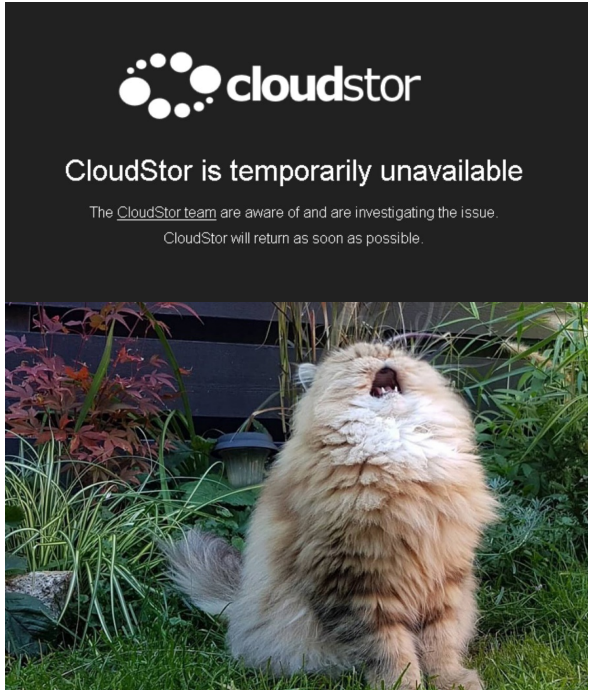
HERDING K8S

Crystal Chua, AARNet

A BIT OF BACKGROUND

- **Last year, we had issues with our S3 gateways taking CloudStor offline**
  - The working hypothesis was resource contention between eosd and xrd+webdav
- **We tried to work through these issues with a lot of help from CERN!!**
- **In the end, we decided to move the S3 gateways to their own storage**
- **And then decided that we would redesign the way we do storage**
  - More resilient
  - Smaller instances instead of one large instance
  - Move to single disk per FST (currently 1 server = 1 FST)
- **We called these new, smaller instances “shards”**



REBUILDING FROM SCRATCH

- The first decision we made was moving to Kubernetes
- Problem: no one on the team had concrete experience with Kubernetes
- Spent over a month just learning how to run and operate Kubernetes

THERE WERE A LOT OF QUESTIONS

deployment??

persistent volumes?

persistent volumes claims??

daemonset??

rancher ??

pod??

statefulset???

ingress???

service??

calico???

or canal?

helm charts??

or maybe flannel??

namespaces?

kubedns?

coredns??

aarnet

DEPLOYING KUBERNETES

- We tried a few deployment methods, ended up going with Rancher 2.2
- RKE (Rancher Kubernetes Engine) deploys the undercloud
- Helm deploys the Rancher application to the undercloud
- Rancher creates & manages other Kubernetes clusters
- How to deploy in production, test, etc?
- `./deploy-everything.sh`

“DEPLOY EVERYTHING”

- Everything is templated, only a few variables (eg. cluster name) needs to change
- We recently ran up a new development cluster and only had one issue that wasn’t previously documented (Rancher web UI needs a proper cert)
- Updated doco:

Setup + REQUIREMENTS

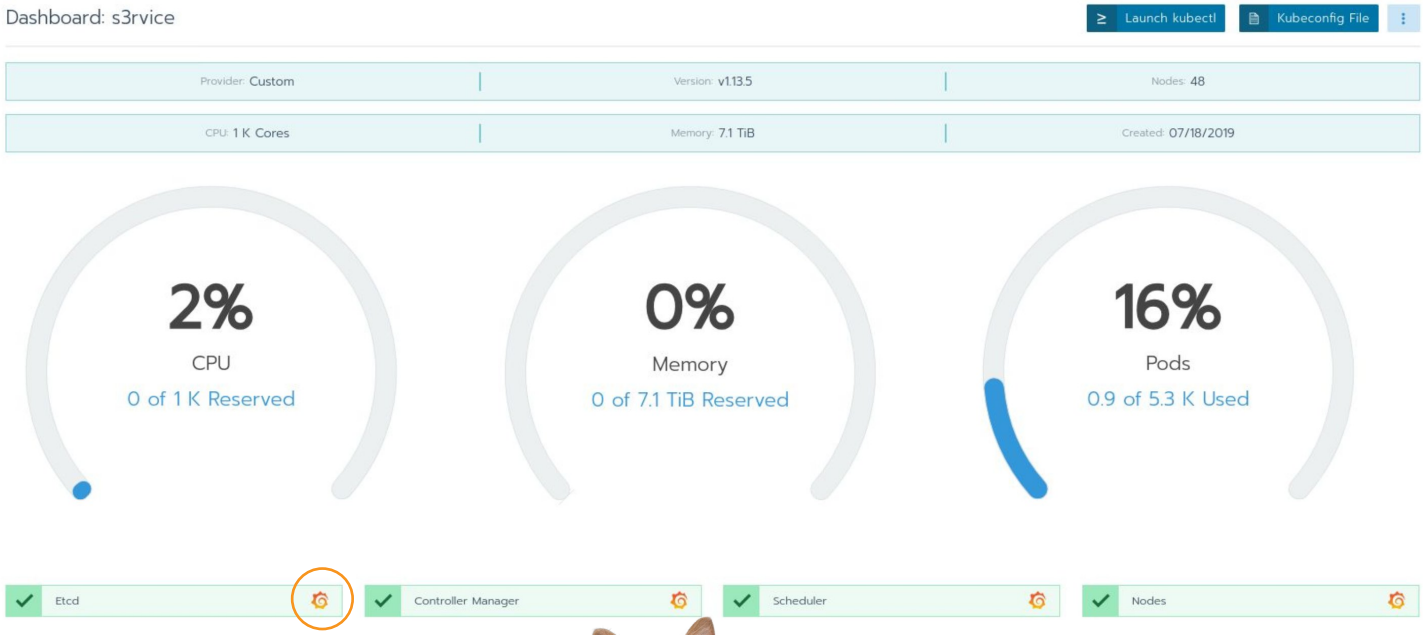
```
./do-everything
```

Specifying `rancher.yml` in the Rancher helm install command is ***crucial*** - this file contains all customisations/configuration options we require.

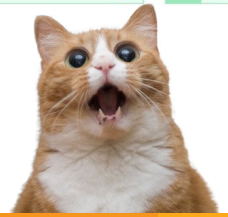
REQUIREMENTS (like i mentioned before)

For a new environment you will need to specify a separate hostname and also proper certs for that hostname. Rancher will absolutely hate your guts if you don't give it a proper cert. The cert has to be a full chain, don't forget that. DO NOT FORGET THAT

RANCHER 2.2 IS NICE



in-built grafana + prometheus



REALLY NICE

Cluster Logging

We will use fluentd to collect stdout/stderr logs from each container and the log files which exist under path `/var/log/containers/` on each host. The logs can be shipped to a target you configure below.

Edit as File

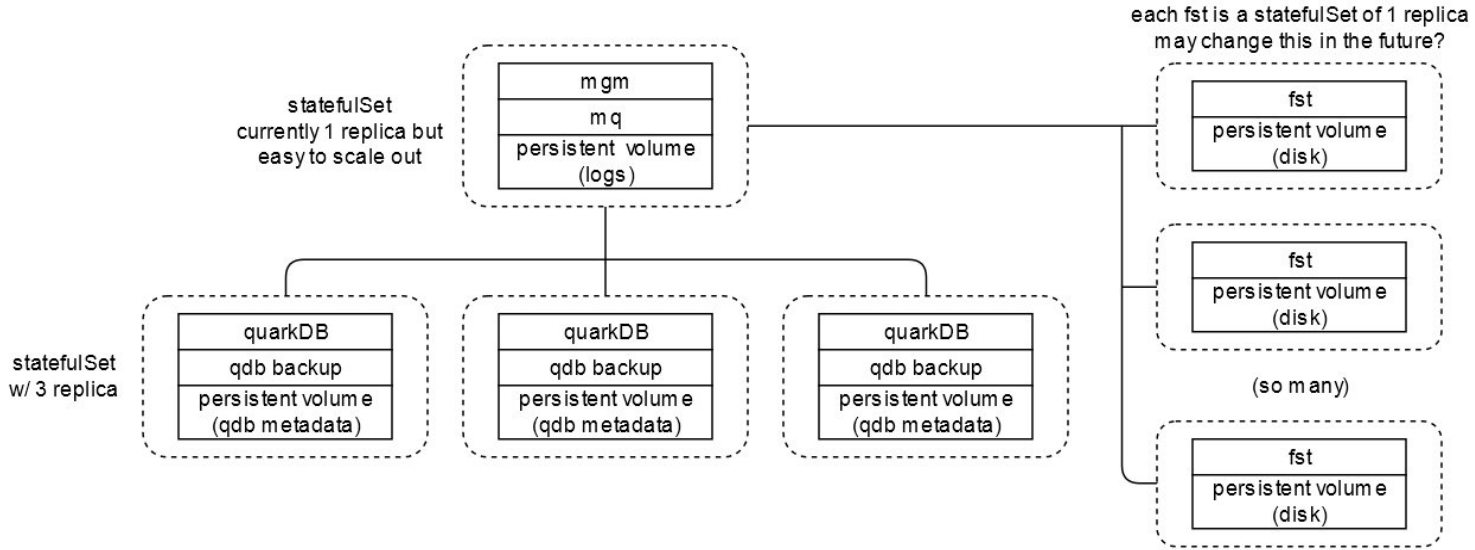
None | Elasticsearch | Splunk | Kafka | Syslog | Fluentd

No logging target, click the Save button below to set `Elasticsearch` as the logging target.

DEPLOYING EOS

- We deploy EOS using helm charts
- Helm charts are essentially templated Kubernetes resources
- The only manual step in deploying currently is disk encryption
  - That said, we just run a script on the storage servers to do this
- Config file is templated, each instance only needs minimal edits
  - Instance name
  - List of allocated disks
- `./deploy.sh –s <instance-name>`

EOS IN KUBERNETES @ AARNET



THE BENEFITS OF SHARDING

mostly stable but single large point of failure

upgrades take a long time (esp. with in-memory namespace)

complicated to replicate setup

templated, standardized

deployment of a new instance takes a couple of minutes

individually easy to manage, but also easy to customise

WHAT WE HAVE NOW

- A dev shard for testing new features, upgrades etc
- A test shard for pre-prod testing & offering trials to interested people
- A “general” shard for the majority of users
- Customer-specific shards for institutions/organisations that pay for it

NEXT STEPS

- Make everything Kubernetes!!!
- Working towards deploying OCIS directly to Kubernetes
- Moving our last, giant EOS instance to QuarkDB
- And then splitting that instance up into shards too
  - (kind of like the EOSHOME migrations!)

THANK YOU