Deployment & Betrieb von Ceph mit (ceph-)ansible

Frühjahrsfachgespräch GUUG 2018

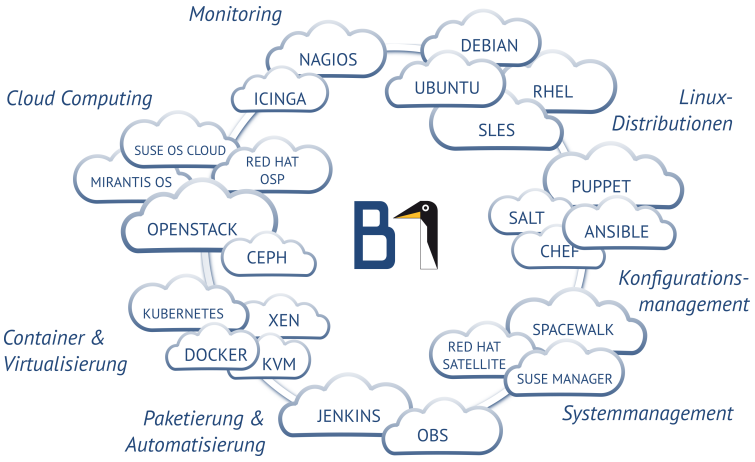
1. März 2018

Michel Raabe
Linux Consultant & Trainer
B1 Systems GmbH
raabe@b1-systems.de

Vorstellung B1 Systems

- gegründet 2004
- primär Linux/Open Source-Themen
- national & international tätig
- ca. 100 Mitarbeiter
- unabhängig von Soft- und Hardware-Herstellern
- Leistungsangebot:
  - Beratung & Consulting
  - Support
  - Entwicklung
  - Training
  - Betrieb
  - Lösungen
- Büros in Rockolding, Köln, Berlin & Dresden

Schwerpunkte



Ceph

Ceph: Distributed Storage

- Software-Defined-Storage
- Object, Block & File – alles in einem
- „...economical commodity hardware...“
- kein Single-Point-Of-Failure
- self-manage
- LTS Releases (Jewel, Luminous, Nautilus, ...)
- CERN, Telekom, SAP, Western Digital , Cisco, ...

Ceph: Daemons 1/2

Monitor

- 3 per Cluster
- Quorum
- Cluster/PG Status
- unabhängig von Daten/Tree

OSD

- 10s bis 1000s per Cluster
- pro Storage Device (SSD, HDD, NVMe)
- Clientzugriff
- Replikation

Ceph: Daemons 2/2

Weitere Daemons:

- Ceph Manager – mgr
- Ceph Metadata Server – mds
- Radow Gateway (S3) – rgw

Ceph: Deployment

- ceph-deploy
- ceph Puppet Module
- Salt (z.B. DeepSea)
- kolla-ansible
- ceph-ansible

ceph-ansible

ceph-ansible

- Github Project
- stable-3.0 (v3.0.26)
- Ceph Luminous ab stable-2.2 Tag
- ansible-2.1/2.2.1
- `site.yml` oder `site-docker.yml`
- Cluster via Inventory: mons, osds, rgws, ...

ceph-ansible: Inventory

```
[mons]
ceph-mon1
ceph-mon2
ceph-mon3
[mgrs]
ceph-mon1
ceph-mon2
ceph-mon3
[ods]
ceph-osd1
ceph-osd2
ceph-osd3
```

Groups:

- mons, agents, osds, mdss, rgws, nfss, restapis, rbdmirrors, mgrs, ...

ceph-ansible: Groups

Ansible-spezifische Konfiguration:

```
group_vars/
    all.yml
    osds.yml
    clients.yml
    mons.yml


host_vars/
    $hostname.yml
```

ceph-ansible: config II - docker vs „normal“

site-docker.yml

```
containerized_deployment: true
```

site.yml

```
ceph_origin: repository
ceph_repository: community
ceph_stable_release: luminous
```

ceph-ansible: osds 1/3

```
osd_objectstore: bluestore || filestore

osd_scenario: collocated

devices:
    - /dev/vdb
    - /dev/vdc
```

ceph-ansible: osds 2/3

```
osd_objectstore: bluestore

osd_scenario: non-collocated

devices:
    - /dev/vdb
    - /dev/vdc
dedicated_devices:
    - /dev/vdd
    - /dev/vdd
bluestore_wal_devices: "{{ dedicated_devices }}"
```

ceph-ansible: osds 3/3

```
osd_objectstore: filestore (!)

osd_scenario: lvm

lvm_volume:
 - data: datalv1
   data_vg: datavg1
   journal: journallv1
   journal_vg: journalvg1
 - data: datalv2
   data_vg: datavg2
   journal: /dev/vdd
```

ceph-ansible: pools 1/2

mons.yml

```
openstack_config: true
openstack_glance_pool:
    name: images
    pg_num: "{{ osd_pool_default_pg_num }}"
    rule_name: ""

openstack_pools:
  -   "{{ openstack_glance_pool }}"
openstack_keys:
  -   { name: client.glance, [...], mode: "0600", acls: [] }
```

- `openstack_config: true` aktiviert vordefinierte Openstack Keys (cinder, glance, openstack, ...)
- Ermitteln von `osd_pool_default_pg_num` per:

  `# ceph daemon mon.<name> config get osd_pool_default_pg_num`

ceph-ansible: pools 2/2 [1]

inventory

```
[clients]
$hostname
```

clients.yml

```
user_config: true
copy_admin_key: true
pools:
- { name: foobar, pgs: "{{ ceph_conf_overrides.global.osd_pool_default_pg_num }}" }
keys:
- { name: client.foobar, key: "#KEY", mon_cap: "allow r", osd_cap: "allow *", mode: "0600", acls: [] }
```

[1] Nicht für Container Deployments!

ceph-ansible: (crush-)rules

Nur via `create-simple`, kein `create-replicated` !

```
crush_rule_config: true
crush_rule_ssd:
    name: SSD
    root: SSD
    type: host
    default: false
crush_rules:
 - "{{ crush_rule_ssd }}"
```

ceph-ansible: (osd-)crush-location [2]

- Konfiguration über Inventory
- flexibel pro Host
- per default nur `rack`:

```
osd_crush_location: "\"root={{ ceph_crush_root }} rack={{ ceph_crush_rack }} host={{ ansible_hostname }}\""
```

osds.yml

```
crush_location: true
```

Inventory

```
[osds]
osd0 ceph_crush_root=default ceph_crush_rack=rack1
```

[2] Nicht für Container Deployments!

ceph-ansible: bads

- fetch Verzeichnis
- ceph-keys
- ceph user/group fehlt
- Containerized Deployment

Vielen Dank für Ihre Aufmerksamkeit!

Bei weiteren Fragen wenden Sie sich bitte an info@b1-systems.de oder +49 (0)8457 - 931096