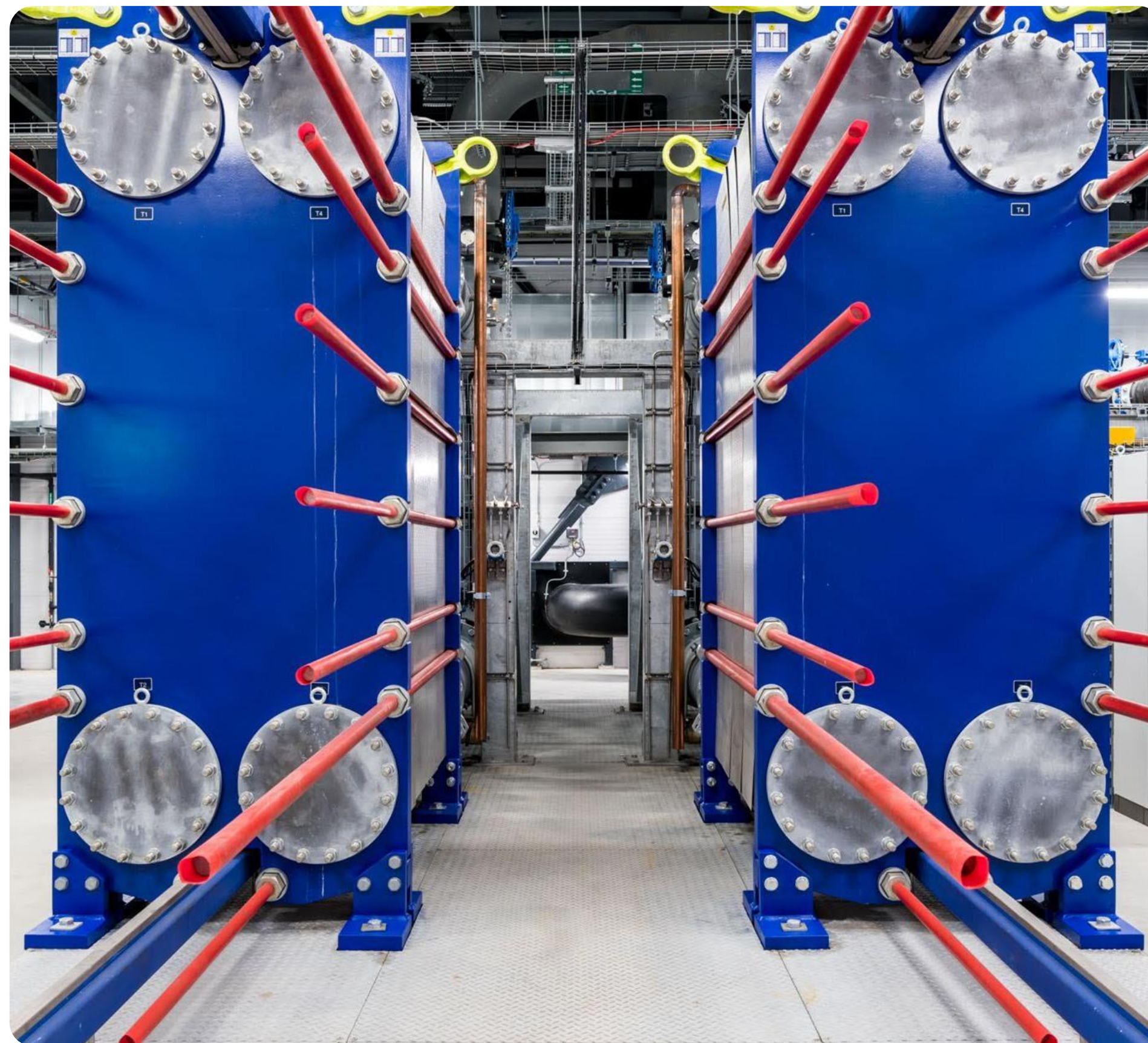





# Intel 如何結合 Google Cloud 進行全新數位轉型

Benson Yeh  
Senior Customer Engineer  
Google Cloud



# Intel + Google Cloud Solution pillars

Infrastructure & Application modernization

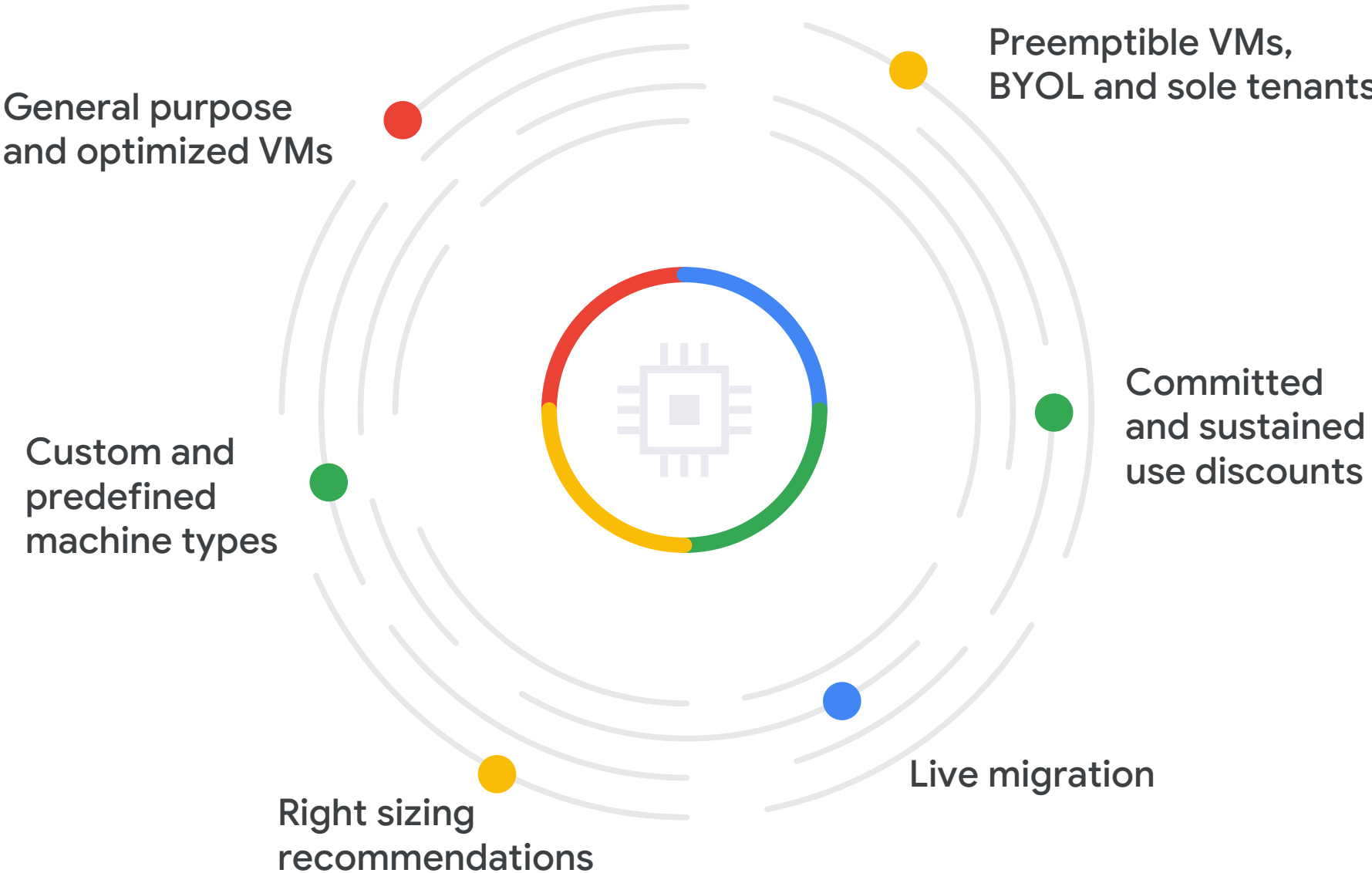
Pillar	Infrastructure Modernization		Application Modernization	
Solution	<b>1</b> Compute	<b>2</b> Enterprise Workloads	<b>3</b> Anthos Hybrid/ Multi-Cloud	<b>4</b> 5G/LTE Edge solutions for industry
Objectives	<p>Accelerate consumption of <b>General Purpose</b> (N1/N2), <b>Compute Optimized</b> (C2) and <b>Memory Optimized</b> (M1/M2) Instances</p> <p>Drive VM Migration with <b>Partners</b> to grow GCP revenue</p> <p>Develop differentiated GCP services with <b>Intel Technologies</b></p>	<p>Develop and grow joint programs to drive Enterprise Workloads (<b>VMware, SAP, Oracle</b>) on GCP</p> <p>Collaborate to <b>optimize performance of enterprise workloads</b></p>	<p>Develop and grow strategic partnership to enable customers and partners to consume <b>joint Hybrid and Multi-cloud solutions with Anthos</b></p> <p> Anthos</p>	<p>Develop strategic partnership to <b>accelerate 5G/LTE and Edge solutions</b> for Enterprises/industry and drive monetization for telcos and the ecosystem</p>



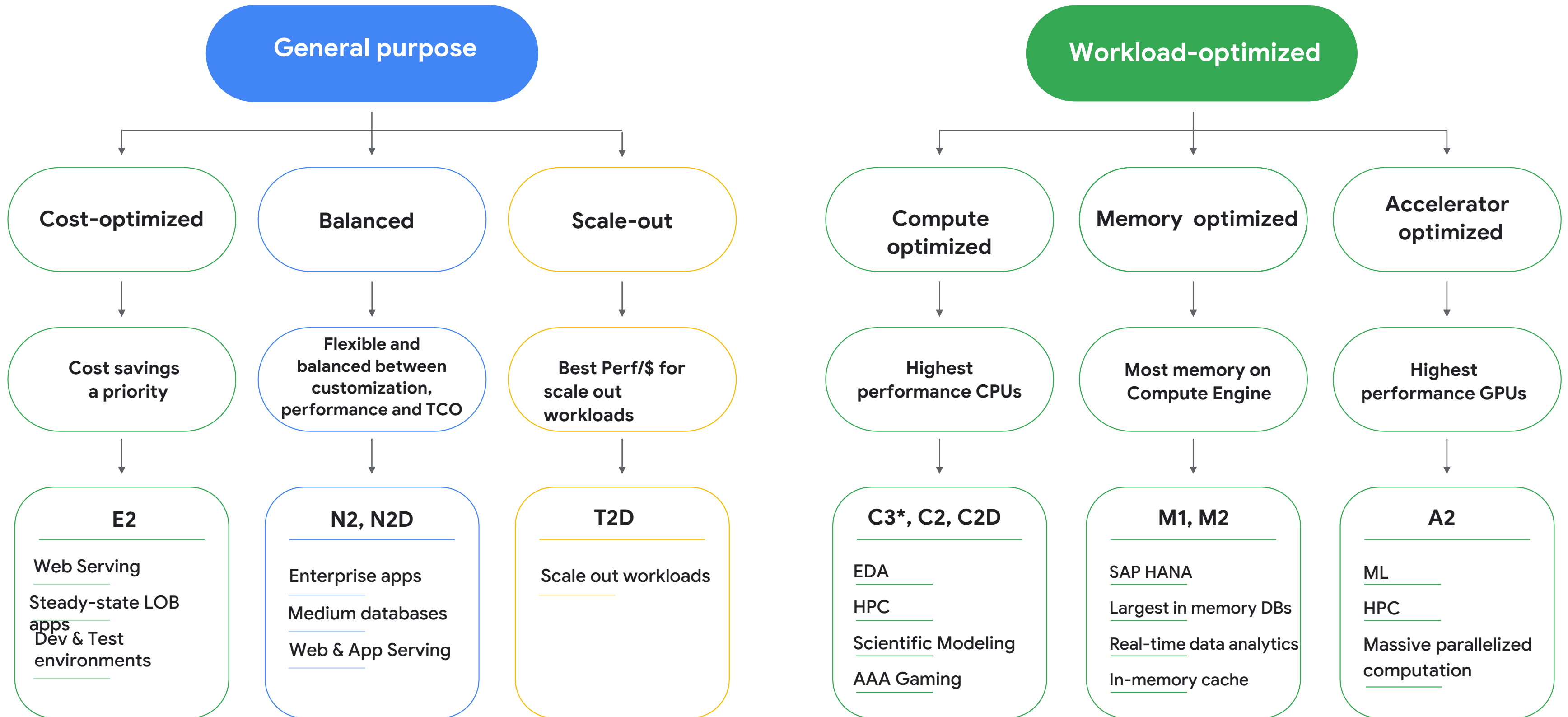
# Introducing Compute Engine

Compute Engine lets you create and run **virtual machines** on Google infrastructure.

Get access to a **variety of predefined and customizable VM families** coupled with consumption and pricing models, as well as functionality for **all of your application and workload requirements**.



# Google Cloud VM families in context





# General-purpose

Google Cloud



# General-purpose VMs: N2

Best fit for general purpose workloads that prioritize feature flexibility and leading price/perf

- ✓ Intel 2nd Generation Xeon Scalable Processors (IceLake & Cascade Lake with Up to 3.4Ghz all-core-turbo)
- ✓ Leading price/perf leveraging the latest hardware
- ✓ Enterprise application, medium databases, web & app serving
- ✓ Most flexible with the widest feature sets

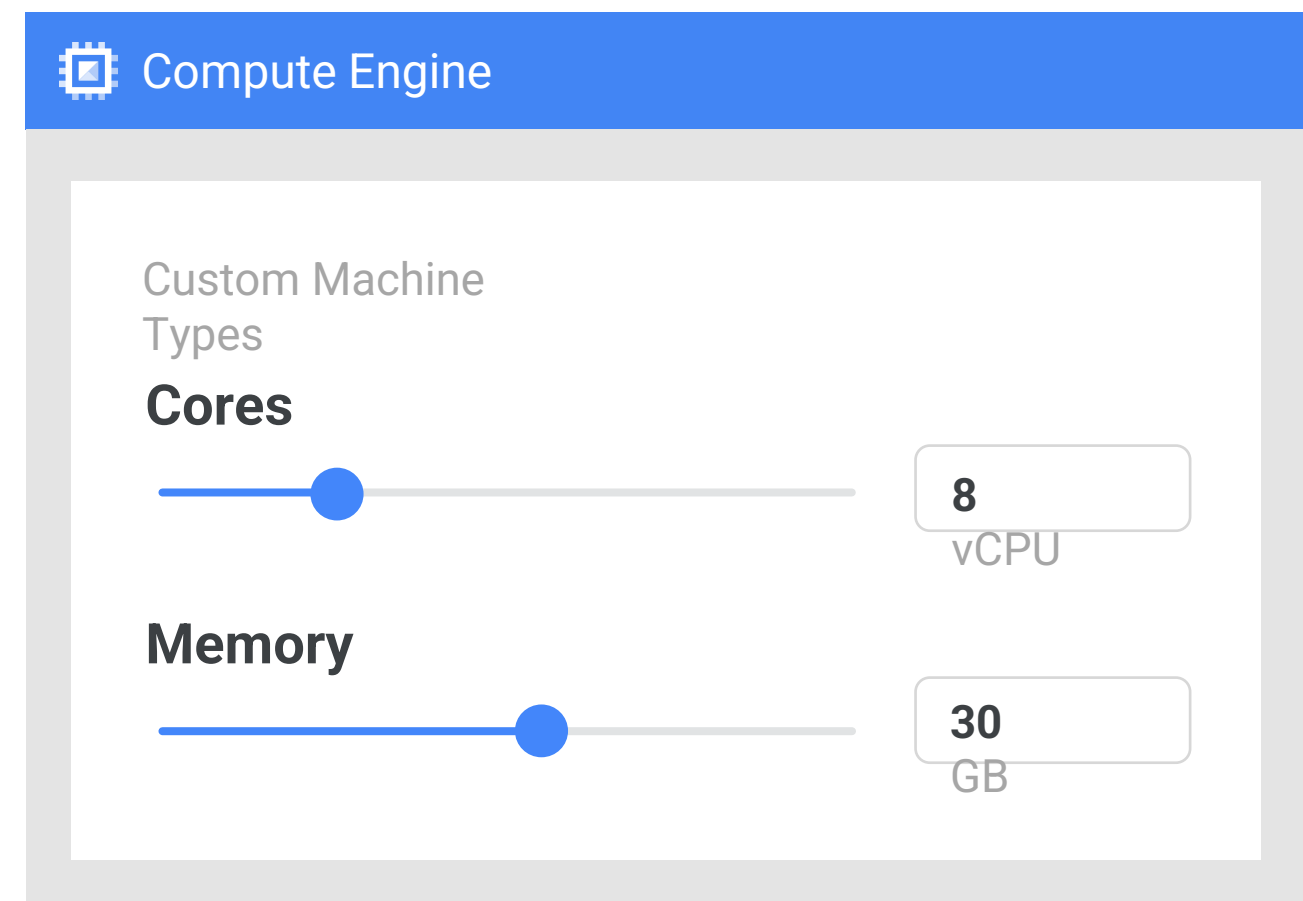


# Custom Machine Types

The exact fit for your resource needs

Configure your own machine types, with average 19% savings

- Create a machine type with 1 vCPU and up to 128 vCPU, 8GB memory per core.
  - Or any even number of vCPUs in between
- Use Compute Engine's stop/start feature
  - Move your workload to a smaller or larger Custom Machine Type instance or predefined configuration
- Priced by the resources they use (i.e. vCPU, memory)



Predictably calculate the cost of Custom Machine Type shapes in any configuration.

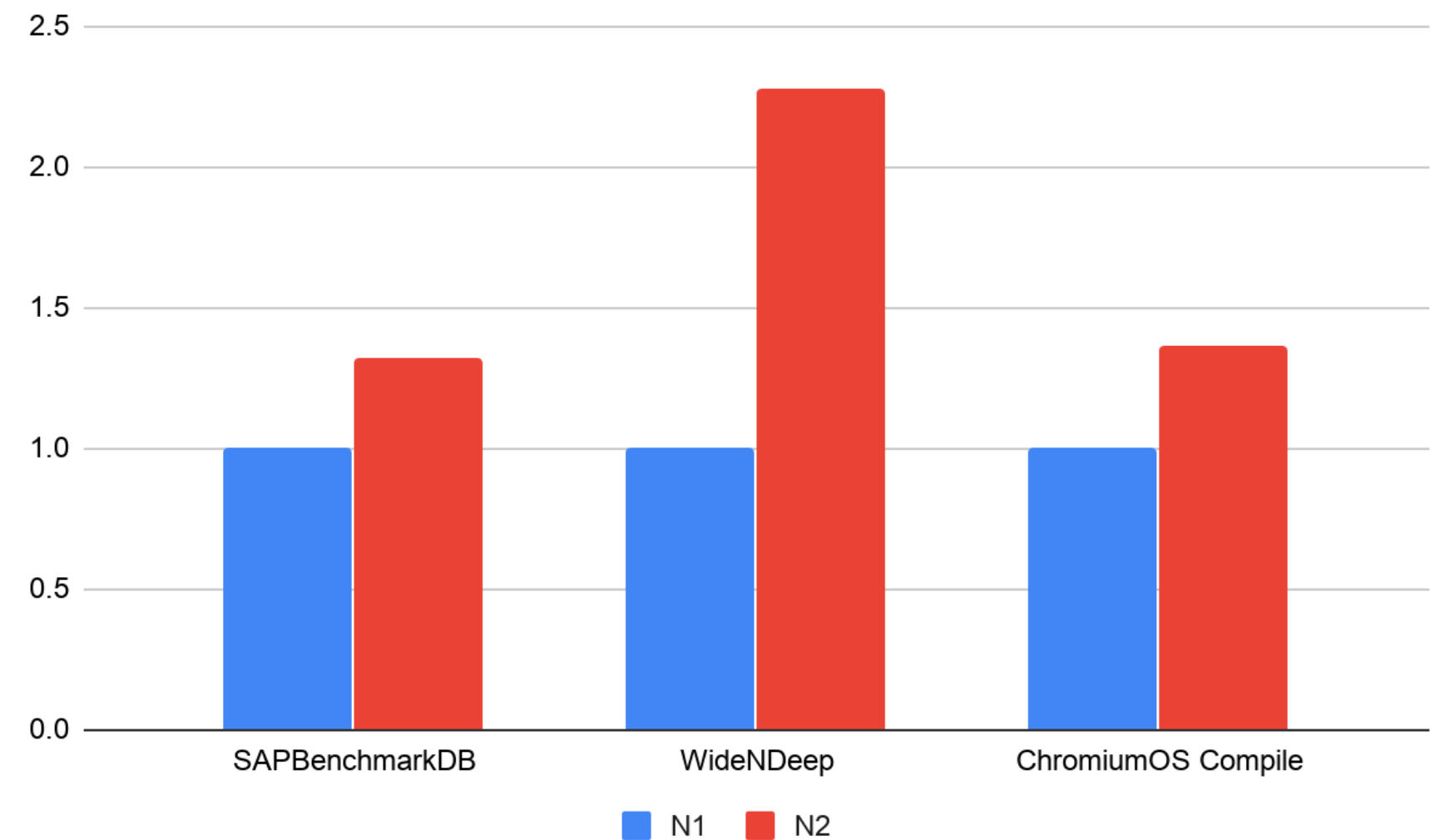
# General-purpose (N2) VM

## Benchmarks

N2 provides performance improvements for:

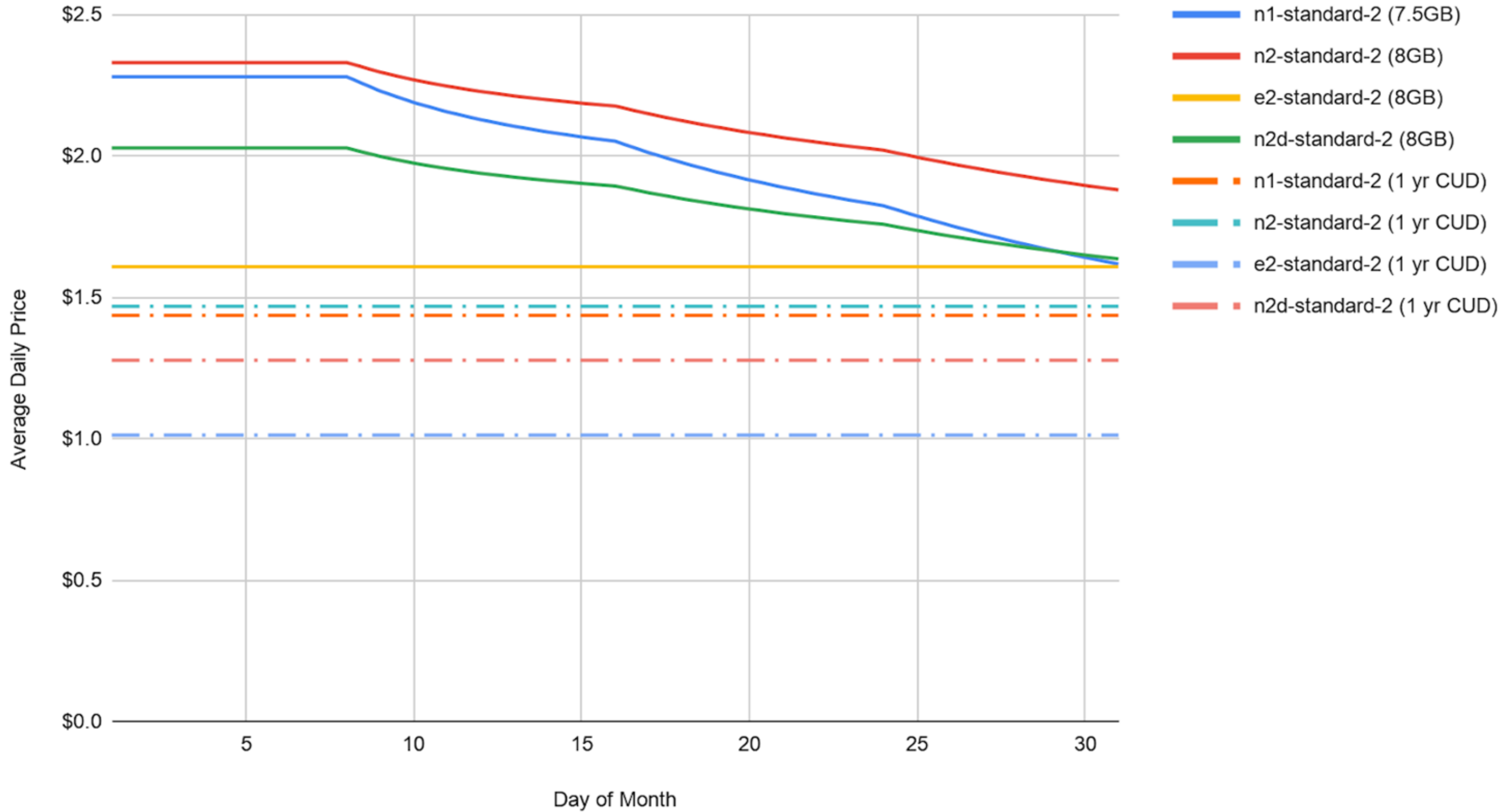
- **High end DBs including mid-sized SAP Hana and SQL instances**
- **General Purpose applications that can leverage N2's higher per core performance**
- **AI inference of a Wide and Deep model using Intel-optimized Tensorflow**

### N2 Relative Performance Improvements





### Average price based on lifetime of instance (after SUDs)





# Compute- optimized

Google Cloud



# Compute-optimized workloads : C2

Performance sensitive for CPU workloads, or licensed applications that may benefit from more powerful cores

- High-performance web servers
- AAA Gaming
- High Performance Computing (HPC)
  - Simulations, Financial modeling, Electronic Design Automation etc
- Media Transcoding



# Compute-optimized workloads

Performance sensitive for CPU workloads, or licensed applications that may benefit from more powerful cores

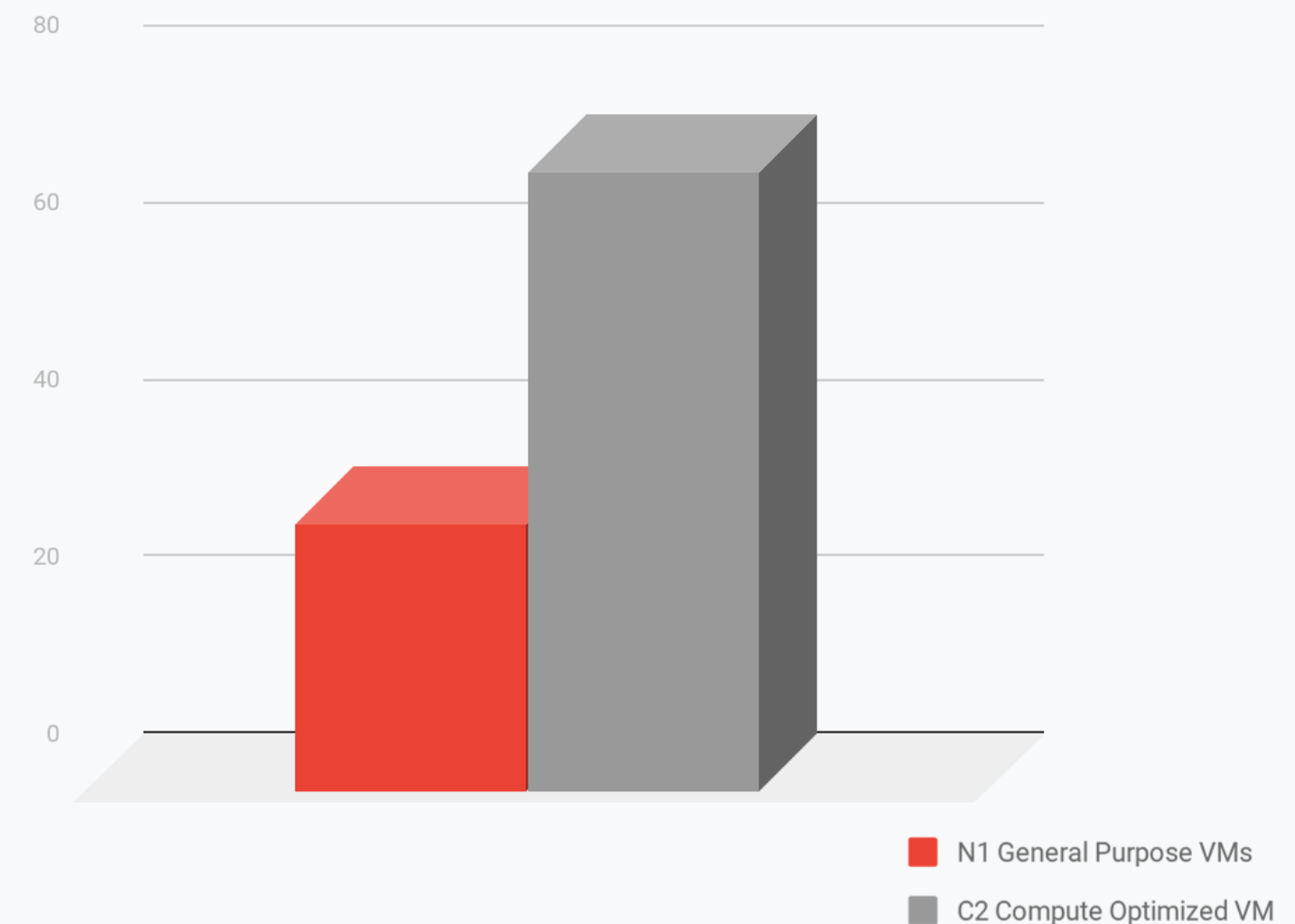
## Product Specs

- Intel Cascade Lake (14 nm)
- High frequency 3.9GHz all-core-turbo
- Fixed Sizes C2-standard-4/8/16/30/60
- 4 GB / vCPU ratio (up to 240 GB)
- Up to 3TB local SSD
- Up to 100 Gbps Network Bandwidth

## Highlights

- NUMA-aware for Performance
- Higher Performance per thread
- Isolation for latency sensitive workloads

40% higher performance  
per core



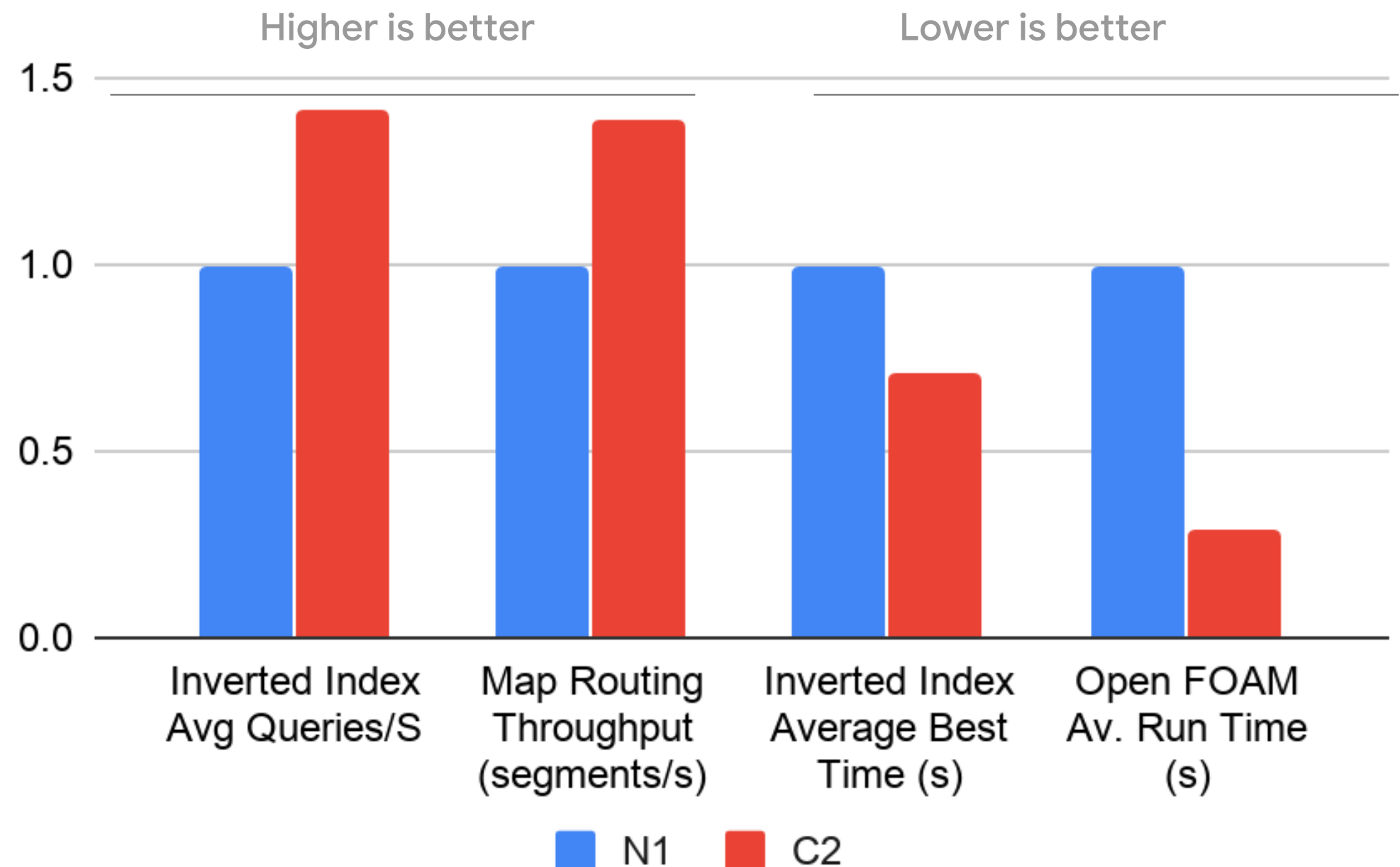
# Compute-optimized workloads

## Benchmarks

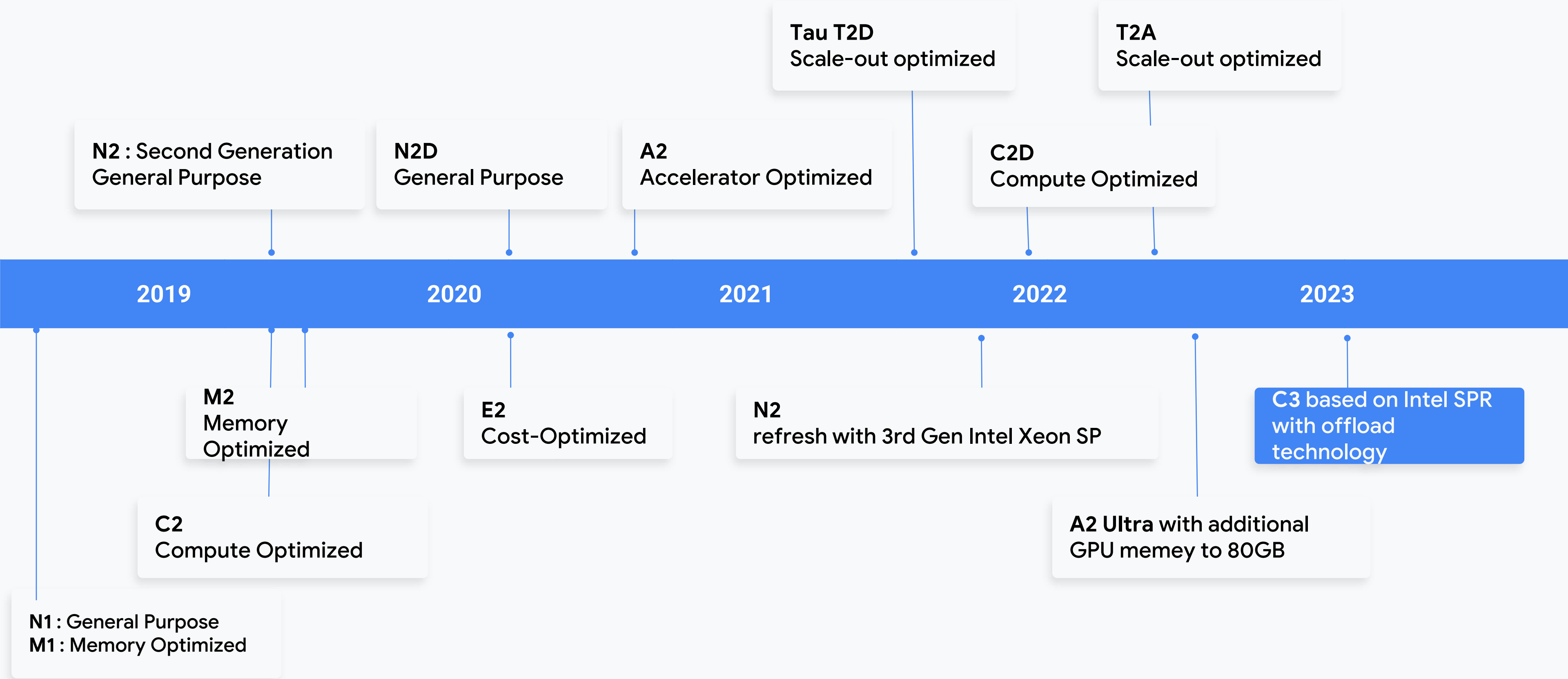
C2 provides great performance for

- **Highly interactive applications** including gaming and high perf web serving
- **HPC workloads** requiring high and consistent performance with visibility into the underlying hardware

## C2 Performance Improvements



# Compute Engine on Google Cloud



# Compute Optimized

## C3 on Intel Sapphire Rapids

Coming Soon

Designed to enable:

- High performance web and app serving, data processing workloads, and high throughput HPC workloads

Features:

- 4th generation Intel<sup>®</sup> Xeon<sup>®</sup> Scalable
- New: co-designed IPU with Intel
- New: 200 Gbps low latency networking
- New: up to 10x faster storage IOPS than previous gen



# C3 VMs

## Public Preview (us-central1, us-east4, europe-west4)

Consistently high performance for a broad range of general-purpose workloads.

- Latest Intel Sapphire Rapids CPU with up to 176 vCPUs
- VM shapes optimized for NUMA isolation, performance and consistency.
- Google's custom Intel Infrastructure Processing Unit (IPU) with up to 200Gbps
- Up to 4x higher throughput and 10x higher IOPS than previous gen C2 with Hyperdisk (coming March/April)
- Improved maintenance experience

### C3 highcpu shapes

Machine types	vCPUs	Memory (GB)	Network Bandwidth (Gbps)	Storage IOPS (PD-SSD/ Hyperdisk Extreme)	Storage Throughput (MB/s) (PD-SSD/ Hyperdisk Extreme)
c3-highcpu-4	4	8	Up to 23	15,000	240
c3-highcpu-8	8	16	Up to 23	15,000	240
c3-highcpu-22	22	44	Up to 23	25,000	800
c3-highcpu-44	44	88	Up to 32	60,000	800
c3-highcpu-88	88	176	Up to 62	80,000 / 350,000	1,200 / 5,000
c3-highcpu-176	176	352	Up to 100	80,000 / 350,000	1,200 / 5,000

Standard (4GB/vCPU) and highmem (8GB/vCPU) coming Q2'23







## Snap

"We were impressed to see an approximate 20% increase in C3 performance over the current generation C2 VMs from Google Cloud for a key workload."

– Aaron Sheldon, Sr. Software Engineer



## Parallel Works

"Based on the initial performance data, WRF on C3 can deliver as much as 10x faster results for about the same computational cost. This will significantly accelerate R&D for our customers in weather, environment, and engineering domains."

– Michael Wilde, CEO



Parallel Works



## ANSYS

"In early testing with our flagship products, including Ansys Fluent, Mechanical and LS-DYNA, on the new Google Cloud C3 VMs, we're seeing up to 3x performance gains over C2 VMs due to higher memory bandwidth and lower network latency."

– Wim Slagter, Strategic Partnerships Director at Ansys





# Integrated with Google Cloud

Google Cloud



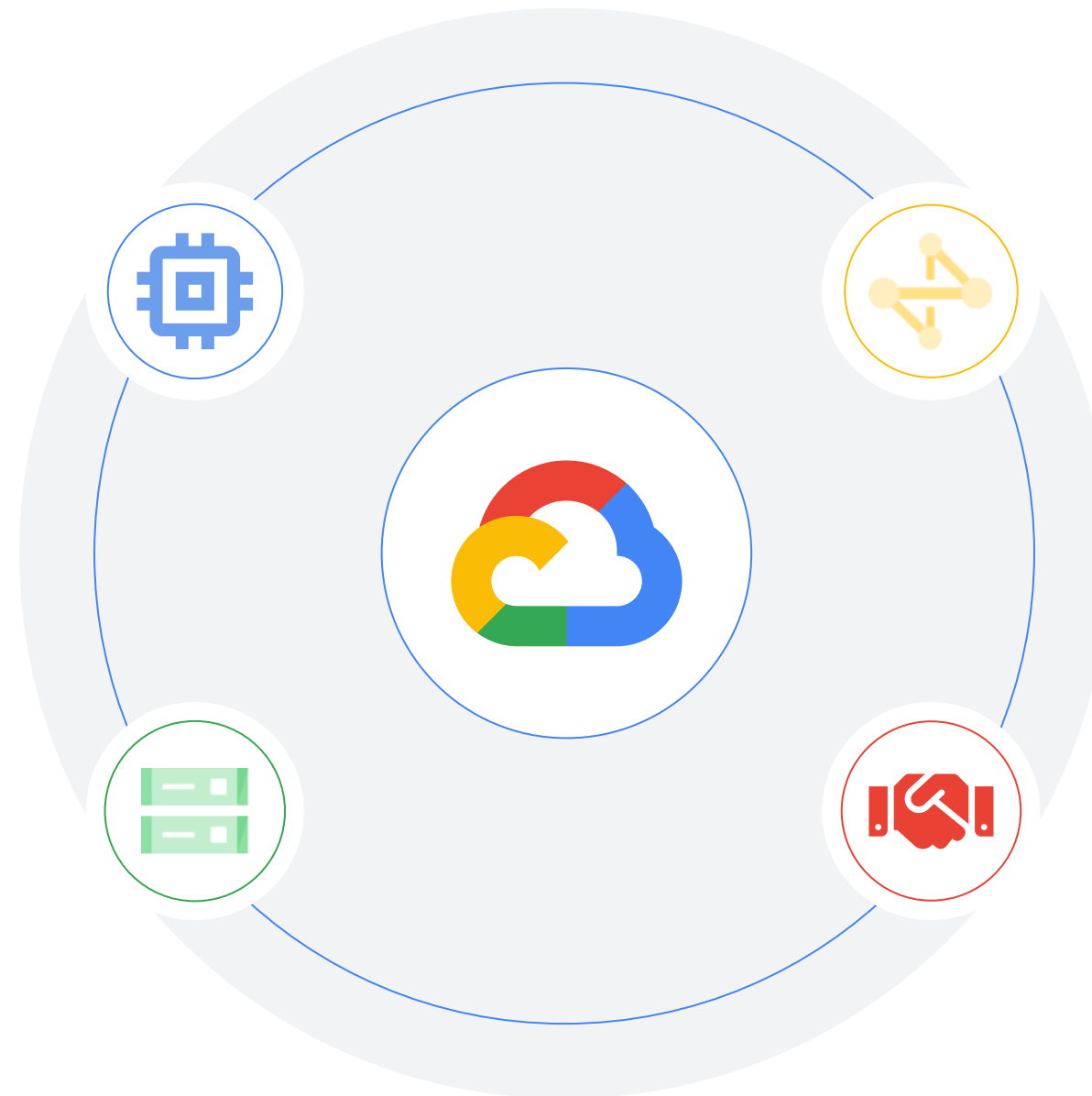
# High Performance Computing on Google Cloud

## Compute

Google Compute Engine's **VMs boot in seconds**, are built for consistently high performance, and have security built in.

## Storage

**Various storage service offerings** remove much of the burden of building and managing storage and infrastructure.



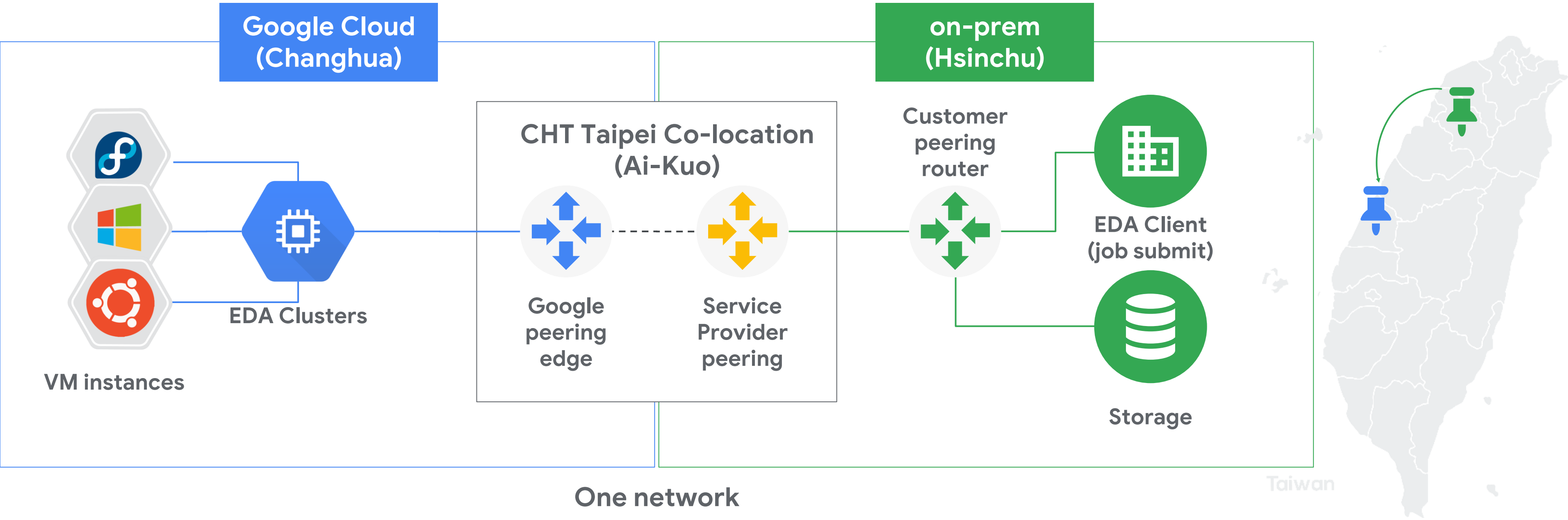
## Network

Google's high performance **private network** connects VMs with **high throughput, low latency** interconnects.

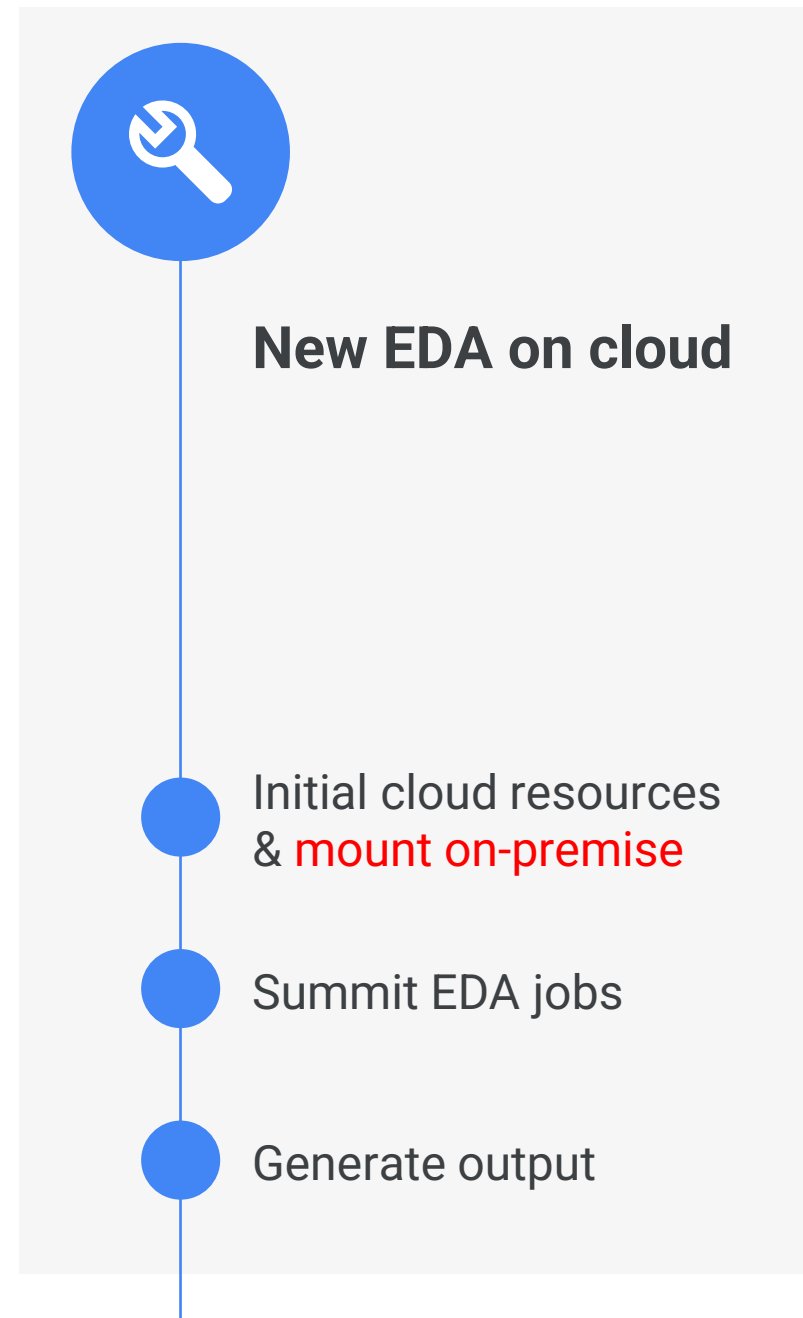
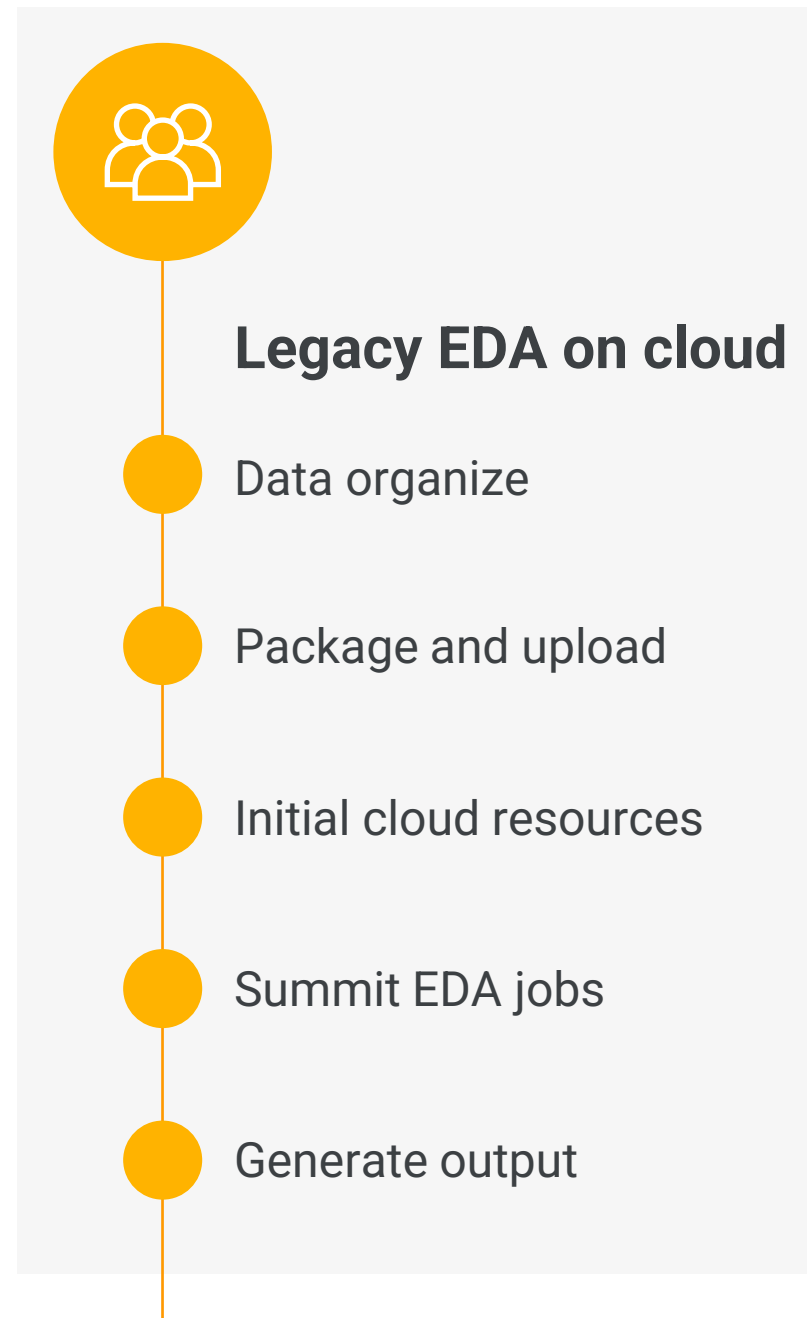
## HPC Software

Google Cloud offers **native HPC tooling**, and supports a broad portfolio of **HPC software** from our **HPC partners** and **open source projects**.

# Google Cloud Interconnect - As-IS (3ms latency)



# Democratize EDA by optimize EDA pipeline on Google cloud

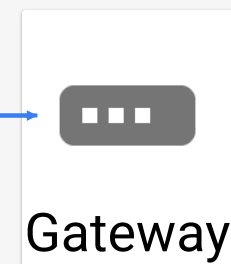
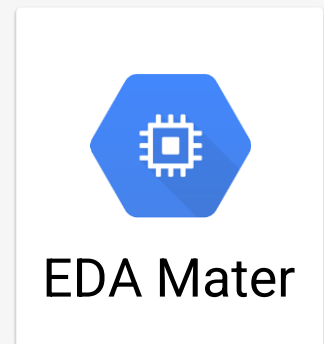
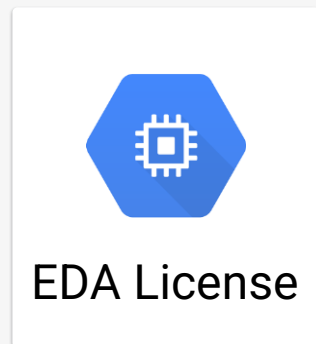


1. No pre-work required
2. 25% ~ 40% TCO off

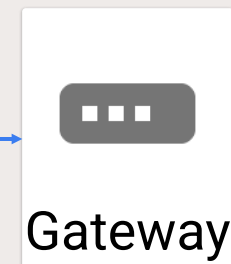


# EDA Hybrid Cloud Architecture

Google Cloud Platform (Changhua, Taiwan)

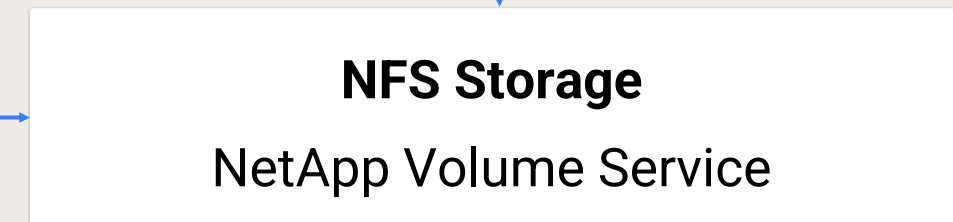
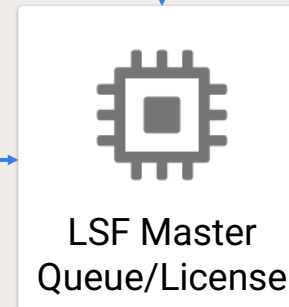
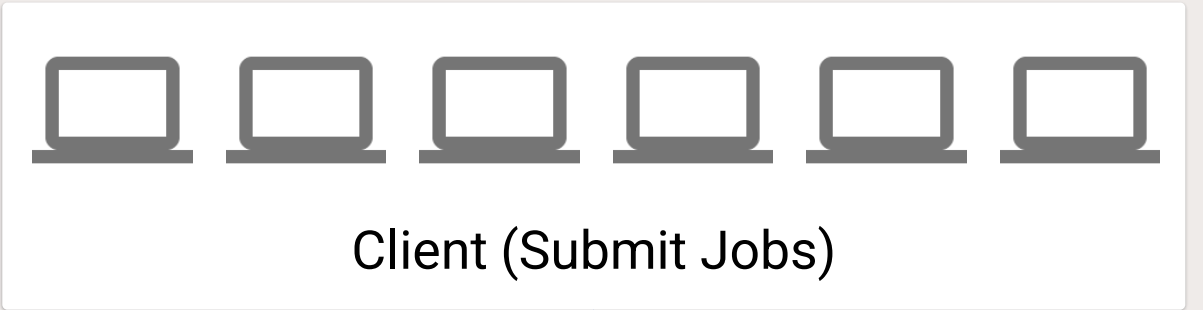


Interconnect



3ms network latency,  
with up to 100 Gbps  
throughput

On-Premises (Hsinchu)



# Summary

Let us summarize the 3 key steps for identifying and migrating workloads to Intel instance:



Technical Deep  
dive

Understand Intel instance's super power



Workload  
identification

Assess the work workloads and identify the ideal  
workloads for Intel instance



Proof of  
Concept

Let's start to use Intel instance on GCP to make  
your work better!



**Thank you**

Google Cloud