SUSE® Enterprise Storage 6 on 华为 泰山 Implementation Guide

SUSE® Enterprise Storage 6 on 华为 泰山 Implementation Guide

Kai Liu

出版日期：Mar 15 2020

SUSE LLC
10 Canal Park Drive
Suite 200
Cambridge MA 02141
USA
https://documentation.suse.com

目录

1 Tips

1.1 Use a different NTP server

The default time server is the admin node. To change it, add

```
time_server: <server address>
```

in cluster.yml

1.2 Copy files to all cluster nodes

`salt-cp` command can be used to copy files from the salt master node to minion nodes. This can be very convenient, for example, to keep /etc/hosts file in sync on all nodes.

```
salt-cp '*' /etc/hosts /etc/hosts
```

1.3 Important files

/etc/salt/minion
: Salt minion configuration file

/etc/salt/minion_id
: Salt minion name. Useful if changed host name and need to change minion name accordingly.

/srv/pillar/ceph/deepsea_minions.sls
: Deepsea minion targets

/srv/pillar/ceph/stack/ceph/cluster.yml
: Deepsea cluster configuration for the cluster "ceph" (the default cluster name). Need to run stage 2 configure after modification. Or refresh pillar and check:

```
# salt <target> saltutil.pillar_refresh
# salt <target> pillar.items
```

CLUSTER CONFIGURATION FILES

/srv/pillar/ceph/stack/global.yml
: Affects all minions in the Salt cluster.

/srv/pillar/ceph/stack/ceph/cluster.yml
: Affects all minions in the cluster named "ceph".

/srv/pillar/ceph/stack/ceph/roles/role.yml
: Affects all minions that are assigned the specific role in the ceph cluster.

/srv/pillar/ceph/stack/cephminions/<minion ID>/yml
: Affects the individual minion.

1.4 How to completely uninstall the cluster for reinstall

In case you did something wrong and would like to start over without re-installing the whole OS.

```
# salt-run disengage.safety
# salt-run state.orch ceph.purge
```

1.5 Erase OSD drives

Sometimes it's necessary to completely erase the information left from other programs or previous installtions.

Besides of removing the partition table, the first few hundred MBs must also be cleared because that's the Bluestore small partition that has a filesystem, which holds some key information of that Bluestore disk. If it is not erased, the filesystem will be intact and it still holds the old information so reinstall will fail.

```
salt '<osd target>' cmd.run 'for d in {<start letter>..<end letter>}; do ceph-
disk zap /dev/sd$d; dd if=/dev/zero of=/dev/sd$d bs=1M count=500; done'
```

1.6 How to get salt pillar information

```
# salt '*' pillar.items
```

This can only give information after running stage 1 AKA the discovery stage

1.7 SES built-in network benchmark

https://www.suse.com/documentation/suse-enterprise-storage-5/singlehtml/book_storage_admin/book_storage_admin.html#storage.bp.performance.net_issues

```
# salt-run net.iperf cluster=ceph output=full
```

1.8 Ceph built-in OSD benchmark

https://www.suse.com/documentation/suse-enterprise-storage-5/singlehtml/book_storage_admin/book_storage_admin.html#storage.bp.performance.slowosd

```
# ceph tell osd.<id> bench
```

1.9 Ceph built-in pool scope benchmark

```
# rados -p <pool name> bench 60 write
```

1.10 Interface bonding

Use following parameters for bonding in 802.3ad mode (need switch support).
mode=802.3ad miimon=100 lacp_rate=fast xmit_hash_policy=layer3+4

Recommended Size for the BlueStore's WAL and DB Device https://www.suse.com/documentation/suse-enterprise-storage-5/singlehtml/book_storage_deployment/book_storage_deployment.html#about.bluestore

https://www.suse.com/documentation/suse-enterprise-storage-5/singlehtml/book_storage_deployment/book_storage_deployment.html#rec.waldb.size

1.11 Offline setup

Setup a SMT or RMT server, and mirror below repositories from SCC.

- SLE-Product-SLES15-SP1-Pool
- SLE-Product-SLES15-SP1-Updates
- SLE-Module-Server-Applications15-SP1-Pool
- SLE-Module-Server-Applications15-SP1-Updates
- SLE-Module-Basesystem15-SP1-Pool
- SLE-Module-Basesystem15-SP1-Updates
- SUSE-Enterprise-Storage-6-Pool
- SUSE-Enterprise-Storage-6-Updates

Then point all nodes to the SMT/RMT server.

1.12 Change node roles

After change of node roles by editing policy.cfg, need to run Stage 2 Configure to refresh configuration files.

```
# deepsea stage run ceph.stage.2
```

1.13 More tips

Check the SES 6 Administration Guide (https://documentation.suse.com/ses/6/single-html/ses-admin/#part-troubleshooting) for more hints & tips, FAQ, and troubleshooting techniques.

2 参考信息

SUSE Enterprise Storage Technical Overview
https://www.suse.com/media/white-paper/suse_enterprise_storage_technical_overview_wp.pdf

SUSE Enterprise Storage Tech Specs
https://www.suse.com/products/suse-enterprise-storage/#tech-specs

SUSE Enterprise Storage 6 - Release Notes
https://www.suse.com/releasenotes/x86_64/SUSE-Enterprise-Storage/6/

SUSE Enterprise Storage 6 - Deployment Guide
https://documentation.suse.com/ses/6/single-html/ses-deployment/#book-storage-deployment

SUSE Enterprise Storage 6 - Administration Guide
https://documentation.suse.com/ses/6/single-html/ses-admin/#book-storage-admin

SUSE Linux Enterprise Server 15 SP1 - Deployment Guide
https://documentation.suse.com/sles/15-SP1/single-html/SLES-deployment/#book-sle-deployment

SUSE Linux Enterprise Server 15 SP1 - Administration Guide
https://documentation.suse.com/sles/15-SP1/single-html/SLES-admin/#book-sle-admin

SUSE Linux Enterprise Server 15 SP1 - Storage Administration Guide
https://documentation.suse.com/sles/15-SP1/single-html/SLES-storage/#book-storage

SUSE Linux Enterprise Server 15 SP1 - Repository Mirroring Tool Guide
https://documentation.suse.com/sles/15-SP1/single-html/SLES-rmt/#book-rmt

A 设备清单

作用	数量	组件	说明
Admin Node	1	A VM on a 华为 x86 machine	The node consists of: • 8x vCPU cores • 64GB RAM • 100GB virtual disk for OS • 2x virtual NIC port to public and cluster network
OSD hosts, MON, MGR shared the OSD hosts	4	华为 泰山 5280	Each node consists of: • 2x Kunpeng 920 • 256GB • 1x Kunpeng 920 Integrated SAS Controller • 2x 300GB SAS HDD for OS • 34x 4TB 7.2k SATA HDD for OSD • 2x 华为 1TB NVMe SSD for db and journal • 1x Kunpeng 920 Embedded Network Controller - 1GbE Quad Port • 1x Dual Port Intel 82599 10Gb Ethernet adapter

作用	数量	组件	说明
Software	1	SUSE Enterprise Storage 6 Subscription Base configuration	Allows for 4 storage nodes and 6 infrastructure nodes
Switches	2	华为 CE6851-48S6Q-HI	32 Ports of 100GbE

B policy.cfg example

```
## Cluster Assignment
cluster-ceph/cluster/*.sls

## Roles
# ADMIN
role-master/cluster/admin*.sls
role-admin/cluster/admin*.sls

# Monitoring
role-prometheus/cluster/admin*.sls
role-grafana/cluster/admin*.sls

# MON
role-mon/cluster/ceph[123]*.sls

# MGR (mgrs are usually colocated with mons)
role-mgr/cluster/ceph[123]*.sls

# MDS
role-mds/cluster/ceph2*.sls

# IGW
role-igw/cluster/ceph3*.sls

# RGW
role-rgw/cluster/ceph4*.sls

# NFS
# role-ganesha/cluster/ganesha*.sls

# COMMON
config/stack/default/global.yml
config/stack/default/ceph/cluster.yml

# Storage
```

```
role-storage/cluster/ceph[1234]*.sls
```

C drive_groups.yml example

```
default:
  target: 'I@roles:storage'
  data_devices:
    # Use all hard disks as data device
    rotational: 1
  db_devices:
    # Use solid state drives as db device
    rotational: 0
```