Whitepaper

Futurewei Solution for OpenStack and Software Defined Storage (SDS) Integration Plan

Trademarks and Permissions

FUTUREWEI Technologies and other Futurewei trademarks are trademarks of Futurewei Technologies. Huawei trademarks are trademarks of Huawei Technologies Co., Ltd. All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Futurewei Technologies, Inc.

Boston Research Center

Address: 111 Speen Street, Suite 114
Framingham, MA 01701
United States of America

Website: http://www.futurewei.com/

Contents

Table of Figures

Revision History

V1	Dec, 2019

Introduction

Enterprise deployments require reliable, high performance solutions that are also flexible, agile, and cost efficient. In response to the rapidly changing requirements of the data center, many IT leaders have found the promise of open-source and commoditized infrastructure attractive to deliver the features that they require, but are concerned about integration, operational costs, and maintaining enterprise-level reliability.

Cloud computing provides an efficient way to organize different resources and provides abundant services for users of an IT infrastructure. The Infrastructure as a service (IaaS) layer supports essential capabilities for computation, network, and storage. OpenStack is one of the most popular open-source software platforms for cloud infrastructure. It has been widely adopted by customers and cloud providers all over the world. There is also a trend among service providers to provide software-defined features. For instance, Software Defined Storage (SDS) is intended to reduce the total cost of ownership by utilizing commodity hardware and installing storage functional software on it. Compared to traditional storage arrays, SDS provides better scalability and agility, and also can provide a higher total throughput than traditional storage solutions.

At the same time, SDS systems are trying to meet the needs that customers requested for traditional arrays. Huawei's FusionStorage is designed and tested for enterprise workloads, supporting features that open source solutions are lack of. With many levels of Erasure Coding support, customer's critical data is protected with cost in mind. With snapshots and synchronous replications, customer's business continuity is ensured. With the traditional feature support but with a distributed architecture, the FusionStorage nodes can be serviced, or be upgraded and replaced online while the cluster is running. Therefore, the life of the cluster can be prolonged tremendously, thus protecting customer's investment.

OpenStack and SDS can be naturally combined to provide a flexible yet powerful solution for customers and cloud vendors. If deployed and tuned correctly, OpenStack software running on SDS will provide great value to the enterprise, and business will benefit from its performance and lower TCO. A well-designed solution plan is the key to bring out its value and a successful deployment.

At Futurewei, we evaluated several configurations for OpenStack using SDS with the goal to find successful options for enterprise deployment. In this white paper, we list the detailed components needed to create an enterprise-grade solution and recommend options for common use cases. We understand that an OpenStack deployment project may not start from scratch, so we also consider interoperability, making this solution a seamless integration with existing systems. We discuss the tradeoffs and best practices when running OpenStack using SDS in different scenarios, such as with iSCSI, or with special clients, and on top of different software stacks. Two major SDS offerings that we cover are Ceph and Huawei's FusionStorage. To ensure the business continuity of the system, we discuss disaster recovery and efficient backup. Finally, we discuss many other integration topics in depth, such as security, monitoring, and automation.

1. OpenStack with software defined storage

OpenStack Introduction

As an open source project, OpenStack was first released in October 2010. After several years of continuous enhancement, OpenStack had become a powerful infrastructure platform with a broad range of features, including Nova (for VMs), Neutron (for SDN), Ironic (for bare metal), Swift (for object storage), and Cinder (for block storage). Figure 1 shows the core OpenStack components. More detailed information about each component can be found on their website [1].

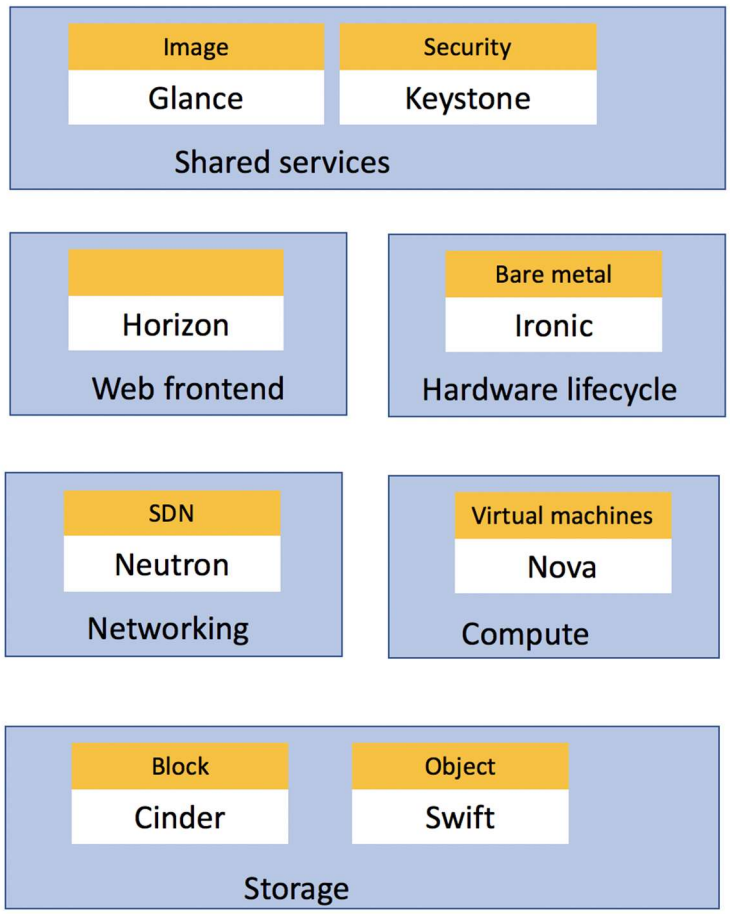

Figure 1 Core OpenStack components

OpenStack enables its users to build up a cloud that provides a rich set of services. Users can also customize the selections to fit their needs. OpenStack can be used to build a private cloud, or to build a public cloud that provides cloud services to the public.

For readers of this white paper, we list some OpenStack components and terms in the following table:

Name		Comments
Horizon	Horizon is the frontend interface to let administrators control the components of OpenStack.	Horizon panels are web based.
Ironic	Ironic is the hardware lifecycle service to bootstrap a baremetal machine.	Ironic can be used to provision baremetal servers.
Nova	Nova is the component to provision compute instances, such as a virtual machine.	Nova runs as daemons on Linux servers. Purposely built Nova clients are available besides OpenStack CLI commands.
Neutron	Neutron is the component to provide "Network as a service" for cloud tenants between their interfaces, such as vNICs.	Use Neutron to build software defined network (SDN) with plug-ins provided by different vendors.
Cinder	Cinder is the component to provide block storage service. Cinder defines a set of APIs for how to consume the block service.	OpenStack has a reference implementation (LVM) and most storage vendors provide Cinder support.
Swift	Swift is component to provide object storage service.	Swift API is partially compatible with S3 API and differences exist.
Keystone	Keystone is the component to provide identity authentication, authorization, and service discovery.	Keystone manages projects, users, and roles. Keystone can use built-in authentication mechanisms or use external authentication.
Glance	Glance is component for image service. Users can use Glance to discover, register, and retrieve VM images	The images can be stored in a variety of storage devices.

Software Defined Storage (SDS) introduction

Software defined storage (SDS) is an industry effort to build abstraction of storage functionalities and provide storage as a service. Therefore, the storage functionalities can be provided on a variety of hardware platforms and configurations. Through this storage virtualization layer, storage functionalities can be configured on demand, and deployed quickly. The cost of total ownership is lower not only through savings due to commercial off-the-shelf hardware, but also through the agility and time saving that SDS brings in.

SDS's vision is complement to what OpenStack Cinder is trying to achieve. Cinder defines a list of APIs so that a variety of storage systems can be connected to OpenStack as block storage. In this white paper, we chose to build solutions on both OpenStack Cinder and SDS because together they provide the scalability, performance, and cost-base for our scalable system. We chose two popular

SDS systems, Ceph and Huawei's FusionStorage 8.0, to illustrate the best practices to compose solutions using OpenStack and SDS.

Ceph

Ceph is an open source project that provides block, object, and file services. Although Ceph project is separate from the OpenStack project, Ceph is often used to provide storage in an OpenStack environement because it is one of the best open-source storage products on the market. Ceph's versatility has been proven in that it has been deployed on many platforms, including Kubernetes.

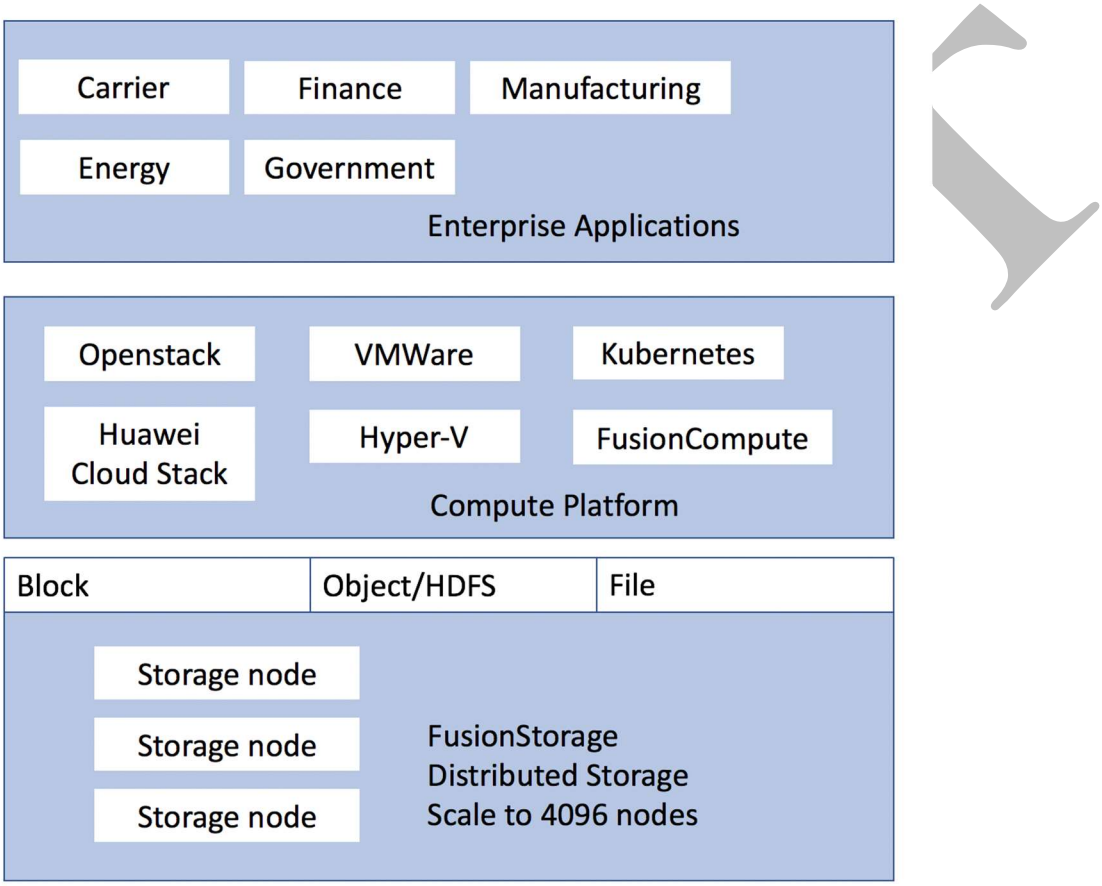

Figure 2 FusionStorage SDS integration diagram

FusionStorage

Huawei's FusionStorage 8.0 is Huawei's latest product to provide SDS functionalities. It can be scaled to 4096 nodes, so that the total storage pool can meet the most demanding sizing requirement. FusionStorage can be deployed on servers from different vendors and on two architectures -- x86 and ARM. FusionStorage provides more storage features provided by open-source projects. For example, FusionStorage provides many Erasure Coding options. FusionStorage also provides a rich set of Enterprise features, from snapshot, clone, backup, to remote replication. This product has been successfully deployed on many mission critical systems, in both private and public clouds.

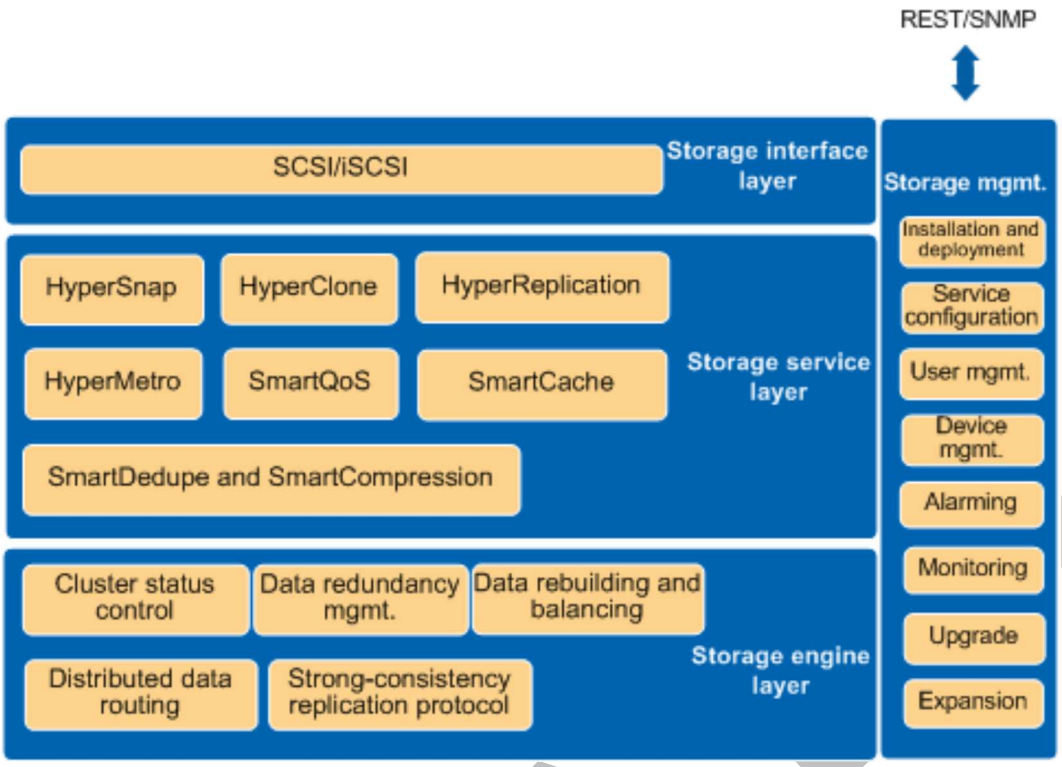

Figure 3 Functional architecture of FusionStorage 8.0

With both products included in our solution, we are confident that our solution contains best products from both worlds – Ceph for open-source projects, and FusionStorage for vendor-provided products. Futurewei solution's goal is to maximize the capability of OpenStack SDS software suite and lower the TCO of customer with selective options.

OpenStack Cinder and SDS requirements

Beyond the requirements for SDS to support data path protocols such as iSCSI, in order to be controlled by OpenStack, an SDS must support OpenStack's control path protocols. Cinder is OpenStack's component that defines the interface for storage providers. The detailed features will be determined by the block storage driver and the driver provided by vendors. The following diagram shows the high-level workflow of how to control Ceph and FusionStorage via Cinder drivers. The Admin user may use the Horizon panel to send control commands to Nova and Cinder. The Cinder scheduler will execute the commands with the support of storage provided drivers.

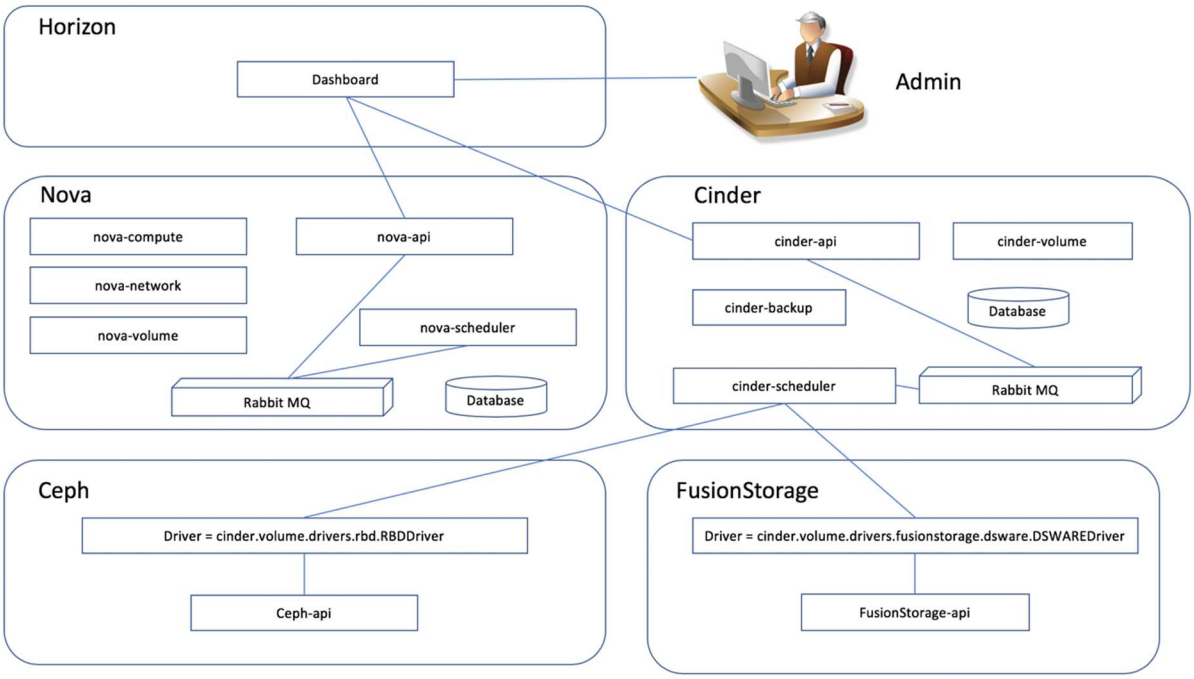

Figure 4 OpenStack SDS integration

Next, we use FusionStorage as an example to show what a Cinder driver provides. The following table shows a supporting matrix of what is supported by FusionStorage's current Cinder driver.

Features	Pike	Queens	Rocky	Stein
Create Volume	√	√	√	√
Delete Volume	√	√	√	√
Attach Volume	√	√	√	√
Detach Volume	√	√	√	√
Extend Volume	√	√	√	√
Create Snapshot	√	√	√	√
Delete Snapshot	√	√	√	√
Create Volume from Snapshot	√	√	√	√
Create Volume from Image	√	√	√	√
Create Volume from Volume	√	√	√	√
Create Image from Volume	√	√	√	√
SmartThin	√	√	√	√
Manage/Unmanage Volume	√	√	√	√
Manage/Unmanage Snapshot	√	√	√	√
Multipath	√	√	√	√
QoS	√	√	√	√
Retype	√	√	√	√
iSCSI	√	√	√	√
SCSI	x	x	√	√
MultiAttach	x	√	√	√
SmartThick	x	x	x	x
SmartPartition	x	x	x	x
HyperMetro	x	x	x	x

Replication V2.1	x	x	x	x
HyperMetro Consistency Group	x	x	x	x
Backup Snapshot	x	x	x	x
Snapshot Consistency Group	x	x	x	x
Consistency Group	x	x	x	x

SmartThin is to support Thin provisioning and SmartThick is for thick devices. SmartPartition is an intelligent cache partition feature to help QoS SLA levels. Finally, HyperMetro is FusionStorage's active-active replication feature set. Some of the features will require the customer to use FusionStorage commands instead of Cinder commands. But the most common Cinder commands are covered.

Detailed steps to install and configure FusionStorage's Cinder driver can be found in FusionStorage OpenStack Driver Configuration Guide [6].

SDS functional components

FusionStorage and Ceph both feature separated control plane and data plane, it is recommended to put data plane on a high-speed network such as at least 10g ethernet or InfiniBand, while control plane can use lower speed network such as traditional gigabit network. Redundancy is required in both control and data plane in production environment. On data nodes, each physical storage device (e.g. disks) are managed by a dedicated daemon (OSD), user data is duplicated and distributed across all data nodes within a failure domain. No single point failure should cause any data service interruption.

In addition to OSD, FusionStorage includes additional storage services to provide enterprise features such as migration and replication, see later chapters for more details, as well as Fusion Storage 8.0 documentation [8].

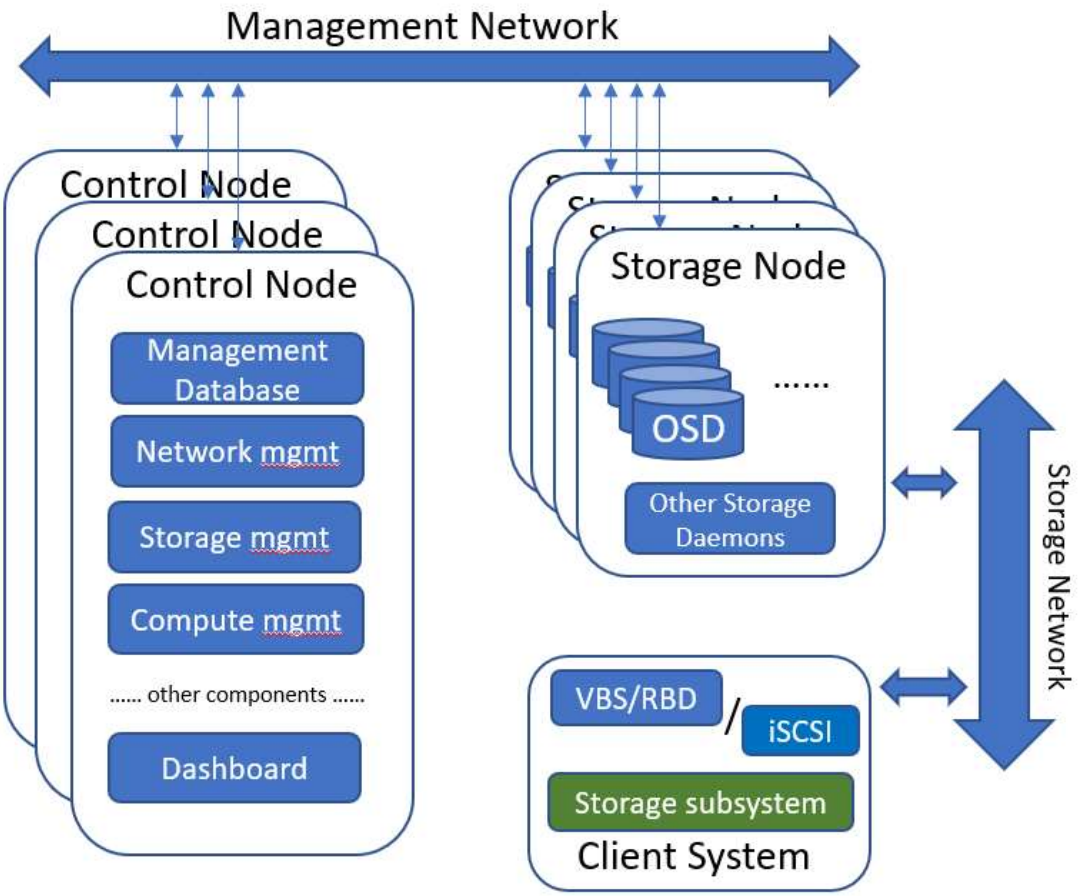

Figure 5 SDS functional components

2. Hardware and Software Selections

OpenStack + FusionStorage configuration

OpenStack SDS is a distributed scale-out solution for a wide range of deploy sizes. Huawei's recommended hardware for this solution is the FusionServer 2288H V5, which has 2 CPUs and can be scaled up to 1.5TiB memory, supporting12-36 disks per node.

Here is a list of certified hardware that are tested for FusionStorage 8.0:

Storage nodes:

Model	Image	CPU	Memory	Storage
2288H V5 12-Slot for HDDs or 12 NVMe SSD		2 * Xeon	Up to 1.5 TiB (64GiB * 24)	12 * 3.5″ HDD / NVME SSD
2288H V5 25-Slot or 24 NVME SSDs		2 * Xeon	Up to 1.5 TiB (64GiB * 24)	25 * 2.5″ HDD / 24 * NVME SSD
5288 V5 36-Slot		2 * Xeon	Up to 1.5 TiB (64GiB * 24)	36 * 3.5″ HDD / NVME SSD
TaiShan 2280 V2 12-Slot		2 * Hi1616 ARM SOC		12 * 3.5″ HDD / NVME SSD
TaiShan 2280 V2 25-Slot		2 * Hi1616 ARM SOC		25 * 2.5″ HDD / 24 * NVME SSD
TaiShan 2280 V2 36-Slot		2 * Hi1616 ARM SOC		36 * 3.5″ HDD / NVME SSD

Network Switches: 

Model	Number of Ports	Primary Speed	Connection Type
CE6855-48S6Q-HI	48	10G	Copper
CE6851-48S6Q-HI	48	10G	Copper
CE6865-48S8CQ-EI	48	25G	Copper
CE6860-48S8CQ-EI	48	25G	Copper
CE6855-48S6Q-HI	48	10G	SFP+
CE6851-48S6Q-HI	48	10G	SFP+
CE6865-48S8CQ-EI	48	25G	SFP+
CE6860-48S8CQ-EI	48	25G	SFP+

OpenStack + Ceph configuration

Ceph supports a wide range of commodity x86 servers, however for the solution's consistency, we recommend using same hardware as Huawei's FusionStorage option.

Best Practice

- A minimum of 2 control nodes per cluster is recommended for FusionStorage, while 3 or more is recommended.
- It is recommended to isolate management network, IPMI network and data network. It is recommended to have a redundant pair of 10G or higher speed switch for data network. Management and IPMI network can use gigabit switch. When distance between nodes are long enough (such as cross rack), SFP+ is preferred over copper connection.
- Number of storage node should be calculated based on planed storage need divided by single node capacity. It is recommended to leave enough spare capacity for failure coverage and redundancy.
-

Following example illustrates an example OpenStack SDS hardware configuration:

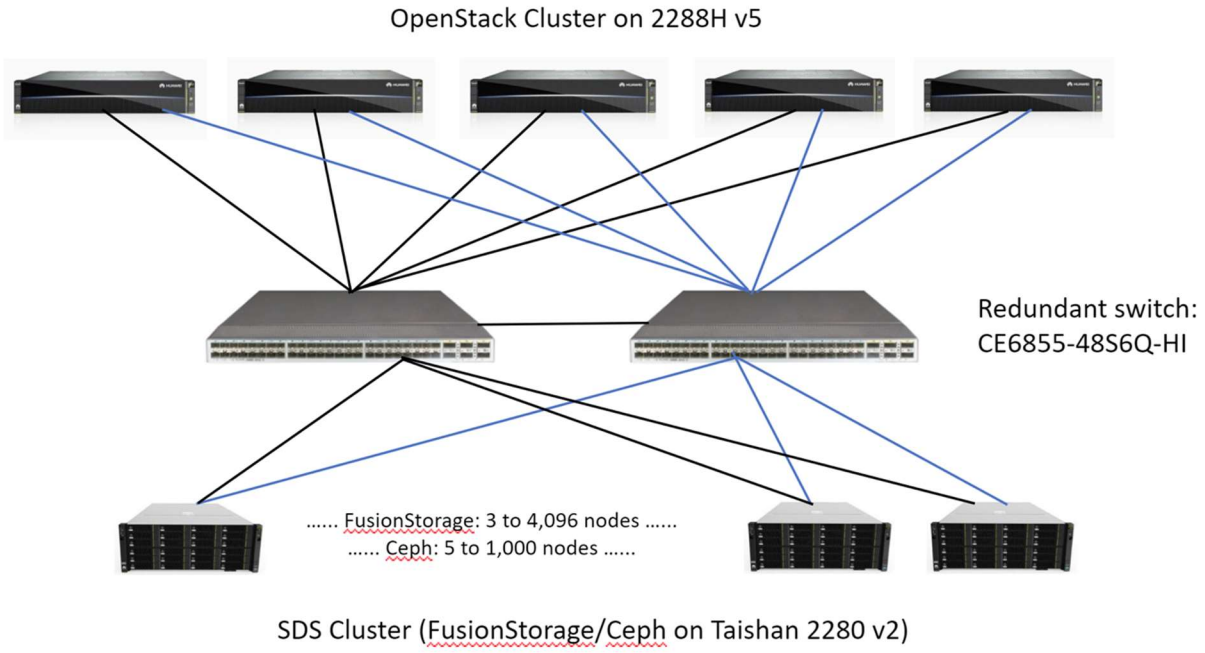

Figure 6 FusionStorage sample hardware configuration

3. Deployment and Integration

In Futurewei, to help customers deploy OpenStack SDS solutions, we developed automated tools and release them publicly under XX license. Anyone can use the auto deployment tools to deploy OpenStack SDS into a new environment or an existing environment. The deploy tool will recognize the system environment, then auto configure a variety of parameters on switches, servers, and storage.

The auto deployment tool is compatible with products from multiple vendors but is best compatible with Futurewei solutions.

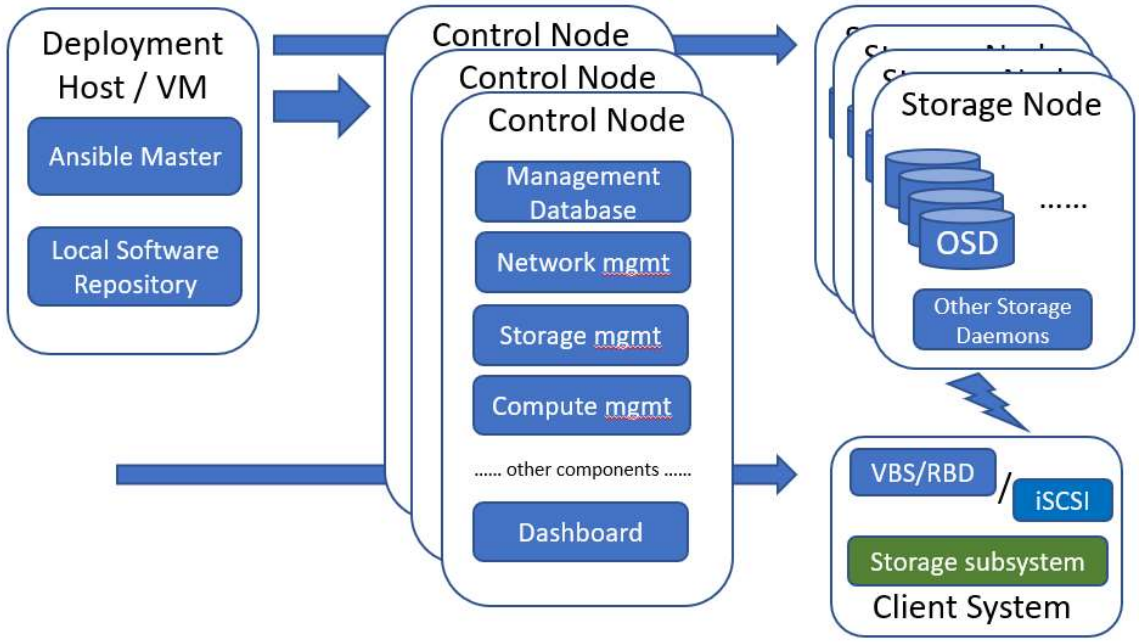

Figure 7 SDS Deploy Environment – Blue components are deployed by FustionStorage / Ceph installation tool

The recommended way to deploy OpenStack SDS is using Ansible playbook. For Ceph, the open source Ceph-Ansible [7] is the recommended tool. For FusionStorage, a custom playbook will be available upon request, and a rich set of CLI/REST API is available for developing your own playbook.

In addition to Ansible, FusionStorage also provides a user-friendly GUI deploy tool in case manual installation is preferred. See "Installation Guide" of FusionStorage [8] for more details.

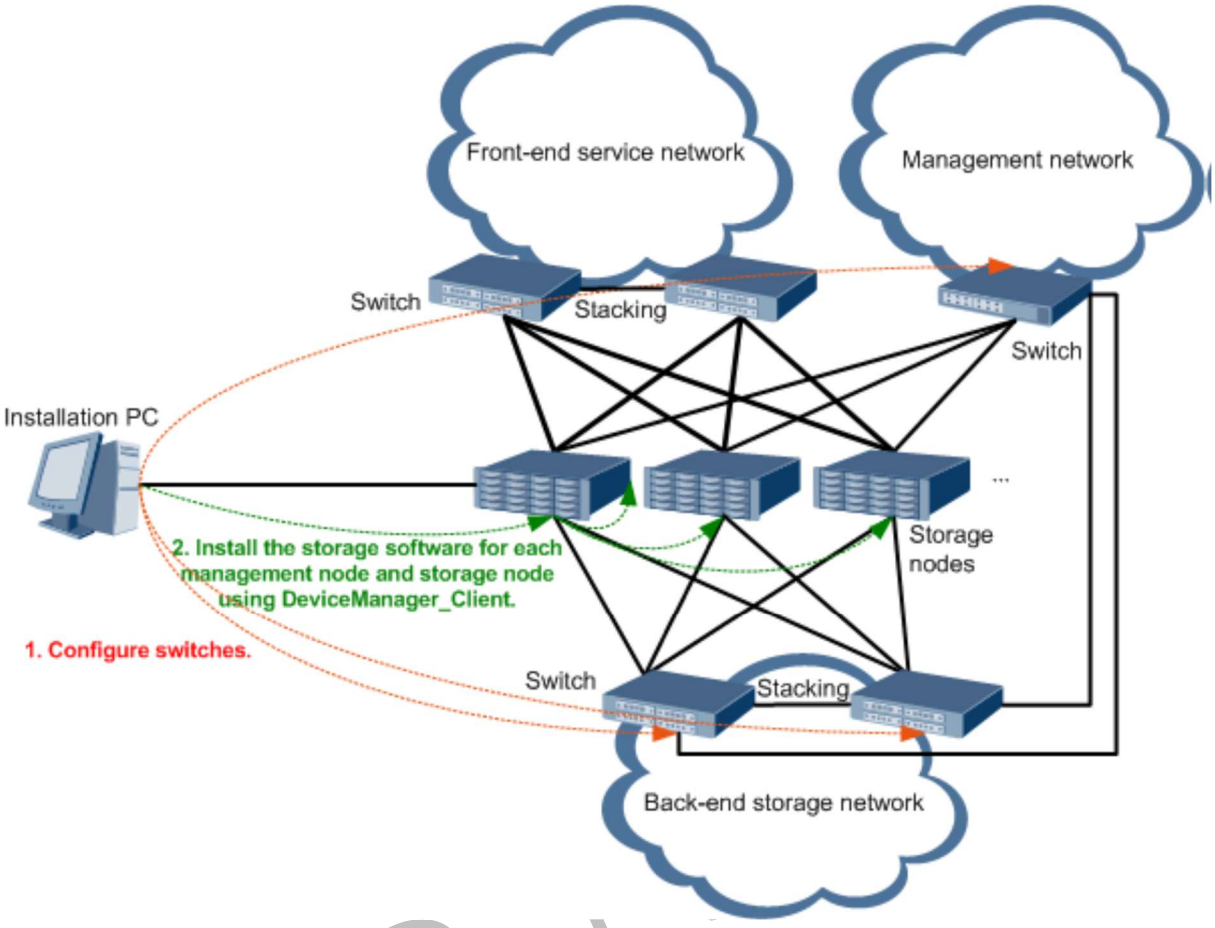

Figure 8 FusionStorage 8.0 manual deploy workflow

OpenStack SDS iSCSI solution

Both Ceph and Huawei's FustionStorage SDS solution support iSCSI storage interface. The iSCSI interface provides a versatile but slightly lower performance approach compared to native connections offered by Ceph's RBD and FusionStorage's VBS.

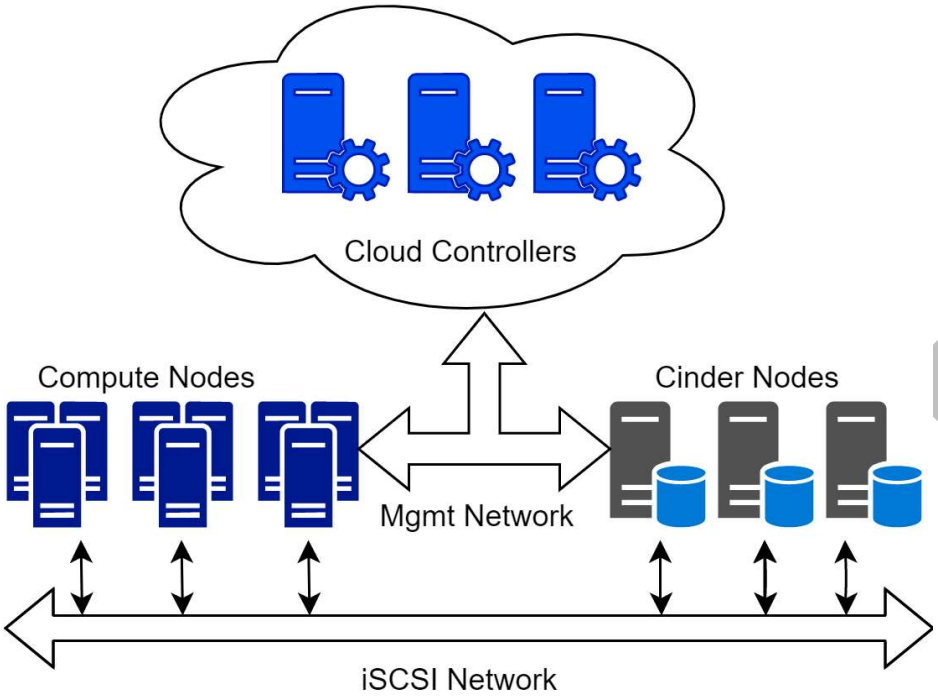

Figure 9 OpenStack components connecting to FusionStorage or Ceph using iSCSI.

OpenStack SDS client-based native block solution

Ceph block client deployment option

The block storage technology in Ceph is RADOS (Reliable Autonomic Distributed Object Store), which can scale to thousands of devices across thousands of nodes by using an algorithm to calculate where the data is stored and provide the scaling you need. RADOS Block Devices, or RBDs, are thin-provisioned images that can be shared via the rbd kernel module or an iSCSI gateway.

FusionStorage block client deployment option

FusionStorage support native block storage interface through a kernel driver called VBS, like Ceph's RBD. It can be installed on block client device or iSCSI gateway.

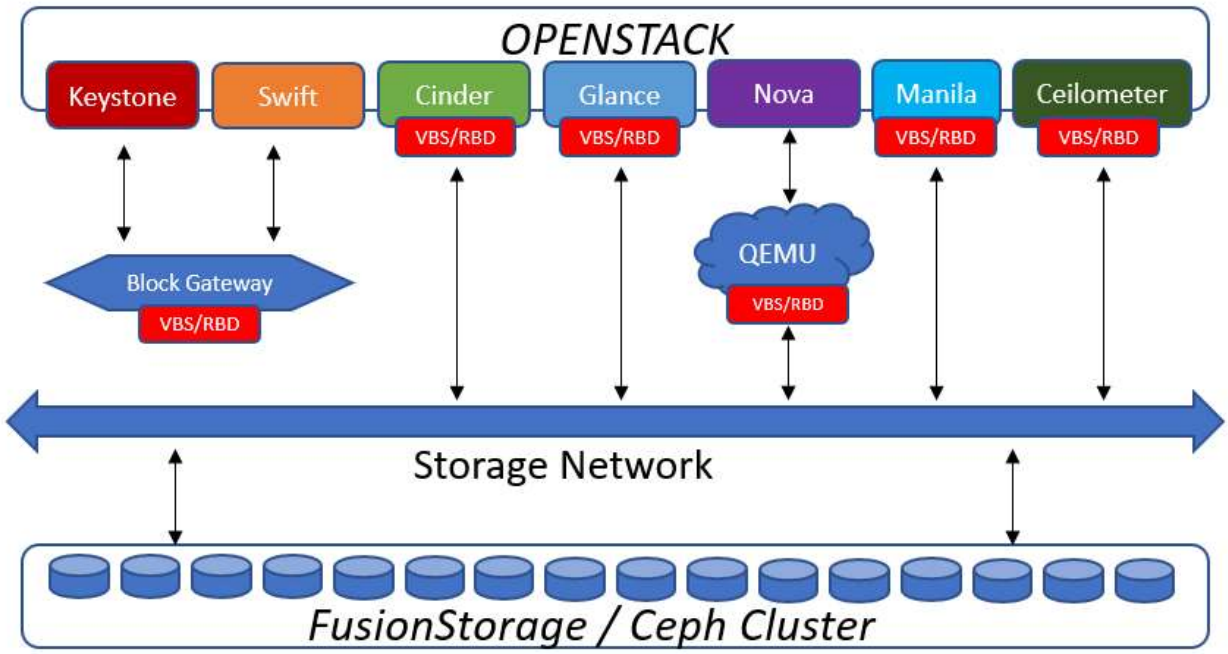

Figure 10 OpenStack SDS native block storage integration (red components indicate where native block driver is installed)

Best Practice

Following preparations are necessary for both FusionStorage and Ceph ansible deploy:

- A deployment host with ansible and other dependencies installed.
- In production environment, the best practice is to have separate control and data network, both networks should have been configured with static IP either manually or through DHCP server prior to the deploy.
- Inventory file is compiled and verified for the target platform.
- Ansible accessibility verified using the prepared inventory file. In case of Ceph, playbook can be validated using ceph-validate role prior to the actual deploy.
- Deploy source should avoid dependency on internet access in production environment to reduce security risk and complexity. For FusionStorage, it is the default mode; for Ceph, the origin should be set to 'local' and the corresponding local resource should be ready

4. OpenStack and SDS Business continuity (DR and backup solution)

There are several ways for OpenStack to achieve higher business continuity level. OpenStack defines both replication and snapshot capabilities. Depending on the SDS software used, these capabilities can be implemented as host-based or storage based.

Volume Snapshot and clone

Cinder defines the snapshot user interface for OpenStack. Both Ceph and FusionStorage provide drivers to support snapshot functionality. In Cinder's context, a volume snapshot is defined as a point-in-time image of a volume, and the snapshot is read only. In contrast, a clone is a volume that can be accessed by read and write. One can create multiple clones based on a snapshot.

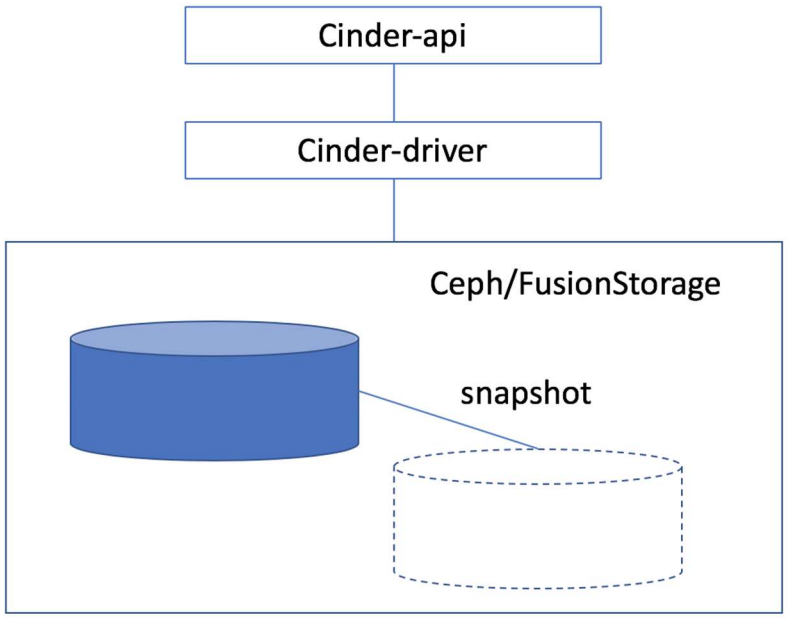

Figure 11 Taking storage-based snapshots using Cinder APIs

Compared to a host-based snapshot scheme, a storage-based snapshot scheme has many advantages. Because the snapshot is taken on the storage subsystem, the complicated logic -- such as maintaining the differences, how to be space efficient – is offloaded to the storage implementation. In OpenStack, through the cinder-API, users can issue a series of commands to create, delete, and revert to snapshots. One can take snapshots periodically so that any logical errors can be reverted. The snapshot strategy can also be combined with backup strategy to backup data locally and remotely.

With HyperSnap and HyperClone features in FusionStorage 8.0, users can flexibly create a tree of replicas that reflect different point in time. This can be achieved via OpenStack Cinder APIs or via FusionStorage APIs directly.

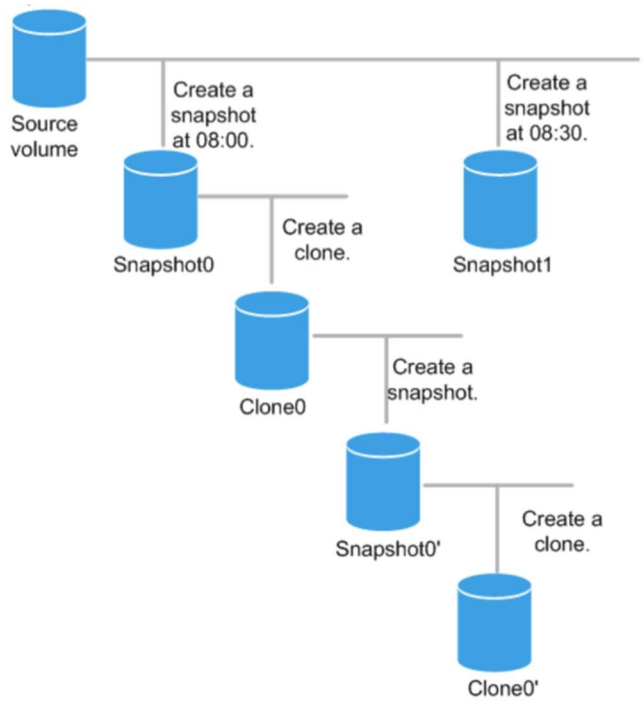

Figure 12 An example of snapshot and clone tree rooted from one source volume

Volume Replication

OpenStack has a volume replication feature that provides business continuity of data. The spec can be found at the OpenStack website [10].

In summary, Cinder defines the state transitions with states – error, copying, active, active-stopped, and inactive. The host-based replication solution needs to be integrated with Nova, while the storage-based solution must be provided by the storage vendors. In this white paper, we will describe the storage-based solution with two examples: Ceph and FusionStorage.

Ceph has an RBD driver that supports volume replication. It uses the RBD mirror feature to transfer data from the primary OpenStack site to the secondary site asynchronously.

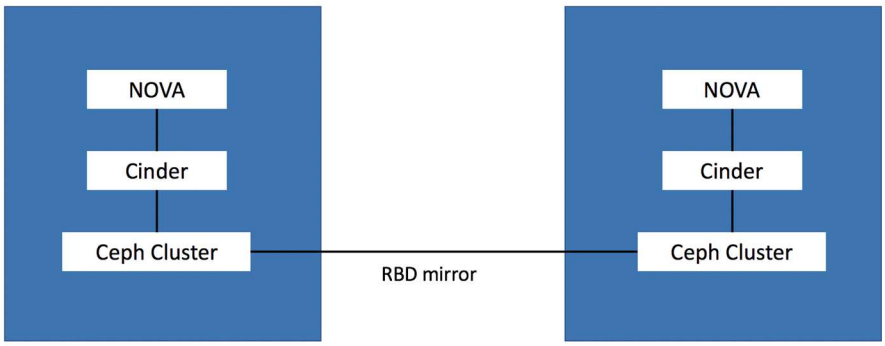

Figure 13 Using Ceph's RBD mirror for volume replication

FusionStorage provides stronger replication features than Ceph, by providing not only the asynchronous replication, but also synchronous replication and active-active HyperMetro features.

However, since FusionStorage provides a richer set of features than Ceph and Cinder, users must rely on FusionStorage APIs and CLIs to control the replication behavior.

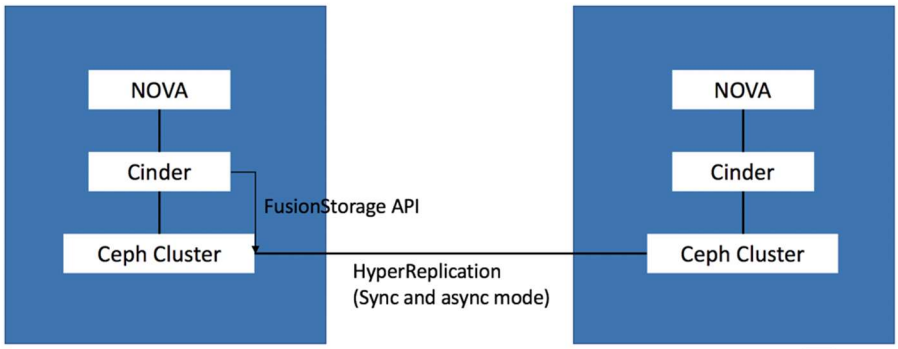

Figure 14 Using FusionStorage's API and HyperReplication for volume replication

FusionStorage can be used to create a cross-AZ synchronous volume replication to provide a max availability solution. A 3-site multiple AZ DR plan with a mixed synchronous and asynchronous replication can be built using several FusionStorage clusters. A multi-AZ solution has more requirements and conditions that must be met. More details of the HyperReplication feature can be found in the feature guide [11].

Backup

Cinder provides a backup interface to allow users to back up a volume. FusionStorage is bundled with eBackup to provide the backup service for OpenStack. eBackup driver must be installed on each OpenStack Cinder node, and then Cinder backup interface can be used to control backup needs.

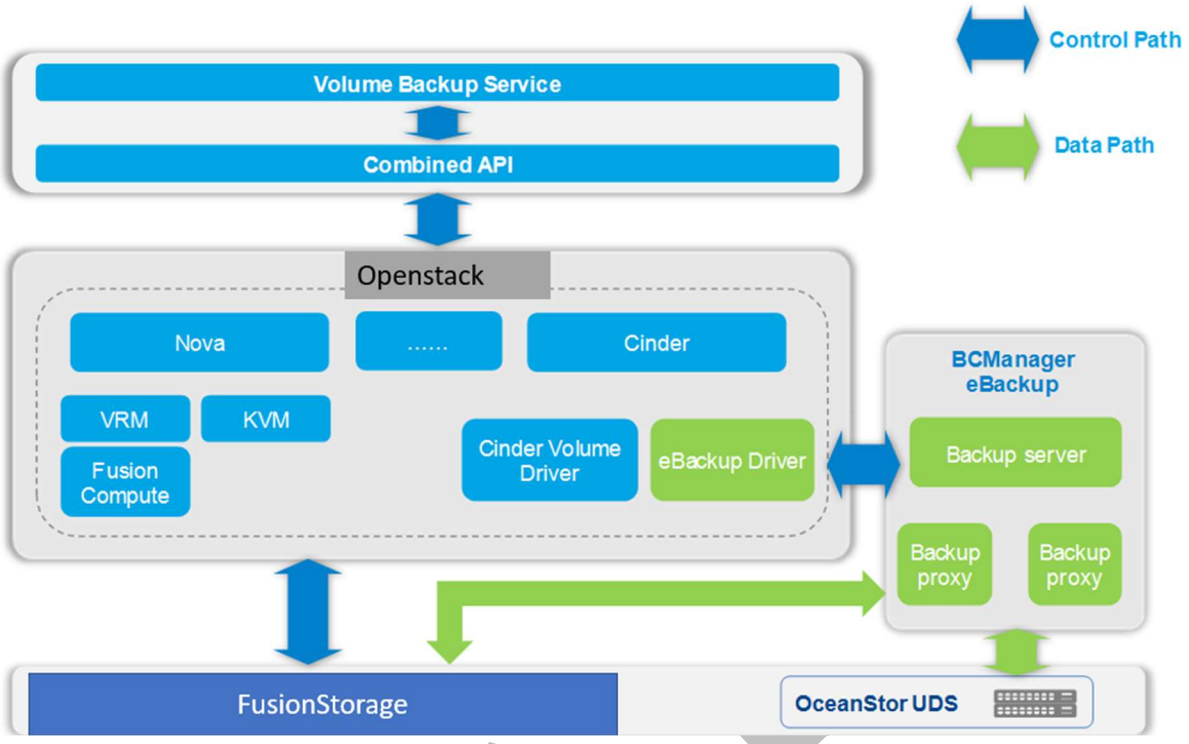

Figure 15 Using eBackup to provide backup feature in the OpenStack environment

Besides Cinder, eBackup also provides its own CLI and RESTful APIs to be included into customer's backup plan. It provides efficient backup features for incremental backup to storage and incremental restore from the storage, optionally with compressed data. For detailed features of eBackup, please refer to eBackup Product Document [9].

Best Practice

The best practice for OpenStack DR is to use the DR features provided by the storage layer.

- Use Cinder provide APIs and different drivers
- Use vendor specific APIs if needed
- Use snapshot to protect any logical data corruption
- Use remote replication to protect any physical data corruption
- Use synchronous replication and multi-site solution if required
- Use incremental backup and restore to shorten the backup window
- Use compression to reduce the backup storage usage

5. OpenStack SDS operation and monitoring

In the OpenStack environment, many monitoring tools can be used to monitor the system states, such as workload, usage information. Some vendors provide a customized operation and monitoring tools. For example, Huawei provides its own enhancement for operation and monitoring tools in FusionSphere OpenStack [2], which is based on the open-source version of OpenStack. Recently, open-source monitoring tools like Prometheus [3] become attractive because of their flexibility and extensibility. Prometheus can monitor a broad range of services from application layer to storage layer. In the following sections, we will use two SDS systems, Ceph and FusionStorage, as examples to show how Prometheus can be used as monitoring tool for SDS. Prometheus is often used together with Grafana, which is a visualization tool that can query Prometheus and present the result in customized dashboards.

OpenStack storage Prometheus Integration

Prometheus relies on plug-ins called exporters to report system status for each software component. Exporters can be categorized to official exporters and third-party exporters. A list of exporters can be found at [4].

Digital Ocean provided a third-party exporter called "Ceph exporter", which exports telemetry to Prometheus. The repository for Ceph exporter can be found at [5]. The following diagram shows the basic flow of how Ceph telemetry is propagated to Prometheus.

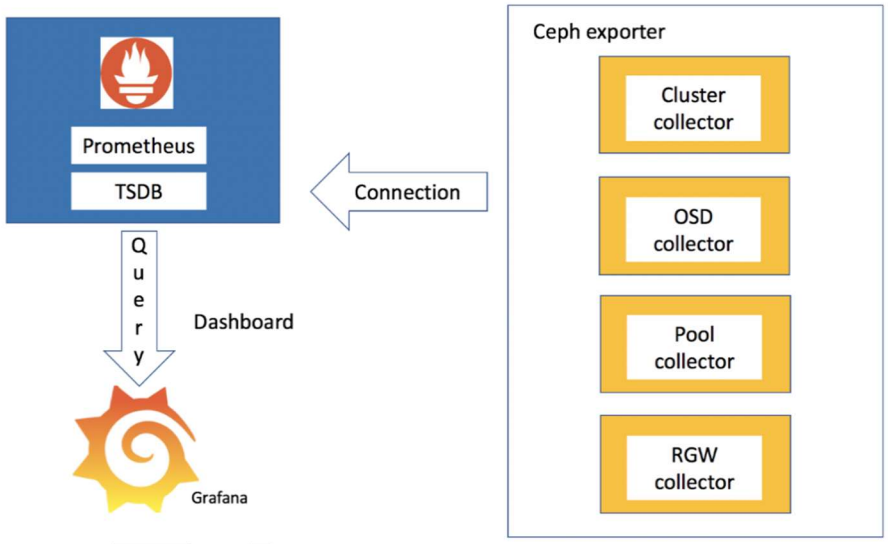

Figure 16 Using Ceph exporter to export telemetry to Prometheus and view it in Grafana

Customers can use Prometheus and custom-made FusionStorage exporters to export information and warnings about the storage. The level of monitoring and information exposing is based on customer needs, and it is a common need for customers to know OSD/disk information, Pool information, IO information, and cluster information.

Logging service integration

Currently, there are several open source based logging service solution available, such as logstash [13] and fluentd [14]. For example, Fluentd has many plug-ins available listed on their website [15]. Ceph is one of the plug-in in the list.

The benefits of consolidating logs to form a centralized logging service are: (a) all events are centralized into one place to be displayed. (b) all logs have an understandable format, (c) the quality of the logs collected is high due to transformation, and (d) one can use big data and machine learning tools to process data collected for insight.

From SDS perspective, providing integration interface to these services will allow them to be monitored by a bigger ecosystem, which does not limit to OpenStack. For example, Prometheus and Fluentd are also recommended by Cloud Native Computing Foundation (CNCF) and are part of CNCF roadmap [16].

Best Practice

Monitoring can be a complicated issue as customers may already have a cloud-wide monitoring infrastructure. New software and systems will need to fit into the existing framework. Therefore, SDS systems should provide different plug-ins to support different monitoring frameworks.

- SDS provided monitoring tool is a good starting point
- Prometheus plus Grafana is one popular choice for open source solutions
- Different adaptations and customizations may be needed for exporters
- If needed, integrate SDS with open sourced logging services to centralize the logs

Of course, one SDS system may not have plug-ins ready for all available monitoring software. It is a common practice to prepare for popular software (e.g., Prometheus, Fluentd, etc.) and build support for others on demand.

6. Data migration issues

With the growth of the cluster size, and the building of new IaaS infrastructure, there is often a need to migrate data stored in the previous infrastructure to the new infrastructure, which contains the lastest hardware and software. In this chapter, we discuss the best practices to migrate between different SDS platforms.

Host-based migration

A host-based migration is to trigger host commands to copy data from old volumes to new volumes. Both volumes will be mounted to the host executing the migration commands. Then data will be first read into the host memory, then write into the new volume. The advantage of a host-based migration is simplicity, because no special equipment or features is needed. However, it is very difficult to implement a live migration in this case, and the time between the migration start to migration ending could be long. Therefore, for critical systems the host-based migration is not an option.

From any block storage to FusionStorage

To satisfy the need to migrate data from existing storage to FusionStorage, FusionStorage provides a feature called SmartMigration to access data remotely from other storage systems. Since the external LUN feature uses common protocols like iSCSI, any block storage systems can be supported. The SmartMigration feature for FusionStorage will be released soon.

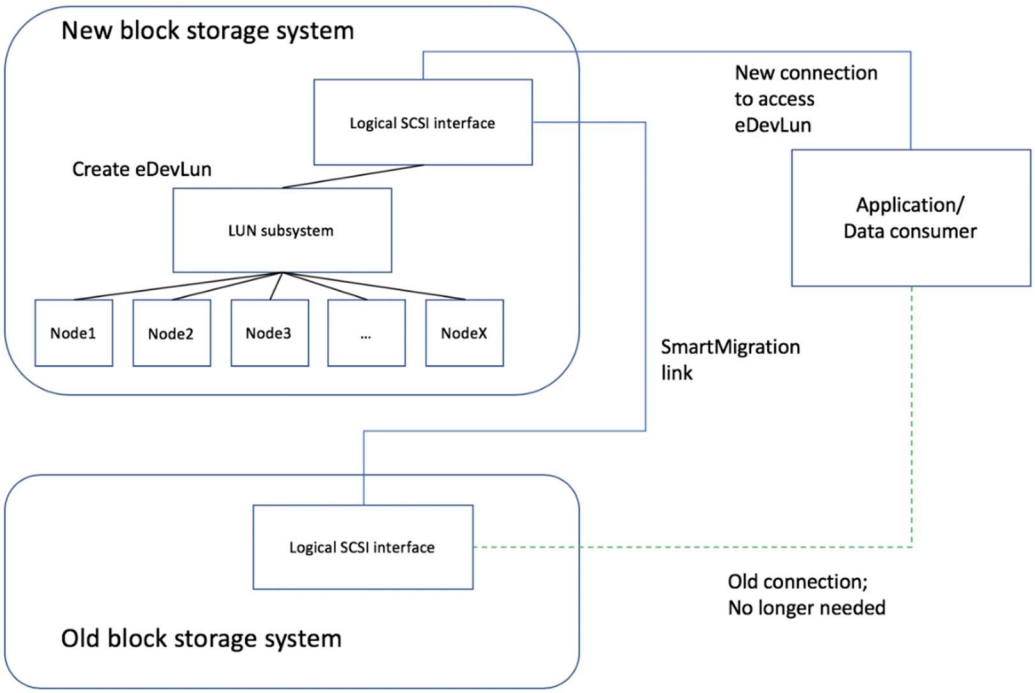

Figure 17 Using the SmartMigration feature to provide a fast data migration

This simplified diagram shows how SmartMigration can help applications switch from one system to another. After creating the SmartMigration link, users can access eDevLun in the new system and SmartMigration will make sure correct data is obtained via the link.

The following diagram shows a screen shot (for Dorado) to create a link to the remote storage system, referred to as heterogeneous storage system, and then create an external device LUN, referred to as eDevLUN.

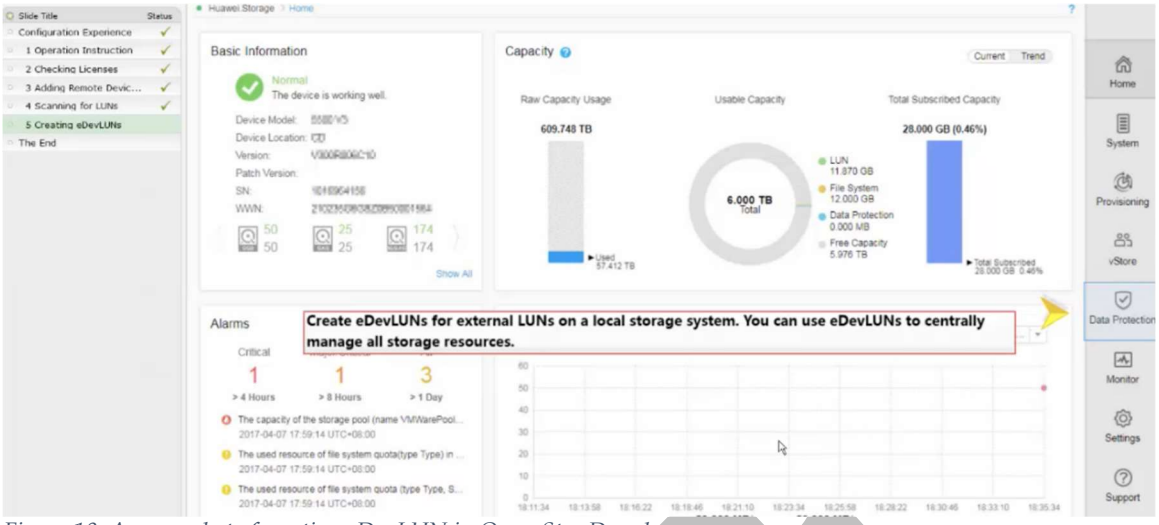

Figure 18 A screenshot of creating eDevLUN in OceanStor Dorado

With this feature, any value on the remote LUN can be instantly accessible from the new FusionStorage system. To facilitate the smooth migration, automation tools based on CLI and Ansible can be provided to customers upon request.

From FusionStorage to FusionStorage

Migrating data from FusionStorage to other FusionStorage can use its HyperReplication feature. HyperReplication can be firstly used to provide a background copy, which will move the majority of the data. Then HyperReplication synchronous mode can be used to completely migrate the data in a consistent way.

The following diagram depicts the simplified data migration process.

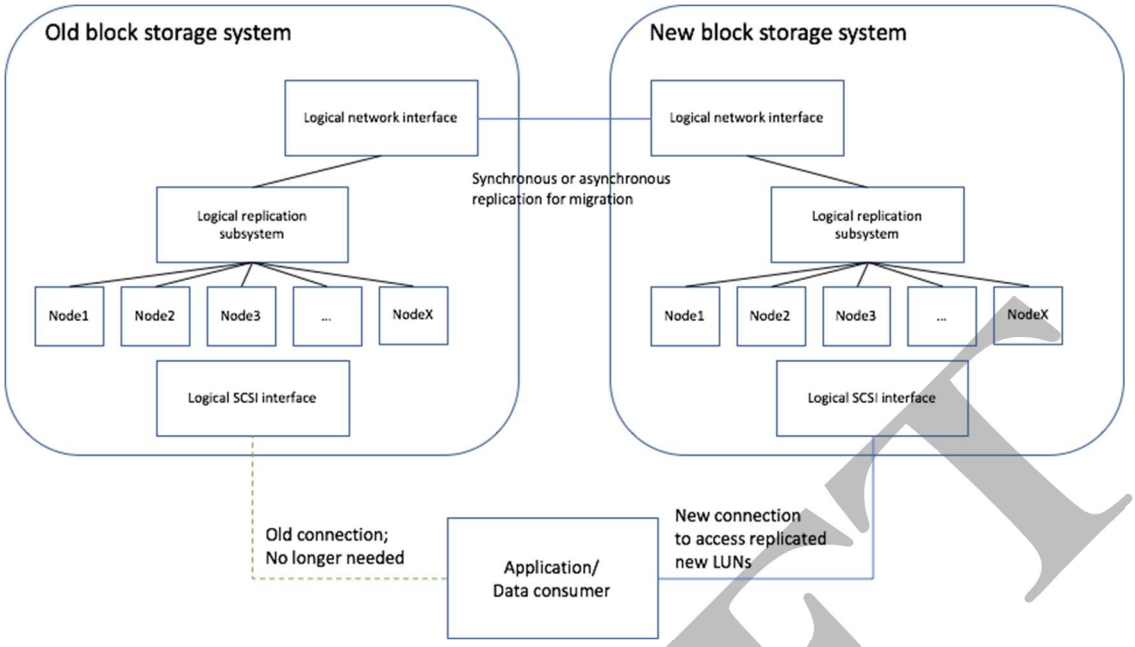

Figure 19 Using FusionStorage replication to provide data migration

The storage replication subsystem can copy data from one FusionStorage system to another. Data can be copied to the new system while the old system still provides service. Then we can cut off the application to use the new FusionStorage system.

To help customers migrate, tools and ansible playbooks can be provided to customers for automation.

Best Practice

For critical systems and systems that need to limit the down time, live migration or migration with minimally down time would be an ideal choice:

- Applications can be migrated to the new storage system either instantly or with interruption within seconds.
- Data copy rate can be controlled and changed on demand so that the copy can be background or in high priority to meet the IO needs.
- Migration should be automated to minimize user impact

FusionStorage provides a set of tools that can help customers meet their migration needs.

7. OpenStack SDS User Management and Security Integration

User Management

There are 2 types of user managements and authentications in a typical SDS system: Data path authentication and control path authentication.

Data Path User Management and Authentication

Data path user management and authentication is less important in OpenStack environment. And often, it is vendor and protocol specific. Supporting AD/LDAP is also specific to each individual protocol (NFS, CIFS, Swift/S3, etc.) and often a vendor's decision.

Since SDS primarily serves as storage backend for OpenStack Nova, Glance etc., it is considered in a fairly contained environment. Security concerns are masked by physical and logical separations of hardware, network and compute. Data authentication is commonly not a concern for service providers (OpenStack operators).

However, if VM hosts of tenants (Customers of OpenStack) want to access SDS storage directly (for example, via Swift, S3, NAS, ISCSI, FC etc.), data authentication is needed. Based on OpenStack overall architecture and Cinder native support, it is unlikely for service providers to expose other protocols other than object and file to VM hosts. Multi-tenancy support for user management and authentication is also a must.

Object Access Authentication

SDS shall support S3 and Swift native authentication, which are AWS {access-key}:{hash-of-header-and-secret} based authentication and X-Storage-Token based authentication.

For better integration with OpenStack, SDS shall offer more integration with Keystone, which means SDS shall be able to recognize and authenticate token issued by Keystone.

Another way to authenticate S3 or Swift is to use LDAP authentication. In this scheme, an SDS provides utilities to create an SDS recognizable token based on user id and password. It also requires SDS to support LDAP and having a LDAP service account.

Ceph object gateway supports LDAP authentication with easy configuration setup instructions. For more information, please refer to the Ceph LDAP authentication documentation [12]. It requires administrators to manually create LDAP users, and Ceph provided a utility **radosgw-token** to export access token to S3 clients.

NAS Access Authentication

SDS shall support all user management and authentications native to protocols, including, Active Directory for Samba and NFSv4, Linux authentication for NFS.

Block Access Authentication

Block access is an exception that SDS providers can have a freedom to choose whatever user authentication and user management of their choices. The reason behind that is due to physical and logical separation of network, hardware and compute mentioned previously.

Different vendors may choose different authentication mechanisms. As an example, Ceph uses its own "shared secret"- based authentication and user management for RBD and CHAP for iSCSI. In contrast, FusionStorage adopted client IP for VBS authentication and CHAP for iSCSI.

Control Path User Management and Authentication
For control path management, single sign-on is a preferred feature for SDS system management and control.

Ceph Dashboard or FusionStorage Device Manager offers system management GUI via single sign-on. There are several possible choices: SAML2, LDAP, Oauth2.0, and OpenID Connect. Among the choices, SAML2 is most friendly to web applications so it is often chosen. But due to LDAP's popularity, it is still a common choice for OpenStack users. Oauth2.0 and OpenID Connect are more focused on mobile applications, therefore, they are least likely to be chosen by OpenStack users.

Ceph supports SAML2 and FusionStorage supports LDAP.

Identity and Security Management Tool Integration

LDAP
A standard LDAP directory, such as Red Hat Directory Server, is a general-purpose directory: it can be customized to fit a broad range of use cases.

- Schema: a flexible schema that can be customized for a vast array of entries, such as users, machines, network entities, physical equipment, or buildings.
- Typically used as: a back-end directory to store data for other applications, such as business applications that provide services on the Internet.

LDAP integration for SDS providers shall mainly be focused on single sign-on for protocol access management if possible (Such as S3/Swift, NFS, etc.) and SDS system management. Since SDS systems seldom host services other than storage services, there is no such requirement for host OS to support LDAP for other purposes (for example, to login SDS host).

Keystone
Keystone is the identity service used by OpenStack for authentication (authN) and high-level authorization (authZ). It currently supports token-based authN and user-service authorization. It has recently been rearchitected to allow for expansion to support proxying external services and AuthN/AuthZ mechanisms such as oAuth, SAML and openID in future versions.

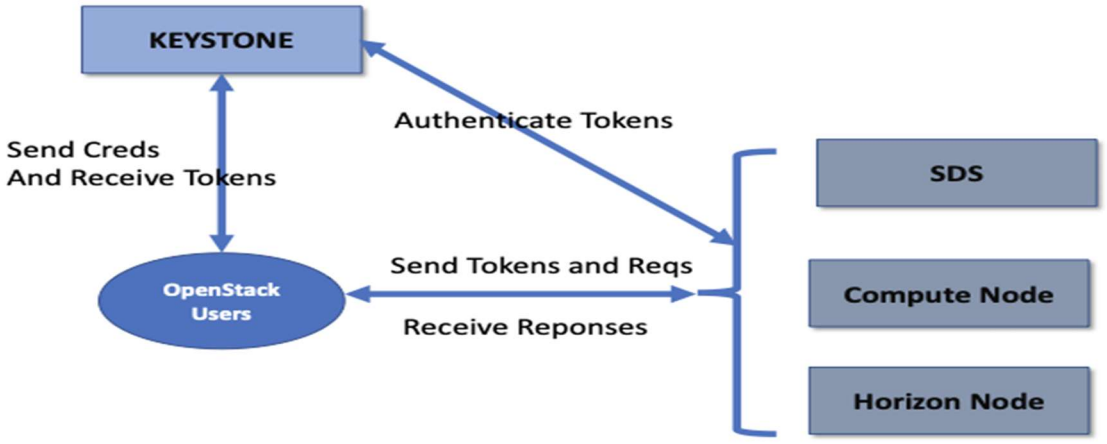

Figure 20 Using Keystone to provide authentication for OpenStack components

Although Keystone is not very popular in OpenStack providers choices for IDP, there is still some merit to support Keystone by SDS providers for better integration. SDS providers shall recognize tokens generated by Keystone and use user credentials derived from tokens to authenticate with Keystone using Keystone APIs.

IDM

Red Hat Identity Management (IDM) provides a centralized and unified way to manage identity stores, authentication, policies, and authorization policies in a Linux-based domain. IDM significantly reduces the administrative overhead of managing different services individually and using different tools on different machines.

IDM is one of the few centralized identity, policy, and authorization software solutions that support:

- Advanced features of Linux operating system environments
- Unifying large groups of Linux machines
- Native integration with Active Directory

Supporting IDM on SDS nodes implies that SDS OS must support IDM clients. However, since SDS barely hosts other services other than storage services and SDS OS is only installed for the purpose of providing infrastructure for storage service, one can argue that there is no such need for IDM support on SDS nodes. But OpenStack provider can choose to enable the support for the purpose of security purpose(Auditing, etc).

However, even though it is optional, it is desirable for SDS to add its management service as IDM service (KRB5 service) or support user authentication through LDAP (provided by IDM), because this integration simplified identity management for users. A typical integration with IDM can be implemented by an OpenStack Provider as follows:

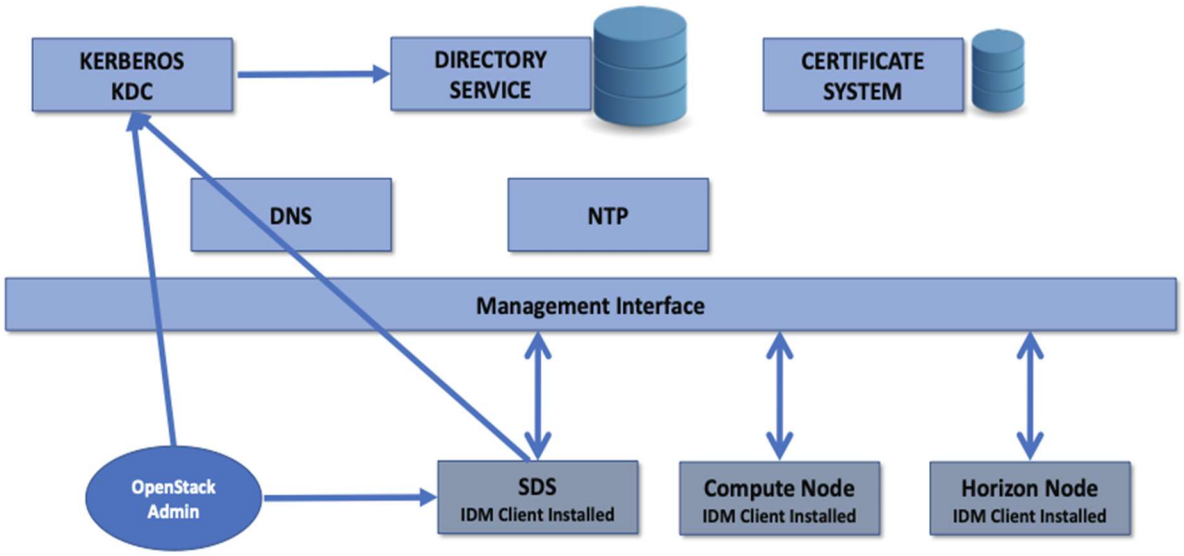

Figure 21 Integrating OpenStack services with IDM

An SDS provider shall provide the capability to interact with KDC/ID Provider and Directory Services (LDAP) to create SDS service principles and authenticate users for system management, user management, and operations. It may need to support SAML or LDAP protocol.

Best Practice

For data path authentication with integration with OpenStack, SDS object implementation shall support integration with Keystone and LDAP authentication.

For management path authentication, single sign-on (SSO) is preferred. OpenStack users can choose either SAML2 or LDAP. To support future enhancement and integration in OpenStack environment, SDS provider shall consider supporting both and integrate with IDPs (like OpenStack Keystone or Redhat IDM) to support SSO.

For better integration with OpenStack, SDS provider shall choose to integrate with IDPs (Keystone or Redhat IDM is preferred) to support single sign-on for SDS system management, data access, user management and operations. This integration helps the administrators to consolidate the user management in the OpenStack solution. Of course, the real solution must be based on OpenStack provider's own requirements. As mentioned before, SDS user management, operations are usually exposed through web management GUI or Restful APIs. Providing an SSO service will be nice to have for the OpenStack provider but also potentially putting on more burdens on OpenStack provider's infrastructure team to support such frameworks (Kerberos, OAUTH2.0, SAML, etc.).

8. Summary

In this white paper, we discussed a broad range of topics about integrating software defined storage systems with OpenStack. As examples, we selected two popular SDS implementations, Ceph and FusionStorage. We discussed the best practices to select hardware and software, to install SDS software on common servers, to enable DR features, to monitor the operations, to migrate data from other storage solutions, and integrate with user management systems. This document can be served as a guide to anyone who would like to integrate SDS systems into an OpenStack environment. Many references and links are provided so that more detailed information can be obtained.

Many SDS systems are designed to be integrated with a variety of user environment, including OpenStack but not limited to OpenStack. It is known that not all features can be called in an "OpenStack native" way. Since the OpenStack community is making progress on many APIs, and SDS vendors will keep investing on new features. We hope the integration will get more seamless in the future. In this white paper, we are not limiting ourselves to implement everything as OpenStack native, and we adopt vendor APIs and CLIs as needed. We believe this stance reflects the reality and is the best practice.

Another issue that we discussed is how to balance the vendor provided tools and other open source tools. SDS vendors provided a set of tools that reflect vendor's view of how the system should be managed. At the same time, an OpenStack infrastructure builder might want to integrate all the vendors' solution and tools into one single panel. With more and more vendors embracing open-source tools, many SDS system provide scripts or software for integration, such as Ansible playbooks, Prometheus exporters, etc. We expect the trend to continue and more and more vendors will follow. However, even with the provided software, the OpenStack infrastructure builder will need to customize the software to adapt into their own existing OpenStack environment. Sometimes the job can be significant.

Overall, we believe OpenStack and SDS are a natural fit for each other and these two communities will continue cooperation and generate more value for infrastructure builders and users.

References

[1] OpenStack software overview, OpenStack project, https://OpenStack.org/software

[2] FusionSphere OpenStack, https://e.huawei.com/en/cloud-computing/fusionsphere-OpenStack

[3] Prometheus, https://prometheus.io/

[4] Exporters and integration, Prometheus, https://prometheus.io/docs/instrumenting/exporters/

[5] Ceph exporter repository, https://github.com/digitalocean/ceph_exporter

[6] FusionStorage OpenStack Driver Configuration Guide, https://github.com/Huawei/FusionStorage_OpenStack_Driver

[7] Ceph ansible deployment guide: https://docs.ceph.com/ceph-ansible/master/

[8] FusionStorage 8.0 Documentation: https://support.huawei.com/enterprise/en/cloud-storage/FusionStorage-pid-21905727

[9] OceanStor BCManager eBackup Product Document, https://support.huawei.com/hedex/hdx.do?docid=EDOC1000109632

[10] OpenStack volume replication specification, https://specs.OpenStack.org/OpenStack/cinder-specs/specs/juno/volume-replication.html

[11] FusionStorage 8.0.0 Block Storage HyperReplication Feature Guide, https://support.huawei.com/enterprise/en/doc/EDOC1100081417/31411ab/data-replication

[12] Ceph LDAP Authentication, https://support.huawei.com/enterprise/en/doc/EDOC1100081417/31411ab/data-replication

[13] Logstash: centralize, transform and stash your data, https://www.elastic.co/products/logstash

[14] Fluentd.org, https://www.fluentd.org/

[15] Fluentd plug-in list, https://www.fluentd.org/plugins/all

[16] CNCF announces fluentd graduation, https://www.cncf.io/announcement/2019/04/11/cncf-announces-fluentd-graduation/