OpenStack Cinder drive for Lustre

DataDirect Networks, Inc

Shuichi Ihara, Shilong Wang

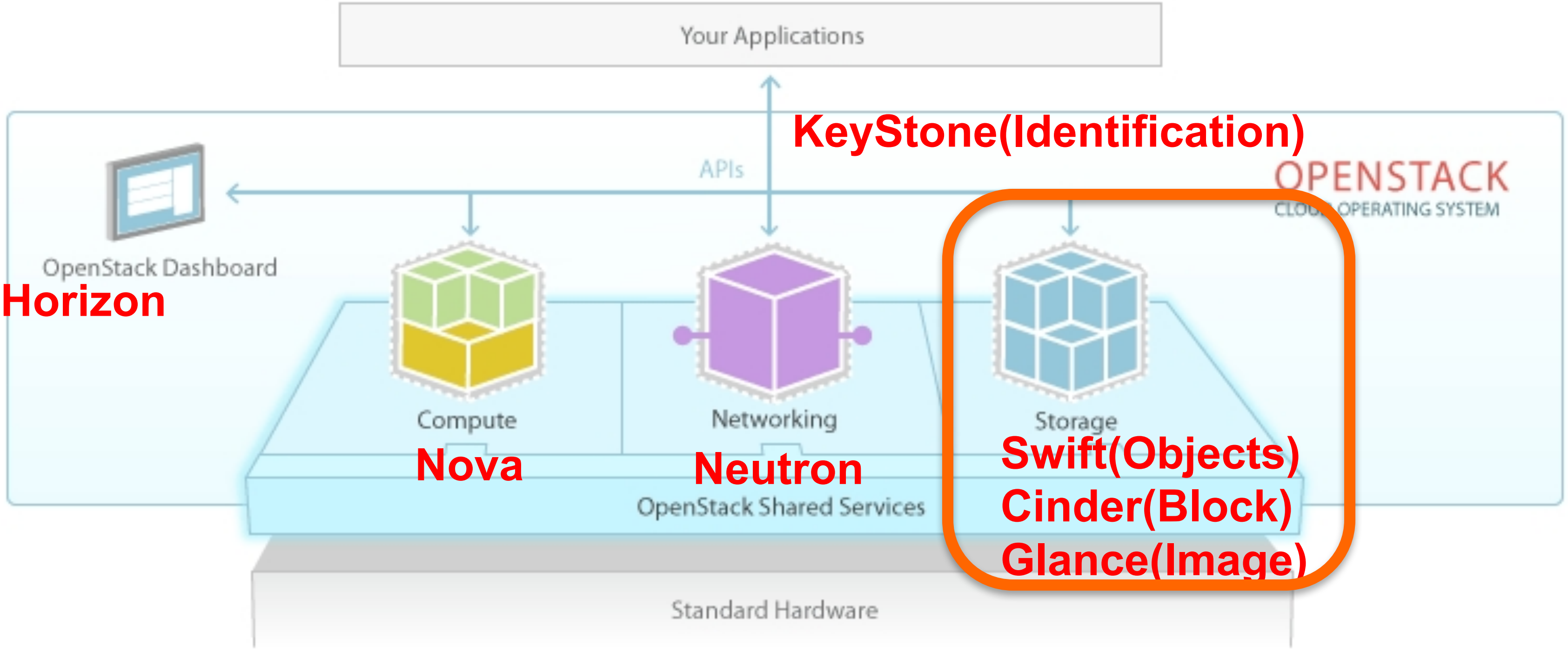
2017/05/31

What's OpenStack?

- OpenStack was an open-source project started in 2010 by RackSpace and NASA and large community many people and companies involved.
- One of widely known software stack at Enterprise system
- A set of software tool for building and managing cloud environment(private or public)
- Provides compute, storage and network so on and API-compatible with AWS

3

Comportment of OpenStack



Storage Service

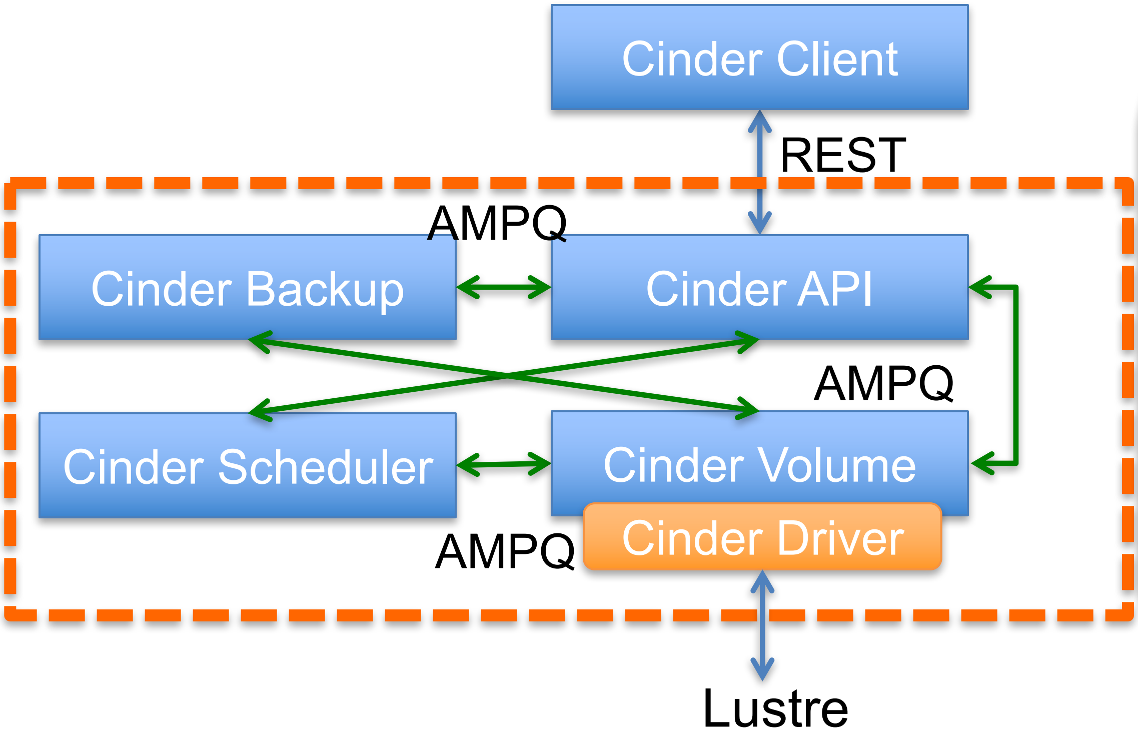
Object Storage Service (SWIFT)

- Full distributed REST API-accessible storage platform
- Supports Multi Tenancy

Block Storage Service (Cinder)

- Provide traditional block level storage resources to other Openstack services. e.g. OpenStack Nova compute instances
- Manage the creation, attach/deatach of volumes between host servers
- Many cinder drivers are available
  - https://wiki.openstack.org/wiki/CinderSupportMatrix
- No Cinder driver for Lustre available Today!

Cinder Architecture Overview



Cinder Functions

Volume Management

- Create, Delete, Show
- Attach/Detach
- Extend, etc

Snapshot

- Create, Delete, Show, Update

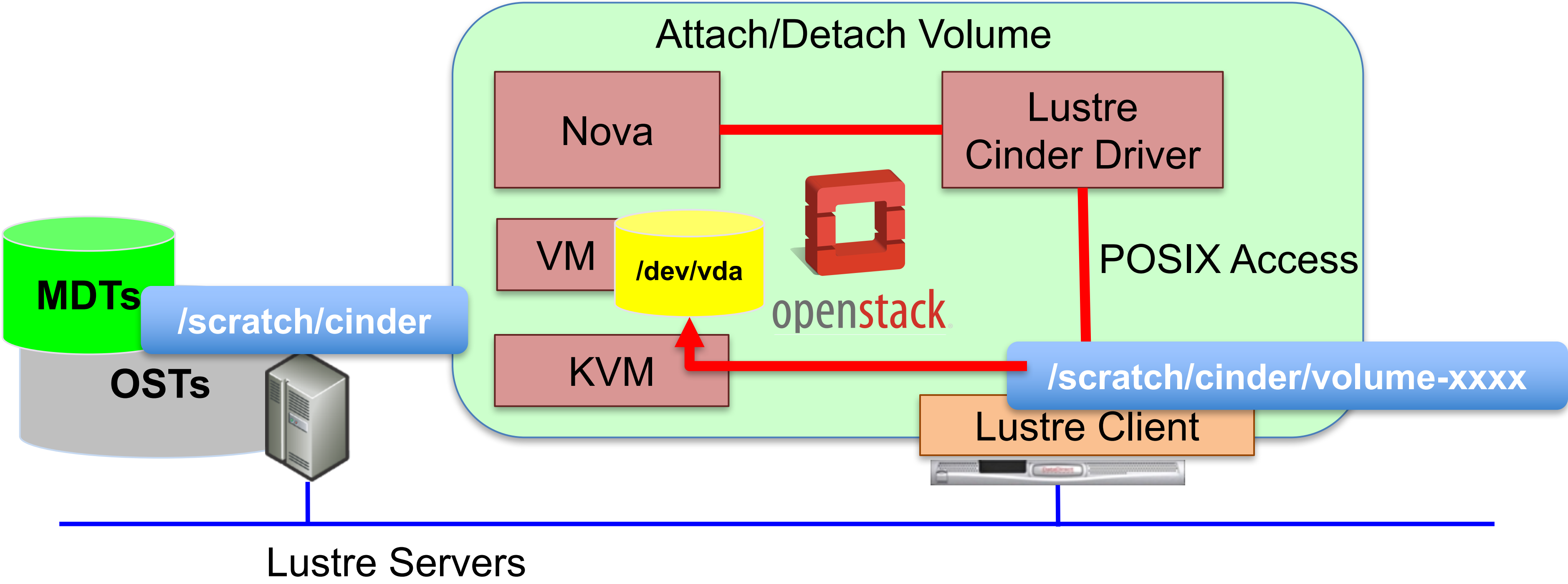
Backup

- Create, Delete, Show, Restore

What does Lustre Cinder driver do?

- Lustre Cinder driver provides block storage to OpenStack's compute service as well as other 3rd party Cinder driver.
- Expose scalable Lustre namespace to multiple VMs on multiple OpenStack hosts
- Bridge on HPC and OpenStack with Lustre. It could make many use case for HPC and Enterprise system
- Buildup Lustre Ecosystem for OpenStack

Architecture of Lustre Cinder driver



How Lustre Cinder driver works(1)

- **Cinder Configuration (/etc/cinder/cinder.conf)**

```
[lustre]
volume_driver = cinder.volume.drivers.lustre.LustreDriver
lustre_share_host = 10.0.10.193@o2ib30:10.0.10.192@o2ib30
lustre_share_path = /scratch/cinder
volume_backend_name = lustre
```

- **Lustre automatically mounted for OpenStack**

```
[root@devstack~]# mount -t lustre
10.0.10.193@o2ib30:10.0.10.192@o2ib30:/scratch/cinder on
/opt/stack/data/cinder/mnt/71ee0200412a18cf142a396734dbb1a4 type lustre
(rw,lazystatfs)
```

How Lustre Cinder driver works(2)

Enabled Lustre Cinder Driver

```
[root@devstack~]# openstack volume service list
+------------------+-----------------+------+---------+-------+----------------------------+
| Binary           | Host            | Zone | Status  | State | Updated At                 |
+------------------+-----------------+------+---------+-------+----------------------------+
| cinder-backup    | devstack        | nova | enabled | up    | 2017-05-21T22:39:31.000000 |
| cinder-scheduler | devstack        | nova | enabled | up    | 2017-05-21T22:39:36.000000 |
| cinder-volume    | devstack@lustre | nova | enabled | up    | 2017-05-21T22:39:30.000000 |
+------------------+-----------------+------+---------+-------+----------------------------+
```

Volume Creation

```
[root@devstack~]# openstack volume create --size 1024 --image CentOS7.3
\ devstack-vm01-vda
```

How Lustre Cinder driver works(3)

Volume List

```
[root@devstack~]# openstack volume list
[root@devstack~]# ls -lh
/opt/stack/data/nova/mnt/71ee0200412a18cf142a396734dbb1a4/volume-*
-rw-rw-rw- 1 qemu qemu 1.0T May 24 00:24
/opt/stack/data/nova/mnt/71ee0200412a18cf142a396734dbb1a4/volume-fbb18151-4f9f-
40e0-a7f7-72f902f752a9
```

Create VM and Attach Volume

```
[root@devstack~]# openstack server create --volume devstack-vm01-vda \
--flavor lustre.client devstack-vm01
[root@devstack~]# ssh devstack-vm01 df -h /dev/vda1
Filesystem      Size  Used Avail Use% Mounted on
/dev/vda1       1.0T  958M  1.0T   1% /
```

Benchmark Configuration

MDS and MDT

- 1 x SuperMicro Server(2 x E5-2690v3, 128GB DIMM, 1 x FDR)
- 1 x SFA7700 and 4 x Toshiba 200GB RI SSD

OSS and OST

- SFA14KXE (ES14K), Single OST (SSD, 8D+1P)
- 1 x OSS included inside of controller/w FDR
- DDN Lustre Distribuution(IEEL3.0 + DDN patches)

Client

- 1 x Dell R620 (2 x E5-2650v2, 128GB DIMM, 1 x FDR)
- Upstream DevStack
- Created 8 x VM (4 CPU cores, 4GB memory, 256GB Volume)

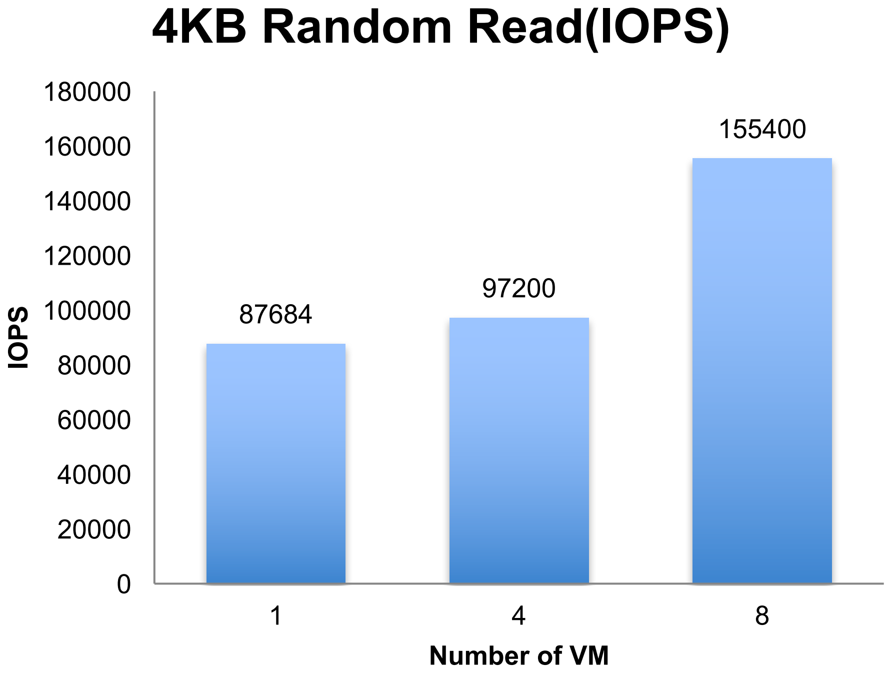
Benchmark Results

4KB Random Read with FIO

- Created large file on 'root' filesystem on each VM (An file to Lustre)
- Run FIO to it on 8 x VMs simultaneously

Removed all read cache

- Disabled Lustre OSS read cache
- VM's cache mode is 'none' which means O_DIRECT to Lustre
- Enabled 'directio' with FIO



Development Status

- **Pushed all patches to gerrit for upstream 'devstack' in OpenStack and under review**
  - Add Cidner driver and support "Lustre" to Nova(VM)
    - https://review.openstack.org/#/c/395572 (397473, 446288 and 446365)
- **Built up Jenkins/CI environment for Lustre Cinder driver**
  - OpenStack requires codes inspections and regression tests pass (same as Lustre), but requires CI infrastructure
  - Many 3rdParty vendors provide CI environment to Openstack community to run tests for Cinder driver
  - DDN contributes and provide resources one of 3rdParty CI infrastructure for general cinder tests

Future plans

- **Merging patches into upstream openstack is first priority**
- **Will add additional features later**
  - Lustre Striping (as well as PFL) support
  - Snapshot support
  - Cloning support
  - JOB Stats integration for performance monitoring and QoS

Lustre Ecosystem for OpenStack

Security and Isolation

- Secured VM environment
  - Subdir mount
  - Authorized data access with Lustre security and Node Map
- Isolated resource management
  - Project Quota, I/O QoS(NRS/TBF), etc

Performance and Performance Management

- Flexible stripe layout with PFL for VM image
- I/O QoS of VMs by Lustre NRS and TBF
- Lustre Performance monitoring for OpeStack

Conclusions

- Developed Lustre Cinder driver to connect OpenStack and Lustre
- Demonstrated minimum required functionalities are working well
- Contributing all patches to OpenStack community and working on merging all patches into upstream OpenStack
- Will extend functions in Lustre Cinder driver and integrate with other Lustre features