Terraform - Cloud Q on AWS

Qumulo Sizing & Performance on AWS

Dack Busch/Gokul Kuppuraj, July 15th, 2022

Overview

The information presented here is provided in the context of AWS infrastructure and the associated Terraform for deployment. To contain permutations cluster sizes from 4 to 10 instances (nodes) are graphically represented. However, the Terraform provided supports up to 20 instances. The flow of this document starts with deriving usable capacity, then throughput, followed by IOPS.

Note: most of the graphs that follow leverage a vertical log axis for better readability. Scale in perf and capacity is linear but may appear otherwise due to the log scale utilized.

Sizing

Erasure Coding

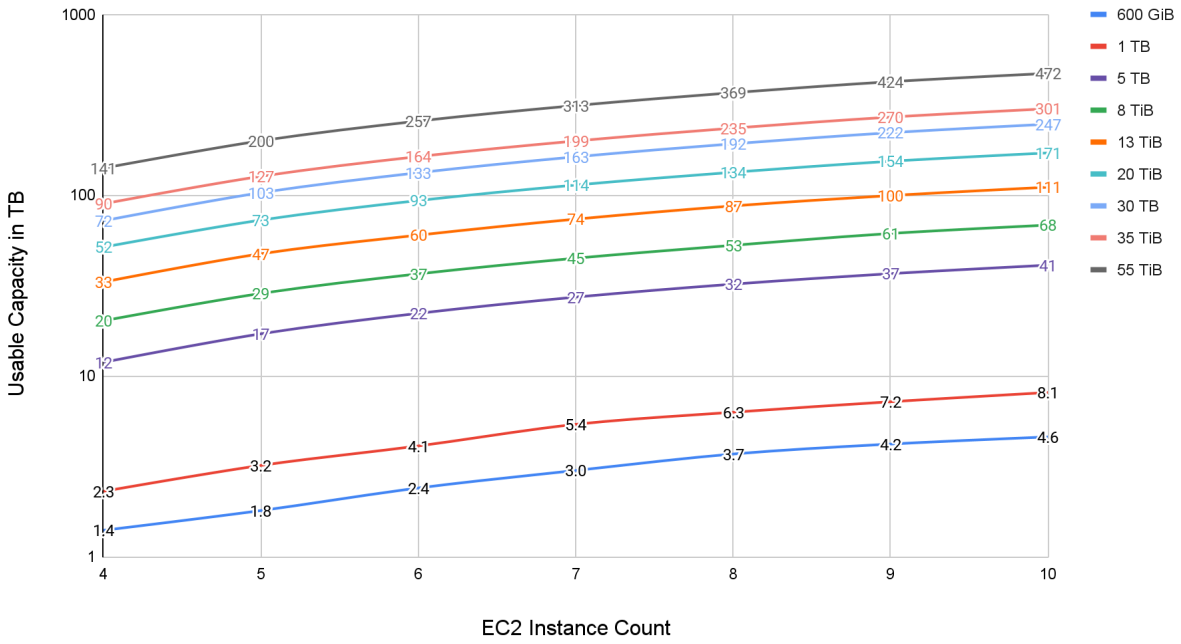
Qumulo Core erasure codes all data placed on EBS volumes. Erasure codes grow in 'width' as the number of instances in the cluster grows. The erasure code is established when the instances are created and the first quorum is formed. Since the Terraform provided supports 4 to 20 nodes, with 2 EBS volume failures or one EC2 instance failure, the erasure codes leveraged are:

# EC2 Instances	EC Code	Data Blocks	Parity Blocks	Efficiency
4	6,4	4	2	66.6%
5	8,6	6	2	75.0%
6	10,8	8	2	80.0%
7	12,10	10	2	83.3%
8	14,12	12	2	85.7%
9-20	16,14	14	2	87.5%

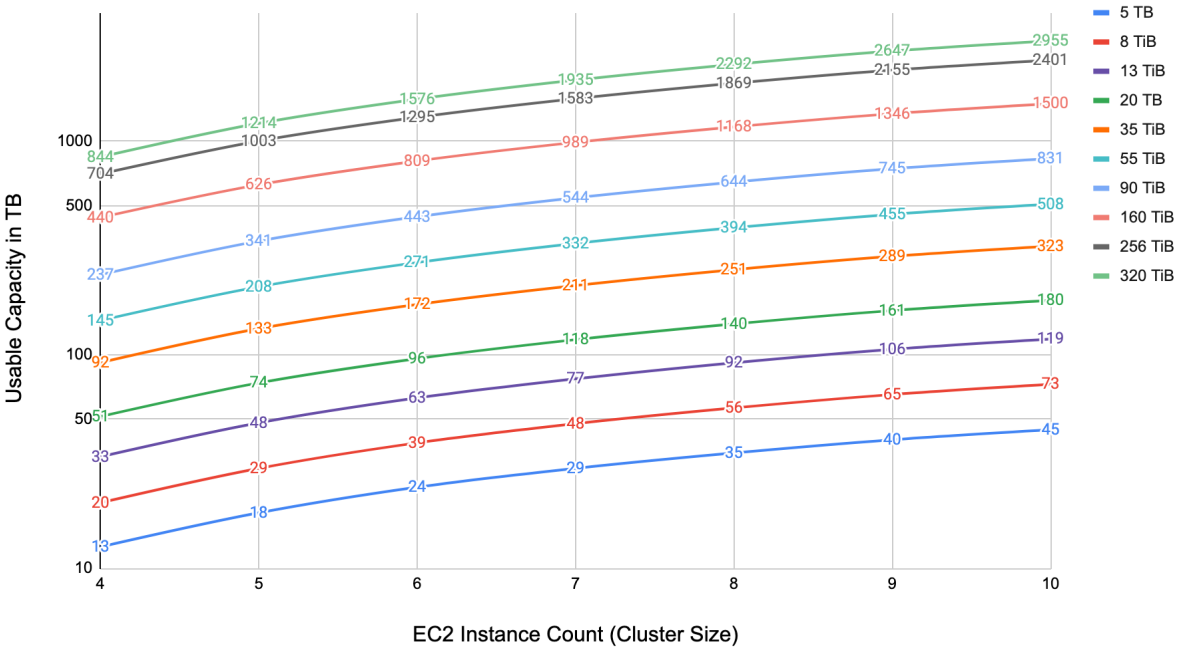
Usable Capacity

Clusters formed with more instances make more efficient use of the EBS volumes provisioned because the Erasure Code is more efficient. Configurations as small as 1.4TB and as large as 6PB are supported with the Terraform provided. The graphs that follow show the usable capacity across All-Flash and Hybrid EBS configurations for 4 to 10 instances with the available Qumulo disk configurations.

Cloud Q All-Flash - Usable Capacity vs Instance Count - by EBS Config



Cloud Q Hybrid - Usable Capacity vs Instance Count - by EBS Config



Performance - Throughput

Single Stream Throughput

Single stream throughput, read or write, is limited to 600MB/s or lower if an instance type and EBS configuration won't support that upper bound. This number equates to the AWS 5 Gbps single TCP flow rate-limit enforced outside of an EC2 placement group.

All-Flash Multi-stream Read Throughput

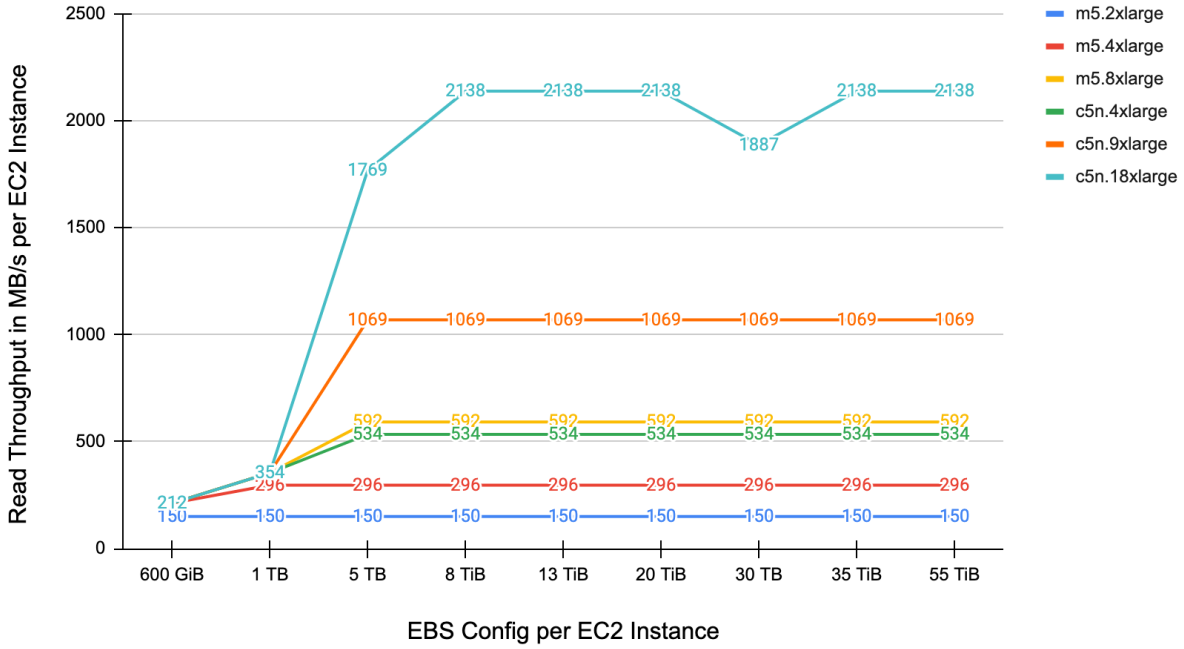
The graphs that follow show the multi-stream read performance per EC2 instance in the cluster. Qumulo requires a minimum of 4 instances (nodes) in a cluster. Performance varies with EBS volume configuration and EC2 instance type. Smaller instance types have less network bandwidth and less EBS bandwidth subjecting them to burst credits. Smaller EBS configurations are also subject to burst credits. For guaranteed performance, respective of baseline IOPS, choose at least a c5n.4xlarge instance type. Then adjust the instance type to increase throughput.
All Flash architectures should be chosen for high throughput workloads, especially in smaller usable capacity clusters, or highly random workloads. IOPS is another factor to consider for small file workloads, small usable capacity clusters, or long sustained workloads. See the IOPS section.

All-Flash Multi-stream Write Throughput

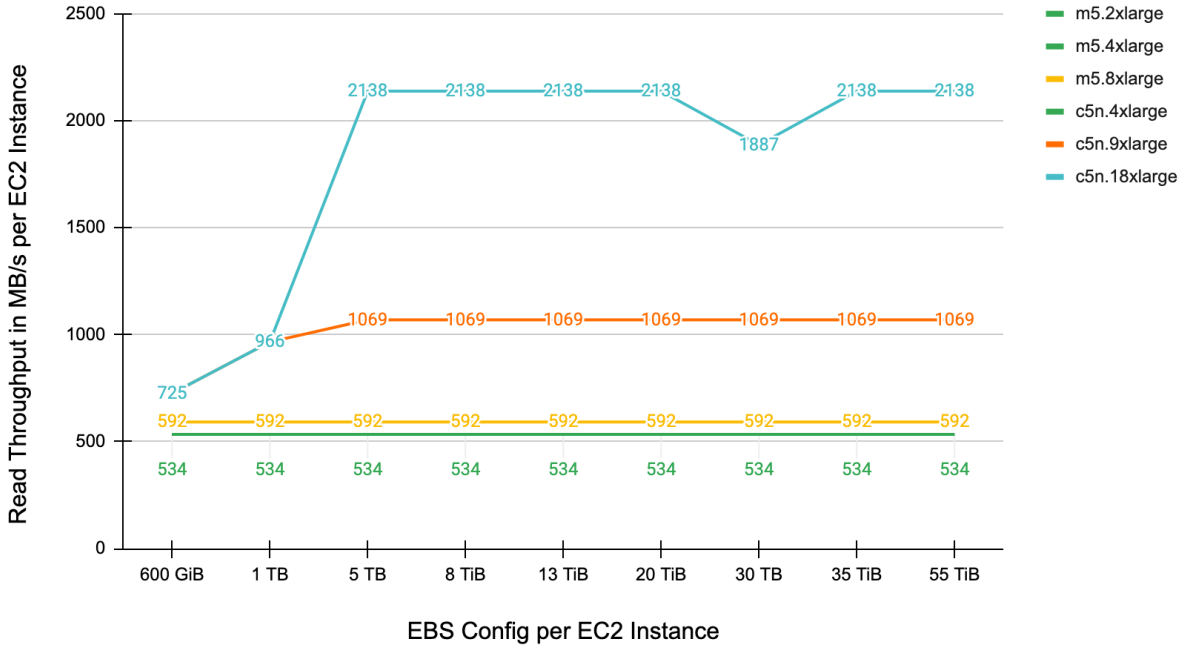
Read performance scales linearly with instance count. Write performance is impacted by the erasure code efficiency factor because the cluster has to write the parity blocks for the code. Clusters coded with 16,14 vs 6,4 not only make more efficient use of the underlying EBS volumes, but write performance is also enhanced. Write performance can be derived for a given erasure code by multiplying the read performance by the EC efficiency factor:

Write Perf for existing files = Read Perf x EC Eff

Cloud Q All-Flash - BASELINE Read Perf per Instance vs EBS Config - By Instance Type

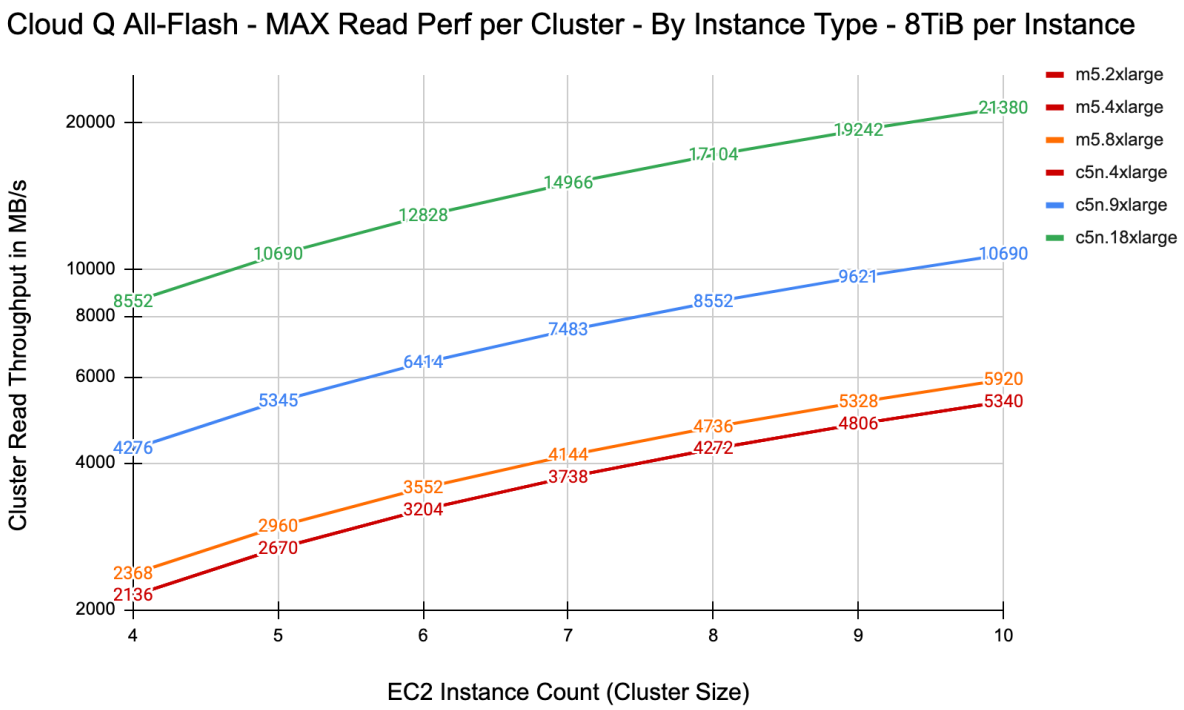


Cloud Q All-Flash - BURST Read Perf per Instance vs EBS Config - By Instance Type



8TiB All-Flash Multi-stream Read MAX Throughput Example

To exemplify how Qumulo clusters perform as they scale, an 8TiB All Flash EBS config is used in the charts that follow for 4 to 10 node cluster configurations. EC2 instance types are graphed across the cluster sizes showing the performance differences.



Hybrid Multi-stream Cached Read Throughput

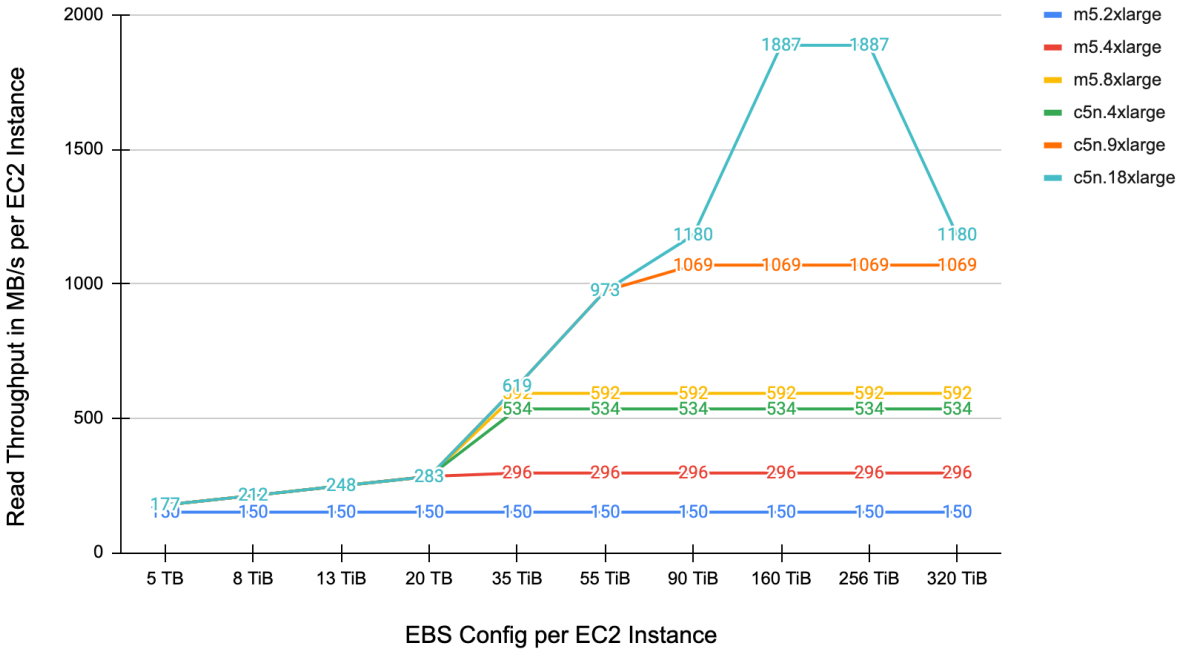
The graphs that follow show the multi-stream cached read performance per EC2 instance in the cluster. Qumulo requires a minimum of 4 instances (nodes) in a cluster. The cluster is a flash-first architecture with all writes landing on flash. The flash is not included in the usable capacity and statistically at least 80% of reads will come from cache due to Qumulo's AI/ML read promotion algorithms. Performance varies with EBS volume configuration and EC2 instance type. Smaller instance types have less network bandwidth and less EBS bandwidth subjecting them to burst credits. Smaller EBS configurations are also subject to burst credits. For guaranteed performance, respective of baseline IOPS, choose at least a c5n.4xlarge instance type. Then adjust the instance type to increase throughput.
Hybrid architectures reduce AWS infrastructure costs and offer performance similar to all-flash for most workloads. This is especially true for large usable capacity hybrid clusters. IOPS for hybrid clusters are lower than equivalent all-flash clusters. See the IOPS section.

Hybrid Multi-stream Write Throughput

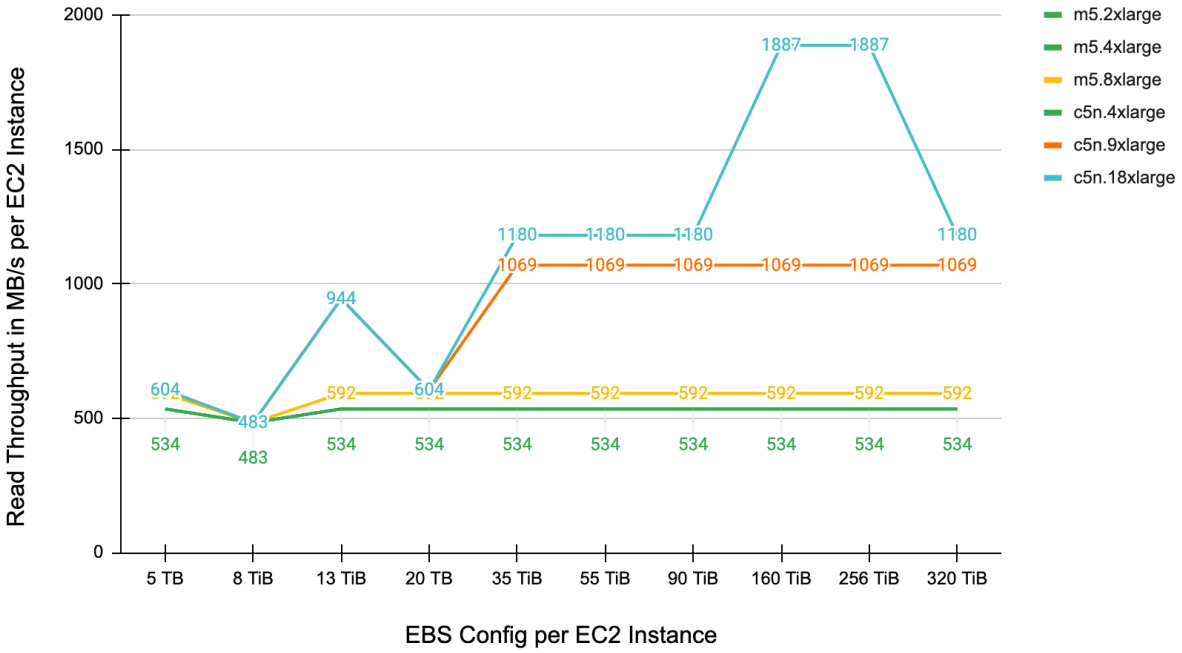
Read performance scales linearly with instance count. Write performance is impacted by the erasure code efficiency factor because the cluster has to write the parity blocks for the code. Clusters coded with 16,14 vs 6,4 not only make more efficient use of the underlying EBS volumes, but write performance is also enhanced. Write performance can be derived for a given erasure code by multiplying the read performance by the EC efficiency factor:

Write Perf for existing files = Read Perf x EC Eff

Cloud Q Hybrid - BASELINE Cached Read Perf per Instance vs EBS Config - By Instance Type

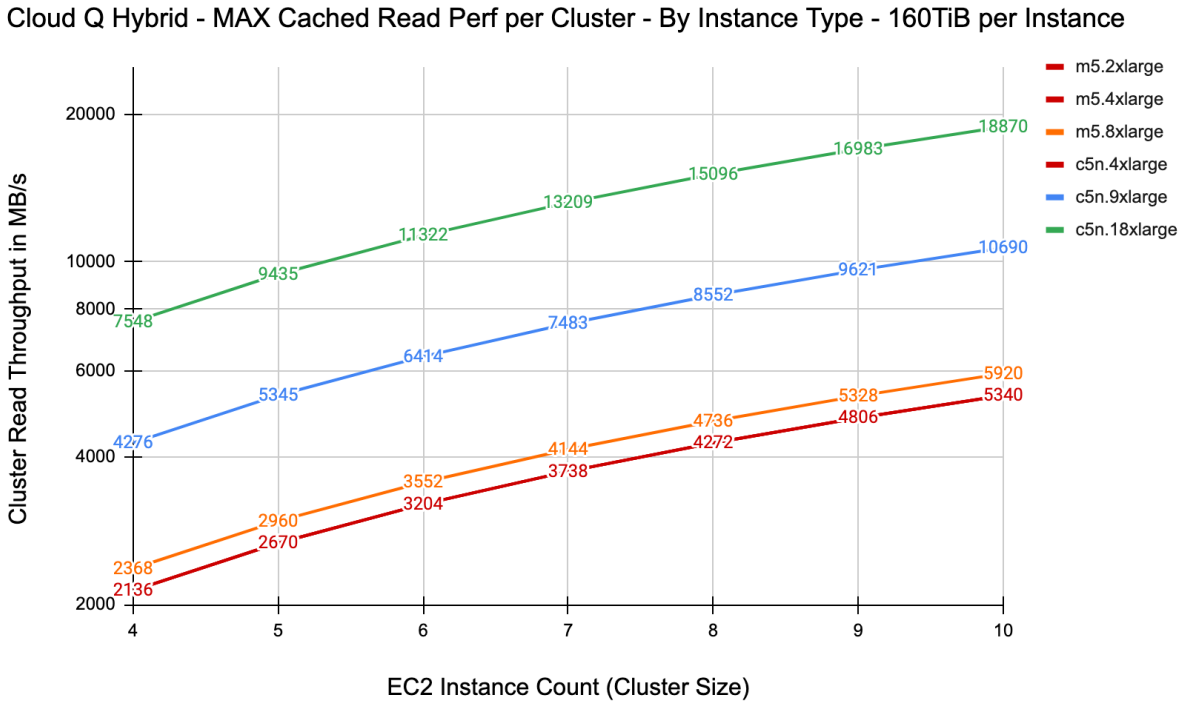


Cloud Q Hybrid- BURST Cached Read Perf per Instance vs EBS Config - By Instance Type



160TiB Hybrid Multi-stream Cached Read MAX Throughput Example

To exemplify how Qumulo clusters perform as they scale an 160TiB EBS config is used in the chart below for 4 to 10 node cluster configurations. EC2 instance types are graphed across the cluster sizes showing the performance differences.

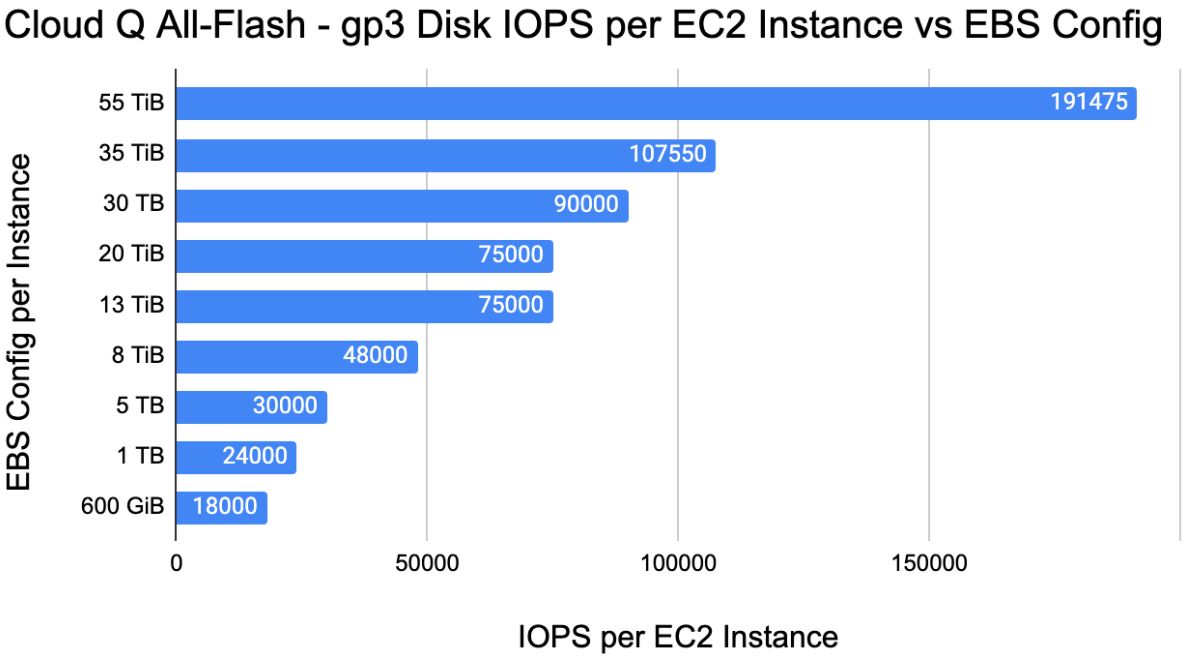


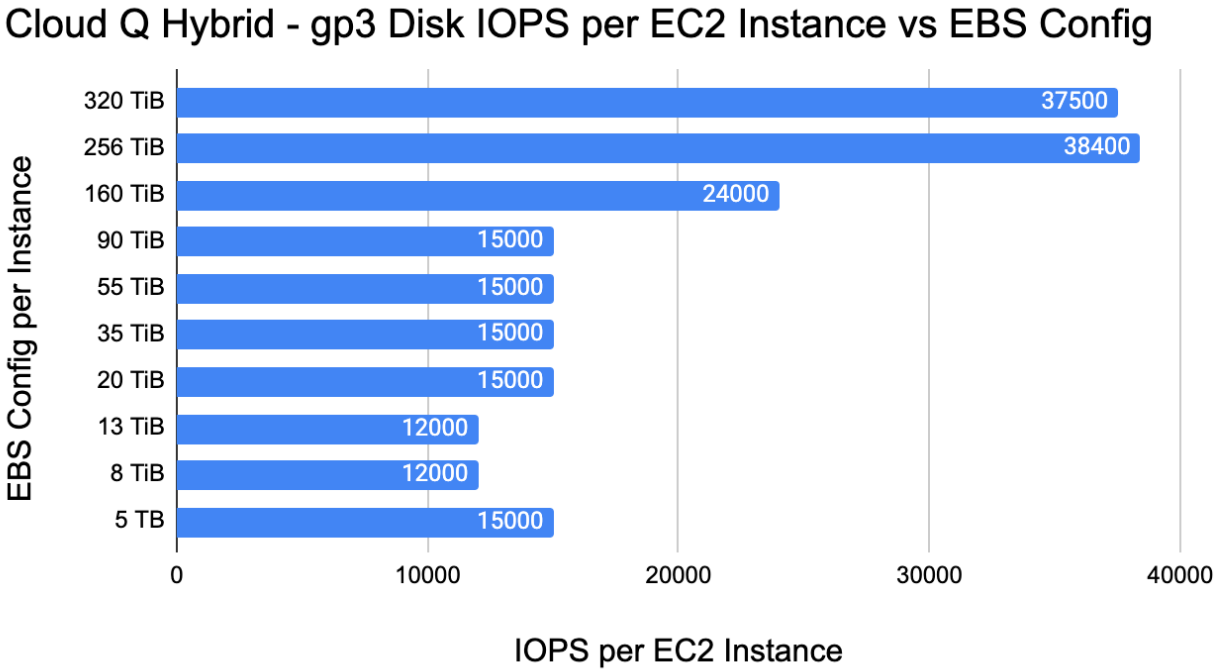
Performance - EBS Disk IOPS

EBS gp3 Disk IOPS

Qumulo uses gp3 EBS volumes by default for flash based EBS volumes. The default IOPS for gp3 is 3000 and throughput is 125MB/s. Qumulo always adjusts the throughput to 250MB/s to match gp2 throughput. Larger gp3 volumes require some additional IOPS to meet gp2 full burst IOPS. AWS Disk IOPS can be inspected in CloudWatch with the resource groups built by the Terraform provided. The charts below illustrate the gp3 EBS Disk IOPS available per EC2 instance for each disk configuration.

Note: EBS Disk IOPS DO NOT equal protocol IOPS shown in the Qumulo UI.





The table below summarizes the IOPS and throughput Qumulo provisions for gp3 for all of the supported EBS configurations.

Type	Disk Config	gp3 EBS Volume Settings			# gp3 vols per EC2	Total per EC2	
		Size GiB	IOPS	Tput MB/s		IOPS	Size TiB
All Flash	600 GiB	100	3000	250	6	18000	0.586
	1 TiB	125	3000	250	8	24000	1.0
	5 TB	500	3000	250	10	30000	4.883
	8 TiB	512	3000	250	16	48000	8.0
	13 TiB	533	3000	250	25	75000	13.0
	20 TiB	820	3000	250	25	75000	20.0
	30 TB	3750	11250	250	8	90000	29.297
	35 TiB	1434	4302	250	25	107550	35.0
	55 TiB	2553	7659	250	25	191475	55.0
Hybrid-st1	5 TB	100	3000	250	5	15000	0.488
	8 TiB	150	3000	250	4	12000	0.586
	13 TiB	175	3000	250	4	12000	0.684
	20 TB	160	3000	250	5	15000	0.781
	35 TiB	350	3000	250	5	15000	1.709
	55 TiB	550	3000	250	5	15000	2.686
	90 TiB	900	3000	250	5	15000	4.395
	160 TiB	1000	3000	250	8	24000	7.813
	256 TiB	1600	4800	250	8	38400	12.5
	320 TiB	2500	7500	250	5	37500	12.207
Hybrid-sc1	8 TiB	150	3000	250	4	12000	0.586
	13 TiB	175	3000	250	4	12000	0.684
	20 TB	160	3000	250	5	15000	0.781
	35 TiB	350	3000	250	5	15000	1.709
	55 TiB	550	3000	250	5	15000	2.686
	90 TiB	900	3000	250	5	15000	4.395
	160 TiB	1000	3000	250	8	24000	7.813
	256 TiB	1600	4800	250	8	38400	12.5
	320 TiB	2500	7500	250	5	37500	12.207

EBS gp2 Disk IOPS

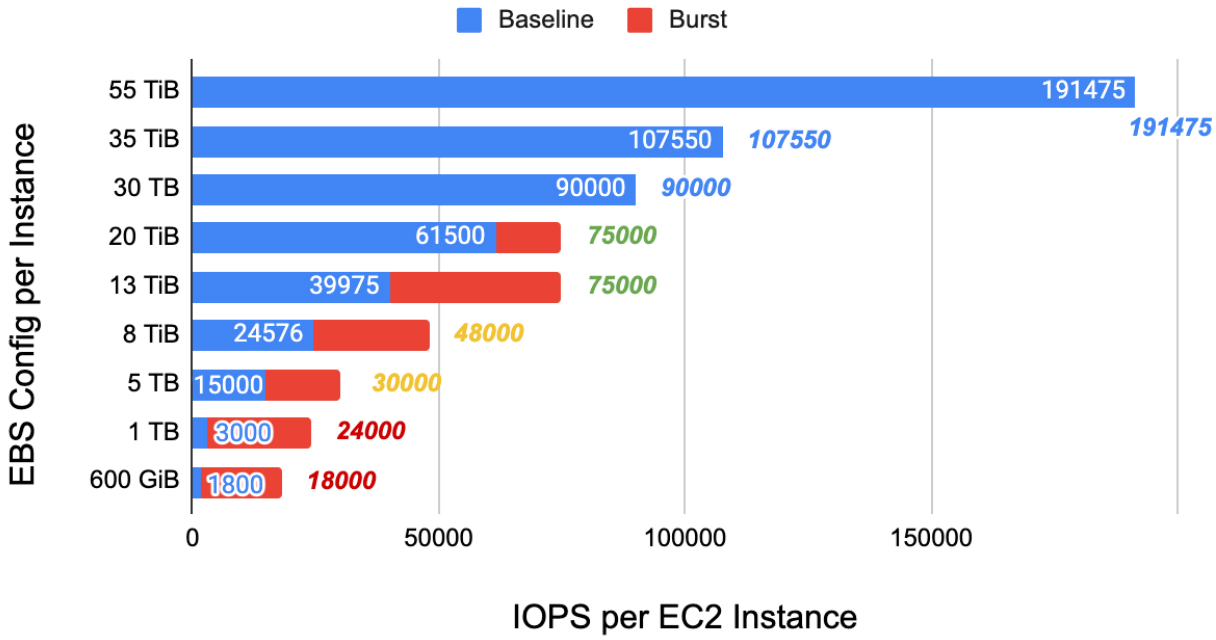
Qumulo leverages AWS EBS gp2 for flash-based EBS volumes in regions, local-zones, or Outposts where gp3 may not be available. gp2 EBS volumes are subject to burst credits. This is even more pronounced for small volumes. The EBS configurations offered by Qumulo balance volume size, the number of volumes, and the baseline vs burst credits available. Every EBS configuration leveraged by a Qumulo node (EC2 instance) has multiple EBS volumes per EC2 instance. The specific configurations complicate the topic, so the focus here is on the disk IOPS offered per instance for the given EBS configurations. Larger EBS configurations completely eliminate burst credits for disk IOPS. AWS Disk IOPS can be inspected in CloudWatch with the resource groups built by the Terraform provided.

Note: EBS Disk IOPS DO NOT equal protocol IOPS shown in the Qumulo UI.

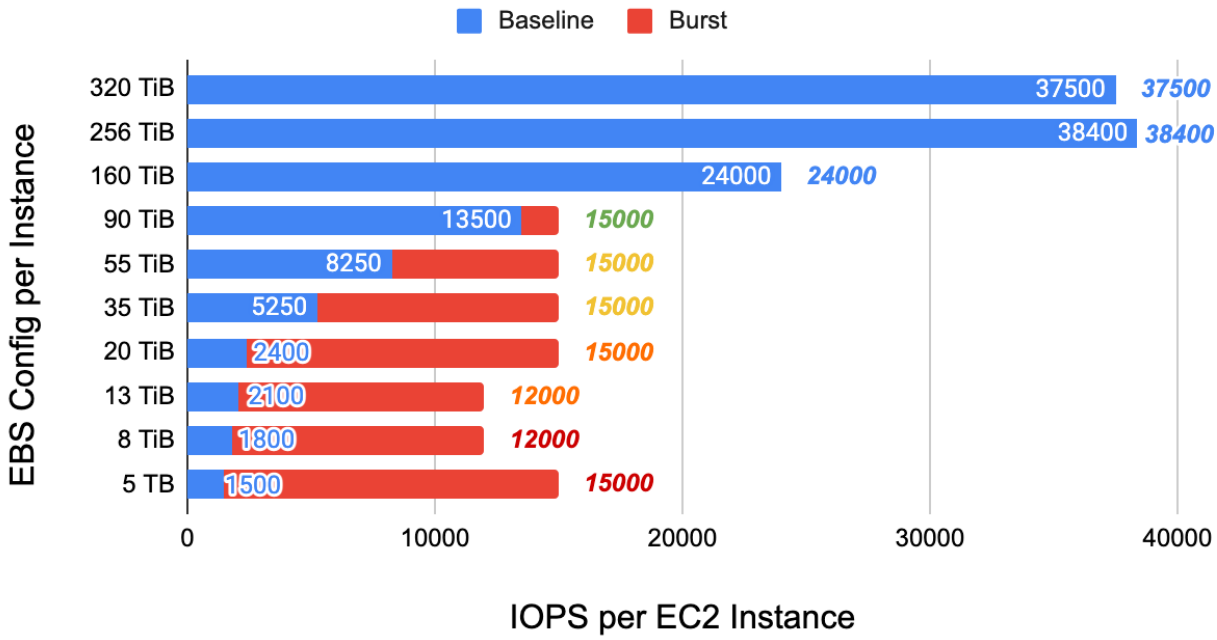
In the charts that follow total disk IOPS are shown italicized to the right of each bar. These numbers are shown per EC2 instance in the cluster across all the supported EBS configurations. The color coding for total disk IOPS indicates the duration that max IOPS can be sustained before burst credits will be exhausted.

Text Color Code	Max IOPS duration
RED	30 minutes
ORANGE	35 minutes
YELLOW	45 minutes
GREEN	2+ hours
BLUE	No limit

Cloud Q All-Flash - Disk IOPS per EC2 Instance vs EBS Config

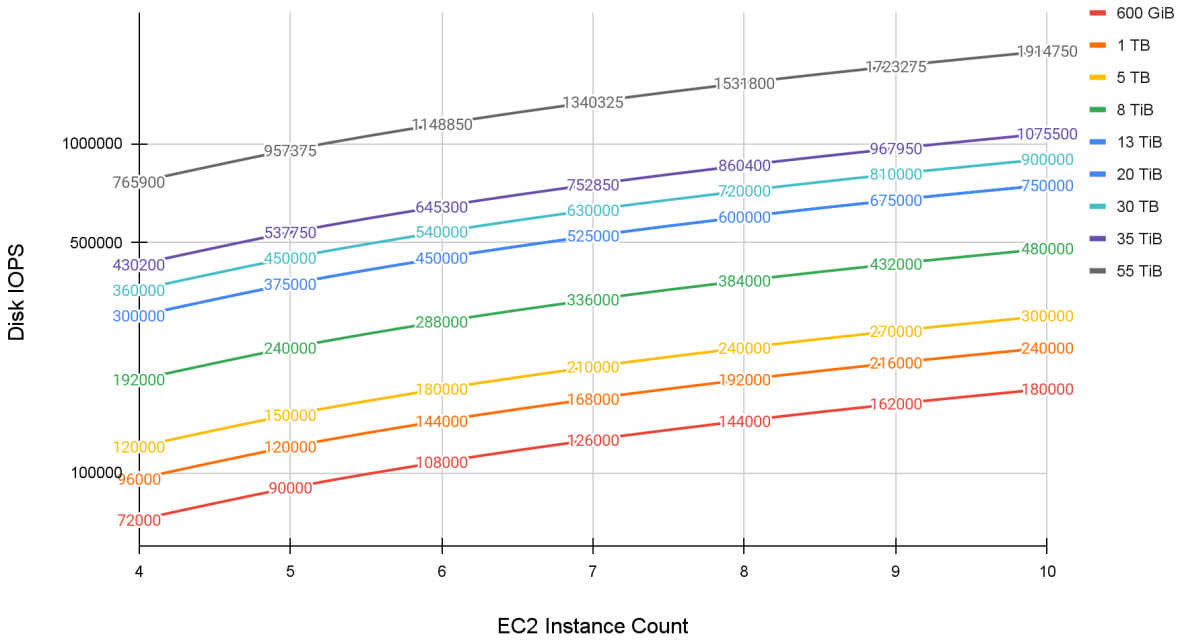


Cloud Q Hybrid - Disk IOPS per EC2 Instance vs EBS Config



MAX gp2 or Guaranteed gp3 EBS Disk IOPs per Cluster

Cloud Q All-Flash - MAX Disk IOPs per Cluster - by EBS Config



Cloud Q Hybrid - MAX Disk IOPs per Cluster - by EBS Config