

# DAOS on Google Cloud

## DUG22 @ SC22

Dean Hildebrand, Technical Director, Office of the CTO

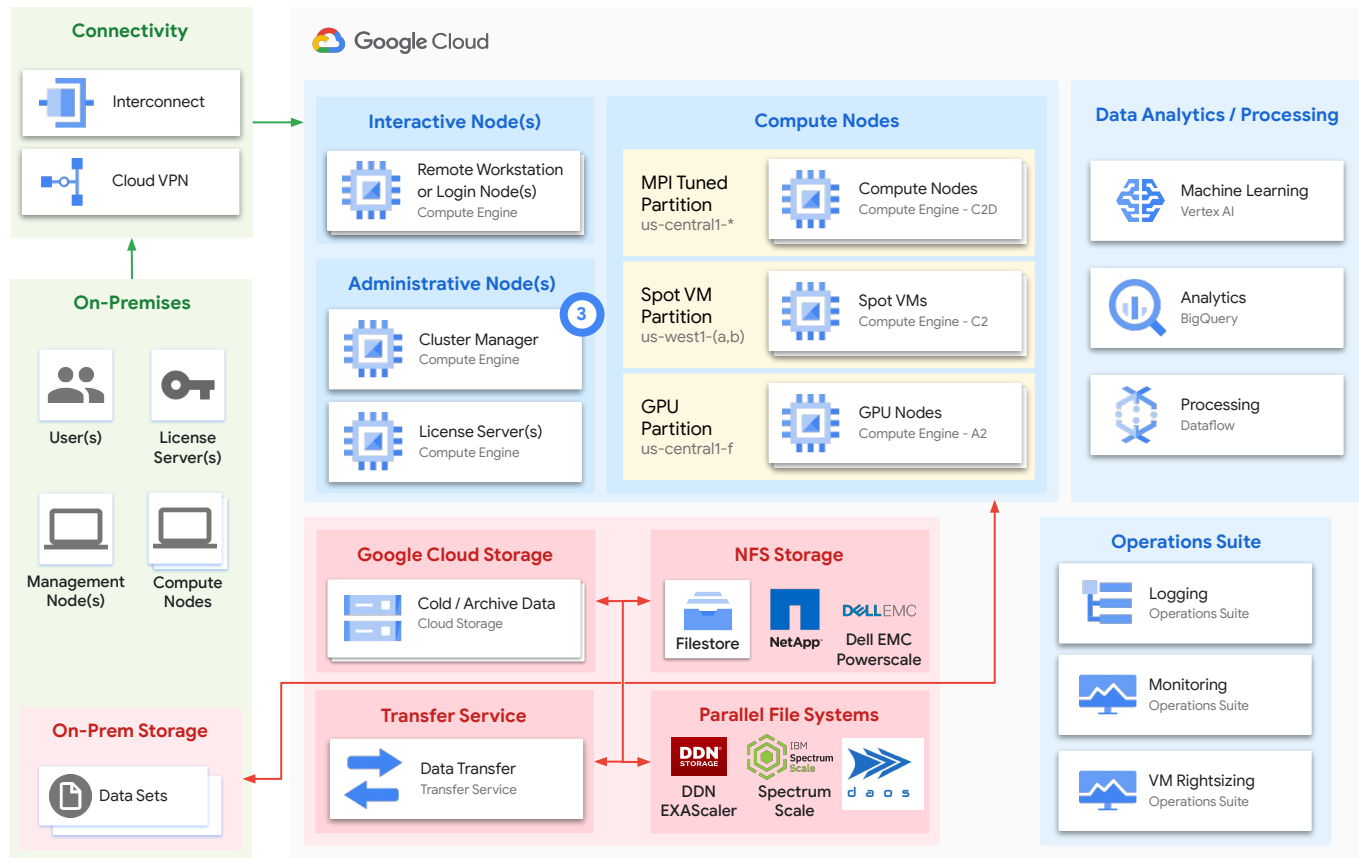
# Open, Standards-Based Architecture for Cloud HPC

On-Premises, Hybrid,  
Connectivity

High Performance Computing  
Architecture

Workload Manager

Data Storage



# Google Cloud Focus Areas for **DAOS**

01

# Deliver High Performance for Cloud and Hybrid Workflows

# DAOS on Google Cloud

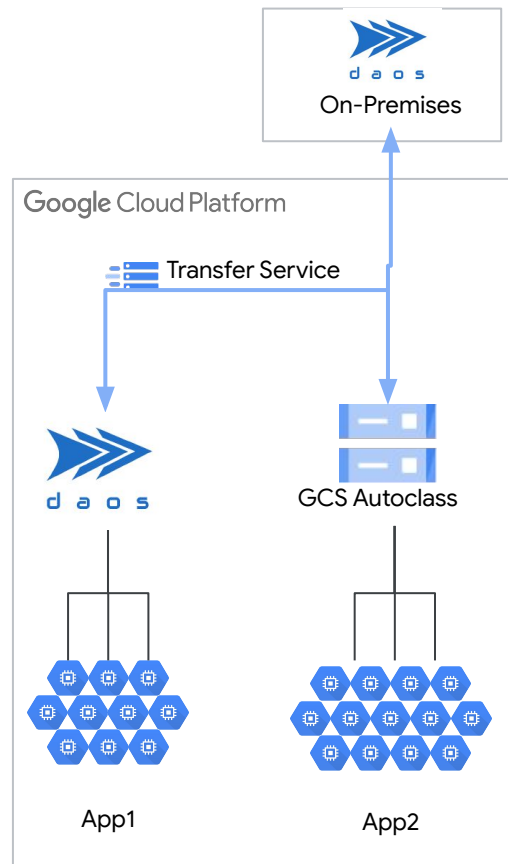
## Accelerate HPC and AI Applications

### Combine DAOS and Object Storage (GCS)

- Use GCS for latency tolerant apps and archive
- Use DAOS to provide
  - File API
  - 100x lower latency
  - Increased small file, small I/O and metadata IOPs
  - Increased single client performance

### Details

- Keep GPUs/TPUs fed with data
- User-level DAOS client compatible with Google Kubernetes Engine (GKE)
- Software-managed data protection across servers to improve availability
- Simplified deployment via Terraform and Google HPC Toolkit
- Currently support up to 6TB NVMe per storage server plus variable amounts of RAM
  - Cloud can keep rebuilding with additional servers until zone lacks capacity



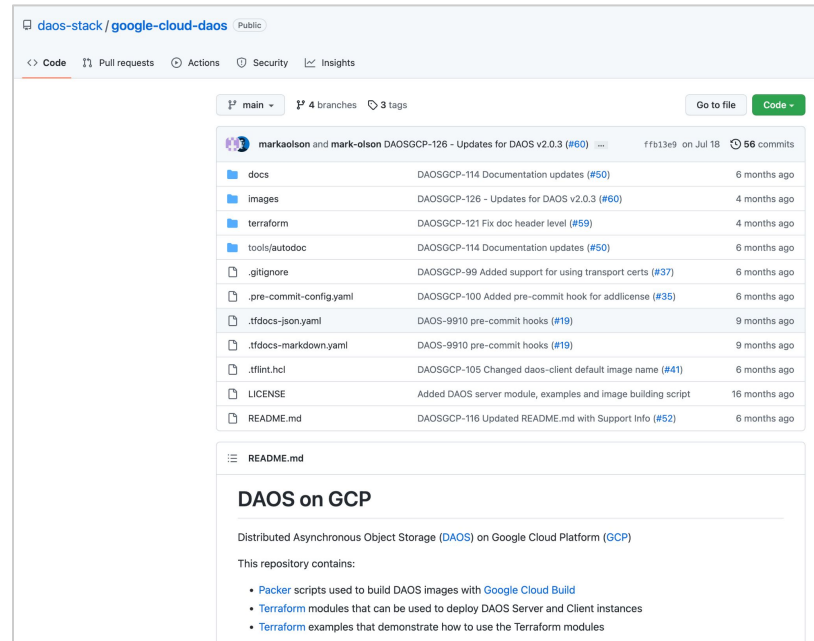
02

# Easy, Repeatable, and Integrated Deployment

# Deploying DAOS on GCP

## Terraform

Standalone DAOS deployment that can be integrated with your current workflow



### Installation and Setup

RHEL and clones

SUSE

DAOS in Docker

[DAOS in the Cloud](#)

[Deploying DAOS in Google Cloud](#)

Build from Scratch

Admin/Client Tools

## Deploying DAOS in the Cloud

### Deploying DAOS in Google Cloud

DAOS can be installed in GCP. Please refer to the [How To Deploy DAOS in Google Cloud](#) section of the Google Cloud HPC Toolkit documentation for details.

# Deploying DAOS on GCP

## Cloud HPC Toolkit

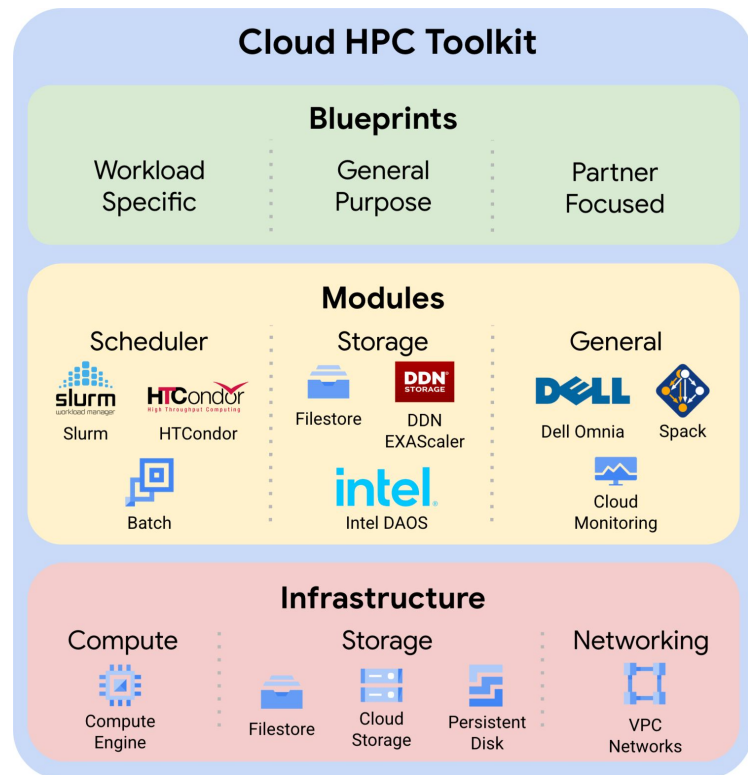
Modular, composable, terraform-based toolkit designed to make it easy to deploy repeatable, turnkey HPC environments that follow Google Cloud's HPC best practices.

Fully open-source

Predefined configs to ensure best Price/Perf for key workloads

### Key components:

- **Blueprints** defines an HPC environment
- **Modules** are code to deploy specific components such as a scheduler, a storage system, and network
- **Infrastructure** used by the deployed HPC system






# Simple DAOS Blueprint Example

```
blueprint_name: daos-cluster
vars:
  project_id: ## Set GCP Project ID Here ##
  deployment_name: daos-cluster
  region: us-central1
  zone: us-central1-c
deployment_groups:
- group: primary
  modules:
  - id: network1
    source: modules/network/pre-existing-vpc

# This module creates a DAOS server. Server images MUST be created first
- id: daos-server
  source:
  github.com/daos-stack/google-cloud-daos.git//terraform/modules/daos_server?ref=v0.2.1
  use: [network1]
  settings:
    number_of_instances: 2
    labels: {ghpc_role: file-system}

# This module creates a MIG with DAOS clients. Client images MUST be created first
- id: daos-client
  source:
  github.com/daos-stack/google-cloud-daos.git//terraform/modules/daos_client?ref=v0.2.1
  use: [network1, daos-server]
  settings:
    number_of_instances: 2
    labels: {ghpc_role: compute}
```



```
ghpc create community/examples/intel/daos-cluster.yaml
terraform -chdir=daos-cluster/primary init
terraform -chdir=daos-cluster/primary validate
terraform -chdir=daos-cluster/primary apply
```



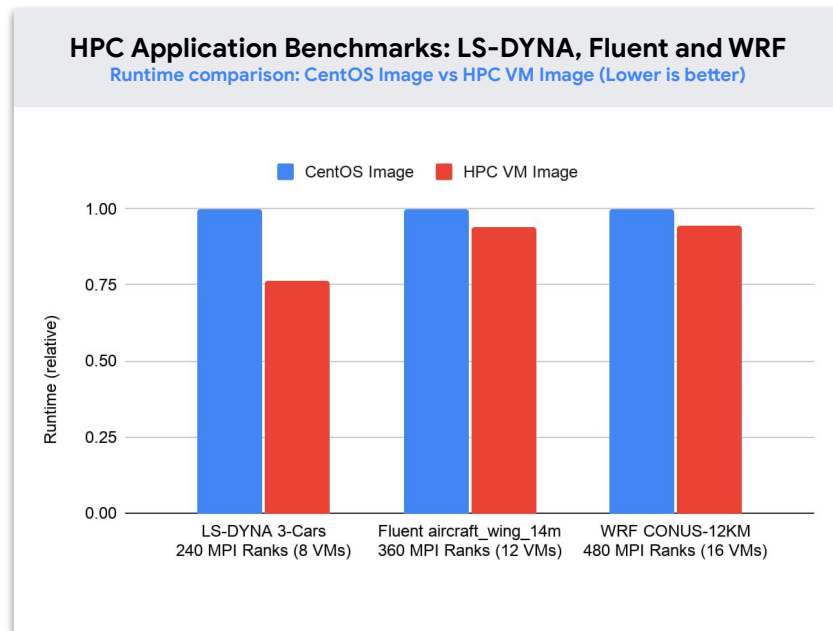
Can be your favourite scheduler  
e.g., Slurm, Cloud Batch

02

# Fast out of the Box

# Pre-Tuned HPC VM Images

- Bundle DAOS client/servers into optimized images
  - **Focus on TF, POSIX**
- Pre-set tunings for max performance
  - **Focus on TCP in cloud (with and w/o RDMA)**
- Additional Tunings and Optimizations Included
  - Adjust user limits on system resources
  - Increase tcp \*mem settings
  - Use the network-latency profile
  - Disable Linux firewalls
  - Disable SELinux
  - Intel MPI collective tunings
  - (Optional) Disable Spectre/Meltdown patches
- Available as a stand-alone image, in the Marketplace, or as individual tunings you can apply to your own images



<https://cloud.google.com/blog/topics/hpc/introducing-hpc-vm-images>

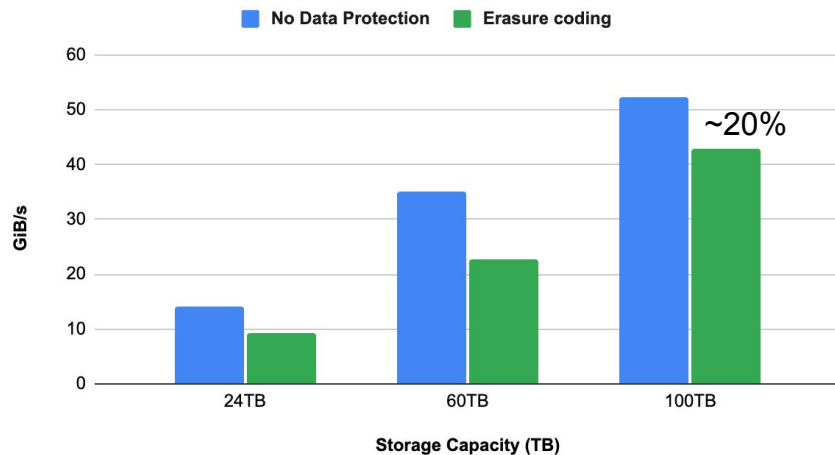
03

**Maximize Perf/\$**

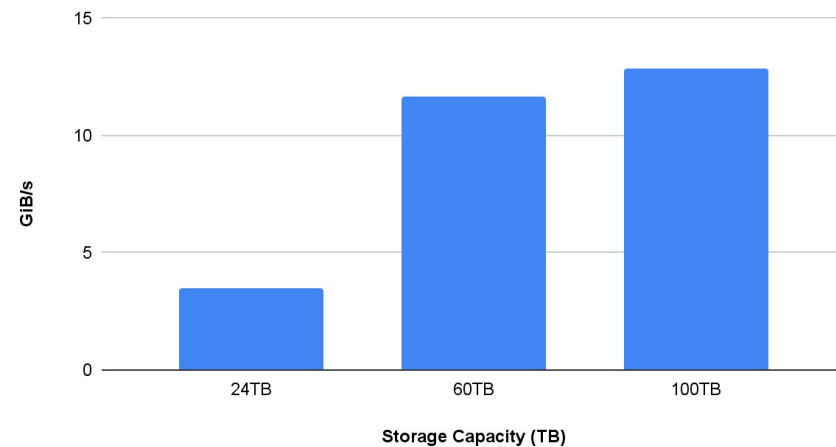
# IO500 - IOR Write

Servers: n2-custom-36-262GB  
Clients: c2-standard-16-64GB  
(2x clients than servers)  
Protection - 8+2EC

## IOR Easy Write



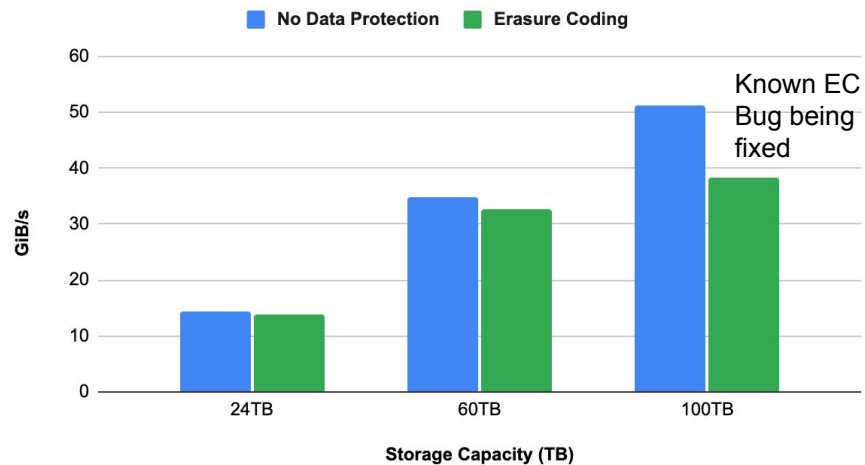
## IOR Hard Write



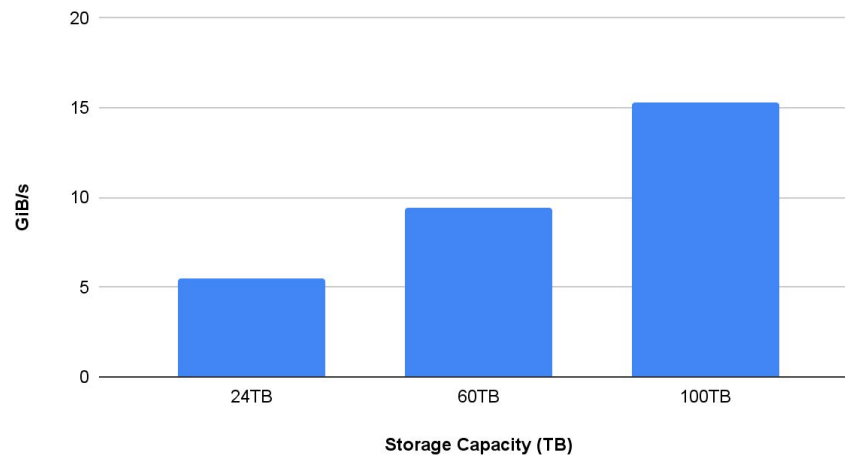
# IO500 - IOR Read

Servers: n2-custom-36-262GB  
Clients: c2-standard-16-64GB  
(2x clients than servers)  
Protection - 8+2EC

## IOR Easy Read



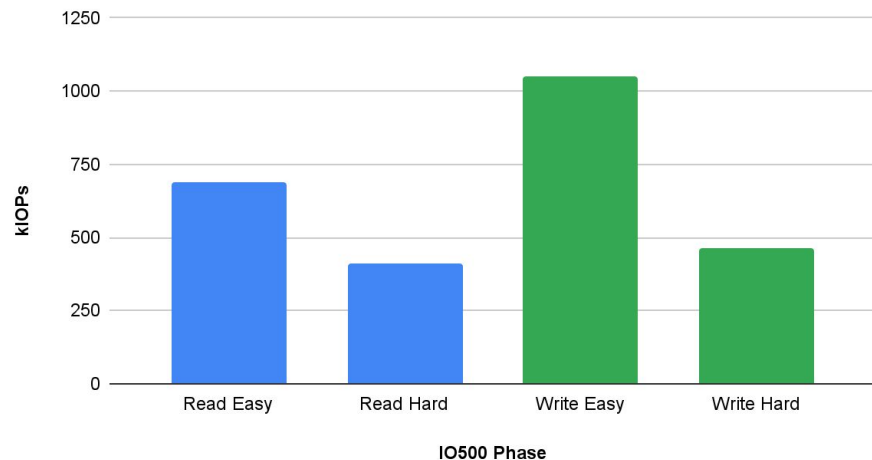
## IOR Hard Read



# IO500 - Metadata Performance

Servers: n2-custom-36-262GB  
Clients: c2-standard-16-64GB  
(2x clients than servers for 100TB)  
Protection - 8+2EC

### MetaData 100TB



**~50% drop from the easiest to hardest workload**

### 24TB MD Test Client Tests



**Performance scaling with clients**

# DAOS Resiliency Features for Cloud

## Happy to Collaborate with Community

- Non-disruptive upgrades to deliver cloud expectations of high availability
- Client/Server compatibility through several generations to support long running jobs while upgrading servers
- Data rebuild policies
  - Server failures may recover quickly without data loss and don't need rebuild
  - NVMe Reservations for fast and evenly distributed failover of networked NVMe devices
- Support for all-NVMe (no pmem)
  - Niche market for pmem meant it never was deployed widely in cloud
  - Memory is expensive, so all NVMe with \*some\* mem caching delivers better perf/\$
- Object store integration for hybrid, flexibility, and cost savings



# DAOS in GCP

## Two Primary Models

### 1. DAOS is a cache, GCS as source of truth

- Hybrid workflows using GCS as a secure transient storage
- Enables data access from all cloud services using a variety of semantics
- Avoids lockin to a specific storage service (as data can be easily copied elsewhere)
- Can be ephemeral (cheaper) or persistent with all-NVMe solution (costlier)
- Support existing buckets without modification
- Pay for hot data stored in both DAOS and GCS

**Greater flexibility  
and cloud/hybrid  
integration**

### 2. DAOS as source of truth, optional tiering to GCS for cost savings

- Optimized tiering performance (e.g., small object gathering)
- Optimized for DAOS client and semantics (potentially limit access to cloud services)
- Persistent only (costly)
- Avoid paying for hot data in both DAOS and GCS
- Cannot support existing buckets without additional data copies

**Traditional  
familiarity**

# Try DAOS on GCP Today!

<https://docs.daos.io/v2.0/cloud/>



**Thank you.**

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

Google Cloud