Rethinking Gluster Management using k8s

Aravinda VK

Principal Software Engineer, Red Hat

Co-maintainer for many of the Gluster components. Like Geo-replication, EventsAPI, Glusterfind, Changelog, etc.

•••

Amar Tumballi

Founding engineering member of Gluster project.

Consultant, Maintainer @ Gluster.org

Gluster Meetup - 25th Sep, 2019

How many of you know

- Kubernetes ?
- CSI ?
- GlusterFS ?

GlusterFS

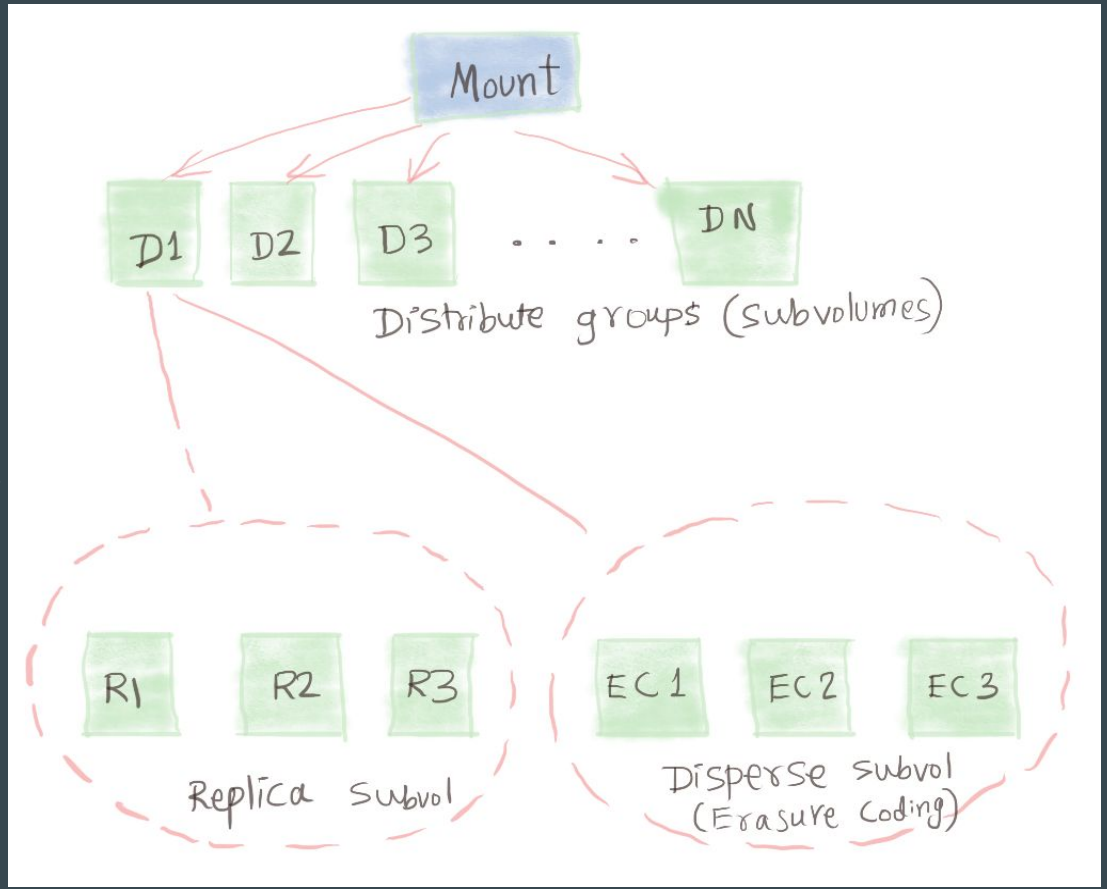
Free and open source,

Scalable,

Network filesystem.

Software defined storage (SDS).

Started in 2006.

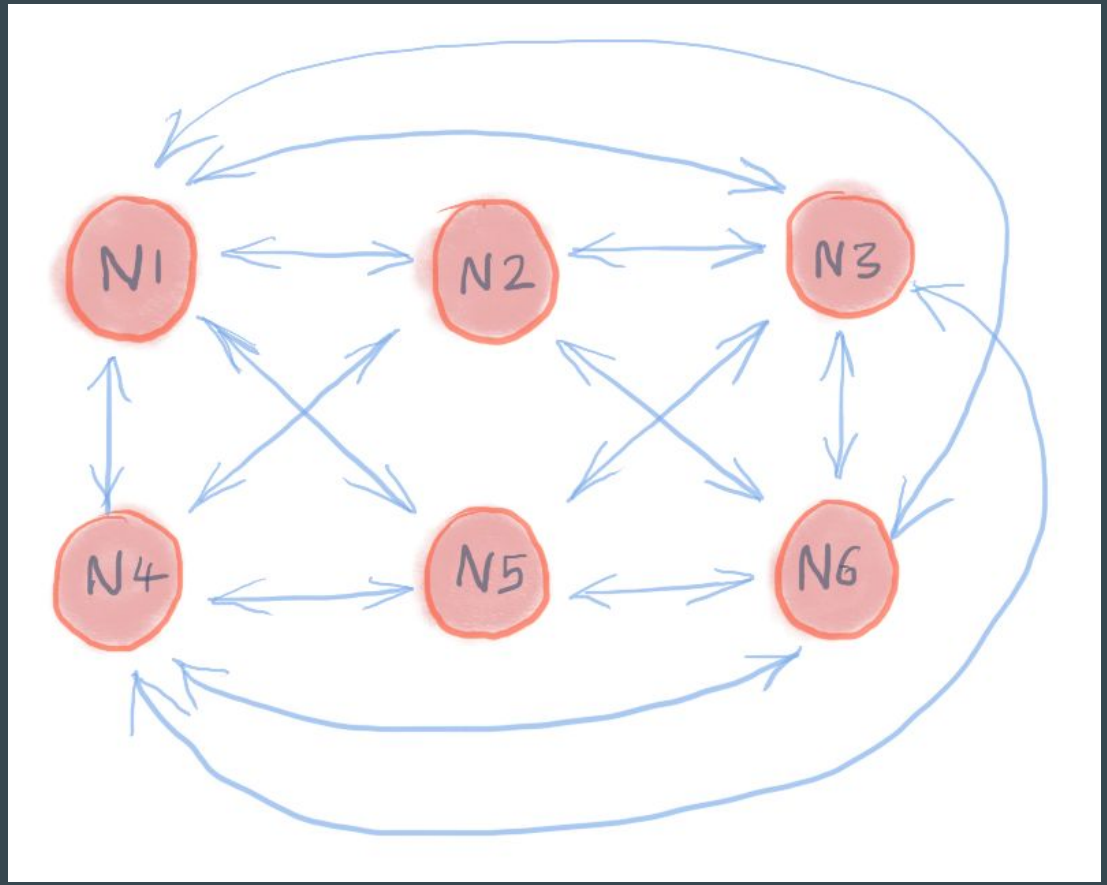


GlusterD

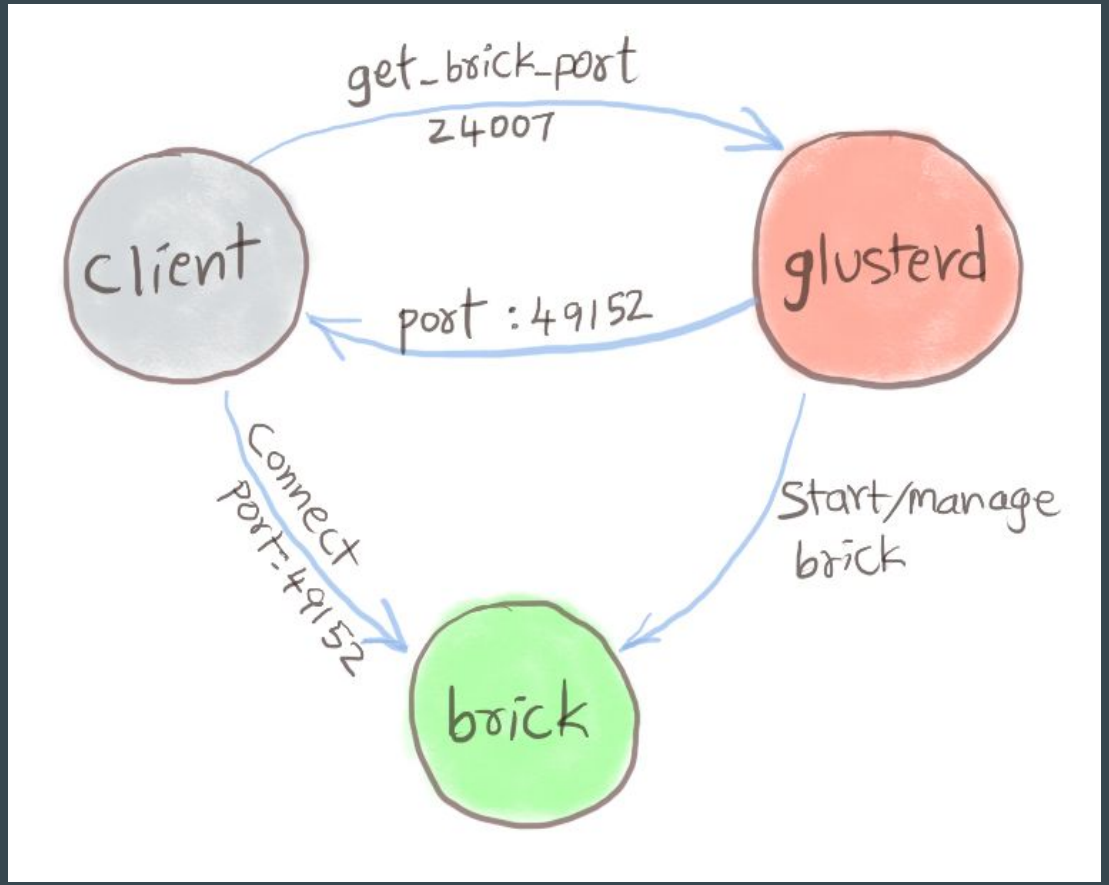
Management daemon for GlusterFS

- Clustering
- Volume Management
- Brick processes management
- Portmap for bricks
- Manages the services(Brick, Self-heal, Rebalance etc)
- Volfile for Bricks , Self heal etc
- Managing Quota, Snapshot, Geo-replication and others

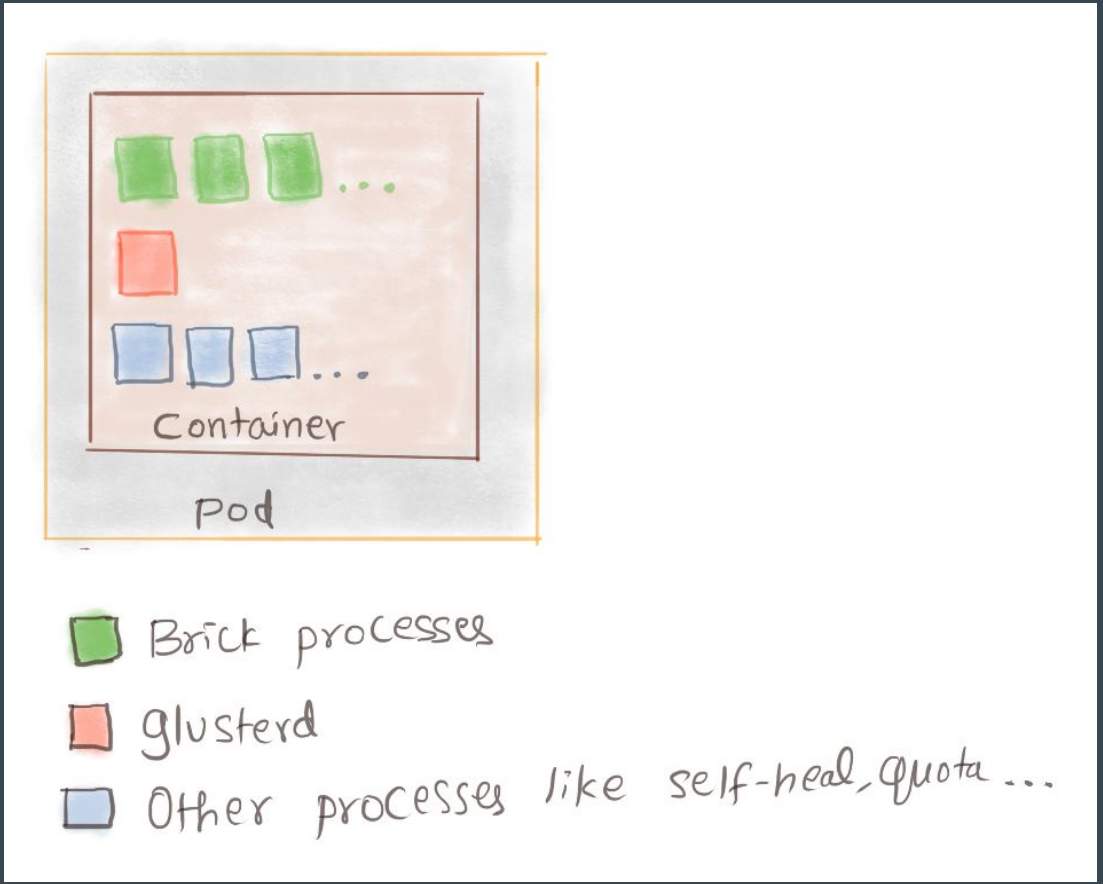
Cluster



Ports


get_brick_port
24007
client
glusterd
port : 49152
Connect
port-49152
Start/manage
brick
brick

Processes in a Container



Why change when it is working?

- Too many layers, hard to debug.
- Duplication of task is bad, and can cause in-consistency.
- K8s can provide added infrastructure like process management, cluster authentication, monitoring and centralized logging.
- Running more than one process in a container defeats the purpose of microservices.
- Currently, no ideal solution with Gluster for storage in k8s :-)

Kubernetes Operator is new Glusterd!

Introducing KaDalu project

- KaDalu in Kannada language means Ocean
- https://github.com/kadalu/kadalu
- Easy to install on an already running kubernetes cluster using Kadalu Operator.
- Best way to export your existing storage arrays into k8s ecosystem.
- Two steps to get storage setup complete.

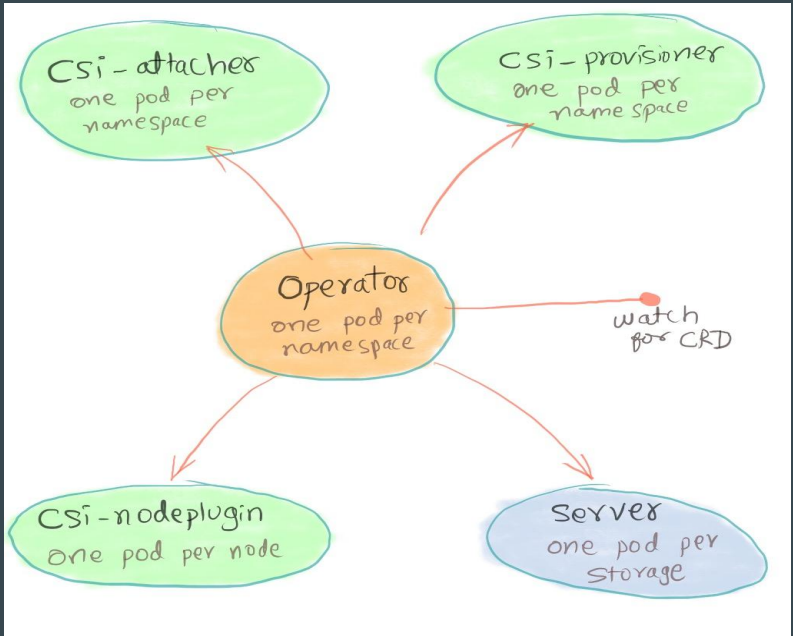
```
$ kubectl create -f kadalu-operator.yaml #<- no change required for this file
$ kubectl create -f kadalu-storage.yaml  #<- add your storage type and host:device details
      here
```

Demo

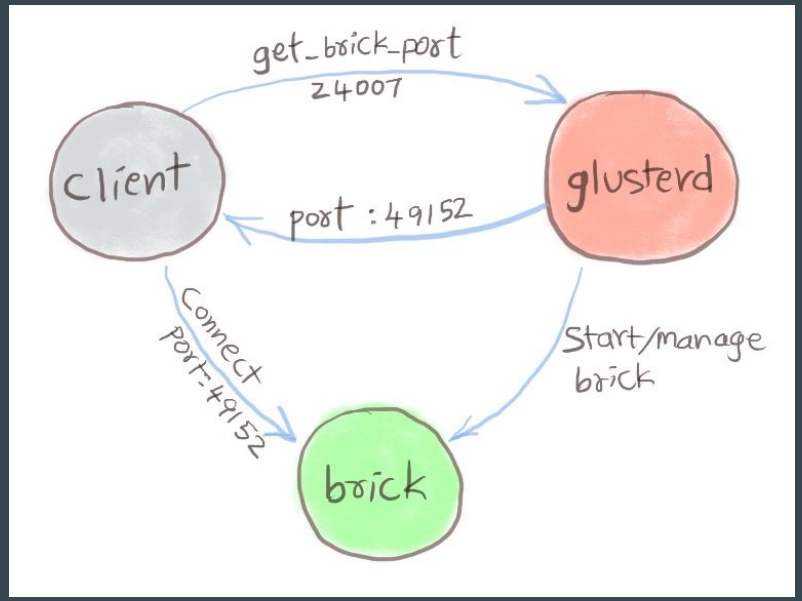
https://asciinema.org/a/259949

Architecture

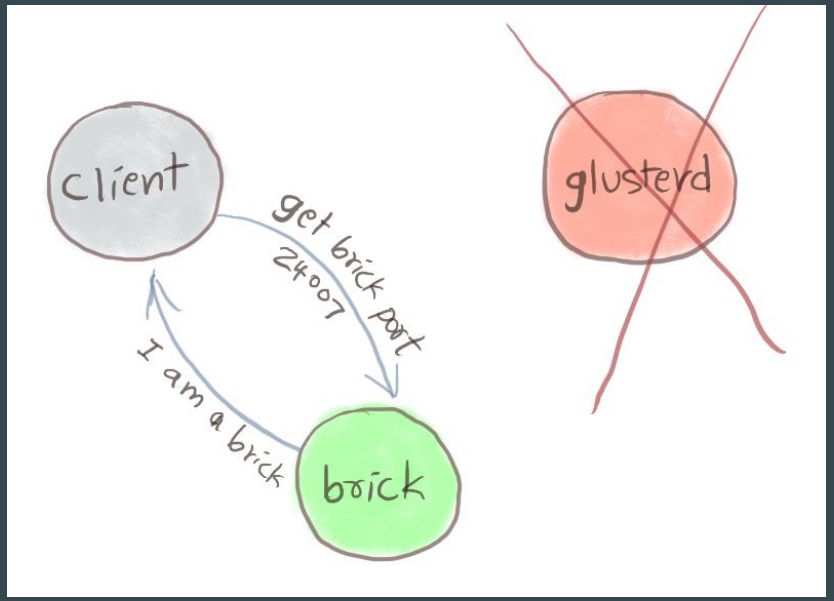
```
$ kubectl get pods -n kadalu
NAME                              READY   STATUS    RESTARTS   AGE
operator-6dfb65dcdd-r664t         1/1     Running   0          30m
csi-attacher-0                    2/2     Running   0          30m
csi-provisioner-0                 3/3     Running   0          30m
csi-nodeplugin-5hfms              2/2     Running   0          30m
csi-nodeplugin-924cc              2/2     Running   0          30m
csi-nodeplugin-cbjl9              2/2     Running   0          30m
server-storage-pool-1-kube1-0     1/1     Running   0          84s
```



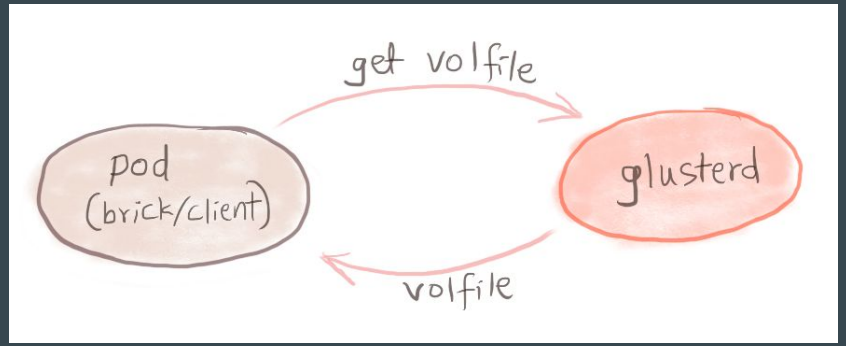
Portmap : No more required



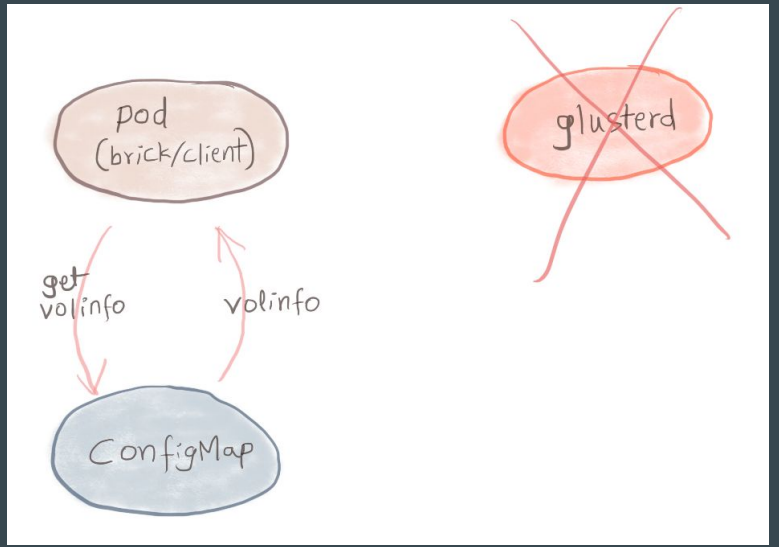
Vs



Volfile: Not a glusterd job anymore



Vs



KaDalu- Operator for GlusterFS in k8s

GlusterD	**kaDalu**
Clustering / Peer Management	k8s
Volume Management	ConfigMap, `kubectl apply`
Brick process management	K8s's pod management
Portmap for Bricks	Not required in new model
Service Management (brick, self-heal, etc)	Runs as another container in same pod. So, managed by k8s as any other pod.
Volfile for Bricks, self-heal etc	ConfigMap
Quota, Snapshot, Geo-Replication	CSI / SideCar containers

Demo

https://asciinema.org/a/259951

More on Internals

- Fixed Templates, instead of volgen depending on volume ‘create’.
- Quota is set directly on backend from a sidecar container.
- Self-heal daemon runs as sidecar container
- A single Gluster volume provides multiple PVs.

Questions?

- Install and test it with different configs.
- Provide a github star
- Help enhance CI/CD
- Contribute documentation and code to project
- File Issues @ https://github.com/kadalu/kadalu/issues
- Contribute to Gluster project to make it simpler for this use-case.

Aravinda VK
@aravindavk

Amar Tumballi
@tumballi