

Flash Memory Summit

SNIA STORAGE
SECURITY SUMMIT
Wednesday, May 11, 2022 • Virtual

Balancing Performance and Spend in the Public Cloud

Kyle Grossmiller
Senior Solutions Architect
Pure Storage

Data Interaction is Different in the Cloud

On-Prem & Hosted



Resiliency

Highly Reliable Arrays, Built-in Snaps & DR

Efficiency

Thin Provisioning, Deduplication, Compression

Cost Considerations

Capacity Planning, resources once purchased are fixed

Cloud



Resiliency

High Availability for \$, Higher Durability, Globally Replicated

Efficiency

Thick Provisioned Capacity & Performance

Cost Considerations

Ingress/egress charges, Cold vs Warm, performance (IOPS) can be wasted

Different Location, Similar Problems

Choices, trade-offs, and sprawl

The public cloud has an ever-increasing number of storage options, snapshot limitations, and configuration parameters

Makes it difficult to understand, track, and implement the right option at the right time.

Storage option sprawl is a significant and potentially costly problem—especially at scale



Flash Memory Summit

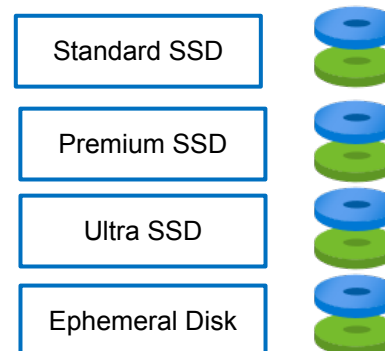


Azure Managed Disk

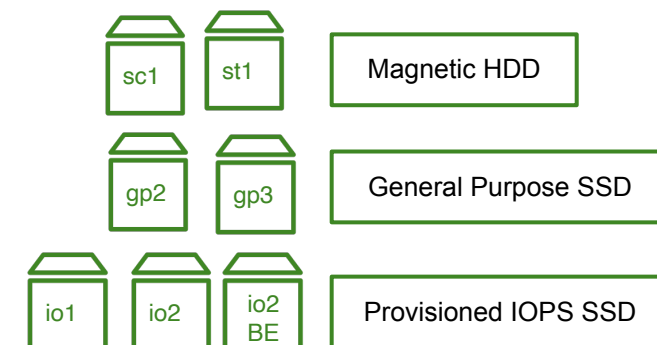


Amazon Elastic Block Store

Options



Tradeoffs



- Capacity vs. Throughput
- Capacity vs. IOPS
- Performance vs Latency
- Multi-Attach: Shared Volumes
- Snapshot capabilities
- Durability

Performance Provisioning in AWS?

AWS Native Storage Solutions

Different block storage options have different advantages and disadvantages within AWS.

Here are a few examples:

- **GP2:** Performance comes with capacity. For example: 100GB gets 100 IOPS and a 5TB disk gets 15000 IOPS. These two values being relational can become cost-ineffective since you might have to overprovision capacity in order to get required performance.
- **GP3:** IOPS can scale independently of capacity, but it can quickly become a costly option. Relative low durability (2 to 3 9's) means multiple copies are likely required.
- **IO1, IO2, IO2BE:** IOPS can scale independently of capacity, but like GP3 it also can be costly. IO2BE requires r5.xlarge VMs and above to use.

What about Performance Provisioning in Azure?

Azure Native Storage Solutions

In some scenarios over provisioning for performance (e.g. you have a 1TB volume but require 20,000 IOPs) is the only option if certain features are also required like snapshots.

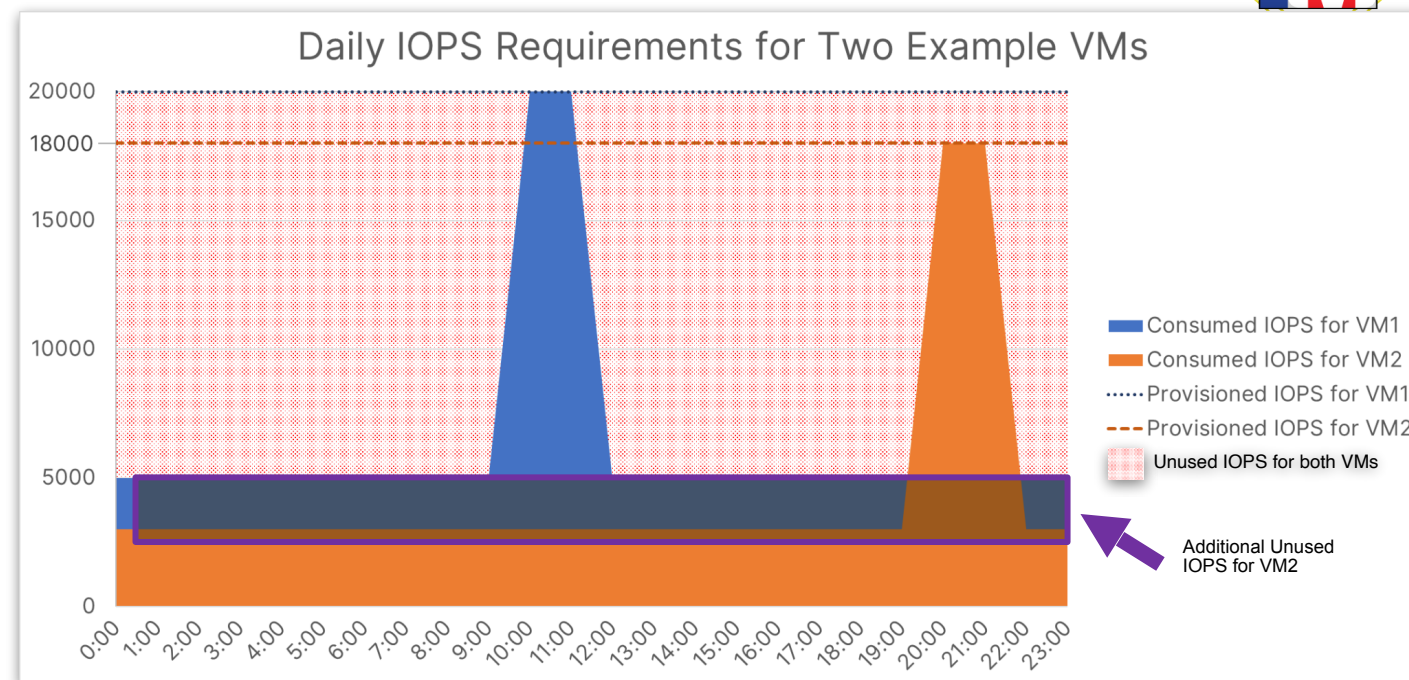
- **Premium SSD:** 1TB is the minimum needed for maximum IOPS performance. Premium SSD capacity points are binary values so if, for example, you needed 4.1TB of space - an 8TB Premium SSD is required. That leads to 3.9 TB of unused capacity you still must pay for.
- **Ultra SSD:** Performs well but you provision and pay for capacity, peak IOPs, peak bandwidth and VM vCPU reservation. Does not have the full set of enterprise features.

Underuse, Overspend Example

Overprovisioning for peak use on Azure with Ultra SSD tier

Monthly Cost	
Capacity cost (GiB)	\$ 0.11972
IOPS cost	\$ 0.04964
Bandwidth Cost (MB/s)	\$ 0.34967

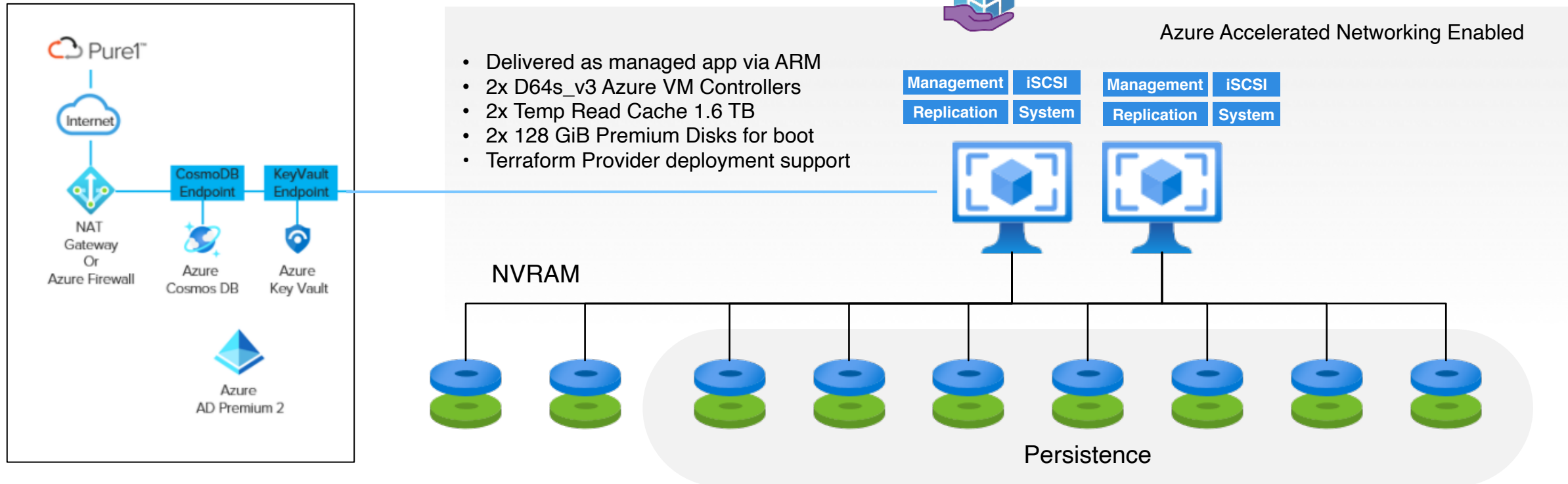
Monthly Cost							
	Provisioned Capacity (GiB)	Capacity Cost	Provisioned IOPS	IOPS Cost	Provisioned BW (MB/s)	BW Cost	Total Cost
VM1 12 TB Ultra SSD Tier	11175.84	\$ 1,337.97	20,000	\$ 992.80	1,000	\$ 349.67	\$ 2,680.44
VM2 8 TB Ultra SSD Tier	7450.58	\$ 891.98	18,000	\$ 893.52	1,000	\$ 349.67	\$ 2,135.17



Monthly Surplus		
Unused IOPS	Unused Bandwidth	Wasted Spend
67%	67%	\$ 899.45
74%	74%	\$ 919.96
Total Wasted Spend		\$ 1,819.41

Pure Cloud Block Store Azure Architecture

Highly-Available, Deeply-Integrated Design



5,714+ Provisioned IOPs & 86 MB/s per Ultra SSD for Persistence (14x)
 15,000 Provisioned IOPs & 450 MB/s per Ultra SSD for NVRAM (2x)
 >6TB = 1 IOP/GB

Pure Data Services Make Your Public Cloud Better

Purity^{FA}



Data
Reduction



Thin
Provisioning



Ransomware
Remediation



ActiveCluster
Replication



Async
Replication



ActiveDR



High
Availability



Always-on
Encryption



Snapshots



Azure

Azure Managed Disk

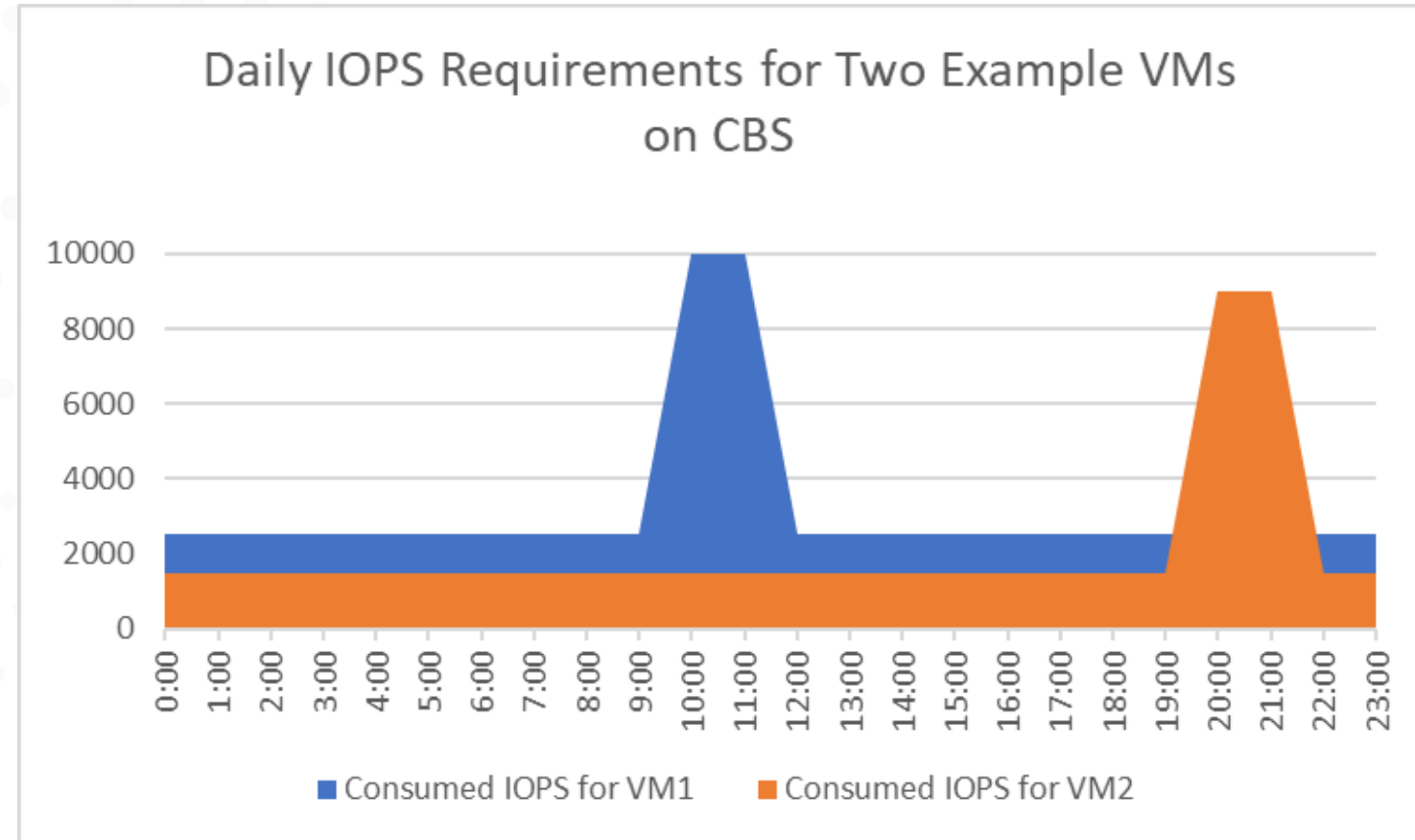
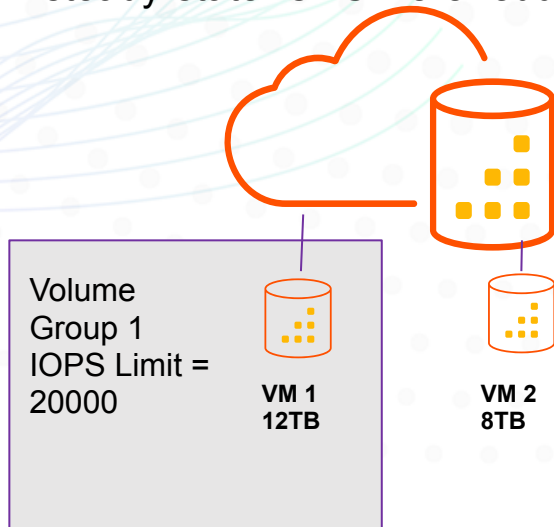


AWS Elastic Block Store

Provisioned IOPS with Pure Cloud Block Store



- Cloud Block Store will coalesce write IOPS from different workloads, meaning that you gain performance efficiencies and can fully utilize your provisioned resources.
- In our two VM example, both peak and steady-state IOPS were reduced by half.



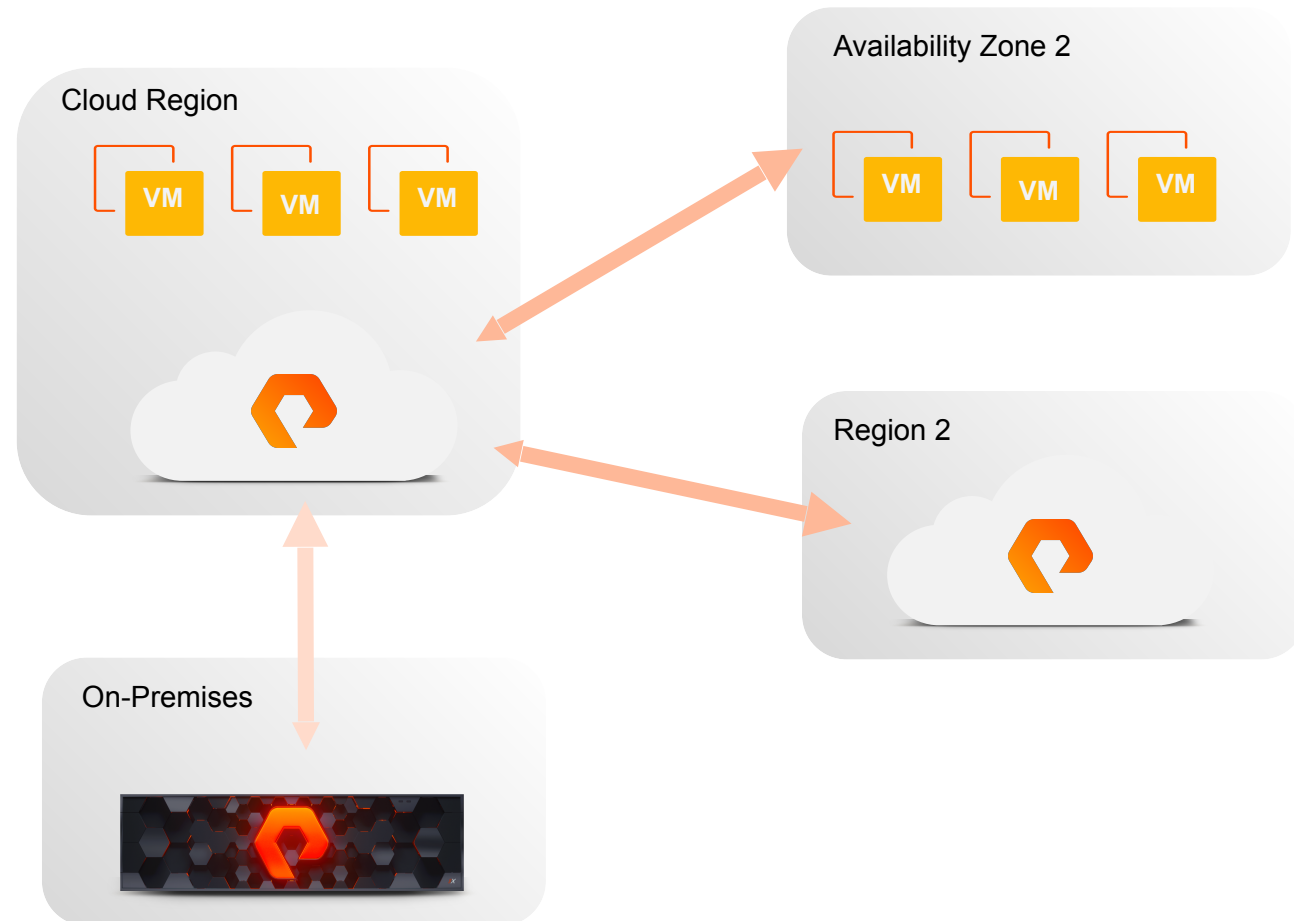
- Volumes can also be added to Volume Groups which enable QoS to be set for IOPS and/or Bandwidth to ensure workload fairness.



Ingress / Egress Charges from Data Mobility

Moving data in the cloud isn't always free

- Both AWS and Azure can charge on both sides of a data migration depending on situation
- Managing cost requires knowing the nuances of each scenario
- Minimize the impact with data reduction



*In AWS, there is no transit cost as long as VPCs are in the same availability zone



Optimizing Replication Reduces Bandwidth

Preserve data compression and deduplication when transferring between Pure products

- The storage footprint is reduced, lowering Cloud costs
- Data transfer costs and network utilization are minimized
- Replication times are shorter

