Red Hat OpenShift Container Storage

External-mode performance characterization for databases and analytics

Red Hat OpenShift Container Storage external mode benefits:

Accelerate performance for input/output (I/O)-intensive apps like those based on PostgreSQL databases.

Access pre-existing Red Hat Ceph Storage clusters directly, eliminating the need to move data.

Share the same Red Hat Ceph Storage cluster across multiple Red Hat OpenShift Container Platform clusters.

Exploit petabyte scalability for data-intensive apps.

Isolate storage fault domains to meet governance needs while offering consistent OpenShift Container Storage semantics.

Table of contents

Executive summary

The growing embrace of hybrid cloud environments is placing increased focus on cloud-native storage. Developers and data scientists need to architect and deploy container-native storage rapidly and per their organization's governance needs. They need the ability to launch their data-intensive apps and machine learning (ML) pipelines where it makes sense within the hybrid cloud, with a fundamental assumption of consistent data storage services. Those services must be inherently reliable and able to scale both in capacity and performance independent of the applications they support.

Red Hat® OpenShift® Container Storage now offers solution architects multiple deployment options to suit their specific needs. Traditional internal mode storage offers simplified management, deployment speed, and agility by deploying OpenShift Container Storage pods within the Red Hat OpenShift Container Platform cluster. OpenShift Container Storage external mode[1] supports Red Hat Ceph® Storage clusters that are separate and independent from the Red Hat OpenShift Container Platform cluster that consumes the storage. OpenShift Container Storage external mode option can serve multiple purposes, including:

▸ Providing centralized storage for multiple Red Hat OpenShift Container Platform clusters.

▸ Scaling or tuning Red Hat Ceph Storage outside the specifications available within an internal mode OpenShift Container Storage cluster.

▸ Utilizing an existing Red Hat Ceph Storage cluster.

▸ Sharing a Red Hat Ceph Storage cluster between multiple Red Hat OpenShift Container Platform clusters.

To characterize performance for different OpenShift Container Storage modes, Red Hat engineers deployed both database and analytics workloads on a Red Hat OpenShift Container Platform cluster built with the latest Intel Xeon processors and Intel solid state drive (SSD) technology. For PostgreSQL testing, the external Red Hat Ceph Storage cluster achieved greater performance and scalability, aided by high-performance Intel Optane SSD Data Center (DC) Series for storing Ceph metadata. For less I/O-intensive analytics workloads, testing showed that both internal and external mode configurations provided roughly equal performance when normalized for the number of drives.

OpenShift Container Storage internal and external mode deployment options

Earlier releases of OpenShift Container Storage were focused on a fully containerized Ceph cluster run with a Red Hat OpenShift Container Platform cluster, optimized as necessary to provide block, file, or object storage with standard 3x replication. While this approach made it easy to deploy a fully integrated Ceph cluster within a Red Hat OpenShift environment, it presented several limitations:

▸ External Red Hat Ceph clusters could not be accessed, potentially requiring redundant storage.

▸ A single OpenShift Container Storage cluster was deployed per Red Hat OpenShift Container Platform instance, so the storage layer could not be mutualized across multiple clusters.

▸ Internal mode storage had to respect some limits in terms of total capacity and number of drives, keeping organizations from exploiting Ceph's petabyte-scale capacity.

---

1 *External mode is available with Red Hat OpenShift Container Storage 4.5 or later. See the OpenShift Container Storage documentation for more information.*

OpenShift Container Storage external mode overcomes these issues by allowing OpenShift Container Platform to access a separate and independent Red Hat Ceph Storage cluster (Figure 1). Together, with traditional internal-mode storage, solutions architects now have multiple deployment options to address their specific workload needs. These options preserve a common, consistent storage services interface to applications and workloads, while each provides distinct benefits:

Internal mode schedules applications and OpenShift Container Storage pods on the same Red Hat OpenShift cluster, offering simplified management, deployment, speed, and agility. OpenShift Container Storage pods can either be converged onto the same nodes or disaggregated on different nodes, allowing organizations to balance Red Hat OpenShift compute and storage resources as they like.

External mode decouples storage from Red Hat OpenShift clusters, allowing multiple Red Hat OpenShift clusters to consume storage from a single, external Red Hat Ceph Storage cluster that can be scaled and optimized as needs dictate.
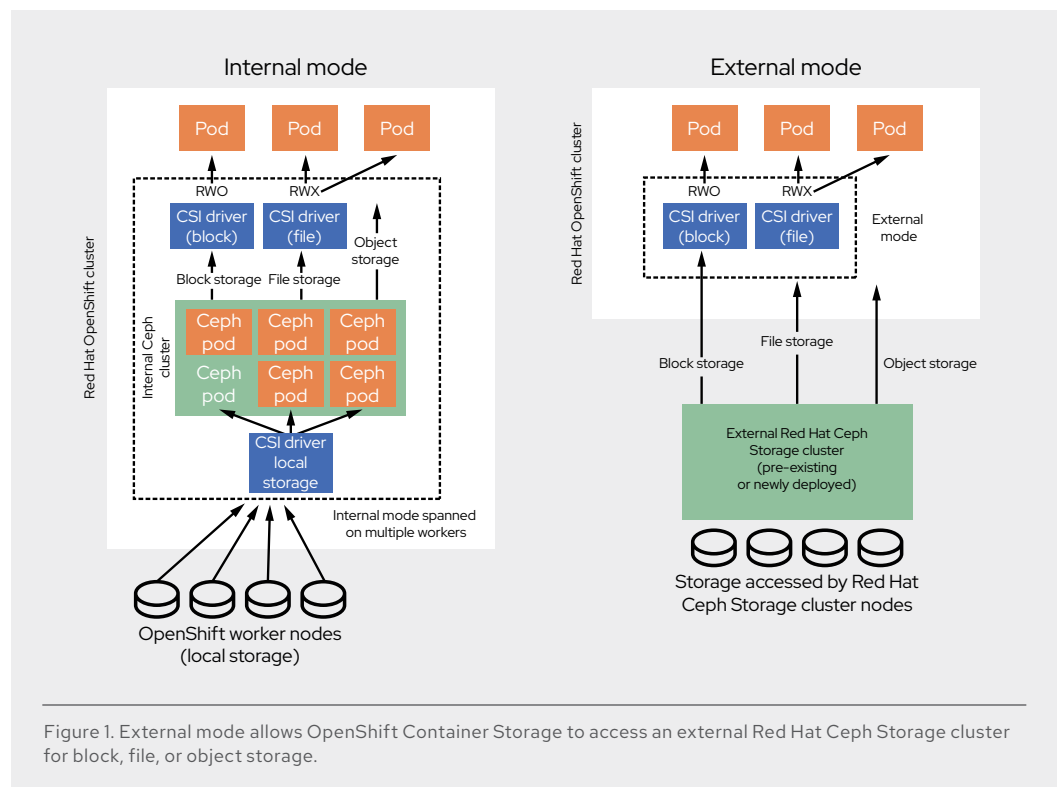

Figure 1. External mode allows OpenShift Container Storage to access an external Red Hat Ceph Storage cluster for block, file, or object storage.

Workloads and cluster configuration

Because OpenShift Container Storage internal and external mode architectures differ in component configuration, data path, and networking, engineers wanted to verify that they provide the same predictable and scalable Ceph performance across multiple workloads.

Workload descriptions

To help guide organizations with architectural choices, Red Hat engineers wanted to compare the performance of two types of workloads — databases and analytics — across both internal- and external-mode configurations. The following workloads were evaluated:

▸ **Database workload.** The database workload consisted of multiple PostgreSQL database instances queried by a Sysbench multithreaded workload generator tool called Sherlock. The workload uses a block-based access method, and the principal measurement metric is aggregated transactions per second (TPS) across all databases.

▸ **Analytics workload.** The analytics workload consisted of a selection of 56 I/O-intensive workloads from the TPC-DS benchmark, run by Starburst Enterprise for Presto. This workload exercised object performance, with the primary metric being the time taken to complete all the queries (execution time).

Cluster configuration

Both workloads were run on the same Red Hat OpenShift Container Platform cluster, configured alternately with internal mode and external mode (Table 1 and Figure 2). The two configurations were chosen based on the following criteria:

▸ Internal mode OpenShift Container Storage clusters are deployed within a Red Hat OpenShift cluster, either integrated within the same worker nodes that are serving other applications, or, as in this case, disaggregated on infrastructure nodes to provide separate application and data scalability. A three-node OpenShift Container Storage cluster was used for internal mode testing.

▸ External mode storage clusters are massively scalable, support mixed media types, expanded tuning options, and are able to support multiple Red Hat OpenShift clusters. A six-node OpenShift Container Storage cluster was used for external-mode testing, with different numbers of nodes used for different tests. The external mode configuration also evaluated placing Red Hat Ceph Storage metadata on separate high-performance Intel Optane SSDs.

Table 1. Test cluster with internal mode and external mode storage plane options

	Control plane	Compute plane	Storage plane	
Resources	Helper node, bootstrap node, Red Hat OpenShift Container Platform master node	Red Hat OpenShift Container Platform worker nodes	OpenShift Container Storage internal cluster nodes	OpenShift Container Storage external cluster nodes

	Control plane	Compute plane	Storage plane	
Node count	1+1+3	4	3	6
CPU	2x Intel Xeon CPU E5-2699 v4 @2.20 GHz		2x Intel Xeon Gold 6142 CPU @2.60 GHz	
Memory	192GB	256GB	192GB	192GB
Metadata drives	N/A	N/A	2x Intel SSD D5-P4320 Series (7.68TB, 2.5-inch, PCIe 3.1 x4, 3D2, QLC)	2x Intel Optane SSD DC P4800X Series (750GB, ½ height PCIe x4, 3D XPoint)
Data drives	N/A	N/A		8x Intel SSD D5-P4320 Series (7.68TB, 2.5-inch PCIe 3.1 x4, 3D2, QLC)
Red Hat OpenShift cluster network	1x 25Gb Ethernet	1x 25Gb Ethernet	1x 25Gb Ethernet	2x 25Gb Ethernet SFP28 (bonded
Provisioning network	1 x 1Gb Ethernet	1 x 1Gb Ethernet	1 x 1Gb Ethernet	
Operating system	Red Hat Enterprise Linux® CoreOS	Red Hat Enterprise Linux® CoreOS	Red Hat Enterprise Linux® CoreOS	Red Hat Enterprise Linux 8.2
Red Hat OpenShift Container Platform version	4.5.2	4.5.2	4.5.2	N/A
OpenShift Container Storage version	N/A	N/A	4.5	4.5

Figure 2. Red Hat test cluster configuration
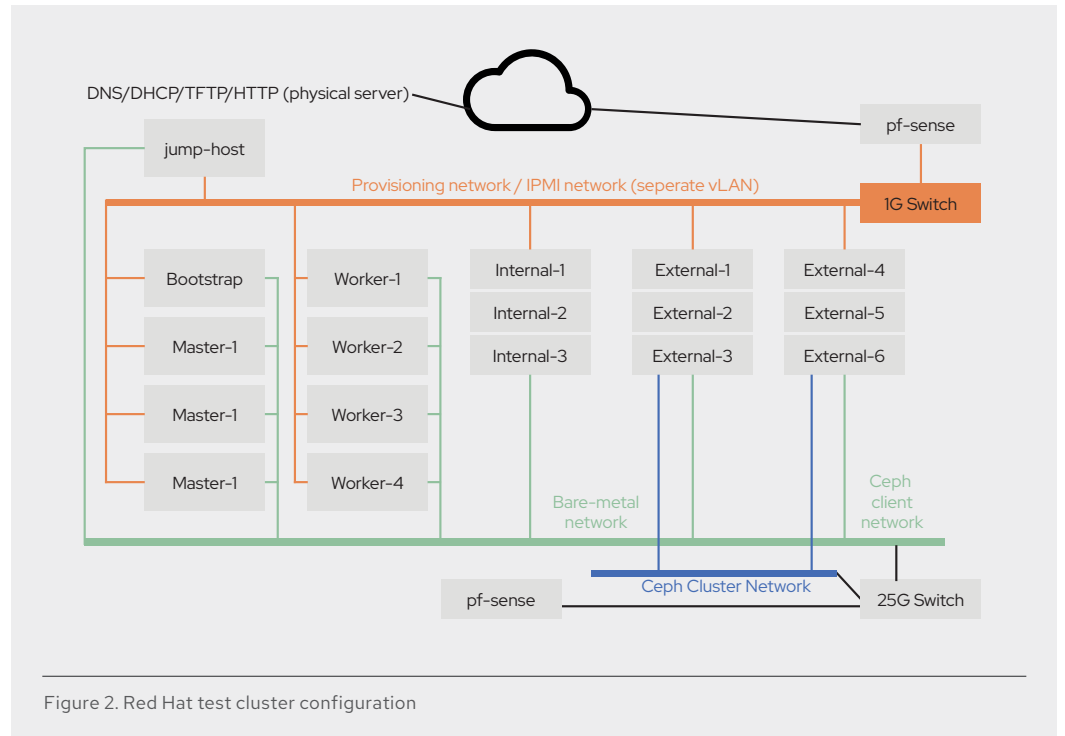
Intel Xeon processors and SSDs

Intel technologies played a significant role in Red Hat testing. The test cluster employed the latest Intel Xeon Scalable processors and Intel Solid State Drive (SSD) technology, including:

▸ **Intel Xeon processor E5 family.** Designed for architecting next-generation data centers, Intel Xeon processor E5 family delivers versatility across diverse workloads in the datacenter or cloud. The Red Hat OpenShift Container Platform cluster under test ran on servers powered by Intel Xeon processor E5-2699 v4.

▸ **Intel Xeon Scalable processors.** OpenShift Container Storage servers used servers equipped with two Intel Xeon Gold 6142 CPUs. These processors deliver workload-optimized performance and advanced reliability with the highest memory speed, capacity, and interconnects.

▸ **Intel SSD D5-P4320 Series.** Intel SSD D5-P4320 Series were used as combined data and meta-data drives for the internal-mode testing, and as data drives for the external-mode testing. As the industry's first PCIe-enabled QLC drive for the datacenter, the Intel SSD D5-P4320 Series delivers big, affordable, and reliable storage. With 33% more bits than TLC, the Intel QLC NAND SSDs enable triple the storage consolidation compared to hard disk drives (HDDs), leading to lower operational costs.

▸ **Intel Optane SSD DC P4800X Series.** For maximum scalability and reliability, Intel Optane SSD DC P4800X Series were used for metadata storage in external-mode testing. Hosting metadata on high-speed storage is an effective means of enhancing Red Hat Ceph Storage performance. The Intel Optane SSD DC P4800X represents an ideal storage tier that provides high performance, low latency, and predictably fast service to improve response times. Important for Ceph metadata, the Intel Optane SSD DC P4800X Series provides high endurance suitable for write-intensive applications like online transaction processing (OLTP). Able to withstand heavy write traffic, the Intel SSD DC P4800X Series is ideal for a Red Hat Ceph Storage cluster utilized by OpenShift Container Storage external mode.

PostgreSQL testing

To evaluate database performance within Red Hat OpenShift Container Platform, engineers employed PostgreSQL on the test cluster, measuring performance using both OpenShift Container Storage internal mode and external mode. Testing simulated multiple databases being queried by multiple clients for a variety of query types. To generate the workload, open source Sherlock scripts were used to launch multiple PostgreSQL databases and to exercise them using the scriptable Sysbench multi-threaded benchmark tool.

Default and tuned OpenShift Container Storage internal-mode testing

Even within a Red Hat OpenShift Container Platform cluster, tuning options can result in performance gains for data-intensive applications like PostgreSQL. Engineers compared a default OpenShift Container Storage configuration with a tuned configuration. The two internal-mode configurations were equipped as follows:

▸ **Default configuration.** 2x CPUs, 6GB per Ceph object storage daemon (OSD)

▸ **Tuned configuration.** 6x CPUs, 16GB per Ceph OSD

Figure 3 shows that aggregated TPS across all of the database instances was dramatically higher with the tuned configuration than with the default configuration. Moreover, performance in the tuned configuration scaled nicely as the number of worker nodes and PostgreSQL databases quadrupled while performance for the default configuration remained roughly the same. These results demonstrate that tuning can use high-performance devices such as Intel SSD D5-P4320 Series to achieve higher storage performance for PostgreSQL.

Figure 3. Tuned internal-mode configurations can provide performance scalability for PostgreSQL.
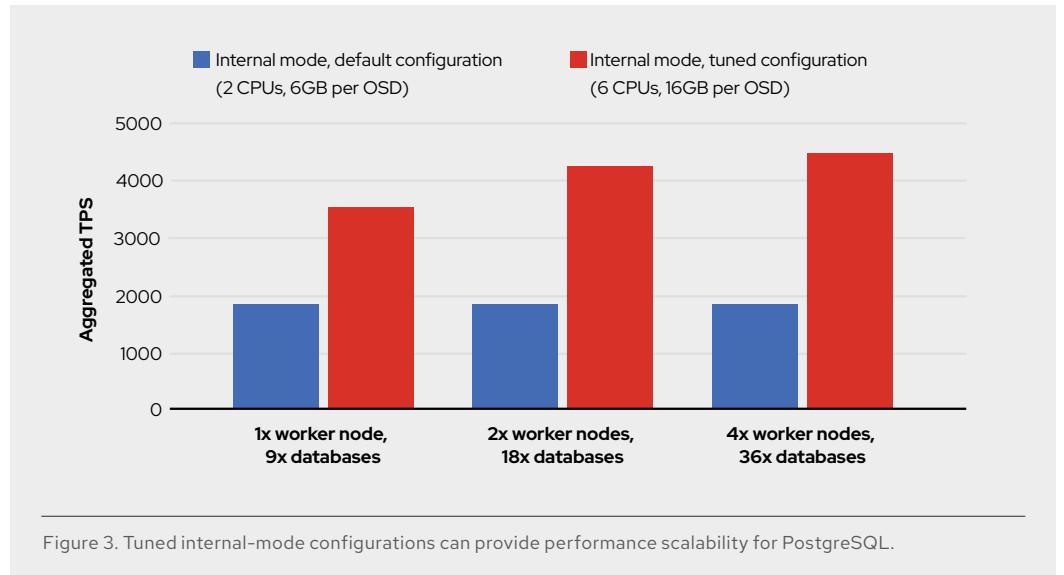
Comparing OpenShift Container Storage internal and external modes

Engineers then compared the same two internal-mode performance results with an OpenShift Container Storage external mode configuration, using the external Red Hat Ceph Storage cluster.[2] For the external-mode cluster, resources were increased to 8 CPUs and 18GB per OSD. Ceph meta-data was stored on two Intel Optane SSD DC P4800X on each server.

The resulting performance testing yielded higher aggregated TPS above the tuned internal-mode configuration for a given number of databases (Figure 4). PostgreSQL performance also scaled as the number of worker nodes and databases was quadrupled.

---

2 *Though the external Red Hat Ceph Storage cluster contained six nodes, engineers only used three nodes for performance testing in the interest of consistency with the internal-mode tests.*

Figure 4. PostgreSQL on different OpenShift Container Storage deployment modes
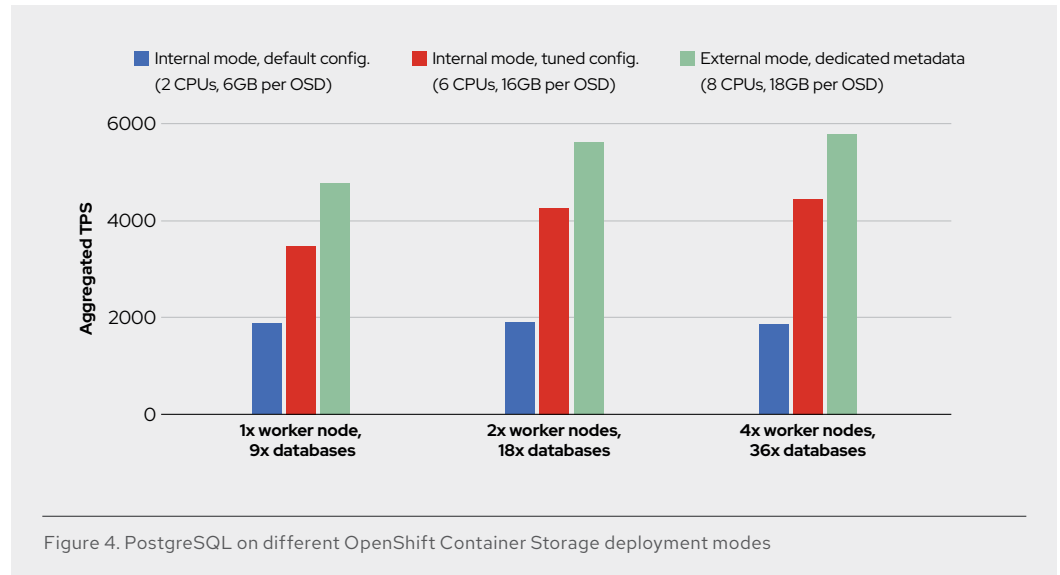
TPC-DS testing

The TPC Benchmark DS (TPC-DS) is a standardized benchmark that provides a representative evaluation of performance as a general-purpose decision support system. It includes data sets and queries that allow benchmarking of infrastructure by reporting query response times. This test is a good fit for simulating analytics workloads in a reproducible and objective manner. Out of the 100 queries provided by TPC-DS, Red Hat engineers selected 56 of the most I/O-intensive queries to evaluate OpenShift Container Storage.

Starburst Enterprise for Presto

In Red Hat testing, the selected TPC-DS queries were run by Starburst Enterprise for Presto—a modern solution built on the open source Presto distributed SQL query engine. It harnesses the value of open source Presto, the fastest distributed query engine available today. Presto was designed and written for interactive analytic queries against data sources of all sizes, ranging from gigabytes to petabytes. Starburst Enterprise for Presto approaches the speed of commercial data warehouses while scaling dramatically.

The Starburst Enterprise for Presto platform provides distributed query support for varied data sources, including:

▸ Apache Cassandra

▸ Hive (Apache Hadoop Distributed File System, HDFS)

▸ Amazon Simple Storage Service (S3, HDFS)

▸ Microsoft SQL Server

▸ MySQL

▸ PostgreSQL

TPC-DS data generation

Before the TPC-DS queries could be run, the cluster needed to be populated with data. The data generation methodology applied was the same for both OpenShift Container Storage internal mode and external mode, as follows:

▸ A Presto cluster was deployed on the worker nodes using the Starburst Presto Kubernetes operator (running on four nodes).

▸ Eight Presto workers with 40 CPUs (fixed number, request=limit) and 120GB of RAM were used to generate the data, using the TPCDS connector included with Starburst Enterprise for Presto. This connector creates synthetic data sets at various scale factors (data set sizes).

▸ The data sets created were of various sizes (1GB, 10GB, 100GB, 1TB, 3TB, 10TB), and creation time was measured for each of these.

Figure 5 illustrates data generation times measured on both the internal-mode cluster and the external-mode cluster. In general, data generation times were found to be highly similar for the two storage options.
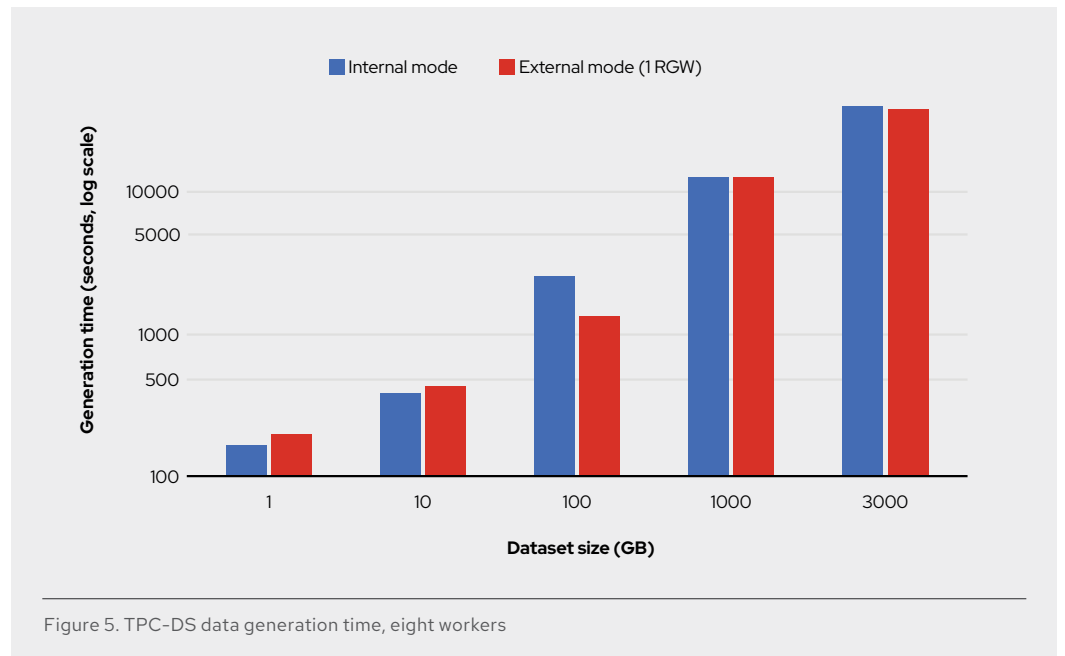

Figure 5. TPC-DS data generation time, eight workers

TPC-DS query performance

Once the data was generated, engineers ran the subset of 56 TPC-DS queries with the two different Presto configurations against the data sets created on the OpenShift Container Storage internal-mode and external-mode clusters. Engineers used two configurations to run the queries:

▸ Eight Presto workers with 40 CPUs and 120GB of RAM

▸ Four Presto workers with 80 CPUs and 240GB of RAM

For the OpenShift Container Storage external mode tests, two different configurations of RADOS Gateway (RGW) access were then tested: a single RGW as an S3 endpoint, and 12 RGWs behind a load-balancer acting as the S3 endpoint. As shown in Figure 6, external mode runs generally performed at parity or slightly better than the results from all of these runs, demonstrating that organizations can reap the benefits of OpenShift Container Storage external mode without sacrificing performance.

Figure 6. TPC-DS benchmark, various runs (lower is better)
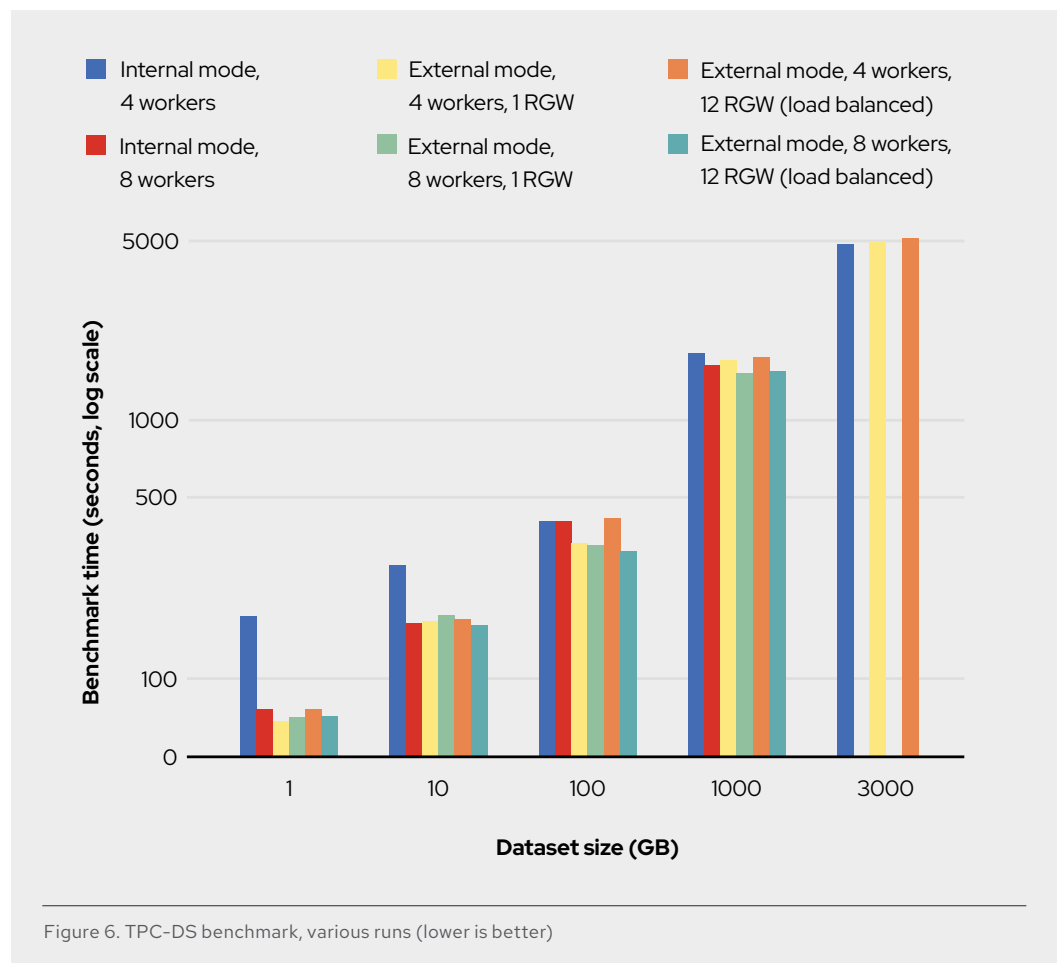
Conclusion

Data scientists and cloud-native app developers need their apps to work the same, wherever they are on the hybrid cloud, without variability in storage services capabilities. By offering a choice of internal-mode and external-mode storage for OpenShift Container Storage, Red Hat gives organizations the flexibility they need without sacrificing performance.

Optimized external mode configurations can help data-intensive apps like those based on PostgreSQL to improve performance and scale predictably as the number of worker nodes and databases grows. For less data-intensive applications such as the TPC-DS queries evaluated herein, organizations can enjoy the flexibility of external mode operation without compromising performance.

About Red Hat

Red Hat is the world's leading provider of enterprise open source software solutions, using a community-powered approach to deliver reliable and high-performing Linux, hybrid cloud, container, and Kubernetes technologies. Red Hat helps customers integrate new and existing IT applications, develop cloud-native applications, standardize on our industry-leading operating system, and automate, secure, and manage complex environments. Award-winning support, training, and consulting services make Red Hat a trusted adviser to the Fortune 500. As a strategic partner to cloud providers, system integrators, application vendors, customers, and open source communities, Red Hat can help organizations prepare for the digital future.

facebook.com/redhatinc
@RedHat
linkedin.com/company/red-hat

NORTH AMERICA	**EUROPE, MIDDLE EAST, AND AFRICA**	**ASIA PACIFIC**	**LATIN AMERICA**
1 888 REDHAT1	00800 7334 2835 europe@redhat.com	+65 6490 4200 apac@redhat.com	+54 11 4329 7300 info-latam@redhat.com

redhat.com
#F26686_0121