The Comparison of Ceph and Commercial Server SAN

Yuting Wu
wuyuting@awcloud.com
AWcloud

Keep an open mind
Lead the cloud world

海云捷迅 AWCLOUD

Agenda

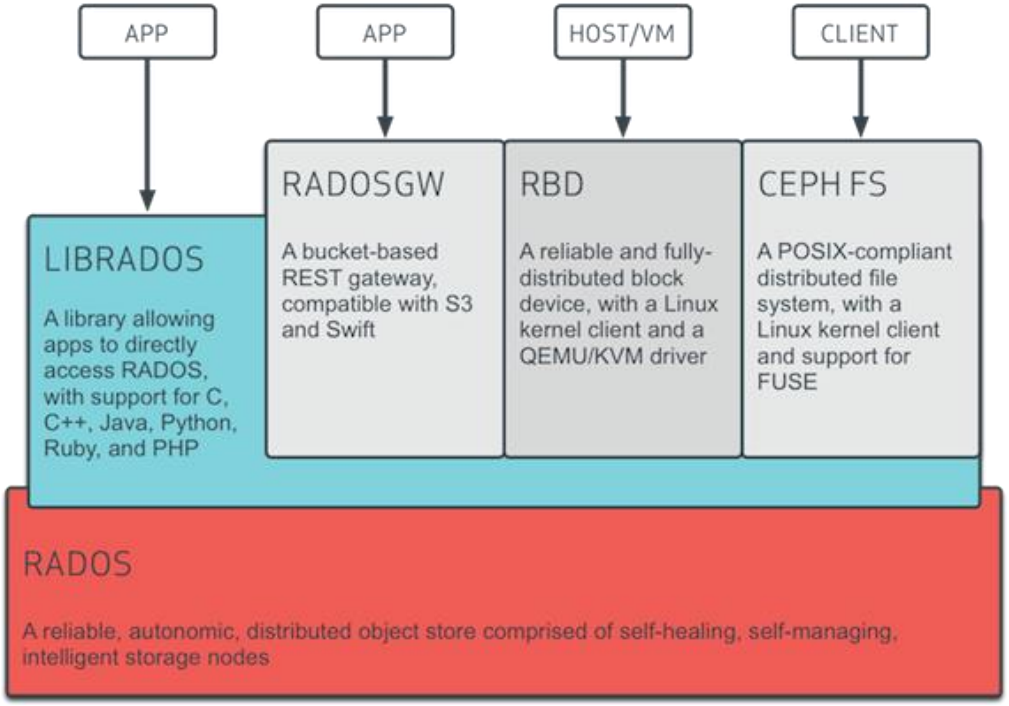
- Introduction to AWcloud
- Introduction to Ceph Storage
- Introduction to ScaleIO and SolidFire
- Comparison of Ceph and Server SAN
- Performance test of Ceph and ScaleIO
- Summary

海云捷迅 AWCLOUD

Who are we

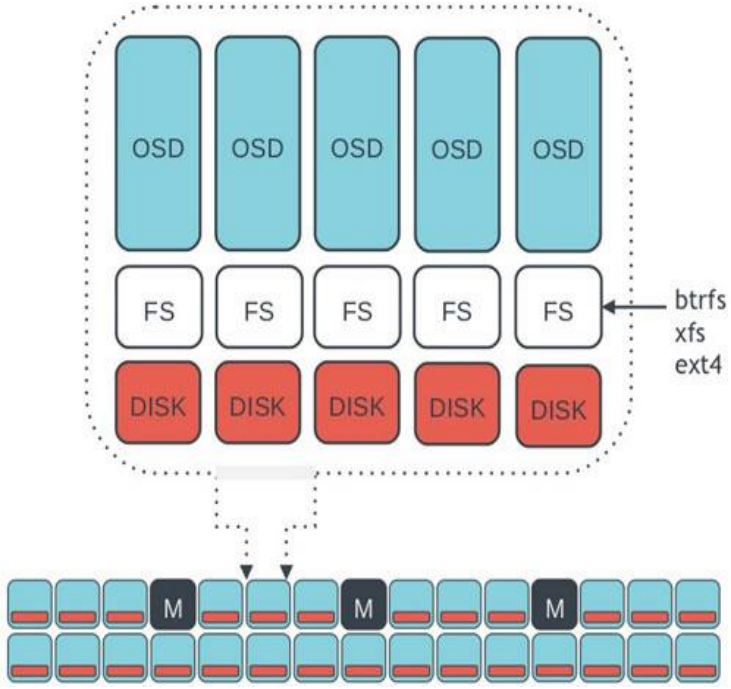
- Pure OpenStack player in China
- China’s leading enterprise cloud service provider
- Broad deployment of production clouds in China
- Highly diverse set of workloads and use cases

Ceph Architectural



Ceph uniquely delivers **object, block, and file storage** in one unified system.

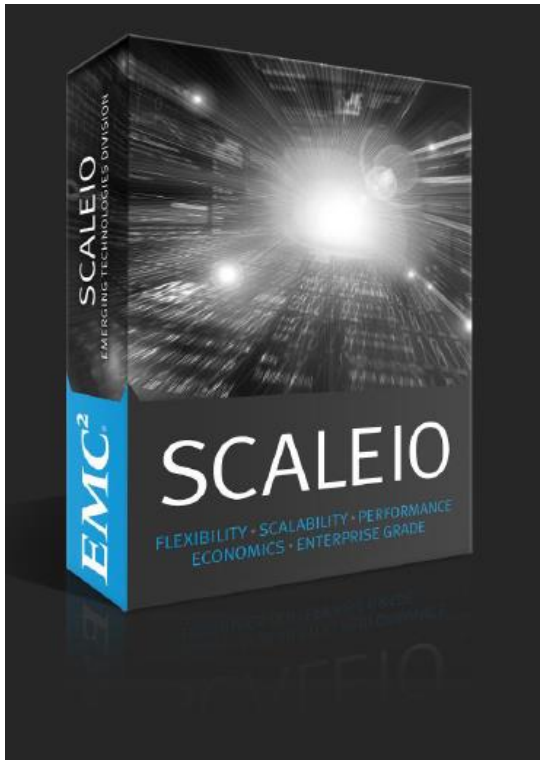
Ceph Components



OSDs: A *OSD* stores data, handles data replication, recovery, backfilling, rebalancing, and provides some monitoring information to Ceph Monitors by checking other Ceph OSD Daemons for a heartbeat.

Monitors: A *Ceph Monitor* maintains maps of the cluster state, including the monitor map, the OSD map, the Placement Group (PG) map, and the CRUSH map.

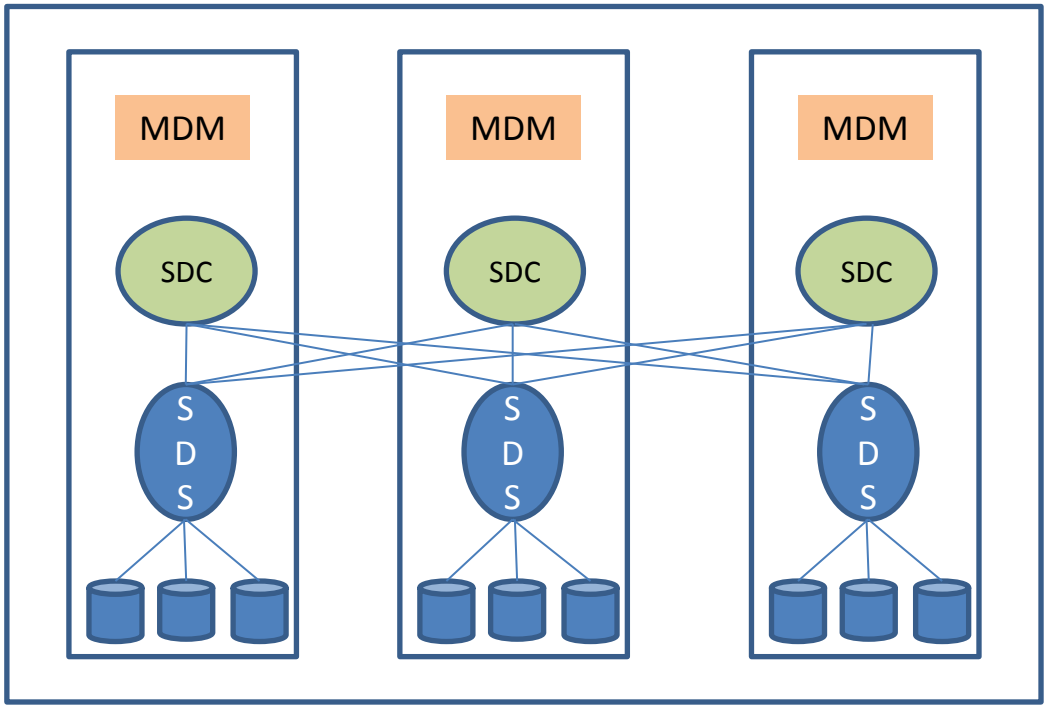
ScaleIO Introduction



ScaleIO :

- A software-only solution
- Installed on common commodity servers
- Turn existing DAS storage into shared block storage

ScaleIO Architecture



MDM：
Configures and monitors the ScaleIO system
Cluster Mode or Single Mode
SDS：
Manages the capacity of a single server
Data access
SDC：
Exposes ScaleIO volumes as block device

Solidfire Introduction

Node：
A collection of Solid State Drives

Cluster：

- Make up of a collection of nodes
- At least four nodes in a cluster（five or more nodes are recommended）
- Nodes connect to each other with 10GbE
- 1GbE for management
- 10GbE for storage iSCSI
- Scaled-out by adding nodes

Deployment

Deployment-Ceph

CeTune
Intel

Puppet

Ceph-deplpy

ceph

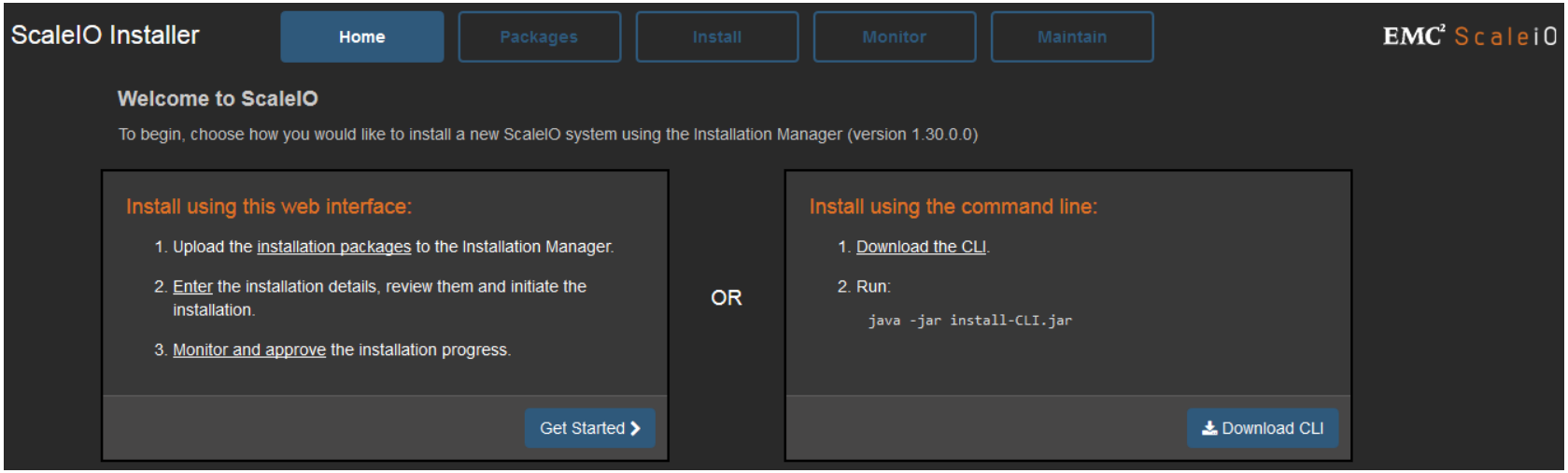
Chef

ANSIBLE

Ceph-disk

Deployment-ScaleIO

Installation Manager Window



Installation topology file

Domain	Username	Password	Operating	Is MDM/TB	MDM Mgmt	MDM IPs	Is SDS	SDS Name	SDS All I	SDS-SDS (	SDS-SDC (	Protecti	Fault Set	SDS Devi	SDS Pool	Optimize	Is SDC
domain1	root	password	linux	Primary	10.20.0.7	10.20.0.7	Yes	SDS1	192.168.1.8			domain1		/dev/sdb	pool1,poc	Yes	Yes
domain1	root	password	linux	Secondary	10.20.0.8	10.20.0.8	Yes	SDS2	192.168.1.5			domain1		/dev/sdb	pool1,poc	Yes	Yes
domain1	root	password	linux	TB		10.20.0.9	Yes	SDS3	192.168.1.1			domain1		/dev/sdb	pool1,poc	Yes	Yes
domain1	root	password	linux			10.20.0.6	Yes	SDS5	192.168.1.7			domain1		/dev/sdb	pool1,poc	Yes	Yes

Deployment-SolidFire

SolidFire storage nodes are delivered as an appliance with SolidFire Element OS installed and ready to be configured. After configured, each node can be added to a SolidFire Cluster。

Operations and Management

Operations and Management

	Ceph	ScaleIO	Solidfire
CLI	√	√	√
Web UI		√	√
REST Gateway		√	√
Other	VSM		TUI

Features and Volume Methods

Features

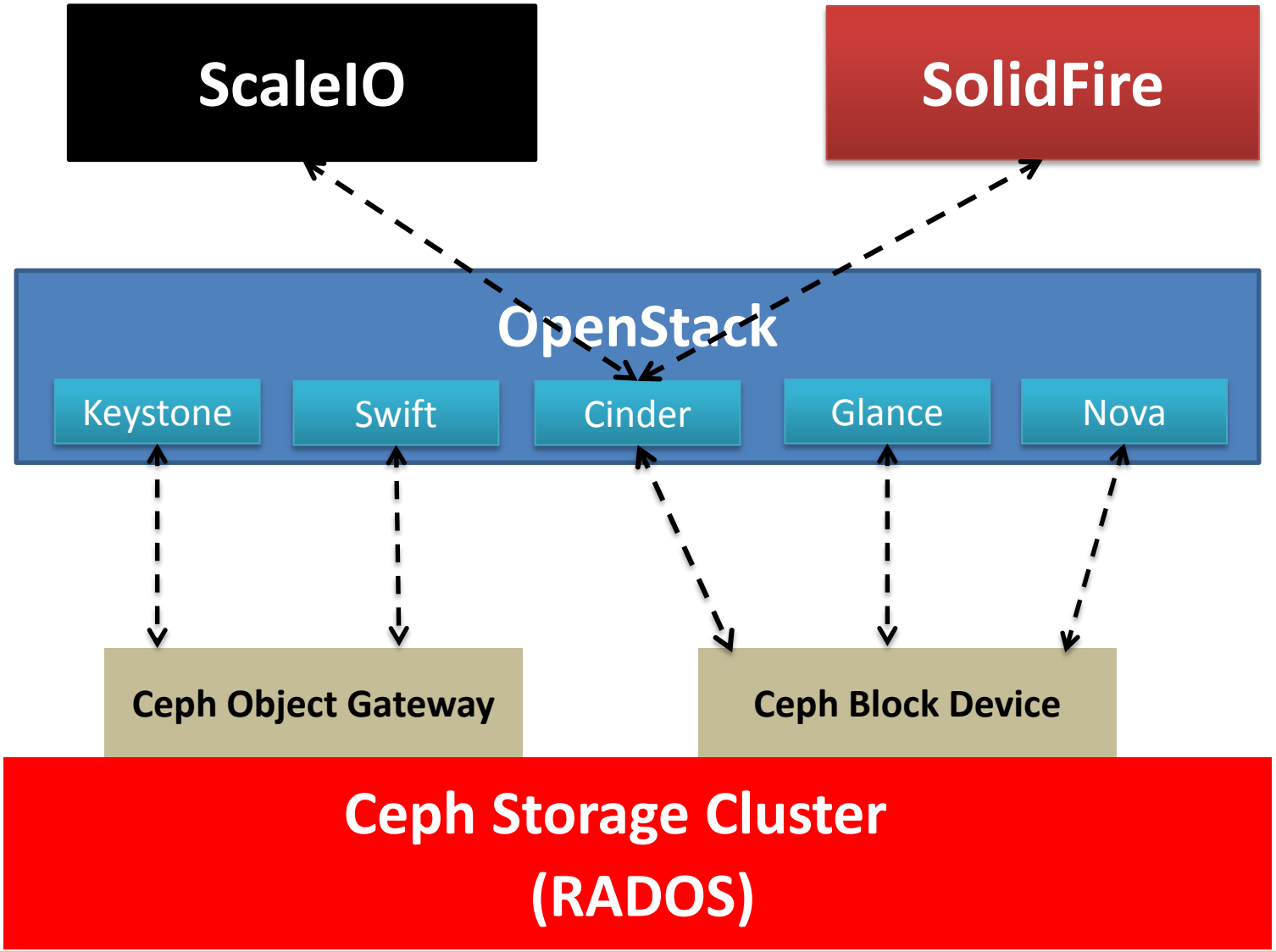
	Ceph	ScaleIO	SolidFire
Shared Filesystem	√		
Object Storage	√		
Block Storage	√	√	√
Deduplication			√
Compression			√
Thin provision	√	√	√
Minimum IOPS			√
Maximum IOPS		√	√
IOPS burst control			√
Bandwidth control		√	
Replication			√
OpenStack	√	√	√
Hyper-V		√	√
VMware		√	√
XenServer	√	√	√

Volume Methods

	Ceph	ScaleIO	SolidFire
Creating Volume	√	√	√
Resizing Volume	√	√	√
Modifying Volume(Qos)		√	√
Deleting Volume	√	√	√
Viewing Deleted Volumes			√
Restoring a Deleted Volume			√
Cloning a Volume	√	√	√
Viewing Running Tasks			√
Volume Access Group			√
Creating a Volume Snapshot	√	√	√
Rolling Back to a Snapshot	√	√	√
Backing up a Snapshot	√	√	√
Cloing a Volume from a Snapshot	√	√	√
Backing up a Volume to an Object Store			√
Restoring a Volume from an Object Store			√

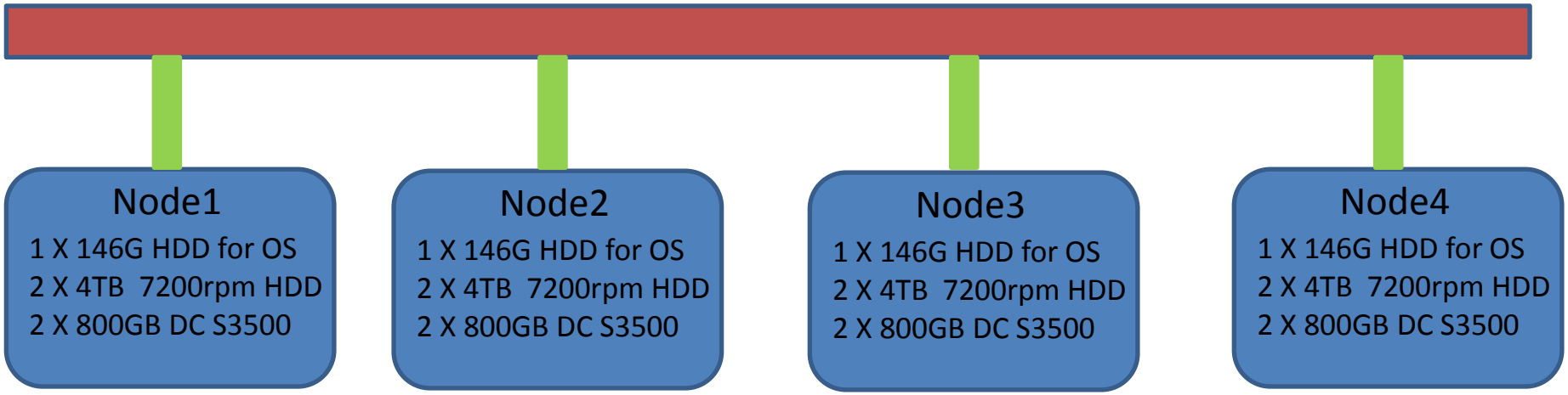
Integrating with OpenStack

Integrating with OpenStack



Performance Test

Test Environment



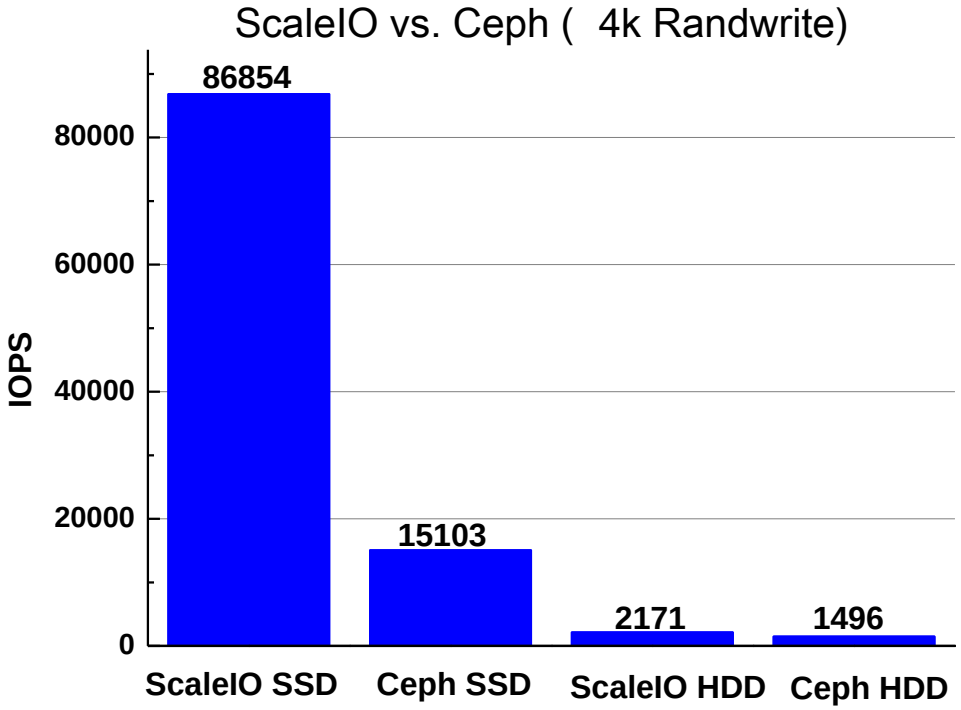
Node Hardware

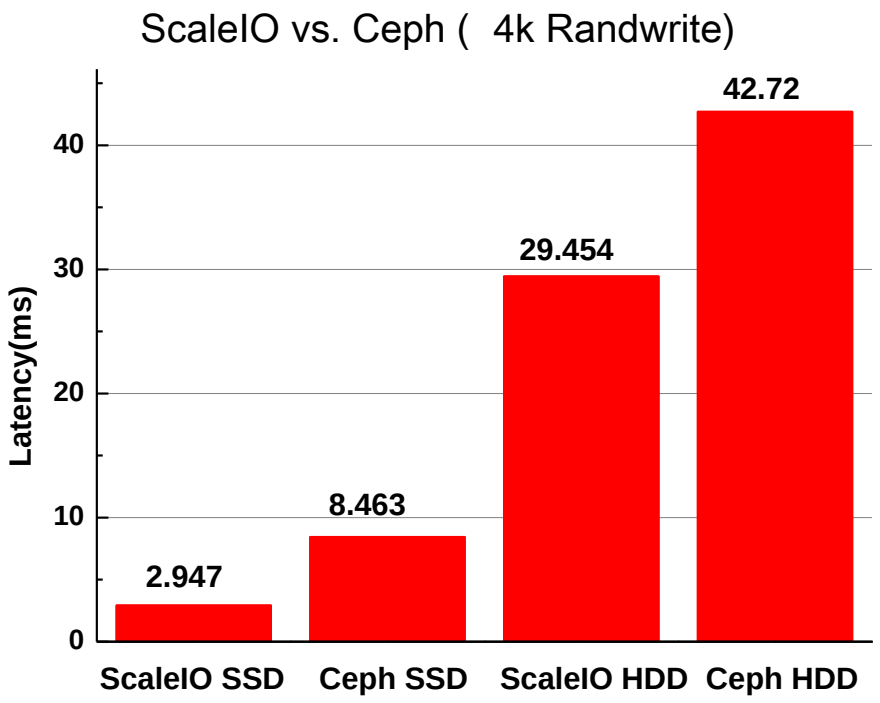
CPU(s)	2*2*10
Memory	128GB
Storage	2 X 4TB HDD, 2X800GB SSD
Network	4 X 10Bb NIC

Software

Ceph	0.94.4
ScaleIO	1.32
Kernel	3.14
OS	Centos 6.5
Tools	fio, vdbench

IOPS and Latency





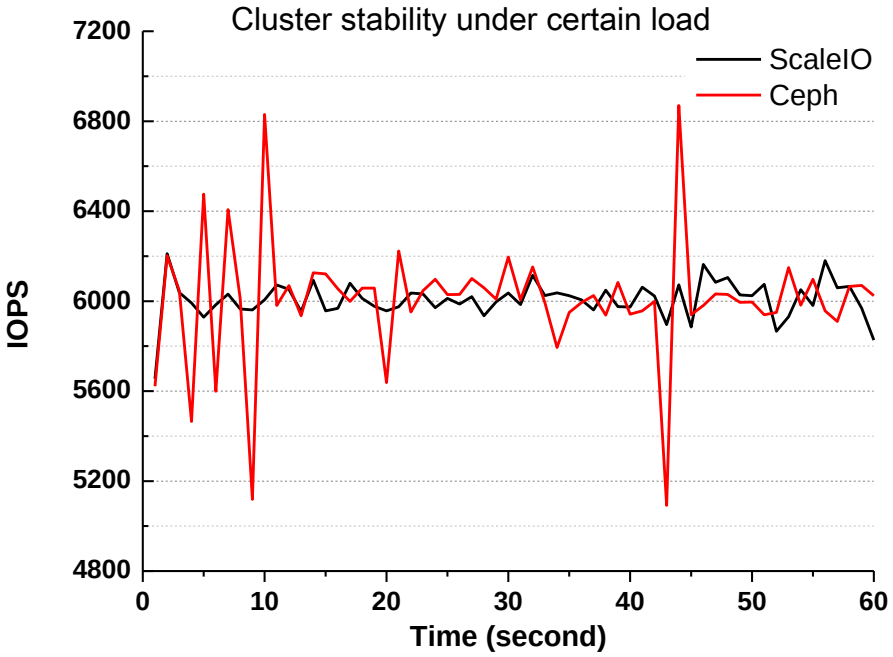
fio -ioengine=libaio -bs=4k -direct=1 -thread -rw=randwrite -size=10G -filename=/dev/scinia -name="test" -iodepth=64 -runtime=30

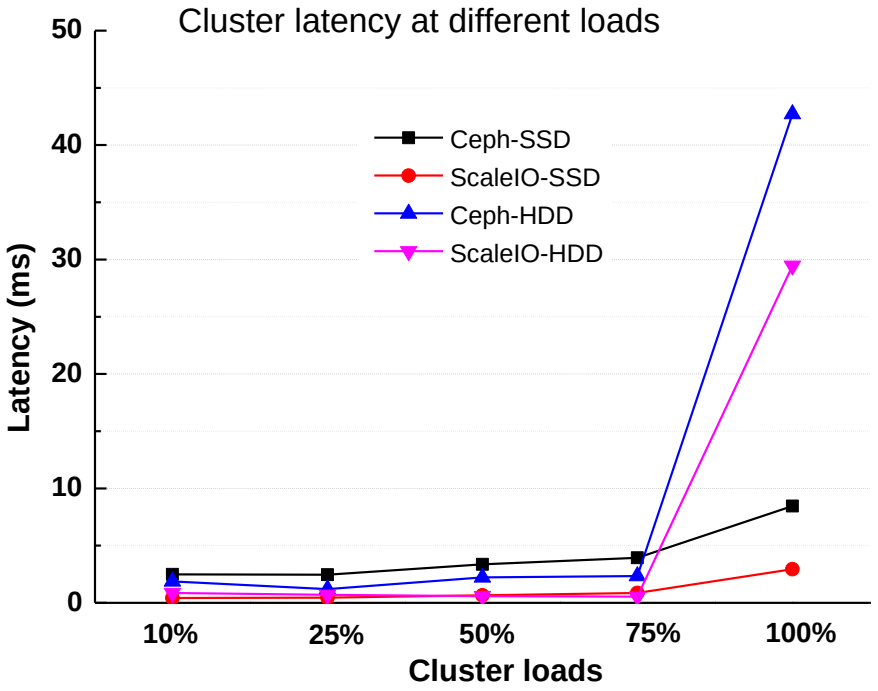
```
[global]
ioengine=rbd
clientname=admin
pool=ssdpool
rbdname=image1
invalidate=0
rw=randwrite
bs=4K
runtime=30
[rbd_iodepth32]
iodepth=64
```

Cluster stability and latency

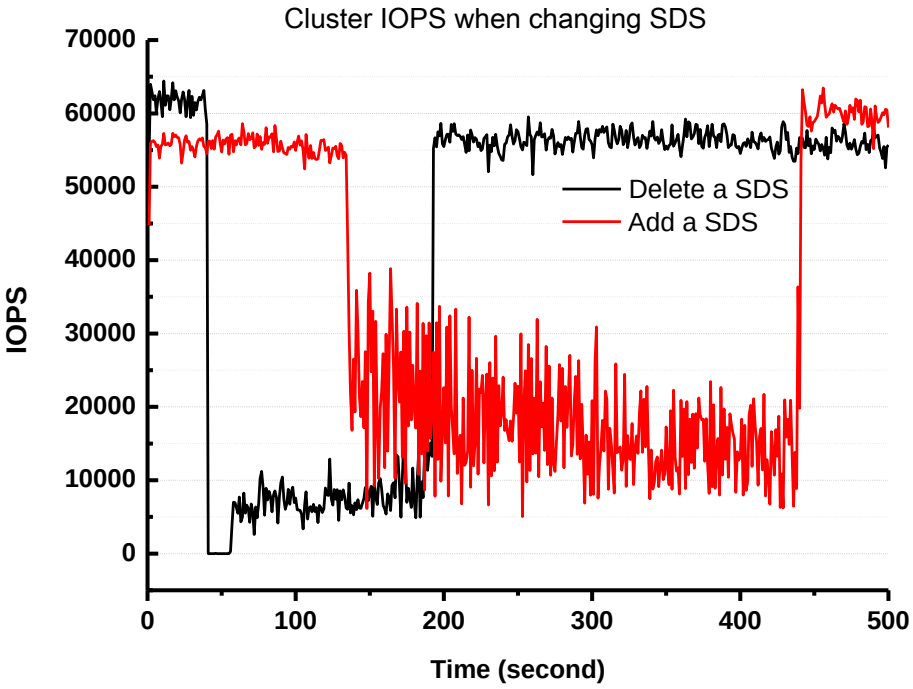
Vdbench

SD: openflags=o_direct,threads=64
WD: xfersize=4096,rdpct=0,seekpct=100
RD: iorate=6000,elapsed=60, interval=1





What happens when scaling out/in



OSD recovery limit in ceph.conf

[global]

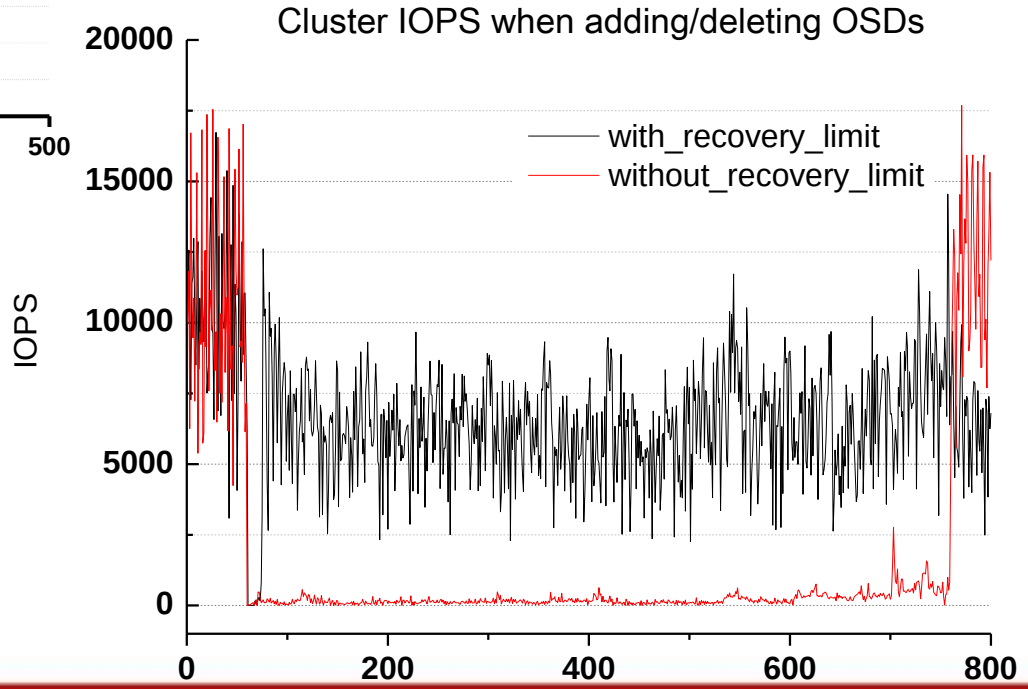
……

[osd]

……

osd_max_backfills = 1

osd_recovery_max_active = 1



Summary

- Ceph need a more user-friendly deployment and management tool
- Ceph lacks of advanced storage features (Qos guarantee, Deduplication, Compression)
- Ceph is the best integration for OpenStack
- Ceph is acceptable for HDD but not good enough for high-performance disk
- Ceph has a lot of configuration parameters, but lacks of relevant instructions and visualization tools

Thank you for your listening.

wuyuting@awcloud.com

Keep an open mind

Lead the cloud world

海云捷迅 AWCLOUD