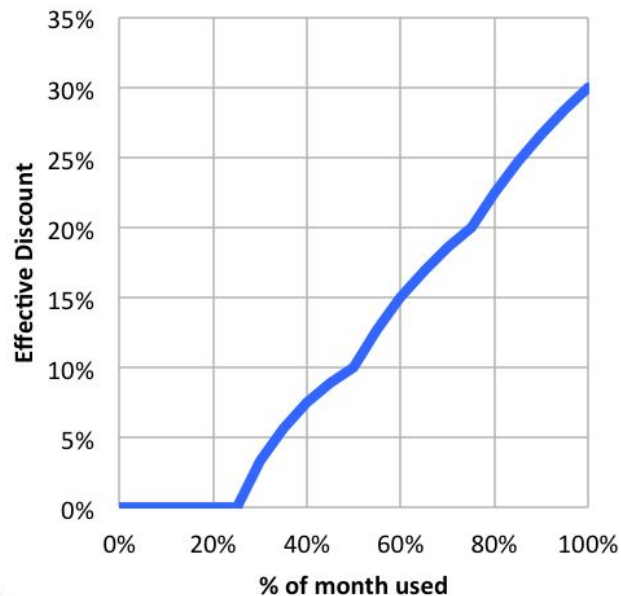


#3 Autoscaling

Sustained use discounts

Up to 30% net discount for instances that run the entire month.

Usage Level (% of month)	% at which incremental is charged
0% - 25%	100% of base rate
25% - 50%	80% of base rate
50% - 75%	60% of base rate
75% - 100%	40% of base rate



#4 Automatic Discounts

Ce qui existe en standard...
“Pay-as-you-use” model

Serverless

Deploy and scale applications fast and securely in a fully managed environment



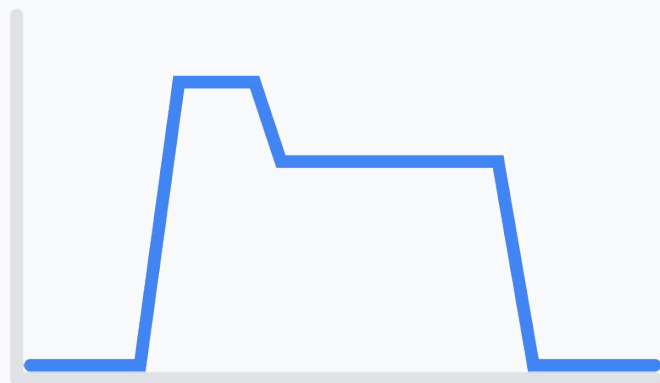
Auto-scaling



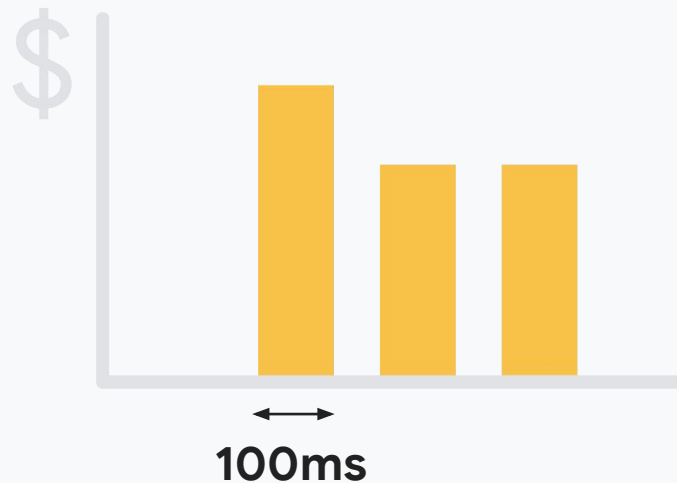
No Infra
Management

#3 Autoscaling

Pay per use

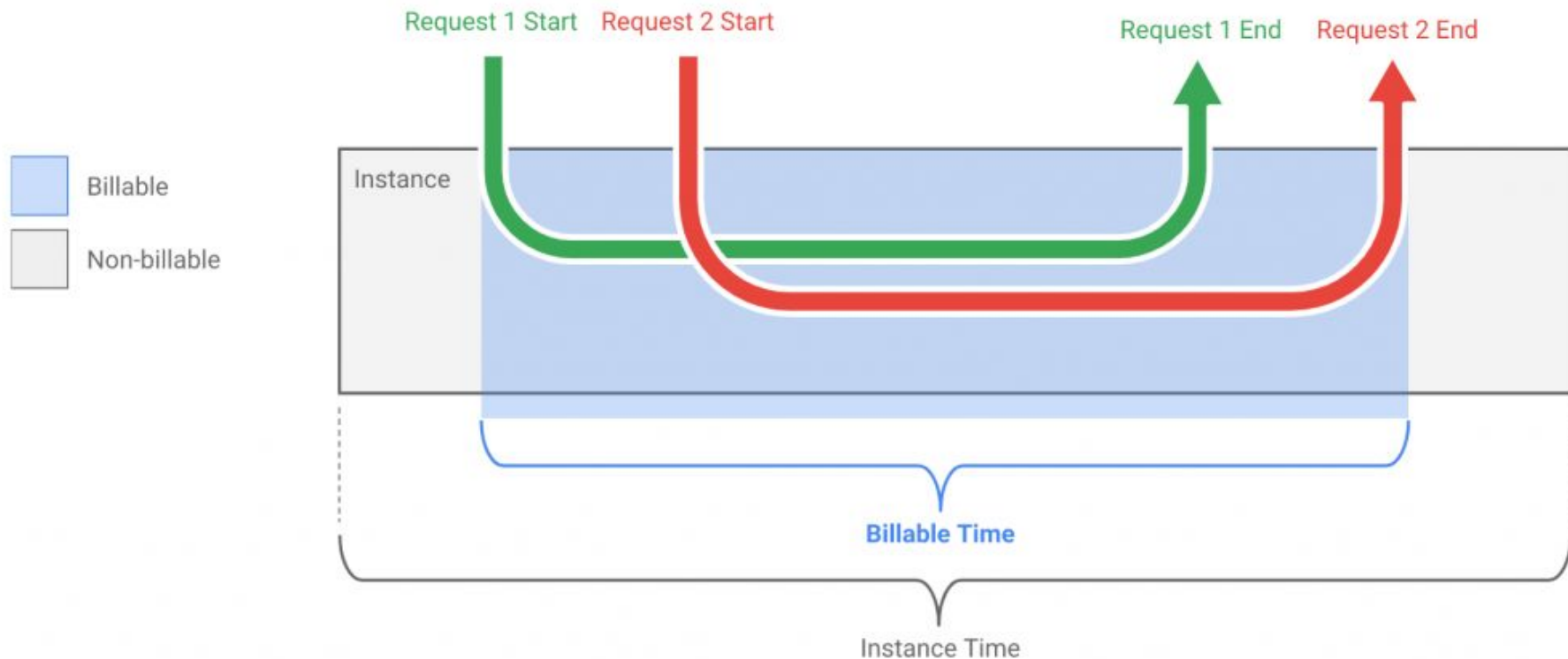


CPU / Memory / Requests



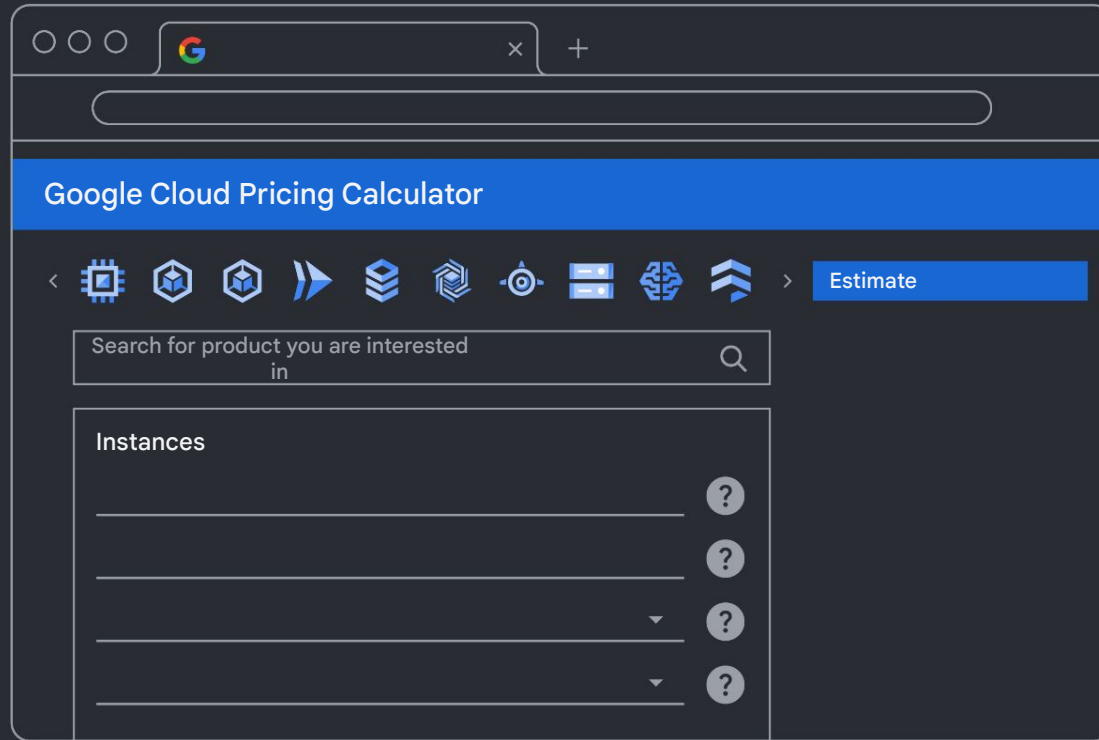
100ms

Cloud Run – Billable container instance time



#5 Serverless Pricing Model

Ce qui existe en standard...
Pour mesurer et comprendre

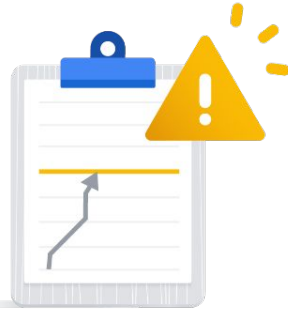


cloud.google.com/products/calculator

Billing tools help to budget and monitor usage



Budgets



Alerts



Reports



Quotas

Ce qui existe en standard...

Free tiers

Free tier

Compute Engine

Machines virtuelles évolutives et hautes performances

1 instance e2-micro par mois



Cloud Storage

Des performances exceptionnelles, une fiabilité exemplaire et des tarifs compétitifs pour tous vos besoins de stockage

5 Go-mois de stockage standard



BigQuery

Entrepôt de données d'analyse entièrement géré, à l'échelle du pétaoctet

1 To de requêtes par mois



Google Kubernetes Engine

Orchestration de conteneurs en un clic via des clusters Kubernetes, gérés par Google

Un cluster Autopilot ou zonal par mois



App Engine

Plate-forme pour créer des applications Web et des backends mobiles adaptables

28 instances-heure par jour



Cloud Run

Un environnement entièrement géré permettant d'exécuter des conteneurs sans état

2 millions de requêtes par mois



Cloud Build

Des compilations rapides, cohérentes et fiables sur Google Cloud

120 compilations-minute par jour



Suite Operations (anciennement Stackdriver)

Surveillance, journalisation et diagnostics pour les applications sur Google Cloud

Attributions mensuelles pour la journalisation et la surveillance



Firestore

Base de données de documents NoSQL qui simplifie le stockage, la synchronisation et l'interrogation des données d'applications

1 Go d'espace de stockage



Pub/Sub

Un service de messagerie et de diffusion des données fiable et en temps réel, à l'échelle mondiale

10 Go de messages par mois



Cloud Functions

Un environnement sans serveur pour créer et connecter des services cloud avec du code

2 millions d'appels par mois



Vision AI

Détection de thèmes, reconnaissance optique de caractères, détection de visages et bien plus encore

1 000 unités par mois



<https://cloud.google.com/free>



FinOps

<https://www.finops.org>



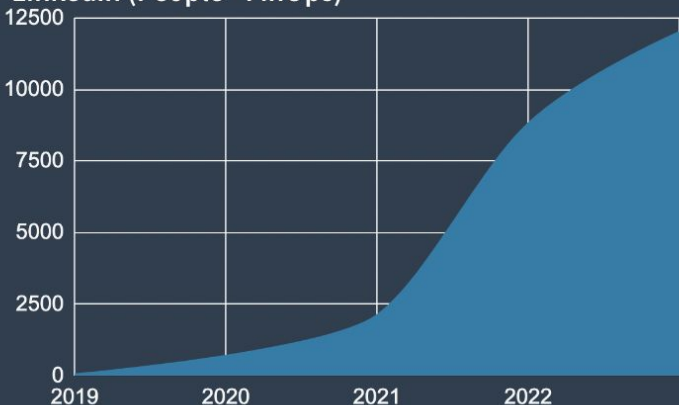
Trends



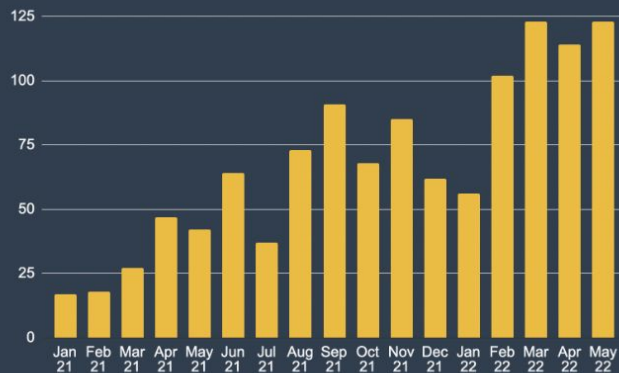
The FinOps industry has hit an inflection point this year... (we just launched jobs.finops.org too)

Social proof of those offering or seeking skills

Linkedin (People - FinOps)



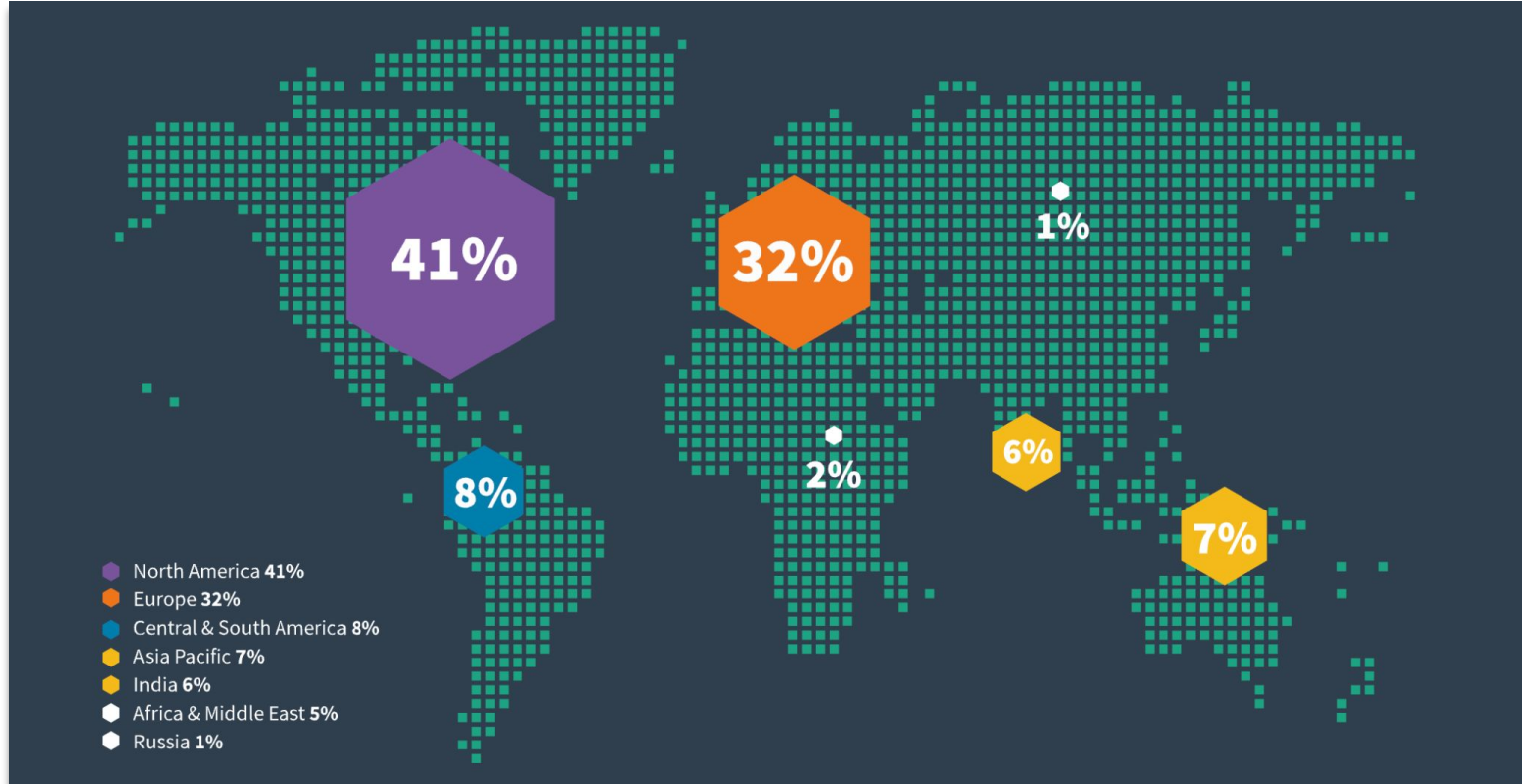
Posts in #jobs in the Slack 2021-2022



7000+ Community Members



Identified FinOps practitioners





Definition

L'approche **FinOps**, contraction des termes de finance et d'opérations, vise au monitoring et à l'optimisation des coûts en matière de cloud computing. Il est question, dans le même esprit que la stratégie DevOps, d'encourager la coopération et l'assimilation des moyens IT par les métiers. De ce fait, les pratiques FinOps ajoutent une dimension financière dans la gestion des coûts du Cloud.

DevOps

- ✗ a standard
- ✗ a certification
- ✗ a goal

- ✓ a culture
 - ✓ a philosophy
 - ✓ a journey
-

Key principles in FinOps

1. Teams need to collaborate
2. Everyone takes ownership for their cloud usage
3. A centralized team drives FinOps
4. Reports should be accessible and timely
5. Decisions are driven by business value of cloud
6. Take advantage of the variable cost model of the cloud.

Cloud FinOps is an evolving cloud financial management discipline and cultural practice that:



enables organizations to get **maximum business** value



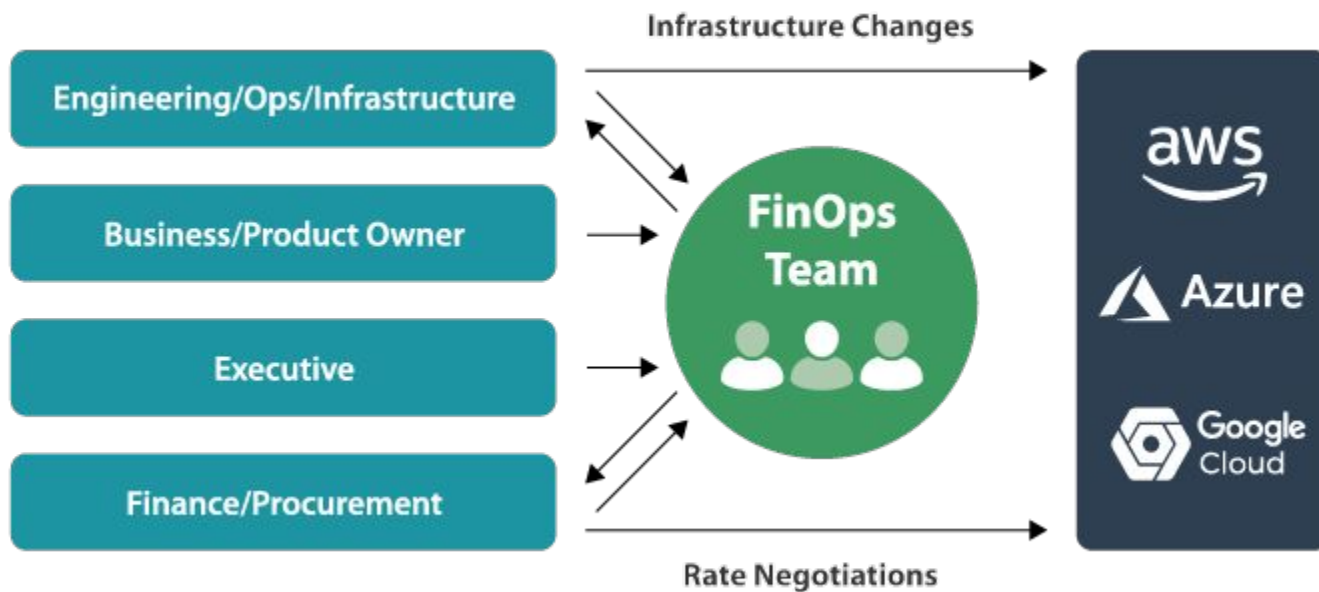
by **helping** engineering, finance and business teams

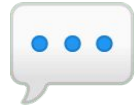


to **collaborate** on data-driven spending **decisions**

FinOps.org

FinOps Principles by FinOps Foundation - [FinOps Foundation under the ccbv4.0 license](#)





Ecosystem





12,000+

Strong Community

Join the
Community



7,500+

Certified Practitioners

Get Trained or
Certified



5,000+

Companies Participate

Enterprise
Membership



47

of the Fortune 50

Join the
Community



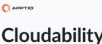
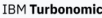
CLOUD COMPUTING > PERFORMANCE MANAGEMENT

FinOps Foundation Brings New Focus to Cloud Costs

The FinOps Foundation is joining the Linux Foundation to accelerate its efforts to bring the costs of the cloud in line with financial accounting within companies.

[Sean Michael Kerner](#) | Jun 30, 2020

25 Premier Members & 51 general members



Upskilling



FinOps Personas



🎓 Training & Certification

FinOps Certification and Training

Advance your knowledge of FinOps with a range of training and certification



♥ Community





Events



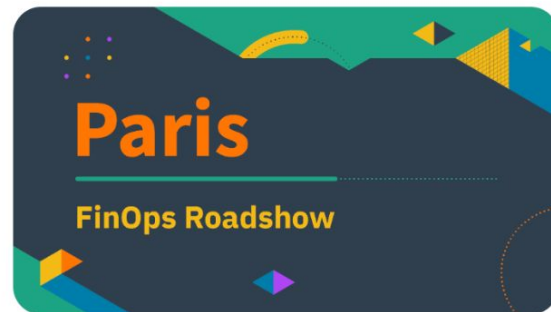
September FinOps Summit: FinOps X Sessions

Sep 14, 2023



Washington DC FinOps Roadshow

Oct 17, 2023



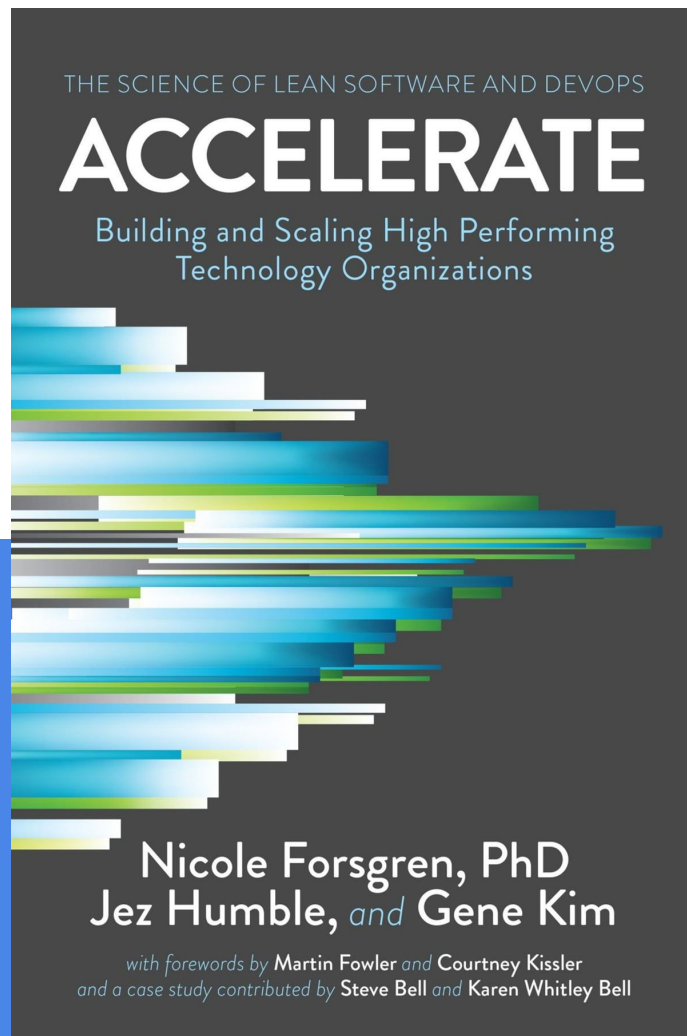
Paris FinOps Roadshow

Oct 18, 2023



Capabilities

Accelerate / State of DevOps Report



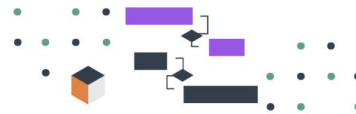
Domain & Capabilities



Understanding Cloud Usage and Cost



Performance Tracking & Benchmarking



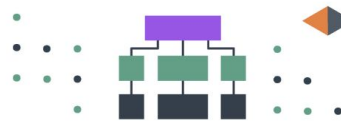
Real-Time Decision Making



Cloud Rate Optimization



Cloud Usage Optimization



Organizational Alignment

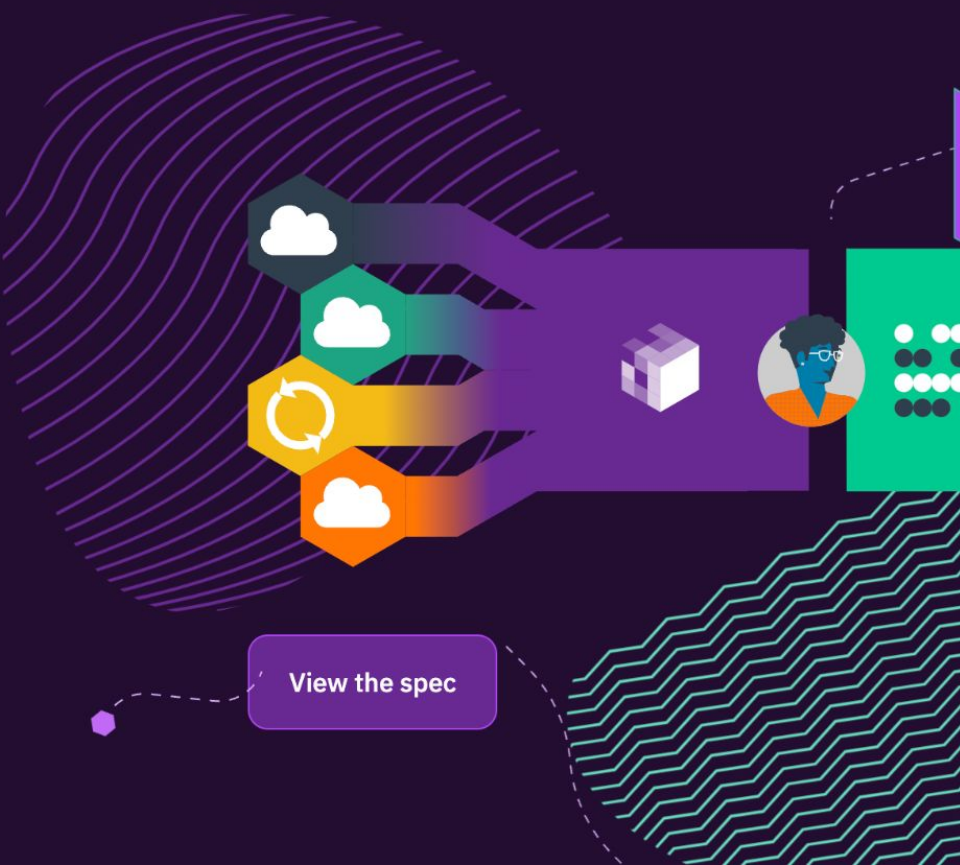


Tooling



FinOps Open Cost & Usage Specification

FOCUS™ is a technical specification to build and maintain an open standard for cloud cost, usage, and billing data.

[View the spec](#)



Love your cloud bill! 💰📈

Cloud cost estimates for Terraform in pull requests

Get started

- ✓ Open source
- ✓ Backed by Y Combinator & Sequoia



Infracost commented 2 hours ago

💰 **Monthly cost will increase by 27%** 📈

Infracost output ▼

~ ebs_volume.my_vol	
+ Provisioned IOPS	+\$286
- aws_lambda_function.my_func	-\$47

Amount: +\$239

📌 Join us Feb 2nd to learn about monitoring your K8s costs in different environments

Monitor & reduce Kubernetes spend

Kubecost provides real-time cost visibility and insights for teams using Kubernetes, helping you continuously reduce your cloud costs.

[▶ Watch a Demo](#)[Get Started](#)

Overview



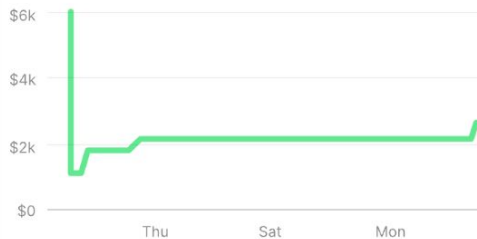
Monthly savings of \$1,886.98 identified [Learn More >](#)

Total Cost

\$3,310.28

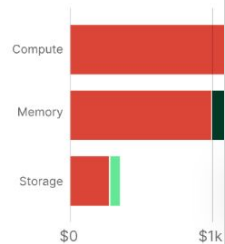
Monthly cluster costs

Snapshot of run rate expenses based on current market rates



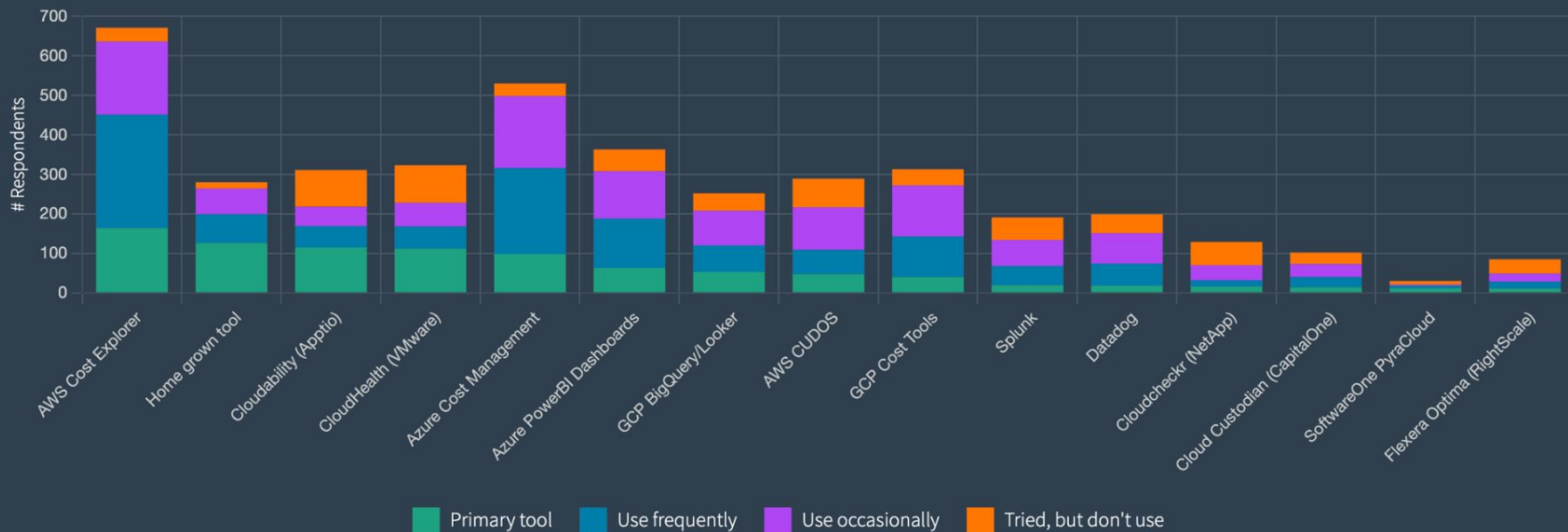
Resource Efficiency

Expense allocation based on currently



Mix of Tooling Used to Support FinOps [↗](#)

Having the right FinOps tools available can smooth the most tumultuous parts of a journey (especially as cloud usage, cost, and complexity increase). Here are the top tools our respondents report they use daily to manage and forecast cloud costs. n=838, multiple choice





Examples of measurement

BUSINESS (Sales based on the way products are sold)	FINANCIAL (Service component value indicator)	TECHNICAL (Sum of FinOps tactics and engineering choices at infra value layer)
Cost per weekly active user	Cost per core hour	Cost variance reductions
Cost per million transactions	Cost per GB of RAM	Spend risk reductions
Cost per million in revenue		



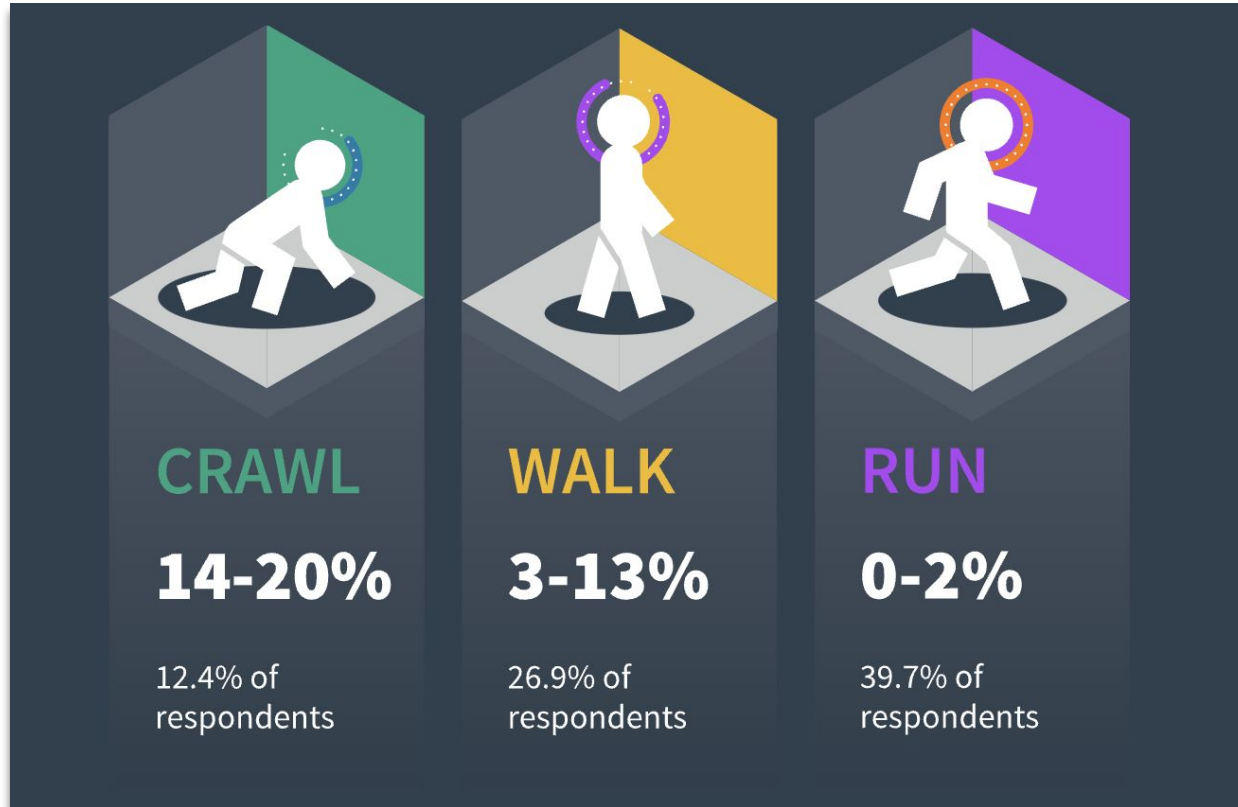
Adopt



⚡ What is the difference between forecasted spend and actual spend?



Percentage of commitment-based discounts that are estimated to go unused each month





How do you promote FinOps adoption in the organization? [↗](#)

Crawl-stage respondents are in a learning state and rely on organic adoption, whereas run-stage respondents push executive directives, grass-roots initiatives, gamifications, and reward systems.

n=692

[VIEW 'OTHER' ANSWERS →](#)



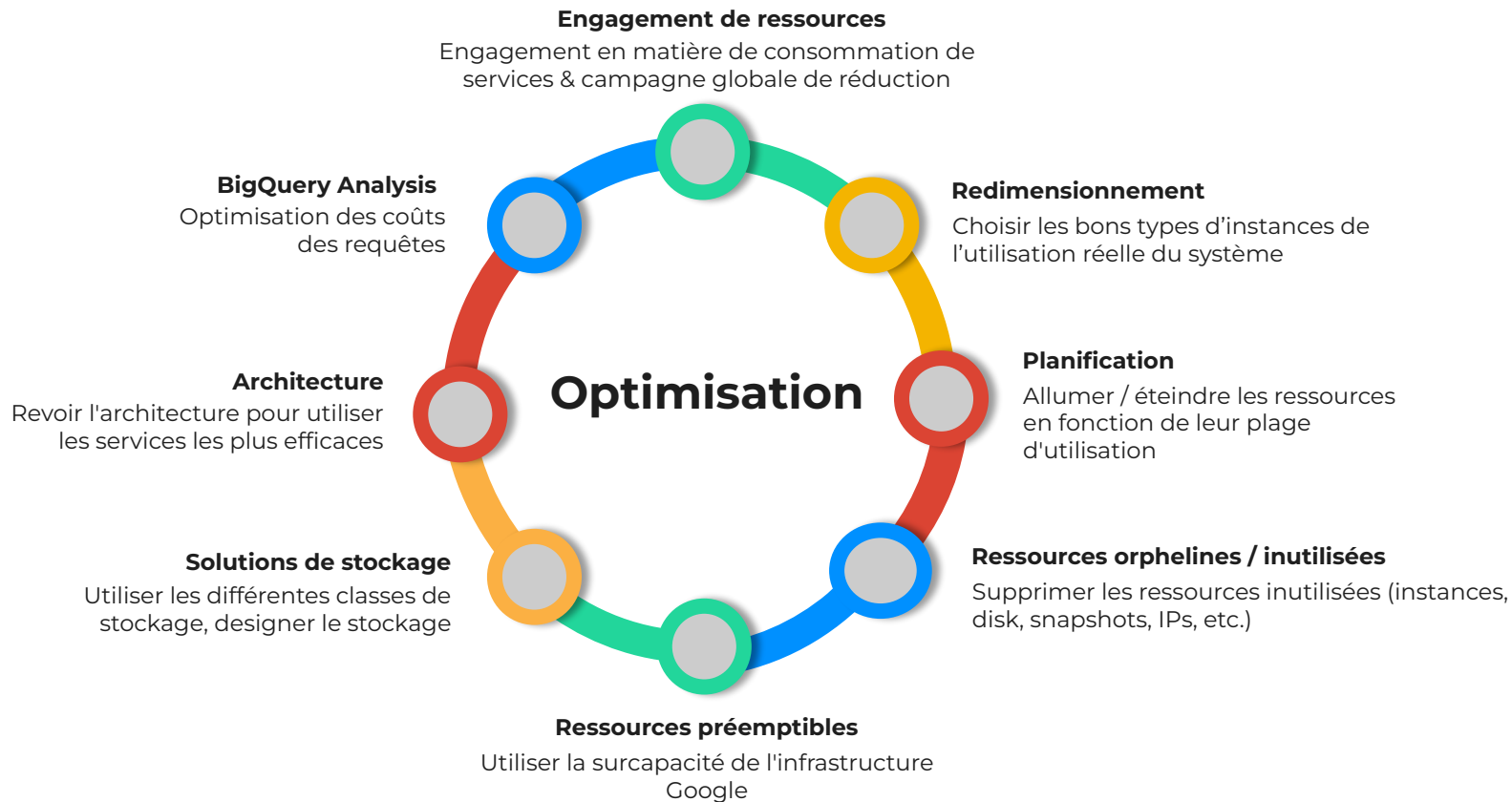
Pourquoi s'y mettre ?

Value of Cloud FinOps

1. Accelerate business value realization and innovation
2. Drive financial accountability and visibility
3. Optimize cloud usage and cost efficiency
4. Enable cross-organizational trust and collaboration
5. Prevent sprawl of cloud spend

**Comment l'appliquer au
monde Google Cloud ?**

Des leviers disponibles



Cost & Value Optimization



I'm ready to start optimizing my cloud costs, but where do I begin?!?



Wait! It's not just about optimizing costs — it's about knowing where to invest to maximize value.



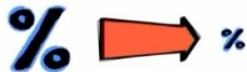
That's right! There are 3 key areas to optimize; Resources, Rate, and Architecture. Let's take a closer look...



1

OPTIMIZE YOUR RESOURCES

by eliminating idle resources, using autoscaling, and rightsizing your compute & databases for the best fit. Work with your Product and Engineering teams to take these steps.



2

OPTIMIZE YOUR RATES

by leveraging resource & spend-based committed use, BigQuery editions and volume & contract discounting. A Cloud FinOps team can help to save by making decisions centrally.



3

OPTIMIZE YOUR ARCHITECTURE

by embracing managed services & serverless technologies like AutoML, Cloud Run, Vertex AI and more. Use these technologies to focus on your business value and go-to-market capability.

HIGH EFFORT & LOW SAVINGS

HIGH EFFORT & HIGH SAVINGS

Rightsizing
(Post-Deployment)

Re-Platforming

Re-Architecting

Spot
VMsStorage
Lifecycle
Management

Autoscaling

Managed Services
& Serverless for
New WorkloadsRightsizing
(Pre-Deployment)

Idle Resources

Custom
Machine TypeCommitted
Use Discounts

Big Query Slots

LOW EFFORT & LOW SAVINGS

LOW EFFORT & HIGH SAVINGS

Check out this Cost Optimization Matrix! It really reveals how options & effort can greatly affect your bottom line!



Oui d'accord et ?



Exemple 1 : Firebase





Flavien Normand for Zenika

Posted on Jan 18, 2021 • Updated on Jan 25, 2021 • Originally published at blog.zenika.com



1

Optimisations Firebase, moins cher et plus performant

[#firebase](#) [#finops](#)

Exemple 2 : GKE





<https://brennrm.github.io/posts/wasting-money-with-kubernetes.html>

FinOps Kubernetes : 10 indicateurs à surveiller pour rationaliser les coûts



Benoit 🐱💜 COUETIL
Ingénieur DevOps/Delivery

11 articles

✓ Following



GKE Enterprise annoncé à Next

Introducing GKE Enterprise edition

Google Cloud

gle-enterprise-101

Search (f) for resources, docs, products, and more

Search

SEND FEEDBACK

REFRESH

Kubernetes Engine

ENTERPRISE

Fleet

gle-enterprise-101-fleet

Resource Management

Overview

Clusters

Workloads

Teams

Services & Ingress

Applications

Secrets & ConfigMaps

Storage

Object Browser

Backup for GKE

Features

Service Mesh

Security Posture

Config

Marketplace

Release Notes

Overview (gke-enterprise-101-fleet)

Feature management

Policy Controller status

6 clusters

6 without Policy Controller

0 without Google constraints

0 without Google constraint library violations

0 with Google constraint library violations

Automate and assess compliance across your clusters with Google Policy Essentials

Try Policy Controller

Config Delivery

0 packages

0 error

0 stalled

0 pending

0 reconciling

0 synced

0 unknown

Use Config Delivery to manage configurations consistently across clusters and environments.

Enable Config Sync

Fleet efficiency

CPU utilization by fleet

20 CPUs

15 CPUs

10 CPUs

5 CPUs

0 CPUs

Allocatable: 8.46 CPUs

Requested: 4.24 CPUs

Memory utilization by fleet

50GB

40GB

30GB

20GB

10GB

0GB

Allocatable: 26.52GB

Requested: 6.89GB

Disk utilization by fleet

800GB

600GB

400GB

200GB

0GB

Allocatable: 394.45GB

Used: 0.08GB

Fleet Management

2:59 / 6:15

Scroll for details

CC

HD

Exemple 3 : CUD



Committed Use Discounts (CUDs)

Resource-based CUDs

Ex: 50 vCPU pour N2D dans us-central1

Services pris en charge : Compute Engine, Dataproc, Google Kubernetes Engine (Standard)

Spend-based CUDs

Exemple: \$50/hour spend in Cloud SQL (PgSQL) dans us-central1

Services pris en charge : Cloud SQL, Cloud Spanner, Cloud Run, Google Cloud VMware Engine, Google Kubernetes Engine (Mode Autopilot)

Billing

Billing account

Committed use discount test

Overview

Reports

Cost table

Cost breakdown

Commitments

Commitment analysis

Budgets & alerts

Billing export

Pricing

Transactions

Payment settings

Account management

Committed use discounts

PURCHASE

RECOMMENDATIONS

This is a summary of your committed use discounts by commitment type. [Learn more](#)

Filter

State : Active

Enter property name or value

Type	Region	Scope	Active commitment	State	Term	Start date	End date	Recommendation	Trailing 30-day commitment cost	Trailing 30-day savings	Trailing 30-day utilization	Trailing 30-day coverage
▼ Compute Engine Resource-based us-central1	us-central1											
▼ Accelerator Optimized GPU	us-central1		1 GPUs									
commitment-2 (NVIDIA Tesla A100)	us-central1	billingAccounts/01667...	1 GPUs	Active	3 years	2021-06-12	2024-06-12					
▼ General Purpose Cores	us-central1		9 vCPU									
commitment-2	us-central1	billingAccounts/01667...	1 vCPU	Active	1 year	2021-03-20	2022-03-20					
commitment-2	us-central1	billingAccounts/01667...	1 vCPU	Active	3 years	2020-07-21	2022-07-25					
commitment-1	us-central1	billingAccounts/01667...	4 vCPU	Active	3 years	2021-05-28	2024-05-28					
commitment-3	us-central1	billingAccounts/01667...	3 vCPU	Active	3 years	2021-07-10	2024-07-10					
▼ General Purpose Memory	us-central1		6.5 GB					💡 SAVE \$13.15/MO				
commitment-2	us-central1	billingAccounts/01667...	6.5 GB	Active	3 years	2020-07-21	2022-07-25					
▶ Compute Engine Resource-based us-east1	us-east1											
▼ Cloud SQL Database VM	europa-north1	billingAccounts/01667...	\$0.01/hr						\$5.40	-\$5.40		0.00% 0.00%
Cloud SQL Database VM	europa-north1	billingAccounts/01667...	\$0.01/hr	Active	1 year	2020-12-30	2021-12-30					
▼ Cloud SQL Database VM	europa-west1	billingAccounts/01667...	\$0.07/hr					💡 SAVE \$22.46/MO	\$28.54	\$27.38		100.00% 57.49%
name	europa-west1	billingAccounts/01667...	\$0.06/hr	Active	3 years	2021-08-26	2024-08-26					
Recommended 3 year Cloud SQL Database VM comr	europa-west1	billingAccounts/01667...	\$0.01/hr	Active	3 years	2021-09-09	2024-09-09					
▼ Cloud SQL Database VM	us-central1	billingAccounts/01667...	\$0.02/hr					💡 SAVE \$666.43/MO	\$10.16	\$3.52		100.00% 1.02%
Cloud SQL Database VM	us-central1	billingAccounts/01667...	\$0.01/hr	Active	1 year	2020-12-28	2021-12-28					
Recommended 1 year Cloud SQL Database VM comr	us-central1	billingAccounts/01667...	\$0.01/hr	Active	1 year	2021-09-08	2022-09-08					

Recommended Compute Engine Commitments

You could be saving money with Compute Engine Commitments. Turn on discount sharing to see additional savings from all applicable cost recommendations. [Learn more](#)

Cost savings

\$79.30/month

Recommendations

DISMISS

Filter Enter property name or value

Est. savings ↓	Recommendation type	Region	Project
☐ Save \$43.95/month	New commitment	europe-west3	ze
☐ Save \$22.08/month	New commitment	europe-west3	ze
☐ Save \$8.64/month	New commitment	us-west1	ze
☐ Save \$4.63/month	New commitment	us-west1	ze

Google Cloud Platform

Recommended Compute Engine Commitments

Commitment scope Projects Edit

Compute Engine
Based on your historical data, you could be saving money with Compute Engine Commitments lower the cost of stable Compute Engine usage.

Filter recommendations

Est. savings ↑	Recommendation type
Save \$4,560/month	New commitment
Save \$2,942/month	Additional commitment
Save \$2,190/month	Additional commitment
Save \$2,100/month	New commitment
Save \$1,021/month	New commitment

Save with a new commitment in us-east1

I want this commitment to...

☒ Maximize savings

☐ Cover minimum stable usage

Recommended commitment

Duration

☐ 1-year ☒ 3-year

Scope

Project : project-bat

Region

us-east1 (South Carolina)

Type

General purpose

vCPUs

100

Memory

65 GB (Standard ratio)

Estimated savings

Based on your December 2018 Compute Engine usage

Est. 3-year savings

\$164,160.00

Actual December 2018 cost

\$10,857.14

Est. cost with recommended commitment

\$6,297.14

Est. savings with recommended commitment

\$4,560.00

us-east1 Compute Engine

vCPU

Last month Last 3 months Last year

Daily average vCPU usage



December 2018 costs



Réserver un engagement de 3 ans pour économiser \$164,160 sur ce projet

CONTINUE TO PURCHASE

Select commitment type

You have usage eligible for commitments on the following:

Commitment type

Compute Engine Flexible Committed Use Discounts

AlloyDB UCS Subscription

Cloud Spanner Committed Use Discounts

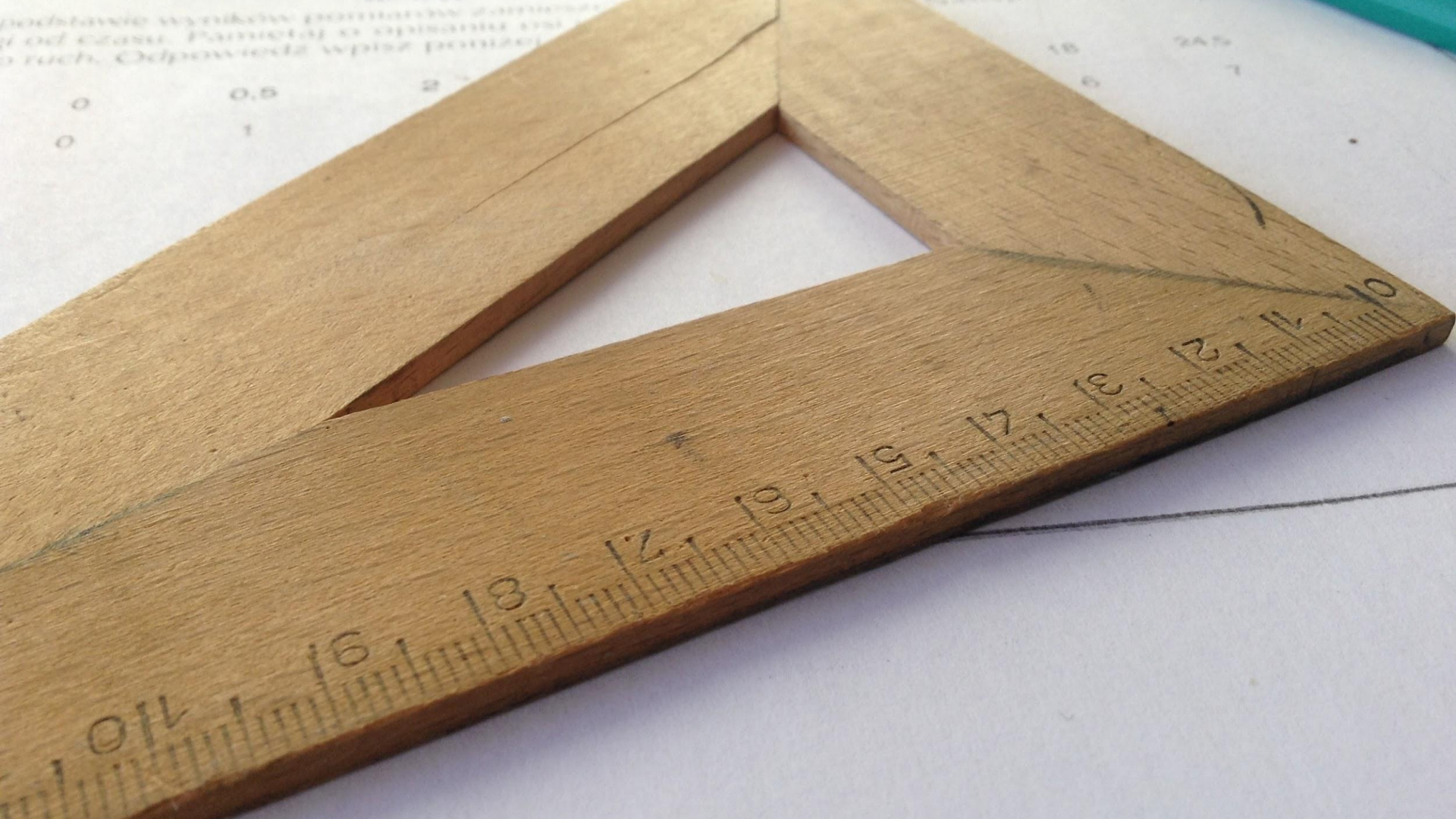
Cloud SQL Database VM

Kubernetes Engine Autopilot Committed Use Discounts

Cloud Run Committed Use Discounts

Exemple 4 : Cloud Native Dev

podstawie wyników pomiarów zamieszczonego od czasu. Pamiętaj o opisanu osi o ruch. Odpowiedź wpisz poniżej.



Développer en respectant les Cloud Patterns



Cloud Patterns : Writing Applications That Scale - Julien Landuré - WEB2DAY 2017

Web2day • 200 vues • il y a 5 ans

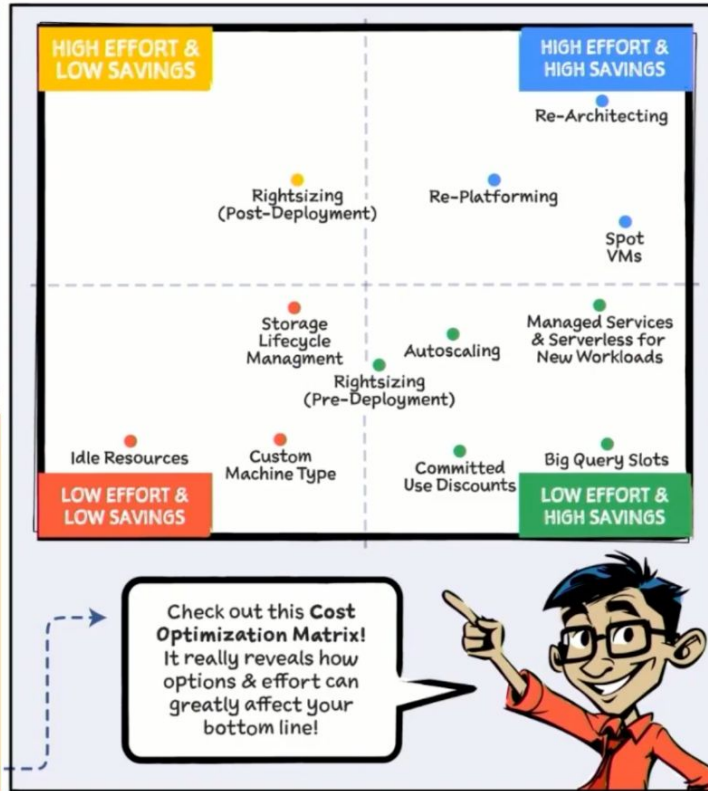
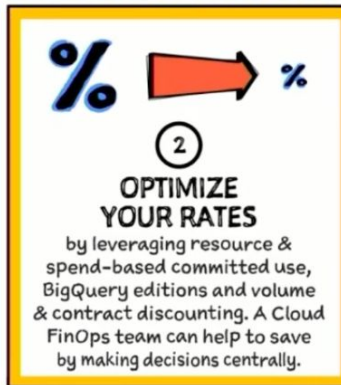


[DevFest Nantes 2016] Cloud patterns

GDG France • 329 vues • il y a 6 ans

Conclusion

Cost & Value Optimization



Les points à retenir

- 👍 Un sujet tendance avec de vrais concepts
- 🔑 Démarche essentielle comme le SRE et la Sécurité
- 💪 Une culture “FinOps” à diffuser
 - Autoscaling de 0 à N – par env (hors-prod / prod)
 - Respect de 12-Factors côté dev
 - Profiter au maximum du pricing model du Serverless
 - Choisir le bon outil pour faire la bonne chose
 - Observer, mesurer et maîtriser son architecture
 - Savoir quand se “commit” et sur quelle durée



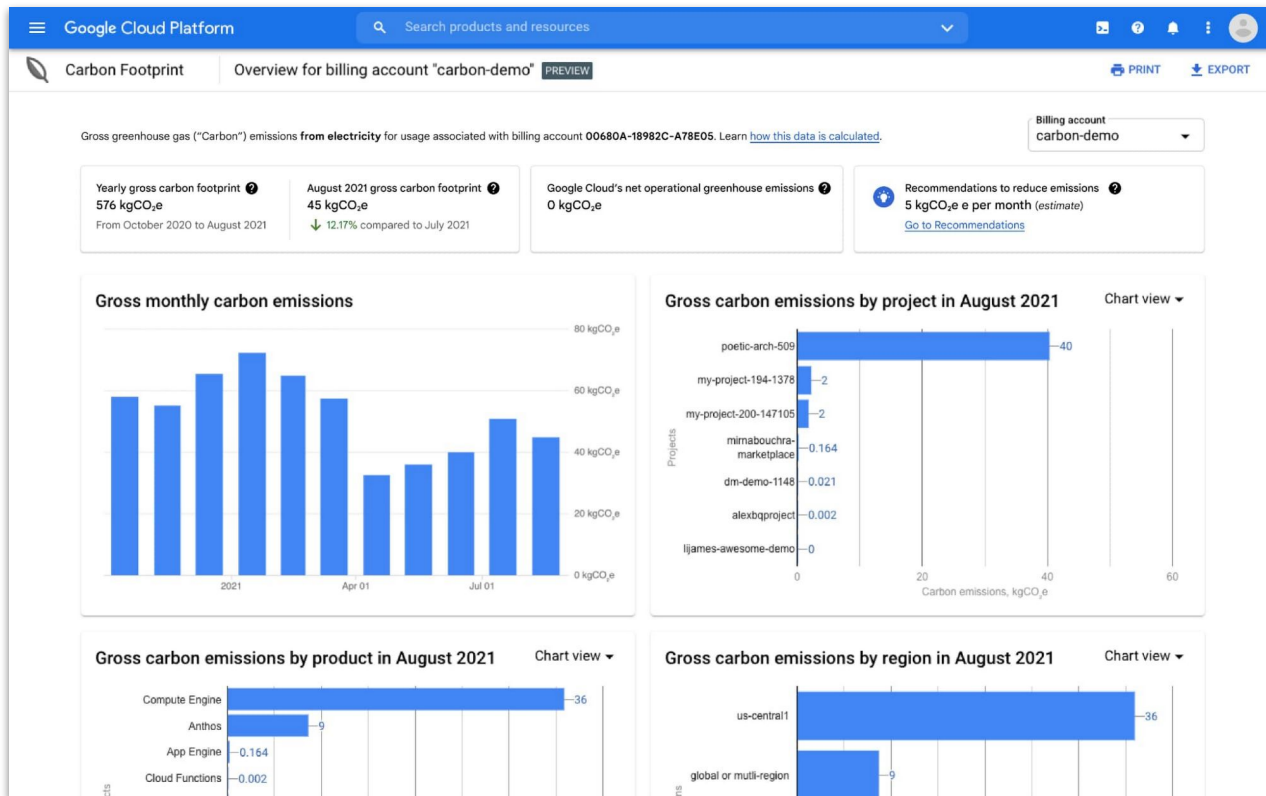
Next



#GreenIT #GreenOps



Carbon Footprint



<https://greensoftware.foundation>

[About](#)[Working Groups](#)[Projects](#)[Resources](#)[Articles](#)[JOIN US](#)

We are building a trusted ecosystem of people,
standards, tooling and best practices for

GREEN SOFTWARE





Merci !

Introduction au FinOps avec Google Cloud ☁️



jlandure.dev
@jlandure