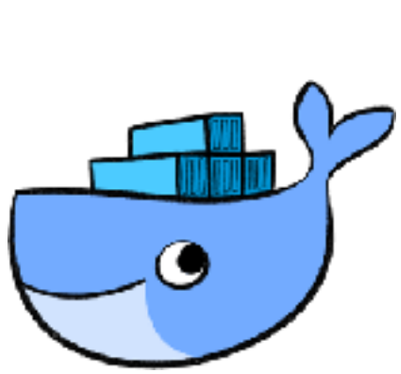
Kubernetes Storage And GlusterFS

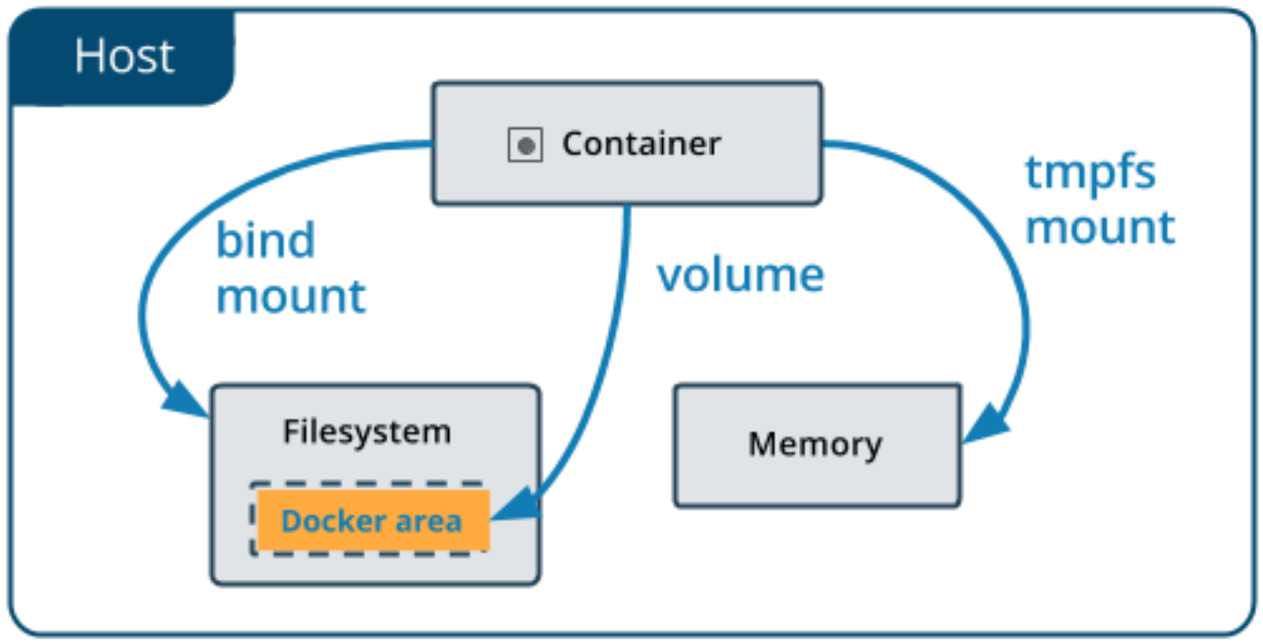
David Chang

- Docker Storage
- Kubernetes Storage
- GlusterFS on Kubernetes



Docker Storage

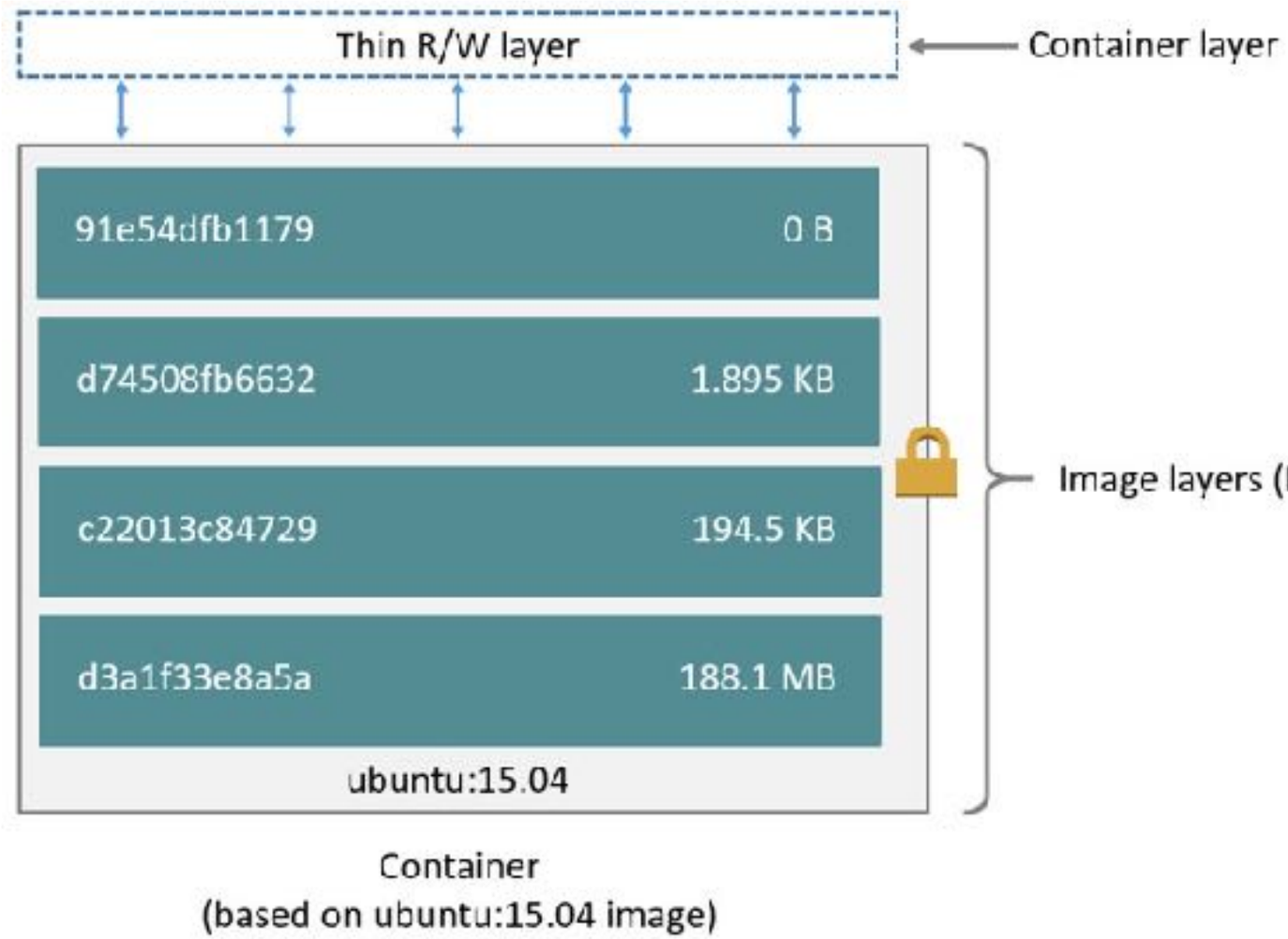
- Writable layer of container
- Tmpfs
- Docker Volume
- …



https://docs.docker.com/v17.09/engine/admin/volumes/

Writable layer

- Inside container
- Live with container
- Hard to send in/out data



Writable layer

- Live (and die) with container

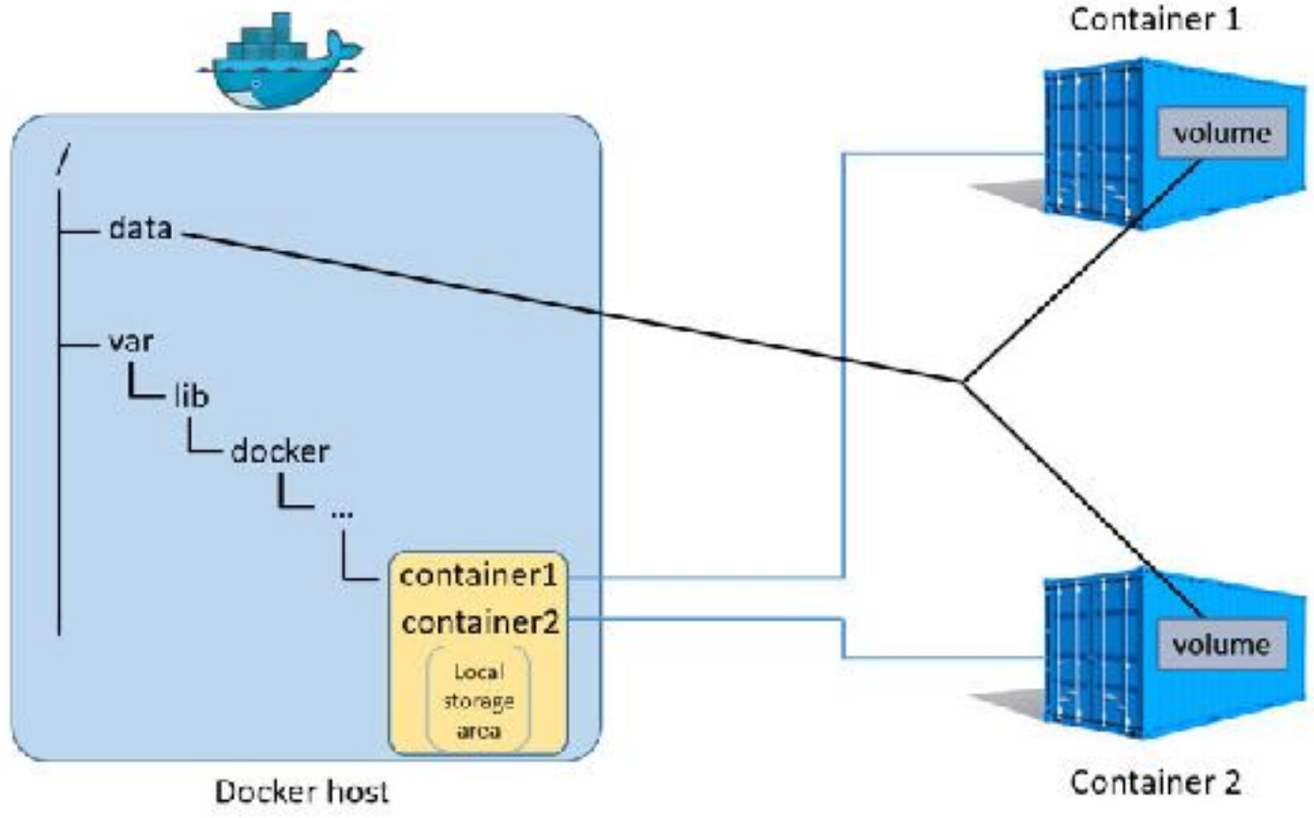
- docker ps -s
  size = writable layer
  virtual size = writable layer + container

~ (zsh)

```
Last login: Fri Feb  9 15:02:05 on ttys008
➜  ~ git:(master) ✗ docker ps -s
CONTAINER ID        IMAGE                  COMMAND                  CREATED             STATUS              PORTS                              NAMES               SIZE
cedc3915ac60        nginx                  "nginx -g 'daemon of…"   3 hours ago         Up 3 hours          80/tcp                             nginx               2B (virtual 108MB)
0f1cfe835044        ubuntu                 "/bin/bash"              22 hours ago        Up 22 hours                                            ubuntu              57B (virtual 112MB)
91f2b18232b9        artefactual/gearmand   "docker-entrypoint.s…"   22 hours ago        Up 22 hours         0.0.0.0:4730->4730/tcp             gearmand            0B (virtual 20.8MB)
d1da54541f55        influxdb               "/entrypoint.sh infl…"   22 hours ago        Up 22 hours         0.0.0.0:8086->8086/tcp             influxdb            0B (virtual 198MB)
50ff28ea304f        redis                  "docker-entrypoint.s…"   22 hours ago        Up 22 hours         0.0.0.0:6379->6379/tcp             redis               0B (virtual 107MB)
34f38ca394e6        mongo                  "docker-entrypoint.s…"   22 hours ago        Up 22 hours         0.0.0.0:27017->27017/tcp           mongo               1.02MB (virtual 360MB)
➜  ~ git:(master) ✗
```

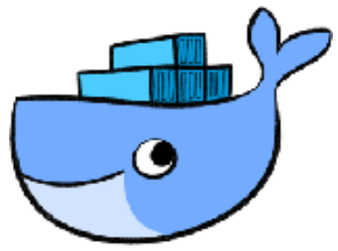
Docker Volume

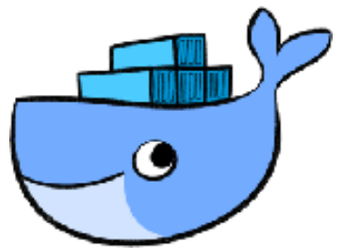
- Anonymous Volume
  - Mount to docker storage
- Bind Mount
  - Mount to host filesystem
  - docker run -v

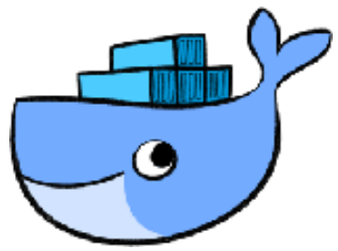


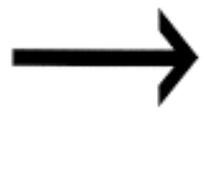
From Docker to K8s

- Reliability
- Scalability
- Resource Management
- Performance









https://kubernetes.io/docs/tasks/configure-pod-container/configure-volume-storage/

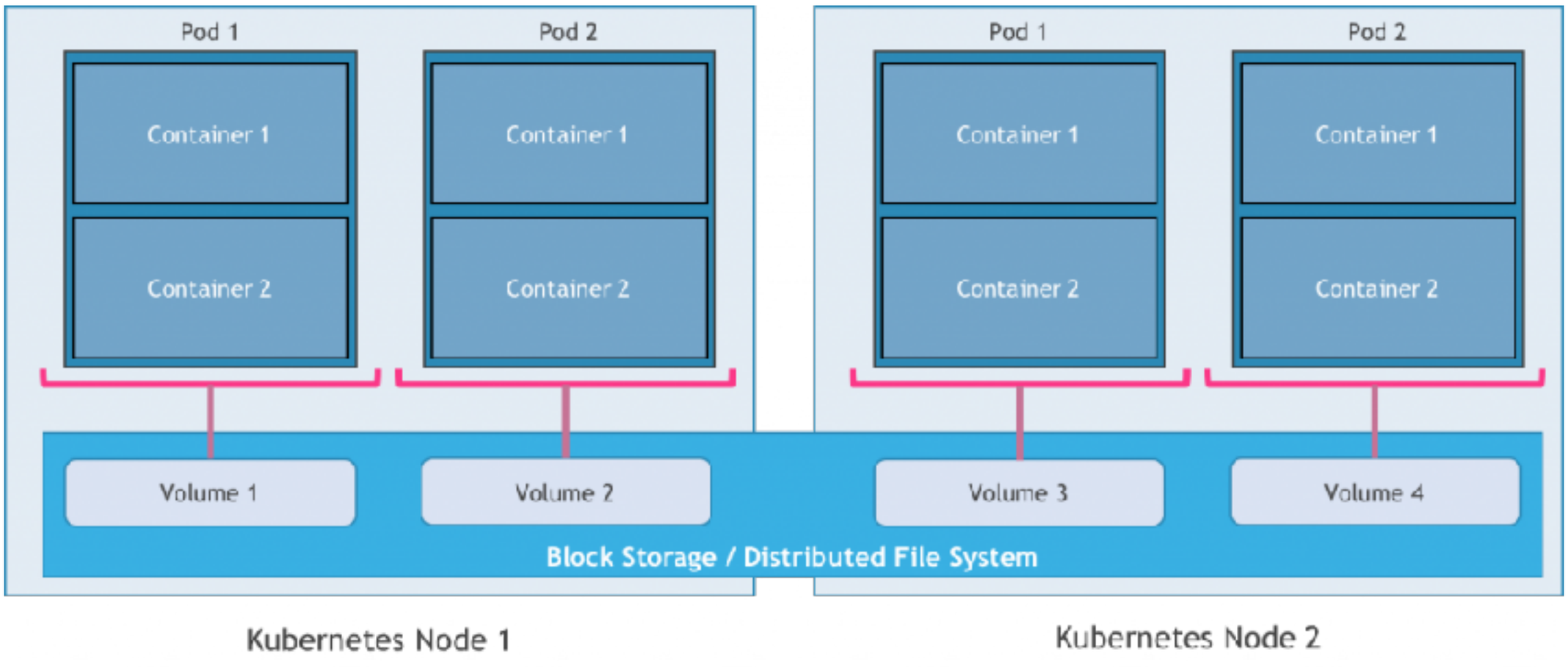
Kubernetes Storage

- On-disk file
  - Inside container
  - Hard to share
  - Live with container
- Kubernetes Volume

https://kubernetes.io/docs/concepts/storage/volumes/

Kuberentes Volume

- Explicit lifetime: outlive containers
- A directory, with different medium behind



https://thenewstack.io/strategies-running-stateful-applications-kubernetes-volumes/

K8s Volume Types

- emptyDir
- AWSStore, AzureDisk, GCEDisk
- PersistentVolumeClaim, configMap, secret
- NFS, CephFS, GlusterFS,
- …

https://kubernetes.io/docs/concepts/storage/volumes

emptyDir

- Provision: Pod assigned to Node
- Lifetime: with Pod
- Sharing: containers in Pod
- Delete: with Pod
- Medium: whatever on the Node: disk, SSD, tempts…

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: test-pd
spec:
  containers:
  - image: k8s.gcr.io/test-webserver
    name: test-container
    volumeMounts:
    - mountPath: /cache
      name: cache-volume
volumes:
- name: cache-volume
  emptyDir: {}
```

GCEPersistentDisk

- Provision: GCP
  gcloud compute disks create
  —size=500GB my-data-disk

- Lifetime: GCP

- Sharing:
  multi-Pod read or one-Pod R/W

- Delete: GCP

- Medium: whatever on the Node: disk, SSD, tempfs…

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: test-pd
spec:
  containers:
  - image: k8s.gcr.io/test-webserver
    name: test-container
    volumeMounts:
    - mountPath: /test-pd
      name: test-volume
  volumes:
  - name: test-volume
    # This GCE PD must already exist.
    gcePersistentDisk:
      pdName: my-data-disk
      fsType: ext4
```

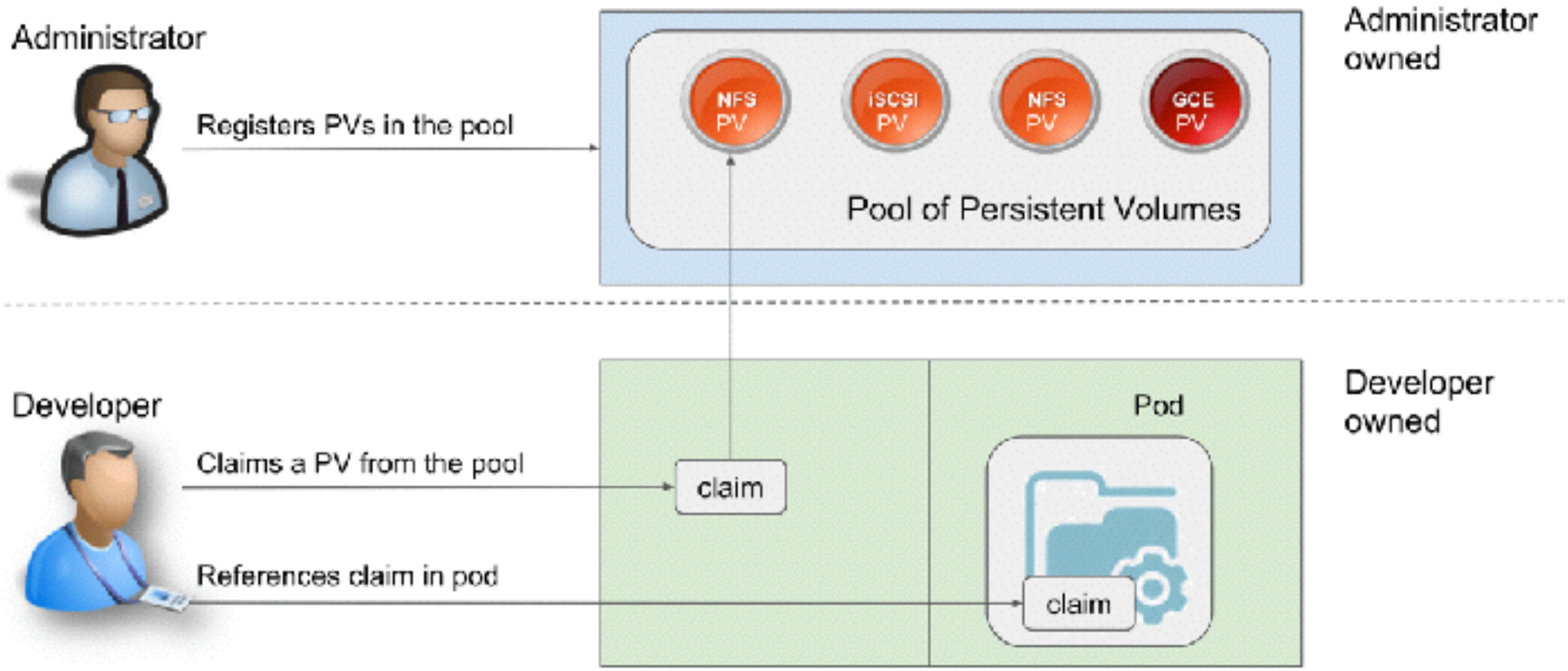
PersistentVolumeClaim

- A K8s resource request
- A type of K8s Volume
- Depends on PV impl
- StorageClass

```yaml
kind: PersistentVolumeClaim
apiVersion: v1
metadata:
  name: myclaim
spec:
  accessModes:
    - ReadWriteOnce
  volumeMode: Filesystem
  resources:
    requests:
      storage: 8Gi
  storageClassName: slow
  selector:
    matchLabels:
      release: "stable"
    matchExpressions:
      - {key: environment, operator: In, values: [dev]}
```

PV & PVC

- Abstract object with k8s APIs
- Many implementations: ex GCEDisk, NFS, GlusterFS…



https://thenewstack.io/strategies-running-stateful-applications-kubernetes-persistent-volumes-claims/

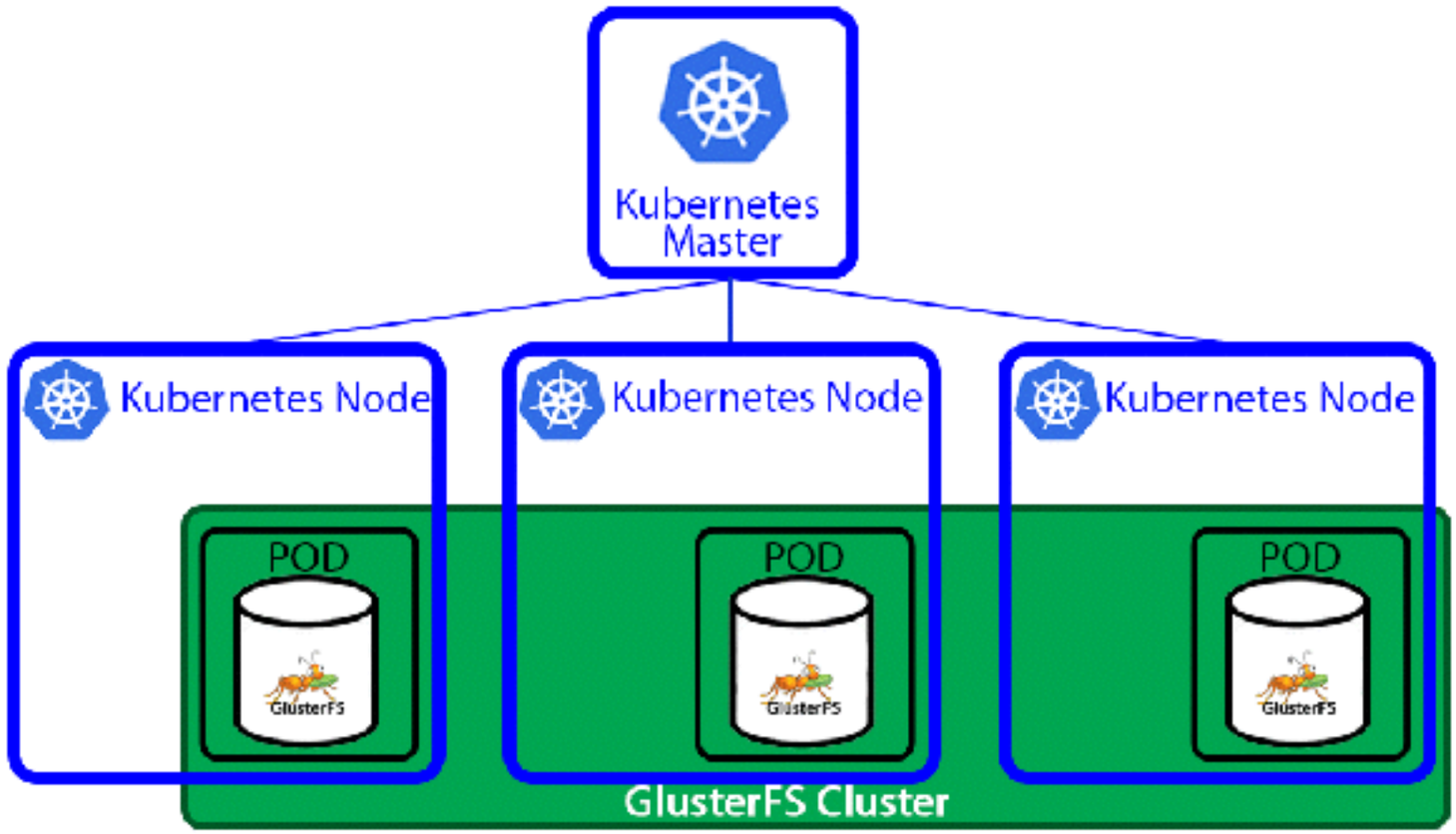
StorageClass

```yaml
apiVersion: storage.k8s.io/v1
kind: StorageClass
metadata:
  name: slow
provisioner: kubernetes.io/glusterfs
parameters:
  resturl: "http://127.0.0.1:8081"
  clusterid: "630372ccdc720a92c681fb928f27b53f"
  restauthenabled: "true"
  restuser: "admin"
  secretNamespace: "default"
  secretName: "heketi-secret"
  gidMin: "40000"
  gidMax: "50000"
  volumetype: "replicate:3"
```

```yaml
kind: PersistentVolumeClaim
apiVersion: v1
metadata:
  name: myclaim
spec:
  accessModes:
    - ReadWriteOnce
  volumeMode: Filesystem
  resources:
    requests:
      storage: 8Gi
  storageClassName: slow
  selector:
    matchLabels:
      release: "stable"
    matchExpressions:
      - {key: environment, operator: In, values: [dev]}
```

GlusterFS

- Provision: GlusterFS
- Lifetime: GlusterFS
- Sharing: multi-Pod R/W
- Delete: GlusterFS
- Medium: GlusterFS



https://github.com/kubernetes/examples/tree/master/staging/volumes/glusterfs

PV to GlusterFS

- Admin: provides PVs
- User: Pod request PVClaim
- K8s “get” a PVC from GlusterFS
- Mount PVC to Pod
- Release PVC when Pod terminated
- Reuse PV

https://github.com/kubernetes/examples/tree/master/staging/volumes/glusterfs

GlusterFS

- Software defined distributed storage
- Scalable networking filesystem
- Free & open source

- Heketi
- LVM

Quick Start

- Install glusterfs
- Prepare a clean filesystem on host
- Establish trusted pool
- Create glusterFS volume
- Use volume through network
- http://docs.gluster.org/en/latest/Quick-Start-Guide/Quickstart/

Heketi

- RESTful volume management for GlusterFS
- Automatically bricks placement
- Health check
- Multiple GlusterFS clusters support

https://github.com/heketi/heketi

LVM

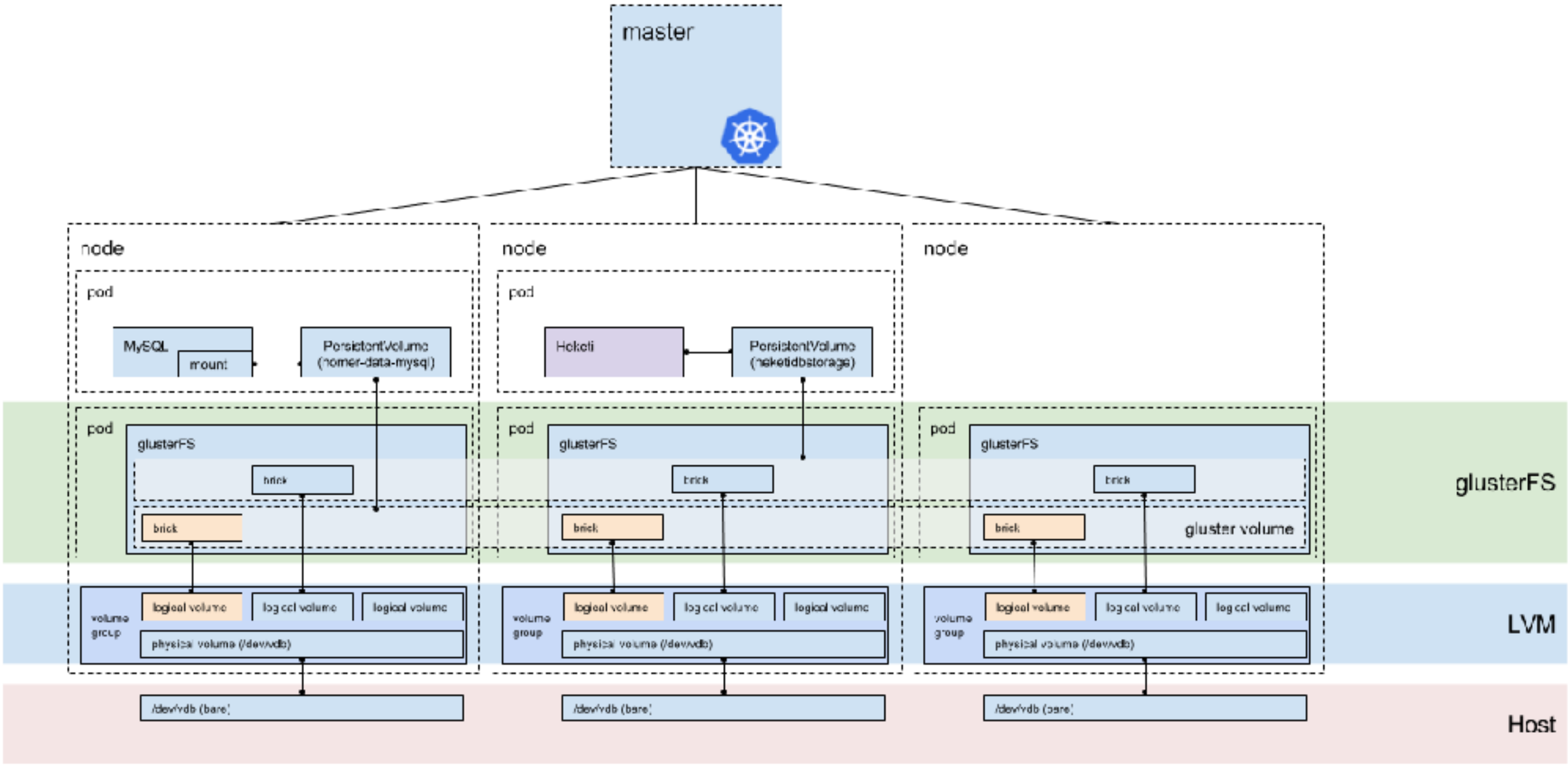
- Logical Volume Manager
- Manage volumes, filesystems
- Create filesystems without partitioning a disk

https://wiki.ubuntu.com/Lvm

Put Things Together

- My-apps
- Kubernetes
- Heketi
- GlusterFS
- LVM
- Host

Put Things Together



http://blog.leifmadsen.com/blog/2017/09/19/persistent-volumes-with-glusterfs/

Put Things Together

- Admin creates GlusterFS cluster with overall 100Gi capacity
- Admin starts Heketi and applies Heketi secret to K8s
- Admin creates a StorageClass of GlusterFS provisioner with replicate:3
- In K8s, User creates a Pod request a PVC of 10 Gi
- K8s forward request to Heketi, Heketi provisions 3 replica of 10Gi volume on GlusterFS(Occupies 30Gi)
- K8s bind PVC 10Gi to Pod
- Pod status from Creating to Running
- Pod terminated. K8s Reclaim PVC back to PV
- Heketi reclaim provisioned volume

Q & A

Linker Networks

- https://www.yourator.co/companies/LinkerNetworks
- ML/DL
- Networking virtualization
- Kubernetes
- dchang@linkernetworks.com

That’s all, for today

Thank you very much