OpenStack Magnum @CERN

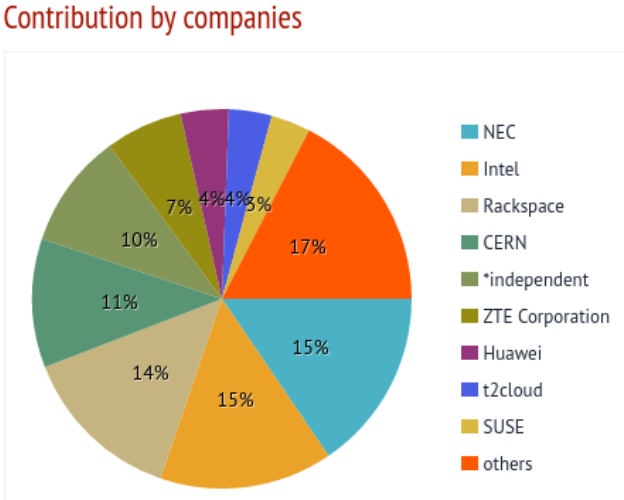
Scaling Container Clusters to thousand of nodes

Spyros Trigazis @strigazi

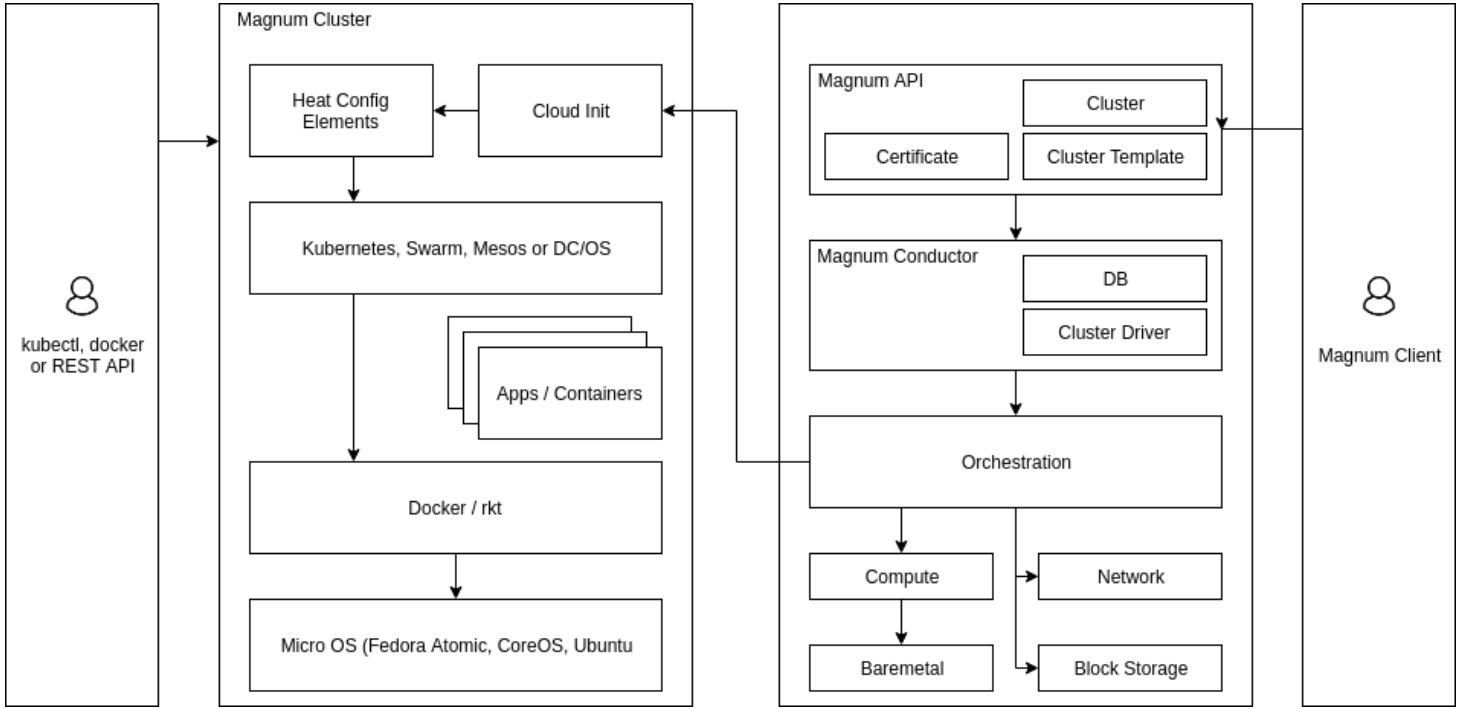
OpenStack Magnum #openstack-containers

Kubernetes, Docker Swarm, Apache Mesos, DC/OS (experimental) aaS
Deep integration of OpenStack with Container technologies:

- Compute Instances
- Networks, Load Balancers
- Storage
- Security
- Native Container API
- Lifecycle cluster operations
  - Scale cluster up and down
  - More WIP



OpenStack Magnum Architecture



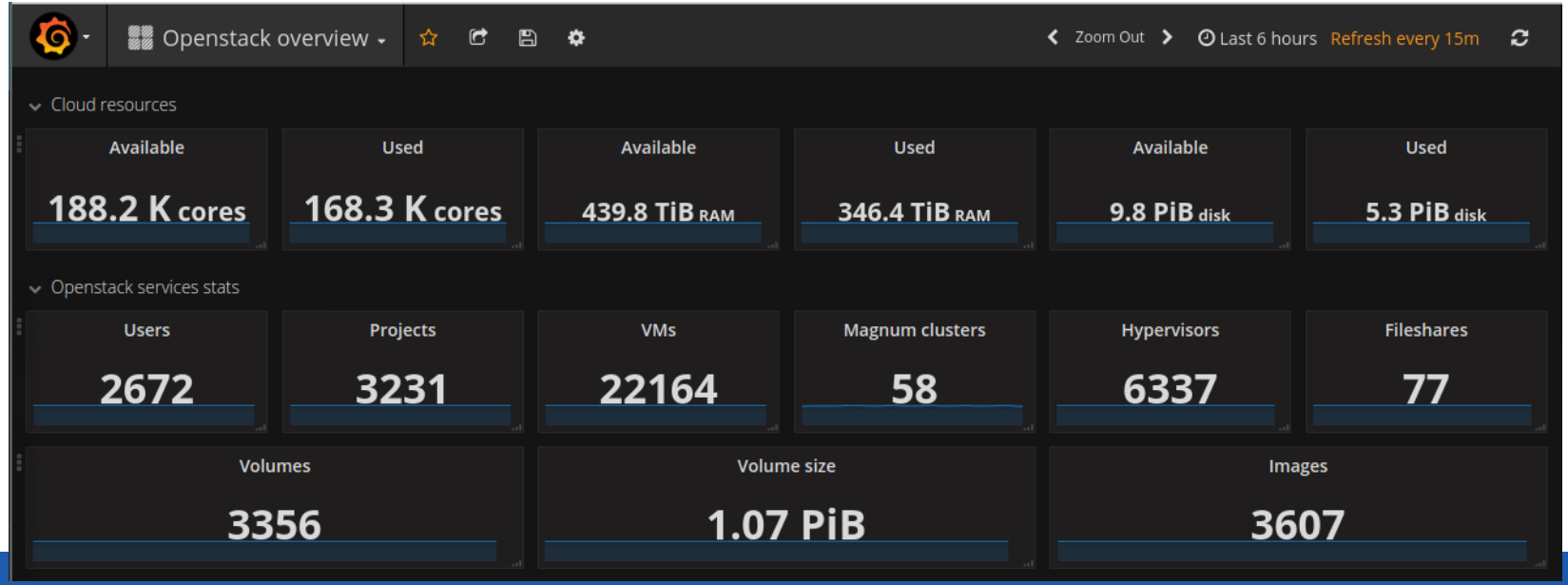
Plans for Magnum Pike

- Rolling upgrades of clusters
  - Upgrade to new versions of Kubernetes
- Heterogeneous Clusters
  - Mix of VMs and Baremetal, spread across AZs
- Docker Swarm Mode
- Container Monitoring
  - Work in Progress for a cadvisor, prometheus and grafana stack
- Full support for custom cluster drivers
  - Allow ops deploy easier driver with independent packages
- Baremetal support for all drivers

CERN OpenStack Infrastructure

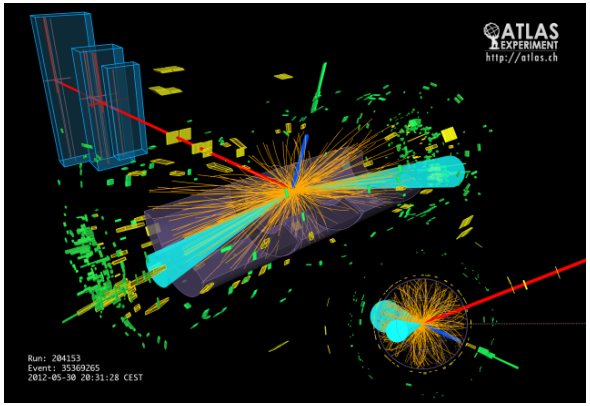
Production since 2013

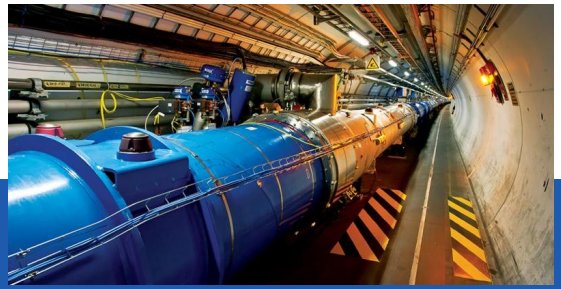
~ 190.000 cores ~4 million vms created ~200 vms per hour

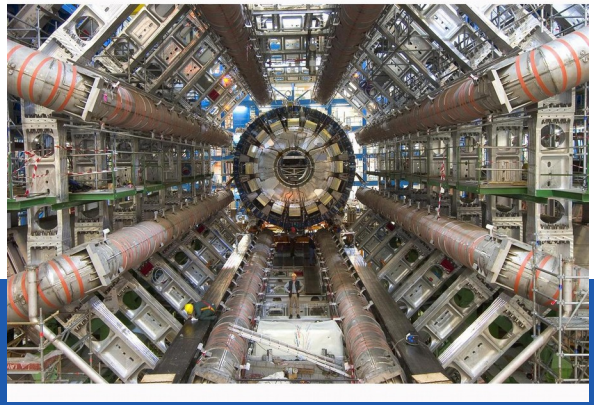


CERN Container Use Cases

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

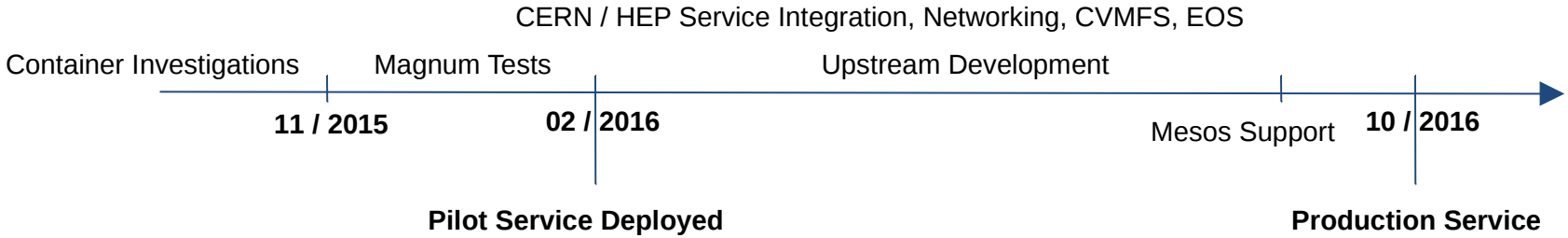






CERN Magnum Deployment

- Integrate containers in the CERN cloud
  - Shared identity, networking integration, storage access, …
- Add CERN services in *system* containers with atomic
- **Fast, Easy to use**



CERN Magnum Deployment

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

```
$ magnum cluster-template-list
+------+--------------------------+
| uuid | name                     |
+------+--------------------------+
| .... | swarm                    |
| .... | swarm-ha                 |
| .... | kubernetes               |
| .... | kubernetes-ha            |
| .... | mesos                    |
| .... | mesos-ha                 |
+------+--------------------------+
```

CERN Magnum Deployment

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

```
$ magnum cluster-create --name myswarmcluster --cluster-template swarm --node-count 100
     ~ 5 mins later
$ magnum cluster-list
+------+----------------+------------+--------------+----------------------------+
| uuid | name           | node_count | master_count | keypair  | status          |
+------+----------------+------------+--------------+----------------------------+
| .... | myswarmcluster | 100        | 1            | mysshkey | CREATE_COMPLETE |
+------+----------------+------------+--------------+----------------------------+

$ $(magnum cluster-config myswarmcluster --dir magnum/myswarmcluster)

$ docker info / ps / ...
$ docker run --volume-driver cvmfs -v atlas.cern.ch:/cvmfs/atlas -it centos /bin/bash
[root@32f4cf39128d /]#
```

Rally Benchmarks and Kubernetes scalability

- Benchmark the Magnum service
  - How fast can I get my container cluster?
  - Use Rally to measure to performance like any other OpenStack service
- Benchmark the resources
  - Ok, it was reasonably fast, what can I do with it?
  - Use a demo provided by Google to measure the performance of the cluster
    - Rally tests for container are under development and near completion

Deployment Setup at CERN and CNCF

CERN

- 240 hypervisors
  - 32 cores, 64 GB RAM, 10Gb inks
- Container storage in our CEPH cluster
- Magnum / Heat setup
  - Dedicated 3 node controllers, dedicated 3 node RabbitMQ cluster
- Flat Network for vms

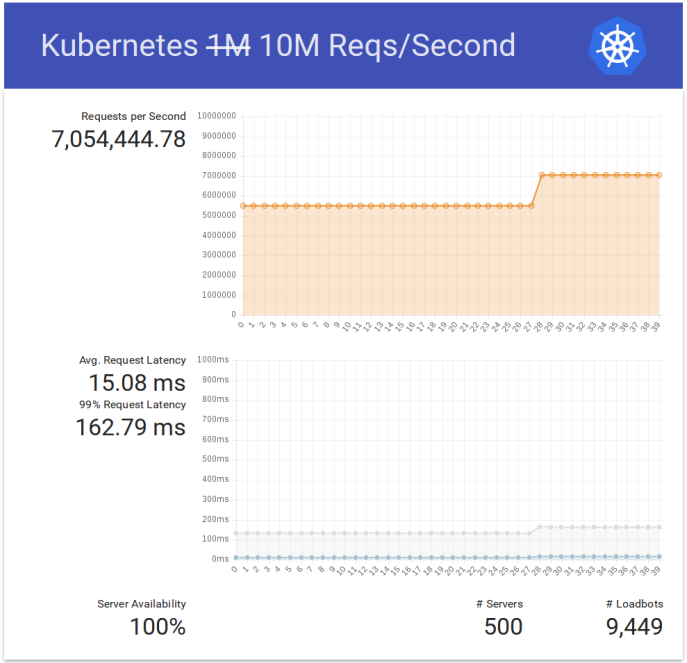
CNCF

- 100 hypervisors
  - 24 cores, 128 GB RAM
- Container storage in local disk
- Magnum / Heat setup
  - Shared 3 node controllers, shared 5 node RabbitMQ cluster
- Private networks with linux bridge

CERN Results

- Second go: rally and 7 million requests / sec
  - Kubernetes **7 million requests / sec**

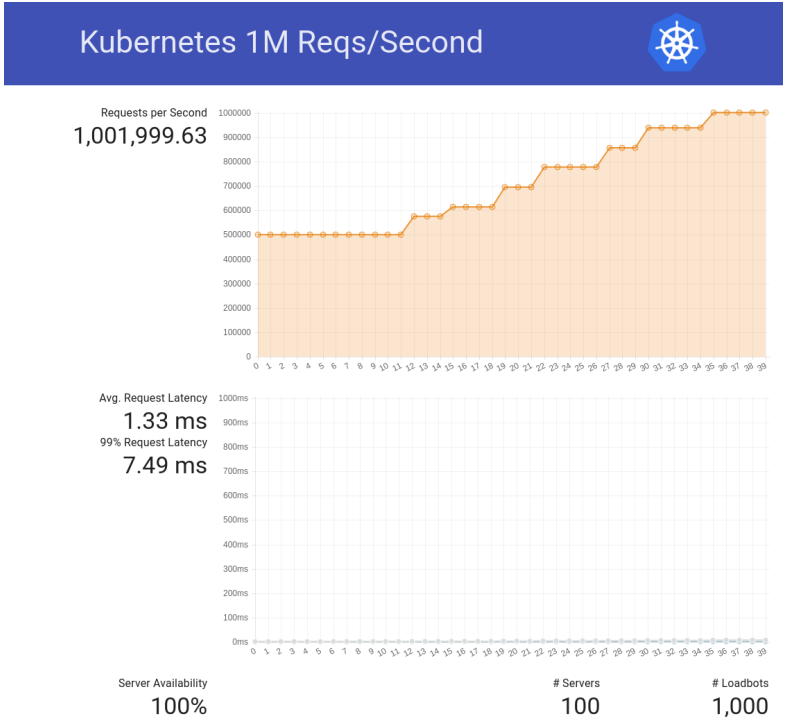
Cluster Size (Nodes)	Concurrency	Deployment Time (min)
2	50	2.5
16	10	4
32	10	4
128	5	5.5
512	1	14
1000	1	23



CNCF Results

Cluster Size (Nodes)	Concurrency	Number of Clusters	Deployment Time (min)
2	10	100	3.02
2	10	1000	Able to create 219 clusters
32	5	100	Able to create 28 clusters

nodes	containers	reqs/sec	latency	flannel
35	1100	1M	83.2 ms	udp
80	1100	1M	1.33 ms	host-gw
80	3100	3M	26.1 ms	host-gw



CERN