OpenStack & ClusterLabs

Michele Baldessari “bandini”
Software Engineer

Red Hat

Agenda

- Introduction
- OpenStack
- Deployment
- Evolution
- Thoughts
- Q&A

- Thoughts
- Next steps
- Q&A

Red Hat

Introduction

- Member of OpenStack Engineering Team
- Working on the HA bits of OpenStack (w/dciabrin)
- We had the brilliant idea of calling our team PID1

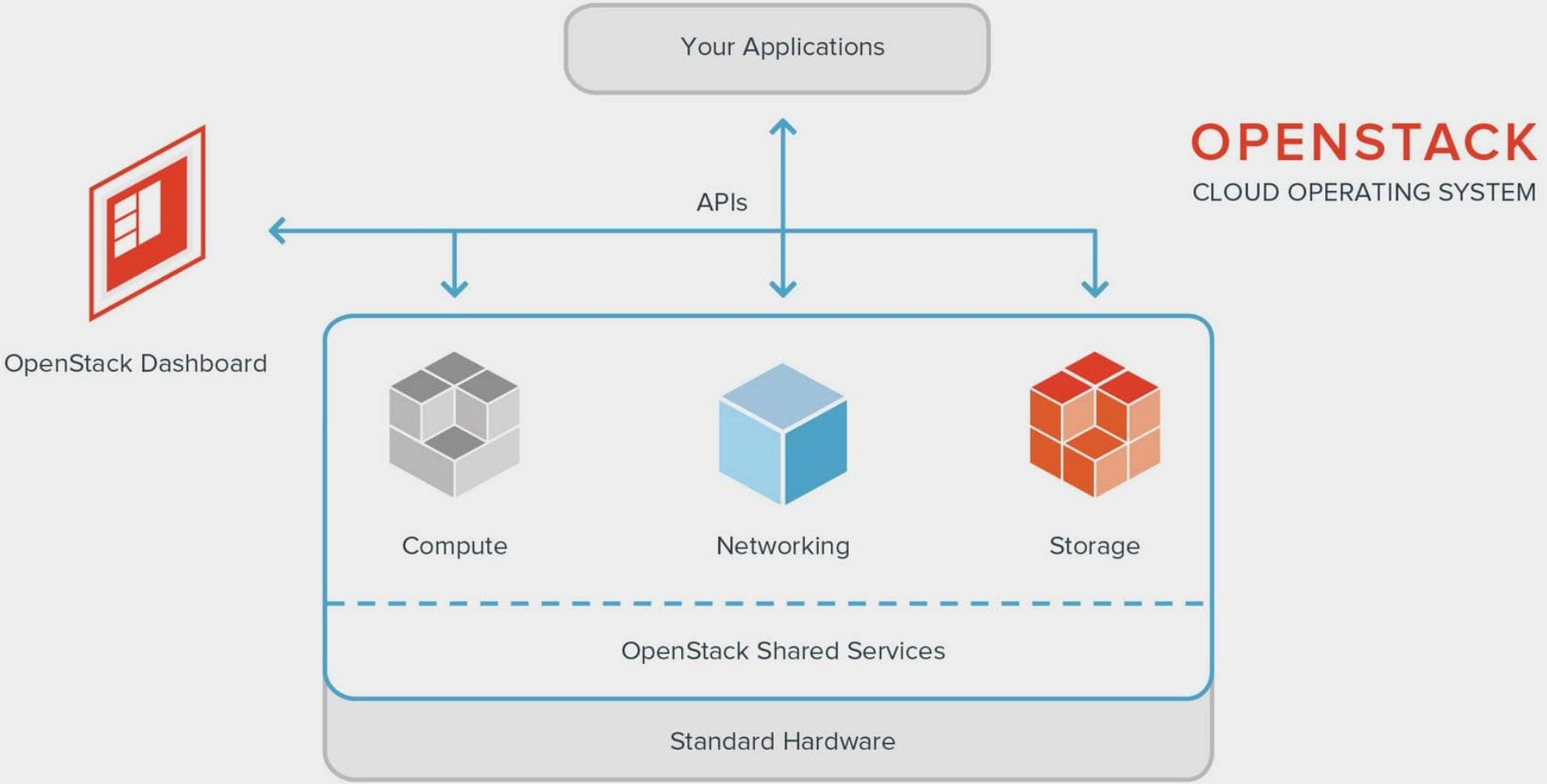
Red Hat

What is OpenStack

OpenStack is a cloud operating system that controls large pools of compute, storage, and networking resources throughout a datacenter, all managed and provisioned through APIs with common authentication mechanisms.

Red Hat

OpenStack



Red Hat

Deployment

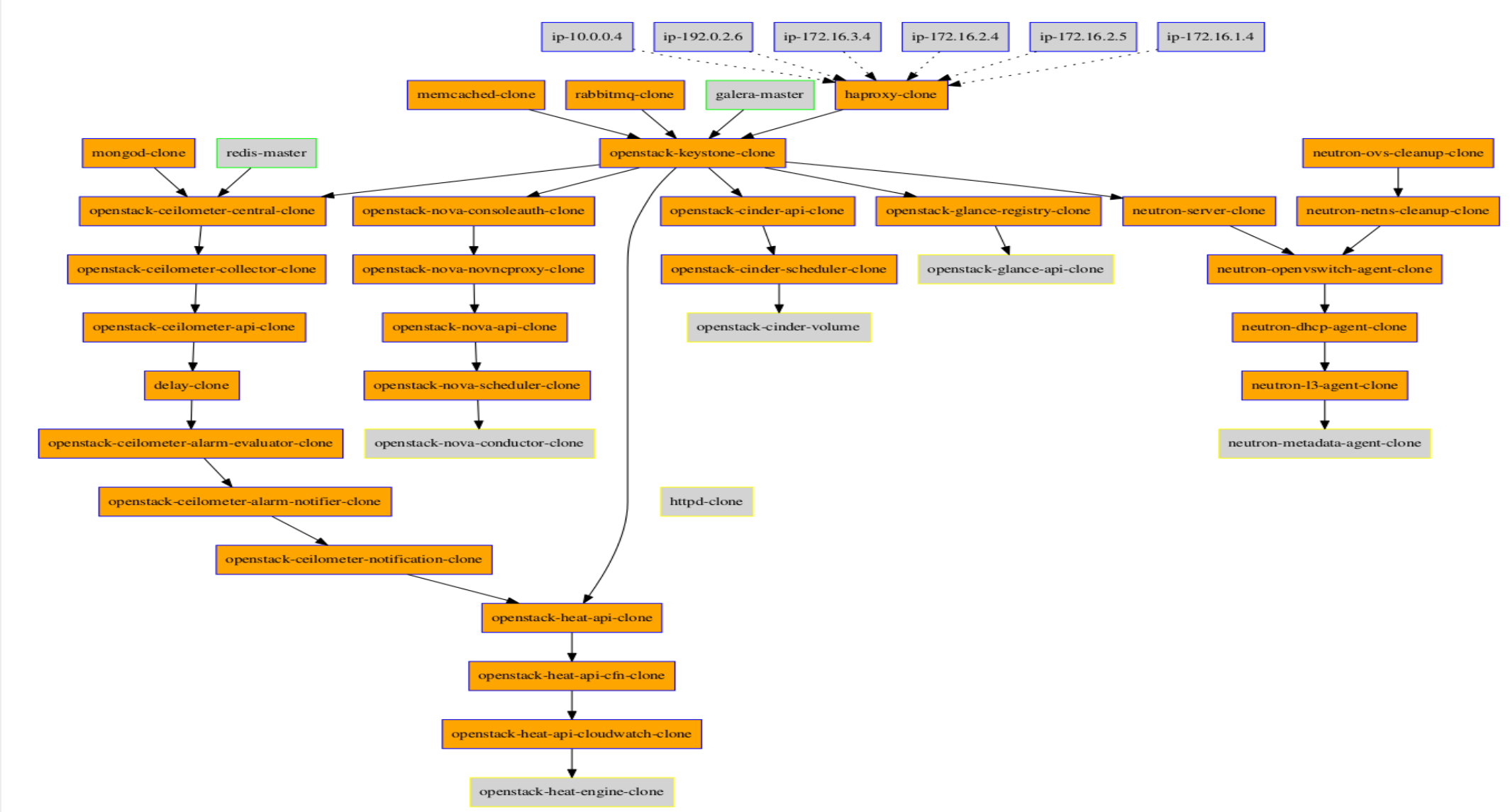
- Deployment is done via TripleO
- Reuses OpenStack components to deploy the Openstack Cloud
- Undercloud/Overcloud
- Mostly puppet + shell, transitioning to ansible

Red Hat

Usage of Pacemaker and friends?

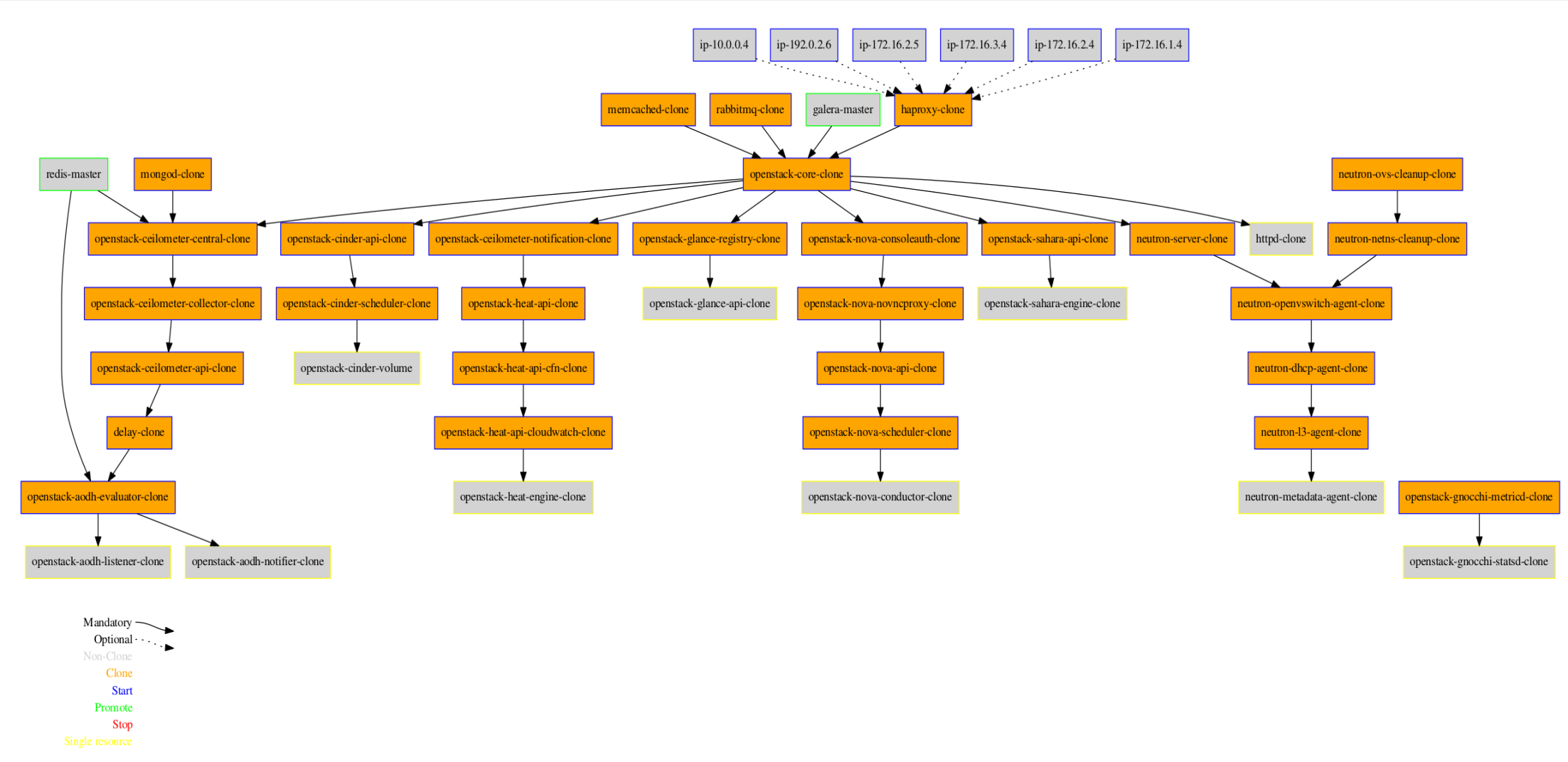
Red Hat

How we started (Liberty)

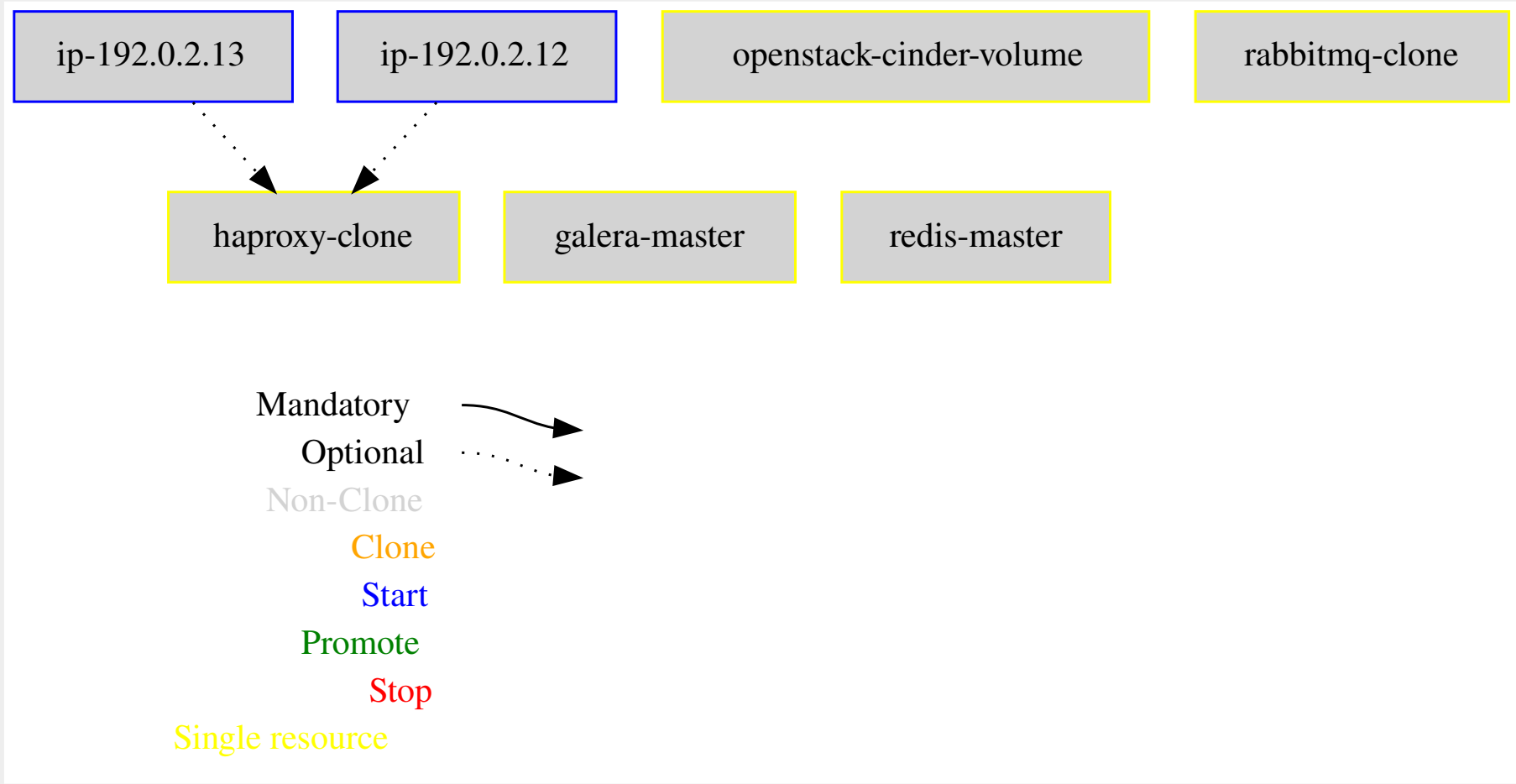


Red Hat

Maybe improved? (Mitaka)



Saner (Newton)



Red Hat

Composable HA (Ocata)

- Same as Newton
- But we can split off services to their own nodes

```
pcs property set --node controller-1 galera-
role=true

pcs constraint location galera rule resource-
discovery=exclusive score=0 galera-role eq true
```


Red Hat

Containers anyone?

Red Hat

Containerized (Pike, Queens, Rocky)

- Like in Ocata
- HA services run inside containers called "bundles" (aka containers in pcmk-speak)
- Except VIPs
- Uses docker daemon as the container engine (CentOS/RHEL 7.x)

Red Hat

Containerized (Pike, Queens, Rocky)

- There are two types of bundles:

1. Those with an OCF resource agent (+pcmk remote) inside

(Galera, rabbitmq, redis, ovn-dbs)

```
[root@controller-0 ~]# docker exec -it galera-bundle-docker-0 sh -c 'ps -ef'
UID          PID    PPID  C STIME TTY          TIME CMD
root           1       0  0 Jan31 ?        00:00:00 dumb-init -- /bin/bash /usr/local/bin/kolla_start
root           7       1  0 Jan31 ?        00:01:12 /usr/sbin/pacemaker_remoted
mysql        374       1  0 Jan31 ?        00:00:00 /bin/sh /usr/bin/mysqld_safe ...
mysql        678     374  1 Jan31 ?        00:57:04 /usr/libexec/mysqld --defaults-file=/etc/my.cnf ....
root      943696       7  0 09:56 ?        00:00:00 /bin/sh /usr/lib/ocf/resource.d/heartbeat/galera
monitor
```

(*) Need ‘touch /etc/libqb/force-filesystem-sockets’

Red Hat

Containerized (Pike, Queens, Rocky)

2. Simple bundles (haproxy, cinder-volume, ...)

```
[root@controller-0 ~]# docker exec -it haproxy-bundle-docker-0 sh -c 'ps -ef'
UID          PID    PPID  C STIME TTY          TIME CMD
root           1       0  0 Jan31 ?        00:00:00 dumb-init --single-child -- /bin/bash
/usr/local/bin/kolla_start
root           7       1  0 Jan31 ?        00:00:00 /usr/sbin/haproxy -f /etc/haproxy/haproxy.cfg -Ws
haproxy       14       7  0 Jan31 ?        00:22:04 /usr/sbin/haproxy -f /etc/haproxy/haproxy.cfg -Ws
```


Red Hat

Containerized (Stein, Train, ...)

- CentOS/RHEL 8.x
- Same as before but uses podman instead of docker
- No daemons involved
- s/docker/podman/g for most commands

Red Hat

So what does the cluster look like these days?

```
[root@controller-0 ~]# pcs status
Online: [ controller-0 controller-1 controller-2 ]
GuestOnline: [ galera-bundle-0@controller-0 galera-bundle-1@controller-1 galera-bundle-
2@controller-2 ovn-dbs-bundle-0@controller-0 ovn-dbs-bundle-1@controller-1 ovn-dbs-bundle-
2@controller-2 rabbitmq-bundle-0@controller-0 rabbitmq-bundle-1@controller-1 rabbitmq-bundle-
2@controller-2 redis-bundle-0@controller-0 redi
s-bundle-1@controller-1 redis-bundle-2@controller-2 ]

 Container bundle set: galera-bundle [cluster.common.tag/rhosp16-openstack-mariadb:pcmklatest]
   galera-bundle-0      (ocf::heartbeat:galera):        Master controller-0
   galera-bundle-1      (ocf::heartbeat:galera):        Master controller-1
   galera-bundle-2      (ocf::heartbeat:galera):        Master controller-2
 Container bundle set: rabbitmq-bundle [cluster.common.tag/rhosp16-openstack-
rabbitmq:pcmklatest]
   rabbitmq-bundle-0    (ocf::heartbeat:rabbitmq-cluster):      Started controller-0
   rabbitmq-bundle-1    (ocf::heartbeat:rabbitmq-cluster):      Started controller-1
   rabbitmq-bundle-2    (ocf::heartbeat:rabbitmq-cluster):      Started controller-2
 Container bundle set: redis-bundle [cluster.common.tag/rhosp16-openstack-redis:pcmklatest]
   redis-bundle-0       (ocf::heartbeat:redis): Master controller-0
   redis-bundle-1       (ocf::heartbeat:redis): Slave controller-1
   redis-bundle-2       (ocf::heartbeat:redis): Slave controller-2
 ip-192.168.24.38       (ocf::heartbeat:IPaddr2):       Started controller-0
 ip-10.0.0.101  (ocf::heartbeat:IPaddr2):       Started controller-1
 ip-172.17.1.84 (ocf::heartbeat:IPaddr2):       Started controller-0
 ip-172.17.1.18 (ocf::heartbeat:IPaddr2):       Started controller-1
 ip-172.17.3.122        (ocf::heartbeat:IPaddr2):       Started controller-0
 ip-172.17.4.90 (ocf::heartbeat:IPaddr2):       Started controller-1
Container bundle set: haproxy-bundle [cluster.common.tag/rhosp16-openstack-
haproxy:pcmklatest]
   haproxy-bundle-podman-0      (ocf::heartbeat:podman):        Started controller-0
   haproxy-bundle-podman-1      (ocf::heartbeat:podman):        Started controller-1
   haproxy-bundle-podman-2      (ocf::heartbeat:podman):        Started controller-2
 Container bundle set: ovn-dbs-bundle [cluster.common.tag/rhosp16-openstack-ovn-
northd:pcmklatest]
   ovn-dbs-bundle-0     (ocf::ovn:ovndb-servers):       Master controller-0
   ovn-dbs-bundle-1     (ocf::ovn:ovndb-servers):       Slave controller-1
   ovn-dbs-bundle-2     (ocf::ovn:ovndb-servers):       Slave controller-2
 ip-172.17.1.74 (ocf::heartbeat:IPaddr2):       Started controller-0
 stonith-fence_ipmilan-5254000dbdcc     (stonith:fence_ipmilan):        Started
controller-0
 stonith-fence_ipmilan-5254000c18dc     (stonith:fence_ipmilan):        Started
controller-0
 stonith-fence_ipmilan-5254004ed481     (stonith:fence_ipmilan):        Started
controller-1
 Container bundle: openstack-cinder-volume [cluster.common.tag/rhosp16-openstack-cinder-
volume:pcmklatest]
   openstack-cinder-volume-podman-0     (ocf::heartbeat:podman):        Started
controller-1
```


Red Hat

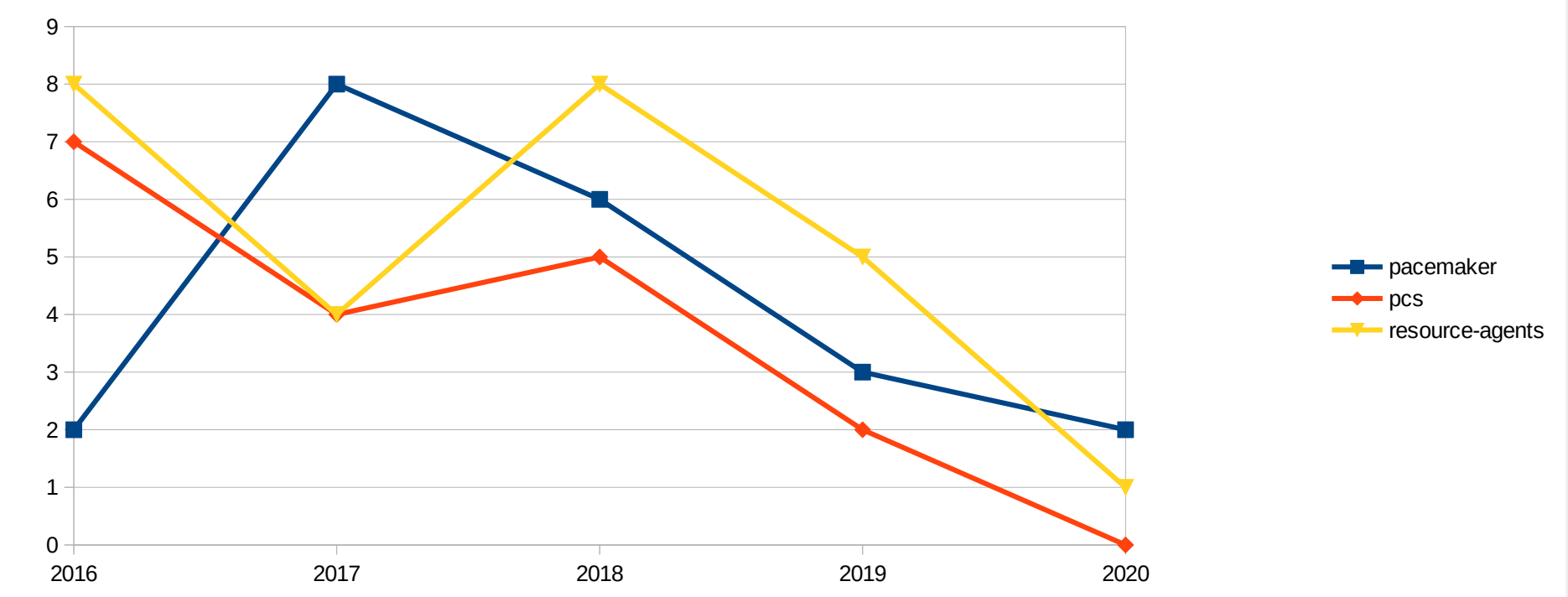
Thoughts

So from our POV how are things with the ClusterLabs projects in general?

Red Hat

The Good

- Pacemaker & co. are **stable** and very rarely regress



(We deal with openvswitch, ovn, ceph, podman, docker, baseos, puppet+modules, keepalived, ansible, python modules, haproxy, cloud-init, ..., so we **know**)

- Great support from HA team in general

The Ugly

Bundles support is a bit incomplete?

- Bundles with pcmk-remote require **perfect** sync of pacemaker binaries between host containers (breaks containers expectations, makes my life a misery)

- crm_<commands> inside a bundle with pcmk-remote are a lottery (best to be avoided?)
- Hard to try and bind mount things from the host (FHS)

Red Hat

The Bad

- Corosync breaking protocol. No cluster compat across versions
- Lack of a single script-focused APIs (single-status resource is finally coming, would love API/commands for rolling restarts, easy status scripting for resources)
- Bet there are X other projects doing/in need of our scripting functionality (e.g. Will this resource change make it restart in pcmk?)

Red Hat

The Nice to Have

- A means to safely guard against concurrent resource updates/restart (say TLS certificate changed for a resource)
- We're moving to ansible, so we will explore the existing ansible-pacemaker modules hoping not to write a new one?
- Per-bundle custom operations (timeouts, on-fail=block, etc.)
- Clear guidance on what --wait means for crm/pcs commands (or in general what guarantees do we have with CLI?)
- Changing secrets without having to restart the whole cluster

Questions?

Red Hat

Thank you

Never hassle the Hoff!

Red Hat