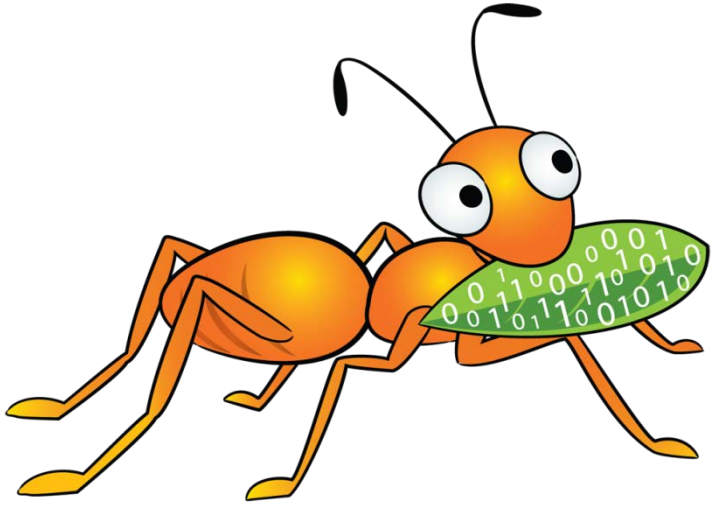
Software Defined Storage with Gluster



Vijay Bellur
Lead Gluster Architect
Red Hat
Twitter: @vbellur

redhat.

Agenda

- Software Defined Storage (SDS)
- Gluster as SDS - 4Ws and a H
  - Why Gluster?
  - What is Gluster?
  - How does Gluster work?
  - Where is Gluster used?
  - What next in Gluster?
- Q & A

What is Software Defined Storage?

redhat.

What is Software Defined Storage?

- All Storage Solutions have Software!
- General Consensus -
  - “Marketing buzzword” [1].
- Evolving terminology

[1] http://www.snia.org/sites/default/files/SNIA%20Software%20Defined%20Storage%20White%20Paper-%20v1.0k-DRAFT.pdf

SDS - Characteristics

- Runs on Commodity Hardware
- Scale-out with resource aggregation
- Elasticity
- Automated Management
- Various data services
- Storage as a Platform & Storage as a Service

SDS with Gluster

Why Gluster?

Why Gluster?

- **2.5+ exabytes** of data produced every day!
- 90% of data in last two years
- Data needs to be stored somewhere
- Often data needs to outlive us!

source: http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

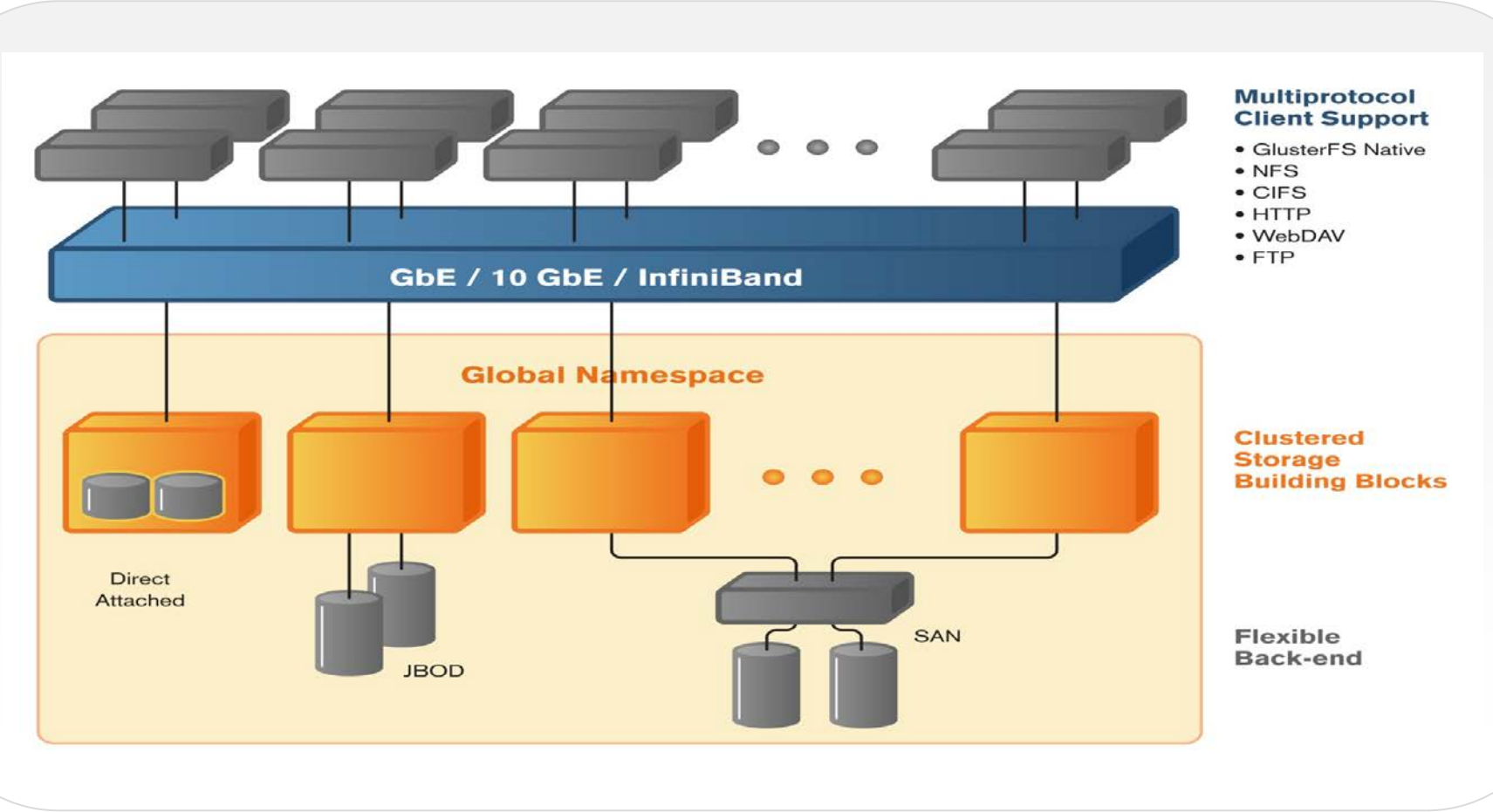
What is Gluster?

What is Gluster?

- Open Scale-out distributed storage system.
- Aggregates storage exports over network interconnects to provide an unified namespace.
- Layered on disk file systems that support extended attributes.
- Provides file, object and block interfaces for data access.

How does Gluster work?

Typical Gluster Deployment



redhat.

Gluster Architecture – Foundations

- Software only, runs on commodity hardware
- Scale-out with Elasticity
- Extensible and modular
- Deployment agnostic
- No external metadata servers

Gluster Volumes

- Logical collection of exports (bricks) from various servers
- Default entity for storage policy definition
- Volume or a part of the volume can be accessed by clients

Volume Types

Volume type determines:

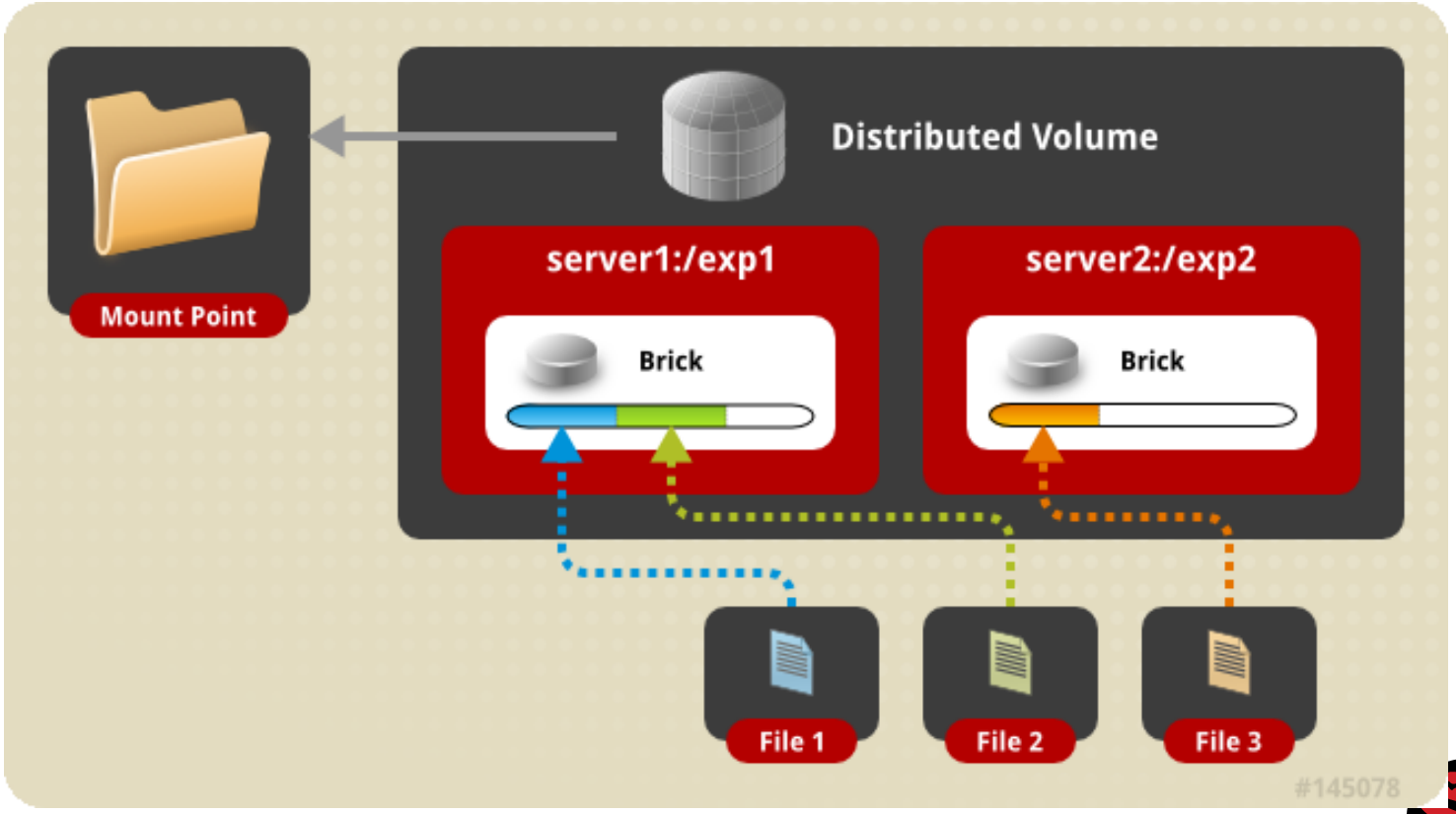
- Data Placement
- Redundancy
- More..

Available volume types:

- Distribute
- Striped
- Replicated
- Distributed Replicate
- Striped Replicate
- Distributed Striped Replicate
- Dispersed

Distributed Volume

- Distributes files across various bricks of the volume.
- Directories are present on all bricks of the volume.
- Removes the need for an external meta data server, provides O(1) lookup.

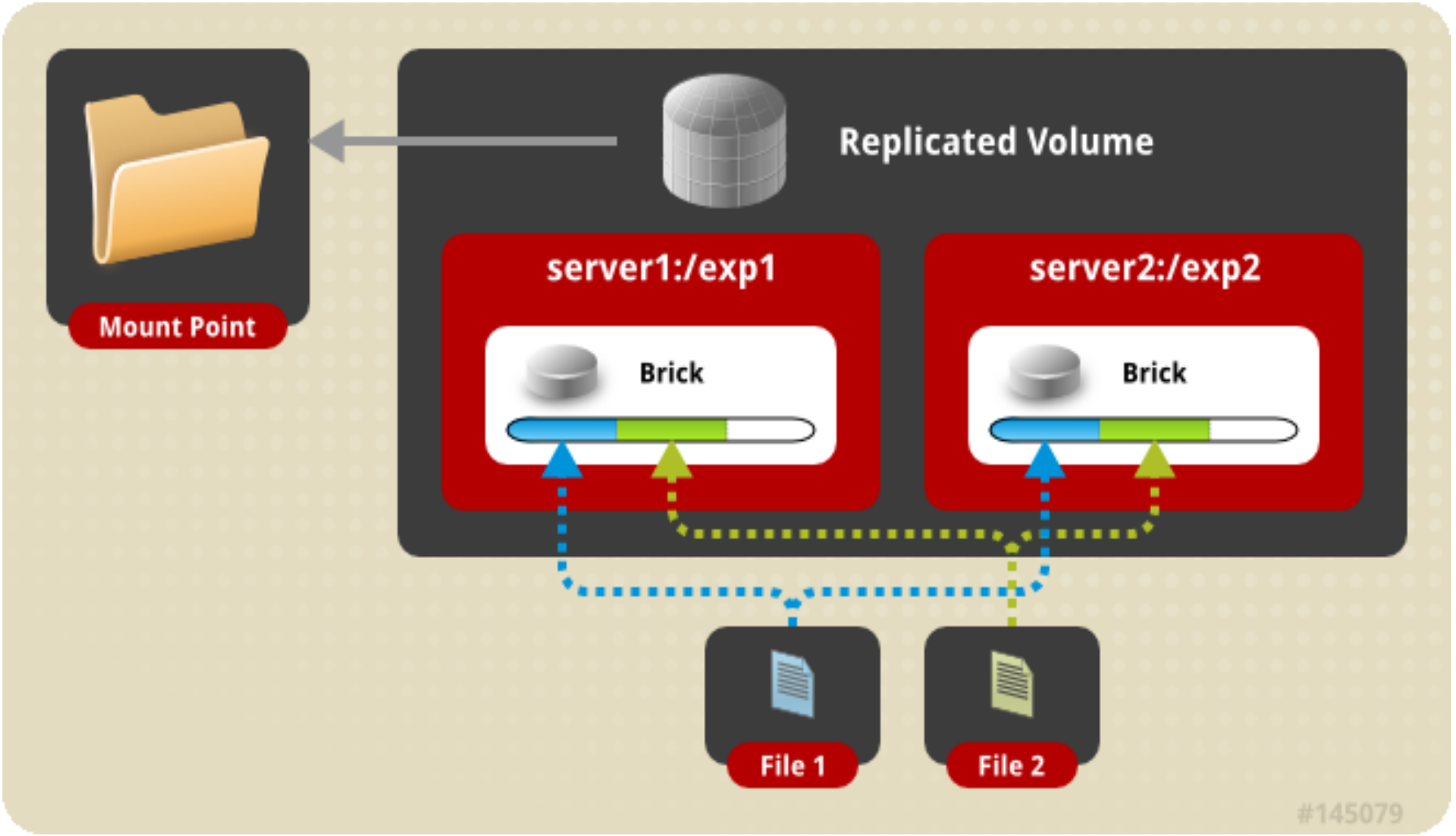


redhat.

Replicated Volume

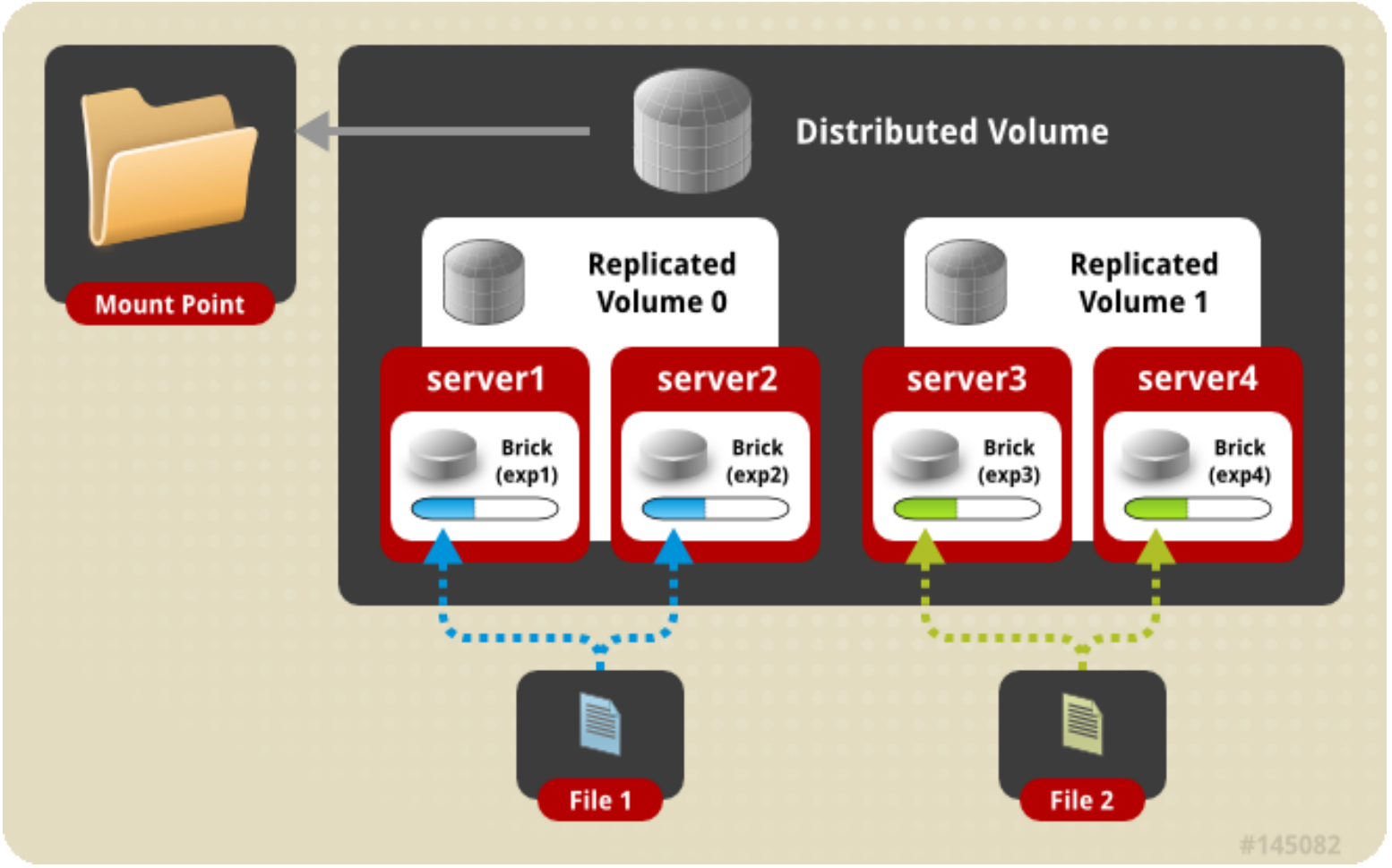
- Synchronous replication of all updates.
- Provides HA for data.
- Transaction driven for ensuring consistency.
- Changelogs maintained for re-conciliation.
- Any number of replicas can be configured.

How does a replicated volume work?



redhat.

Distributed Replicated Volume



redhat.

Dispersed Volume

- Erasure Coding / RAID 5 over the network
- “Disperses” data on to various bricks
- Algorithm: Reed solomon
- Non–systematic erasure coding

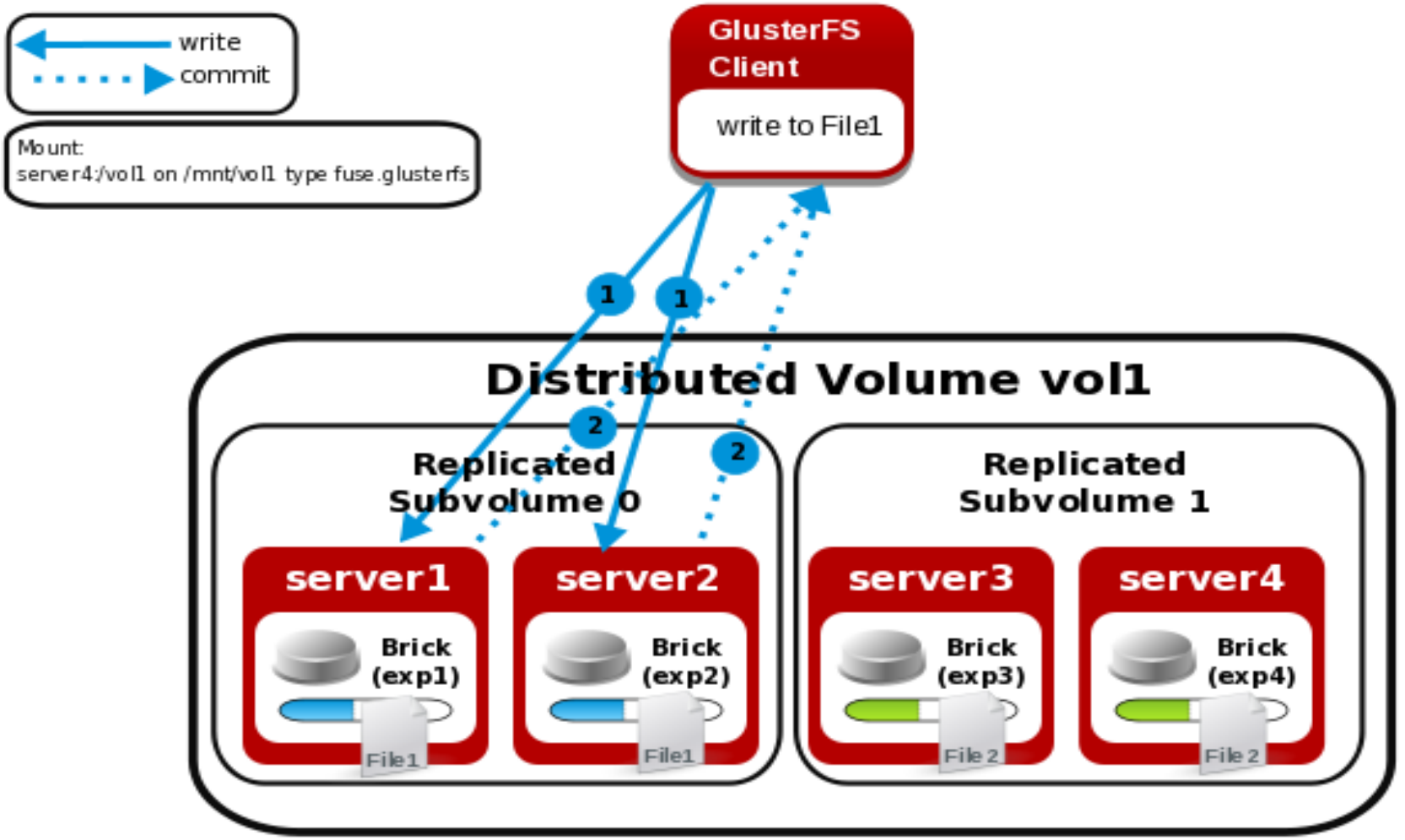
Access Mechanisms

Access Mechanisms

Gluster volumes can be accessed via the following mechanisms:

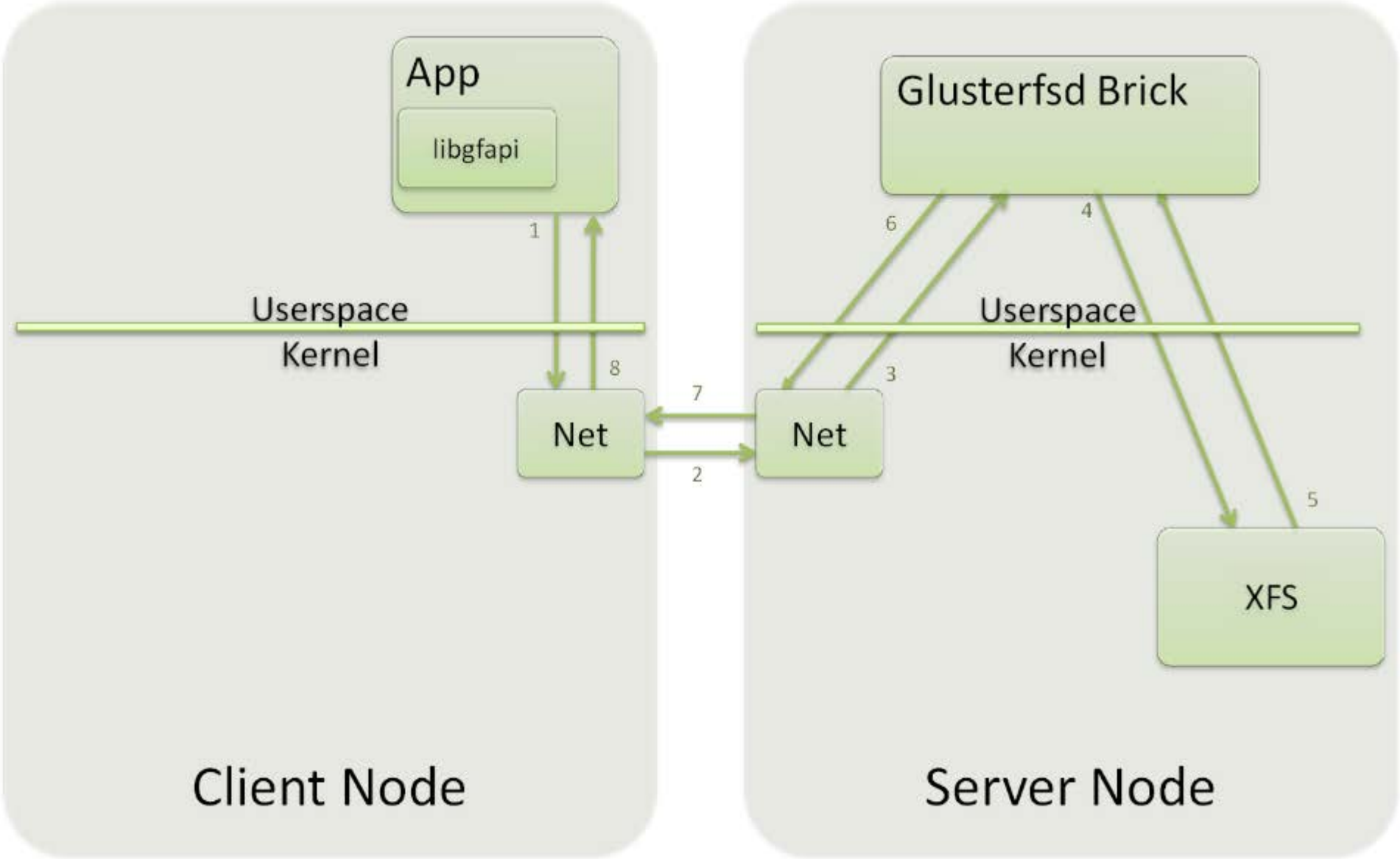
- FUSE based Native protocol
- NFSv3
- SMB
- libgfapi
- ReST/HTTP (object)
- HDFS
- iSCSI (block)

FUSE based native access

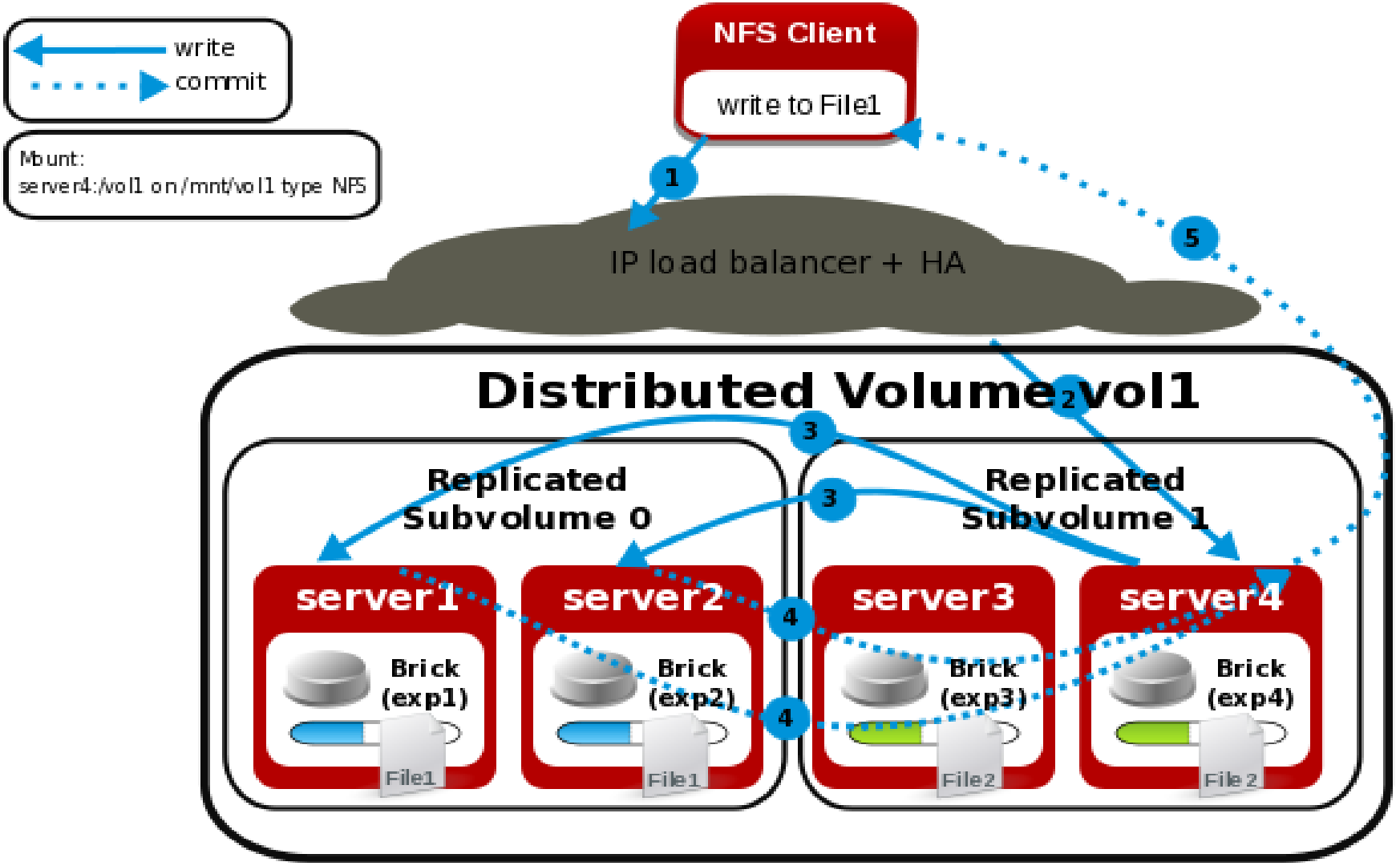


redhat.

libgfapi access

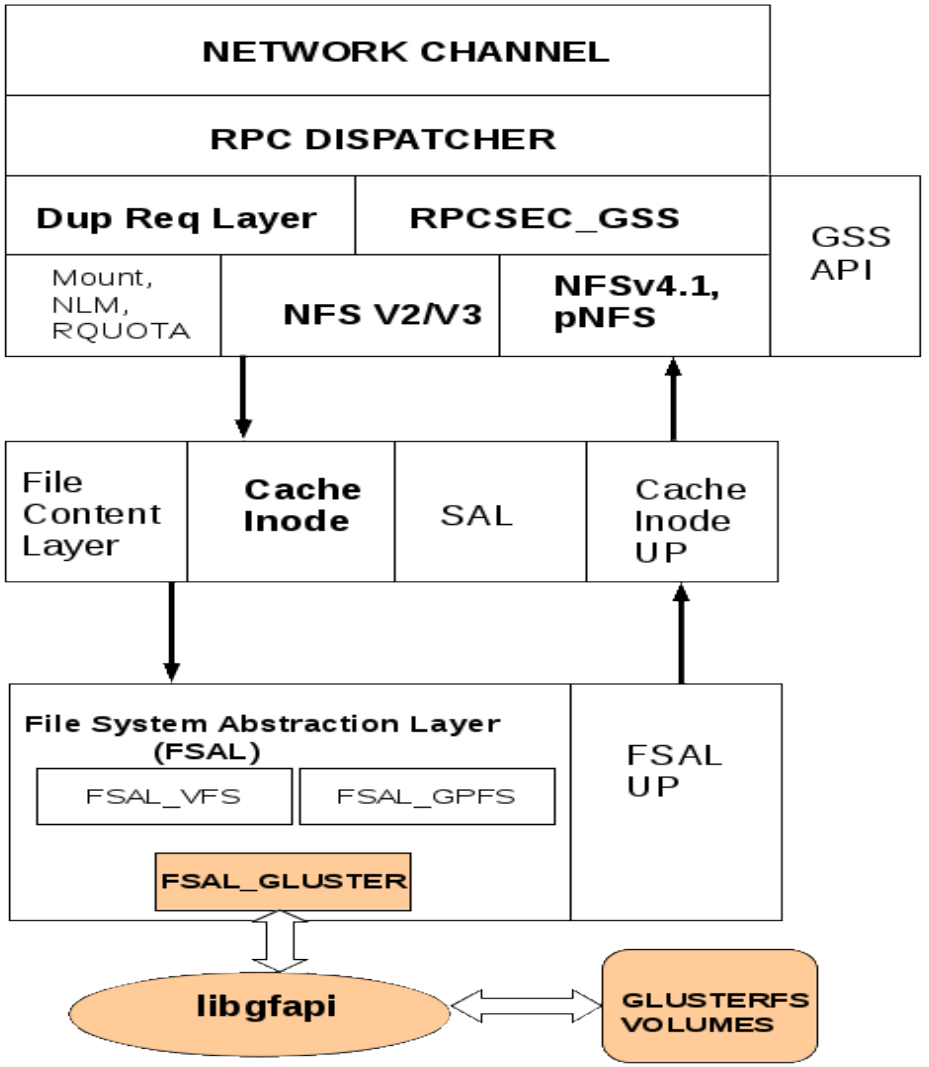


NFSv3 access with Gluster NFS



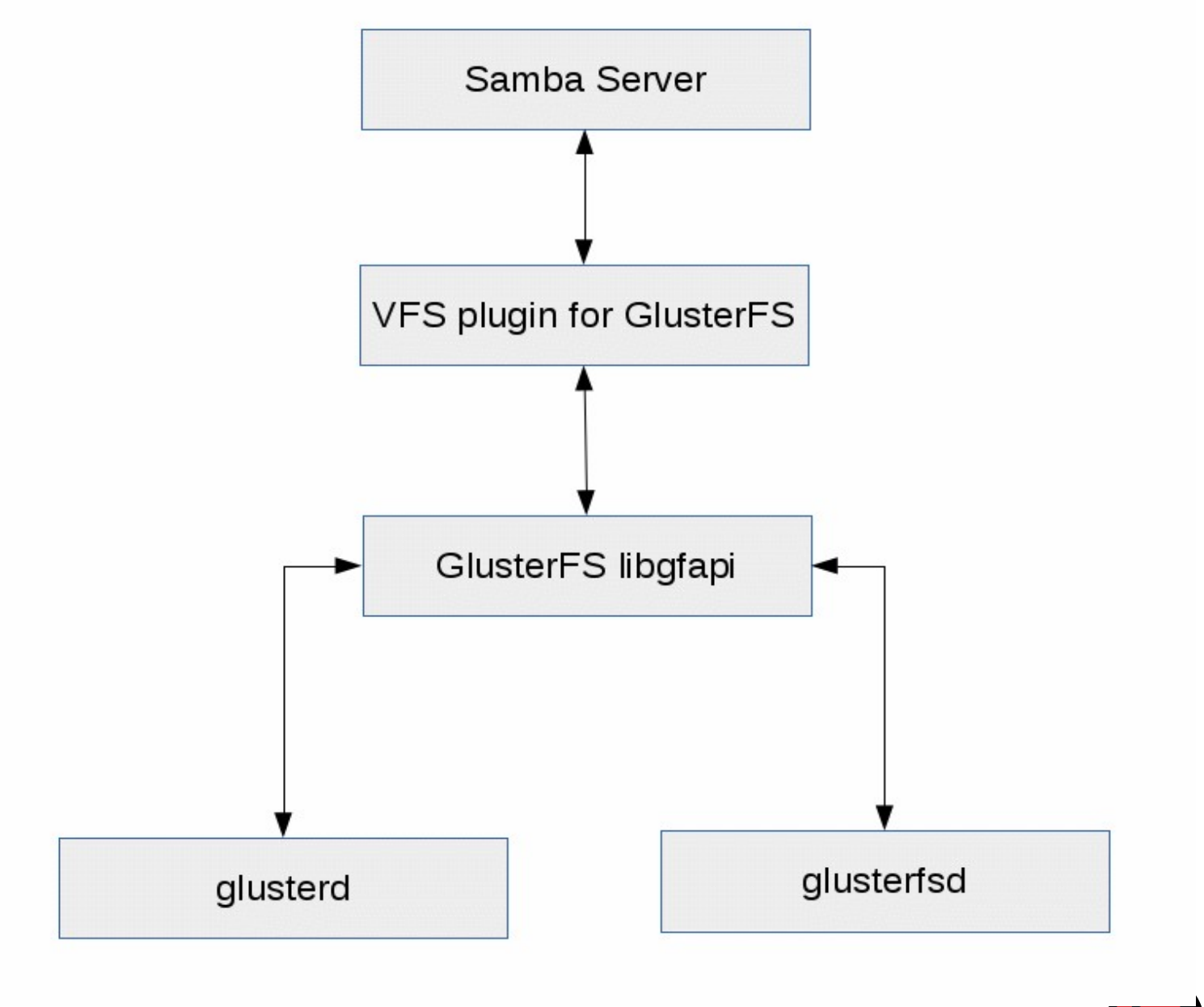
redhat.

Nfs-Ganesha with Gluster



redhat.

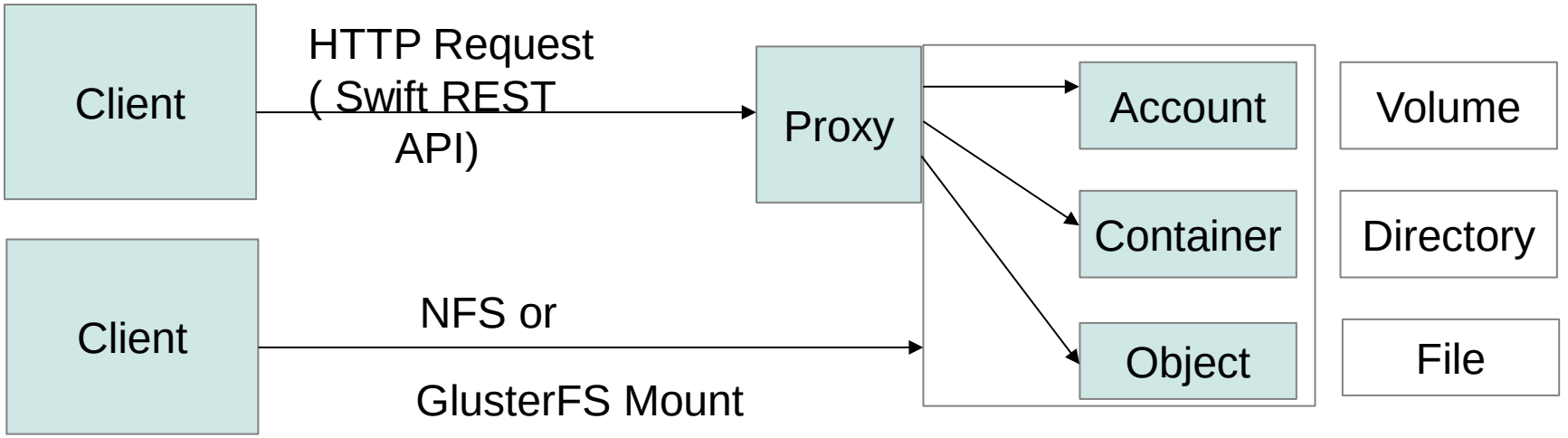
SMB with Gluster



redhat.

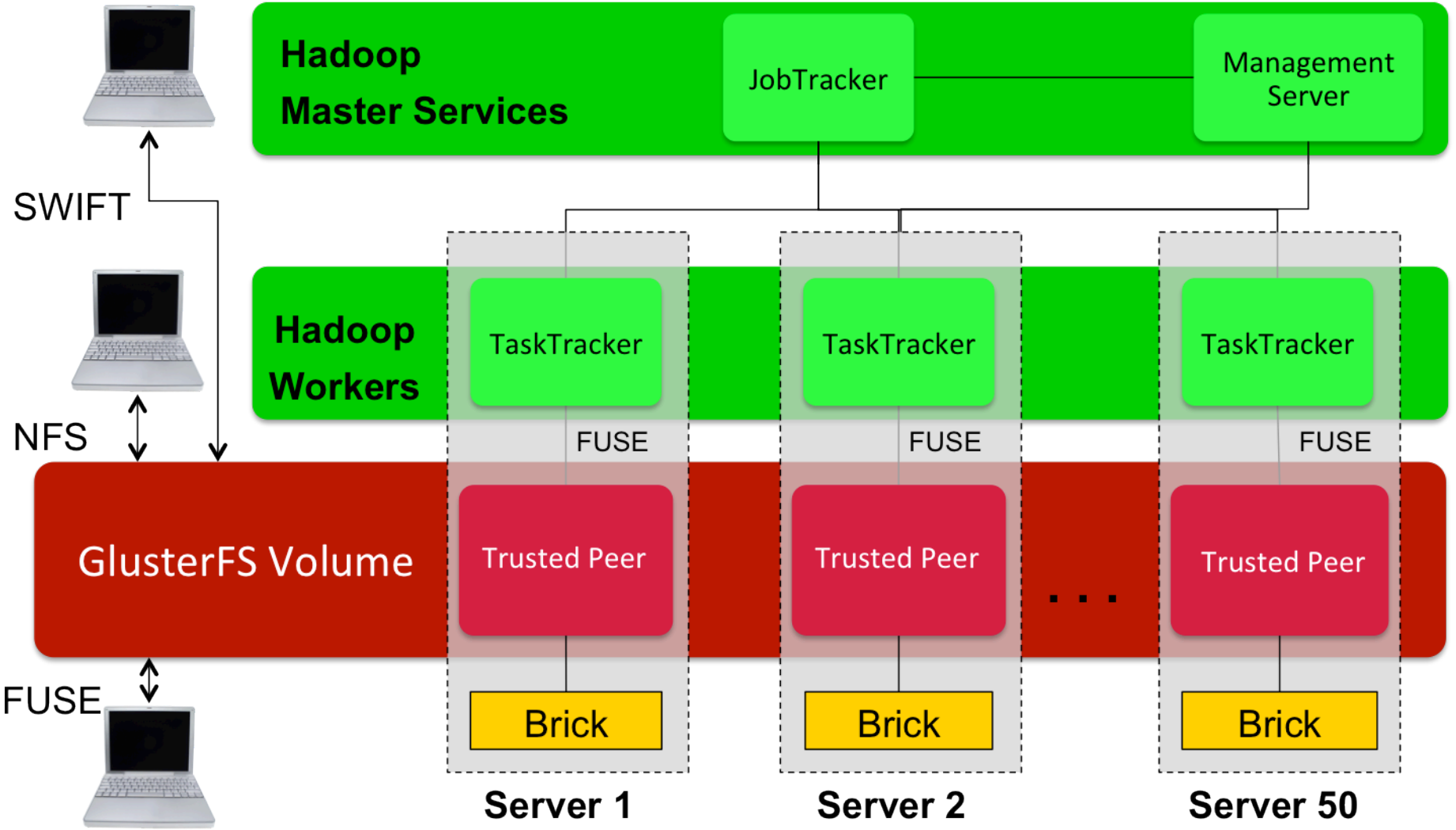
Object/ReST - SwiftonFile

- Unified File and object view
- Entity mapping between file and object building blocks



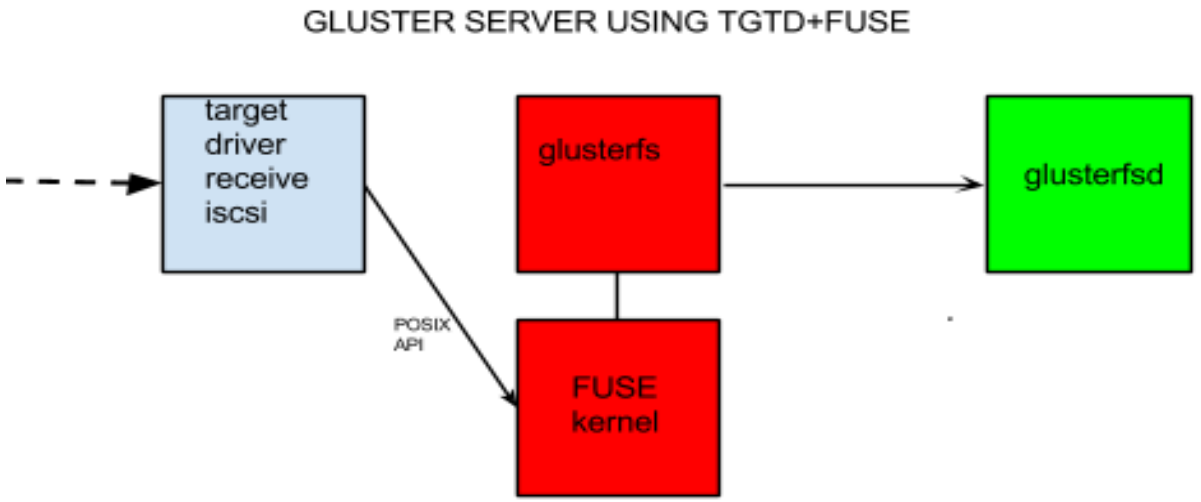
redhat.

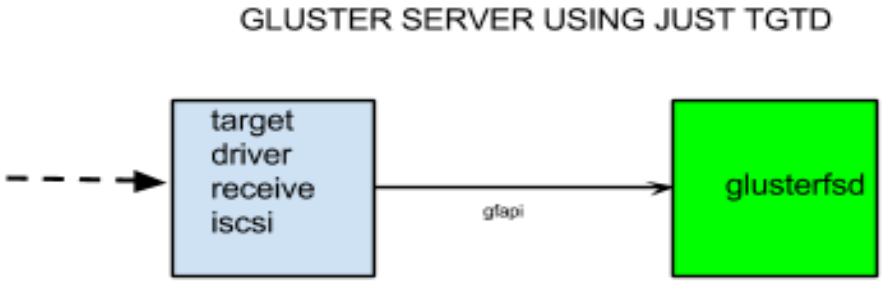
HDFS access



redhat.

Block/iSCSi access





redhat.

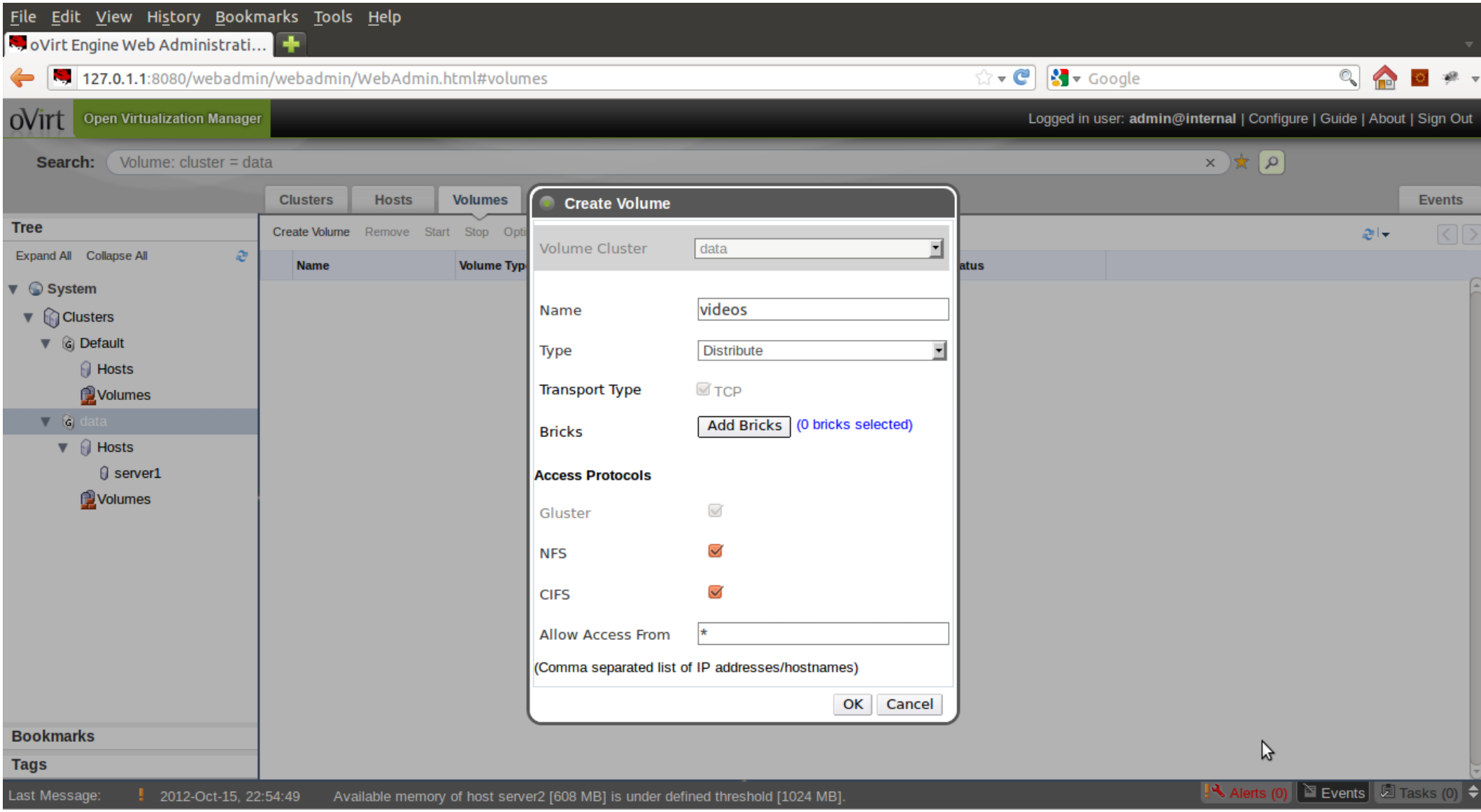
Gluster Features

- Scale-out NAS
  - Elasticity
  - Directory, Volume & Inode quotas
- Data Protection and Recovery
  - Volume and File Snapshots
  - User Serviceable Snapshots
  - Geographic/Asynchronous replication
- Archival
  - Read-only
  - WORM

Gluster Features

- Isolation for multi-tenancy
  - SSL based data transfer
  - Encryption at rest
- Performance
  - Client side in memory caching for performance
  - Data, metadata and readdir caching
- Monitoring
  - Built in io statistics
  - /proc like interface for introspection
- Provisioning
  - puppet-gluster
  - gluster-deploy
- More..

Web based Management - oVirt



redhat.

Gluster Monitoring with Nagios

	Disk Utilization		OK	06-25-2014 06:05:07	1d 1h 21m 55s	1/3	OK : 60.0% used (6.0GB out of 10.0GB)
	Gluster Management		OK	06-25-2014 05:03:29	1d 1h 21m 26s	1/3	Process glusterd is running
	Memory Utilization		OK	06-25-2014 06:05:07	0d 0h 59m 30s	1/3	OK- 70.43% used(0.68GB out of 0.97GB)
	NFS		OK	06-25-2014 05:08:37	1d 1h 21m 26s	1/3	Process glusterfs-nfs is running
	Network Utilization		OK	06-25-2014 06:05:07	0d 6h 3m 26s	1/3	OK: eth0:UP
	Quota		OK	06-25-2014 05:13:46	1d 1h 21m 26s	1/3	OK: Quota not enabled
	SMB		CRITICAL	06-25-2014 04:48:07	1d 1h 21m 26s	3/3	CRITICAL: Process smb is not running
	Self-Heal		OK	06-25-2014 04:48:55	1d 1h 21m 26s	1/3	Gluster Self Heal Daemon is running
	Swap Utilization		OK	06-25-2014 06:05:07	0d 6h 3m 26s	1/3	OK- 5.15% used(0.05GB out of 1.00GB)
dhcp43-97.lab.eng.blr.redhat.com	Brick - /bricks/b1		OK	06-25-2014 06:04:24	1d 1h 21m 9s	1/3	OK: Brick /bricks/b1 is up
	Brick - /bricks/b2		OK	06-25-2014 06:04:24	1d 1h 21m 9s	1/3	OK: Brick /bricks/b2 is up
	Brick Utilization - /bricks/b1		OK	06-25-2014 06:04:24	1d 1h 21m 9s	1/3	OK : 53.0% used (5.0GB out of 9.0GB)
	Brick Utilization - /bricks/b2		OK	06-25-2014 06:04:24	1d 1h 21m 9s	1/3	OK : 53.0% used (5.0GB out of 9.0GB)
	CTDB		OK	06-25-2014 05:04:20	1d 1h 21m 9s	1/3	CTDB ignored as SMB and NFS are not running
	Cpu Utilization		OK	06-25-2014 06:04:55	0d 0h 13m 9s	1/3	CPU Status OK: Total CPU:4.73% Idle CPU:95.27%
	Disk Utilization		OK	06-25-2014 06:04:24	1d 1h 21m 9s	1/3	OK : 60.0% used (6.0GB out of 10.0GB)
	Gluster Management		OK	06-25-2014 05:12:03	1d 1h 21m 9s	1/3	Process glusterd is running
	Memory Utilization		OK	06-25-2014 06:04:37	0d 1h 20m 9s	1/3	OK- 79.04% used(0.77GB out of 0.97GB)
	NFS		CRITICAL	06-25-2014 04:47:12	1d 1h 22m 41s	3/3	CRITICAL: Process glusterfs-nfs is not running
	Network Utilization		OK	06-25-2014 06:04:46	0d 6h 3m 9s	1/3	OK: eth0:UP
	Quota		OK	06-25-2014 04:52:20	1d 1h 21m 9s	1/3	OK: Quota not enabled
	SMB		CRITICAL	06-25-2014 04:54:55	1d 1h 21m 9s	3/3	CRITICAL: Process smb is not running
	Self-Heal		OK	06-25-2014 04:57:29	1d 1h 21m 9s	1/3	Gluster Self Heal Daemon is running
	Swap Utilization		OK	06-25-2014 06:05:03	0d 6h 3m 30s	1/3	OK- 1.82% used(0.02GB out of 1.00GB)
test-cluster	Cluster - Quorum	?	PENDING	N/A	1d 1h 22m 38s+	1/3	Service is not scheduled to be checked...
	Cluster Auto Config		OK	06-25-2014 05:02:37	1d 1h 20m 56s	1/3	Cluster configurations are in sync
	Cluster Utilization		OK	06-25-2014 06:05:12	1d 1h 18m 56s	1/3	OK - used 58% of available 16.4980621338 GB
	Volume Self-Heal - rep-vol		OK	06-25-2014 06:04:37	1d 1h 20m 56s	1/3	No unsynced entries present
	Volume Status - rep-vol		OK	06-25-2014 06:04:37	1d 1h 20m 56s	1/3	OK: Volume : DISTRIBUTED_REPLICATE type - All bricks are Up
	Volume Utilization - rep-vol		OK	06-25-2014 06:04:37	1d 1h 20m 56s	1/3	OK: Utilization:58.42%

http://www.ovirt.org/Features/Nagios_Integration

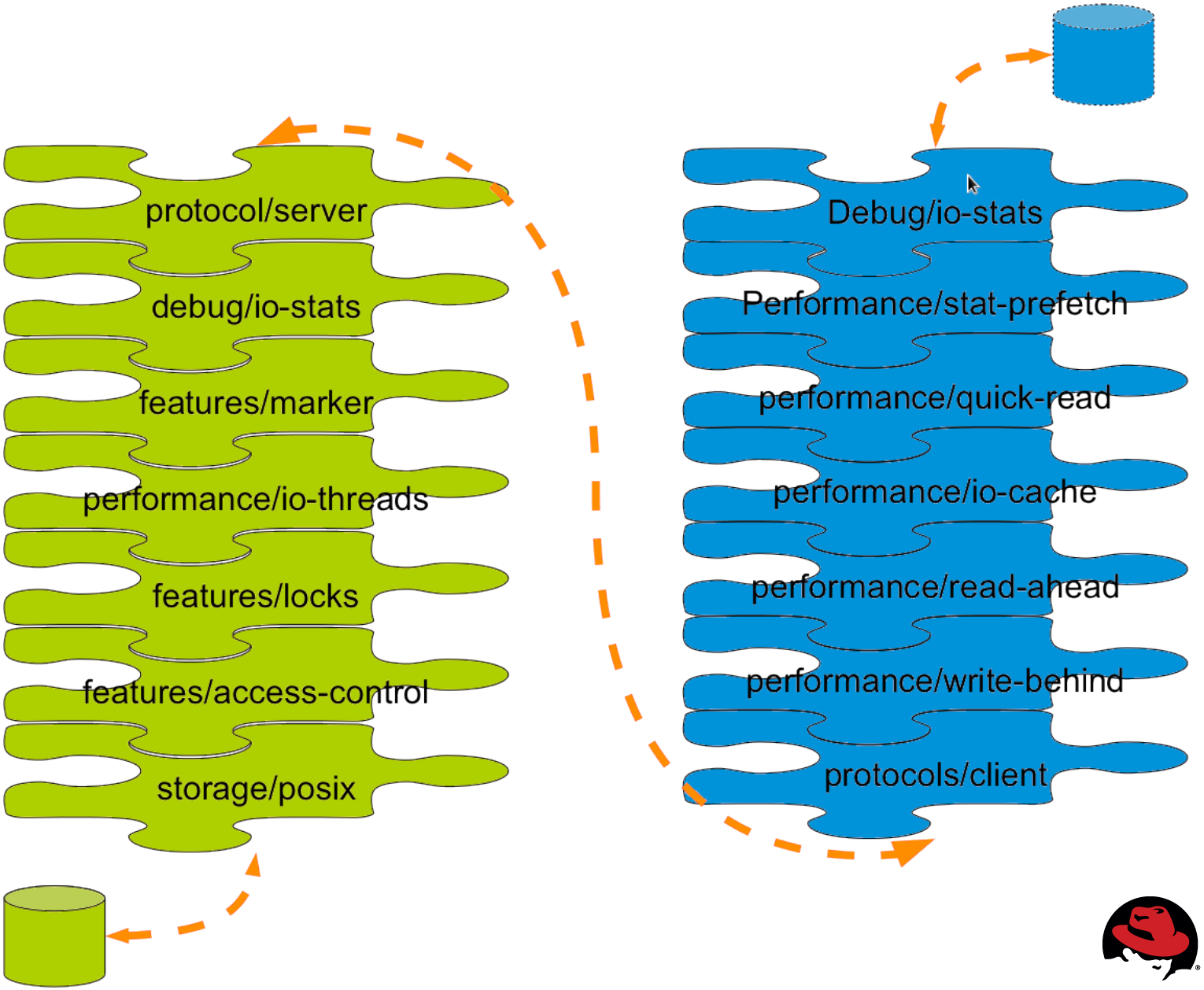
redhat.

How is it implemented?

Translators in Gluster

- Building blocks for a Gluster process.
- Based on Translators in GNU HURD.
- Each translator is a functional unit.
- Translators can be stacked together for achieving desired functionality.
- Translators are deployment agnostic – can be loaded in either the client or server stacks.

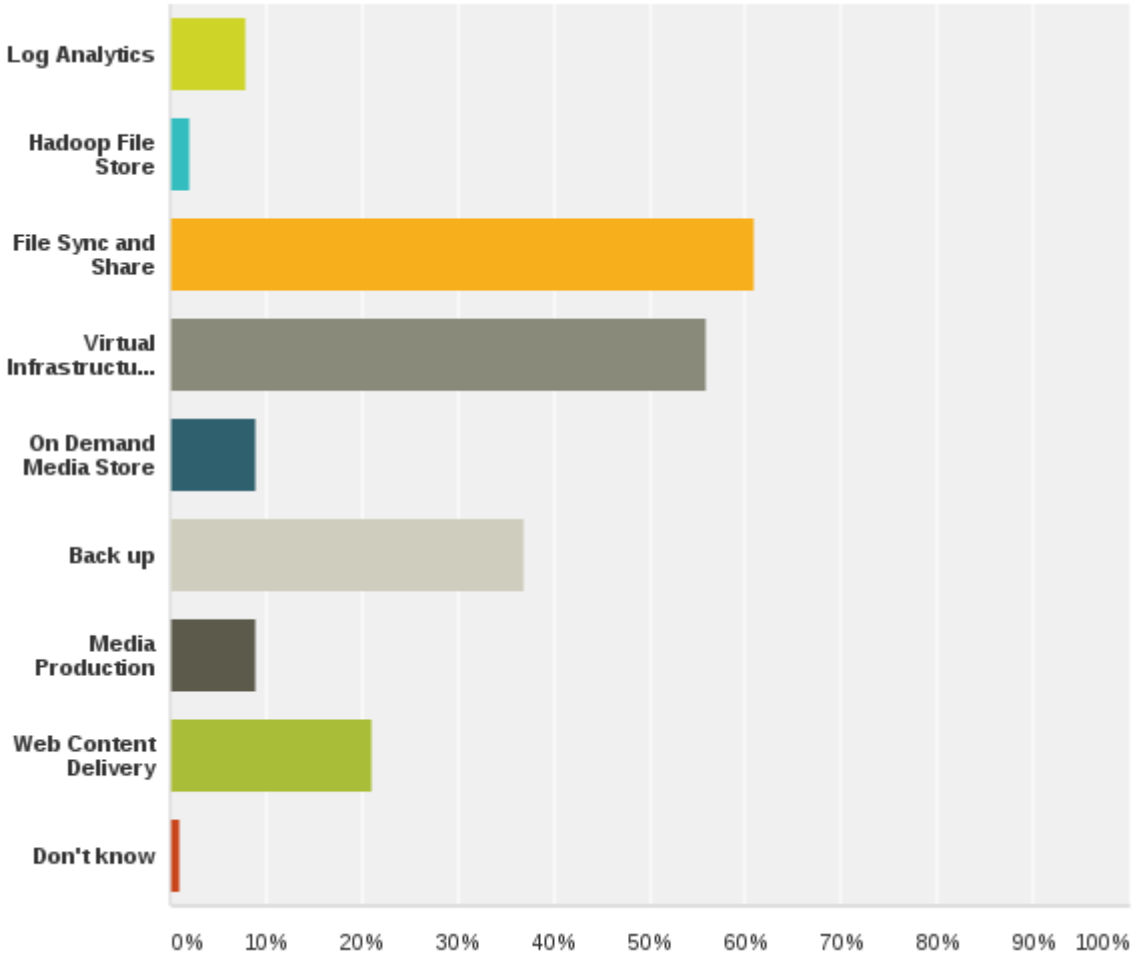
Customizable Translator Stack



redhat.

Where is Gluster used?

Gluster Use Cases



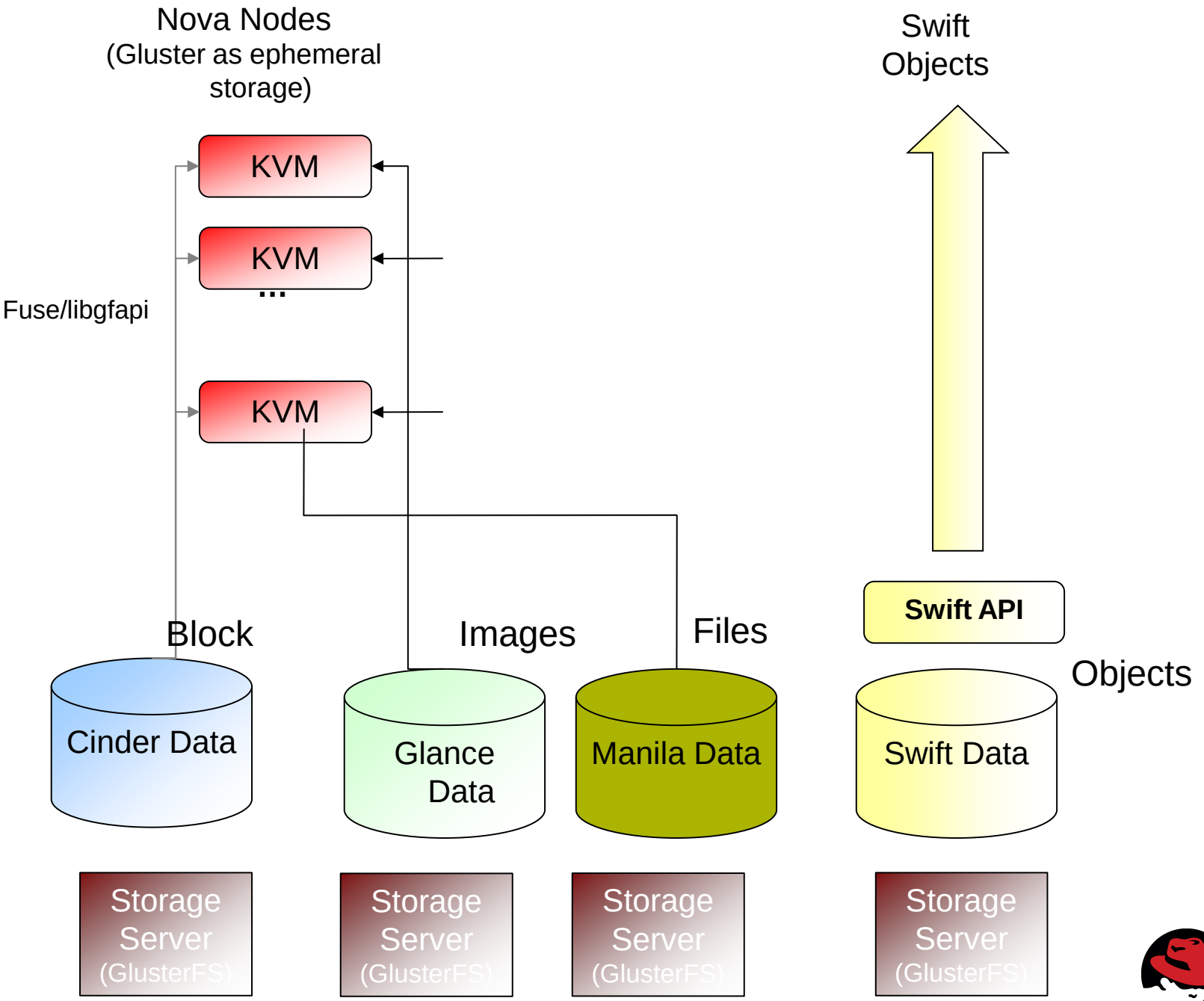
Source: 2014
GlusterFS user survey

redhat.

Ecosystem Integration

- Currently used with various ecosystem projects
  - Virtualization
    - OpenStack
    - oVirt
    - Qemu
    - CloudStack
  - Big Data Analytics
    - Hadoop
    - Tachyon
  - File Sync and Share
    - ownCloud

Openstack Kilo + Gluster – Current Integration



redhat.

What next in Gluster?

New Features in GlusterFS 3.7

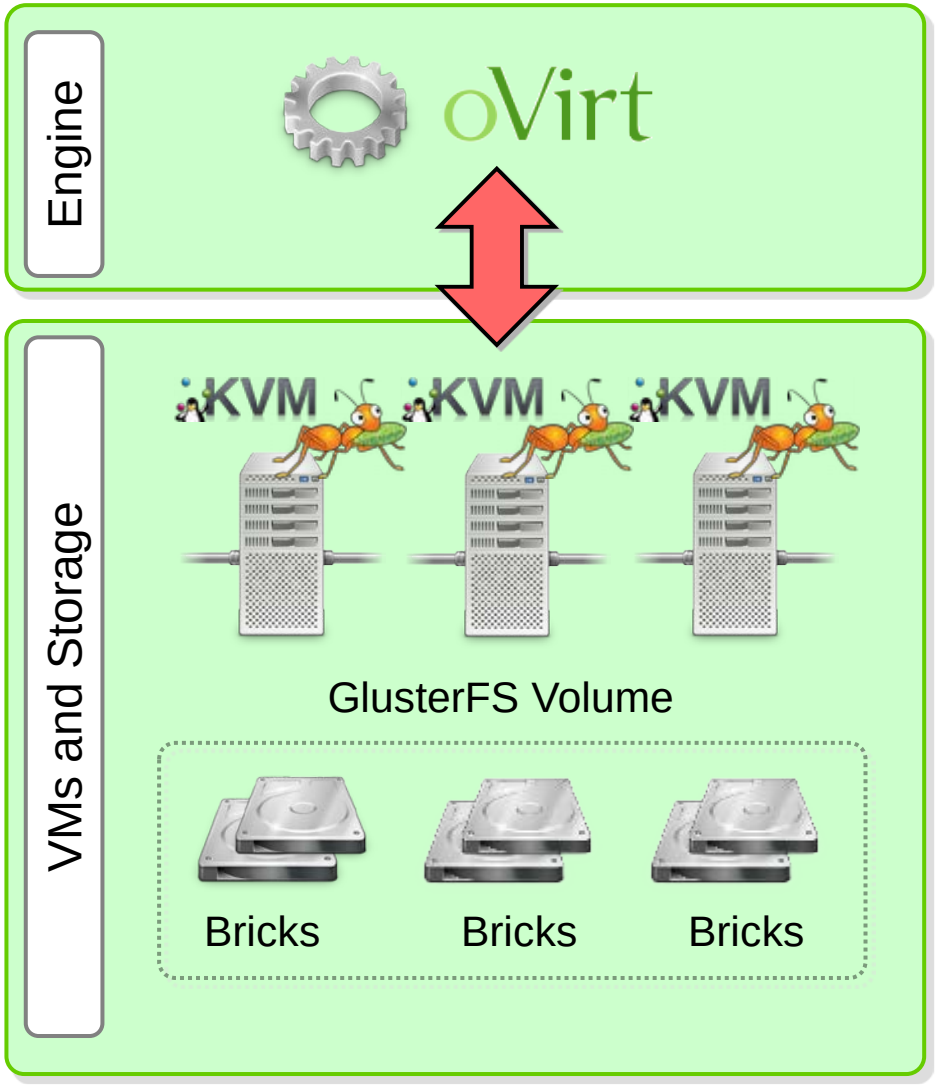
- Data Tiering
- Bitrot detection
- Sharding
- NFSv4, 4.1 and pNFS access using NFS Ganesha
- Netgroups style configuration for NFS
- Performance improvements

Features beyond GlusterFS 3.7

- HyperConvergence with oVirt
- Compression at rest
- De-duplication
- Multi-protocol support with NFS, FUSE and SMB
- Native ReST APIs for gluster management
- More integration with OpenStack, Containers

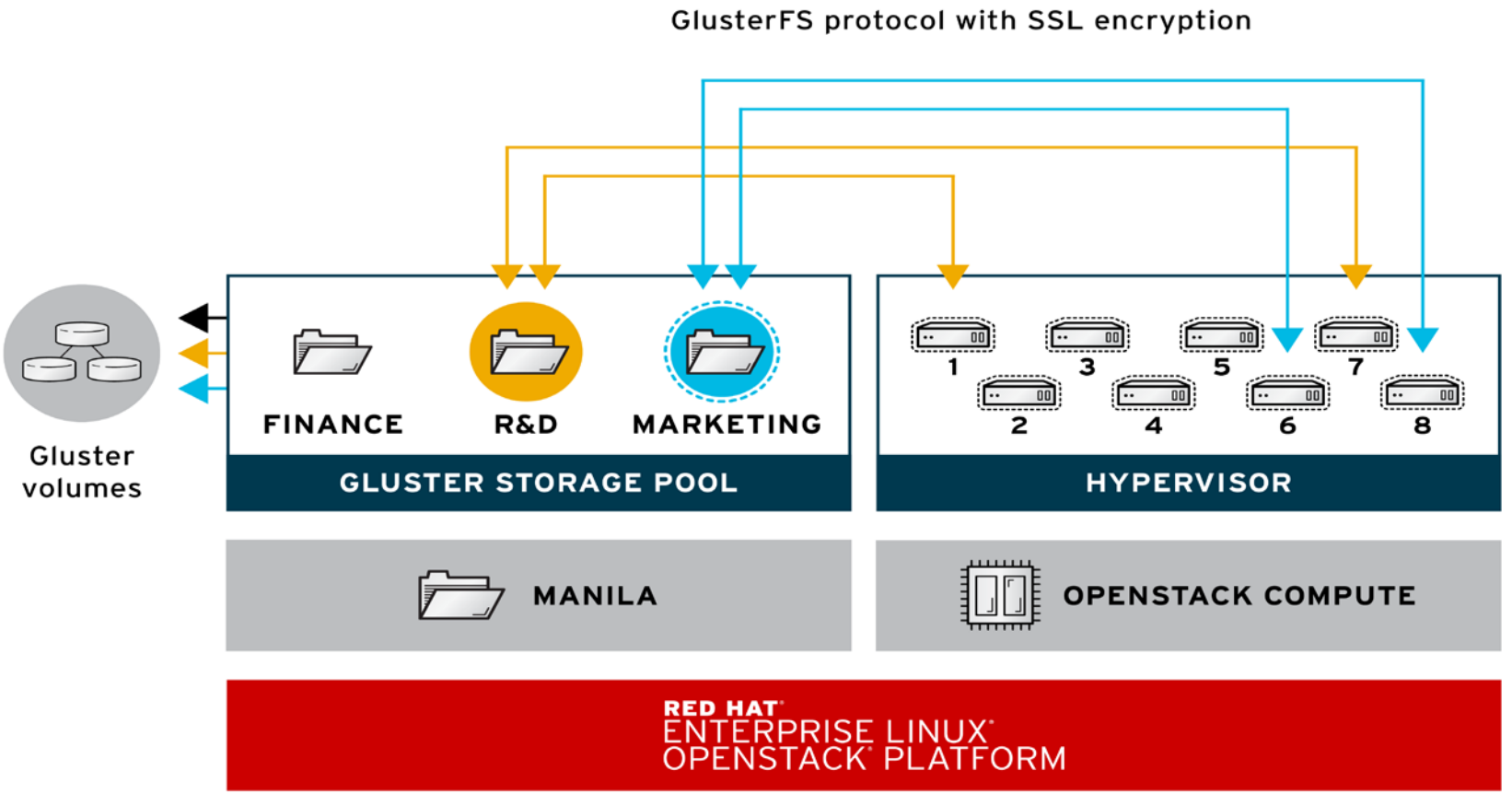
Hyperconverged oVirt – Gluster

- Server nodes are used both for virtualization and serving replicated images from Gluster Volumes
- Support for both scaling up, adding more disks, and scaling out, adding more hosts



redhat.

Gluster Native Driver – OpenStack Manila



redhat

Gluster 4.0

- Not Evolutionary anymore
- Intended for massive scalability and manageability improvements, remove known bottlenecks
- Make it easier for devops to provison, manage and monitor
- Enable larger deployments and new use cases

Gluster 4.0

- New Style Replication
- Improved Distributed hashing Translator
- Composite operations in the GlusterFS RPC protocol
- Support for multiple networks
- Coherent client-side caching
- Advanced data tiering
- ... and much more

Resources

Mailing lists:

gluster-users@gluster.org

gluster-devel@gluster.org

IRC:

#gluster and #gluster-dev on freenode

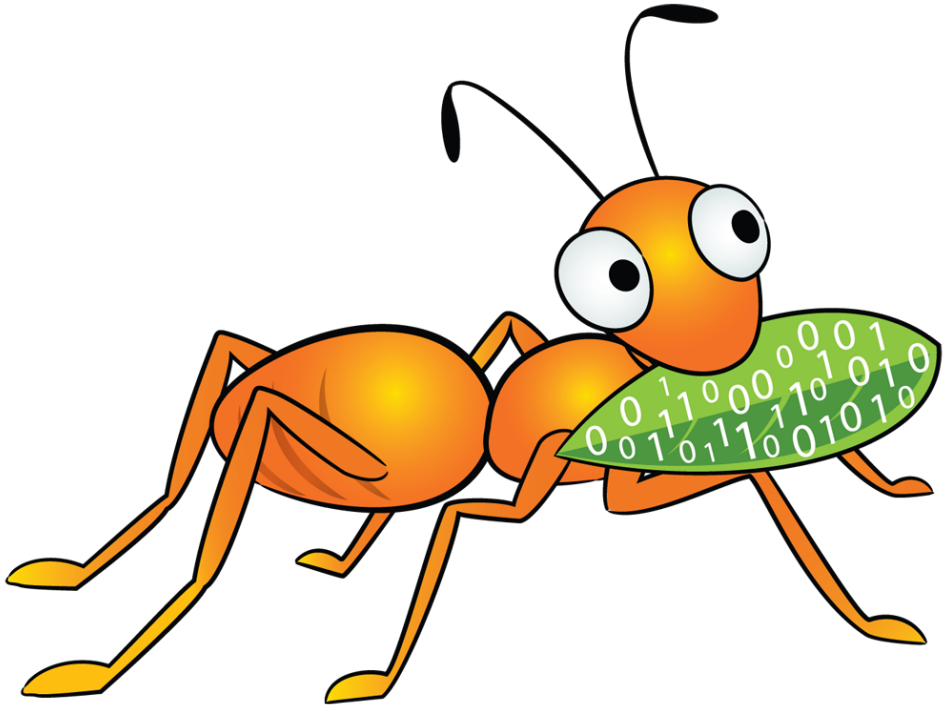
Web:

http://www.gluster.org

Conclusions

- SDS evolution on the way
- Data explosion unbounded
- Great time to be in Storage!

Thank you!



redhat.