SUSE Enterprise Storage 7.1

Guia de Implantação

Guia de Implantação

SUSE Enterprise Storage 7.1

por Tomáš Bažant, Alexandra Settle, e Liam Proven

Data de Publicação: 20/03/2025

https://documentation.suse.com

Copyright © 2020–2025 SUSE LLC e colaboradores. Todos os direitos reservados.

Exceto quando especificado de outra forma, este documento está licenciado nos termos da Creative Commons Attribution-ShareAlike 4.0 International (CC-BY-SA 4.0): https://creativecommons.org/licenses/by-sa/4.0/legalcode.

Para ver as marcas registradas da SUSE, visite http://www.suse.com/company/legal/. Todas as marcas registradas de terceiros pertencem aos seus respectivos proprietários. Os símbolos de marca registrada (®, ™ etc.) indicam as marcas registradas da SUSE e de suas afiliadas. Os asteriscos (*) indicam marcas registradas de terceiros.

Todas as informações deste manual foram compiladas com a maior atenção possível aos detalhes. Entretanto, isso não garante uma precisão absoluta. A SUSE LLC, suas afiliadas, os autores ou tradutores não serão responsáveis por possíveis erros nem pelas consequências resultantes de tais erros.

Conteúdo

Sobre este guia ix

1 Documentação disponível ix

2 Inserindo comentários x

3 Convenções da documentação xi

4 Suporte xiii

Declaração de suporte do SUSE Enterprise Storage xiii • Prévias de tecnologia xiv

5 Colaboradores do Ceph xv

6 Comandos e prompts de comando usados neste guia xv

Comandos relacionados ao Salt xv • Comandos relacionados ao Ceph xvi • Comandos gerais do Linux xvii • Informações adicionais xvii

I APRESENTANDO O SUSE ENTERPRISE STORAGE (SES) 1

1 SES e Ceph 2

1.1 Recursos do Ceph 2

1.2 Componentes básicos do Ceph 3

RADOS 3 • CRUSH 4 • Nós e daemons do Ceph 5

1.3 Estrutura de armazenamento do Ceph 6

Pools 6 • Grupos de posicionamento 7 • Exemplo 7

1.4 BlueStore 9

1.5 Informações adicionais 10

2 Requisitos e recomendações de hardware 11

2.1 Visão geral da rede 11

Recomendações de rede 12

2.2 Configurações de várias arquiteturas 14

2.3 Configuração de hardware 15

Configuração de cluster mínima 15 • Configuração recomendada do cluster de produção 17 • Configuração de múltiplos caminhos 18

2.4 Nós de armazenamento de objetos 19

Requisitos mínimos 19 • Tamanho mínimo do disco 20 • Tamanho recomendado para o dispositivo WAL e BD do BlueStore 20 • SSD para partições WAL/BD 21 • Número máximo recomendado de discos 22

2.5 Nós do monitor 22

2.6 Nós do Gateway de Objetos 23

2.7 Nós do servidor de metadados 23

2.8 Nó de Admin 23

2.9 Nós do Gateway iSCSI 23

2.10 SES e outros produtos da SUSE 24

SUSE Manager 24

2.11 Limitações de nome 24

2.12 OSD e monitor que compartilham um servidor 24

3 Configuração de alta disponibilidade do nó de admin 26

3.1 Estrutura do cluster de HA para o nó de admin 26

3.2 Criando um cluster de HA com o nó de admin 27

II IMPLANTANDO O CLUSTER DO CEPH 29

4 Introdução e tarefas comuns 30

4.1 Ler os detalhes da versão 30

5 Instalando e configurando o SUSE Linux Enterprise Server 32

6 Implantando o Salt 33

7 Implantando o cluster de boot usando `ceph-salt` 36

7.1 Instalando `ceph-salt` 36

7.2 Configurando as propriedades do cluster 36
Usando o shell do `ceph-salt` 37 • Adicionando minions Salt 39 • Especificando minions Salt gerenciados pelo cephadm 40 • Especificando o nó de admin 40 • Especificando o primeiro nó MON/MGR 41 • Especificando perfis ajustados 41 • Gerando um par de chaves SSH 41 • Configurando o servidor de horário 42 • Configurando as credenciais de login do Ceph Dashboard 43 • Usando o registro do container 43 • Habilitando a criptografia de dados em trânsito (msgr2) 49 • Configurando a rede do cluster 52 • Verificando a configuração do cluster 52 • Exportando as configurações do cluster 53

7.3 Atualizando os nós e o cluster mínimo de boot 54

7.4 Revisando as etapas finais 55

7.5 Desabilitar clientes não seguros 57

8 Implantando os serviços principais restantes com o cephadm 58

8.1 O comando **`ceph orch`** 58
Exibindo o status do orquestrador 58 • Listando dispositivos, serviços e daemons 58

8.2 Especificação de serviço e posicionamento 59
Criando especificações de serviço 60 • Criando a especificação de posicionamento 61 • Aplicando a especificação de cluster 61 • Exportando a especificação de um cluster em execução 62

8.3 Implantar serviços do Ceph 62
Implantando Ceph Monitors e Ceph Managers 63 • Implantando Ceph OSDs 64 • Implantando servidores de metadados 65 • Implantando Gateways de Objetos 66 • Implantando Gateways iSCSI 69 • Implantando o NFS Ganesha 70 • Implantação `rbd-mirror` 71 • Implantando a pilha de monitoramento 72

9 Implantação de serviços adicionais 74

9.1 Instalação do gateway iSCSI 74
Armazenamento em blocos iSCSI 74 • Informações gerais sobre `ceph-iscsi` 76 • Considerações de implantação 78 • Instalação e configuração 78 • Exportando imagens de dispositivo de blocos RADOS por meio do `tcmu-runner` 86

III FAZENDO UPGRADE DE VERSÕES ANTERIORES 88

10 Upgrade do SUSE Enterprise Storage 6 para 7.1 89

10.1 Antes de fazer upgrade 89
Pontos a serem considerados 90 • Fazendo backup da configuração e dos dados do cluster 91 • Verificando etapas do upgrade anterior 92 • Atualizando os nós e verificando a saúde do cluster 92 • Verificando o acesso a repositórios do software e imagens de container 93

10.2 Fazendo upgrade do Master Salt 94

10.3 Fazendo upgrade dos nós MON, MGR e OSD 96

10.4 Fazendo upgrade de nós de gateway 98

10.5 Instalando o `ceph-salt` e aplicando a configuração do cluster 98

10.6 Fazendo upgrade e adotando a pilha de monitoramento 100

10.7 Reimplantação do serviço de gateway 103
Fazendo upgrade do Gateway de Objetos 103 • Fazendo upgrade do NFS Ganesha 104 • Fazendo upgrade do servidor de metadados 108 • Fazendo upgrade do iSCSI Gateway 109

10.8 Limpeza após o upgrade 110

11 Upgrade do SUSE Enterprise Storage 7 para 7.1 112

11.1 Antes de fazer upgrade 112

Pontos a serem considerados 112 • Fazendo backup da configuração e dos dados do cluster 113 • Verificando o acesso a repositórios do software e imagens de container 113

11.2 Migrar o SUSE Linux Enterprise Server em cada nó do cluster para a versão SUSE Linux Enterprise Server 15 SP3 114

11.3 Atualizar os pacotes relacionados ao SUSE Enterprise Storage em cada nó do cluster 115

11.4 Fazer upgrade dos serviços de cluster existentes do Ceph 115

A Atualizações de manutenção do Ceph baseadas nos point releases de upstream do “Pacific” 116

Glossário 117

Sobre este guia

O foco deste guia é a implantação de um cluster básico do Ceph e de serviços adicionais. Ele também abrange as etapas de upgrade da versão anterior do produto para o SUSE Enterprise Storage 7.1.

SUSE Enterprise Storage 7.1 é uma extensão do SUSE Linux Enterprise Server 15 SP3. Ele combina os recursos do projeto de armazenamento do Ceph (http://ceph.com/) com a engenharia corporativa e o suporte da SUSE. O SUSE Enterprise Storage 7.1 oferece às organizações de TI o recurso para implantar uma arquitetura de armazenamento distribuído capaz de suportar inúmeros casos de uso por meio de plataformas de hardware convencional.

1 Documentação disponível

Nota: Documentação online e atualizações mais recentes

A documentação para nossos produtos está disponível no site https://documentation.suse.com, onde você também pode encontrar as atualizações mais recentes e procurar ou fazer download da documentação em vários formatos. As atualizações mais recentes da documentação estão disponíveis em inglês.

Além disso, a documentação do produto está disponível no seu sistema instalado em `/usr/share/doc/manual`. Ela está incluída em um pacote RPM chamado `ses-manual_LANG_CODE`. Instale esse pacote caso ainda não esteja em seu sistema, por exemplo:

```
# zypper install ses-manual_en
```

A seguinte documentação está disponível para este produto:

Guia de Implantação **(https://documentation.suse.com/ses/html/ses-all/book-storage-deployment.html)**

O foco deste guia é a implantação de um cluster básico do Ceph e de serviços adicionais. Ele também abrange as etapas de upgrade para o SUSE Enterprise Storage 7.1 a partir da versão anterior do produto.

Guia de Administração e Operações (https://documentation.suse.com/ses/html/ses-all/book-storage-admin.html)

O foco deste guia são as tarefas de rotina que você, como administrador, precisa realizar após a implantação do cluster básico do Ceph (operações do dia 2). Ele também descreve todos os meios possíveis para acessar os dados armazenados em um cluster do Ceph.

Guia de Reforço da Segurança (https://documentation.suse.com/ses/html/ses-all/book-storage-security.html)

O foco deste guia é a garantia da segurança do cluster.

Troubleshooting Guide (Guia de Solução de Problemas) (https://documentation.suse.com/ses/html/ses-all/book-storage-troubleshooting.html)

Este guia aborda os vários problemas comuns durante a execução do SUSE Enterprise Storage 7.1 e outras questões relacionadas a componentes relevantes, como o Ceph ou o Gateway de Objetos.

Guia do SUSE Enterprise Storage para Windows (https://documentation.suse.com/ses/html/ses-all/book-storage-windows.html)

Este guia descreve a integração, a instalação e a configuração dos ambientes Microsoft Windows e SUSE Enterprise Storage que usam o Driver do Windows.

2 Inserindo comentários

Ficamos satisfeitos com seus comentários e contribuições para esta documentação. E você pode contribuir por meio de vários canais:

Solicitações de serviço e suporte

Para conhecer os serviços e as opções de suporte disponíveis para o seu produto, consulte http://www.suse.com/support/.

Para abrir uma solicitação de serviço, você precisa de uma assinatura do SUSE registrada no SUSE Customer Center. Vá para https://scc.suse.com/support/requests, efetue login e clique em *Criar Novo*.

Relatórios de bugs

Relate os problemas com a documentação em https://bugzilla.suse.com/. A geração de relatórios de problemas requer uma conta do Bugzilla.

Para simplificar esse processo, você pode usar os links *Relatar Bug na Documentação* ao lado dos cabeçalhos na versão HTML deste documento. Isso pré-seleciona o produto e a categoria certos no Bugzilla e adiciona um link à seção atual. Você pode começar a digitar o relatório do bug imediatamente.

Contribuições

Para contribuir com esta documentação, use os links *Editar Fonte* ao lado dos cabeçalhos na versão HTML deste documento. Eles levarão você até o código-fonte no GitHub, onde é possível abrir uma solicitação pull. A contribuição requer uma conta do GitHub.

Para obter mais informações sobre o ambiente da documentação usado para este documento, consulte o README do repositório em https://github.com/SUSE/doc-ses.

E-mail

É possível também relatar erros e enviar comentários sobre a documentação para `doc-team@suse.com`. Inclua o título do documento, a versão do produto e a data de publicação do documento. Mencione também o número e o título da seção relevante (ou inclua o URL) e insira uma breve descrição do problema.

3 Convenções da documentação

Os seguintes avisos e convenções tipográficas são usados nesta documentação:

- `/etc/passwd`: Nomes de diretório e arquivo
- *`PLACEHOLDER`*: Substitua *`PLACEHOLDER`* pelo valor real
- `PATH`: Uma variável de ambiente
- **`ls`**, `--help`: Comandos, opções e parâmetros
- `user`: O nome do usuário ou grupo
- `package_name`: O nome de um pacote de software
- `Alt` , `Alt`–`F1` : Uma tecla para pressionar ou uma combinação de teclas. As teclas são mostradas em maiúsculas como no teclado.
- *Arquivo*, *Arquivo* › *Gravar Como* : itens de menu, botões
- `AMD/Intel` Este parágrafo é relevante apenas para as arquiteturas AMD64/Intel 64. As setas marcam o início e o fim do bloco de texto.

`IBM Z, POWER` Este parágrafo é relevante apenas para as arquiteturas `IBM Z` e `POWER`. As setas marcam o início e o fim do bloco de texto.

- *Capítulo 1, "Capítulo de exemplo"*: Uma referência cruzada a outro capítulo deste guia.
- Comandos que devem ser executados com privilégios `root`. Geralmente, você também pode usar o comando **`sudo`** como prefixo nesses comandos para executá-los como usuário não privilegiado.

```
# command
> sudo command
```

- Comandos que podem ser executados por usuários sem privilégios.

```
> command
```

- Avisos

Atenção: Mensagem de aviso

Informações vitais que você deve saber antes de continuar. Avisa sobre problemas de segurança, potencial perda de dados, danos no hardware ou perigos físicos.

Importante: Aviso importante

Informações importantes que você deve saber antes de continuar.

Nota: Nota

Informações adicionais, por exemplo, sobre diferenças nas versões do software.

Dica: Aviso de dica

Informações úteis, como uma diretriz ou informação prática.

- Avisos compactos

Informações adicionais, por exemplo, sobre diferenças nas versões do software.

Informações úteis, como uma diretriz ou informação prática.

4 Suporte

Encontre a declaração de suporte do SUSE Enterprise Storage e as informações gerais sobre as prévias de tecnologia a seguir. Para obter detalhes sobre o ciclo de vida do produto, consulte https://www.suse.com/lifecycle.

Se você tiver direito a suporte, encontre os detalhes de como coletar informações para um ticket de suporte em https://documentation.suse.com/sles-15/html/SLES-all/cha-adm-support.html.

4.1 Declaração de suporte do SUSE Enterprise Storage

Para receber suporte, você precisa de uma inscrição apropriada na SUSE. Para ver as ofertas de suporte específicas que estão disponíveis para você, acesse https://www.suse.com/support/ e selecione seu produto.

Os níveis de suporte são definidos da seguinte forma:

L1

Determinação do problema, que significa suporte técnico designado para fornecer informações de compatibilidade, suporte ao uso, manutenção contínua, coleta de informações e solução básica de problemas usando a documentação disponível.

L2

Isolamento do problema, que significa suporte técnico designado para analisar os dados, reproduzir os problemas dos clientes, isolar a área problemática e resolver os problemas não resolvidos no Nível 1 ou preparar-se para o Nível 3.

L3

Resolução do problema, que significa suporte técnico designado para resolver os problemas com a participação da engenharia para solucionar defeitos nos produtos que foram identificados pelo Suporte de Nível 2.

Para clientes e parceiros contratados, o SUSE Enterprise Storage é entregue com suporte L3 para todos os pacotes, com as seguintes exceções:

- Prévias de tecnologia.
- Som, gráficos, fontes e arte.
- Pacotes que requerem um contrato de cliente adicional.
- Alguns pacotes enviados como parte do módulo *Workstation Extension* contam apenas com o suporte L2.
- Os pacotes com nomes que terminam em `-devel` (contendo arquivos de cabeçalho e recursos de desenvolvedor semelhantes) serão suportados apenas junto com seus pacotes principais.

A SUSE apenas oferecerá suporte ao uso dos pacotes originals. Isto é, pacotes que não foram modificados nem recompilados.

4.2 Prévias de tecnologia

As Prévias de tecnologia são pacotes, pilhas ou recursos fornecidos pela SUSE como amostras de inovações futuras. As prévias de tecnologia foram incluídas para sua conveniência e para que você possa testar as novas tecnologias em seu ambiente. Agradecemos seus comentários! Se você testar uma prévia de tecnologia, contate seu representante SUSE e conte sobre sua experiência e seus casos de uso. Suas informações são úteis para o desenvolvimento futuro.

As prévias de tecnologia têm as seguintes limitações:

- As prévias de tecnologia ainda estão em desenvolvimento. Portanto, elas podem ter funcionalidades incompletas, instáveis ou, de alguma outra maneira, *inadequadas* para uso em produção.
- As prévias de tecnologia *não* contam com suporte.
- As prévias de tecnologia talvez estejam disponíveis apenas para arquiteturas de hardware específicas.

- Os detalhes e as funcionalidades das prévias de tecnologia estão sujeitos a mudanças. Consequentemente, o upgrade para as versões subsequentes de uma prévia de tecnologia pode ser impossível e exigir uma instalação nova.
- A SUSE pode descobrir que uma prévia não atende às necessidades do cliente ou do mercado, ou não está em conformidade com os padrões da empresa. As prévias de tecnologia podem ser removidas de um produto a qualquer momento. A SUSE não se compromete em oferecer uma versão com suporte desse tipo de tecnologia no futuro.

Para obter uma visão geral das prévias de tecnologia fornecidas com seu produto, consulte as notas de lançamento em https://www.suse.com/releasenotes/x86_64/SUSE-Enterprise-Storage/7.1.

5 Colaboradores do Ceph

O projeto do Ceph e a respectiva documentação são resultados do trabalho de centenas de colaboradores e organizações. Visite https://ceph.com/contributors/ para obter mais detalhes.

6 Comandos e prompts de comando usados neste guia

Como administrador de cluster do Ceph, você configura e ajusta o comportamento do cluster executando comandos específicos. Há vários tipos de comandos que serão necessários:

6.1 Comandos relacionados ao Salt

Esses comandos ajudam você a implantar nós do cluster do Ceph, executar comandos em vários (ou todos) nós do cluster ao mesmo tempo ou a adicionar ou remover nós do cluster. Os comandos usados com mais frequência são **`ceph-salt`** e **`ceph-salt config`**. Você precisa executar os comandos do Salt no nó do Master Salt como `root`. Esses comandos são introduzidos com o seguinte prompt:

```
root@master #
```

Por exemplo:

```
root@master # ceph-salt config ls
```

6.2 Comandos relacionados ao Ceph

Esses são comandos de nível inferior para configurar e ajustar todos os aspectos do cluster e seus gateways na linha de comando, por exemplo, **`ceph`**, **`cephadm`**, **`rbd`** ou **`radosgw-admin`**.

Para executar os comandos relacionados ao Ceph, você precisa ter acesso de leitura a uma chave do Ceph. Os recursos da chave definem seus privilégios no ambiente do Ceph. Uma opção é executar os comandos do Ceph como `root` (ou por meio do **`sudo`**) e usar o chaveiro padrão irrestrito "ceph.client.admin.key".

A opção mais segura e recomendada é criar uma chave individual mais restritiva para cada usuário administrador e colocá-la em um diretório onde os usuários possam lê-la, por exemplo:

```
~/.ceph/ceph.client.USERNAME.keyring
```

Dica: Caminho para as chaves do Ceph

Para utilizar um usuário admin e um chaveiro personalizados, você precisa especificar o nome de usuário e o caminho para a chave toda vez que executar o comando **`ceph`** com as opções `-n client.USER_NAME` e `--keyring PATH/TO/KEYRING`.

Para evitar isso, inclua essas opções na variável `CEPH_ARGS` nos arquivos `~/.bashrc` dos usuários individuais.

É possível executar os comandos relacionados ao Ceph em qualquer nó do cluster, mas a recomendação é executá-los no Nó de Admin. Esta documentação utiliza o usuário `cephuser` para executar os comandos, portanto, eles são introduzidos com o seguinte prompt:

```
cephuser@adm >
```

Por exemplo:

```
cephuser@adm > ceph auth list
```

Dica: Comandos para nós específicos

Se a documentação orientar você a executar um comando em um nó do cluster com uma função específica, isso será processado pelo prompt. Por exemplo:

```
cephuser@mon >
```

6.2.1 Executando o **ceph-volume**

A partir do SUSE Enterprise Storage 7, os serviços do Ceph são executados em containers. Se você precisar executar o **ceph-volume** em um nó OSD, anteceda-o com o comando **cephadm**, por exemplo:

```
cephuser@adm > cephadm ceph-volume simple scan
```

6.3 Comandos gerais do Linux

Os comandos do Linux não relacionados ao Ceph, como **mount**, **cat** ou **openssl**, são introduzidos com os prompts `cephuser@adm >` ou `#`, dependendo dos privilégios exigidos pelo comando relacionado.

6.4 Informações adicionais

Para obter mais informações sobre o gerenciamento de chaves do Ceph, consulte o *Livro "Guia de Administração e Operações", Capítulo 30 "Autenticação com* cephx*", Seção 30.2 "As principais áreas de"*.

I Apresentando o SUSE Enterprise Storage (SES)

1 SES e Ceph 2

2 Requisitos e recomendações de hardware 11

3 Configuração de alta disponibilidade do nó de admin 26

1 SES e Ceph

O SUSE Enterprise Storage é um sistema de armazenamento distribuído projetado para escalabilidade, confiabilidade e desempenho que usa a tecnologia Ceph. É possível executar um cluster do Ceph em servidores convencionais em uma rede comum, como Ethernet. O cluster tem capacidade de dimensionamento para milhares de servidores (posteriormente mencionados como nós) e dentro da faixa de petabytes. Ao contrário dos sistemas convencionais que têm tabelas de alocação para armazenar e buscar dados, o Ceph usa um algoritmo determinístico para alocar armazenamento de dados e não tem nenhuma estrutura centralizada de informações. Nos clusters de armazenamento, o Ceph considera a adição ou remoção de hardware a regra, e não a exceção. O cluster do Ceph automatiza as tarefas de gerenciamento, como distribuição e redistribuição de dados, replicação de dados, detecção de falhas e recuperação. O Ceph é autorreparável e autogerenciável, o que resulta na redução de overhead administrativo e orçamentário.

Este capítulo apresenta uma visão geral resumida do SUSE Enterprise Storage 7.1 e descreve brevemente os componentes mais importantes.

1.1 Recursos do Ceph

O ambiente do Ceph tem os seguintes recursos:

Escalabilidade

O Ceph pode ser dimensionado para milhares de nós e pode gerenciar o armazenamento na faixa de petabytes.

Hardware convencional

Não há necessidade de hardware especial para executar um cluster do Ceph. Para obter os detalhes, consulte a *Capítulo 2, Requisitos e recomendações de hardware*

Autogerenciável

O cluster do Ceph é autogerenciável. Quando os nós são adicionados, removidos ou falham, o cluster redistribui os dados automaticamente. Ele também reconhece discos sobrecarregados.

Sem Ponto Único de Falha

Nenhum nó em um cluster armazena informações importantes separadamente. É possível configurar o número de redundâncias.

Software de código-fonte aberto

O Ceph é uma solução de software de código-fonte aberto e independente de hardware ou de fornecedores específicos.

1.2 Componentes básicos do Ceph

Para aproveitar todas as vantagens do Ceph, é necessário conhecer alguns dos componentes e conceitos básicos. Esta seção apresenta algumas partes do Ceph que são mencionadas com mais frequência em outros capítulos.

1.2.1 RADOS

O componente básico de Ceph é denominado *RADOS (Reliable Autonomic Distributed Object Store)*. Ele é responsável por gerenciar os dados armazenados no cluster. Normalmente, os dados no Ceph são armazenados como objetos. Cada objeto consiste em um identificador e nos dados.

O RADOS oferece os seguintes métodos de acesso aos objetos armazenados que envolvem diversos casos de uso:

Gateway de Objetos

O Gateway de Objetos é um gateway HTTP REST para armazenamento de objetos RADOS. Ele permite acesso direto aos objetos armazenados no cluster do Ceph.

Dispositivo de blocos RADOS

É possível acessar o Dispositivo de Blocos RADOS (RBD, RADOS Block Device) como qualquer outro dispositivo de blocos. Eles podem ser usados em combinação com o `libvirt` para fins de virtualização, por exemplo.

CephFS

O Sistema de Arquivos Ceph é compatível com POSIX.

`librados`

`librados` é uma biblioteca que pode ser usada com várias linguagens de programação para criar um aplicativo capaz de interagir diretamente com o cluster de armazenamento.

O Gateway de Objetos e o RBD usam a `librados`, enquanto o CephFS estabelece interface direta com o RADOS (*Figura 1.1, "Interfaces com o armazenamento de objetos do Ceph"*).

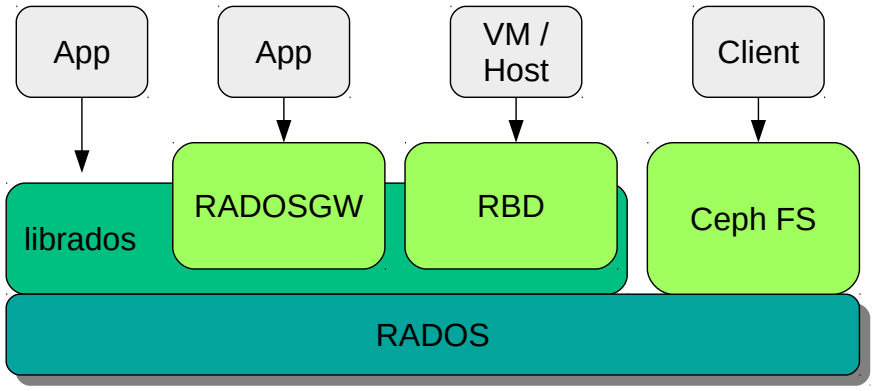

FIGURA 1.1: INTERFACES COM O ARMAZENAMENTO DE OBJETOS DO CEPH

1.2.2 CRUSH

No centro de um cluster do Ceph está o algoritmo *CRUSH*. CRUSH é o acrônimo de *Controlled Replication Under Scalable Hashing*. Trata-se de uma função que processa a alocação de armazenamento e precisa de poucos parâmetros equivalentes. Isso significa que apenas uma pequena quantidade de informações é necessária para calcular a posição de armazenamento de um objeto. Os parâmetros são um mapa atual do cluster, incluindo o estado de saúde, algumas regras de posicionamento definidas pelo administrador e o nome do objeto que precisa ser armazenado ou recuperado. Com essas informações, todos os nós no cluster do Ceph são capazes de calcular onde um objeto e suas réplicas serão armazenados. Isso dificulta a gravação e a leitura de dados muito eficientes. O CRUSH tenta distribuir igualmente os dados a todos os nós do cluster.

O *Mapa CRUSH* contém todos os nós de armazenamento e as regras de posicionamento definidas pelo administrador para armazenar objetos no cluster. Ele define uma estrutura hierárquica que costuma corresponder à estrutura física do cluster. Por exemplo, os discos com dados estão em hosts, os hosts estão em racks, os racks estão em linhas e as linhas estão em data centers. É possível usar essa estrutura para definir *domínios de falha*. Em seguida, o Ceph garante que as replicações sejam armazenadas em ramificações diferentes de um domínio de falha específico.

Se o domínio de falha for definido como rack, as replicações dos objetos serão distribuídas para racks diferentes. Isso pode reduzir as interrupções causadas por um switch com falha em um rack. Se uma unidade de distribuição de energia fornece uma linha de racks, o domínio de falha pode ser definido como linha. Quando a unidade de distribuição de energia falha, os dados replicados ainda ficam disponíveis em outras linhas.

1.2.3 Nós e daemons do Ceph

No Ceph, os nós são servidores que trabalham para o cluster. Eles podem executar vários tipos de daemons. Recomendamos executar apenas um tipo de daemon em cada nó, exceto os daemons do Ceph Manager, que podem ser combinados com Ceph Monitors. Cada cluster requer pelo menos os daemons do Ceph Monitor, Ceph Manager e Ceph OSD:

Nó de Admin

O *Nó de Admin* é um nó de cluster do Ceph do qual você executa comandos para gerenciar o cluster. O Nó de Admin é um ponto central do cluster do Ceph porque ele gerencia o restante dos nós do cluster, consultando e instruindo os serviços do Minion Salt.

Ceph Monitor

Os nós do *Ceph Monitor* (geralmente abreviado como *MON*) mantêm informações sobre o estado de saúde do cluster, um mapa de todos os nós e as regras de distribuição de dados (consulte a *Seção 1.2.2, "CRUSH"*).

Em caso de falhas ou conflitos, os nós do Ceph Monitor no cluster decidem, por maioria, quais informações estão corretas. Para formar uma maioria qualificada, é recomendável ter um número ímpar de nós do Ceph Monitor e, pelo menos, três deles.

Se for usado mais de um site, os nós do Ceph Monitor deverão ser distribuídos para um número ímpar de sites. O número de nós do Ceph Monitor por site deve ser suficiente para que mais do que 50% dos nós do Ceph Monitor continuem funcionando em caso de falha de um site.

Ceph Manager

O Ceph Manager coleta as informações de estado do cluster inteiro. O daemon do Ceph Manager é executado com os daemons do Ceph Monitor. Ele fornece monitoramento adicional e estabelece interface com os sistemas externos de monitoramento e gerenciamento. Ele também inclui outros serviços. Por exemplo, a IU da Web do Ceph Dashboard é executada no mesmo nó que o Ceph Manager.

O Ceph Manager não requer configuração adicional, além de garantir que esteja em execução.

Ceph OSD

O *Ceph OSD* é um daemon que processa *Dispositivos de Armazenamento de Objetos*, que são unidades físicas ou lógicas de armazenamento (discos rígidos ou partições). Os Dispositivos de Armazenamento de Objetos podem ser discos/partições físicos ou volumes lógicos. O daemon também se encarrega da replicação e redistribuição de dados em caso de nós adicionados ou removidos.

Os daemons Ceph OSD comunicam-se com os daemons do monitor e informam a eles o estado dos outros daemons OSD.

Para usar o CephFS, o Gateway de Objetos, o NFS Ganesha ou o iSCSI Gateway, são necessários outros nós:

MDS (Metadata Server – Servidor de Metadados)

Os metadados do CephFS são armazenados em seu próprio pool RADOS (consulte a *Seção 1.3.1, "Pools"*). Os Servidores de Metadados atuam como uma camada de cache inteligente para os metadados e serializam o acesso quando necessário. Isso permite acesso simultâneo de vários clientes sem sincronização explícita.

Gateway de Objetos

O Gateway de Objetos é um gateway HTTP REST para armazenamento de objetos RADOS. Ele é compatível com o OpenStack Swift e o Amazon S3 e tem seu próprio gerenciamento de usuários.

NFS Ganesha

O NFS Ganesha concede acesso de NFS ao Gateway de Objetos ou CephFS. Ele é executado no espaço do usuário, em vez do kernel, e interage diretamente com o Gateway de Objetos ou o CephFS.

iSCSI Gateway

iSCSI é um protocolo de rede de armazenamento que permite aos clientes enviar comandos SCSI para dispositivos de armazenamento SCSI (destinos) em servidores remotos.

Gateway do Samba

O Gateway do Samba concede ao Samba acesso aos dados armazenados no CephFS.

1.3 Estrutura de armazenamento do Ceph

1.3.1 Pools

Os objetos armazenados em um cluster do Ceph são agrupados em *pools*. Os pools representam as partições lógicas do cluster para o mundo externo. É possível definir um conjunto de regras para cada pool. Por exemplo, o número necessário de replicações de cada objeto. A configuração padrão dos pools chama-se *pool replicado*.

Normalmente, os pools contêm objetos, mas também podem ser configurados para agir como um RAID 5. Nessa configuração, os objetos são armazenados em pacotes com outros pacotes de codificação. Os pacotes de codificação contêm as informações redundantes. O número de dados e de pacotes de codificação pode ser definido pelo administrador. Nessa configuração, os pools são denominados *pools codificados para eliminação* ou *pools EC*.

1.3.2 Grupos de posicionamento

Os PGs (*Placement Groups – Grupos de Posicionamento*) são usados para distribuição de dados em um pool. Ao criar um pool, determinado número de grupos de posicionamento é definido. Os grupos de posicionamento são usados internamente para agrupar objetos e são um fator importante para o desempenho de um cluster do Ceph. O PG de um objeto é determinado pelo nome do objeto.

1.3.3 Exemplo

Esta seção mostra um exemplo simplificado de como o Ceph gerencia os dados (consulte a *Figura 1.2, "Exemplo de Ceph de pouco dimensionamento"*). Esse exemplo não representa uma configuração recomendada para um cluster do Ceph. A configuração de hardware consiste em três nós de armazenamento ou Ceph OSDs (`Host 1`, `Host 2`, `Host 3`). Cada nó tem três discos rígidos que são usados como OSDs (`osd.1` a `osd.9`). Os nós do Ceph Monitor são ignorados neste exemplo.

Nota: Diferença entre Ceph OSD e OSD

Enquanto *Ceph OSD* ou *daemon Ceph OSD* refere-se a um daemon executado em um nó, a palavra *OSD* refere-se ao disco lógico com o qual o daemon interage.

O cluster tem dois pools: `Pool A` e `Pool B`. Enquanto o Pool A replica os objetos apenas duas vezes, a resiliência do Pool B é mais importante e efetua três replicações de cada objeto.

Quando um aplicativo coloca um objeto em um pool (por exemplo, pela API REST), um Grupo de Posicionamento (`PG1` a `PG4`) é selecionado com base no nome do pool e do objeto. Em seguida, o algoritmo CRUSH calcula em quais OSDs o objeto é armazenado, com base no Grupo de Posicionamento que contém o objeto.

Neste exemplo, o domínio de falha foi definido como host. Isso garante que as replicações dos objetos sejam armazenadas em hosts diferentes. Dependendo do nível de replicação definido para um pool, o objeto é armazenado em dois ou três OSDs, que são usados pelo Grupo de Posicionamento.

Um aplicativo que grava um objeto interage apenas com um Ceph OSD: o principal. O Ceph OSD principal efetua a replicação e confirma a conclusão do processo de gravação depois que todos os outros OSDs armazenaram o objeto.

Em caso de falha no `osd.5`, todos os objetos no `PG1` ainda estarão disponíveis no `osd.1`. Logo que o cluster reconhece a falha em um OSD, outro OSD entra em ação. Neste exemplo, o `osd.4` é usado como substituto do `osd.5`. Os objetos armazenados no `osd.1` são replicados para o `osd.4` para restaurar o nível de replicação.

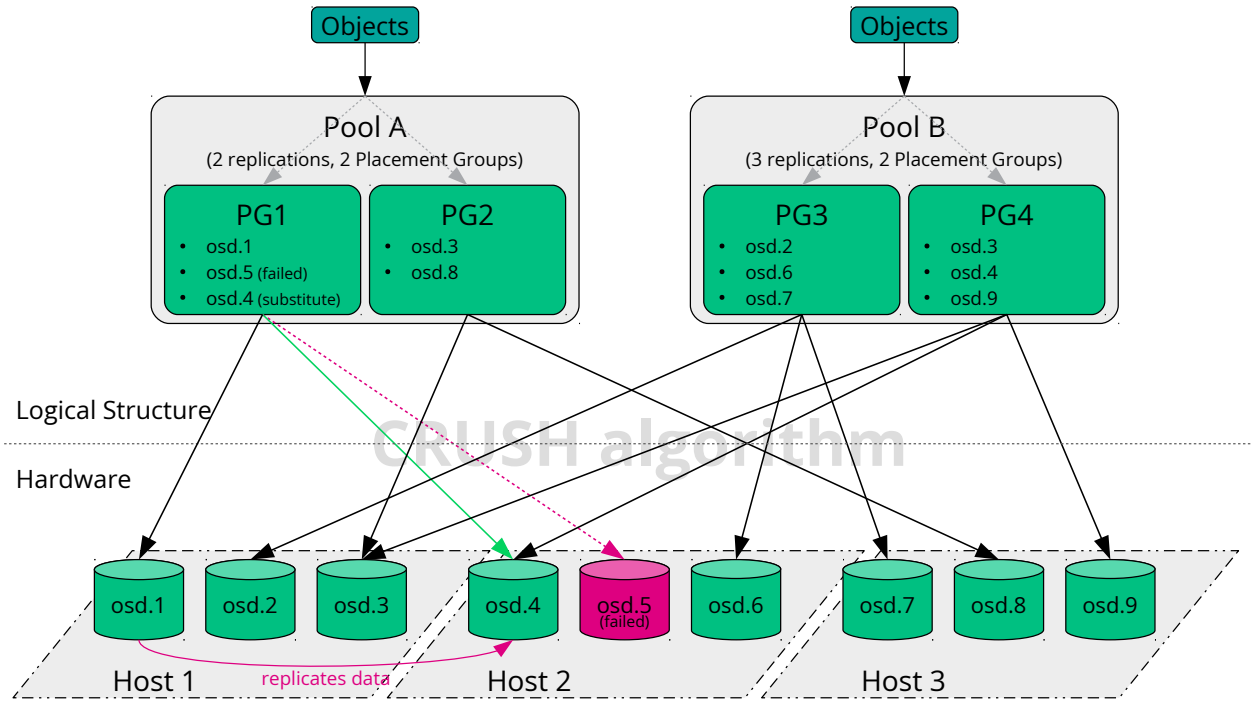

FIGURA 1.2: EXEMPLO DE CEPH DE POUCO DIMENSIONAMENTO

Se um novo nó com novos OSDs for adicionado ao cluster, o mapa do cluster será modificado. Na sequência, a função CRUSH retorna locais diferentes para os objetos. Os objetos que recebem os novos locais serão realocados. Esse processo resulta no uso equilibrado de todos os OSDs.

1.4 BlueStore

BlueStore é um novo back end de armazenamento padrão para Ceph a partir do SES 5. Ele apresenta melhor desempenho do que o FileStore, checksum completo de dados e compactação incorporada.

O BlueStore gerencia um, dois ou três dispositivos de armazenamento. No caso mais simples, o BlueStore consome um único dispositivo de armazenamento principal. Normalmente, o dispositivo de armazenamento é particionado em duas partes:

1. Uma pequena partição denominada BlueFS que implementa funcionalidades do tipo do sistema de arquivos que o RocksDB exige.

2. Normalmente, o restante do dispositivo é ocupado por uma partição grande. Ele é gerenciado diretamente pelo BlueStore e contém todos os dados reais. Normalmente, esse dispositivo principal é identificado por um link simbólico de bloco no diretório de dados.

É possível também implantar o BlueStore em dois dispositivos adicionais:

É possível usar o *dispositivo WAL* para o diário interno ou o registro write-ahead do BlueStore. Ele é identificado pelo link simbólico `block.wal` no diretório de dados. O uso de um dispositivo WAL separado será útil apenas se ele for mais rápido do que o dispositivo principal ou o dispositivo de BD. Por exemplo, quando:

- O dispositivo WAL é um NVMe, o dispositivo de BD é uma SSD e o dispositivo de dados é uma SSD ou HDD.

- Ambos os dispositivos WAL e de BD são SSDs separadas, e o dispositivo de dados é uma SSD ou HDD.

É possível usar um *dispositivo de BD* para armazenar metadados internos do BlueStore. O BlueStore (ou, em vez dele, o RocksDB incorporado) colocará o máximo de metadados que puder no dispositivo de BD para melhorar o desempenho. Mais uma vez, apenas será útil provisionar um dispositivo de BD compartilhado se ele for mais rápido do que o dispositivo principal.

Dica: Planejar o tamanho do BD

Planeje com cuidado para garantir o tamanho suficiente do dispositivo de BD. Se o dispositivo de BD ficar cheio, os metadados serão despejados no dispositivo principal, o que prejudica bastante o desempenho do OSD.

Você pode verificar se uma partição WAL/BD está ficando lotada e sendo despejada ao executar o comando **`ceph daemon osd.`***`ID`* **`perf dump`**. O valor `slow_used_bytes` mostra a quantidade de dados que está sendo despejada:

```
cephuser@adm > ceph daemon osd.ID perf dump | jq '.bluefs'
"db_total_bytes": 1073741824,
"db_used_bytes": 33554432,
"wal_total_bytes": 0,
"wal_used_bytes": 0,
"slow_total_bytes": 554432,
"slow_used_bytes": 554432,
```

1.5 Informações adicionais

- Como um projeto comunitário, o Ceph tem sua própria documentação online completa. Para os tópicos não encontrados neste manual, visite https://docs.ceph.com/en/pacific/.
- A publicação original *CRUSH: Controlled, Scalable, Decentralized Placement of Replicated Data* por *S.A. Weil, S.A. Brandt, E.L. Miller, C. Maltzahn* apresenta informações úteis sobre o funcionamento interno do Ceph. Principalmente na implantação de clusters em grande escala, trata-se de uma leitura recomendada. Você encontra essa publicação em http://www.ssrc.ucsc.edu/papers/weil-sc06.pdf.
- É possível usar o SUSE Enterprise Storage com distribuições que não são do SUSE OpenStack. Os clientes Ceph precisam estar em um nível que seja compatível com o SUSE Enterprise Storage.

Nota

A SUSE suporta o componente de servidor da implantação do Ceph, e o cliente é suportado pelo fornecedor de distribuição do OpenStack.

2 Requisitos e recomendações de hardware

Os requisitos de hardware do Ceph dependem bastante da carga de trabalho de E/S. Os requisitos e as recomendações de hardware a seguir devem ser considerados um ponto de partida para o planejamento detalhado.

Em geral, as recomendações apresentadas nesta seção são baseadas em cada processo. Se houver vários processos localizados na mesma máquina, os requisitos de CPU, RAM, disco e rede deverão ser incrementados.

2.1 Visão geral da rede

O Ceph tem várias redes lógicas:

- Uma rede de front end chamada `public network`.
- Uma rede interna confiável, a rede de back end, chamada `cluster network`. Isso é opcional.
- Uma ou mais redes de cliente para gateways. Isso é opcional e está fora do escopo deste capítulo.

A rede pública é a rede pela qual os daemons do Ceph se comunicam entre si e com seus clientes. Isso significa que todo o tráfego de cluster do Ceph passa por essa rede, exceto quando há uma rede de cluster configurada.

A rede de cluster é a rede de back end entre os nós OSD para replicação, reequilíbrio e recuperação. Se configurada, essa rede opcional deve oferecer o dobro da largura de banda da rede pública com replicação de três vias padrão, já que o OSD primário envia duas cópias para outros OSDs por meio dessa rede. A rede pública está em um lado entre os clientes e os gateways para se comunicar com monitores, gerenciadores, nós MDS e nós OSD. Ela também é usada por monitores, gerenciadores e nós MDS para se comunicar com os nós OSD.

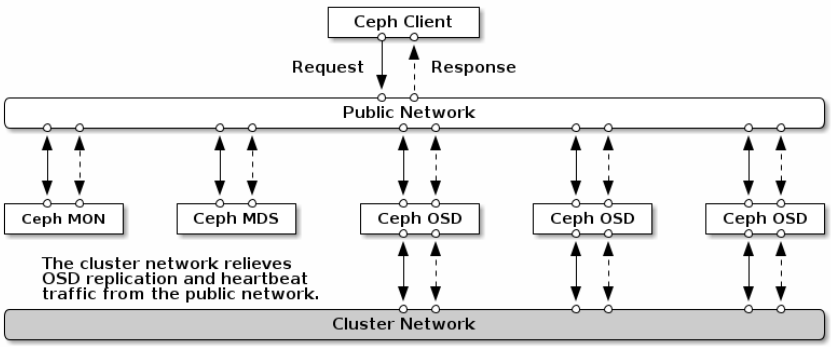

FIGURA 2.1: VISÃO GERAL DA REDE

2.1.1 Recomendações de rede

Recomendamos uma única rede tolerante a falhas com largura de banda suficiente para atender aos seus requisitos. Para o ambiente de rede pública do Ceph, recomendamos duas interfaces de rede de 25 GbE (ou mais rápidas) vinculadas por meio de 802.3ad (LACP). Essa é considerada a configuração mínima do Ceph. Se você também usa uma rede de cluster, recomendamos quatro interfaces de rede de 25 GbE vinculadas. A vinculação de duas ou mais interfaces de rede proporciona melhor throughput por meio da agregação de links e, considerando os links e switches redundantes, melhora a tolerância a falhas e a capacidade de manutenção.

Você também pode criar VLANs para isolar diferentes tipos de tráfego em uma vinculação. Por exemplo, você pode criar uma vinculação para fornecer duas interfaces VLAN: uma para a rede pública e a segunda para a rede de cluster. No entanto, isso *não* é necessário ao configurar o projeto de rede do Ceph. Os detalhes sobre a vinculação das interfaces estão disponíveis em https://documentation.suse.com/sles/15-SP3/html/SLES-all/cha-network.html#sec-network-iface-bonding.

É possível aprimorar a tolerância a falhas por meio do isolamento dos componentes em domínios de falha. Para melhorar a tolerância a falhas da rede, a vinculação de uma interface de duas Placas de Interface de Rede (NIC, Network Interface Cards) separadas oferece proteção contra falhas de uma única NIC. Da mesma forma, criar uma vinculação entre dois switches protege contra falhas de um switch. Recomendamos consultar o fornecedor do equipamento de rede para projetar o nível de tolerância a falhas necessário.

Importante: Rede de administração não suportada

A configuração de rede de administração adicional, que permite, por exemplo, separar projetos de rede SSH, Salt ou DNS, não foi testada nem é suportada.

Dica: Nós configurados por DHCP

Se os nós de armazenamento foram configurados por DHCP, os tempos de espera padrão talvez não sejam suficientes para que a rede seja configurada corretamente antes que os vários daemons do Ceph sejam iniciados. Se isso acontecer, os Ceph MONs e OSDs não serão iniciados corretamente (a execução do `systemctl status ceph\*` resultará em erros do tipo "impossível vincular"). Para evitar esse problema, recomendamos aumentar o tempo de espera do cliente DHCP para pelo menos 30 segundos em cada nó no cluster de armazenamento. Para fazer isso, mude as seguintes configurações em cada nó:

Em `/etc/sysconfig/network/dhcp`, defina

```
DHCLIENT_WAIT_AT_BOOT="30"
```

Em `/etc/sysconfig/network/config`, defina

```
WAIT_FOR_INTERFACES="60"
```

2.1.1.1 Adicionando uma rede privada a um cluster em execução

Se você não especificar uma rede de cluster durante a implantação do Ceph, ele assumirá um ambiente de rede pública única. Enquanto o Ceph funciona bem com uma rede pública, seu desempenho e segurança melhoram quando você define uma segunda rede privada de cluster. Para suportar duas redes, cada nó do Ceph precisa ter pelo menos duas placas de rede.

Você precisará aplicar as mudanças a seguir a cada nó do Ceph. É relativamente rápido fazer isso em um cluster pequeno, mas pode ser muito demorado se você tem um cluster com centenas ou milhares de nós.

1. Defina a rede de cluster usando o seguinte comando:

```
# ceph config set global cluster_network MY_NETWORK
```

Reinicie os OSDs para vinculação à rede de cluster especificada:

```
# systemctl restart ceph-*@osd.*.service
```

2. Verifique se a rede privada de cluster funciona conforme esperado no nível do OS.

2.1.1.2 Nós de monitoramento em sub-redes diferentes

Se os nós do monitor estiverem em várias sub-redes, por exemplo, localizados em salas distintas e processados por switches diferentes, você precisará especificar o endereço da rede pública na notação CIDR:

```
cephuser@adm > ceph config set mon public_network
 "MON_NETWORK_1, MON_NETWORK_2, MON_NETWORK_N
```

Por exemplo:

```
cephuser@adm > ceph config set mon public_network "192.168.1.0/24, 10.10.0.0/16"
```

Atenção

Se você especificar mais de um segmento de rede para a rede pública (ou cluster), conforme descrito nesta seção, cada uma dessas sub-redes deverá ser capaz de rotear para todas as outras. Do contrário, os MONs e outros daemons do Ceph em diferentes segmentos de rede não poderão se comunicar, e o resultado será um cluster dividido. Além disso, se você usa um firewall, certifique-se de incluir cada endereço IP ou sub-rede em seu iptables e abrir as portas para eles em todos os nós, conforme necessário.

2.2 Configurações de várias arquiteturas

O SUSE Enterprise Storage suporta as arquiteturas x86 e Arm. Ao considerar cada arquitetura, é importante notar que, de uma perspectiva de núcleo por OSD, frequência e RAM, não há diferença real entre as arquiteturas de CPU quanto ao dimensionamento.

Assim como nos processadores x86 menores (sem servidor), os núcleos de desempenho mais baixo com base em Arm podem não oferecer uma experiência ideal, principalmente quando usados para pools codificados para eliminação.

Nota

Em toda a documentação, *SYSTEM-ARCH* é usado no lugar de x86 ou Arm.

2.3 Configuração de hardware

Para obter a melhor experiência do produto, recomendamos começar com a configuração de cluster recomendada. Para um cluster de teste ou um cluster com menos requisitos de desempenho, documentamos uma configuração de cluster mínima suportada.

2.3.1 Configuração de cluster mínima

Uma configuração de cluster mínima do produto consiste em:

- Pelo menos quatro nós físicos (nós OSD) com colocation de serviços
- Ethernet dupla de 10 Gb como rede vinculada
- Um Nó de Admin separado (pode ser virtualizado em um nó externo)

Veja abaixo uma configuração detalhada:

- Nó de Admin separado com 4 GB de RAM, quatro núcleos, capacidade de armazenamento de 1 TB. Normalmente, esse é o nó do Master Salt. Os serviços e gateways do Ceph, como Ceph Monitor, Servidor de Metadados, Ceph OSD, Gateway de Objetos ou NFS Ganesha, não são suportados no Nó de Admin, pois ele precisa orquestrar os processos de atualização e upgrade do cluster de forma independente.
- Pelo menos quatro nós OSD físicos, com oito discos OSD cada. Consulte a *Seção 2.4.1, "Requisitos mínimos"* para obter os requisitos.
  A capacidade total do cluster deve ser dimensionada de modo que, mesmo com um nó indisponível, a capacidade total usada (incluindo redundância) não exceda 80%.
- Três instâncias do Ceph Monitor. Os monitores precisam ser executados do armazenamento SSD/NVMe, não de HDDs, por motivos de latência.
- É possível fazer colocation dos monitores, do Servidor de Metadados e dos gateways nos nós OSD. Consulte a *Seção 2.12, "OSD e monitor que compartilham um servidor"* para saber sobre colocation do monitor. Se você fizer colocation dos serviços, os requisitos de memória e CPU precisarão ser somados.

- O iSCSI Gateway, o Gateway de Objetos e o Servidor de Metadados exigem pelo menos 4 GB de RAM incremental e quatro núcleos.
- Se você usa CephFS, S3/Swift ou iSCSI, serão necessárias pelo menos duas instâncias das respectivas funções (Servidor de Metadados, Gateway de Objetos, iSCSI) para redundância e disponibilidade.
- Os nós devem ser dedicados ao SUSE Enterprise Storage e não devem ser usados para nenhuma outra carga de trabalho física, em containers ou virtualizada.
- Se qualquer um dos gateways (iSCSI, Gateway de Objetos, NFS Ganesha, Servidor de Metadados etc.) for implantado em VMs, essas VMs não deverão ser hospedadas em máquinas físicas que processam outras funções do cluster. (Isso é desnecessário, pois eles são suportados como serviços com colocation.)
- Ao implantar serviços como VMs em hipervisores fora do cluster físico principal, os domínios de falha devem ser respeitados para garantir a redundância.
  Por exemplo, não implante várias funções do mesmo tipo no mesmo hypervisor, como várias instâncias de MONs ou MDSs.
- Na implantação em VMs, é importante principalmente garantir que os nós tenham uma forte conectividade de rede e uma boa sincronização de horário.
- Os nós do hipervisor devem ser dimensionados adequadamente para evitar a interferência de outras cargas de trabalho que consomem recursos de CPU, RAM, rede e armazenamento.

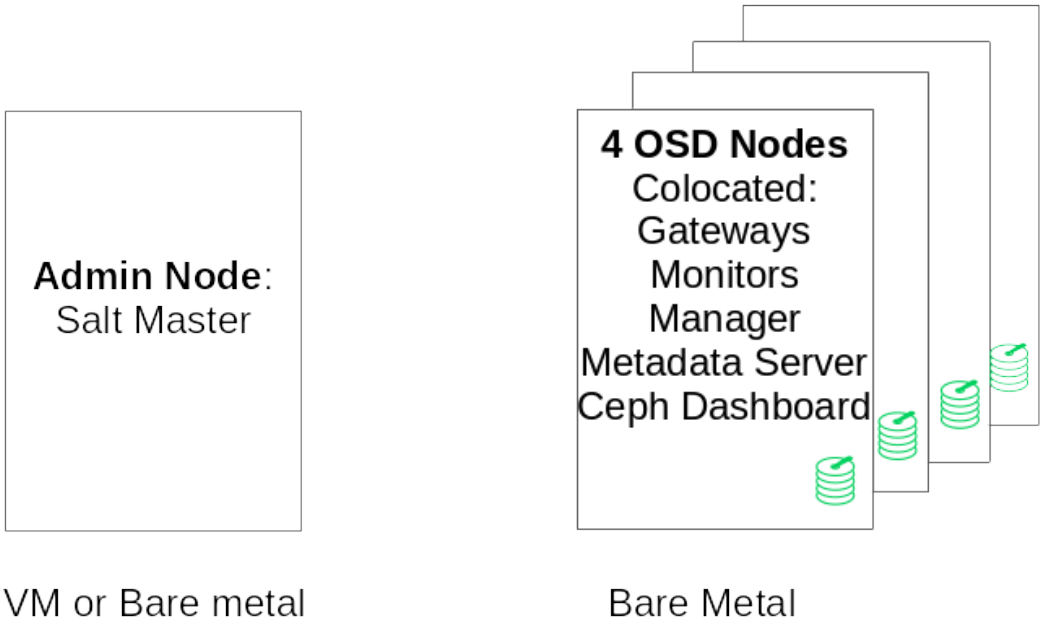

FIGURA 2.2: CONFIGURAÇÃO DE CLUSTER MÍNIMA

2.3.2 Configuração recomendada do cluster de produção

Depois que você aumentar o cluster, recomendamos realocar os Ceph Monitors, os Servidores de Metadados e os Gateways para nós separados para melhorar a tolerância a falhas.

- Sete Nós de Armazenamento de Objetos
  - Nenhum nó único excede ~ 15% do total de armazenamento.
  - A capacidade total do cluster deve ser dimensionada de modo que, mesmo com um nó indisponível, a capacidade total usada (incluindo redundância) não exceda 80%.
  - Ethernet de 25 Gb ou superior, cada uma vinculada ao cluster interno e à rede pública externa.
  - Mais de 56 OSDs por cluster de armazenamento.
  - Consulte a *Seção 2.4.1, "Requisitos mínimos"* para ver outras recomendações.
- Nós dedicados de infraestrutura física.

- Três nós do Ceph Monitor: 4 GB de RAM, processador de 4 núcleos, SSDs RAID 1 para disco.
  Consulte a *Seção 2.5, "Nós do monitor"* para ver outras recomendações.
- Nós do Gateway de Objetos: 32 GB de RAM, processador de 8 núcleos, SSDs RAID 1 para disco.
  Consulte a *Seção 2.6, "Nós do Gateway de Objetos"* para ver outras recomendações.
- Nós do iSCSI Gateway: 16 GB de RAM, processador de 8 núcleos, SSDs RAID 1 para disco.
  Consulte a *Seção 2.9, "Nós do Gateway iSCSI"* para ver outras recomendações.
- Nós do Servidor de Metadados (um ativo/um standby ativo): 32 GB de RAM, processador de 8 núcleos, SSDs RAID 1 para disco.
  Consulte a *Seção 2.7, "Nós do servidor de metadados"* para ver outras recomendações.
- Um Nó de Admin do SES: 4 GB de RAM, processador de 4 núcleos, SSDs RAID 1 para disco.

2.3.3 Configuração de múltiplos caminhos

Para usar um hardware de múltiplos caminhos, verifique se o LVM vê `multipath_component_detection = 1` no arquivo de configuração na seção `devices`. É possível executar o comando **`lvm config`** para verificar isso.

Se preferir, verifique se o LVM filtra os componentes mpath de um dispositivo por meio da configuração do filtro do LVM. Esse procedimento será específico do host.

Nota

Ele não é recomendado e deverá ser considerado apenas se `multipath_component_detection = 1` não puder ser definido.

Para obter informações sobre configuração de múltiplos caminhos, consulte https://documentation.suse.com/sles/15-SP3/html/SLES-all/cha-multipath.html#sec-multipath-lvm.

2.4 Nós de armazenamento de objetos

2.4.1 Requisitos mínimos

- As seguintes recomendações de CPU consideram os dispositivos independentemente do uso pelo Ceph:
  - 1x Thread de CPU de 2 GHz por spinner.
  - 2x Threads de CPU de 2 GHz por SSD.
  - 4x Threads de CPU de 2 GHz por NVMe.
- Redes separadas de 10 GbE (pública/cliente e interna), 4x 10 GbE necessário, 2x 25 GbE recomendado.
- RAM total necessária = número de OSDs x (1 GB + `osd_memory_target`) + 16 GB. Consulte o *Livro "Guia de Administração e Operações", Capítulo 28 "Configuração do cluster do Ceph", Seção 28.4.1 "Configurando o dimensionamento automático do cache"* para obter mais detalhes sobre `osd_memory_target`.
- Discos OSD nas configurações de JBOD ou nas configurações individuais de RAID-0.
- Diário OSD pode residir no disco OSD.
- Os discos OSD devem ser usados exclusivamente pelo SUSE Enterprise Storage.
- Disco e SSD dedicados para o sistema operacional, preferencialmente em uma configuração de RAID 1.
- Aloque pelo menos 4 GB de RAM adicional se esse host OSD for hospedar parte de um pool de cache usado para camadas de cache.
- Ceph Monitors, Ceph Gateway e Servidores de Metadados podem residir em Nós de Armazenamento de Objetos.
- Por motivos de desempenho do disco, os nós OSD estão completamente vazios. Nenhuma outra carga de trabalho deve ser executada em um nó OSD, a menos que seja uma configuração mínima de Ceph Monitors e Ceph Managers.
- SSDs para Diário com o diário de SSD de proporção 6:1 para OSD.

Nota

Verifique se os nós OSD não têm dispositivos de blocos em rede mapeados, como iSCSI ou imagens de Dispositivo de Blocos RADOS.

2.4.2 Tamanho mínimo do disco

Há dois tipos de espaço em disco necessários para execução no OSD: o espaço para o dispositivo WAL/BD e o espaço principal para os dados armazenados. O valor mínimo (e padrão) para o WAL/BD é de 6 GB. O espaço mínimo para os dados é de 5 GB, pois as partições menores do que 5 GB recebem automaticamente o peso de 0.

Portanto, embora o espaço mínimo em disco para um OSD seja de 11 GB, não é recomendável um disco menor do que 20 GB, mesmo para fins de teste.

2.4.3 Tamanho recomendado para o dispositivo WAL e BD do BlueStore

Dica: Mais informações

Consulte a *Seção 1.4, "BlueStore"* para obter mais informações sobre o BlueStore.

- Recomendamos reservar 4 GB para o dispositivo WAL. O tamanho mínimo do BD é 64 GB para cargas de trabalho apenas RBD, mas o tamanho recomendado do BD para cargas de trabalho do Gateway de Objetos e do CephFS é 2% da capacidade do dispositivo principal (sendo no mínimo 196 GB).

Importante

Recomendamos volumes de BD maiores para implantações de alta carga, principalmente se houver uso elevado do RGW ou CephFS. Reserve parte da capacidade (slots) para instalar mais hardware e ter mais espaço no banco de dados, se necessário.

- Se você pretende colocar o dispositivo WAL e BD no mesmo disco, recomendamos usar uma única partição para ambos os dispositivos, em vez de ter uma partição separada para cada um. Dessa forma, o Ceph pode usar o dispositivo BD também para operação de WAL. Portanto, o gerenciamento do espaço em disco é mais eficaz, pois o Ceph usa a partição BD para o WAL apenas quando há necessidade dela. Outra vantagem é que há pouca probabilidade de a partição WAL ficar cheia, e quando ela não é totalmente usada, o espaço dela não é desperdiçado, mas usado para operação do BD.
  Para compartilhar o dispositivo BD com o WAL, *não* especifique o dispositivo WAL, mas apenas o dispositivo BD.
  Encontre mais informações sobre como especificar um layout de OSD no *Livro "Guia de Administração e Operações", Capítulo 13 "Tarefas operacionais", Seção 13.4.3 "Adicionando OSDs por meio da especificação DriveGroups"*.

2.4.4 SSD para partições WAL/BD

As SSDs (Solid-State Drives – Unidades de Estado Sólido) não têm partes móveis. Isso reduz o tempo de acesso aleatório e a latência de leitura enquanto acelera o throughput de dados. Como o preço delas por 1 MB é significativamente maior do que o preço dos discos rígidos giratórios, as SSDs são adequadas apenas para armazenamento menor.

Os OSDs podem observar um aprimoramento significativo no desempenho ao armazenar as partições WAL/BD em um SSD e os dados de objetos em um disco rígido separado.

Dica: Compartilhando um SSD com várias partições WAL/BD

Como as partições WAL/BD ocupam relativamente pouco espaço, você pode compartilhar um disco SSD com várias partições WAL/BD. Com cada partição WAL/BD, lembre-se de que o desempenho do disco SSD é reduzido. Não recomendamos compartilhar mais do que seis partições WAL/BD no mesmo disco SSD, e 12 nos discos NVMe.

2.4.5 Número máximo recomendado de discos

Você poderá ter quantos discos forem permitidos em um servidor. Há algumas coisas que devem ser consideradas na hora de planejar o número de discos por servidor:

- *Largura de banda de rede.* Quanto mais discos você tiver em um servidor, mais dados deverão ser transferidos por placa(s) de rede para as operações de gravação em disco.
- *Memória.* RAM acima de 2 GB é usada para o cache do BlueStore. Com o `osd_memory_target` padrão de 4 GB, o sistema tem um tamanho de cache inicial razoável para mídia giratória. Se você usa SSD ou NVME, considere aumentar o tamanho do cache e a alocação de RAM por OSD para maximizar o desempenho.
- *Tolerância a falhas.* Em caso de falha no servidor inteiro, quanto mais discos ele tiver, mais OSDs o cluster perderá temporariamente. Além disso, para manter as regras de replicação em execução, você precisa copiar todos os dados do servidor com falha para os outros nós no cluster.

2.5 Nós do monitor

- Pelo menos três nós MON são necessários. O número de monitores deve ser sempre ímpar ($1 + 2n$).
- 4 GB de RAM.
- Processador com quatro núcleos lógicos.
- Uma SSD ou outro tipo de armazenamento rápido o bastante é altamente recomendado para monitores, principalmente para o caminho `/var/lib/ceph` em cada nó do monitor, pois o quorum pode ficar instável com altas latências de disco. É recomendada a configuração de dois discos no RAID 1 para redundância. É recomendável o uso de discos separados ou, pelo menos, de partições de disco separadas para que os processos do monitor protejam o espaço em disco disponível do monitor contra eventos como arquivo de registro muito grande.
- Deve haver apenas um processo do monitor por nó.

- A combinação de nós OSD, MON ou do Gateway de Objetos apenas será suportada se houver recursos de hardware suficientes disponíveis. Isso significa que os requisitos para todos os serviços precisam ser incrementados.
- Duas interfaces de rede vinculadas a vários switches.

2.6 Nós do Gateway de Objetos

Os nós do Gateway de Objetos devem ter pelo menos seis núcleos de CPU e 32 GB de RAM. Quando outros processos são colocalizados na mesma máquina, os requisitos precisam ser incrementados.

2.7 Nós do servidor de metadados

O dimensionamento apropriado dos nós do Servidor de Metadados depende do caso de uso específico. Em geral, quanto mais arquivos abertos o Servidor de Metadados tiver que processar, mais CPU e RAM serão necessárias. Veja a seguir os requisitos mínimos:

- 4 GB de RAM para cada daemon do Servidor de Metadados.
- Interface de rede vinculada.
- CPU de 2.5 GHz com, no mínimo, 2 núcleos.

2.8 Nó de Admin

Pelo menos 4 GB de RAM e uma CPU quad-core são necessárias. Isso inclui a execução do Master Salt no Nó de Admin. Para clusters grandes com centenas de nós, 6 GB de RAM é a sugestão.

2.9 Nós do Gateway iSCSI

Os nós do Gateway iSCSI devem ter pelo menos seis núcleos de CPU e 16 GB de RAM.

2.10 SES e outros produtos da SUSE

Esta seção contém informações importantes sobre a integração do SES com outros produtos da SUSE.

2.10.1 SUSE Manager

O SUSE Manager e o SUSE Enterprise Storage não estão integrados, portanto, o SUSE Manager não pode gerenciar um cluster do SES neste momento.

2.11 Limitações de nome

Em geral, o Ceph não suporta caracteres não ASCII em arquivos de configuração, nomes de pool, nomes de usuário, etc. Ao configurar um cluster do Ceph, é recomendável usar apenas caracteres alfanuméricos simples (A-Z, a-z, 0-9) e pontuação mínima ('.', '-', '_') em todos os nomes de objeto/configuração do Ceph.

2.12 OSD e monitor que compartilham um servidor

Embora seja tecnicamente possível executar OSDs e MONs no mesmo servidor em ambientes de teste, é altamente recomendável ter um servidor separado para cada nó do monitor em produção. O principal motivo é o desempenho: quanto mais OSDs o cluster tiver, mais operações de E/S os nós do MON precisarão executar. E, quando um servidor for compartilhado entre um nó do MON e o(s) OSD(s), as operações de E/S do(s) OSD(s) serão um fator limitador para o nó do monitor.

Outra consideração é quando se deve compartilhar discos entre um OSD, um nó do MON e o sistema operacional no servidor. A resposta é simples: se possível, dedique um disco separado ao OSD e um servidor separado a um nó de monitor.

Embora o Ceph suporte OSDs com base em diretório, um OSD deve sempre ter um disco dedicado diferente do sistema operacional.

Dica

Se for *realmente* necessário executar um OSD e um nó do MON no mesmo servidor, execute o MON em um disco separado por meio da montagem do disco no diretório `/var/lib/ceph/mon` para melhorar um pouco o desempenho.

3 Configuração de alta disponibilidade do nó de admin

O *Nó de Admin* é um nó de cluster do Ceph em que o serviço do Master Salt é executado. Ele gerencia o restante dos nós do cluster consultando e instruindo os serviços do Minion Salt. Normalmente, ele também inclui outros serviços, por exemplo, o painel de controle do *Grafana* com suporte do kit de ferramentas de monitoramento do *Prometheus*.

Em caso de falha no Nó de Admin, geralmente você precisa fornecer um novo hardware ativo para o nó e restaurar a pilha completa de configuração do cluster de um backup recente. Esse é um método demorado e provoca interrupção no cluster.

Para evitar o tempo de espera no desempenho do cluster do Ceph causado pela falha no Nó de Admin, é recomendável usar o cluster de Alta Disponibilidade (HA, High Availability) para o Nó de Admin do Ceph.

3.1 Estrutura do cluster de HA para o nó de admin

A ideia de um cluster de HA é que, em caso de falha em um nó do cluster, o outro nó assume automaticamente a função dele, incluindo o Nó de Admin virtualizado. Dessa forma, os outros nós do cluster do Ceph não percebem a falha no Nó de Admin.

A solução de HA mínima para o Nó de Admin requer o seguinte hardware:

- Dois servidores completamente vazios capazes de executar o SUSE Linux Enterprise com a extensão de Alta Disponibilidade e de virtualizar o Nó de Admin.
- Dois ou mais caminhos de comunicação de rede redundantes. Por exemplo, via Ligação de Dispositivo de Rede.
- Armazenamento compartilhado para hospedar a(s) imagem(ns) de disco da máquina virtual do Nó de Admin. O armazenamento compartilhado precisa ser acessível aos dois servidores. Por exemplo, ele pode ser uma exportação NFS, um compartilhamento do Samba ou um destino iSCSI.

Encontre mais detalhes sobre os requisitos de cluster no site https://documentation.suse.com/sle-ha/15-SP3/single-html/SLE-HA-install-quick/#sec-ha-inst-quick-req.

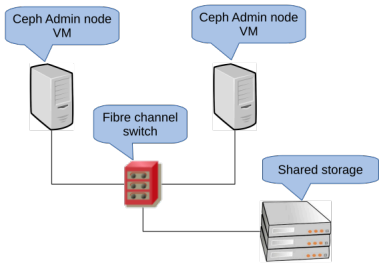

FIGURA 3.1: CLUSTER DE HA DE 2 NÓS PARA O NÓ DE ADMIN

3.2 Criando um cluster de HA com o nó de admin

O procedimento a seguir resume as etapas mais importantes de criação do cluster de HA para virtualização do Nó de Admin. Para obter detalhes, consulte os links indicados.

1. Configure um cluster de HA básico de 2 nós com armazenamento compartilhado, conforme descrito em https://documentation.suse.com/sle-ha/15-SP3/single-html/SLE-HA-install-quick/#art-sleha-install-quick.

2. Em ambos os nós do cluster, instale todos os pacotes necessários para executar o hipervisor KVM e o kit de ferramentas `libvirt`, conforme descrito em https://documentation.suse.com/sles/15-SP3/single-html/SLES-virtualization/#sec-vt-installation-kvm.

3. No primeiro nó do cluster, crie uma nova VM (Virtual Machine – Máquina Virtual) KVM usando o `libvirt`, conforme descrito em https://documentation.suse.com/sles/15-SP3/single-html/SLES-virtualization/#sec-libvirt-inst-virt-install. Use o armazenamento compartilhado pré-configurado para armazenar as imagens de disco da VM.

4. Após o término da configuração da VM, exporte sua configuração para um arquivo XML no armazenamento compartilhado. Use a seguinte sintaxe:

```
# virsh dumpxml VM_NAME > /path/to/shared/vm_name.xml
```

5. Crie um recurso para a VM do Nó de Admin. Consulte https://documentation.suse.com/sle-ha/15-SP3/single-html/SLE-HA-guide/#cha-conf-hawk2 para obter informações gerais sobre a criação de recursos de HA. Há informações detalhadas sobre a criação de recursos para uma máquina virtual KVM descritas em http://www.linux-ha.org/wiki/VirtualDomain_%28resource_agent%29.

6. No convidado da VM recém-criado, implante o Nó de Admin, incluindo os serviços adicionais necessários nele. Siga as etapas relevantes na *Capítulo 6, Implantando o Salt*. Ao mesmo tempo, implante os nós de cluster do Ceph restantes nos servidores de cluster não HA.

II Implantando o cluster do Ceph

4 Introdução e tarefas comuns 30

5 Instalando e configurando o SUSE Linux Enterprise Server 32

6 Implantando o Salt 33

7 Implantando o cluster de boot usando `ceph-salt` 36

8 Implantando os serviços principais restantes com o cephadm 58

9 Implantação de serviços adicionais 74

4 Introdução e tarefas comuns

Desde o SUSE Enterprise Storage 7, os serviços do Ceph são implantados como containers, em vez de pacotes RPM. O processo de implantação tem duas etapas básicas:

Implantando o cluster de boot

Essa fase é chamada de *Implantação do dia 1* e consiste nas seguintes tarefas: Ela inclui a instalação do sistema operacional de base, a configuração da infraestrutura do Salt e a implantação do cluster mínimo, que é composto de um serviço MON e de um serviço MGR.

- Instale e faça a configuração básica do sistema operacional subjacente, SUSE Linux Enterprise Server 15 SP3, em todos os nós do cluster.
- Implante a infraestrutura do Salt em todos os nós do cluster para executar as preparações iniciais de implantação por meio do `ceph-salt`.
- Configure as propriedades básicas do cluster por meio do `ceph-salt` e implante-o.

Implantando serviços adicionais

Durante a *Implantação do dia 2*, os outros serviços essenciais e não essenciais do Ceph, por exemplo, gateways e pilha de monitoramento, são implantados.

Importante

Observe que a documentação da comunidade do Ceph usa o comando **`cephadm bootstrap`** durante a implantação inicial. O `ceph-salt` chama o comando **`cephadm bootstrap`** automaticamente. O comando **`cephadm bootstrap`** não deve ser executado diretamente. Qualquer implantação manual de cluster do Ceph que use o **`cephadm bootstrap`** não será suportada.

4.1 Ler os detalhes da versão

Nos detalhes da versão, você encontra mais informações sobre o que mudou desde a versão anterior do SUSE Enterprise Storage. Consulte os detalhes da versão para verificar se:

- seu hardware precisa de considerações especiais.
- qualquer pacote de software usado foi significativamente modificado.
- são necessárias precauções especiais para a instalação.

Os detalhes da versão também apresentam informações que não puderam ser incluídas a tempo no manual. Eles também incluem notas sobre problemas conhecidos.

Após a instalação do pacote `release-notes-ses`, encontre os detalhes da versão no diretório local `/usr/share/doc/release-notes` ou online em https://www.suse.com/releasenotes/.

5 Instalando e configurando o SUSE Linux Enterprise Server

1. Instale e registre o SUSE Linux Enterprise Server 15 SP3 em cada nó do cluster. Durante a instalação do SUSE Enterprise Storage, é necessário ter acesso aos repositórios de atualização, portanto, o registro é obrigatório. Inclua pelo menos os seguintes módulos:

   - Basesystem Module
   - Server Applications Module

   Encontre mais detalhes sobre como instalar o SUSE Linux Enterprise Server em https://documentation.suse.com/sles/15-SP3/html/SLES-all/cha-install.html.

2. Instale a extensão do *SUSE Enterprise Storage 7.1* em cada nó do cluster.

   ### Dica: Instalar o SUSE Enterprise Storage junto com o SUSE Linux Enterprise Server

   Você poderá instalar a extensão do SUSE Enterprise Storage 7.1 separadamente após instalar o SUSE Linux Enterprise Server 15 SP3 ou adicioná-la durante o procedimento de instalação do SUSE Linux Enterprise Server 15 SP3.

   Encontre mais detalhes sobre como instalar extensões em https://documentation.suse.com/sles/15-SP3/html/SLES-all/cha-register-sle.html.

3. Defina as configurações de rede incluindo a resolução de nome DNS em cada nó. Para obter mais informações sobre como configurar uma rede, consulte https://documentation.suse.com/sles/15-SP3/single-html/SLES-admin/#sec-network-yast. Para obter mais informações sobre como configurar um servidor DNS, consulte https://documentation.suse.com/sles/15-SP3/single-html/SLES-admin/#cha-dns.

6 Implantando o Salt

O SUSE Enterprise Storage usa o Salt e o `ceph-salt` para a preparação inicial do cluster. O Salt ajuda você a configurar e executar comandos em vários nós do cluster simultaneamente, de um host dedicado chamado *Master Salt*. Antes de implantar o Salt, considere os seguintes pontos importantes:

- Os *Minions Salt* são os nós controlados por um nó dedicado denominado Master Salt.
- Se o host Master Salt fizer parte do cluster do Ceph, ele precisará executar seu próprio Minion Salt, mas isso não é um requisito.

Dica: Compartilhando várias funções por servidor

O desempenho do cluster do Ceph é melhor quando cada função é implantada em um nó separado. Porém, as implantações reais às vezes exigem que um nó seja compartilhado com várias funções. Para evitar problema de desempenho e de procedimento de upgrade, não implante a função Ceph OSD, Servidor de Metadados ou Ceph Monitor no Nó de Admin.

- Os Minions Salt precisam resolver corretamente o nome de host do Master Salt na rede. Por padrão, eles procuram o nome de host `salt`, mas você pode especificar qualquer outro nome de host acessível por rede no arquivo `/etc/salt/minion`.

1. Instale o `salt-master` no nó do Master Salt:

```
root@master # zypper in salt-master
```

2. Verifique se o serviço `salt-master` está habilitado e foi iniciado. Se necessário, habilite-o e inicie-o:

```
root@master # systemctl enable salt-master.service
root@master # systemctl start salt-master.service
```

3. Se você pretende usar o firewall, verifique se o nó do Master Salt tem as portas 4505 e 4506 abertas para todos os nós do Minion Salt. Se as portas estiverem fechadas, você poderá abri-las usando o comando **`yast2 firewall`** e permitindo o serviço *salt-master* para a zona apropriada. Por exemplo, `public`.
4. Instale o pacote `salt-minion` em todos os nós do minion.

```
root@minion > zypper in salt-minion
```

5. Edite o `/etc/salt/minion` e remova o comentário da seguinte linha:

```
#log_level_logfile: warning
```

Mude o nível de registro de `warning` para `info`.

Nota: `log_level_logfile` e `log_level`

`log_level` controla quais mensagens de registro serão exibidas na tela, e `log_level_logfile` controla quais mensagens de registro serão gravadas no `/var/log/salt/minion`.

Nota

Verifique se você mudou o nível de registro em *todos* os nós do cluster (minion).

6. Verifique se o *nome de domínio completo e qualificado* de cada nó pode ser resolvido para um endereço IP na rede pública do cluster por todos os outros nós.

7. Configure todos os minions para se conectarem ao master. Se o Master Salt não puder ser acessado pelo nome de host `salt`, edite o arquivo `/etc/salt/minion` ou crie um novo arquivo `/etc/salt/minion.d/master.conf` com o seguinte conteúdo:

```
master: host_name_of_salt_master
```

Se você efetuou quaisquer mudanças nos arquivos de configuração mencionados acima, reinicie o serviço Salt em todos os Minions Salt relacionados:

```
root@minion > systemctl restart salt-minion.service
```

8. Verifique se o serviço `salt-minion` está habilitado e foi iniciado em todos os nós. Se necessário, habilite-o e inicie-o:

```
# systemctl enable salt-minion.service
# systemctl start salt-minion.service
```

9. Verifique a impressão digital de cada Minion Salt e aceite todas as chaves do Salt no Master Salt, se as impressões digitais forem correspondentes.

Nota

Se a impressão digital do Minion Salt retornar vazia, verifique se o Minion Salt tem uma configuração do Master Salt e pode se comunicar com o Master Salt.

Ver a impressão digital de cada minion:

```
root@minion > salt-call --local key.finger
local:
3f:a3:2f:3f:b4:d3:d9:24:49:ca:6b:2c:e1:6c:3f:c3:83:37:f0:aa:87:42:e8:ff...
```

Após coletar as impressões digitais de todos os Minions Salt, liste as impressões digitais de todas as chaves de minion não aceitas no Master Salt:

```
root@master # salt-key -F
[...]
Unaccepted Keys:
minion1:
3f:a3:2f:3f:b4:d3:d9:24:49:ca:6b:2c:e1:6c:3f:c3:83:37:f0:aa:87:42:e8:ff...
```

Se houver correspondência de impressões digitais dos minions, aceite-as:

```
root@master # salt-key --accept-all
```

10. Verifique se as chaves foram aceitas:

```
root@master # salt-key --list-all
```

11. Faça o teste para verificar se todos os Minions Salt respondem:

```
root@master # salt-run manage.status
```

7 Implantando o cluster de boot usando `ceph-salt`

Esta seção orienta você pelo processo de implantação de um cluster básico do Ceph. Leia as subseções a seguir com atenção e execute os comandos incluídos na ordem indicada.

7.1 Instalando `ceph-salt`

O `ceph-salt` inclui ferramentas para implantar clusters do Ceph gerenciados pelo cephadm. O `ceph-salt` usa a infraestrutura do Salt para executar o gerenciamento de OS; por exemplo, atualizações de software ou sincronização de horário, e definir funções para os Minions Salt.

No Master Salt, instale o pacote `ceph-salt`:

```
root@master # zypper install ceph-salt
```

O comando acima instalou `ceph-salt-formula` como uma dependência que modificou a configuração do Master Salt inserindo arquivos adicionais no diretório `/etc/salt/master.d`. Para aplicar as mudanças, reinicie o `salt-master.service` e sincronize os módulos do Salt:

```
root@master # systemctl restart salt-master.service
root@master # salt \* saltutil.sync_all
```

7.2 Configurando as propriedades do cluster

Use o comando **`ceph-salt config`** para configurar as propriedades básicas do cluster.

Importante

O arquivo `/etc/ceph/ceph.conf` é gerenciado pelo cephadm e os usuários *não devem* editá-lo. Os parâmetros de configuração do Ceph devem ser definidos usando o novo comando **`ceph config`**. Consulte o *Livro "Guia de Administração e Operações", Capítulo 28 "Configuração do cluster do Ceph", Seção 28.2 "Banco de dados de configuração"* para obter mais informações.

7.2.1 Usando o shell do `ceph-salt`

Se você executar o **`config`** sem nenhum caminho ou subcomando, digitará um shell interativo do `ceph-salt`ceph-salt. O shell é prático se você precisa configurar várias propriedades em um lote e não deseja digitar a sintaxe completa do comando.

```
root@master # ceph-salt config
/> ls
o- / ..................................................................... [...]
  o- ceph_cluster ........................................................ [...]
  | o- minions ..................................... [no minions]
  | o- roles ..................................... [...]
  |   o- admin ..................................... [no minions]
  |   o- bootstrap ..................................... [no minion]
  |   o- cephadm ..................................... [no minions]
  |   o- tuned ..................................... [...]
  |     o- latency ..................................... [no minions]
  |     o- throughput ..................................... [no minions]
  o- cephadm_bootstrap ..................................... [...]
  | o- advanced ..................................... [...]
  | o- ceph_conf ..................................... [...]
  | o- ceph_image_path ..................... [ no image path]
  | o- dashboard ..................................... [...]
  | | o- force_password_update ..................... [enabled]
  | | o- password ..................................... [admin]
  | | o- ssl_certificate ..................... [not set]
  | | o- ssl_certificate_key ..................... [not set]
  | | o- username ..................................... [admin]
  | o- mon_ip ..................................... [not set]
  o- containers ..................................... [...]
  | o- registries_conf ..................... [enabled]
  | | o- registries ..................................... [empty]
  | o- registry_auth ..................................... [...]
  |   o- password ..................................... [not set]
  |   o- registry ..................................... [not set]
  |   o- username ..................................... [not set]
  o- ssh ..................................... [no key pair set]
  | o- private_key ..................... [no private key set]
  | o- public_key ..................... [no public key set]
  o- time_server ..................... [enabled, no server host set]
    o- external_servers ..................... [empty]
    o- servers ..................................... [empty]
    o- subnet ..................................... [not set]
```

Como você pode ver na saída do comando `ceph-saltls do`, a configuração do cluster está organizada em uma estrutura de árvore. Para configurar uma propriedade específica do cluster no shell do `ceph-salt`, você tem duas opções:

- Execute o comando a partir da posição atual e digite o caminho absoluto para a propriedade como o primeiro argumento:

```
/> /cephadm_bootstrap/dashboard ls
o- dashboard ..................................................... [...]
  o- force_password_update ..................................... [enabled]
  o- password ................................................... [admin]
  o- ssl_certificate ........................................... [not set]
  o- ssl_certificate_key ....................................... [not set]
  o- username ................................................... [admin]
/> /cephadm_bootstrap/dashboard/username set ceph-admin
Value set.
```

- Mude para o caminho com a propriedade que você precisa configurar e execute o comando:

```
/> cd /cephadm_bootstrap/dashboard/
/ceph_cluster/minions> ls
o- dashboard ..................................................... [...]
  o- force_password_update ..................................... [enabled]
  o- password ................................................... [admin]
  o- ssl_certificate ........................................... [not set]
  o- ssl_certificate_key ....................................... [not set]
  o- username ..................................................[ceph-admin]
```

Dica: Preenchimento automático de trechos da configuração

Enquanto estiver em um shell do `ceph-salt`, você poderá usar o recurso de preenchimento automático semelhante ao do shell normal do Linux (Bash). Ele preenche os caminhos de configuração, subcomandos ou nomes de Minion Salt. Ao preencher automaticamente um caminho de configuração, você tem duas opções:

- Para permitir que o shell termine um caminho relativo à sua posição atual, pressione a tecla TAB `→|` duas vezes.
- Para permitir que o shell termine um caminho absoluto, digite `/` e pressione a tecla TAB `→|` duas vezes.

Dica: Navegando com as teclas do cursor

Se você digitar **cd** no shell do `ceph-salt` sem nenhum caminho, o comando imprimirá uma estrutura de árvore da configuração do cluster com a linha do caminho atual ativo. Você pode usar as teclas do cursor para cima e para baixo para navegar pelas linhas individuais. Depois de confirmar com `Enter`, o caminho de configuração mudará para o último caminho ativo.

Importante: Convenção

Para manter a documentação consistente, usaremos uma única sintaxe de comando sem inserir o shell do `ceph-salt`. Por exemplo, você pode listar a árvore de configuração do cluster usando o seguinte comando:

```
root@master # ceph-salt config ls
```

7.2.2 Adicionando minions Salt

Inclua todos ou um subconjunto de Minions Salt que implantamos e aceitamos na *Capítulo 6, Implantando o Salt* na configuração do cluster do Ceph. Você pode especificar os Minions Salt usando os nomes completos ou as expressões glob “*” e “?” para incluir vários Minions Salt de uma vez. Use o subcomando **add** no caminho `/ceph_cluster/minions`. O seguinte comando inclui todos os Minions Salt aceitos:

```
root@master # ceph-salt config /ceph_cluster/minions add '*'
```

Verifique se os Minions Salt especificados foram adicionados:

```
root@master # ceph-salt config /ceph_cluster/minions ls
o- minions ..................................................... [Minions: 5]
  o- ses-master.example.com ..................................... [no roles]
  o- ses-min1.example.com ....................................... [no roles]
  o- ses-min2.example.com ....................................... [no roles]
  o- ses-min3.example.com ....................................... [no roles]
  o- ses-min4.example.com ....................................... [no roles]
```

7.2.3 Especificando minions Salt gerenciados pelo cephadm

Especifique os nós que pertencerão ao cluster do Ceph e serão gerenciados pelo cephadm. Inclua todos os nós que executarão os serviços do Ceph e também o Nó de Admin:

```
root@master # ceph-salt config /ceph_cluster/roles/cephadm add '*'
```

7.2.4 Especificando o nó de admin

O Nó de Admin é o nó em que o arquivo de configuração `ceph.conf` e o chaveiro admin do Ceph estão instalados. Geralmente, você executa os comandos relacionados ao Ceph no Nó de Admin.

Dica: Master Salt e nó de admin no mesmo nó

Em um ambiente homogêneo onde todos ou a maioria dos hosts pertencem ao SUSE Enterprise Storage, recomendamos manter o Nó de Admin no mesmo host que o Master Salt.

Em um ambiente heterogêneo onde uma infraestrutura do Salt hospeda mais de um cluster, por exemplo, o SUSE Enterprise Storage junto com o SUSE Manager, *não* coloque o Nó de Admin no mesmo host que o Master Salt.

Para especificar o Nó de Admin, execute o seguinte comando:

```
root@master # ceph-salt config /ceph_cluster/roles/admin add ses-master.example.com
1 minion added.
root@master # ceph-salt config /ceph_cluster/roles/admin ls
o- admin ..................................................... [Minions: 1]
  o- ses-master.example.com ..................... [Other roles: cephadm]
```

Dica: Instalar o `ceph.conf` e o chaveiro admin em vários nós

Você pode instalar o arquivo de configuração do Ceph e o chaveiro admin em vários nós, se sua implantação exigir isso. Por motivos de segurança, evite instalá-los em todos os nós do cluster.

7.2.5 Especificando o primeiro nó MON/MGR

Você precisa especificar qual dos Minions Salt do cluster inicializará o cluster. Esse minion será o primeiro a executar os serviços Ceph Monitor e Ceph Manager.

```
root@master # ceph-salt config /ceph_cluster/roles/bootstrap set ses-min1.example.com
Value set.
root@master # ceph-salt config /ceph_cluster/roles/bootstrap ls
o- bootstrap ..................................... [ses-min1.example.com]
```

Além disso, você precisa especificar o endereço IP do MON de boot na rede pública para garantir que o parâmetro `public_network` seja definido corretamente, por exemplo:

```
root@master # ceph-salt config /cephadm_bootstrap/mon_ip set 192.168.10.20
```

7.2.6 Especificando perfis ajustados

Você precisa especificar quais dos minions do cluster têm os perfis ajustados ativamente. Para fazer isso, adicione estas funções explicitamente com os seguintes comandos:

Nota

Um minion não pode ter as duas funções `latency` e `throughput`.

```
root@master # ceph-salt config /ceph_cluster/roles/tuned/latency add ses-min1.example.com
Adding ses-min1.example.com...
1 minion added.
root@master # ceph-salt config /ceph_cluster/roles/tuned/throughput add ses-
min2.example.com
Adding ses-min2.example.com...
1 minion added.
```

7.2.7 Gerando um par de chaves SSH

O cephadm usa o protocolo SSH para se comunicar com os nós do cluster. Uma conta do usuário chamada `cephadm` é criada automaticamente e usada para comunicação por SSH.

Você precisa gerar a parte particular e a pública do par de chaves SSH:

```
root@master # ceph-salt config /ssh generate
```

```
Key pair generated.
root@master # ceph-salt config /ssh ls
o- ssh ................................................ [Key Pair set]
  o- private_key ..... [53:b1:eb:65:d2:3a:ff:51:6c:e2:1b:ca:84:8e:0e:83]
  o- public_key ...... [53:b1:eb:65:d2:3a:ff:51:6c:e2:1b:ca:84:8e:0e:83]
```

7.2.8 Configurando o servidor de horário

Todos os nós do cluster precisam ter o horário sincronizado com uma fonte de horário confiável. Há vários cenários para realizar a sincronização de horário:

- Se todos os nós do cluster já estiverem configurados para sincronizar o horário usando um serviço NTP de sua escolha, desabilite completamente o processamento do servidor de horário:

  ```
  root@master # ceph-salt config /time_server disable
  ```

- Se o seu site já tiver uma fonte de horário única, especifique o nome de host dessa fonte:

  ```
  root@master # ceph-salt config /time_server/servers add time-server.example.com
  ```

- Se preferir, o `ceph-salt` pode configurar um dos Minions Salt para agir como o servidor de horário para o restante do cluster. Esse recurso às vezes é chamado de "servidor de horário interno". Nesse cenário, o `ceph-salt` configura o servidor de horário interno (que deve ser um dos Minions Salt) para sincronizar seu horário com um servidor de horário externo, como `pool.ntp.org`, e configura todos os outros minions para obter o horário do servidor de horário interno. Isso pode ser feito da seguinte maneira:

  ```
  root@master # ceph-salt config /time_server/servers add ses-master.example.com
  root@master # ceph-salt config /time_server/external_servers add pool.ntp.org
  ```

  A opção `/time_server/subnet` especifica a sub-rede da qual os clientes NTP têm permissão para acessar o servidor NTP. Ela é definida automaticamente quando você especifica `/time_server/servers`. Se você precisar mudá-la ou especificá-la manualmente, execute:

  ```
  root@master # ceph-salt config /time_server/subnet set 10.20.6.0/24
  ```

Verifique as configurações do servidor de horário:

```
root@master # ceph-salt config /time_server ls
```

```
o- time_server ..................................... [enabled]
  o- external_servers ....................................... [1]
  | o- pool.ntp.org ........................................ [...]
  o- servers ................................................ [1]
  | o- ses-master.example.com ............................... [...]
  o- subnet ..................................... [10.20.6.0/24]
```

Encontre mais informações sobre como configurar a sincronização de horário em https://documentation.suse.com/sles/15-SP3/html/SLES-all/cha-ntp.html#sec-ntp-yast.

7.2.9 Configurando as credenciais de login do Ceph Dashboard

O Ceph Dashboard estará disponível após a implantação do cluster básico. Para acessá-lo, você precisa definir um nome de usuário e uma senha válidos, por exemplo:

```
root@master # ceph-salt config /cephadm_bootstrap/dashboard/username set admin
root@master # ceph-salt config /cephadm_bootstrap/dashboard/password set PWD
```

Dica: Forçando atualização da senha

Por padrão, o primeiro usuário do painel de controle será forçado a mudar sua senha ao efetuar o primeiro login. Para desabilitar esse recurso, execute o seguinte comando:

```
root@master # ceph-salt config /cephadm_bootstrap/dashboard/force_password_update
 disable
```

7.2.10 Usando o registro do container

O cluster do Ceph precisa ter acesso a um registro de container para que possa fazer download e implantar os serviços do Ceph em container. Há duas maneiras de acessar o registro:

- Se o cluster puder acessar o registro padrão em `registration.suse.com` (diretamente ou por proxy), você poderá apontar o `ceph-salt` diretamente para esse URL sem criar um registro local. Continue seguindo as etapas na *Seção 7.2.10.2, "Configurando o caminho para imagens de container"*.
- Se o cluster não puder acessar o registro padrão, por exemplo, para uma implantação isolada (air-gapped), você precisará configurar um registro de container local. Depois que o registro local for criado e configurado, você precisará apontar o `ceph-salt` para ele.

7.2.10.1 Criando e configurando o registro local (opcional)

Importante

Há vários métodos de criação de um registro local. As instruções nesta seção são exemplos de criação de registros seguros e não seguros. Para obter informações gerais sobre a execução de um registro de imagem de container, consulte https://documentation.suse.com/sles/15-SP3/single-html/SLES-container/#sec-docker-registry-installation.

Dica: Posicionamento e uso de porta

Implante o registro em uma máquina acessível a todos os nós no cluster. Recomendamos o Nó de Admin. Por padrão, o registro escuta na porta 5000.

No nó de registro, use o seguinte comando para garantir que a porta esteja livre:

```
ss -tulpn | grep :5000
```

Se outros processos (como `iscsi-tcmu`) já estiverem escutando na porta 5000, determine outra porta livre que possa ser usada para mapear para a porta 5000 no container de registro.

PROCEDIMENTO 7.1: CRIANDO O REGISTRO LOCAL

1. Verifique se a extensão `Containers Module` está habilitada:

```
> SUSEConnect --list-extensions | grep -A2 "Containers Module"
Containers Module 15 SP3 x86_64 (Activated)
```

2. Verifique se os seguintes pacotes estão instalados: `apache2-utils` (no caso de habilitar um registro seguro), `cni`, `cni-plugins`, `podman`, `podman-cni-config` e `skopeo`.

3. Colete as seguintes informações:

   - Nome de domínio completo e qualificado do host de registro (`REG_HOST_FQDN`).
   - Um número de porta disponível usado para mapear para a porta 5000 do container de registro (`REG_HOST_PORT`).
   - Se o registro será seguro ou não seguro (`insecure=[true|false]`).

4. Para iniciar um registro não seguro (sem criptografia SSL), siga estas etapas:

   a. Configure o `ceph-salt` para o registro não seguro:

```
cephuser@adm > ceph-salt config containers/registries_conf enable
cephuser@adm > ceph-salt config containers/registries_conf/registries \
 add prefix=REG_HOST_FQDN insecure=true \
 location=REG_HOST_PORT:5000
```

   b. Inicie o registro não seguro criando o diretório necessário (por exemplo, `/var/lib/registry`) e iniciando o registro com o comando **`podman`**:

```
# mkdir -p /var/lib/registry
# podman run --privileged -d --name registry \
 -p REG_HOST_PORT:5000 -v /var/lib/registry:/var/lib/registry \
 --restart=always registry:2
```

   c. Para que o registro seja iniciado após uma reinicialização, crie um arquivo da unidade `systemd` para ele e habilite-o:

```
> sudo podman generate systemd --files --name registry
> sudo mv container-registry.service /etc/systemd/system/
> sudo systemctl enable container-registry.service
```

5. Para iniciar um registro seguro, siga estas etapas:

   a. Crie os diretórios necessários:

```
# mkdir -p /var/lib/registry/{auth,certs}
```

   b. Gere um certificado SSL:

```
# openssl req -newkey rsa:4096 -nodes -sha256 \
 -keyout /var/lib/registry/certs/domain.key -x509 -days 365 \
 -out /var/lib/registry/certs/domain.crt
```

Nota

Defina o valor `CN=[value]` como o nome de domínio completo e qualificado do host (`[REG_HOST_FQDN]`).

c. Copie o certificado para todos os nós do cluster e atualize o cache do certificado:

```
# salt-cp '*' /var/lib/registry/certs/domain.crt \
 /etc/pki/trust/anchors/
# salt '*' cmd.shell "update-ca-certificates"
```

d. Gere uma combinação de nome de usuário e senha para autenticação no registro:

```
# htpasswd2 -bBc /var/lib/registry/auth/htpasswd \
 REG_USERNAME REG_PASSWORD
```

e. Inicie o registro seguro. Use o flag `REGISTRY_STORAGE_DELETE_ENABLED=true` para que você possa apagar imagens posteriormente com o comando **`skopeo delete`**.

```
podman run --name myregistry -p REG_HOST_PORT:5000 \
 -v /var/lib/registry:/var/lib/registry \
 -v /var/lib/registry/auth:/auth:z \
 -e "REGISTRY_AUTH=htpasswd" \
 -e "REGISTRY_AUTH_HTPASSWD_REALM=Registry Realm" \
 -e REGISTRY_AUTH_HTPASSWD_PATH=/auth/htpasswd \
 -v /var/lib/registry/certs:/certs:z \
 -e "REGISTRY_HTTP_TLS_CERTIFICATE=/certs/domain.crt" \
 -e "REGISTRY_HTTP_TLS_KEY=/certs/domain.key" \
 -e REGISTRY_STORAGE_DELETE_ENABLED=true \
 -e REGISTRY_COMPATIBILITY_SCHEMA1_ENABLED=true -d registry:2
```

f. Teste o acesso seguro ao registro:

```
> curl https://REG_HOST_FQDN:REG_HOST_PORT/v2/_catalog \
 -u REG_USERNAME:REG_PASSWORD
```

6. Quando o registro local é criado, você precisa sincronizar as imagens de container do registro oficial do SUSE em `registry.suse.com` com o registro local. Você pode usar o comando **`skopeo sync`** incluído no pacote `skopeo` para essa finalidade. Para obter mais detalhes, consulte a página de manual (**`man 1 skopeo-sync`**). Considere estes exemplos:

EXEMPLO 7.1: VENDO ARQUIVOS DE MANIFESTO

```
skopeo inspect docker://registry.suse.com/ses/7.1/ceph/ceph | jq .RepoTags
skopeo inspect docker://registry.suse.com/ses/7.1/ceph/grafana | jq .RepoTags
skopeo inspect docker://registry.suse.com/ses/7.1/ceph/prometheus-server:2.32.1 |
 jq .RepoTags
skopeo inspect docker://registry.suse.com/ses/7.1/ceph/prometheus-node-
exporter:1.1.2 | jq .RepoTags
```

```
skopeo inspect docker://registry.suse.com/ses/7.1/ceph/prometheus-
alertmanager:0.21.0 | jq .RepoTags
```

EXEMPLO 7.2: SINCRONIZAR COM UM DIRETÓRIO

Sincronizar todas as imagens do Ceph:

```
skopeo sync --src docker --dest dir registry.suse.com/ses/7.1/ceph/ceph /root/
images/
```

Sincronizar apenas as imagens mais recentes:

```
skopeo sync --src docker --dest dir registry.suse.com/ses/7.1/ceph/
ceph:latest /root/images/
```

EXEMPLO 7.3: SINCRONIZAR AS IMAGENS DO GRAFANA:

```
skopeo sync --src docker --dest dir registry.suse.com/ses/7.1/ceph/grafana /
root/images/
```

Sincronizar apenas as imagens mais recentes do Grafana:

```
skopeo sync --src docker --dest dir registry.suse.com/ses/7.1/ceph/
grafana:latest /root/images/
```

EXEMPLO 7.4: SINCRONIZAR AS IMAGENS MAIS RECENTES DO PROMETHEUS

```
skopeo sync --src docker --dest dir registry.suse.com/ses/7.1/ceph/prometheus-
server:2.32.1 /root/images/
skopeo sync --src docker --dest dir registry.suse.com/ses/7.1/ceph/prometheus-node-
exporter:1.1.2 /root/images/
skopeo sync --src docker --dest dir registry.suse.com/ses/7.1/ceph/prometheus-
alertmanager:0.21.0 /root/images/
```

PROCEDIMENTO 7.2: CONFIGURAR O REGISTRO LOCAL E AS CREDENCIAIS DE ACESSO

1. Configure o URL do registro local:

```
cephuser@adm > ceph-salt config /containers/registry_auth/registry set REG_HOST_URL
```

2. Configure o nome de usuário e a senha para acessar o registro local:

```
cephuser@adm > ceph-salt config /containers/registry_auth/username set REG_USERNAME
```

```
cephuser@adm > ceph-salt config /containers/registry_auth/password set REG_PASSWORD
```

Dica: Cache de registro

Para evitar a ressincronização do registro local quando novos containers atualizados forem exibidos, você pode configurar um *cache de registro*.

7.2.10.2 Configurando o caminho para imagens de container

Importante

Esta seção ajuda você a configurar o caminho para as imagens de container do cluster de boot (implantação do primeiro par de Ceph Monitor e Ceph Manager). O caminho não se aplica a imagens de container de serviços adicionais, por exemplo, a pilha de monitoramento.

Dica: Configurando o proxy HTTPS

Se você precisa usar um proxy para se comunicar com o servidor de registro do container, execute as seguintes etapas de configuração em todos os nós do cluster:

1. Copie o arquivo de configuração dos containers:

```
> sudo cp /usr/share/containers/containers.conf /etc/containers/
containers.conf
```

2. Edite o arquivo recém-copiado e adicione a configuração `http_proxy` a esta seção `[engine]`, por exemplo:

```
> cat /etc/containers/containers.conf
 [engine]
 http_proxy=proxy.example.com
 [...]
```

O cephadm precisa saber um caminho de URI válido para as imagens de container. Verifique a configuração padrão executando:

```
root@master # ceph-salt config /cephadm_bootstrap/ceph_image_path ls
```

Se você não precisa de um registro alternativo ou local, especifique o registro de container do SUSE padrão:

```
root@master # ceph-salt config /cephadm_bootstrap/ceph_image_path set registry.suse.com/
ses/7.1/ceph/ceph
```

Se a implantação exigir um caminho específico, por exemplo, um caminho para um registro local, configure-o da seguinte maneira:

```
root@master # ceph-salt config /cephadm_bootstrap/ceph_image_path set LOCAL_REGISTRY_PATH
```

7.2.11 Habilitando a criptografia de dados em trânsito (msgr2)

O Messenger v2 (MSGR2) é o protocolo on-wire do Ceph. Ele oferece um modo de segurança que criptografa todos os dados que passam pela rede, encapsulamento de payloads de autenticação e habilitação da integração futura de novos modos de autenticação (como Kerberos).

Importante

Atualmente, o msgr2 não é suportado pelos clientes Ceph do kernel do Linux, como CephFS e Dispositivo de Blocos RADOS.

Os daemons do Ceph podem se vincular a várias portas, permitindo que os clientes Ceph legados e os novos clientes compatíveis com a versão 2 se conectem ao mesmo cluster. Por padrão, os MONs agora se vinculam à nova porta 3300 atribuída pela IANA (CE4h ou 0xCE4) para o novo protocolo v2 e também à porta antiga padrão 6789 para o protocolo v1 legado.

O protocolo v2 (MSGR2) suporta dois modos de conexão:

modo crc

Uma autenticação inicial forte quando a conexão é estabelecida e uma verificação de integridade CRC32C.

modo seguro

Uma autenticação inicial forte quando a conexão é estabelecida e a criptografia completa de todo o tráfego pós-autenticação, incluindo uma verificação de integridade criptográfica.

Para a maioria das conexões, há opções que controlam os modos que são usados:

ms_cluster_mode
: O modo de conexão (ou modos permitidos) usado para comunicação intracluster entre os daemons do Ceph. Se houver vários modos na lista, a preferência será dos que forem listados primeiro.

ms_service_mode
: Uma lista de modos permitidos para os clientes usarem na conexão com o cluster.

ms_client_mode
: Uma lista de modos de conexão, em ordem de preferência, para os clientes usarem (ou permitirem) na comunicação com um cluster do Ceph.

Há um conjunto paralelo de opções que se aplicam especificamente aos monitores, permitindo que os administradores definam requisitos diferentes (geralmente mais seguros) para comunicação com os monitores.

ms_mon_cluster_mode
: O modo de conexão (ou modos permitidos) que será usado entre os monitores.

ms_mon_service_mode
: Uma lista de modos permitidos para clientes ou outros daemons do Ceph usarem na conexão com monitores.

ms_mon_client_mode
: Uma lista de modos de conexão, em ordem de preferência, para clientes ou daemons que não são de monitor usarem na conexão com monitores.

Para habilitar o modo de criptografia MSGR2 durante a implantação, você precisa adicionar algumas opções à configuração do `ceph-salt` antes de executar o **`ceph-salt apply`**.

Para usar o modo `secure`, execute os comandos a seguir.

Adicione a seção global ao `ceph_conf` na ferramenta de configuração do `ceph-salt`:

```
root@master # ceph-salt config /cephadm_bootstrap/ceph_conf add global
```

Defina as seguintes opções:

```
root@master # ceph-salt config /cephadm_bootstrap/ceph_conf/global set ms_cluster_mode
 "secure crc"
```

```
root@master # ceph-salt config /cephadm_bootstrap/ceph_conf/global set ms_service_mode
 "secure crc"
root@master # ceph-salt config /cephadm_bootstrap/ceph_conf/global set ms_client_mode
 "secure crc"
```

Nota

Certifique-se de que `secure` venha antes de `crc`.

Para *forçar* o modo `secure`, execute os seguintes comandos:

```
root@master # ceph-salt config /cephadm_bootstrap/ceph_conf/global set ms_cluster_mode
 secure
root@master # ceph-salt config /cephadm_bootstrap/ceph_conf/global set ms_service_mode
 secure
root@master # ceph-salt config /cephadm_bootstrap/ceph_conf/global set ms_client_mode
 secure
```

Dica: Atualizando as configurações

Para mudar qualquer uma das configurações acima, defina as mudanças de configuração no armazenamento de configuração do monitor. Isso é feito usando o comando **`ceph config set`**.

```
root@master # ceph config set global CONNECTION_OPTION CONNECTION_MODE [--force]
```

Por exemplo:

```
root@master # ceph config set global ms_cluster_mode "secure crc"
```

Para verificar o valor atual, incluindo o valor padrão, execute o seguinte comando:

```
root@master # ceph config get CEPH_COMPONENT CONNECTION_OPTION
```

Por exemplo, para obter o `ms_cluster_mode` dos OSD's, execute:

```
root@master # ceph config get osd ms_cluster_mode
```

7.2.12 Configurando a rede do cluster

Opcionalmente, se você executar uma rede de cluster separada, talvez seja necessário definir o endereço IP da rede do cluster seguido pela parte da máscara de sub-rede após a barra, por exemplo, `192.168.10.22/24`.

Execute os seguintes comandos para habilitar `cluster_network`:

```
root@master # ceph-salt config /cephadm_bootstrap/ceph_conf add global
root@master # ceph-salt config /cephadm_bootstrap/ceph_conf/global set
 cluster_network NETWORK_ADDR
```

7.2.13 Verificando a configuração do cluster

A configuração mínima do cluster foi concluída. Verifique se há erros óbvios:

```
root@master # ceph-salt config ls
o- / ..................................................................... [...]
  o- ceph_cluster ........................................................ [...]
  | o- minions ........................................... [Minions: 5]
  | | o- ses-master.example.com ................................ [admin]
  | | o- ses-min1.example.com ....................... [bootstrap, admin]
  | | o- ses-min2.example.com ................................ [no roles]
  | | o- ses-min3.example.com ................................ [no roles]
  | | o- ses-min4.example.com ................................ [no roles]
  | o- roles ..................................................... [...]
  |   o- admin ........................................... [Minions: 2]
  |   | o- ses-master.example.com ...................... [no other roles]
  |   | o- ses-min1.example.com ................. [other roles: bootstrap]
  |   o- bootstrap ............................... [ses-min1.example.com]
  |   o- cephadm ........................................... [Minions: 5]
  |   o- tuned ..................................................... [...]
  |     o- latency ....................................... [no minions]
  |     o- throughput .................................... [no minions]
  o- cephadm_bootstrap ..................................................... [...]
  | o- advanced ..................................................... [...]
  | o- ceph_conf .................................................... [...]
  | o- ceph_image_path .............. [registry.suse.com/ses/7.1/ceph/ceph]
  | o- dashboard .................................................... [...]
  |   o- force_password_update ............................... [enabled]
  |   o- password ................................ [randomly generated]
  |   o- username ................................................ [admin]
  | o- mon_ip ........................................... [192.168.10.20]
  o- containers ................................................... [...]
  | o- registries_conf ..................................... [enabled]
```

```
  | | o- registries ..................................... [empty]
  | o- registry_auth ..................................... [...]
  |   o- password ..................................... [not set]
  |   o- registry ..................................... [not set]
  |   o- username ..................................... [not set]
  o- ssh ..................................... [Key Pair set]
  | o- private_key ..... [53:b1:eb:65:d2:3a:ff:51:6c:e2:1b:ca:84:8e:0e:83]
  | o- public_key ...... [53:b1:eb:65:d2:3a:ff:51:6c:e2:1b:ca:84:8e:0e:83]
  o- time_server ..................................... [enabled]
    o- external_servers ..................................... [1]
    | o- 0.pt.pool.ntp.org ..................................... [...]
    o- servers ..................................... [1]
    | o- ses-master.example.com ..................................... [...]
    o- subnet ..................................... [10.20.6.0/24]
```

Dica: Status da configuração do cluster

Você pode verificar se a configuração do cluster é válida executando o seguinte comando:

```
root@master # ceph-salt status
cluster: 5 minions, 0 hosts managed by cephadm
config: OK
```

7.2.14 Exportando as configurações do cluster

Depois de configurar o cluster básico e sua configuração estiver válida, convém exportá-la para um arquivo:

```
root@master # ceph-salt export > cluster.json
```

Atenção

A saída do comando **`ceph-salt export`** inclui a chave privada SSH. Se você estiver preocupado com as implicações de segurança, não execute esse comando sem tomar as devidas precauções.

Caso você danifique a configuração do cluster e tenha de reverter para um estado de backup, execute:

```
root@master # ceph-salt import cluster.json
```

7.3 Atualizando os nós e o cluster mínimo de boot

Antes de implantar o cluster, atualize todos os pacotes de software em todos os nós:

```
root@master # ceph-salt update
```

Se um nó relatar que a `Reinicialização é necessária` durante a atualização, os pacotes importantes do OS, como o kernel, foram atualizados para uma versão mais recente, e você precisa reinicializar o nó para aplicar as mudanças.

Para reinicializar todos os nós que exigem reinicialização, anexe a opção `--reboot`

```
root@master # ceph-salt update --reboot
```

Se preferir, reinicialize-os em uma etapa separada:

```
root@master # ceph-salt reboot
```

Importante

O Master Salt nunca é reinicializado pelos comandos **`ceph-salt update --reboot`** ou **`ceph-salt reboot`**. Se o Master Salt precisar ser reinicializado, você deverá reinicializá-lo manualmente.

Após a atualização dos nós, inicialize o cluster mínimo:

```
root@master # ceph-salt apply
```

Nota

Quando a inicialização for concluída, o cluster terá um Ceph Monitor e um Ceph Manager.

O comando acima abrirá uma interface do usuário interativa que mostra o andamento atual de cada minion.

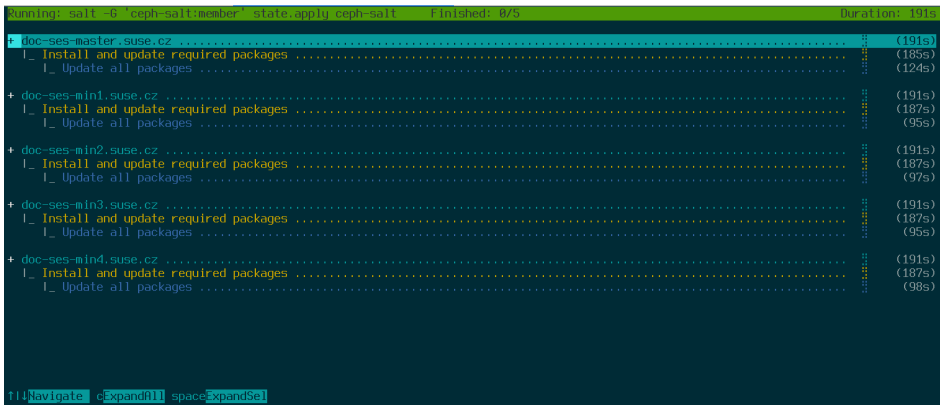

FIGURA 7.1: IMPLANTAÇÃO DE UM CLUSTER MÍNIMO

Dica: Modo não interativo

Se você precisa aplicar a configuração de um script, também há um modo de implantação não interativo. Ele também é útil para implantar o cluster de uma máquina remota, porque a atualização constante das informações de andamento na tela pela rede pode provocar distração:

```
root@master # ceph-salt apply --non-interactive
```

7.4 Revisando as etapas finais

Após a conclusão do comando **`ceph-salt apply`**, você deverá ter um Ceph Monitor e um Ceph Manager. Você deve conseguir executar o comando **`ceph status`** com êxito em qualquer um dos minions que receberam a função `admin` como `root` ou o usuário `cephadm` por meio do `sudo`.

As próximas etapas envolvem o uso do cephadm para implantar mais Ceph Monitor, Ceph Manager, OSDs, Pilha de Monitoramento e Gateways.

Antes de continuar, revise as novas configurações de rede do cluster. Neste ponto, a configuração `public_network` foi preenchida com base no que foi inserido para `/cephadm_bootstrap/mon_ip` na configuração do `ceph-salt`. No entanto, essa configuração foi aplicada apenas ao Ceph Monitor. Você pode revisá-la com o seguinte comando:

```
root@master # ceph config get mon public_network
```

Esse é o mínimo necessário para o Ceph funcionar, mas recomendamos tornar essa configuração `public_network` `global`, o que significa que ela será aplicada a todos os tipos de daemons do Ceph, e não apenas aos MONs:

```
root@master # ceph config set global public_network "$(ceph config get mon
 public_network)"
```

Nota

Essa etapa não é obrigatória. No entanto, se você não usar essa configuração, os Ceph OSDs e outros daemons (exceto o Ceph Monitor) escutarão em *todos os endereços*.

Para que seus OSDs se comuniquem entre si usando uma rede completamente separada, execute o seguinte comando:

```
root@master # ceph config set global cluster_network
 "cluster_network_in_cidr_notation"
```

A execução desse comando garante que os OSDs criados em sua implantação usem a rede de cluster pretendida desde o início.

Se o cluster estiver definido para ter nós densos (mais de 62 OSDs por host), atribua portas suficientes aos Ceph OSDs. Atualmente, a faixa padrão (6800-7300) não permite mais do que 62 OSDs por host. Para um cluster com nós densos, ajuste a configuração `ms_bind_port_max` para um valor adequado. Cada OSD consumirá oito portas adicionais. Por exemplo, um host definido para executar 96 OSDs requer 768 portas. `ms_bind_port_max` deve ser definido, no mínimo, como 7568 executando o seguinte comando:

```
root@master # ceph config set osd.* ms_bind_port_max 7568
```

Você precisará ajustar as configurações de firewall de acordo para que isso funcione. Consulte o *Livro “Troubleshooting Guide”, Capítulo 13 “Hints and tips”, Seção 13.7 “Firewall settings for Ceph”* para obter mais informações.

7.5 Desabilitar clientes não seguros

Desde o Pacific v15.2.11, um novo aviso de saúde foi implementado para informar a você que clientes não seguros têm permissão para ingressar no cluster. Por padrão, esse aviso está *ativado*. O Ceph Dashboard mostrará o cluster no status `HEALTH_WARN`, e a verificação do status do cluster na linha de comando informará o seguinte:

```
cephuser@adm > ceph status
cluster:
  id:     3fe8b35a-689f-4970-819d-0e6b11f6707c
  health: HEALTH_WARN
  mons are allowing insecure global_id reclaim
[...]
```

Esse aviso significa que os Ceph Monitors ainda permitem que clientes antigos e sem patch se conectem ao cluster. Isso garante que os clientes existentes ainda consigam se conectar durante o upgrade do cluster, mas avisa você de que há um problema que precisa ser resolvido. Quando o upgrade do cluster e de todos os clientes for feito para a versão mais recente do Ceph, execute o seguinte comando para não permitir os clientes sem patch:

```
cephuser@adm > ceph config set mon auth_allow_insecure_global_id_reclaim false
```

8 Implantando os serviços principais restantes com o cephadm

Após implantar o cluster básico do Ceph, implante os serviços principais em mais nós do cluster. Para tornar os dados do cluster acessíveis aos clientes, implante também mais serviços.

Atualmente, oferecemos suporte à implantação de serviços do Ceph na linha de comando usando o orquestrador do Ceph (subcomandos **`ceph orch`**).

8.1 O comando **`ceph orch`**

O comando do orquestrador do Ceph **`ceph orch`**, uma interface com o módulo cephadm, processará a listagem de componentes do cluster e a implantação dos serviços do Ceph em novos nós do cluster.

8.1.1 Exibindo o status do orquestrador

O comando a seguir mostra o modo e o status atuais do orquestrador do Ceph.

```
cephuser@adm > ceph orch status
```

8.1.2 Listando dispositivos, serviços e daemons

Para listar todos os dispositivos de disco, execute o seguinte:

```
cephuser@adm > ceph orch device ls
Hostname   Path      Type  Serial  Size   Health   Ident  Fault  Available
ses-master /dev/vdb  hdd   0d8a... 10.7G  Unknown  N/A    N/A    No
ses-min1   /dev/vdc  hdd   8304... 10.7G  Unknown  N/A    N/A    No
ses-min1   /dev/vdd  hdd   7b81... 10.7G  Unknown  N/A    N/A    No
[...]
```

Dica: Serviços e daemons

Serviço é um termo geral para um serviço do Ceph de um tipo específico, por exemplo, Ceph Manager.

Daemon é uma instância específica de um serviço, por exemplo, um processo `mgr.ses-min1.gdlcik` executado em um nó denominado `ses-min1`.

Para listar todos os serviços conhecidos pelo cephadm, execute:

```
cephuser@adm > ceph orch ls
NAME  RUNNING  REFRESHED  AGE  PLACEMENT  IMAGE NAME                  IMAGE ID
mgr       1/0  5m ago     -    <no spec>  registry.example.com/[...]  5bf12403d0bd
mon       1/0  5m ago     -    <no spec>  registry.example.com/[...]  5bf12403d0bd
```

Dica

Você pode limitar a lista a serviços em um nó específico com o parâmetro opcional `--host` e a serviços de um determinado tipo com o parâmetro opcional `--service-type`. Os tipos aceitáveis são `mon`, `osd`, `mgr`, `mds` e `rgw`.

Para listar todos os daemons em execução implantados pelo cephadm, execute:

```
cephuser@adm > ceph orch ps
NAME            HOST     STATUS   REFRESHED AGE VERSION    IMAGE ID     CONTAINER ID
mgr.ses-min1.gd ses-min1 running) 8m ago    12d 15.2.0.108 5bf12403d0bd b8104e09814c
mon.ses-min1    ses-min1 running) 8m ago    12d 15.2.0.108 5bf12403d0bd a719e0087369
```

Dica

Para consultar o status de um daemon específico, use `--daemon_type` e `--daemon_id`. Para os OSDs, o ID é o ID numérico do OSD. Para o MDS, o ID é o nome do sistema de arquivos:

```
cephuser@adm > ceph orch ps --daemon_type osd --daemon_id 0
cephuser@adm > ceph orch ps --daemon_type mds --daemon_id my_cephfs
```

8.2 Especificação de serviço e posicionamento

A maneira recomendada de especificar a implantação dos serviços do Ceph é criar um arquivo no formato YAML com a especificação dos serviços que você pretende implantar.

8.2.1 Criando especificações de serviço

Você pode criar um arquivo de especificação separado para cada tipo de serviço, por exemplo:

```
root@master # cat nfs.yml
service_type: nfs
service_id: EXAMPLE_NFS
placement:
  hosts:
  - ses-min1
  - ses-min2
spec:
  pool: EXAMPLE_POOL
  namespace: EXAMPLE_NAMESPACE
```

Se preferir, você poderá especificar vários (ou todos) tipos de serviço em um arquivo (por exemplo, `cluster.yml`), que descreve quais nós executarão serviços específicos. Lembre-se de separar os tipos de serviço individuais com três traços (`---`):

```
cephuser@adm > cat cluster.yml
service_type: nfs
service_id: EXAMPLE_NFS
placement:
  hosts:
  - ses-min1
  - ses-min2
spec:
  pool: EXAMPLE_POOL
  namespace: EXAMPLE_NAMESPACE
---
service_type: rgw
service_id: REALM_NAME.ZONE_NAME
placement:
  hosts:
  - ses-min1
  - ses-min2
  - ses-min3
---
[...]
```

As propriedades mencionadas acima têm o seguinte significado:

`service_type`

O tipo de serviço. Ele pode ser um serviço do Ceph (`mon`, `mgr`, `mds`, `crash`, `osd` ou `rbd-mirror`), um gateway (`nfs` ou `rgw`) ou parte da pilha de monitoramento (`alertmanager`, `grafana`, `node-exporter` ou `prometheus`).

`service_id`

O nome do serviço. As especificações do tipo `mon`, `mgr`, `alertmanager`, `grafana`, `node-exporter` e `prometheus` não exigem a propriedade `service_id`.

`placement`

Especifica os nós que executarão o serviço. Consulte a *Seção 8.2.2, "Criando a especificação de posicionamento"* para obter mais detalhes.

`spec`

Especificação adicional relevante para o tipo de serviço.

Dica: Aplicando serviços específicos

Geralmente, os serviços de cluster do Ceph têm várias propriedades específicas. Para obter exemplos e detalhes da especificação de cada serviço, consulte a *Seção 8.3, "Implantar serviços do Ceph"*.

8.2.2 Criando a especificação de posicionamento

Para implantar os serviços do Ceph, o cephadm precisa saber em quais nós implantá-los. Use a propriedade `placement` e liste os nomes abreviados de host dos nós aos quais o serviço se aplica:

```
cephuser@adm > cat cluster.yml
[...]
placement:
  hosts:
  - host1
  - host2
  - host3
[...]
```

8.2.3 Aplicando a especificação de cluster

Após criar um arquivo `cluster.yml` completo com as especificações de todos os serviços e seu posicionamento, você poderá aplicar o cluster executando o seguinte comando:

```
cephuser@adm > ceph orch apply -i cluster.yml
```

Para ver o status do cluster, execute o comando **`ceph orch status`**. Para ver mais detalhes, consulte a *Seção 8.1.1, "Exibindo o status do orquestrador"*.

8.2.4 Exportando a especificação de um cluster em execução

Embora você tenha implantado serviços no cluster do Ceph usando os arquivos de especificação, conforme descrito na *Seção 8.2, "Especificação de serviço e posicionamento"*, a configuração do cluster pode divergir da especificação original durante sua operação. Além disso, você pode ter removido os arquivos de especificação por engano.

Para recuperar uma especificação completa de um cluster em operação, execute:

```
cephuser@adm > ceph orch ls --export
placement:
  hosts:
  - hostname: ses-min1
    name: ''
    network: ''
service_id: my_cephfs
service_name: mds.my_cephfs
service_type: mds
---
placement:
  count: 2
service_name: mgr
service_type: mgr
---
[...]
```

Dica

Você pode anexar a opção `--format` para mudar o formato de saída padrão `yaml`. Você pode selecionar entre `json`, `json-pretty` ou `yaml`. Por exemplo:

```
ceph orch ls --export --format json
```

8.3 Implantar serviços do Ceph

Depois que o cluster básico estiver em execução, você poderá implantar os serviços do Ceph nos outros nós.

8.3.1 Implantando Ceph Monitors e Ceph Managers

O cluster do Ceph tem três ou cinco MONs implantados em nós diferentes. Se houver cinco ou mais nós no cluster, recomendamos a implantação de cinco MONs. Uma boa prática é implantar os MGRs nos mesmos nós que os MONs.

Importante: Incluir MON de boot

Ao implantar MONs e MGRs, lembre-se de incluir o primeiro MON que você adicionou durante a configuração do cluster básico na *Seção 7.2.5, "Especificando o primeiro nó MON/MGR"*.

Para implantar MONs, use a seguinte especificação:

```
service_type: mon
placement:
  hosts:
  - ses-min1
  - ses-min2
  - ses-min3
```

Nota

Se você precisar adicionar outro nó, anexe o nome de host à mesma lista YAML. Por exemplo:

```
service_type: mon
placement:
 hosts:
 - ses-min1
 - ses-min2
 - ses-min3
 - ses-min4
```

Da mesma forma, para implantar MGRs, use a seguinte especificação:

Importante

Verifique se a implantação tem pelo menos três Ceph Managers em cada implantação.

```
service_type: mgr
placement:
  hosts:
  - ses-min1
  - ses-min2
  - ses-min3
```

Dica

Se os MONs ou MGRs *não* estiverem na mesma sub-rede, você precisará anexar os endereços das sub-redes. Por exemplo:

```
service_type: mon
placement:
  hosts:
  - ses-min1:10.1.2.0/24
  - ses-min2:10.1.5.0/24
  - ses-min3:10.1.10.0/24
```

8.3.2 Implantando Ceph OSDs

Importante: Quando o dispositivo de armazenamento está disponível

Um dispositivo de armazenamento será considerado *disponível* se todas as condições abaixo forem verdadeiras:

- O dispositivo não tem partições.
- O dispositivo não tem nenhum estado de LVM.
- O dispositivo não está montado.
- O dispositivo não contém um sistema de arquivos.
- O dispositivo não contém um OSD com BlueStore.
- O dispositivo é maior do que 5 GB.

Se as condições acima não forem verdadeiras, o Ceph se recusará a provisionar esses OSDs.

A implantação de OSDs pode ser feita de duas maneiras:

- Instrua o Ceph a consumir todos os dispositivos de armazenamento disponíveis e não utilizados:

```
cephuser@adm > ceph orch apply osd --all-available-devices
```

- Use o DriveGroups (consulte o *Livro "Guia de Administração e Operações", Capítulo 13 "Tarefas operacionais", Seção 13.4.3 "Adicionando OSDs por meio da especificação DriveGroups"*) para criar uma especificação de OSD que descreva os dispositivos que serão implantados com base em suas propriedades, como tipo de dispositivo (SSD ou HDD), nomes de modelo de dispositivo e tamanho, ou os nós onde os dispositivos residem. Em seguida, aplique a especificação executando o seguinte comando:

```
cephuser@adm > ceph orch apply osd -i drive_groups.yml
```

8.3.3 Implantando servidores de metadados

O CephFS requer um ou mais serviços do Servidor de Metadados (MDS, Metadata Server). Para criar um CephFS, primeiro crie os servidores MDS usando a seguinte especificação:

Nota

Verifique se você já criou pelo menos dois pools (um para os dados do CephFS e outro para os metadados do CephFS) antes de usar a especificação a seguir.

```
service_type: mds
service_id: CEPHFS_NAME
placement:
  hosts:
  - ses-min1
  - ses-min2
  - ses-min3
```

Depois que os MDSs estiverem em operação, crie o CephFS:

```
ceph fs new CEPHFS_NAME metadata_pool data_pool
```

8.3.4 Implantando Gateways de Objetos

O cephadm implanta um Gateway de Objetos como uma coleção de daemons que gerenciam um *domínio* e uma *zona* específicos.

Você pode relacionar um serviço de Gateway de Objetos a um domínio e uma zona existentes (consulte o *Livro "Guia de Administração e Operações", Capítulo 21 "Gateway de Objetos do Ceph", Seção 21.13 "Gateways de Objetos multissite"* para obter mais detalhes) ou especificar um *REALM_NAME* e *ZONE_NAME* não existentes, e eles serão criados automaticamente depois que você aplicar a seguinte configuração:

```
service_type: rgw
service_id: REALM_NAME.ZONE_NAME
placement:
  hosts:
  - ses-min1
  - ses-min2
  - ses-min3
spec:
  rgw_realm: RGW_REALM
  rgw_zone: RGW_ZONE
```

8.3.4.1 Usando o acesso SSL seguro

Para usar uma conexão SSL segura com o Gateway de Objetos, você precisa de um par de arquivos de chave e certificado SSL válidos (consulte o *Livro "Guia de Administração e Operações", Capítulo 21 "Gateway de Objetos do Ceph", Seção 21.7 "Habilitar HTTPS/SSL para Gateways de Objetos"* para obter mais detalhes). Você precisa habilitar o SSL e especificar um número de porta para conexões SSL e os arquivos de chave e certificado SSL.

Para habilitar o SSL e especificar o número da porta, inclua o seguinte em sua especificação:

```
spec:
  ssl: true
  rgw_frontend_port: 443
```

Para especificar a chave e o certificado SSL, você pode colar o conteúdo deles diretamente no arquivo de especificação YAML. O sinal de barra vertical (|) no final da linha informa ao analisador para esperar um valor de string de várias linhas. Por exemplo:

```
spec:
  ssl: true
```

```
rgw_frontend_port: 443
rgw_frontend_ssl_certificate: |
 -----BEGIN CERTIFICATE-----
 MIIFmjCCA4KgAwIBAgIJAIZ2n35bmwXTMA0GCSqGSIb3DQEBCwUAMGIxCzAJBgNV
 BAYTAkFVMQwwCgYDVQQIDANOU1cxHTAbBgNVBAoMFEV4YW1wbGUgUkdXIFNTTCBp
 [...]
 -----END CERTIFICATE-----
 rgw_frontend_ssl_key: |
 -----BEGIN PRIVATE KEY-----
 MIIJRAIBADANBgkqhkiG9w0BAQEFAASCCS4wggkqAgEAAoICAQDLtFwg6LLl2j4Z
 BDV+iL4AO7VZ9KbmWIt37Ml2W6y2YeKX3Qwf+3eBz7TVHR1dm6iPpCpqpQjXUsT9
 [...]
 -----END PRIVATE KEY-----
```

Dica

Em vez de colar o conteúdo dos arquivos de chave e certificado SSL, você pode omitir as palavras-chave `rgw_frontend_ssl_certificate:` e `rgw_frontend_ssl_key:` e fazer upload deles no banco de dados de configuração:

```
cephuser@adm > ceph config-key set rgw/cert/REALM_NAME/ZONE_NAME.crt \
 -i SSL_CERT_FILE
cephuser@adm > ceph config-key set rgw/cert/REALM_NAME/ZONE_NAME.key \
 -i SSL_KEY_FILE
```

8.3.4.1.1 Configurar o Gateway de Objetos para escutar nas portas 443 e 80

Para configurar o Gateway de Objetos para escutar nas portas 443 (HTTPS) e 80 (HTTP), siga estas etapas:

Nota

Os comandos no procedimento usam domínio e zona `default`.

1. Forneça um arquivo de especificação para implantar o Gateway de Objetos. Consulte a *Seção 8.3.4, "Implantando Gateways de Objetos"* para obter mais detalhes sobre a especificação do Gateway de Objetos. Utilize o seguinte comando:

   ```
   cephuser@adm > ceph orch apply -i SPEC_FILE
   ```

2. Se os certificados SSL não forem fornecidos no arquivo de especificação, adicione-os usando o seguinte comando:

```
cephuser@adm > ceph config-key set rgw/cert/default/default.crt -i certificate.pem
cephuser@adm > ceph config-key set rgw/cert/default/default.key -i key.pem
```

3. Mude o valor padrão da opção `rgw_frontends`:

```
cephuser@adm > ceph config set client.rgw.default.default rgw_frontends \
 "beast port=80 ssl_port=443"
```

4. Remova a configuração específica criada pelo cephadm. Identifique para qual destino a opção `rgw_frontends` foi configurada executando o comando:

```
cephuser@adm > ceph config dump | grep rgw
```

   Por exemplo, o destino é `client.rgw.default.default.node4.yiewdu`. Remova o valor específico atual de `rgw_frontends`:

```
cephuser@adm > ceph config rm client.rgw.default.default.node4.yiewdu rgw_frontends
```

Dica

Em vez de remover um valor de `rgw_frontends`, você pode especificá-lo. Por exemplo:

```
cephuser@adm > ceph config set client.rgw.default.default.node4.yiewdu \
 rgw_frontends "beast port=80 ssl_port=443"
```

5. Reinicie os Gateways de Objetos:

```
cephuser@adm > ceph orch restart rgw.default.default
```

8.3.4.2 Implantação com um subcluster

Os *subclusters* ajudam a organizar os nós nos clusters para isolar as cargas de trabalho e facilitar a expansão elástica. Se a implantação for com um subcluster, use a seguinte configuração:

```
service_type: rgw
service_id: REALM_NAME.ZONE_NAME.SUBCLUSTER
```

```
placement:
  hosts:
  - ses-min1
  - ses-min2
  - ses-min3
spec:
  rgw_realm: RGW_REALM
  rgw_zone: RGW_ZONE
  subcluster: SUBCLUSTER
```

8.3.5 Implantando Gateways iSCSI

O cephadm implanta um Gateway iSCSI, que é um protocolo SAN (Storage Area Network) que permite aos clientes (denominados iniciadores) enviar comandos SCSI para dispositivos de armazenamento SCSI (destinos) em servidores remotos.

Use a seguinte configuração para a implantação. Verifique se `trusted_ip_list` contém os endereços IP de todos os nós do Gateway iSCSI e do Ceph Manager (consulte o exemplo de saída abaixo).

Nota

Verifique se o pool foi criado antes de aplicar a especificação a seguir.

```
service_type: iscsi
service_id: EXAMPLE_ISCSI
placement:
  hosts:
  - ses-min1
  - ses-min2
  - ses-min3
spec:
  pool: EXAMPLE_POOL
  api_user: EXAMPLE_USER
  api_password: EXAMPLE_PASSWORD
  trusted_ip_list: "IP_ADDRESS_1,IP_ADDRESS_2"
```

Nota

Verifique se os IPs listados em `trusted_ip_list` *não* têm espaço após a separação por vírgula.

8.3.5.1 Configuração SSL segura

Para usar uma conexão SSL segura entre o Ceph Dashboard e a API de destino iSCSI, você precisa de um par de arquivos de chave e certificado SSL válidos. Eles podem ser emitidos por CA ou autoassinados (consulte o *Livro "Guia de Administração e Operações", Capítulo 10 "Configuração manual", Seção 10.1.1 "Criando certificados autoassinados"*). Para habilitar o SSL, inclua a configuração `api_secure: true` no arquivo de especificação:

```
spec:
  api_secure: true
```

Para especificar a chave e o certificado SSL, você pode colar o conteúdo diretamente no arquivo de especificação YAML. O sinal de barra vertical (`|`) no final da linha informa ao analisador para esperar um valor de string de várias linhas. Por exemplo:

```
spec:
  pool: EXAMPLE_POOL
  api_user: EXAMPLE_USER
  api_password: EXAMPLE_PASSWORD
  trusted_ip_list: "IP_ADDRESS_1,IP_ADDRESS_2"
  api_secure: true
  ssl_cert: |
    -----BEGIN CERTIFICATE-----
    MIIDtTCCAp2gAwIBAgIYMC4xNzc1NDQxNjEzMzc2MjMyXzxvQ7EcMA0GCSqGSIb3
    DQEBCwUAMG0xCzAJBgNVBAYTAlVTMQ0wCwYDVQQIDARVdGFoMRcwFQYDVQQHDA5T
    [...]
    -----END CERTIFICATE-----
  ssl_key: |
    -----BEGIN PRIVATE KEY-----
    MIIEvQIBADANBgkqhkiG9w0BAQEFAASCBKcwggSjAgEAAoIBAQC5jdYbjtNTAKW4
    /CwQr/7wOiLGzVxChn3mmCIF3DwbL/qvTFTX2d8bDf6LjGwLYloXHscRfxszX/4h
    [...]
    -----END PRIVATE KEY-----
```

8.3.6 Implantando o NFS Ganesha

Importante

O NFS Ganesha suporta o NFS versão 4.1 e mais recente. Ele não suporta o NFS versão 3.

O cephadm implanta o NFS Ganesha usando um pool RADOS predefinido e um namespace opcional. Para implantar o NFS Ganesha, use a seguinte especificação:

Nota

Você precisa ter um pool RADOS predefinido; do contrário, haverá falha na operação **`ceph orch apply`**. Para obter mais informações sobre como criar um pool, consulte o *Livro "Guia de Administração e Operações", Capítulo 18 "Gerenciar pools de armazenamento", Seção 18.1 "Criando um pool"*.

```
service_type: nfs
service_id: EXAMPLE_NFS
placement:
  hosts:
  - ses-min1
  - ses-min2
spec:
  pool: EXAMPLE_POOL
  namespace: EXAMPLE_NAMESPACE
```

- *`EXAMPLE_NFS`* com uma string arbitrária que identifica a exportação do NFS.
- *`EXAMPLE_POOL`* com o nome do pool em que o objeto de configuração RADOS do NFS Ganesha será armazenado.
- *`EXAMPLE_NAMESPACE`* (opcional) com o namespace desejado do NFS do Gateway de Objetos (por exemplo, `ganesha`).

8.3.7 Implantação `rbd-mirror`

O serviço `rbd-mirror` se encarrega de sincronizar as imagens do Dispositivo de Blocos RADOS entre dois clusters do Ceph (para obter mais detalhes, consulte o *Livro "Guia de Administração e Operações", Capítulo 20 "Dispositivo de blocos RADOS", Seção 20.4 "Espelhos de imagens RBD"*). Para implantar o `rbd-mirror`, use a seguinte especificação:

```
service_type: rbd-mirror
service_id: EXAMPLE_RBD_MIRROR
placement:
  hosts:
  - ses-min3
```

8.3.8 Implantando a pilha de monitoramento

A pilha de monitoramento consiste no Prometheus, nos exportadores do Prometheus, no Alertmanager do Prometheus e no Grafana. O Ceph Dashboard usa esses componentes para armazenar e visualizar as métricas detalhadas sobre o uso e o desempenho do cluster.

Dica

Se a implantação exigir imagens de container personalizadas ou exibidas localmente dos serviços de pilha de monitoramento, consulte o *Livro "Guia de Administração e Operações", Capítulo 16 "Monitoramento e alerta", Seção 16.1 "Configurando imagens personalizadas ou locais"*.

Para implantar a pilha de monitoramento, siga estas etapas:

1. Habilite o módulo `prometheus` no daemon do Ceph Manager. Esse procedimento expõe as métricas internas do Ceph para que o Prometheus possa lê-las:

```
cephuser@adm > ceph mgr module enable prometheus
```

Nota

Verifique se esse comando foi executado antes da implantação do Prometheus. Se o comando não foi executado antes da implantação, você deve reimplantar o Prometheus para atualizar a configuração do Prometheus:

```
cephuser@adm > ceph orch redeploy prometheus
```

2. Crie um arquivo de especificação (por exemplo, `monitoramento.yaml`) com um conteúdo semelhante ao seguinte:

```
service_type: prometheus
placement:
  hosts:
  - ses-min2
---
service_type: node-exporter
---
service_type: alertmanager
placement:
```

```
  hosts:
  - ses-min4
---
service_type: grafana
placement:
  hosts:
  - ses-min3
```

3. Aplique os serviços de monitoramento executando este comando:

```
cephuser@adm > ceph orch apply -i monitoring.yaml
```

Pode levar um ou dois minutos para que os serviços de monitoramento sejam implantados.

Importante

O Prometheus, o Grafana e o Ceph Dashboard são todos configurados automaticamente para se comunicarem entre si, resultando em uma integração totalmente funcional do Grafana no Ceph Dashboard, quando implantados conforme descrito acima.

A única exceção a essa regra é o monitoramento com imagens RBD. Consulte o *Livro "Guia de Administração e Operações", Capítulo 16 "Monitoramento e alerta", Seção 16.5.4 "Habilitando o monitoramento de imagens RBD"* para obter mais informações.

9 Implantação de serviços adicionais

9.1 Instalação do gateway iSCSI

iSCSI é um protocolo SAN (Storage Area Network) que permite aos clientes (denominados *iniciadores*) enviar comandos SCSI para dispositivos de armazenamento SCSI (*destinos*) em servidores remotos. O SUSE Enterprise Storage 7.1 inclui um recurso que abre o gerenciamento de armazenamento do Ceph para clientes heterogêneos, como Microsoft Windows* e VMware* vSphere, por meio do protocolo iSCSI. O acesso de múltiplos caminhos do iSCSI fornece disponibilidade e escalabilidade para esses clientes, e o protocolo iSCSI padronizado também proporciona uma camada adicional de isolamento de segurança entre os clientes e o cluster do SUSE Enterprise Storage 7.1. O recurso de configuração é denominado `ceph-iscsi`. Ao usar o `ceph-iscsi`, os administradores de armazenamento do Ceph podem definir volumes replicados, aprovisionados dinamicamente e altamente disponíveis com suporte a instantâneos apenas leitura, clones de leitura-gravação e redimensionamento automático com Dispositivo de Blocos RADOS (RBD, RADOS Block Device) do Ceph. Em seguida, os administradores podem exportar os volumes por um host de gateway `ceph-iscsi` único ou por vários hosts de gateway com suporte a failover de múltiplos caminhos. Os hosts do Linux, Microsoft Windows e VMware podem se conectar aos volumes pelo protocolo iSCSI, que os torna disponíveis como qualquer outro dispositivo de blocos SCSI. Isso significa que os clientes do SUSE Enterprise Storage 7.1 podem executar com eficácia um subsistema completo de infraestrutura de armazenamento de blocos no Ceph, que fornece todos os recursos e benefícios de uma SAN convencional, permitindo um crescimento futuro.

Este capítulo apresenta informações detalhadas sobre como configurar uma infraestrutura de cluster do Ceph juntamente com um iSCSI Gateway para que os hosts de clientes possam usar os dados armazenados remotamente como dispositivos de armazenamento local por meio do protocolo iSCSI.

9.1.1 Armazenamento em blocos iSCSI

iSCSI é uma implementação do comando SCSI (Small Computer System Interface) definido pelo Protocolo IP, especificado na RFC 3720. O iSCSI é implementado como um serviço em que o cliente (iniciador) comunica-se com o servidor (destino) por meio de uma sessão na porta TCP

3260. O endereço IP e a porta do destino iSCSI são chamados de *portal iSCSI*, em que um destino pode ser exposto por meio de um ou mais portais. A combinação de um destino e de um ou mais portais é denominada *grupo de portais de destino* (TPG, Target Portal Group).

O protocolo da camada de link de dados subjacente ao iSCSI costuma ser a Ethernet. Mais especificamente, as infraestruturas modernas do iSCSI usam 10 GigE Ethernet, ou redes mais rápidas, para alcançar o throughput ideal. A conectividade 10 Gigabit Ethernet entre o iSCSI Gateway e o cluster de back end do Ceph é altamente recomendada.

9.1.1.1 Destino iSCSI do kernel do Linux

Originalmente, o destino iSCSI do kernel do Linux era chamado de LIO, referente a `linux-iscsi.org`, o domínio original e o site do projeto na Web. Por algum tempo, nada menos do que quatro implementações de destino iSCSI concorrentes estavam disponíveis para a plataforma Linux, mas o LIO acabou prevalecendo como único destino iSCSI de referência. O código de kernel de linha principal do LIO usa o nome simples, porém um pouco ambíguo, "destino", para diferenciar entre "núcleo de destino" e uma variedade de módulos de destino de front end e back end.

Seguramente, o módulo de front end mais usado é o iSCSI. No entanto, o LIO também suporta FC (Fibre Channel), FCoE (Fibre Channel over Ethernet) e vários outros protocolos de front end. Neste momento, apenas o protocolo iSCSI é suportado pelo SUSE Enterprise Storage.

O módulo de back end de destino usado com mais frequência é aquele capaz de simplesmente reexportar qualquer dispositivo de blocos disponível no host de destino. Esse módulo é denominado *iblock*. No entanto, o LIO também tem um módulo de back end específico do RBD que suporta acesso paralelizado de E/S de múltiplos caminhos às imagens RBD.

9.1.1.2 Iniciadores iSCSI

Esta seção apresenta informações resumidas sobre os iniciadores iSCSI usados nas plataformas Linux, Microsoft Windows e VMware.

9.1.1.2.1 Linux

O iniciador padrão para a plataforma Linux é `open-iscsi`. O `open-iscsi` inicia um daemon, `iscsid`, que o usuário pode utilizar para descobrir destinos iSCSI em qualquer portal especificado, efetuar login em destinos e mapear volumes iSCSI. O `iscsid` comunica-se com a

camada intermediária do SCSI para criar dispositivos de blocos no kernel, e que depois o kernel poderá tratar como qualquer outro dispositivo de blocos SCSI no sistema. É possível implantar o iniciador `open-iscsi` em conjunto com o recurso Device Mapper Multipath (`dm-multipath`) para oferecer um dispositivo de blocos iSCSI altamente disponível.

9.1.1.2.2 Microsoft Windows e Hyper-V

O iniciador iSCSI padrão para o sistema operacional Microsoft Windows é o Microsoft iSCSI. O serviço iSCSI pode ser configurado por meio de uma GUI (Graphical User Interface – Interface Gráfica do Usuário) e suporta E/S de múltiplos caminhos para alta disponibilidade.

9.1.1.2.3 VMware

O iniciador iSCSI padrão para o VMware vSphere e o ESX é o do software VMware ESX: `vmkiscsi`. Quando habilitado, ele pode ser configurado pelo cliente do vSphere ou pelo comando **`vmkiscsi-tool`**. Em seguida, você pode formatar volumes de armazenamento conectados por meio do adaptador de armazenamento iSCSI do vSphere com VMFS e usá-los como qualquer outro dispositivo de armazenamento da VM. O iniciador do VMware também suporta E/S de múltiplos caminhos para alta disponibilidade.

9.1.2 Informações gerais sobre `ceph-iscsi`

O `ceph-iscsi` combina os benefícios dos Dispositivos de Blocos RADOS com a versatilidade universal do iSCSI. Ao executar o `ceph-iscsi` em um host de destino iSCSI (conhecido como iSCSI Gateway), qualquer aplicativo que tenha que usar armazenamento de blocos poderá se beneficiar do Ceph, mesmo que ele não reconheça nenhum protocolo de cliente do Ceph. Em vez disso, os usuários podem utilizar o iSCSI ou qualquer outro protocolo de front end de destino para se conectar a um destino LIO, o que converte todas as operações de E/S de destino em armazenamento RBD.

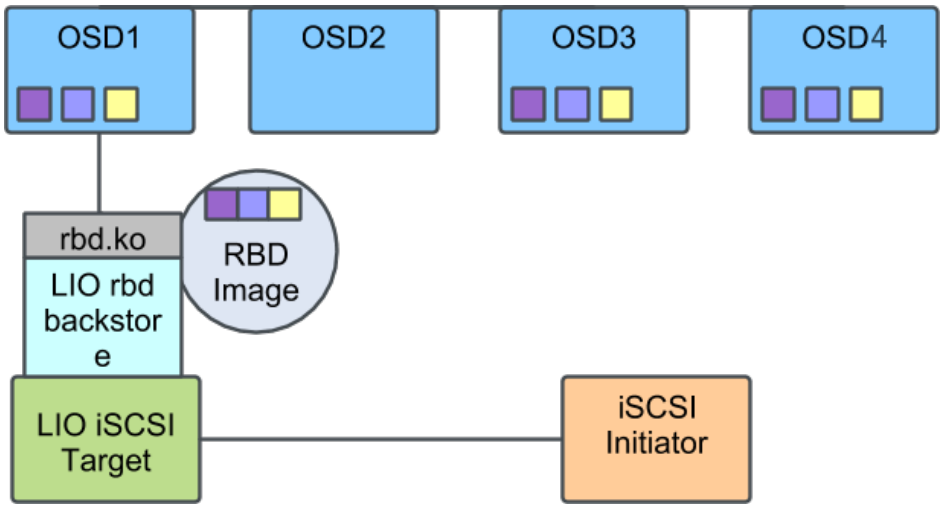

FIGURA 9.1: CLUSTER DO CEPH COM ÚNICO ISCSI GATEWAY

O `ceph-iscsi` apresenta alta disponibilidade inerente e suporta operações de múltiplos caminhos. Portanto, os hosts downstream do iniciador podem usar vários iSCSI Gateways para alta disponibilidade e escalabilidade. Durante a comunicação com uma configuração do iSCSI com mais de um gateway, os iniciadores podem equilibrar a carga das solicitações iSCSI entre vários gateways. Em caso de falha no gateway, estado temporariamente inacessível ou desabilitado para manutenção, a E/S continuará por meio de outro gateway de forma visível.

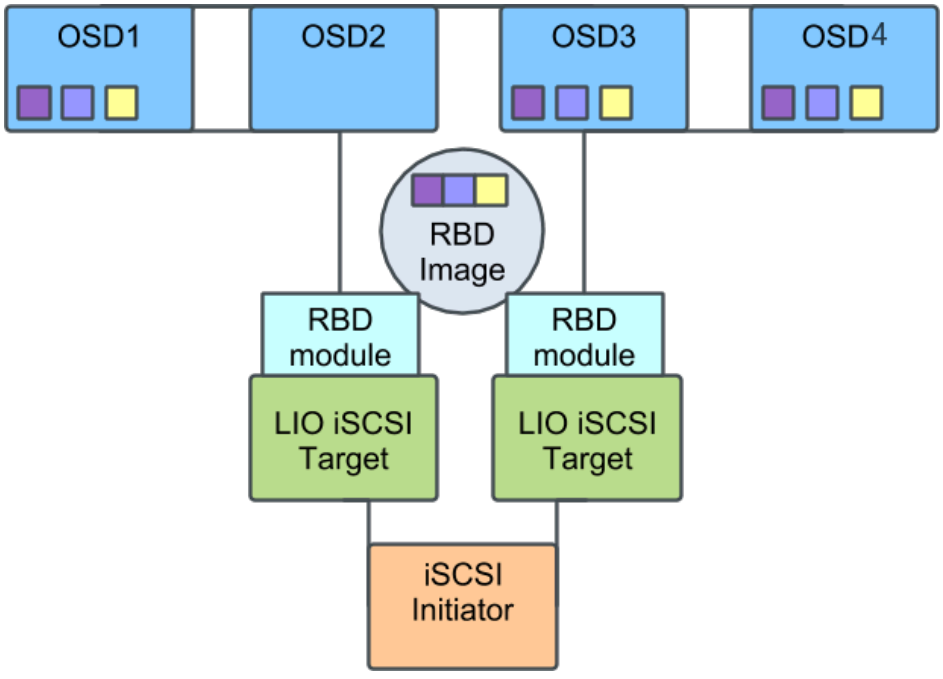

FIGURA 9.2: CLUSTER DO CEPH COM VÁRIOS GATEWAYS ISCSI

9.1.3 Considerações de implantação

Uma configuração mínima do SUSE Enterprise Storage 7.1 com `ceph-iscsi` consiste nos seguintes componentes:

- Um cluster de armazenamento do Ceph. O cluster do Ceph consiste em um mínimo de quatro servidores físicos que hospedam, pelo menos, oito OSDs (Object Storage Daemons – Daemons de Armazenamento de Objetos) cada. Nesse tipo de configuração, três nós OSD também são dobrados como host do monitor (MON).
- Um servidor de destino iSCSI que executa o destino iSCSI do LIO, configurado por meio do `ceph-iscsi`.
- Um host do iniciador iSCSI, executando o `open-iscsi` (Linux), o Iniciador Microsoft iSCSI (Microsoft Windows) ou qualquer outra implementação de iniciador iSCSI compatível.

Uma configuração de produção recomendada do SUSE Enterprise Storage 7.1 com `ceph-iscsi` consiste em:

- Um cluster de armazenamento do Ceph. Um cluster de produção do Ceph consiste em qualquer número de nós OSD (normalmente, mais do que 10), cada um costuma executar de 10 a 12 daemons de armazenamento de objetos (OSDs, Object Storage Daemons), com um mínimo de três hosts MON dedicados.
- Vários servidores de destino iSCSI que executam o destino iSCSI do LIO, configurados por meio do `ceph-iscsi`. Para failover e equilíbrio de carga do iSCSI, esses servidores devem executar um kernel com suporte ao módulo `target_core_rbd`. Os pacotes de atualização estão disponíveis no canal de manutenção do SUSE Linux Enterprise Server.
- Qualquer número de hosts do iniciador iSCSI, executando o `open-iscsi` (Linux), o Iniciador Microsoft iSCSI (Microsoft Windows) ou qualquer outra implementação de iniciador iSCSI compatível.

9.1.4 Instalação e configuração

Esta seção descreve as etapas para instalar e configurar um iSCSI Gateway no SUSE Enterprise Storage.

9.1.4.1 Implantar o gateway iSCSI em um cluster do Ceph

A implantação do Gateway iSCSI do Ceph segue o mesmo procedimento da implantação de outros serviços do Ceph: por meio do cephadm. Para ver mais detalhes, consulte a *Seção 8.3.5, "Implantando Gateways iSCSI"*.

9.1.4.2 Criando imagens RBD

As imagens RBD são criadas no armazenamento do Ceph e, em seguida, exportadas para o iSCSI. É recomendável usar um pool RADOS dedicado para essa finalidade. Você pode criar um volume de qualquer host capaz de se conectar com seu cluster de armazenamento usando o utilitário de linha de comando **`rbd`** do Ceph. Para isso, o cliente deve ter pelo menos um arquivo de configuração mínima `ceph.conf` e as credenciais apropriadas de autenticação no CephX.

Para criar um novo volume para exportação subsequente por meio do iSCSI, use o comando **`rbd create`**, especificando o tamanho do volume em megabytes. Por exemplo, para criar um volume de 100 GB chamado `testvol` no pool denominado `iscsi-images`, execute:

```
cephuser@adm > rbd --pool iscsi-images create --size=102400 testvol
```

9.1.4.3 Exportando imagens RBD por meio do iSCSI

Para exportar imagens RBD por meio do iSCSI, você pode usar a interface da Web Ceph Dashboard ou o utilitário gwcli do `ceph-iscsi`. Nesta seção, vamos nos concentrar apenas no gwcli, demonstrando como criar um destino iSCSI que exporta uma imagem RBD usando a linha de comando.

Nota

As imagens RBD com as seguintes propriedades não podem ser exportadas por meio do iSCSI:

- imagens com o recurso de `registro em diário` habilitado
- imagens com uma `unidade de distribuição` menor do que 4096 bytes

Como `root`, insira o container do Gateway iSCSI:

```
# cephadm enter --name CONTAINER_NAME
```

Como `root`, inicie a interface de linha de comando do Gateway iSCSI:

```
# gwcli
```

Vá para `iscsi-targets` e crie um destino com o nome `iqn.2003-01.org.linux-iscsi.iscsi.SYSTEM-ARCH:testvol`:

```
gwcli >  /> cd /iscsi-targets
gwcli >  /iscsi-targets> create iqn.2003-01.org.linux-iscsi.iscsi.SYSTEM-ARCH:testvol
```

Crie os gateways iSCSI especificando o `nome` e o endereço `ip` deles:

```
gwcli >  /iscsi-targets> cd iqn.2003-01.org.linux-iscsi.iscsi.SYSTEM-ARCH:testvol/
gateways
gwcli >  /iscsi-target...tvol/gateways> create iscsi1 192.168.124.104
gwcli >  /iscsi-target...tvol/gateways> create iscsi2 192.168.124.105
```

Dica

Use o comando `help` para mostrar a lista de comandos disponíveis no nó da configuração atual.

Adicione a imagem RBD com o nome `testvol` no pool `iscsi-images`:

```
gwcli >  /iscsi-target...tvol/gateways> cd /disks
gwcli >  /disks> attach iscsi-images/testvol
```

Mapeie a imagem RBD para o destino:

```
gwcli >  /disks> cd /iscsi-targets/iqn.2003-01.org.linux-iscsi.iscsi.SYSTEM-ARCH:testvol/
disks
gwcli >  /iscsi-target...testvol/disks> add iscsi-images/testvol
```

Nota

Você pode usar ferramentas de nível mais baixo, como **`targetcli`**, para consultar a configuração local, mas sem a modificar.

Dica

Você pode usar o comando **`ls`** para revisar a configuração. Alguns nós da configuração também suportam o comando **`info`**, que podem ser usados para exibir informações mais detalhadas.

Por padrão, a autenticação ACL está habilitada, portanto, esse destino ainda não está acessível. Consulte a *Seção 9.1.4.4, "Autenticação e controle de acesso"* para obter mais informações sobre autenticação e controle de acesso.

9.1.4.4 Autenticação e controle de acesso

A autenticação iSCSI é flexível e inclui muitas possibilidades de autenticação.

9.1.4.4.1 Desabilitando a autenticação ACL

Nenhuma Autenticação significa que qualquer iniciador poderá acessar qualquer LUN no destino correspondente. Você pode habilitar *Nenhuma Autenticação* desabilitando a autenticação ACL:

```
gwcli >  /> cd /iscsi-targets/iqn.2003-01.org.linux-iscsi.iscsi.SYSTEM-ARCH:testvol/hosts
gwcli >  /iscsi-target...testvol/hosts> auth disable_acl
```

9.1.4.4.2 Usando a autenticação ACL

Ao usar a autenticação ACL com base no nome do iniciador, apenas os iniciadores definidos têm permissão para se conectar. Faça o seguinte para definir um iniciador:

```
gwcli >  /> cd /iscsi-targets/iqn.2003-01.org.linux-iscsi.iscsi.SYSTEM-ARCH:testvol/hosts
gwcli >  /iscsi-target...testvol/hosts> create iqn.1996-04.de.suse:01:e6ca28cc9f20
```

Os iniciadores definidos poderão se conectar, mas apenas terão acesso às imagens RBD que foram explicitamente adicionadas aos iniciadores:

```
gwcli >  /iscsi-target...:e6ca28cc9f20> disk add rbd/testvol
```

9.1.4.4.3 Habilitando a autenticação CHAP

Além da ACL, você pode habilitar a autenticação CHAP especificando um nome de usuário e uma senha para cada iniciador:

```
gwcli >  /> cd /iscsi-targets/iqn.2003-01.org.linux-iscsi.iscsi.SYSTEM-ARCH:testvol/
hosts/iqn.1996-04.de.suse:01:e6ca28cc9f20
gwcli >  /iscsi-target...:e6ca28cc9f20> auth username=common12 password=pass12345678
```

Nota

Os nomes de usuário devem ter entre 8 e 64 caracteres e podem incluir caracteres alfanuméricos, `.`, `@`, `-`, `_` ou `:`.

As senhas devem ter entre 12 e 16 caracteres e podem incluir caracteres alfanuméricos, `@`, `-`, `_` ou `/`.

Se preferir, você também poderá habilitar a autenticação CHAP mútua especificando os parâmetros `mutual_username` e `mutual_password` no comando **`auth`**.

9.1.4.4.4 Configurando descoberta e autenticação mútua

A *autenticação Discovery* é independente dos métodos de autenticação anteriores. Ela requer credenciais para navegação, é opcional e pode ser configurada por:

```
gwcli >  /> cd /iscsi-targets
gwcli >  /iscsi-targets> discovery_auth username=du123456 password=dp1234567890
```

Nota

Os nomes de usuário devem ter entre 8 e 64 caracteres e podem incluir apenas letras, `.`, `@`, `-`, `_` ou `:`.

As senhas devem ter entre 12 e 16 caracteres e podem incluir apenas letras, `@`, `-`, `_` ou `/`.

Você também pode especificar os parâmetros `mutual_username` e `mutual_password` no comando **`discovery_auth`**.

Use o seguinte comando para desabilitar a autenticação Discovery:

```
gwcli >  /iscsi-targets> discovery_auth nochap
```

9.1.4.5 Definindo configurações avançadas

É possível configurar o `ceph-iscsi` com parâmetros avançados, que depois são passados para o destino de E/S do LIO. Os parâmetros são divididos nos parâmetros `target` e `disk`.

Atenção

Exceto se indicado de outra forma, não é recomendável mudar a configuração padrão desses parâmetros.

9.1.4.5.1 Visualizando configurações de destino

Você pode ver o valor dessas configurações usando o comando **`info`**:

```
gwcli >  /> cd /iscsi-targets/iqn.2003-01.org.linux-iscsi.iscsi.SYSTEM-ARCH:testvol
gwcli >  /iscsi-target...i.SYSTEM-ARCH:testvol> info
```

E mudar uma configuração usando o comando **`reconfigure`**:

```
gwcli >  /iscsi-target...i.SYSTEM-ARCH:testvol> reconfigure login_timeout 20
```

As configurações de `target` disponíveis são:

default_cmdsn_depth

Profundidade padrão do CmdSN (Command Sequence Number – Número de Sequência do Comando). Limita a quantidade de solicitações pendentes que um iniciador iSCSI pode ter em qualquer momento.

default_erl

Nível de recuperação de erro padrão.

login_timeout

Valor de tempo de espera de login em segundos.

netif_timeout

Tempo de espera de falha da NIC em segundos.

prod_mode_write_protect

Se definida como `1`, impede gravações em LUNs.

9.1.4.5.2 Visualizando configurações de disco

Você pode ver o valor dessas configurações usando o comando **`info`**:

```
gwcli >  /> cd /disks/rbd/testvol
gwcli >  /disks/rbd/testvol> info
```

E mudar uma configuração usando o comando **`reconfigure`**:

```
gwcli >  /disks/rbd/testvol> reconfigure rbd/testvol emulate_pr 0
```

As configurações de `disk` disponíveis são:

block_size
: Tamanho do bloco do dispositivo subjacente.

emulate_3pc
: Se definida como `1`, habilita Cópia de Terceiros.

emulate_caw
: Se definida como `1`, habilita Comparar e Gravar.

emulate_dpo
: Se definida como 1, ativa o recurso Disable Page Out (Desabilitar Saída de Página).

emulate_fua_read
: Se definida como `1` habilita a leitura de Force Unit Access (Forçar Acesso de Unidade).

emulate_fua_write
: Se definida como `1`, habilita a gravação de Force Unit Access (Forçar Acesso de Unidade).

emulate_model_alias
: Se definida como `1`, usa o nome do dispositivo de back end para o álias do modelo.

emulate_pr
: Se definida como 0, o suporte para Reservas SCSI, incluindo Reservas de Grupo Persistentes, será desabilitado. Enquanto estiver desabilitado, o iSCSI Gateway do SES pode ignorar o estado de reserva, resultando em melhor latência de solicitação.

Dica

A recomendação é definir `backstore_emulate_pr` como `0` se os iniciadores iSCSI não exigirem suporte para Reserva SCSI.

emulate_rest_reord

Se definida como `0`, o Modificador de Algoritmo de Fila terá Reordenação Restrita.

emulate_tas

Se definida como `1`, habilita o Status Interrompido da Tarefa.

emulate_tpu

Se definida como `1`, habilita a Anulação do Mapeamento de Aprovisionamento Dinâmico.

emulate_tpws

Se definida como `1`, habilita a Mesma Gravação de Aprovisionamento Dinâmico.

emulate_ua_intlck_ctrl

Se definida como `1`, habilita o Interbloqueio de Atenção de Unidade.

emulate_write_cache

Se definida como `1`, ativa a Habilitação de Gravação de Cache.

enforce_pr_isids

Se definida como `1`, assegura o uso obrigatório de ISIDs de reserva persistentes.

is_nonrot

Se definida como `1`, o backstore será um dispositivo não rotacional.

max_unmap_block_desc_count

Número máximo de descritores de blocos para UNMAP.

max_unmap_lba_count:

Número máximo de LBAs para UNMAP.

max_write_same_len

Tamanho máximo para WRITE_SAME.

optimal_sectors

Tamanho da solicitação ideal em setores.

pi_prot_type

Tipo de proteção DIF.

queue_depth

Profundidade da fila.

unmap_granularity

Granularidade de UNMAP.

unmap_granularity_alignment

Alinhamento da granularidade de UNMAP.

force_pr_aptpl

Quando habilitada, o LIO sempre gravará o estado de *reserva persistente* no armazenamento persistente, mesmo que o cliente não o tenha solicitado usando `aptpl=1`. Isso não tem efeito com o back end do RBD do kernel para LIO, ele sempre mantém o estado de reserva persistente. O ideal é que a opção `target_core_rbd` force-o como "1" e emita um erro se alguém tentar desabilitá-lo pela configuração.

unmap_zeroes_data

Afeta se o LIO divulgará ou não o LBPRZ para os iniciadores SCSI, indicando que os zeros serão lidos novamente de uma região seguindo UNMAP ou WRITE SAME com um bit de anulação de mapeamento.

9.1.5 Exportando imagens de dispositivo de blocos RADOS por meio do `tcmu-runner`

O `ceph-iscsi` suporta os dois backstores `rbd` (com base no kernel) e `user:rbd` (tcmu-runner), tornando todo o gerenciamento transparente e independente do backstore.

Atenção: Prévia de tecnologia

Atualmente, as implantações do iSCSI Gateway com base no `tcmu-runner` estão na versão Technology Preview.

Diferentemente das implantações do iSCSI Gateway baseadas em kernel, os iSCSI Gateways com base em `tcmu-runner` não oferecem suporte a E/S de múltiplos caminhos ou Reservas Persistentes SCSI.

Para exportar a imagem do Dispositivo de Blocos RADOS usando o `tcmu-runner`, tudo o que você precisar fazer é especificar o backstore `user:rbd` ao anexar o disco:

```
gwcli >  /disks> attach rbd/testvol backstore=user:rbd
```

Nota

Ao usar o `tcmu-runner`, o recurso `exclusive-lock` deve estar habilitado na imagem RBD exportada.

III Fazendo upgrade de versões anteriores

10 Upgrade do SUSE Enterprise Storage 6 para 7.1 89

11 Upgrade do SUSE Enterprise Storage 7 para 7.1 112

10 Upgrade do SUSE Enterprise Storage 6 para 7.1

Este capítulo apresenta as etapas para fazer upgrade do SUSE Enterprise Storage 6 para a versão 7.1.

O upgrade inclui as seguintes tarefas:

- Fazer upgrade do Ceph Nautilus para o Pacific.
- Alternar da instalação e execução do Ceph por meio de pacotes RPM para a execução em containers.
- Remoção completa do DeepSea e substituição pelo `ceph-salt` e cephadm.

Atenção

As informações de upgrade neste capítulo são válidas *apenas* para upgrades do DeepSea para o cephadm. Não tente seguir estas instruções para implantar o SUSE Enterprise Storage na Plataforma SUSE CaaS.

Importante

O upgrade de versões do SUSE Enterprise Storage mais antigas do que a 6 não é suportado. Você deve primeiro fazer upgrade para a versão mais recente do SUSE Enterprise Storage 6 e, na sequência, seguir as etapas neste capítulo.

10.1 Antes de fazer upgrade

As tarefas a seguir *devem* ser concluídas antes de você iniciar o upgrade. Isso pode ser feito a qualquer momento durante a vida útil do SUSE Enterprise Storage 6.

- A migração do OSD do FileStore para o BlueStore *deve* ser feita antes do upgrade porque o FileStore é incompatível com o SUSE Enterprise Storage 7.1. Há mais detalhes sobre o BlueStore e como migrar do FileStore em https://documentation.suse.com/ses/6/single-html/ses-deployment/#filestore2bluestore.
- Se você executa um cluster mais antigo que ainda usa OSDs do `ceph-disk`, *precisa* alternar para o `ceph-volume` antes do upgrade. Encontre mais detalhes na https://documentation.suse.com/ses/6/single-html/ses-deployment/#upgrade-osd-deployment.

10.1.1 Pontos a serem considerados

Antes de fazer upgrade, leia as seções a seguir para garantir que você entenda todas as tarefas que precisam ser executadas.

- *Ler os detalhes da versão*. Neles, você encontra mais informações sobre o que mudou desde a versão anterior do SUSE Enterprise Storage. Consulte os detalhes da versão para verificar se:

  - Seu hardware precisa de considerações especiais.
  - Qualquer pacote de software usado foi significativamente modificado.
  - São necessárias precauções especiais para a instalação.

  Os detalhes da versão também apresentam informações que não puderam ser incluídas a tempo no manual. Eles também incluem notas sobre problemas conhecidos.
  Você encontra os detalhes da versão do SES 7.1 online em https://www.suse.com/releasenotes/.
  Além disso, após a instalação do pacote `release-notes-ses` do repositório SES 7.1, encontre os detalhes da versão localmente no diretório `/usr/share/doc/release-notes` ou online em https://www.suse.com/releasenotes/.

- Leia a *Parte II, "Implantando o cluster do Ceph"* para se familiarizar com o `ceph-salt` e o orquestrador do Ceph, especialmente as informações sobre as especificações do serviço.

- O upgrade do cluster pode levar algum tempo, aproximadamente o tempo necessário para fazer upgrade de uma máquina multiplicado pelo número de nós do cluster.

- Você precisa primeiro fazer upgrade do Master Salt e depois substituir o DeepSea pelo `ceph-salt` e cephadm. Você *não* poderá começar a usar o módulo de orquestrador cephadm até que seja feito o upgrade de todos os nós do Ceph Manager.

- O upgrade do uso de RPMs do Nautilus para containers do Pacific precisa ser feito em uma única etapa. Isso significa fazer upgrade de um nó inteiro de uma vez, e não um daemon de cada vez.

- O upgrade dos serviços básicos (MON, MGR, OSD) é feito de forma ordenada. Cada serviço fica disponível durante o upgrade. Os serviços de gateway (Servidor de Metadados, Gateway de Objetos, NFS Ganesha, iSCSI Gateway) precisarão ser reimplantados após o upgrade dos serviços básicos. Há um determinado tempo de espera para cada um dos seguintes serviços:

  - **Importante**

    Os Servidores de Metadados e os Gateways de Objetos ficam inativos a partir do momento em que o upgrade dos nós é feito do SUSE Linux Enterprise Server 15 SP1 para o SUSE Linux Enterprise Server 15 SP3 até a reimplantação dos serviços no final do procedimento de upgrade. É importante ter isso em mente principalmente quando esses serviços estão combinados com MONs, MGRs ou OSDs, porque eles podem ficar inativos durante o upgrade do cluster. Se isso for um problema, considere a implantação desses serviços separadamente em nós adicionais antes do upgrade, de modo que eles fiquem inativos pelo menor tempo possível. Essa é a duração do upgrade dos nós de gateway, não a duração do upgrade de todo o cluster.

  - O NFS Ganesha e os iSCSI Gateways ficam inativos apenas no período de reinicialização dos nós durante o upgrade do SUSE Linux Enterprise Server 15 SP1 para o SUSE Linux Enterprise Server 15 SP3, e rapidamente depois quando cada serviço é reimplantado no modo de container.

10.1.2 Fazendo backup da configuração e dos dados do cluster

É altamente recomendável fazer backup de todas as configurações e os dados do cluster antes de iniciar o upgrade para o SUSE Enterprise Storage 7.1. Para obter instruções sobre como fazer backup de todos os seus dados, consulte o https://documentation.suse.com/ses/6/single-html/ses-admin/#cha-deployment-backup.

10.1.3 Verificando etapas do upgrade anterior

Se você fez upgrade da versão 5, verifique se o upgrade para a versão 6 foi concluído com êxito:

Verifique se existe o arquivo `/srv/salt/ceph/configuration/files/ceph.conf.import`.

Esse arquivo é criado pelo processo de integração durante o upgrade do SUSE Enterprise Storage 5 para 6. A opção `configuration_init: default-import` é definida em `/srv/pillar/ceph/proposals/config/stack/default/ceph/cluster.yml`.

Se `configuration_init` ainda está definido como `default-import`, o cluster usa `ceph.conf.import` como arquivo de configuração, e não o `ceph.conf` padrão do DeepSea, que é compilado dos arquivos em `/srv/salt/ceph/configuration/files/ceph.conf.d/`.

Portanto, você precisa inspecionar se há qualquer configuração personalizada em `ceph.conf.import` e, possivelmente, mover a configuração para um dos arquivos em `/srv/salt/ceph/configuration/files/ceph.conf.d/`.

Em seguida, remova a linha `configuration_init: default-import` de `/srv/pillar/ceph/proposals/config/stack/default/ceph/cluster.yml`.

10.1.4 Atualizando os nós e verificando a saúde do cluster

Verifique se todas as atualizações mais recentes do SUSE Linux Enterprise Server 15 SP1 e do SUSE Enterprise Storage 6 foram aplicadas a todos os nós do cluster:

```
# zypper refresh && zypper patch
```

Dica

Consulte https://documentation.suse.com/ses/6/html/ses-all/storage-salt-cluster.html#deepsea-rolling-updates para obter informações detalhadas sobre como atualizar os nós do cluster.

Depois que as atualizações forem aplicadas, reinicie o Master Salt, sincronize os novos módulos do Salt e verifique a saúde do cluster:

```
root@master # systemctl restart salt-master.service
root@master # salt '*' saltutil.sync_all
cephuser@adm > ceph -s
```

10.1.4.1 Desabilitar clientes não seguros

Desde o Nautilus v14.2.20, um novo aviso de saúde foi implementado para informar a você que clientes não seguros têm permissão para ingressar no cluster. Por padrão, esse aviso está *ativado*. O Ceph Dashboard mostrará o cluster no status `HEALTH_WARN`. A linha de comando verifica o status do cluster da seguinte maneira:

```
cephuser@adm > ceph status
cluster:
  id:     3fe8b35a-689f-4970-819d-0e6b11f6707c
  health: HEALTH_WARN
  mons are allowing insecure global_id reclaim
[...]
```

Esse aviso significa que os Ceph Monitors ainda permitem que clientes antigos e sem patch se conectem ao cluster. Isso garante que os clientes existentes ainda consigam se conectar durante o upgrade do cluster, mas avisa você de que há um problema que precisa ser resolvido. Quando o upgrade do cluster e de todos os clientes for feito para a versão mais recente do Ceph, execute o seguinte comando para não permitir os clientes sem patch:

```
cephuser@adm > ceph config set mon auth_allow_insecure_global_id_reclaim false
```

10.1.5 Verificando o acesso a repositórios do software e imagens de container

Verifique se cada nó do cluster tem acesso aos repositórios do software SUSE Linux Enterprise Server 15 SP3 e SUSE Enterprise Storage 7.1 e também ao registro de imagens de container.

10.1.5.1 Repositórios do software

Se todos os nós foram registrados no SCC, você poderá usar o comando **`zypper migration`** para fazer upgrade. Consulte https://documentation.suse.com/sles/15-SP3/html/SLES-all/cha-upgrade-online.html#sec-upgrade-online-zypper para obter mais detalhes.

Se os nós **não** foram registrados no SCC, desabilite todos os repositórios existentes do software e adicione os dois repositórios `Pool` e `Updates` para cada uma das seguintes extensões:

- SLE-Product-SLES/15-SP3
- SLE-Module-Basesystem/15-SP3

- SLE-Module-Server-Applications/15-SP3
- SUSE-Enterprise-Storage-7.1

10.1.5.2 Imagens de container

Todos os nós do cluster precisam acessar o registro de imagens de container. Na maioria dos casos, você usará o registro público do SUSE em `registry.suse.com`. Você precisa das seguintes imagens:

- registry.suse.com/ses/7.1/ceph/ceph
- registry.suse.com/ses/7.1/ceph/grafana
- registry.suse.com/ses/7.1/ceph/prometheus-server
- registry.suse.com/ses/7.1/ceph/prometheus-node-exporter
- registry.suse.com/ses/7.1/ceph/prometheus-alertmanager

Por exemplo, para implantações isoladas (air-gapped), uma alternativa é configurar um registro local e verificar se você tem o conjunto correto de imagens de container disponível. Consulte a *Seção 7.2.10, "Usando o registro do container"* para obter mais detalhes sobre como configurar um registro de imagens de container local.

10.2 Fazendo upgrade do Master Salt

O procedimento a seguir descreve o processo de upgrade do Master Salt:

1. Faça upgrade do OS subjacente para o SUSE Linux Enterprise Server 15 SP3:
   - Para um cluster em que todos os nós foram registrados no SCC, execute **`zypper migration`**.
   - Para um cluster em que os nós têm repositórios de software atribuídos manualmente, execute **`zypper dup`** seguido de **`reboot`**.
2. Desabilite as fases do DeepSea para evitar o uso acidental. Adicione o seguinte conteúdo a `/srv/pillar/ceph/stack/global.yml`:

```
stage_prep: disabled
```

```
stage_discovery: disabled
stage_configure: disabled
stage_deploy: disabled
stage_services: disabled
stage_remove: disabled
```

Grave o arquivo e aplique as mudanças:

```
root@master # salt '*' saltutil.pillar_refresh
```

3. Se você **não** usa imagens de container do `registry.suse.com`, mas sim o registro configurado localmente, edite o `/srv/pillar/ceph/stack/global.yml` para especificar a imagem de container e o registro do Ceph que o DeepSea usará. Por exemplo, para usar `192.168.121.1:5000/my/ceph/image`, adicione as seguintes linhas:

```
ses7_container_image: 192.168.121.1:5000/my/ceph/image
ses7_container_registries:
  - location: 192.168.121.1:5000
```

Se você precisar especificar informações de autenticação para o registro, adicione o bloco `ses7_container_registry_auth:`, por exemplo:

```
ses7_container_image: 192.168.121.1:5000/my/ceph/image
ses7_container_registries:
  - location: 192.168.121.1:5000
ses7_container_registry_auth:
  registry: 192.168.121.1:5000
  username: USER_NAME
  password: PASSWORD
```

Grave o arquivo e aplique as mudanças:

```
root@master # salt '*' saltutil.refresh_pillar
```

4. Observe a configuração existente:

```
cephuser@adm > ceph config assimilate-conf -i /etc/ceph/ceph.conf
```

5. Verifique o status do upgrade. A saída pode ser diferente dependendo da configuração do seu cluster:

```
root@master # salt-run upgrade.status
The newest installed software versions are:
 ceph: ceph version 16.2.7-640-gceb23c7491b
 (ceb23c7491bd96ab7956111374219a4cdcf6f8f4) pacific (stable)
```

```
 os: SUSE Linux Enterprise Server 15 SP3

Nodes running these software versions:
 admin.ceph (assigned roles: master, prometheus, grafana)

Nodes running older software versions must be upgraded in the following order:
 1: mon1.ceph (assigned roles: admin, mon, mgr)
 2: mon2.ceph (assigned roles: admin, mon, mgr)
 3: mon3.ceph (assigned roles: admin, mon, mgr)
 4: data4.ceph (assigned roles: storage, mds)
 5: data1.ceph (assigned roles: storage)
 6: data2.ceph (assigned roles: storage)
 7: data3.ceph (assigned roles: storage)
 8: data5.ceph (assigned roles: storage, rgw)
```

10.3 Fazendo upgrade dos nós MON, MGR e OSD

Faça upgrade do Ceph Monitor, do Ceph Manager e dos nós OSD, um de cada vez. Para cada serviço, siga estas etapas:

1. Antes de adotar qualquer nó OSD, você precisa executar uma conversão de formato dos nós OSD para melhorar a contabilização dos dados OMAP. Você pode fazer isso executando o seguinte comando no Nó de Admin:

   ```
   cephuser@adm > ceph config set osd bluestore_fsck_quick_fix_on_mount true
   ```

   Os nós OSD serão convertidos automaticamente após o término da adoção.

   Nota

   A conversão pode levar de minutos a horas, dependendo da quantidade de dados OMAP contida no disco rígido relacionado. Para obter mais detalhes, consulte o https://docs.ceph.com/en/latest/releases/pacific/#upgrading-non-cephadm-clusters↗.

2. Se o nó do qual você está fazendo upgrade é um OSD, execute o comando a seguir para evitar marcá-lo como `out`:

   ```
   cephuser@adm > ceph osd add-noout SHORT_NODE_NAME
   ```

Substitua *SHORT_NODE_NAME* pelo nome abreviado do nó da forma como ele aparece na saída do comando **ceph osd tree**. Na entrada a seguir, os nomes abreviados de host são `ses-min1` e `ses-min2`

```
root@master # ceph osd tree
ID   CLASS  WEIGHT   TYPE NAME       STATUS  REWEIGHT  PRI-AFF
 -1         0.60405  root default
-11         0.11691      host ses-min1
  4    hdd  0.01949          osd.4       up   1.00000  1.00000
  9    hdd  0.01949          osd.9       up   1.00000  1.00000
 13    hdd  0.01949          osd.13      up   1.00000  1.00000
[...]
 -5         0.11691      host ses-min2
  2    hdd  0.01949          osd.2       up   1.00000  1.00000
  5    hdd  0.01949          osd.5       up   1.00000  1.00000
[...]
```

3. Faça upgrade do OS subjacente para o SUSE Linux Enterprise Server 15 SP3:

   - Se todos os nós do cluster foram registrados no SCC, execute **zypper migration**.
   - Se os nós do cluster tiveram repositórios de software atribuídos manualmente, execute **zypper dup** seguido de **reboot**.

4. Após a reinicialização do nó, coloque em contêiner todos os daemons MON, MGR e OSD existentes nesse nó executando o seguinte comando no Master Salt:

```
root@master # salt MINION_ID state.apply ceph.upgrade.ses7.adopt
```

Substitua *MINION_ID* pelo ID do minion do qual você está fazendo upgrade. Você pode gerar uma lista de IDs de minions ao executar o comando **salt-key -L** no Master Salt.

Dica

Para ver o status e o andamento da *adoção*, verifique o Ceph Dashboard ou execute um dos seguintes comandos no Master Salt:

```
root@master # ceph status
root@master # ceph versions
root@master # salt-run upgrade.status
```

5. Após a conclusão bem-sucedida da adoção, cancele a definição do flag `noout` se o nó do qual você estiver fazendo upgrade for OSD:

```
cephuser@adm > ceph osd rm-noout SHORT_NODE_NAME
```

10.4 Fazendo upgrade de nós de gateway

A seguir, faça upgrade dos nós de gateway separados (Gateway do Samba, Servidor de Metadados, Gateway de Objetos, NFS Ganesha ou iSCSI Gateway). Faça upgrade do OS subjacente para o SUSE Linux Enterprise Server 15 SP3 para cada nó:

- Se todos os nós do cluster foram registrados no SUSE Customer Center, execute o comando **`zypper migration`**.
- Se os nós do cluster tiveram repositórios de software atribuídos manualmente, execute o comando **`zypper dup`** seguido do comando **`reboot`**.

Esta etapa também vale para qualquer nó que faça parte do cluster, mas que ainda não recebeu nenhuma função (em caso de dúvida, verifique a lista de hosts no Master Salt gerada pelo comando **`salt-key -L`** e compare-a com a saída do comando **`salt-run upgrade.status`**).

Quando o upgrade do OS é feito em todos os nós do cluster, a próxima etapa é instalar o pacote `ceph-salt` e aplicar a configuração do cluster. Os serviços de gateway reais são reimplantados no modo em container no final do procedimento de upgrade.

Nota

Os serviços Servidor de Metadados e Gateway de Objetos ficam indisponíveis a partir do momento do upgrade para o SUSE Linux Enterprise Server 15 SP3 até serem reimplantados no final do procedimento de upgrade.

10.5 Instalando o `ceph-salt` e aplicando a configuração do cluster

Antes de iniciar o procedimento de instalação do `ceph-salt` e aplicar a configuração do cluster, verifique o status do cluster e do upgrade executando os seguintes comandos:

```
root@master # ceph status
```

```
root@master # ceph versions
root@master # salt-run upgrade.status
```

1. Remova os Crons `rbd_exporter` e `rgw_exporter` criados pelo DeepSea. No Master Salt, execute o comando `root`crontab -e **`como`** para editar o crontab. Apague os seguintes itens, se houver:

```
# SALT_CRON_IDENTIFIER:deepsea rbd_exporter cron job
*/5 * * * * /var/lib/prometheus/node-exporter/rbd.sh > \
 /var/lib/prometheus/node-exporter/rbd.prom 2> /dev/null
# SALT_CRON_IDENTIFIER:Prometheus rgw_exporter cron job
*/5 * * * * /var/lib/prometheus/node-exporter/ceph_rgw.py > \
 /var/lib/prometheus/node-exporter/ceph_rgw.prom 2> /dev/null
```

2. Exporte a configuração do cluster do DeepSea executando os seguintes comandos:

```
root@master # salt-run upgrade.ceph_salt_config > ceph-salt-config.json
root@master # salt-run upgrade.generate_service_specs > specs.yaml
```

3. Desinstale o DeepSea e instale o `ceph-salt` no Master Salt:

```
root@master # zypper remove 'deepsea*'
root@master # zypper install ceph-salt
```

4. Reinicie o Master Salt e sincronize os módulos do Salt:

```
root@master # systemctl restart salt-master.service
root@master # salt \* saltutil.sync_all
```

5. Importe a configuração do cluster do DeepSea para `ceph-salt`:

```
root@master # ceph-salt import ceph-salt-config.json
```

6. Gere as chaves SSH para comunicação do nó do cluster:

```
root@master # ceph-salt config /ssh generate
```

Dica

Verifique se a configuração do cluster foi importada do DeepSea e especifique as opções que possam ter sido perdidas:

```
root@master # ceph-salt config ls
```

Para ver uma descrição completa da configuração do cluster, consulte a *Seção 7.2, "Configurando as propriedades do cluster"*.

7. Aplique a configuração e habilite o cephadm:

```
root@master # ceph-salt apply
```

8. Se for necessário inserir o URL de registro do container local e as credenciais de acesso, siga as etapas descritas na *Seção 7.2.10, "Usando o registro do container"*.

9. Se você **não** usa imagens de container do `registry.suse.com`, mas usa o registro configurado localmente, informe a imagem de container que o Ceph usará executando o comando:

```
root@master # ceph config set global container_image IMAGE_NAME
```

Por exemplo:

```
root@master # ceph config set global container_image 192.168.121.1:5000/my/ceph/
image
```

10. Pare e desabilite os daemons `ceph-crash` do SUSE Enterprise Storage 6. Novos formulários em container desses daemons são iniciados automaticamente mais tarde.

```
root@master # salt '*' service.stop ceph-crash
root@master # salt '*' service.disable ceph-crash
```

10.6 Fazendo upgrade e adotando a pilha de monitoramento

O procedimento a seguir adota todos os componentes da pilha de monitoramento (consulte o *Livro "Guia de Administração e Operações", Capítulo 16 "Monitoramento e alerta"* para obter mais detalhes).

1. Pause o orquestrador:

```
cephuser@adm > ceph orch pause
```

2. Em qualquer nó que esteja executando o Prometheus, o Grafana e o Alertmanager (por padrão, o Master Salt), execute os seguintes comandos:

```
cephuser@adm > cephadm adopt --style=legacy --name prometheus.$(hostname)
cephuser@adm > cephadm adopt --style=legacy --name alertmanager.$(hostname)
cephuser@adm > cephadm adopt --style=legacy --name grafana.$(hostname)
```

Dica

Se você **não** executa o registro de imagem de container padrão `registry.suse.com`, precisa especificar a imagem que será usada em cada comando, por exemplo:

```
cephuser@adm > cephadm --image 192.168.121.1:5000/ses/7.1/ceph/prometheus-
server:2.32.1 \
  adopt --style=legacy --name prometheus.$(hostname)
cephuser@adm > cephadm --image 192.168.121.1:5000/ses/7.1/ceph/prometheus-
alertmanager:0.21.0 \
  adopt --style=legacy --name alertmanager.$(hostname)
cephuser@adm > cephadm --image 192.168.121.1:5000/ses/7.1/ceph/grafana:7.5.12
 \
 adopt --style=legacy --name grafana.$(hostname)
```

As imagens de container necessárias e suas respectivas versões estão listadas em *Livro "Guia de Administração e Operações", Capítulo 16 "Monitoramento e alerta", Seção 16.1 "Configurando imagens personalizadas ou locais"*.

3. Remova Node-Exporter de **todos** os nós. O Node-Exporter não precisa ser migrado e será reinstalado como um container quando o arquivo `specs.yaml` for aplicado.

```
> sudo zypper rm golang-github-prometheus-node_exporter
```

Se preferir, remova o Node-Exporter de todos os nós simultaneamente usando o Salt no nó de admin:

```
root@master # salt '*' pkg.remove golang-github-prometheus-node_exporter
```

4. Aplique as especificações de serviço que você já exportou do DeepSea:

```
cephuser@adm > ceph orch apply -i specs.yaml
```

Dica

Se você **não** executa o registro de imagem de container padrão `registry.suse.com`, mas um registro de container local, configure o cephadm para usar a imagem de container do registro local para a implantação do Node-Exporter antes de implantá-lo. Do contrário, você poderá ignorar com segurança esta etapa e o aviso a seguir.

```
cephuser@adm > ceph config set mgr mgr/cephadm/
container_image_node_exporter QUALIFIED_IMAGE_PATH
```

Verifique se todas as imagens de container dos serviços de monitoramento apontam para o registro local, não apenas para o Node-Exporter. Esta etapa requer que você faça isso apenas para o Node-Exporter, mas é recomendável definir todas as imagens de container de monitoramento no cephadm para apontar para o registro local neste ponto.

Se você não fizer isso, as novas implantações de serviços de monitoramento e as reimplantações usarão a configuração cephadm padrão, e você talvez não consiga implantar serviços (no caso de implantações isoladas (air-gapped)) ou tenha serviços implantados com versões misturadas.

Encontre uma descrição de como o cephadm precisa ser configurado para usar imagens de container do registro local no *Livro "Guia de Administração e Operações", Capítulo 16 "Monitoramento e alerta", Seção 16.1 "Configurando imagens personalizadas ou locais"*.

5. Continue o orquestrador:

```
cephuser@adm > ceph orch resume
```

10.7 Reimplantação do serviço de gateway

10.7.1 Fazendo upgrade do Gateway de Objetos

No SUSE Enterprise Storage 7.1, os Gateways de Objetos são sempre configurados com um domínio, o que permite multissite no futuro (consulte o *Livro "Guia de Administração e Operações", Capítulo 21 "Gateway de Objetos do Ceph", Seção 21.13 "Gateways de Objetos multissite"* para obter mais detalhes). Se você usou uma configuração de site único do Gateway de Objetos no SUSE Enterprise Storage 6, siga estas etapas para adicionar um domínio. Se você não planeja usar a funcionalidade multissite, pode usar o `padrão` para os nomes de domínio, grupo de zonas e zona.

1. Crie um novo domínio:

```
cephuser@adm > radosgw-admin realm create --rgw-realm=REALM_NAME --default
```

2. Opcionalmente, renomeie a zona padrão e o grupo de zonas.

```
cephuser@adm > radosgw-admin zonegroup rename \
 --rgw-zonegroup default \
 --zonegroup-new-name=ZONEGROUP_NAME
cephuser@adm > radosgw-admin zone rename \
 --rgw-zone default \
 --zone-new-name ZONE_NAME \
 --rgw-zonegroup=ZONEGROUP_NAME
```

3. Configure o grupo de zonas master:

```
cephuser@adm > radosgw-admin zonegroup modify \
 --rgw-realm=REALM_NAME \
 --rgw-zonegroup=ZONEGROUP_NAME \
 --endpoints http://RGW.EXAMPLE.COM:80 \
 --master --default
```

4. Configure a zona master. Para isso, você precisará da ACCESS_KEY e da SECRET_KEY de um usuário do Gateway de Objetos com o flag `system` habilitado. O usuário costuma ser o `admin`. Para obter a ACCESS_KEY e a SECRET_KEY, execute **`radosgw-admin user info --uid admin --rgw-zone=`***`NOME_DA_ZONA`*.

```
cephuser@adm > radosgw-admin zone modify \
 --rgw-realm=REALM_NAME \
 --rgw-zonegroup=ZONEGROUP_NAME \
```

```
--rgw-zone=ZONE_NAME \
--endpoints http://RGW.EXAMPLE.COM:80 \
--access-key=ACCESS_KEY \
--secret=SECRET_KEY \
--master --default
```

5. Confirme a configuração atualizada:

```
cephuser@adm > radosgw-admin period update --commit
```

Para colocar o serviço Gateway de Objetos em container, crie o respectivo arquivo de especificação conforme descrito na *Seção 8.3.4, "Implantando Gateways de Objetos"* e aplique-o.

```
cephuser@adm > ceph orch apply -i RGW.yml
```

10.7.2 Fazendo upgrade do NFS Ganesha

Importante

O NFS Ganesha suporta o NFS versão 4.1 e mais recente. Ele não suporta o NFS versão 3.

Veja a seguir como migrar um serviço NFS Ganesha existente que executa o Ceph Nautilus para um container do NFS Ganesha que executa o Ceph Octopus.

Atenção

A documentação a seguir requer que você já tenha feito upgrade dos serviços básicos do Ceph.

O NFS Ganesha armazena a configuração adicional por daemon e a exporta para um pool RADOS. O pool RADOS configurado pode ser encontrado na linha `watch_url` do bloco `RADOS_URLS` no arquivo `ganesha.conf`. Por padrão, esse pool será denominado `ganesha_config`.

Antes de tentar qualquer migração, é altamente recomendável fazer uma cópia dos objetos de configuração de exportação e daemon localizados no pool RADOS. Para localizar o pool RADOS configurado, execute o seguinte comando:

```
cephuser@adm > grep -A5 RADOS_URLS /etc/ganesha/ganesha.conf
```

Para listar o conteúdo do pool RADOS:

```
cephuser@adm > rados --pool ganesha_config --namespace ganesha ls | sort
  conf-node3
  export-1
  export-2
  export-3
  export-4
```

Para copiar os objetos RADOS:

```
cephuser@adm > RADOS_ARGS="--pool ganesha_config --namespace ganesha"
cephuser@adm > OBJS=$(rados $RADOS_ARGS ls)
cephuser@adm > for obj in $OBJS; do rados $RADOS_ARGS get $obj $obj; done
cephuser@adm > ls -lah
total 40K
drwxr-xr-x 2 root root 4.0K Sep 8 03:30 .
drwx------ 9 root root 4.0K Sep 8 03:23 ..
-rw-r--r-- 1 root root 90 Sep 8 03:30 conf-node2
-rw-r--r-- 1 root root 90 Sep 8 03:30 conf-node3
-rw-r--r-- 1 root root 350 Sep 8 03:30 export-1
-rw-r--r-- 1 root root 350 Sep 8 03:30 export-2
-rw-r--r-- 1 root root 350 Sep 8 03:30 export-3
-rw-r--r-- 1 root root 358 Sep 8 03:30 export-4
```

Para cada nó, qualquer serviço existente do NFS Ganesha precisa ser interrompido e, em seguida, substituído por um container gerenciado pelo cephadm.

1. Pare e desabilite o serviço NFS Ganesha existente:

   ```
   cephuser@adm > systemctl stop nfs-ganesha
   cephuser@adm > systemctl disable nfs-ganesha
   ```

2. Depois que o serviço NFS Ganesha existente for interrompido, um novo serviço poderá ser implantado em um container usando o cephadm. Para fazer isso, você precisa criar uma especificação de serviço que contenha um `service_id`, que será usado para identificar esse novo cluster do NFS, o nome de host do nó que estamos migrando listado como host na especificação de posicionamento e o pool RADOS e o namespace que contêm os objetos de exportação do NFS configurados. Por exemplo:

   ```
   service_type: nfs
   service_id: SERVICE_ID
   placement:
   ```

```
  hosts:
  - node2
  pool: ganesha_config
  namespace: ganesha
```

Para obter mais informações sobre como criar uma especificação de posicionamento, consulte a *Seção 8.2, "Especificação de serviço e posicionamento"*.

3. Aplique a especificação de posicionamento:

```
cephuser@adm > ceph orch apply -i FILENAME.yaml
```

4. Confirme se o daemon do NFS Ganesha está em execução no host:

```
cephuser@adm > ceph orch ps --daemon_type nfs
NAME           HOST   STATUS         REFRESHED  AGE  VERSION  IMAGE NAME
                   IMAGE ID      CONTAINER ID
nfs.foo.node2  node2  running (26m)  8m ago     27m  3.3      registry.suse.com/
ses/7.1/ceph/ceph:latest  8b4be7c42abd  c8b75d7c8f0d
```

5. Repita essas etapas para cada nó do NFS Ganesha. Você não precisa criar uma especificação de serviço separada para cada nó. É suficiente adicionar o nome de host de cada nó à especificação do serviço NFS existente e reaplicá-la.

É possível migrar as exportações existentes de duas maneiras diferentes:

- Recriação ou reatribuição manual usando o Ceph Dashboard.
- Cópia manual do conteúdo de cada objeto RADOS por daemon para a configuração comum recém-criada do NFS Ganesha.

PROCEDIMENTO 10.1: COPIANDO MANUALMENTE AS EXPORTAÇÕES PARA O ARQUIVO DE CONFIGURAÇÃO COMUM DO NFS GANESHA

1. Determine a lista de objetos RADOS por daemon:

```
cephuser@adm > RADOS_ARGS="--pool ganesha_config --namespace ganesha"
cephuser@adm > DAEMON_OBJS=$(rados $RADOS_ARGS ls | grep 'conf-')
```

2. Faça uma cópia dos objetos RADOS por daemon:

```
cephuser@adm > for obj in $DAEMON_OBJS; do rados $RADOS_ARGS get $obj $obj; done
cephuser@adm > ls -lah
total 20K
drwxr-xr-x 2 root root 4.0K Sep 8 16:51 .
drwxr-xr-x 3 root root 4.0K Sep 8 16:47 ..
```

```
-rw-r--r-- 1 root root 90 Sep 8 16:51 conf-nfs.SERVICE_ID
-rw-r--r-- 1 root root 90 Sep 8 16:51 conf-node2
-rw-r--r-- 1 root root 90 Sep 8 16:51 conf-node3
```

3. Classifique e faça a fusão em uma única lista de exportações:

```
cephuser@adm > cat conf-* | sort -u > conf-nfs.SERVICE_ID
cephuser@adm > cat conf-nfs.foo
%url "rados://ganesha_config/ganesha/export-1"
%url "rados://ganesha_config/ganesha/export-2"
%url "rados://ganesha_config/ganesha/export-3"
%url "rados://ganesha_config/ganesha/export-4"
```

4. Grave o novo arquivo de configuração comum do NFS Ganesha:

```
cephuser@adm > rados $RADOS_ARGS put conf-nfs.SERVICE_ID conf-nfs.SERVICE_ID
```

5. Notifique o daemon do NFS Ganesha:

```
cephuser@adm > rados $RADOS_ARGS notify conf-nfs.SERVICE_ID conf-nfs.SERVICE_ID
```

Nota

Essa ação fará com que o daemon recarregue a configuração.

Após a migração bem-sucedida do serviço, será possível remover o serviço NFS Ganesha baseado no Nautilus.

1. Remova o NFS Ganesha:

```
cephuser@adm > zypper rm nfs-ganesha
Reading installed packages...
Resolving package dependencies...
The following 5 packages are going to be REMOVED:
  nfs-ganesha nfs-ganesha-ceph nfs-ganesha-rados-grace nfs-ganesha-rados-urls nfs-
ganesha-rgw
5 packages to remove.
After the operation, 308.9 KiB will be freed.
Continue? [y/n/v/...? shows all options] (y): y
(1/5) Removing nfs-ganesha-
ceph-2.8.3+git0.d504d374e-3.3.1.x86_64 ..........................................................
[done]
(2/5) Removing nfs-ganesha-
rgw-2.8.3+git0.d504d374e-3.3.1.x86_64 ...........................................................
[done]
```

```
(3/5) Removing nfs-ganesha-rados-
urls-2.8.3+git0.d504d374e-3.3.1.x86_64 ..........................................
[done]
(4/5) Removing nfs-ganesha-rados-
grace-2.8.3+git0.d504d374e-3.3.1.x86_64 ..........................................
[done]
(5/5) Removing nfs-
ganesha-2.8.3+git0.d504d374e-3.3.1.x86_64 ..........................................
[done]
Additional rpm output:
warning: /etc/ganesha/ganesha.conf saved as /etc/ganesha/ganesha.conf.rpmsave
```

2. Remova as configurações do cluster legado do Ceph Dashboard:

```
cephuser@adm > ceph dashboard reset-ganesha-clusters-rados-pool-namespace
```

10.7.3 Fazendo upgrade do servidor de metadados

Ao contrário dos MONs, MGRs e OSDs, o Servidor de Metadados não pode ser adotado no local. Em vez disso, você precisa reimplantá-lo em containers usando o orquestrador do Ceph.

1. Execute o comando **`ceph fs ls`** para obter o nome do seu sistema de arquivos, por exemplo:

```
cephuser@adm > ceph fs ls
name: cephfs, metadata pool: cephfs_metadata, data pools: [cephfs_data ]
```

2. Crie um novo arquivo de especificação de serviço `mds.yml`, conforme descrito na *Seção 8.3.3, "Implantando servidores de metadados"*, usando o nome do sistema de arquivos como o `service_id` e especificando os hosts que executarão os daemons MDS. Por exemplo:

```
service_type: mds
service_id: cephfs
placement:
  hosts:
  - ses-min1
  - ses-min2
  - ses-min3
```

3. Execute o comando **`ceph orch apply -i mds.yml`** para aplicar a especificação de serviço e iniciar os daemons MDS.

10.7.4 Fazendo upgrade do iSCSI Gateway

Para fazer upgrade do iSCSI Gateway, é necessário reimplantá-lo em containers usando o orquestrador do Ceph. Se você tiver vários iSCSI Gateways, precisará reimplantá-los um por um para reduzir o tempo de espera do serviço.

1. Pare e desabilite os daemons iSCSI existentes em cada nó do iSCSI Gateway:

```
> sudo systemctl stop rbd-target-gw
> sudo systemctl disable rbd-target-gw
> sudo systemctl stop rbd-target-api
> sudo systemctl disable rbd-target-api
```

2. Crie uma especificação de serviço para o iSCSI Gateway conforme descrito na *Seção 8.3.5, "Implantando Gateways iSCSI"*. Para isso, você precisa das configurações `pool`, `trusted_ip_list` e `api_*` do arquivo `/etc/ceph/iscsi-gateway.cfg` existente. Se você tem o suporte a SSL habilitado (`api_secure = true`), precisa também do certificado SSL (`/etc/ceph/iscsi-gateway.crt`) e da chave (`/etc/ceph/iscsi-gateway.key`).
   Por exemplo, se `/etc/ceph/iscsi-gateway.cfg` contém o seguinte:

```
[config]
cluster_client_name = client.igw.ses-min5
pool = iscsi-images
trusted_ip_list = 10.20.179.203,10.20.179.201,10.20.179.205,10.20.179.202
api_port = 5000
api_user = admin
api_password = admin
api_secure = true
```

   Você precisa criar o seguinte arquivo de especificação de serviço `iscsi.yml`:

```
service_type: iscsi
service_id: igw
placement:
  hosts:
  - ses-min5
spec:
  pool: iscsi-images
  trusted_ip_list: "10.20.179.203,10.20.179.201,10.20.179.205,10.20.179.202"
  api_port: 5000
  api_user: admin
  api_password: admin
  api_secure: true
  ssl_cert: |
    -----BEGIN CERTIFICATE-----
```

```
    MIIDtTCCAp2gAwIBAgIYMC4xNzc1NDQxNjEzMzc2MjMyXzxvQ7EcMA0GCSqGSIb3
    DQEBCwUAMG0xCzAJBgNVBAYTAlVTMQ0wCwYDVQQIDARVdGFoMRcwFQYDVQQHDA5T
    [...]
    -----END CERTIFICATE-----
  ssl_key: |
    -----BEGIN PRIVATE KEY-----
    MIIEvQIBADANBgkqhkiG9w0BAQEFAASCBKcwggSjAgEAAoIBAQC5jdYbjtNTAKW4
    /CwQr/7wOiLGzVxChn3mmCIF3DwbL/qvTFTX2d8bDf6LjGwLYloXHscRfxszX/4h
    [...]
    -----END PRIVATE KEY-----
```

Nota

As configurações `pool`, `trusted_ip_list`, `api_port`, `api_user`, `api_password` e `api_secure` são idênticas às do arquivo `/etc/ceph/iscsi-gateway.cfg`. É possível copiar os valores `ssl_cert` e `ssl_key` dos arquivos de chave e de certificado SSL existentes. Verifique se eles estão indentados corretamente e se o caractere *barra vertical* `|` aparece no final das linhas `ssl_cert:` e `ssl_key:` (veja o conteúdo do arquivo `iscsi.yml` acima).

3. Execute o comando **`ceph orch apply -i iscsi.yml`** para aplicar a especificação de serviço e iniciar os daemons do iSCSI Gateway.

4. Remova o pacote `ceph-iscsi` antigo de cada um dos nós de gateway iSCSI existentes:

```
cephuser@adm > zypper rm -u ceph-iscsi
```

10.8 Limpeza após o upgrade

Após o upgrade, execute as seguintes etapas de limpeza:

1. Verifique a versão atual do Ceph para conferir se o upgrade do cluster foi bem-sucedido:

```
cephuser@adm > ceph versions
```

2. Garanta que nenhum OSD antigo ingresse no cluster:

```
cephuser@adm > ceph osd require-osd-release pacific
```

3. Defina o `pg_autoscale_mode` dos pools existentes, se necessário:

> **Importante**
>
> Por padrão, os pools no SUSE Enterprise Storage 6 tinham o `pg_autoscale_mode` definido como `warn`. Isso resultava em uma mensagem de aviso em caso de número de PGs abaixo do ideal, mas o dimensionamento automático não era feito. O padrão no SUSE Enterprise Storage 7.1 é a opção `pg_autoscale_mode` definida como `on` para que os novos pools e os PGs sejam dimensionados automaticamente. O processo de upgrade não muda automaticamente o `pg_autoscale_mode` dos pools existentes. Para mudá-lo para `on` e aproveitar todos os benefícios do dimensionador automático, consulte as instruções no *Livro "Guia de Administração e Operações", Capítulo 17 "Gerenciamento de dados armazenados", Seção 17.4.12 "Habilitando o dimensionador automático de PG"*.

Encontre mais detalhes no *Livro "Guia de Administração e Operações", Capítulo 17 "Gerenciamento de dados armazenados", Seção 17.4.12 "Habilitando o dimensionador automático de PG"*.

4. Impedir clientes pré-Luminous:

```
cephuser@adm > ceph osd set-require-min-compat-client luminous
```

5. Habilite o módulo de balanceador:

```
cephuser@adm > ceph balancer mode upmap
cephuser@adm > ceph balancer on
```

Encontre mais detalhes no *Livro "Guia de Administração e Operações", Capítulo 29 "Módulos do Ceph Manager", Seção 29.1 "Balanceador"*.

6. Se preferir, habilite o módulo de telemetria:

```
cephuser@adm > ceph mgr module enable telemetry
cephuser@adm > ceph telemetry on
```

Encontre mais detalhes no *Livro "Guia de Administração e Operações", Capítulo 29 "Módulos do Ceph Manager", Seção 29.2 "Habilitando o módulo de telemetria"*.

11 Upgrade do SUSE Enterprise Storage 7 para 7.1

Este capítulo apresenta as etapas para fazer upgrade do SUSE Enterprise Storage 7 para a versão 7.1.

O upgrade inclui as seguintes tarefas:

- Fazer upgrade do SUSE Linux Enterprise Server subjacente e do SUSE Linux Enterprise Server 15 SP2 para a versão SUSE Linux Enterprise Server 15 SP3.
- Fazer upgrade do Ceph Octopus para o Pacific.

11.1 Antes de fazer upgrade

As tarefas a seguir *devem* ser concluídas antes de você iniciar o upgrade. Isso pode ser feito a qualquer momento durante a vida útil do SUSE Enterprise Storage 7.

11.1.1 Pontos a serem considerados

Antes de fazer upgrade, leia as seções a seguir para garantir que você entenda todas as tarefas que precisam ser executadas.

- *Ler os detalhes da versão*. Neles, você encontra mais informações sobre o que mudou desde a versão anterior do SUSE Enterprise Storage. Consulte os detalhes da versão para verificar se:
  - Seu hardware precisa de considerações especiais.
  - Qualquer pacote de software usado foi significativamente modificado.
  - São necessárias precauções especiais para a instalação.

  Os detalhes da versão também apresentam informações que não puderam ser incluídas a tempo no manual. Eles também incluem notas sobre problemas conhecidos.
  Você encontra os detalhes da versão do SES 7.1 online em https://www.suse.com/releasenotes/.

Além disso, após a instalação do pacote `release-notes-ses` do repositório SES 7.1, você encontrará os detalhes da versão no diretório local `/usr/share/doc/release-notes` ou online em https://www.suse.com/releasenotes/.

- Leia a *Parte II, "Implantando o cluster do Ceph"* para se familiarizar com o `ceph-salt` e o orquestrador do Ceph, especialmente as informações sobre as especificações do serviço.

11.1.2 Fazendo backup da configuração e dos dados do cluster

É altamente recomendável fazer backup de todas as configurações e os dados do cluster antes de iniciar o upgrade. Para obter instruções sobre como fazer backup de todos os seus dados, consulte *Livro "Guia de Administração e Operações", Capítulo 15 "Backup e restauração"*.

11.1.3 Verificando o acesso a repositórios do software e imagens de container

Verifique se cada nó do cluster tem acesso aos repositórios do software SUSE Linux Enterprise Server 15 SP3 e SUSE Enterprise Storage 7.1 e também ao registro de imagens de container.

11.1.3.1 Repositórios do software

Se todos os nós foram registrados no SCC, você poderá usar o comando **`zypper migration`** para fazer upgrade. Consulte https://documentation.suse.com/sles/15-SP3/html/SLES-all/cha-upgrade-online.html#sec-upgrade-online-zypper para obter mais detalhes.

Se os nós **não** foram registrados no SCC, desabilite todos os repositórios existentes do software e adicione os dois repositórios `Pool` e `Updates` para cada uma das seguintes extensões:

- SLE-Product-SLES/15-SP3
- SLE-Module-Basesystem/15-SP3
- SLE-Module-Server-Applications/15-SP3
- SUSE-Enterprise-Storage-7.1

11.1.3.2 Imagens de container

Todos os nós do cluster precisam acessar o registro de imagens de container. Na maioria dos casos, você usará o registro público do SUSE em `registry.suse.com`. Você precisa das seguintes imagens:

- registry.suse.com/ses/7.1/ceph/ceph
- registry.suse.com/ses/7.1/ceph/grafana
- registry.suse.com/ses/7.1/ceph/prometheus-server
- registry.suse.com/ses/7.1/ceph/prometheus-node-exporter
- registry.suse.com/ses/7.1/ceph/prometheus-alertmanager

Por exemplo, para implantações isoladas (air-gapped), uma alternativa é configurar um registro local e verificar se você tem o conjunto correto de imagens de container disponível. Consulte a *Seção 7.2.10, "Usando o registro do container"* para obter mais detalhes sobre como configurar um registro de imagens de container local.

11.2 Migrar o SUSE Linux Enterprise Server em cada nó do cluster para a versão SUSE Linux Enterprise Server 15 SP3

Se os nós do cluster foram configurados para usar o SUSE Customer Center, você pode usar o comando **`zypper migration`**.

Se os repositórios de software dos nós do cluster foram configurados manualmente, você precisa fazer upgrade manual dos nós.

Para obter informações detalhadas sobre o upgrade do SUSE Linux Enterprise Server usando o **`zypper`**, consulte https://documentation.suse.com/sles/15-SP3/html/SLES-all/cha-upgrade-online.html#sec-upgrade-online-zypper.

11.3 Atualizar os pacotes relacionados ao SUSE Enterprise Storage em cada nó do cluster

Para atualizar os pacotes do SUSE Enterprise Storage para a versão mais recente, use o comando **`ceph-salt update`**. Para obter mais detalhes, visite o *Livro "Guia de Administração e Operações", Capítulo 13 "Tarefas operacionais", Seção 13.6 "Atualizando os nós do cluster"*.

11.4 Fazer upgrade dos serviços de cluster existentes do Ceph

Faça upgrade de todo o cluster do Ceph para a versão Pacific executando o seguinte comando do Nó de Admin:

```
cephuser@adm > ceph orch upgrade start --image registry.suse.com/ses/7.1/ceph/ceph
```

A Atualizações de manutenção do Ceph baseadas nos point releases de upstream do “Pacific”

Vários pacotes importantes no SUSE Enterprise Storage 7.1 são baseados na série de lançamentos Pacific do Ceph. Quando o projeto do Ceph (https://github.com/ceph/ceph) publica novos point releases na série Pacific, o SUSE Enterprise Storage 7.1 é atualizado para garantir que o produto aproveite as correções de bug de upstream e os backports de recursos mais recentes.

Este capítulo contém resumos das mudanças importantes contidas em cada point release de upstream que foi, ou está planejado para ser incluído no produto.

Glossário

Geral

Alertmanager

Um binário único que processa os alertas enviados pelo servidor Prometheus e notifica o usuário final.

Armazenamento de Objetos do Ceph

O “produto”, o serviço ou os recursos de armazenamento de objetos, que consistem em um Cluster de Armazenamento do Ceph e um Gateway de Objetos do Ceph.

Árvore de roteamento

Um termo que representa qualquer diagrama que mostra as várias rotas que um receptor pode executar.

Ceph Dashboard

Um aplicativo incorporado de monitoramento e gerenciamento do Ceph baseado na Web que administra vários aspectos e objetos do cluster. O painel de controle é implementado como um módulo do Ceph Manager.

Ceph Manager

O Ceph Manager ou MGR é o software de gerenciador do Ceph, que coleta o estado completo de todo o cluster em um único local.

Ceph Monitor

O Ceph Monitor ou MON é o software de monitoração do Ceph.

`ceph-salt`

Inclui ferramentas para implantação de clusters do Ceph gerenciados pelo cephadm por meio do Salt.

cephadm

O cephadm implanta e gerencia um cluster do Ceph conectando-se aos hosts do daemon do gerenciador por SSH para adicionar, remover ou atualizar os containers de daemons do Ceph.

CephFS

O sistema de arquivos do Ceph.

CephX

O protocolo de autenticação do Ceph. O CephX opera como o Kerberos, mas não tem um ponto único de falha.

Cliente do Ceph

A coleção de componentes do Ceph que podem acessar um Cluster de Armazenamento do Ceph. Eles incluem o Gateway de Objetos, o Dispositivo de Blocos do Ceph, o CephFS e as bibliotecas, os módulos de kernel e os clientes FUSE correspondentes.

Cluster de Armazenamento do Ceph

O conjunto principal do software de armazenamento que armazena os dados do usuário. Esse conjunto consiste em Ceph Monitors e OSDs.

Compartimento de memória

Um ponto que agrega outros nós em uma hierarquia de locais físicos.

Conjunto de Regras

Regras para determinar o posicionamento de dados em um pool.

CRUSH, Mapa CRUSH

Controlled Replication Under Scalable Hashing: Um algoritmo que determina como armazenar e recuperar dados calculando os locais de armazenamento de dados. O CRUSH requer um mapa de cluster para armazenar e recuperar dados de forma pseudo-aleatória nos OSDs com uma distribuição uniforme dos dados pelo cluster.

Daemon Ceph OSD

O daemon **`ceph-osd`** é o componente do Ceph responsável por armazenar objetos em um sistema de arquivos local e por conceder acesso a eles pela rede.

Dispositivo de Blocos RADOS (RBD)

O componente de armazenamento de blocos do Ceph. Também conhecido como dispositivo de blocos do Ceph.

DriveGroups

DriveGroups são uma declaração de um ou mais layouts de OSD que podem ser mapeados para unidades físicas. Um layout de OSD define como o Ceph aloca fisicamente o armazenamento OSD na mídia correspondente aos critérios especificados.

Gateway de Objetos

O componente de gateway do S3/Swift do Armazenamento de Objetos do Ceph. Também conhecido como RADOS Gateway (RGW).

Gateway do Samba

O Gateway do Samba ingressa no Active Directory no domínio do Windows para autenticar e autorizar usuários.

Grafana

Análise de banco de dados e solução de monitoramento.

Módulo de sincronização de arquivos

Módulo que permite a criação de uma zona do Gateway de Objetos para manter o histórico das versões de objeto do S3.

Nó

Qualquer máquina ou servidor único em um cluster do Ceph.

Nó de admin

O host do qual você executa os comandos relacionados ao Ceph para administrar os hosts do cluster.

Nó OSD

Um nó do cluster que armazena dados, processa a replicação, a recuperação, o preenchimento e a redistribuição de dados e fornece algumas informações de monitoramento aos Ceph Monitors examinando outros daemons Ceph OSD.

OSD

Dispositivo de Armazenamento de Objetos: Uma unidade de armazenamento física ou lógica.

PG

Grupo de Posicionamento: uma subdivisão de um *pool*, usado para ajuste de desempenho.

Point Release

Qualquer versão Ad Hoc que inclua apenas correções de bug ou segurança.

Pool

Partições lógicas para armazenamento de objetos, como imagens de disco.

Prometheus

Kit de ferramentas de monitoramento e alertas de sistemas.

Regra CRUSH

A regra de posicionamento de dados CRUSH que se aplica a um ou mais pools específicos.

Reliable Autonomic Distributed Object Store (RADOS)

O conjunto principal do software de armazenamento que armazena os dados do usuário (MON + OSD).

Samba

Software de integração do Windows.

Servidor de Metadados

O Servidor de Metadados ou MDS é o software de metadados do Ceph.

Várias zonas

zonegroup