Fedora Atomic host at CERN

CentOS Dojo at CERN 2018

Spyros Trigazis @strigazi

OpenStack Magnum

What is Magnum?

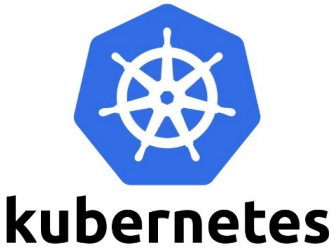
An OpenStack API service that allows creation of container clusters.

- Use your keystone credentials
- You choose your cluster type
- Single-tenant clusters
- Quickly create new clusters with advanced features such as multi-matser

MAGNUM

an OpenStack Community Project

Terminology (1/3): COE



Terminology (2/3): Magnum Cluster

A Magnum cluster is composed of:

- compute instances (virtual or physical)
- neutron networks
- security groups
- cinder volumes
- other resources (eg Load Balancer)

using OpenStack Heat

- Where your containers run
- Lifecycle operations
  - Scale up/down
  - Upgrade
  - Node heal/replace
- Self contained cluster with each own monitoring, data store, additional resources

Terminology (3/3): Native Client

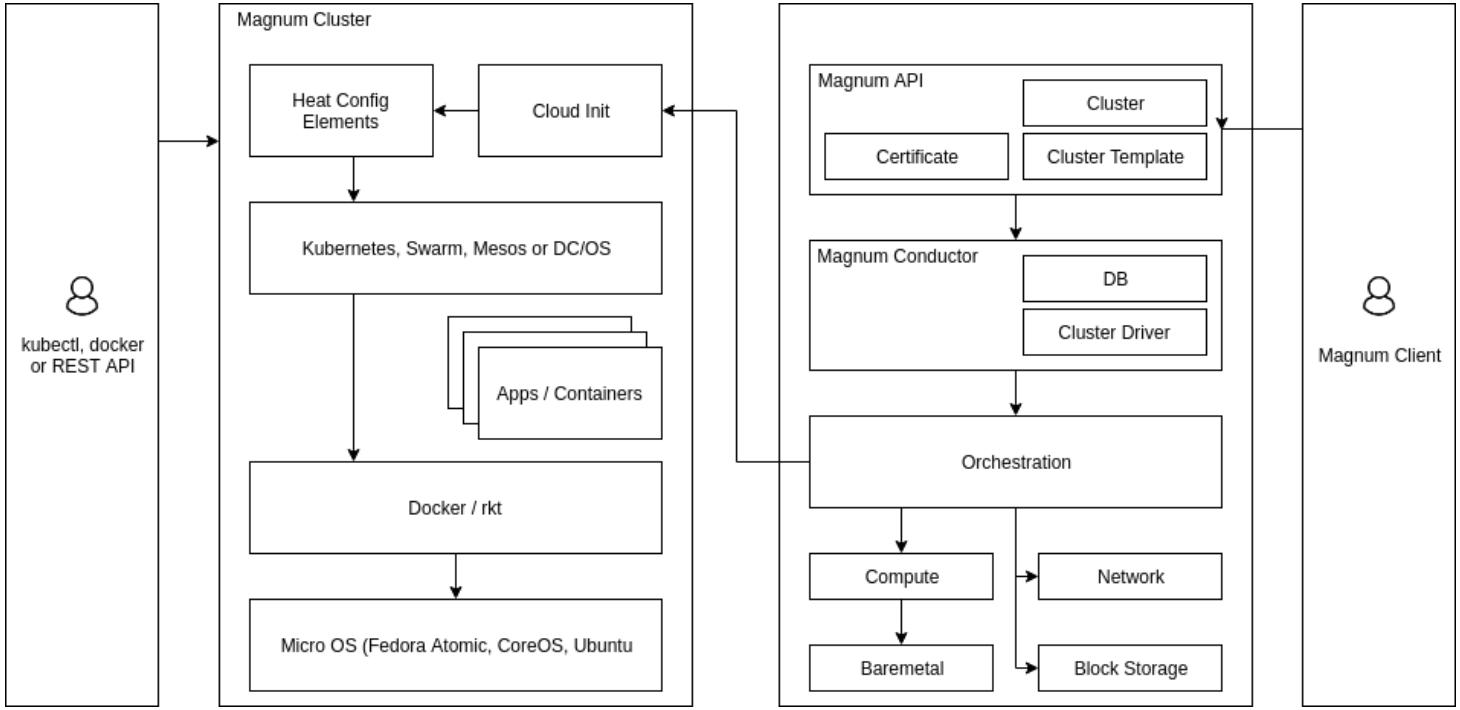
Magnum does not offer a container API, but it allows you to use the COE native client or API to contact your cluster securely over TLS.

Magnum creates a CA for each cluster and stores it in Barbican (recommended but optional). You can store certificates locally or in magnum's DB.

As soon as your cluster is running, you don't have to use the magnum to run containers or even create cinder volumes or Load Balancers. You can use:

- docker
- kubectl
- dcos
- marathon API

OpenStack Magnum Architecture



Why use Magnum to run a container service

- Centrally managed self-service like GKE and EKS
  - Provide clusters to users with one-click deployment (or one API call)
  - You have more than 5 users and more than 10 clusters
- Accounting comes for free if you use quotas in your projects
- Easy entrypoint to containers for new users
- Control which OS your users are running

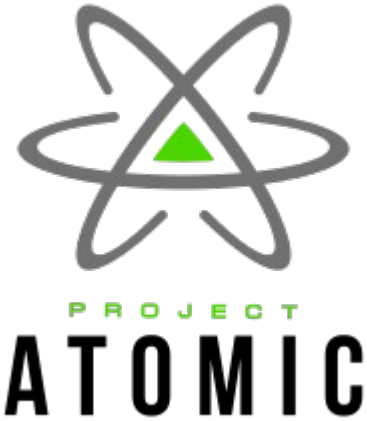
What to consider when running a container service

- Design your network
  - By default, magnum creates a private network per cluster and assigns floating IPs to nodes
  - LBaaS for multi-master clusters
- Run a container registry
  - DockerHub is usually up but latency will always get you
  - Rebuild or mirror the containers used by magnum
- Provide self-service clusters -> Provide software
  - Upgrade magnum regularly, update its configuration regularly
  - Plan which container and glance images are available to users

Working with Atomic Working Group

What is Atomic?

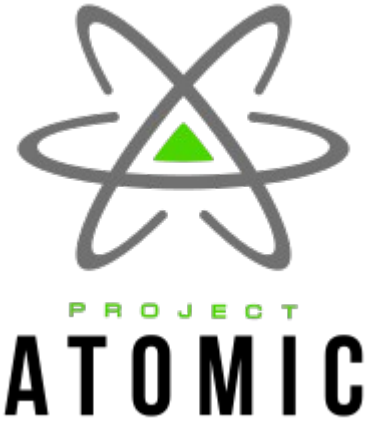
- Immutable Filesystem
- Transactional upgrades (requires reboot)
- Minimal 400mb total, optimized for running linux containers




```
# whoami
root
# dnf
bash: dnf: command not found
# mkdir /foo
mkdir: cannot create directory ‘/foo’: Operation not permitted
# touch /usr/bin/bar
touch: cannot touch '/usr/bin/bar': Read-only file system
```

Running upstream Fedora Atomic

- Freedom for deployments to upgrade when they want
- Forces us to customize only with read-only containers
- Ask the fedora community for help
- No build CI to maintain, OS tested by a bigger community
- https://getfedora.org/en/atomic/



Eventually you become a contributor

- Co-maintain the kubernetes package for fedora and centos
  - Using the same distgit for months :)
- Early testers of skopeo and atomic utilities
- Contribute to system containers by Project Atomic

  https://github.com/projectatomic/atomic-system-containers

Extending Fedora Atomic with System Containers

```
# atomic install --system --storage ostee --name kubelet \
  registry.fedoraproject.org/f28/kubelet
# # edit /etc/kubernetes/kubelet
# systemctl start kubelet
#
# atomic install --system --storage ostee --name docker \
# ${REGISTRY}fedora-docker:18.06
# systemctl start docker
```

```
# atomic containers list --no-trunc
   CONTAINER ID           IMAGE                                    NAME                  STATE      BACKEND
   etcd                   .../etcd:v3.2.7                          etcd                  running    ostree
   kube-apiserver         .../kubernetes-apiserver:v1.11.1         kube-apiserver        running    ostree
   kube-controller-manager .../kubernetes-controller-manager:v1.11.1 kube-controller-manager running    ostree
   kube-scheduler         .../kubernetes-scheduler:v1.11.1         kube-scheduler        running    ostree
   flanneld               .../flannel:v0.9.0                       flanneld              running    ostree
   heat-container-agent   .../heat-container-agent:rawhide         heat-container-agent  running    ostree
```

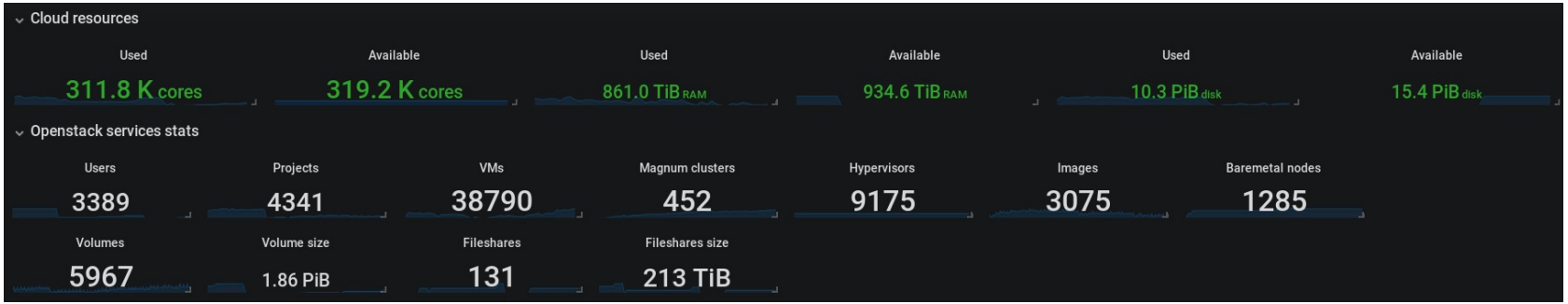
CERN Container service

CERN OpenStack Infrastructure

Production since 2013

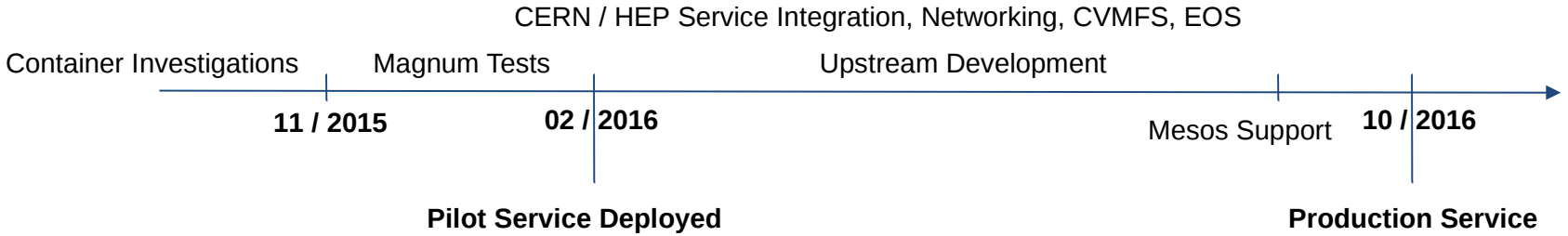
~ 311,000 cores ~600 vms per hour ~36,000 vm running ~450 clusters running

~ 1500 Fedora Atomic 27 VMs



CERN Magnum Deployment

- Integrate containers in the CERN cloud
  - Shared identity, networking integration, storage access, …
- Add CERN services in *system* containers
- Introduce cvfms docker storage driver (file-grained)
- **Fast, Easy to use**



CERN Magnum Deployment

- Clusters are described by *cluster templates*
- Shared/public templates for most common setups, customizable by users

```
$ openstack coe cluster template list
+------+--------------------------+
| uuid | name                     |
+------+--------------------------+
| .... | swarm                    |
| .... | swarm-ha                 |
| .... | kubernetes               |
| .... | kubernetes-ha            |
| .... | mesos                    |
| .... | mesos-ha                 |
| .... | dcos                     |
+------+--------------------------+
```

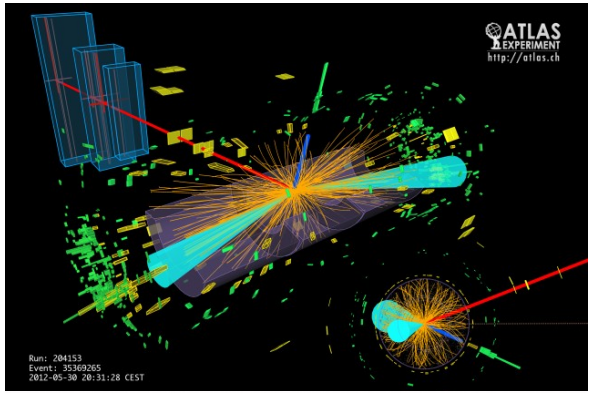
CERN Magnum Deployment

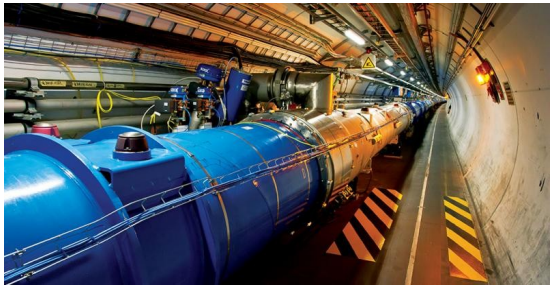
- Clusters are described by *cluster templates*
- Shared/public templates for most common setups, customizable by users

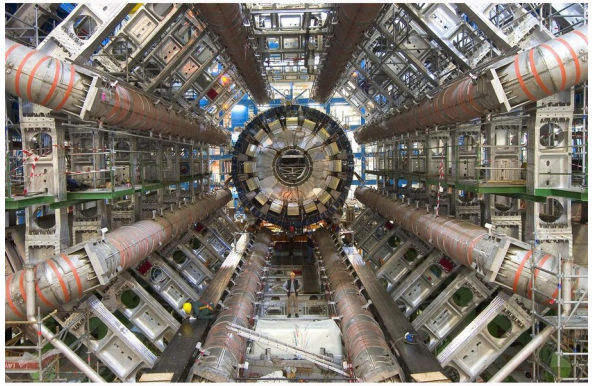
```
$ openstack coe cluster create --name my-k8s --cluster-template kubernetes --node-count 100
     ~ 5 mins later
$ openstack coe cluster list
+------+--------+------------+-------------+----------------------------+
| uuid | name   | node_count | master_count | keypair  | status          |
+------+--------+------------+-------------+----------------------------+
| .... | my-k8s | 100        | 1            | mysshkey | CREATE_COMPLETE |
+------+--------+------------+-------------+----------------------------+
$ $(openstack coe cluster config my-k8s --dir clusters/my-k8s)
$ kubectl get ...
```

CERN Container Use Cases

- Batch Processing
- End user analysis / Jupyter Notebooks
- Machine Learning / TensorFlow / Keras
- Infrastructure Management
  - Data Movement, Web servers, PaaS …
- Continuous Integration / Deployment
- Run OpenStack :-)
- And many others

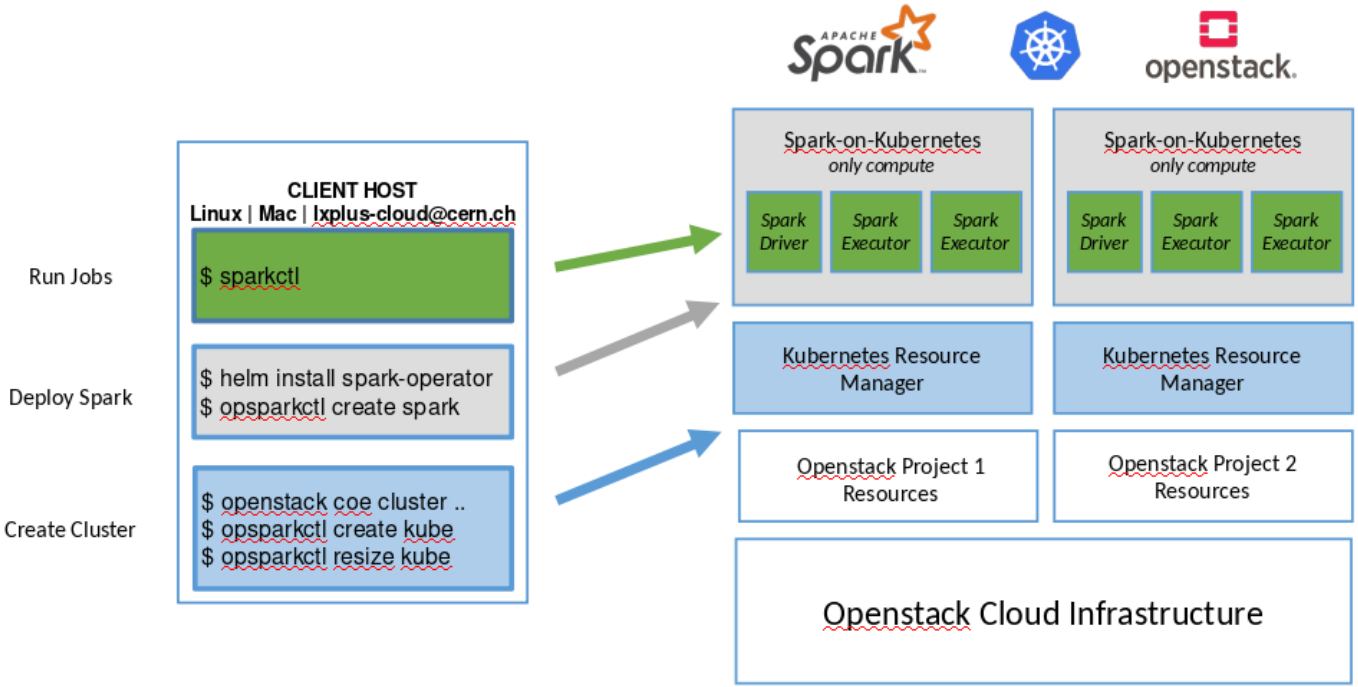






Credit: Ricardo Rocha CERN Cloud

Use Case: Spark on K8s



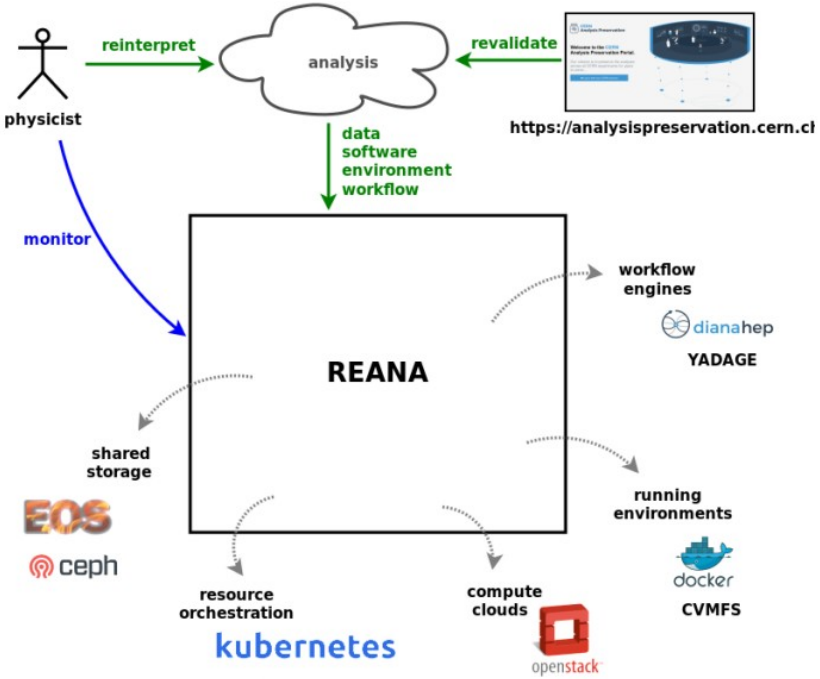
Credit: CERN data analytics working group

Use case: REANA / RECAST

Reusable Analysis Platform

- Workflow Engine (Yadage)
- Each step a Kubernetes Job
- Integrated Monitoring & Logging
- Centralized Log Collection

https://indico.cern.ch/event/557956/



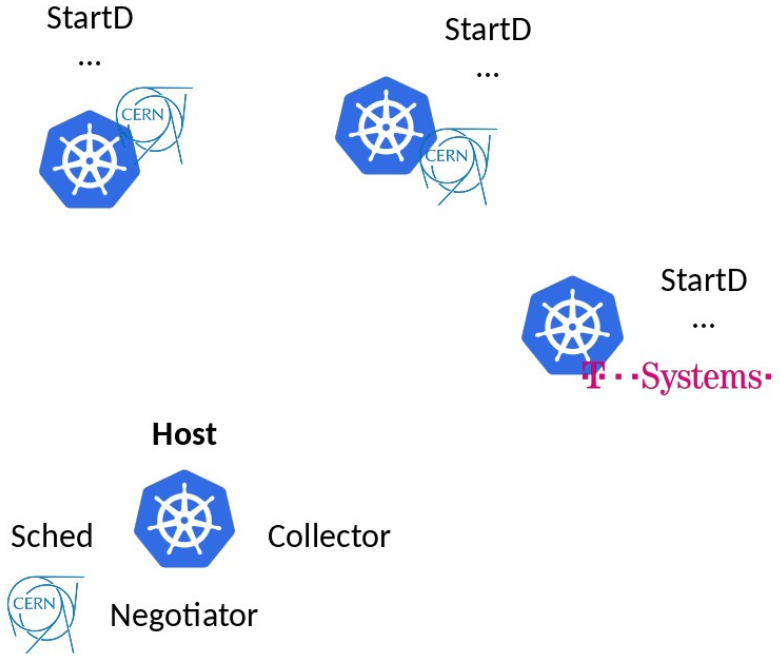
Credit: CERN Invenio User Group Workshop

Use case: Federated Kubernetes

Batch or other jobs on multiple clusters

- Segment the datacenter
- Burst compute capacity
- Same deployment in all clouds

kubefed join --host-cluster-context… --cluster-context … atlas-recast-y
openstack coe federation join cern-condor atlas-recast-x atlas-recast-y



Credit: Ricardo Rocha CERN Cloud

Conclusion

Fedora Atomic is a solid block for Magnum and CERN’s container service.

Its immutable state allows to test once for many users.

Looking forward to Fedora CoreOS!

Questions about Magnum

- Magnum IRC channel: #openstack-containers
- Meeting Tuesdays at 21h00 UTC in #openstack-containers
- Use [magnum] in openstack-operators or openstack-dev ML
- Submit/Follow Stories/Tasks https://storyboard.openstack.org/#!/project_group/magnum
- CERN users https://clouddocs.web.cern.ch/clouddocs/containers/


CERN