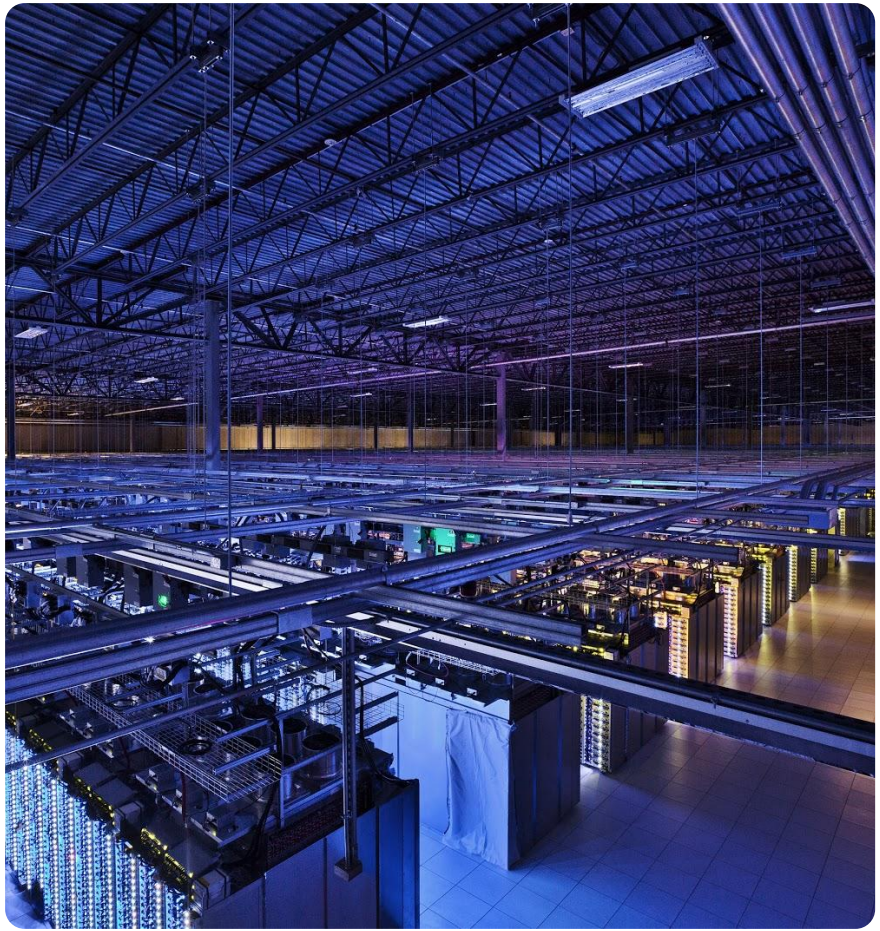


DAOS on Google Cloud

DAOS User Group
November 2021

Carlos Boneti
HPC Software Engineer, Google

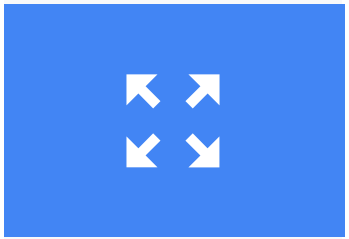
Johann Lombardi
Lead DAOS architect, Intel AXG

Agenda

1. Intro to Google Cloud — 5 minutes
   - Overview
   - Storage Types
   - Typical HPC Architecture
   - Typical Data Path (On-Premises ↔ Cloud)
2. DAOS on Google Cloud — 5 minutes
3. Performance numbers — 8 minutes
   - Overview
   - User Cards and User Paths
   - Product and Service Cards
   - Zones and Nesting
   - Title and Footer Bars

HPC on Google Cloud

Meeting HPC needs with fastest time to insight and simplicity



Fast and Affordable

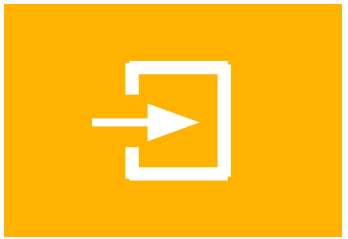
- Latest technologies
- MPI scalability
- Flexible VM shapes
- Spot VMs

Easy and Compatible

- Simple and Easy to Use
- Turnkey HPC environments
- Compatible with leading apps and schedulers

Partner Preferred

- Broad ISV application support
- Broad network of Cloud HPC solution providers



Best of Google

- Leading ML and Analytics
- IoT <-> R&D HPC integration
- Hybrid & multi-cloud via Anthos/K8s
- Best worldwide network

Google Cloud

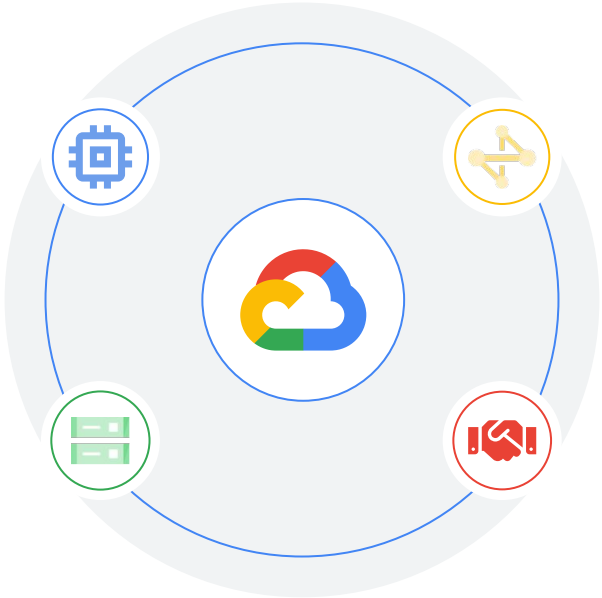
High Performance Computing on Google Cloud

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

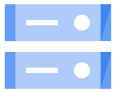
Google's high performance private network connects VMs with high throughput, low latency interconnects.

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

Which storage type?

In Memory	Relational		NoSQL		Analytical	Object	Block		File
									
Memorystore	Cloud SQL	Cloud Spanner	Firestore	Cloud Bigtable	BigQuery	Cloud Storage	Persistent Disk	Local SSD	Filestore & Partners
Managed Redis & Memcached	Managed MySQL and PostgreSQL, and SQL Server	Scalable relational database	Serverless, scalable, document store	Low-latency, scalable key-value and wide-column store	Enterprise DW	Unstructured data, objects or blobs	Flexible VM block storage	High performance NVMe-based block storage	HPC workloads requiring file and POSIX

Instance Max Read / Write 2.2GB/s

Instance Max Read 9.36GB/s Max W rt 4.68GB/s

Block storage performance

https://cloud.google.com/compute/docs/disks/performance

Perf per instance	Zonal standard PD	Regional standard PD	Zonal balanced PD	Regional balanced PD	Zonal SSD PD	Regional SSD PD	Zonal extreme PD	Local SSD (SCSI)	Local SSD (NVMe)
Maximum sustained IOPS									
Read IOPS	7,500*	7,500*	15,000–80,000*	15,000–60,000*	15,000–100,000*	15,000–60,000*	15,000–120,000*	900,000	2,400,000
Write IOPS	15,000*	15,000*	15,000–80,000*	15,000–30,000*	15,000–100,000*	15,000–30,000*	15,000–120,000*	800,000	1,200,000
Maximum sustained throughput (MB/s)									
Read throughput	240–1,200*	240–1,200*	240–1,200*	240–1,200*	240–1,200*	240–1,200*	240–2,200**	9,360	9,360
Write throughput	76–400**	38–200**	240–1,200*	120–600*	240–1,200*	120–600*	240–2,200**	4,680	4,680

* Persistent disk IOPS and throughput performance depends on disk size, instance vCPU count, and I/O block size, among other factors.

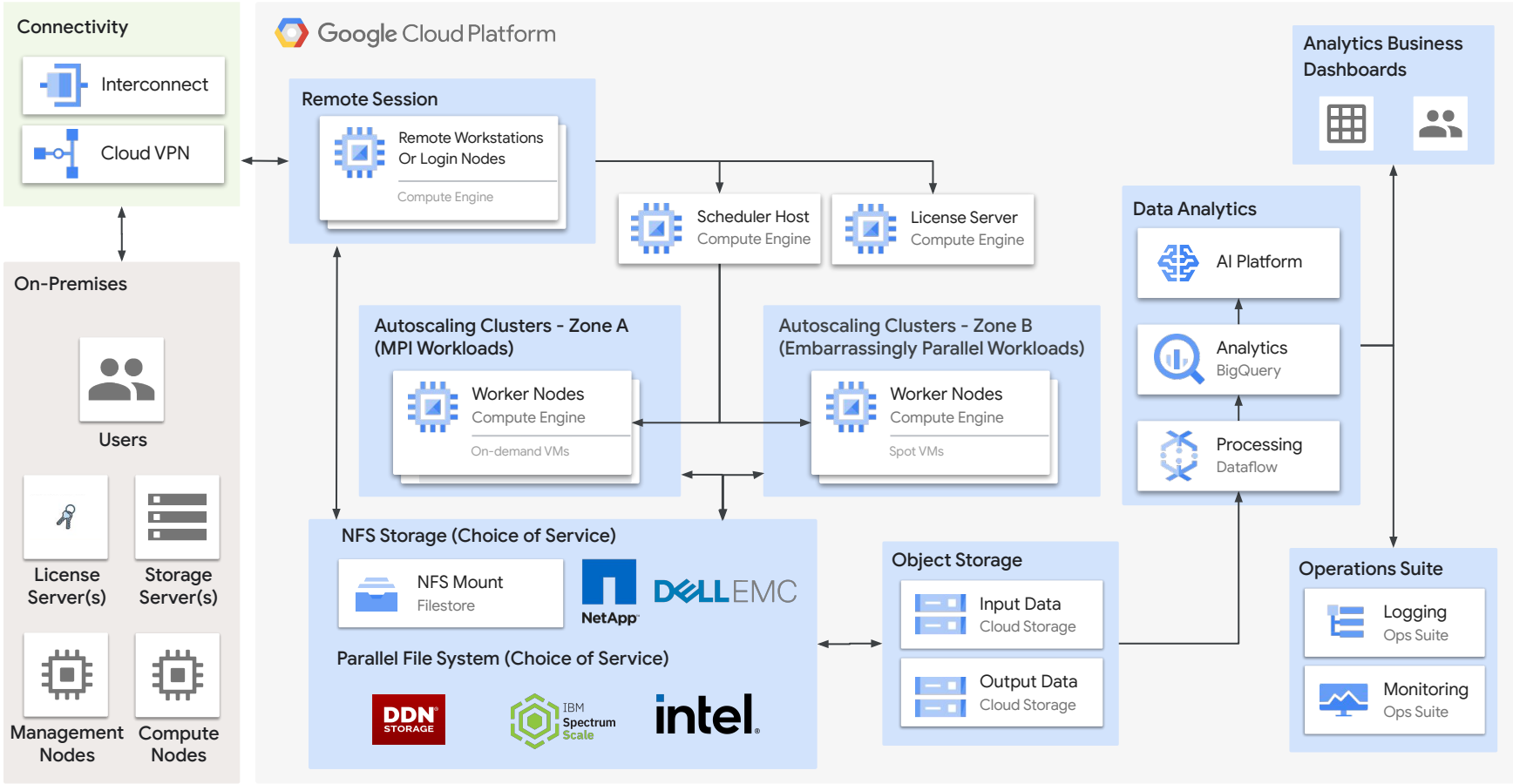
** Persistent disks can achieve greater throughput performance on instances with more vCPUs. Read Network egress caps on write throughput.

Network Bandwidth (N2 VMs)

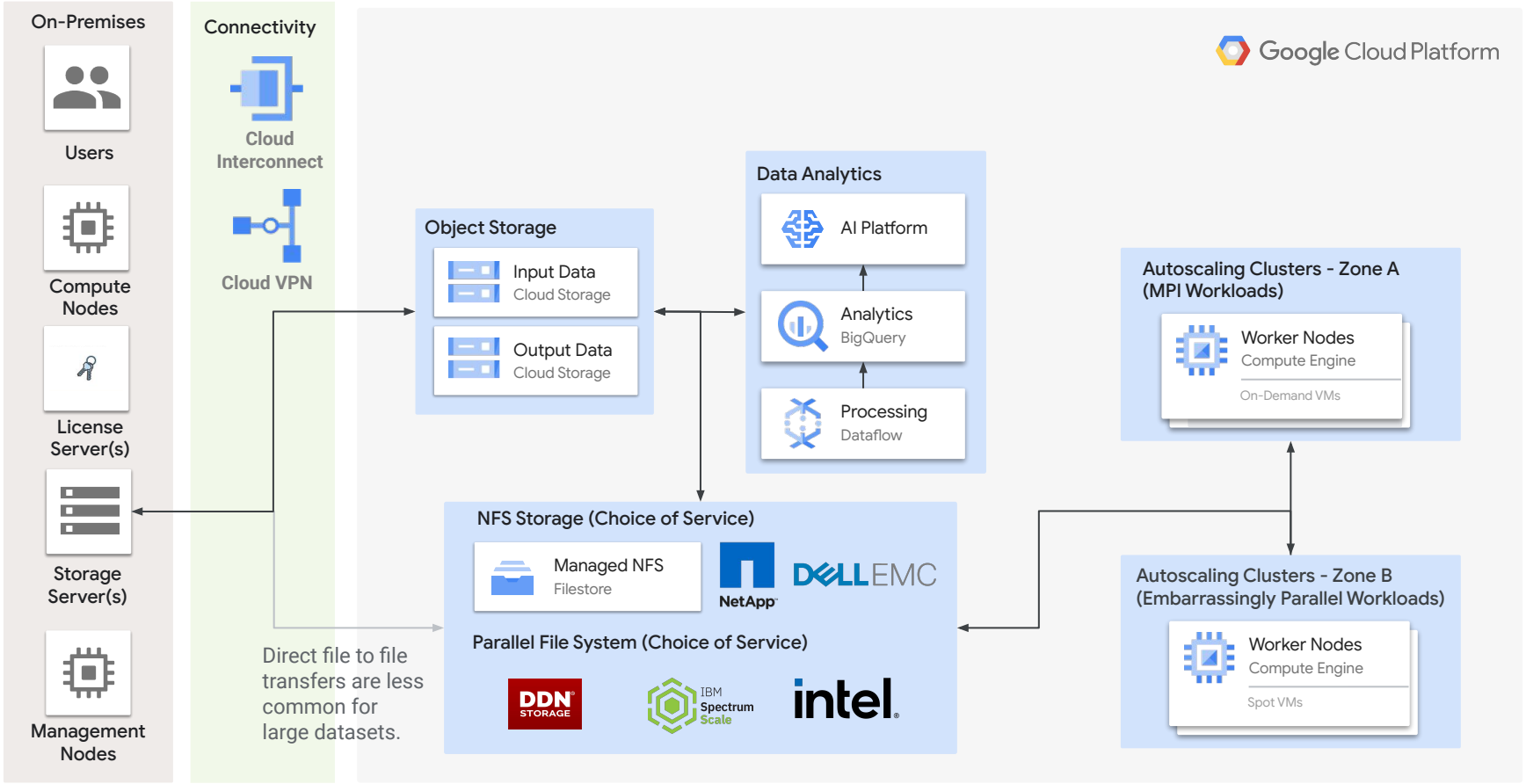
https://cloud.google.com/compute/docs/networking/configure-vm-with-high-bandwidth-configuration

vCPUs	Default	Tier 1
2 - 4	10 Gbps	Not applicable (N/A)
8	16 Gbps	N/A
16	32 Gbps	N/A
32	32 Gbps	50 Gbps
48	32 Gbps	50 Gbps
64	32 Gbps	75 Gbps
80 - 128	32 Gbps	100 Gbps

Open, Standards-Based Architecture for Cloud HPC



Typical data paths



DAOS on Google Cloud: Deployment

https://github.com/daos-stack/google-cloud-daos

First key use case: ephemeral storage (scratch)

- Two terraform modules
  - **daos_server**
  - **daos_client**

- Various examples
  - **simple_daos_server_example**: managed instance group running DAOS servers
  - **daos_client_mig**: managed instance group with client nodes
  - **full_cluster_setup**: both clients and server instances
  - **io500**: automated deployment for io500 runs

DAOS on Google Cloud: Configuration

https://github.com/daos-stack/google-cloud-daos

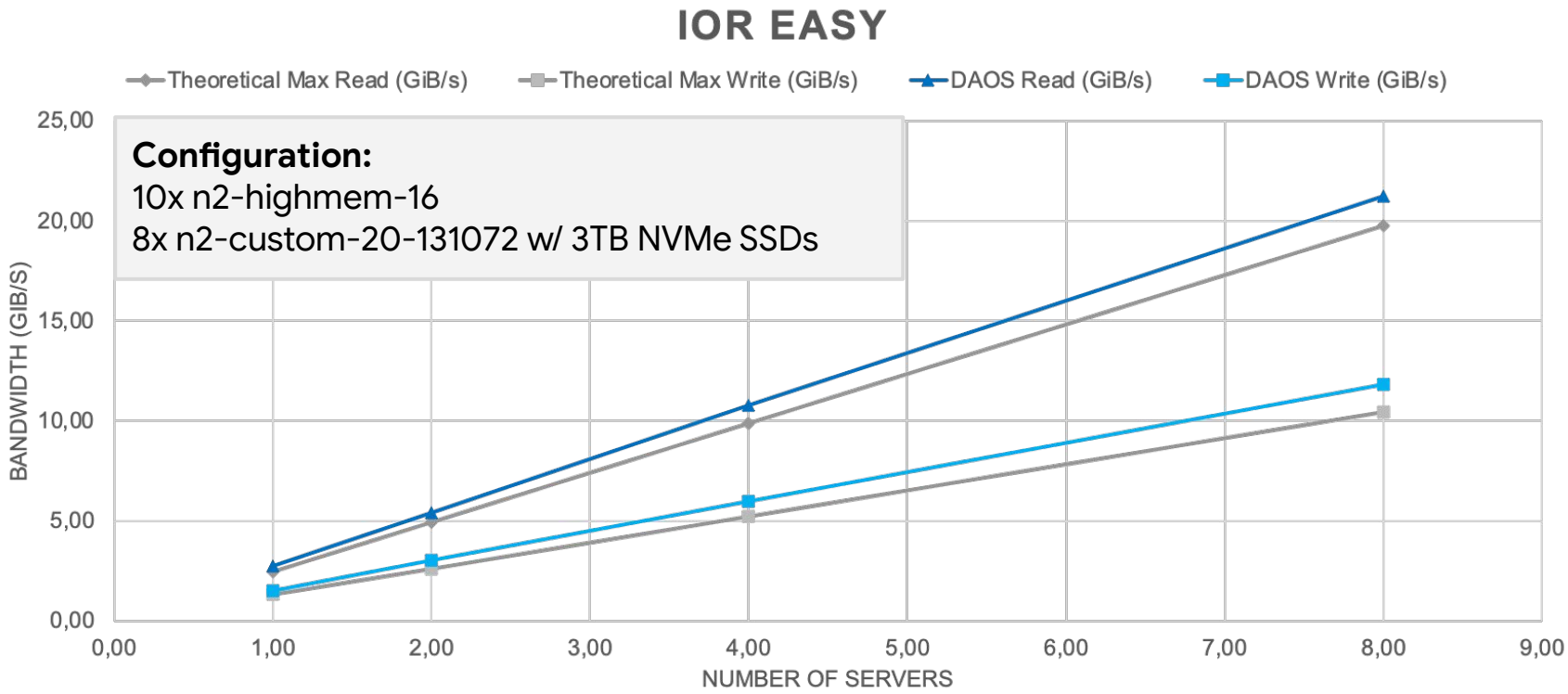
- Recommended configuration
  - 1x target per SSD partition
  - 2x vCPU per target (hyperthread)
  - Adjust DRAM capacity based on number of SSDs
  - n2-highmem-16 with 1.5TB per server for best performance
  - n2-highmem-32 with 6TB per server for best cost per GB
- Interfaces available out of the box
  - native DAOS API (libdaos.so)
  - DAOS File System library (libdfs.so)
  - Fuse daemon (dfuse)
  - Interception library (libioil.so)
  - MPI-IO DAOS driver (Intel MPI)
  - Python module (PyDAOS)
  - To do:
    - HDFS java bindings

DAOS on Google Cloud: Current Limitations

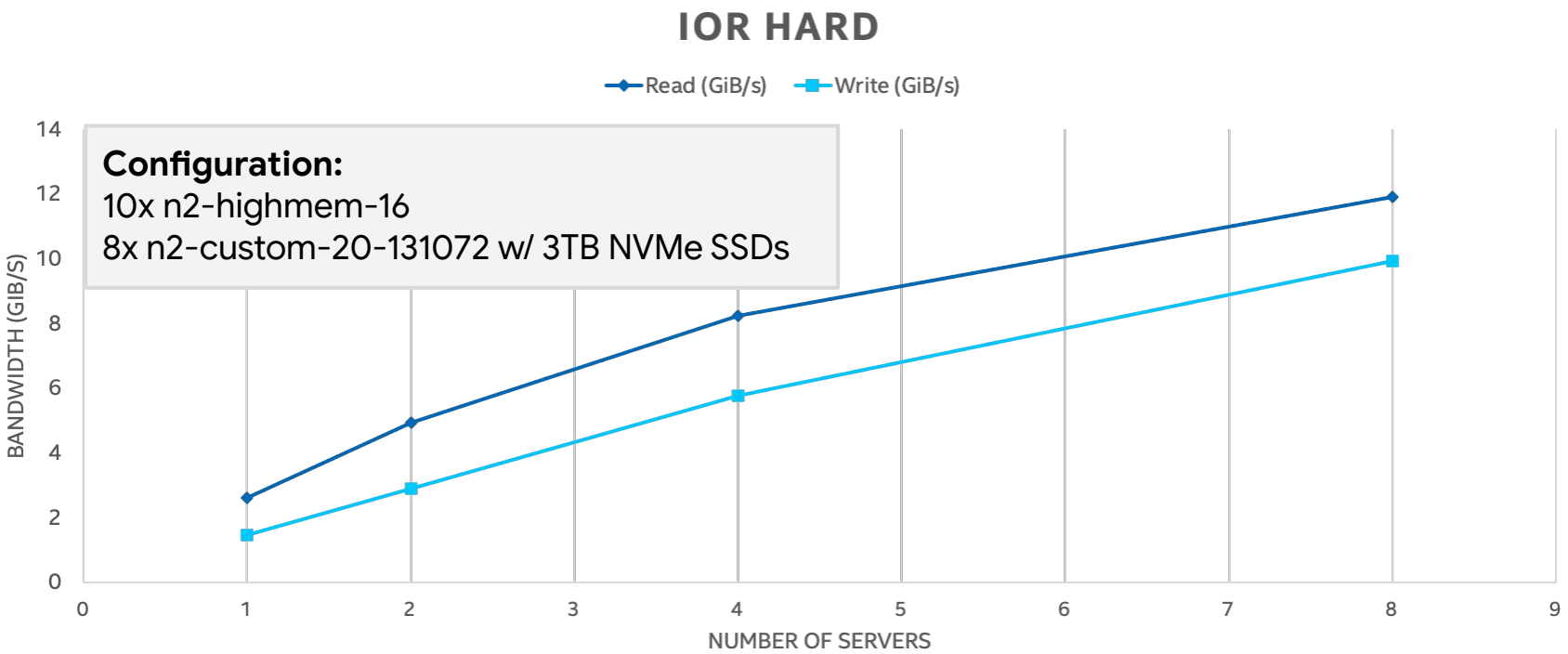
https://github.com/daos-stack/google-cloud-daos

- No persistency (no pmem)
  - RAM for metadata
  - Local SSD (NVMe) for data
  - Cannot survive a full reboot of all the instances at once
  - Cannot survive failures on more server instances than the configured redundancy. However, Google Cloud provides live migration for host maintenance and certain failures https://cloud.google.com/compute/docs/instances/live-migration
- Currently limited to 6TB per node

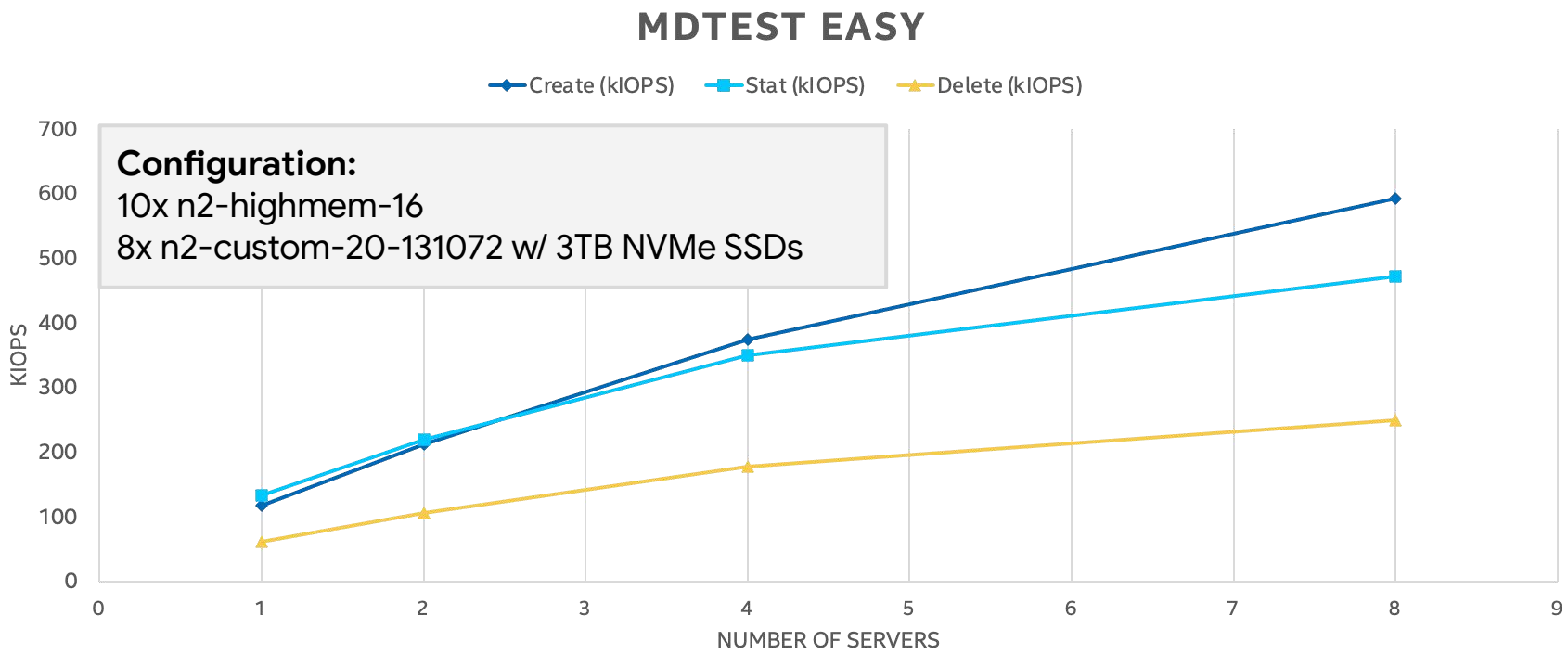
DAOS on Google Cloud: Performance



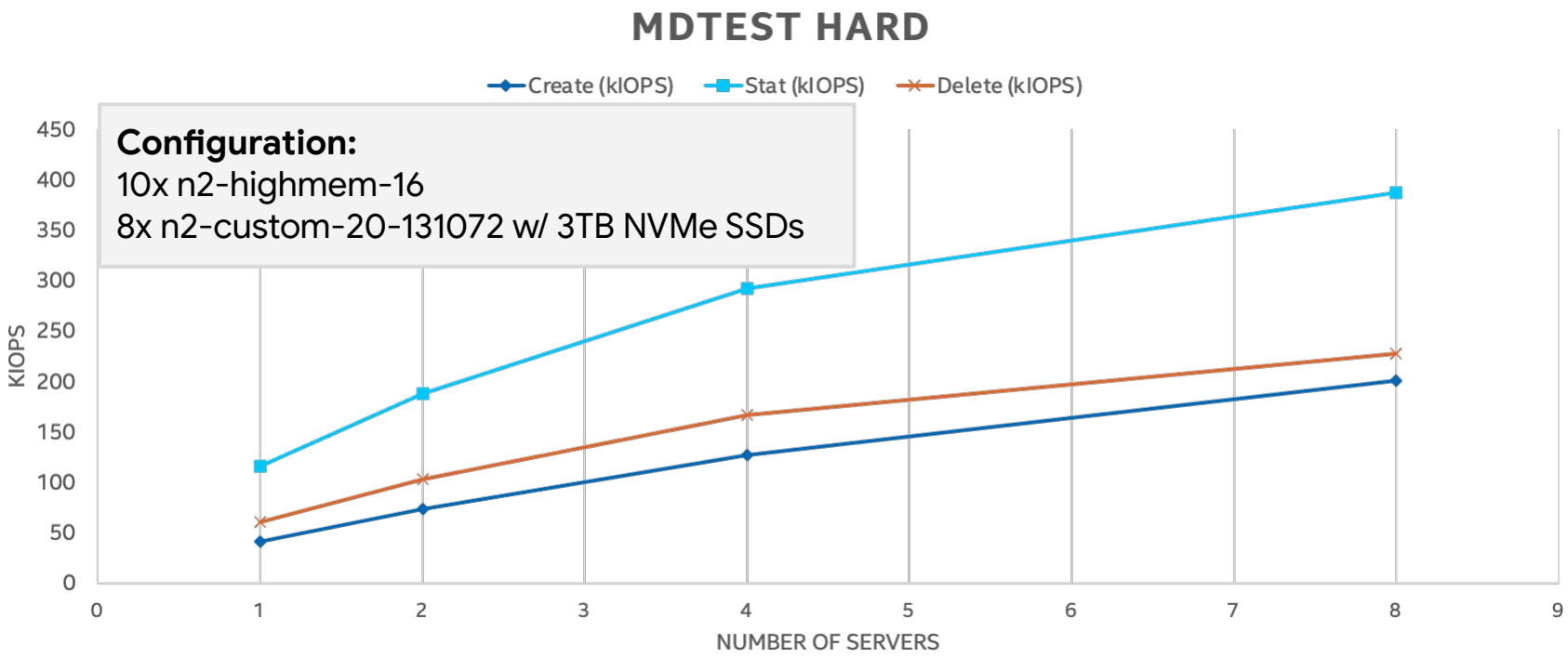
DAOS on Google Cloud: Performance



DAOS on Google Cloud: Performance



DAOS on Google Cloud: Performance



DAOS on Google Cloud: Recap & Future Plans

https://github.com/daos-stack/google-cloud-daos

DAOS on Google Cloud provides a simple and fast way to deploy DAOS with very good performance.

Future plans:

- Automate data movement from GCS
- Telemetry
- Further automation: creation of partitions, mounting in clients, etc.
- Increased storage to 9TB per server node
- Guided initial configurations
  - Best for IOPS
  - Best for bandwidth
  - Lowest cost per GB

Thank you.

https://cloud.google.com/hpc

Google Cloud