Choosing an Interface

SUSE® Enterprise Storage is a versatile, self-healing, and fault-tolerant storage alternative. The Ceph storage system embedded in SUSE Enterprise Storage is an object-based storage environment that hides most the details of the storage process from outside applications and users. The client saves the data, and Ceph takes care of the rest.

SUSE

Choosing an Interface

The only place where the client actually interacts with Ceph is through an interface. Ceph has a powerful collection of interfaces and gateways that let the cluster serve a number of different roles. The Ceph cluster will appear differently to the network depending on what you need it to do and how it will need to fit in with surrounding components. For instance, if the storage cluster is part of a network that is designed to interact with a block storage system, Ceph's RADOS Block Device (RBD) interface is a logical alternative. On the other hand, if client systems on your network access storage through a network-based SMB file server, the SUSE® Enterprise Storage CIFS/SMB gateway might be the best option.

SUSE Enterprise Storage interface and gateway alternatives fall in three basic categories:

- *block storage*
- *filesystem storage*
- *object storage*

This paper offers a summary of some of the most important interface options and provides some insights on when to use each of the alternatives.

Block Storage

The RADOS Block Device (RBD) interface is one of the first interfaces supported by Ceph, and it remains the most popular interface alternative. RBD appears to the client as a network block storage device. RBD interacts directly with the cluster, thus providing better performance than many of the higher-level gateway alternatives. RBD is typically the default option for a SUSE Enterprise Storage interface if the user does not have a specific need for the features provided with the other interface alternatives.

In the most basic configuration, a Linux client system mounts the RBD block device directly, but RBD also serves as a foundation for several other Ceph block-storage configurations. For instance, RBD can act as an interface to OpenStack through the OpenStack Cinder block storage service. Ceph's iSCSI gateway is a useful option if you wish to mount the Ceph cluster as a block-storage device on a Windows and VMware client system.

SUSE Enterprise Storage's block storage options are also used extensively in virtualization scenarios. The Qemu over RBD configuration lets a Linux-based host access the Ceph cluster and make the data available to Windows, Linux, and MacOS virtual guest systems.

Filesystem

SUSE Enterprise Storage also offers tools that let you present the Ceph cluster to the network through a filesystem interface. The filesystem scenario fits in with the conventional model of the network storage system as a file server. If you have legacy applications that expect to look for data in the familiar array of files and folders associated with a filesystem structure, a filesystem interface is the best choice. Internet service providers and other admins who wish to divide the storage space into a group of smaller home directories typically choose a filesystem alternative.

CephFS is a high-performing, POSIX-compliant filesystem specifically designed and optimized to serve as a native filesystem in Ceph environments. The CephFS filesystem, which has been growing in popularity recently due to an expanding feature set and performance improvements, is the standard choice for a filesystem interface to SUSE Enterprise Storage.

CephFS directly supports Linux as a client system. Organizations that wish to provide filesystem access to a diverse collection of Windows, Linux, and Mac OS clients running on a greater network have a pair of options:

- **NFS to CephFS (Ganesha gateway)**—*the cluster appears to the network as an NFS file server.*
- **CIFS/SMB Gateway**—*clients can access the cluster through the popular CIFS/SMB family of file service protocols created by Microsoft and supported by most computer systems.*

NFS and CIFS/SMB are the most common file service protocols on modern networks. These gateway options make it possible for Ceph to support a diverse collection of network clients without the need for retooling or reconfiguration on the client side.

Because these gateway options add an extra layer to the interface stack, the NFS and CIFS/SMB gateways typically are not as fast as a direct connection through CephFS. When SUSE Enterprise Storage provides the back-end storage for High-Performance Computing (HPC) systems, CephFS is the preferred method for making storage resources available to HPC nodes. CephFS is also an ideal interface for data analytics workloads through Hadoop.

The NFS and CIFS/SMB gateways are not used for situations that require low latency or fast response time, such as storage for an HPC application; however, for everyday scenarios that do not require low latency, such as a system of user home directories, these gateways are popular for their convenience and minimal configuration on the client side.

Object Storage

Ceph uses an object storage architecture on the back end, so an object storage interface is one of the more efficient ways to interact with the Ceph cluster (see the box entitled "What is Object Storage?").

SUSE Enterprise Storage offers object-based access to Ceph through a RADOS Gateway interface (called RGW). RGW is actually an HTTP server that interacts with the Ceph cluster and provides a means for accessing storage through a scalable web service architecture.

What is Object Storage?

A conventional block storage environment breaks data into blocks of fixed length and uses a centralized system for storing the blocks. When a file is opened, the filesystem retrieves the blocks one by one, reassembling them into the original file. The process of breaking up and assembling blocks complicates the process of storing and retrieving data. Also, the filesystem must maintain some form of centralized table or broker with knowledge of where the chunks are stored, which can lead to bottlenecks.

Object storage offers a more scalable approach. In an object storage environment, the size of the data can vary, thus eliminating the need to break the file into fixed-size blocks. Knowledge of where the data is stored is derived through hash calculation and distributed throughout the network, which eliminates the need for a central broker or lookup table that could potentially form a bottleneck.

Because the RGW interface is designed to adhere to the principles of REST, a popular methodology for building efficient and scalable services using web components (see the box entitled "Building for REST"), RGW is an ideal choice for scenarios that build Ceph access into automated systems and orchestration architectures. For instance, RGW provides easy and convenient integration with cloud-based S3 (Simple Storage Service) implementations. RGW also integrates easily with OpenStack environments through the OpenStack Swift object-based storage API.

If you wish to connect the Ceph cluster with S3 or OpenStack resources, or if you plan to integrate Ceph with your own automation environment using scripting or another form of middleware system integration, RGW is the best option. RGW with the S3 API

is a popular option for backup and archival use cases because it is efficient at transferring data across long distances.

RGW also supports a gateway to the NFS network filesystem, which makes it possible to integrate the benefits of RGW automation and web services while still providing access to network clients through a conventional filesystem.

Building for REST

Representational State Transfer (REST) is a set of design principles for creating and interfacing web service components. Services that conform to REST guidelines (often referred to as RESTful services) support a uniform and predefined set of stateless operations. Requests for a resource are made using Uniform Resource Identifier (URI) standards. The uniformity of the transactional process, and the prohibition of arbitrary, complex operations, ensures that REST services easily interact with each other with minimal programming time. The stateless nature of REST adds reliability and improves performance at high traffic levels, ensuring that a misbehaving client will not threaten the stability of the overall system.

All or One?

With all the different gateway and interface options available through SUSE Enterprise Storage, one frequent question is whether a Ceph cluster can support more than one of the interface options discussed in this paper. For instance, is it possible for a single cluster to support RESTful gateway connections through S3 and also provide block storage through RBD? The answer is yes, you can organize the Ceph cluster so that it supports connections through multiple interfaces, although the data is stored separately within the cluster. Thus, multiple interfaces can access the same cluster but not the same data.

Ceph organizes storage into data pools. You can assign different data pools to each interface, thus allowing the cluster to support multiple interfaces at once. Supporting multiple interfaces through a single cluster simplifies administration and configuration, allowing all storage to reside within a single environment regardless of the form or function.

Conclusion

SUSE Enterprise Storage is a versatile storage system that can serve a number of different roles on the network. You can mount the SUSE Enterprise Storage cluster as a virtual disk, connect to it as a network filesystem, or access it through a RESTful web service interface. Table 1 on the following page provides a summary of the most common SUSE Enterprise Storage interface options.

The best solution for your organization will depend on the nature of your data and the details of your network configuration, but the following tips apply to many of the most common configurations:

- *For basic access and good performance with Linux client workstations, use the RBD block-storage interface.*
- *If you prefer a block-storage solution, but you have Windows or MacOS clients that will need to mount the cluster directly as a block device, use the iSCSI gateway.*
- *If you would like the storage cluster to appear to the network as a POSIX-compliant filesystem (similar to a file server), use the CephFS filesystem interface. Additional gateways for NFS and CIFS/SMB export the filesystem to a diverse collection of network clients with minimal need for additional client configuration.*
- *If you wish to interface your SUSE Enterprise Storage cluster with S3 or OpenStack Swift, or if you plan to integrate the cluster with an automated web-service infrastructure, use the object-based RGW interface.*

The experts at SUSE will help you understand the SUSE Enterprise Storage interface options and choose a solution that is tailored to the needs of your network.

Storage Type	Scenario	Solution
Block Storage	Linux clients mount the cluster directly as a block device	RBD
	Linux virtualization host makes Ceph block storage available to virtual guest systems	Qemu over RBD
	Clients access Ceph as network-based iSCSI block device	iSCSI gateway
	VMware or Windows-based virtualization host makes Ceph block storage available to virtual guest systems	iSCSI gateway
	Block storage with OpenStack environment	RBD—each storage pool is mapped to an OpenStack Cinder back end
Filesystem	Linux clients mount the cluster as a POSIX-complaint filesystem	CephFS
	NFS clients on the network access Ceph as a network filesystem	NFS to CephFS (Ganesha gateway)
	CIFS/SMB clients on the network access Ceph as a network filesystem	CIFS/Samba gateway
Object Storage	RESTful gateway for remote or cloud access through S3	RGW with S3 API
	Object Storage with OpenStack	RGW with OpenStack Swift API
	NFS network clients access Ceph through object interface	NFS gateway (NFS to RGW)

Table 1. *Storage Interface Options*

Additional contact information and office locations:
www.suse.com

www.suse.com