SK telecom

GlusterFS

가상화WG

Global ICT Leader

목차

1. GlusterFS 개요

2. GlusterFS 구조

3. GlusterFS Volume Type

4. GlusterFS 서버 설치

5. GlusterFS Brick 생성

6. GlusterFS Distributed Vol 생성

7. GlusterFS Replicated Vol 생성

8. GlusterFS Distributed Replicated Vol 생성

9. GlusterFS 추가 명령어(Option, Client)

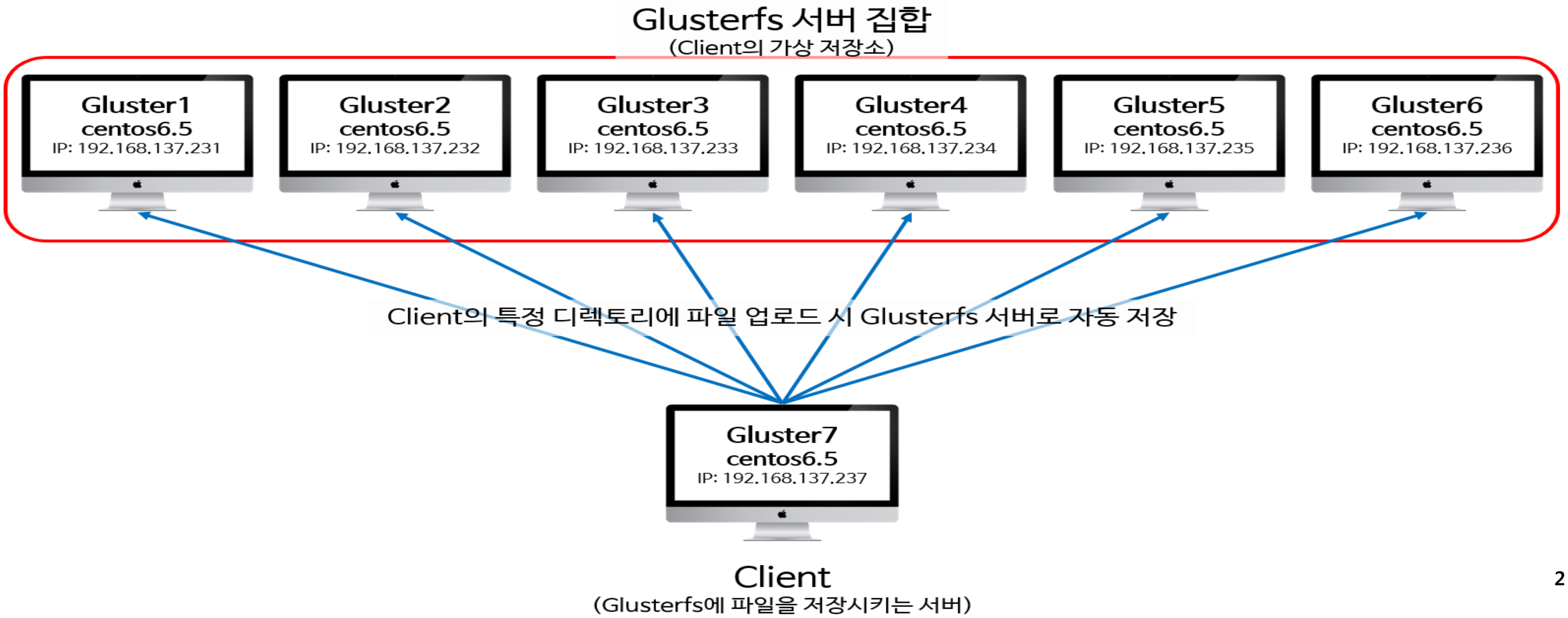
10. GlusterFS 관리 모니터링 tool

11. 오픈스택 연동

1. GlusterFS 개요

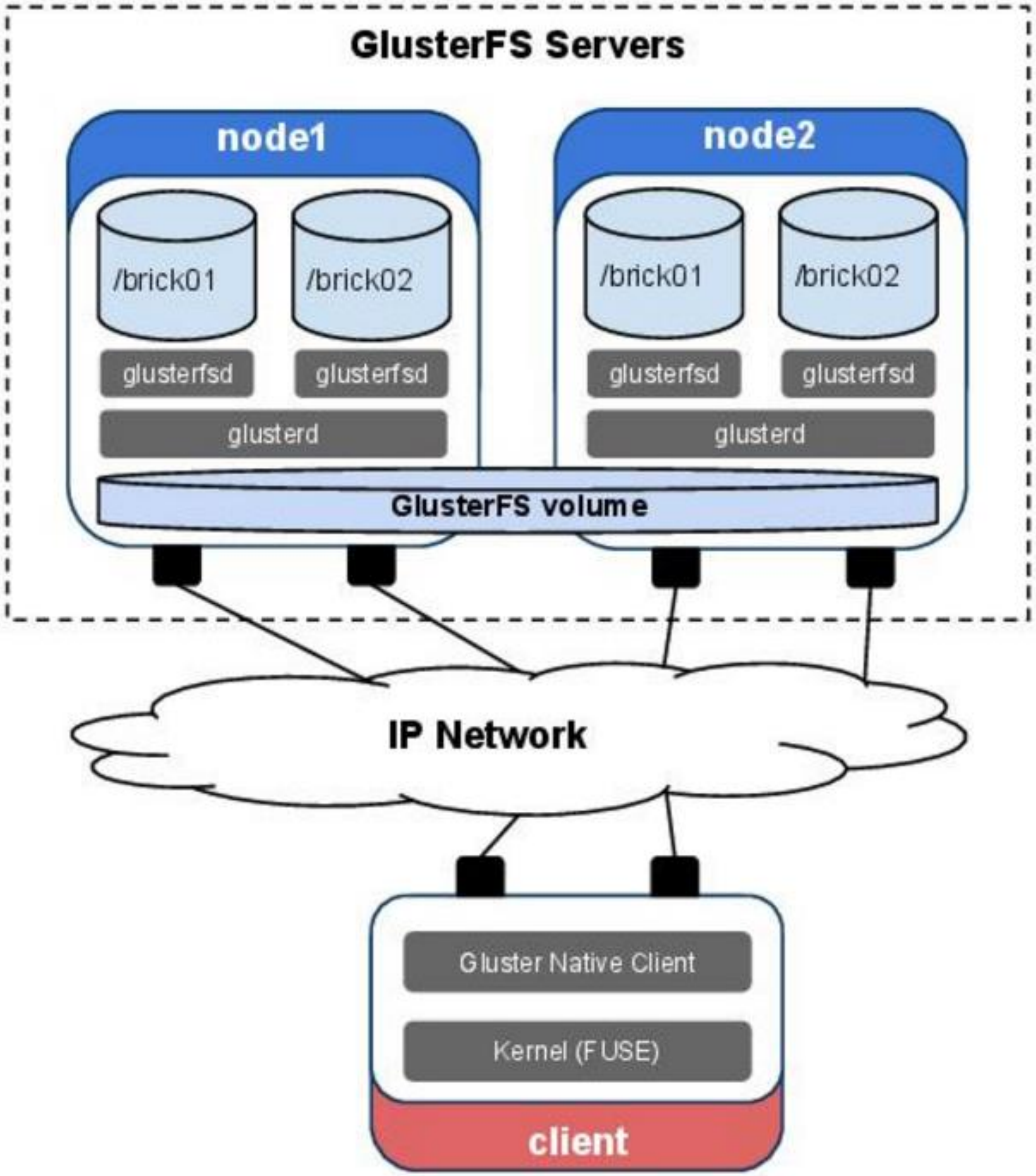
GlusterFS란

- Redhat에서 지원하는 오픈소스 파일시스템
- 수천 Petabyte급의 대용량에 수천 개의 클라이언트가 접속하여 사용 가능
- scale-out 방식 분산 파일 시스템
- 기존의 분산 파일 시스템에 비해 비교적 구성이 간단
- 대용량 및 대규모의 I/O처리 능력이 뛰어남
- Client에서 native(FUSE), NFS, CIFS 방식으로 접근가능



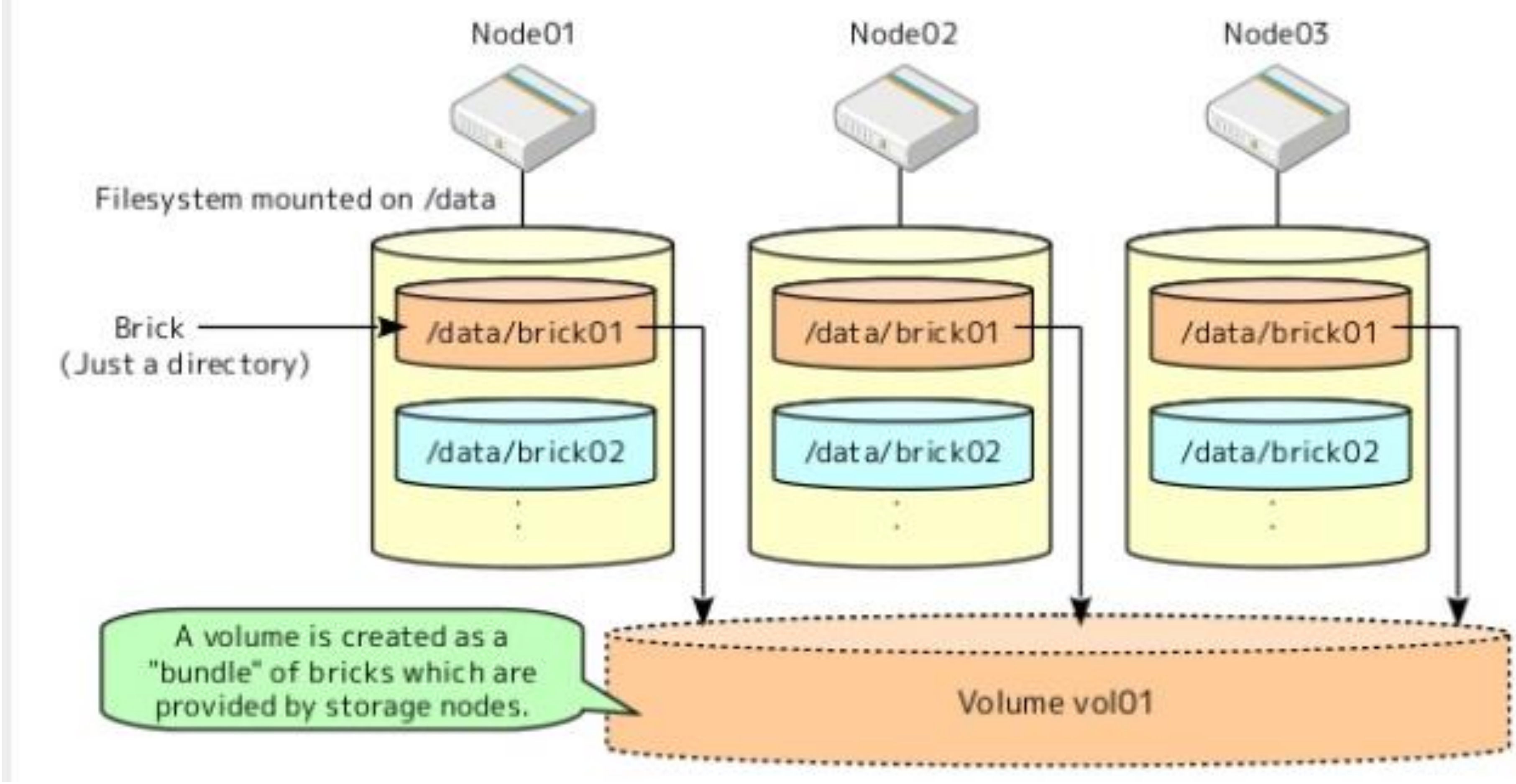
2. GlusterFS 구조

서버 – 클라이언트 연동 구조

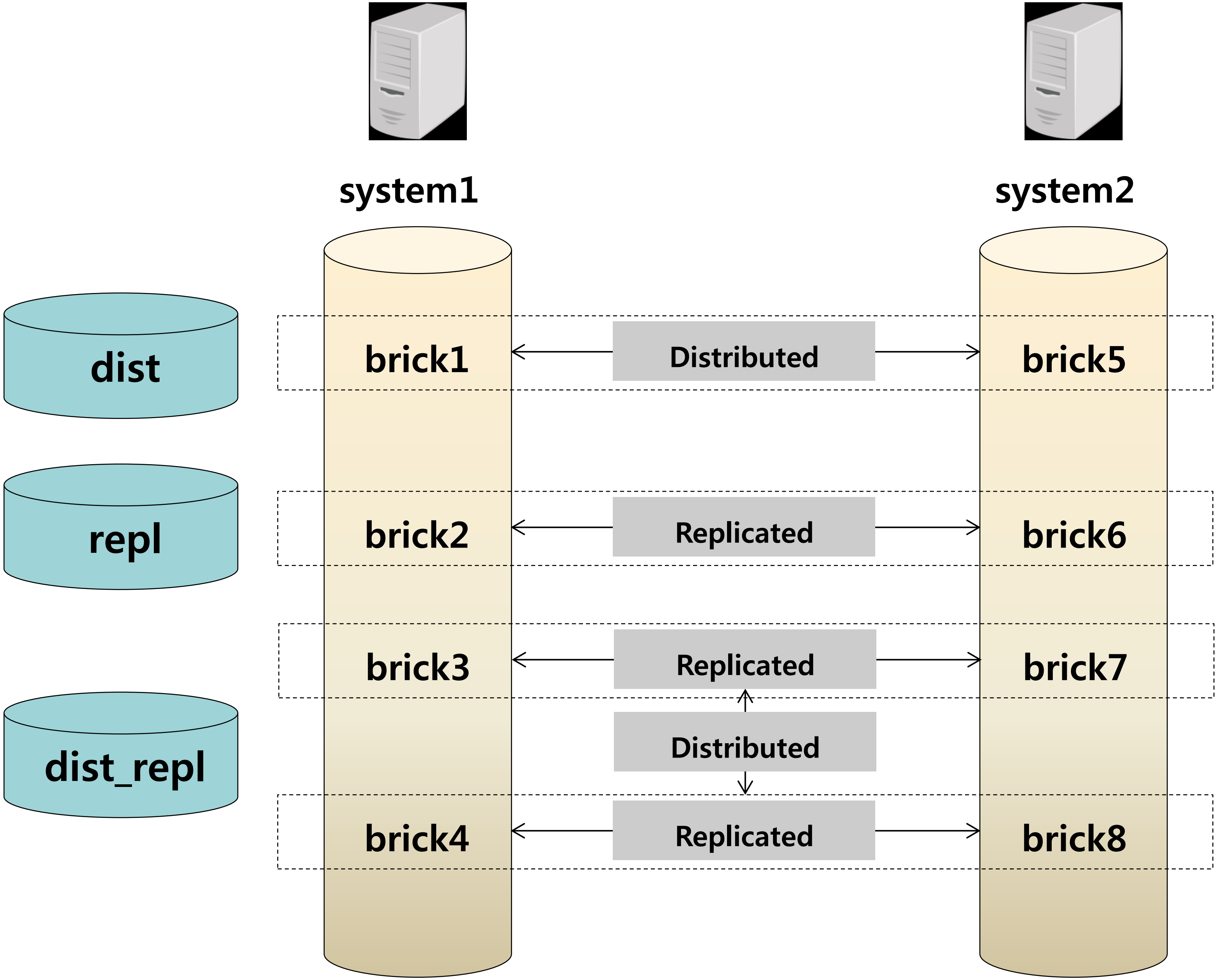


2. GlusterFS 구조

Node는 여러 개의 Brick을 구성할 수 있고 이 중 부분집합을 만들어 Volume을 구성하고 이 Volume이 Client에게 제공된다.

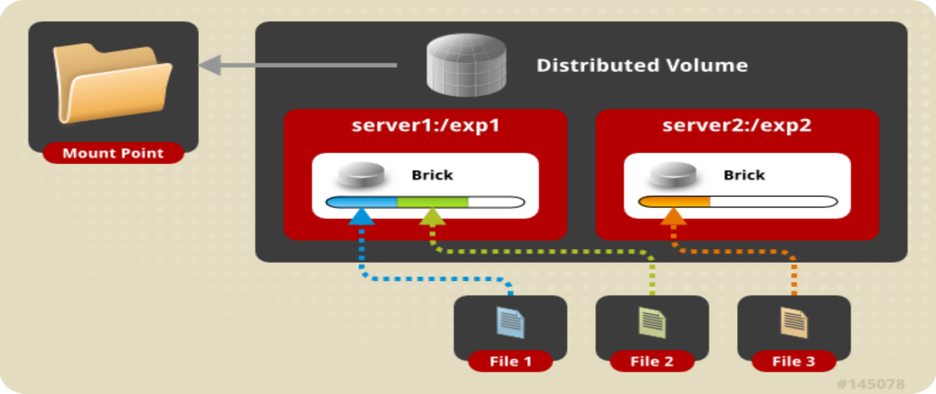


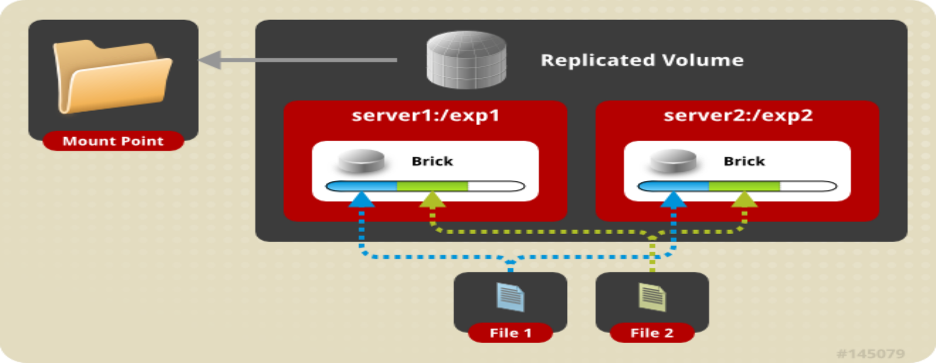
2. GlusterFS 구조

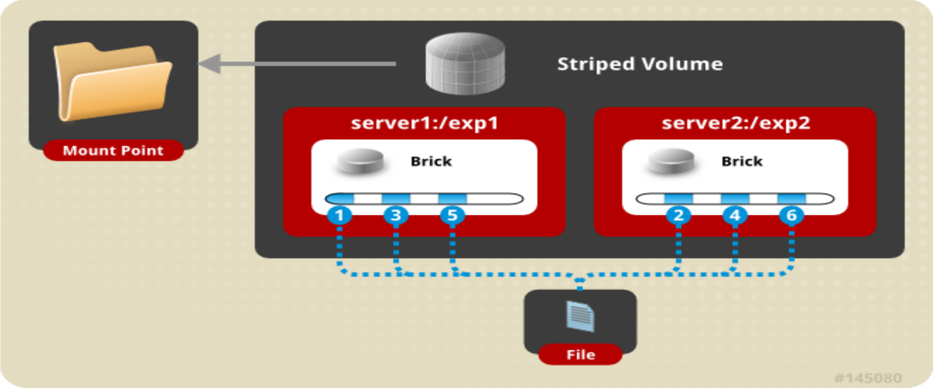


3. GlusterFS volume type

기본 type

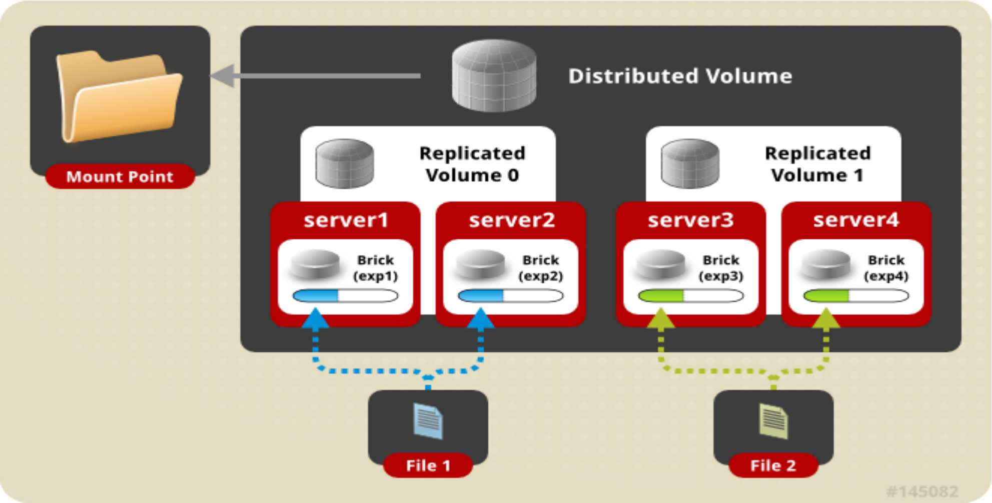




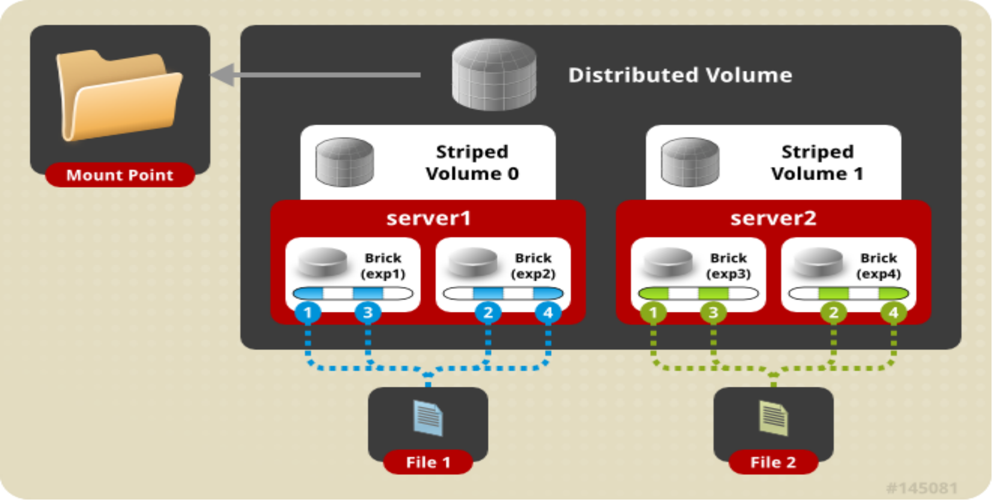


복합 type

Distributed Replicated Volume



Distributed Striped Volume



4. GlusterFS 서버 설치

- **glusterfs 패키지 설치**

```
[root@system01 ~]# yum install -y centos-release-gluster  → glusterfs 패키지 repository 설정
[root@system01 ~]# yum install -y glusterfs-server  → 실제 glusterfs 패키지 설치
[root@system01 ~]# rpm -qa | grep gluster
glusterfs-3.7.1-16.0.1.el7.centos.x86_64
glusterfs-libs-3.7.1-16.0.1.el7.centos.x86_64
```

- **초기 환경 구성시 연동한 dev 2개 확인**

```
[root@system01 ~]# fdisk -l | grep /dev
Disk /dev/sda: 32.2 GB, 32212254720 bytes, 62914560 sectors
/dev/sda1   *        2048     1026047      512000   83  Linux
/dev/sda2         1026048    62914559    30944256   8e  Linux LVM
Disk /dev/sdb: 100 GB, 10737418240 bytes, 20971520 sectors
# lsblk
sdb                           8:16    0    5G  0 disk
 ├─gv01-brickspool_tmeta    253:2    0    8M  0 lvm
 │  └─gv01-brickspool-tpool 253:4    0    5G  0 lvm
 │    ├─gv01-brickspool     253:5    0    5G  0 lvm
 │    ├─gv01-brick1         253:6    0   50G  0 lvm  /brick1
 │    └─gv01-brick2         253:7    0   50G  0 lvm  /brick2
 └─gv01-brickspool_tdata    253:3    0    5G  0 lvm
   └─gv01-brickspool-tpool 253:4    0    5G  0 lvm
     ├─gv01-brickspool     253:5    0    5G  0 lvm
     ├─gv01-brick1         253:6    0   50G  0 lvm  /brick1
     └─gv01-brick2         253:7    0   50G  0 lvm  /brick2
```

4. GlusterFS 서버 설치

- **vi /etc/hosts -> hostname 등록 (ssh로 접속가능 해야함)**

```
[root@system01 ~]# cat /etc/hosts
127.0.0.1   localhost localhost.localdomain localhost4 localhost4.localdomain4
::1         localhost localhost.localdomain localhost6 localhost6.localdomain6
192.168.100.21 system01.example.com system01
192.168.100.22 system02.example.com system02
```

주의 : systemctl stop firewalld.service

- **노드간 연동 테스트**

```
[root@system1 ~]# gluster peer probe system2.example.com
peer probe: success.
[root@system1 ~]# gluster peer status
Number of Peers: 1

Hostname: system2.example.com
Uuid: 351418b0-6d8f-435f-a0d5-335f2ed1abaa
State: Peer in Cluster (Connected)
```

5. GlusterFS Brick 생성

(system1에서)

```
[root@system1 ~]# pvcreate /dev/sdb
  Physical volume "/dev/sdb" successfully created
[root@system1 ~]# vgcreate vg0 /dev/sdb
  Volume group "vg0" successfully created
[root@system1 ~]# lvcreate -l100%free -T vg0/brickspool
  Logical volume "brickspool" created.
[root@system1 ~]# lvcreate -V 100M -T vg0/brickspool -n brick1
  Logical volume "brick1" created.
[root@system1 ~]# lvcreate -V 100M -T vg0/brickspool -n brick2
  Logical volume "brick2" created.
[root@system1 ~]# lvcreate -V 100M -T vg0/brickspool -n brick3
  Logical volume "brick3" created.
[root@system1 ~]# lvcreate -V 100M -T vg0
/brickspool -n brick4
[root@system1 ~]# lvs
  LV         VG     Attr       LSize    Pool       Origin Data%  Meta%  Move Log Cpy%Sync Convert
  root       centos -wi-ao----  11.17g
  swap       centos -wi-ao----   1.30g
  brick1     vg0    Vwi-a-tz--  100.00m brickspool        0.00
  brick2     vg0    Vwi-a-tz--  100.00m brickspool        0.00
  brick3     vg0    Vwi-a-tz--  100.00m brickspool        0.00
  brick4     vg0    Vwi-a-tz--  100.00m brickspool        0.00
  brickspool vg0    twi-aotz-- 1012.00m                   0.00   1.27
```

5. GlusterFS Brick 생성

(system1에서)

```
[root@system1 ~]# mkfs.xfs -i size=512 /dev/vg0/brick1
meta-data=/dev/vg0/brick1        isize=512    agcount=4, agsize=6384 blks
         =                       sectsz=512   attr=2, projid32bit=1
         =                       crc=0        finobt=0
data     =                       bsize=4096   blocks=25536, imaxpct=25
         =                       sunit=16     swidth=16 blks
naming   =version 2              bsize=4096   ascii-ci=0 ftype=0
log      =internal log           bsize=4096   blocks=768, version=2
         =                       sectsz=512   sunit=16 blks, lazy-count=1
realtime =none                   extsz=4096   blocks=0, rtextents=0
[root@system1 ~]# mkdir -p /brick1
vi /etc/fstab 에 추가
/dev/vg0/brick1          /brick1                xfs        rw,noatime,inode64,nouuid        1 2
[root@system1 ~]# mount -a
[root@system1 ~]# df -h
Filesystem               Size  Used Avail Use% Mounted on
/dev/mapper/centos-root   12G  4.7G  6.5G  42% /
devtmpfs                 474M     0  474M   0% /dev
tmpfs                    489M  144K  489M   1% /dev/shm
tmpfs                    489M   14M  476M   3% /run
tmpfs                    489M     0  489M   0% /sys/fs/cgroup
/dev/sda1                497M  177M  321M  36% /boot
tmpfs                     98M   16K   98M   1% /run/user/0
/dev/mapper/vg0-brick1    97M  5.2M   92M   6% /brick1
```

5. GlusterFS Brick 생성

(system2에서)

```
[root@system2 ~]# pvcreate /dev/sdb
  Physical volume "/dev/sdb" successfully created
[root@system2 ~]# vgcreate vg0 /dev/sdb
  Volume group "vg0" successfully created
[root@system2 ~]# lvcreate -l100%free -T vg0/brickspool
  Logical volume "brickspool" created.
[root@system2 ~]# lvcreate -V 100M -T vg0/brickspool -n brick5
  Logical volume "brick5" created.
[root@system2 ~]# lvcreate -V 100M -T vg0/brickspool -n brick6
  Logical volume "brick6" created.
[root@system2 ~]# lvcreate -V 100M -T vg0/brickspool -n brick7
  Logical volume "brick7" created.
[root@system2 ~]# lvcreate -V 100M -T vg0/brickspool -n brick8
  Logical volume "brick8" created.
[root@system2 ~]# lvs
  LV         VG     Attr       LSize    Pool       Origin Data%  Meta%  Move Log Cpy%Sync Convert
  root       centos -wi-ao----   11.17g
  swap       centos -wi-ao----    1.30g
  brick5     vg0    Vwi-a-tz--  100.00m brickspool        0.00
  brick6     vg0    Vwi-a-tz--  100.00m brickspool        0.00
  brick7     vg0    Vwi-a-tz--  100.00m brickspool        0.00
  brick8     vg0    Vwi-a-tz--  100.00m brickspool        0.00
  brickspool vg0    twi-aotz-- 1012.00m                   0.00   1.27
```

5. GlusterFS Brick 생성

(system2에서)

```
[root@system2 ~]# mkfs.xfs -i size=512 /dev/vg0/brick5
meta-data=/dev/vg0/brick5        isize=512    agcount=4, agsize=6384 blks
         =                       sectsz=512   attr=2, projid32bit=1
         =                       crc=0        finobt=0
data     =                       bsize=4096   blocks=25536, imaxpct=25
         =                       sunit=16     swidth=16 blks
naming   =version 2              bsize=4096   ascii-ci=0 ftype=0
log      =internal log           bsize=4096   blocks=768, version=2
         =                       sectsz=512   sunit=16 blks, lazy-count=1
realtime =none                   extsz=4096   blocks=0, rtextents=0
[root@system2 ~]# mkdir -p /brick5
[root@system2 ~]# vi /etc/fstab
/dev/vg0/brick1                 /brick1                         xfs         rw,noatime,inode64,nouuid        1 2
[root@system2 ~]# mount -a
[root@system2 ~]# df -h
Filesystem               Size  Used Avail Use% Mounted on
/dev/mapper/centos-root   12G  4.7G  6.5G  42% /
devtmpfs                 474M     0  474M   0% /dev
tmpfs                    489M  144K  489M   1% /dev/shm
tmpfs                    489M   14M  476M   3% /run
tmpfs                    489M     0  489M   0% /sys/fs/cgroup
/dev/sda1                497M  177M  321M  36% /boot
tmpfs                     98M  4.0K   98M   1% /run/user/42
tmpfs                     98M   16K   98M   1% /run/user/0
/dev/mapper/vg0-brick5    97M  5.2M   92M   6% /brick5
```

6. GlusterFS Distributed Vol 생성

(system1에서)

```
[root@system1 brick1]# gluster vol create dist
system1.example.com:/brick1/brick
system2.example.com:/brick5/brick
volume create: dist: success: please start the volume to access data
[root@system1 brick1]# gluster vol start  dist
volume start: dist: success
[root@system1 brick1]# gluster vol info dist

Volume Name: dist
Type: Distribute
Volume ID: bcb947c3-ec2d-4ba9-9ebc-06eed78c22a4
Status: Started
Snapshot Count: 0
Number of Bricks: 2
Transport-type: tcp
Bricks:
Brick1: system1.example.com:/brick1/brick
Brick2: system2.example.com:/brick5/brick
Options Reconfigured:
transport.address-family: inet
performance.readdir-ahead: on
nfs.disable: on
```

7. GlusterFS Replicated Vol 생성

(system1에서)

```
[root@system1 brick1]# gluster vol create repl replica 2
system1.example.com:/brick2/brick
system2.example.com:/brick6/brick
volume create: repl: success: please start the volume to access data
[root@system1 brick1]# gluster vol start repl
volume start: repl: success
[root@system1 brick1]# gluster vol info repl

Volume Name: repl
Type: Replicate
Volume ID: e6d4035a-597d-4825-ab74-46cb6ca87110
Status: Started
Snapshot Count: 0
Number of Bricks: 1 x 2 = 2
Transport-type: tcp
Bricks:
Brick1: system1.example.com:/brick2/brick
Brick2: system2.example.com:/brick6/brick
Options Reconfigured:
transport.address-family: inet
performance.readdir-ahead: on
nfs.disable: on
```

8. GlusterFS Distributed Replicated Vol 생성

(system1에서)

```
[root@system1 brick1]# gluster vol3 create dist_repl replica 2
system1.example.com:/brick3/brick system2.example.com:/brick7/brick
system1.example.com:/brick4/brick system2.example.com:/brick8/brick
volume create: dist_repl: success: please start the volume to access data
[root@system1 brick1]# gluster vol start dist_repl
volume start: dist_repl: success
[root@system1 brick1]# gluster vol info dist_repl

Volume Name: dist_repl
Type: Distributed-Replicate
Volume ID: f72b5da7-14ad-4358-ad60-9c528c43788f
Status: Started
Snapshot Count: 0
Number of Bricks: 2 x 2 = 4
Transport-type: tcp
Bricks:
Brick1: system1.example.com:/brick3/brick
Brick2: system2.example.com:/brick7/brick
Brick3: system1.example.com:/brick4/brick
Brick4: system2.example.com:/brick8/brick
Options Reconfigured:
transport.address-family: inet
performance.readdir-ahead: on
nfs.disable: on
```

9. GlusterFS 추가 명령어(Option, Client)

```
[root@system1 brick1]# gluster vol set repl nfs.disable on
volume set: success
[root@system1 brick1]# gluster vol info repl

Volume Name: repl
…
nfs.disable: on

[root@system1 brick1]# gluster vol reset repl nfs.disable
volume reset: success: reset volume successful
```

9. GlusterFS Client 구성

GlusterFS native client 방식

```
yum install glusterfs-fuse
mkdir /distvol
vi /etc/fstab
system1.example.com:/dist     /distvol    glusterfs                 _netdev    0 0
mount –a
touch /distvol/file{1..10}
```

서버에서 5개씩 분산되어 생성되는지 확인

```
ls –l /brick1/brick
ls –l /brick5/brick
```

GlusterFS NFS client 방식

```
yum install glusterfs-fuse
mkdir /replvol
vi /etc/fstab
system1.example.com:/repl     /replvol    nfs                       vers=3   0 0
mount –a
touch /replvol/file{1..10}
```

서버에서 10개가 복제되어 생성되는지 확인

```
ls –l /brick2/brick
ls –l /brick6/brick
```

9. GlusterFS Client 구성

GlusterFS CIFS client 방식

[system1]

```
systemctl start smb.service
useradd –s /sbin/nologin cifsuser
smbpasswd –a cifsuser <- 패스워드 2번 입력
mount system1.example.com:/dist_repl                /mnt
chown :cifsuser   /mnt
chmod 775 /mnt
umount /mnt
```

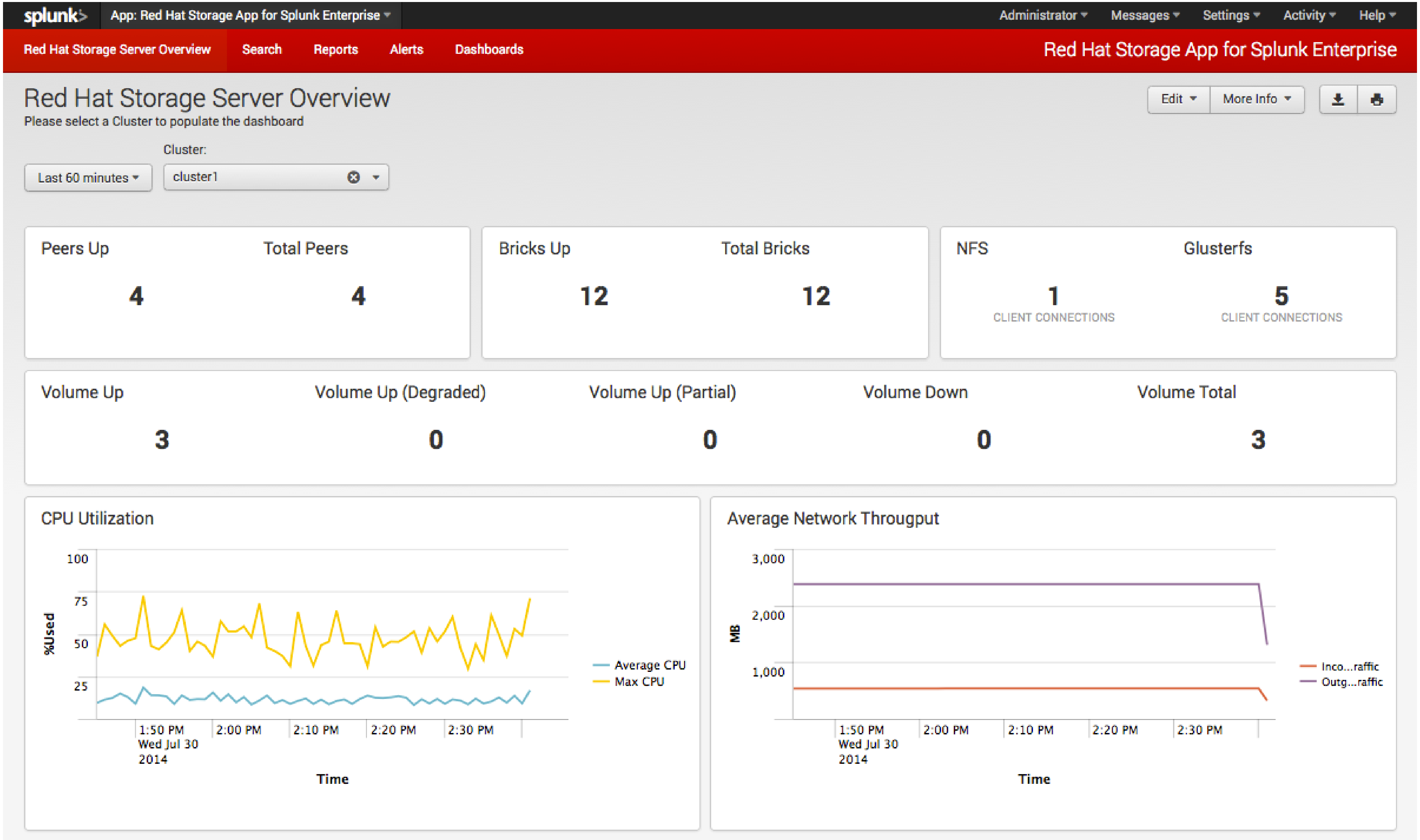
[client]

```
yum install glusterfs-fuse
yum install cifs-utils
mkdir /distreplvol
vi /etc/fstab
//system1.example.com:/gluster-dist_repl       /distreplvol       cifs   user=cifsuser,password=redhat   0 0
mount –a
touch /distreplvol/file{1..10}
```

서버에서 10개가 분배되어 복제되어 생성되는지 확인

```
ls –l /brick3/brick
ls –l /brick4/brick
ls –l /brick7/brick
ls –l /brick8/brick
```

10. GlusterFS 관리 모니터링 tool


splunk>
App: Red Hat Storage App for Splunk Enterprise
Administrator
Messages
Settings
Activity
Help
Red Hat Storage Server Overview
Search
Reports
Alerts
Dashboards
Red Hat Storage App for Splunk Enterprise
Edit
More Info

Red Hat Storage Server Overview

Please select a Cluster to populate the dashboard

Last 60 minutes

Cluster: cluster1

Peers Up 4

Total Peers 4

Bricks Up 12

Total Bricks 12

NFS 1 CLIENT CONNECTIONS

Glusterfs 5 CLIENT CONNECTIONS

Volume Up 3

Volume Up (Degraded) 0

Volume Up (Partial) 0

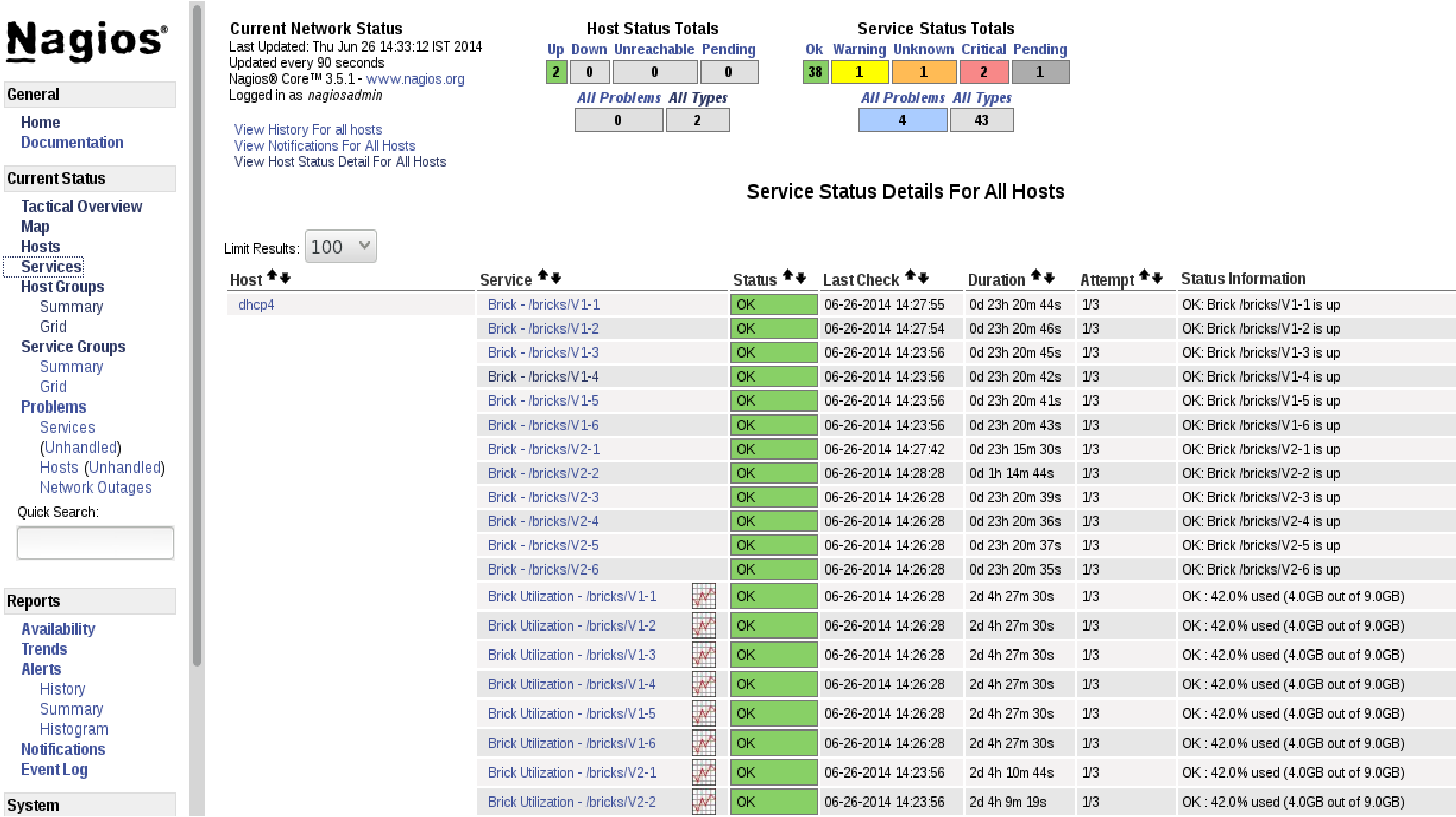
Volume Down 0

Volume Total 3

CPU Utilization

Average Network Througput

10. GlusterFS 관리 모니터링 tool



Current Network Status
Last Updated: Thu Jun 26 14:33:12 IST 2014
Updated every 90 seconds
Nagios® Core™ 3.5.1 - www.nagios.org
Logged in as *nagiosadmin*

View History For all hosts
View Notifications For All Hosts
View Host Status Detail For All Hosts

Host Status Totals

Up	Down	Unreachable	Pending
2	0	0	0

All Problems	*All Types*
0	2

Service Status Totals

Ok	Warning	Unknown	Critical	Pending
38	1	1	2	1

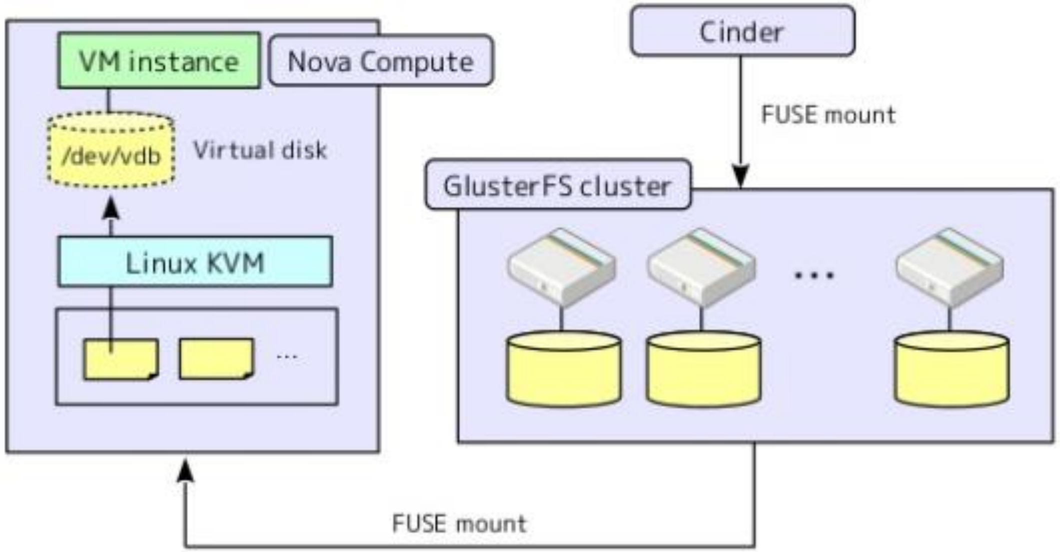
All Problems	*All Types*
4	43

Service Status Details For All Hosts

Limit Results: 100

Host	Service	Status	Last Check	Duration	Attempt	Status Information
dhcp4	Brick - /bricks/V1-1	OK	06-26-2014 14:27:55	0d 23h 20m 44s	1/3	OK: Brick /bricks/V1-1 is up
	Brick - /bricks/V1-2	OK	06-26-2014 14:27:54	0d 23h 20m 46s	1/3	OK: Brick /bricks/V1-2 is up
	Brick - /bricks/V1-3	OK	06-26-2014 14:23:56	0d 23h 20m 45s	1/3	OK: Brick /bricks/V1-3 is up
	Brick - /bricks/V1-4	OK	06-26-2014 14:23:56	0d 23h 20m 42s	1/3	OK: Brick /bricks/V1-4 is up
	Brick - /bricks/V1-5	OK	06-26-2014 14:23:56	0d 23h 20m 41s	1/3	OK: Brick /bricks/V1-5 is up
	Brick - /bricks/V1-6	OK	06-26-2014 14:23:56	0d 23h 20m 43s	1/3	OK: Brick /bricks/V1-6 is up
	Brick - /bricks/V2-1	OK	06-26-2014 14:27:42	0d 23h 15m 30s	1/3	OK: Brick /bricks/V2-1 is up
	Brick - /bricks/V2-2	OK	06-26-2014 14:28:28	0d 1h 14m 44s	1/3	OK: Brick /bricks/V2-2 is up
	Brick - /bricks/V2-3	OK	06-26-2014 14:26:28	0d 23h 20m 39s	1/3	OK: Brick /bricks/V2-3 is up
	Brick - /bricks/V2-4	OK	06-26-2014 14:26:28	0d 23h 20m 36s	1/3	OK: Brick /bricks/V2-4 is up
	Brick - /bricks/V2-5	OK	06-26-2014 14:26:28	0d 23h 20m 37s	1/3	OK: Brick /bricks/V2-5 is up
	Brick - /bricks/V2-6	OK	06-26-2014 14:26:28	0d 23h 20m 35s	1/3	OK: Brick /bricks/V2-6 is up
	Brick Utilization - /bricks/V1-1	OK	06-26-2014 14:26:28	2d 4h 27m 30s	1/3	OK : 42.0% used (4.0GB out of 9.0GB)
	Brick Utilization - /bricks/V1-2	OK	06-26-2014 14:26:28	2d 4h 27m 30s	1/3	OK : 42.0% used (4.0GB out of 9.0GB)
	Brick Utilization - /bricks/V1-3	OK	06-26-2014 14:26:28	2d 4h 27m 30s	1/3	OK : 42.0% used (4.0GB out of 9.0GB)
	Brick Utilization - /bricks/V1-4	OK	06-26-2014 14:26:28	2d 4h 27m 30s	1/3	OK : 42.0% used (4.0GB out of 9.0GB)
	Brick Utilization - /bricks/V1-5	OK	06-26-2014 14:26:28	2d 4h 27m 30s	1/3	OK : 42.0% used (4.0GB out of 9.0GB)
	Brick Utilization - /bricks/V1-6	OK	06-26-2014 14:26:28	2d 4h 27m 30s	1/3	OK : 42.0% used (4.0GB out of 9.0GB)
	Brick Utilization - /bricks/V2-1	OK	06-26-2014 14:23:56	2d 4h 10m 44s	1/3	OK : 42.0% used (4.0GB out of 9.0GB)
	Brick Utilization - /bricks/V2-2	OK	06-26-2014 14:23:56	2d 4h 9m 19s	1/3	OK : 42.0% used (4.0GB out of 9.0GB)

11. 오픈스택 연동



- **Ocata 버전부터는 GlusterFS서버에 추가적인 option 설정필요(수강생은 참고)**

```
# useradd cinder → uid, gid 확인
# gluster volume set VOL_NAME storage.owner-uid CINDER_UID
# gluster volume set VOL_NAME storage.owner-gid CINDER_GID
# gluster volume set VOL_NAME server.allow-insecure on

vi /etc/glusterfs/glusterd.vol 에서 아래옵션 추가 (모든 gluster 서버에서)
 option rpc-auth-allow-insecure on
 systemctl restart glusterd

 gluster vol vol1 set readdir-ahead off
```

[실습] 11. 오픈스택 연동

- **glusterfs 설치 yum install -y glusterfs-fuse**
- **cinder의 경우backend로 LVM, GLUSTERFS 등 구성하여 사용가능**

  - /etc/cinder/cinder.conf에 아래와 같이 설정

```
[defaults]
…
enabled_backends = lvm, glusterfs
…
[glusterfs]
nfs_shares_config=/etc/cinder/glusterfs
volume_driver=cinder.volume.drivers.nfs.NfsDriver
volume_backend_name=glusterfs
```

- /etc/cinder/glusterfs에 아래와 같이 설정후 소유자,권한 수정

```
glusterfs1.example.com:/dist
```

HOST:/VOL_NAME

```
# chown root:cinder /etc/cinder/glusterfs
# chmod 0640 /etc/cinder/glusterfs
# systemctl restart openstack-cinder-api
# systemctl restart openstack-cinder-scheduler
# systemctl restart openstack-cinder-volume
#df –h시 gluster 볼륨이 자동마운트 됨
```

[실습] 11. 오픈스택 연동

- Backend로 잡아놓은 system cinder type 리스트로 등록하여 볼륨으로 사용할 수 있게 등록

```
[root@controller01 ~]# source keystonerc_admin
[root@controller01 ~(keystone_admin)]# cinder type-create glusterfs
…
 [root@controller01 ~(keystone_admin)]# cinder type-key glusterfs set
volume_backend_name=glusterfs
[root@controller01 ~(keystone_admin)]#  cinder type-list
+--------------------------------------+-----------+-------------+-----------+
| ID                                   | Name      | Description | Is_Public |
+--------------------------------------+-----------+-------------+-----------+
| b70d196e-151c-42e1-bd0b-d52c90c21e3c | glusterfs | -           | True      |
| bdd055bc-30e8-4ae0-a936-c65031657f6a | iscsi     | -           | True      |
| f6ca0d79-175b-4da1-8328-432bfc770808 | nfs       | -           | True      |
+--------------------------------------+-----------+-------------+-----------+
[root@controller01 ~(keystone_admin)]#  cinder  create --name test --volume-
type glusterfs 1
```