



企业级分布式存储 解决方案

林培峰

pflin@suse.com

2019.11

SUSE"专家日"2019

中国 成都

目录

- 企业存储的业务挑战
- SUSE 企业级软件定义存储方案
- 案例分享



企业存储的业务挑战

我是谁？

175 ZB
到2025年

ZB = 1024 EB = 1024*1024 PB=1024*1024*1024 TB



我从哪里来？



要到哪里去



企业级分布式存储

SUSE Enterprise Storage 6

SUSE Enterprise Storage 6

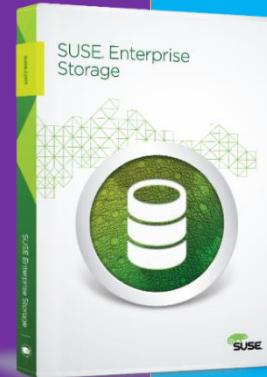
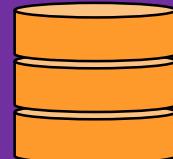
应云而生兼顾传统需求

传统基础架构

运维：GUI

数据服务：NFS, CIFS, iSCSI

应用：备份，归档，虚拟化



软件定义的架构

运维：脚本、API

数据服务：S3, API

架构：分布式

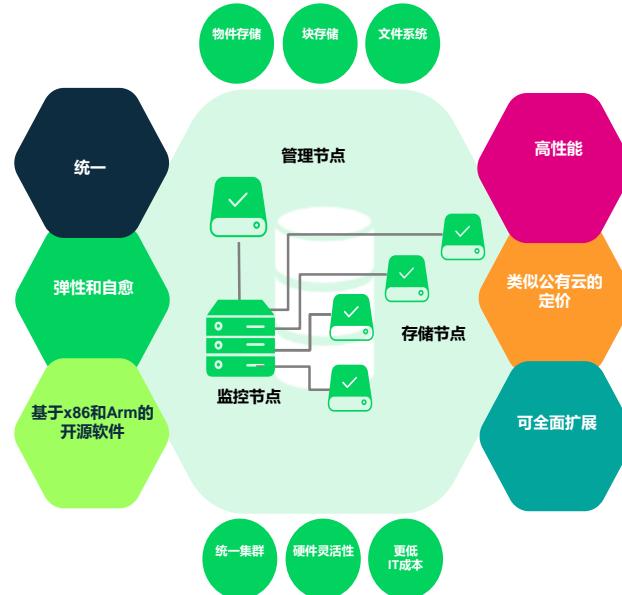
应用：容器，私有云，大数据



SUSE Enterprise Storage

过去12个月的成就

- 发布SUSE Enterprise Storage 5.5
 - 第8版本
- 最新上游版本Ceph Nautilus
 - 前20名Ceph贡献者有8个来自SUSE
- SUSE团队推动Ceph Dashboard上游项目
- 与2017财年相比，SUSE在2018财年的Ceph客户数增加了一倍以上



SUSE Enterprise Storage

核心开发领域



易管理性



互操作性



性能



可用性

- 易于安装
- 基于GUI的监控和管理

- 统一块、文件和物件
- 结构互连

- 缓存分层
- 容器化
- 分层存储管理

- 备份/归档
- 持续数据保护
- 远程复制

SUSE Enterprise Storage 6



易管理性



- 多用户/ 角色管理
- SSO 单点登陆(SAML 2)
- 审计
- 多语言支持
- REST API支持
- 性能监控
- 告警和SNMP

- 磁盘和存储池管理
- 集群配置在线修改
- 块镜像
- QoS
- iSCSI / NFS / CephFS 管理
- 拓扑图管理

SUSE Enterprise Storage 6



互操作性



- IPv6
- RGW Elasticsearch同步模块
- 节点至节点加密**
- 一个集群中有多个CephFS实例**
- NFS-Ganesha双活架构**
- SMB / CIFS for Windows

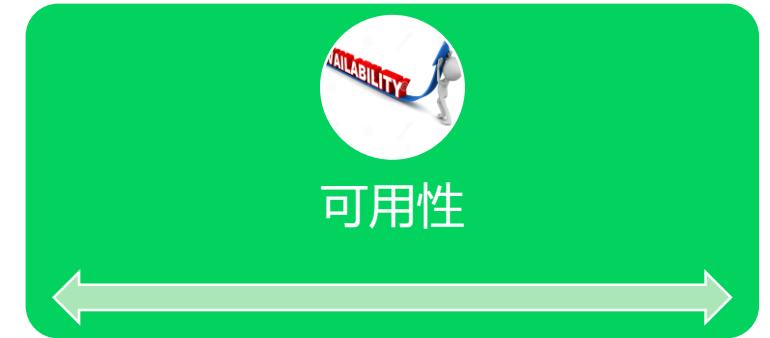
SUSE Enterprise Storage 6



- RBD 的QoS
- 后台运行的QoS
- S3 对象策略驱动的分层**
- Cache 缓存加速**

** 技术预览项

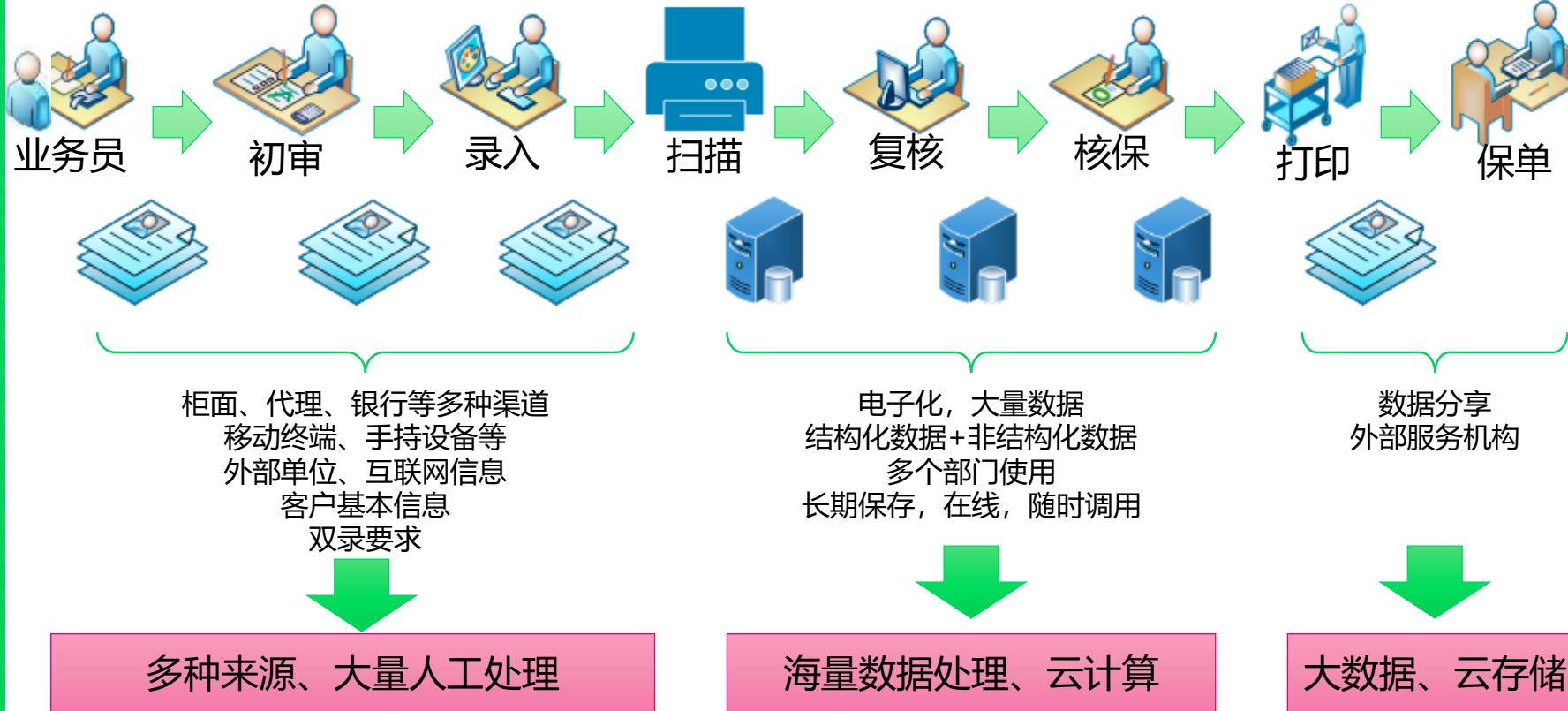
SUSE Enterprise Storage 6



- 通过S3单向同步到外部云
- CephFS快照
- 异步文件复制**

SUSE Enterprise Storage 案例分享

从承保流程分析IT需求



结构化数据备份

COMMVAULT

VERITAS

VEEAM

MICRO
FOCUS
Data Protector

SEP

bareos
open source data protection

备份服务器



消重前
数据



消重
引擎
插件

SUSE



消重后
数据

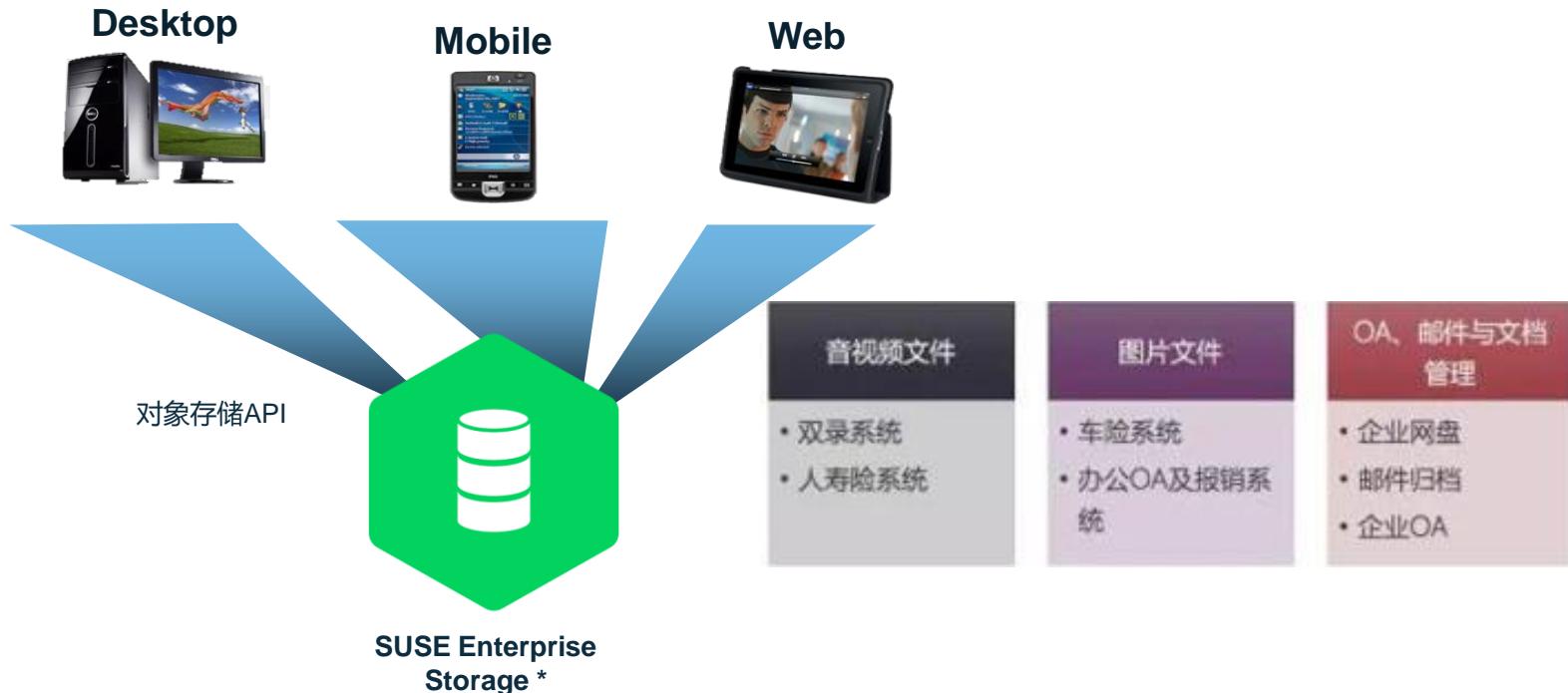


SUSE Enterprise
Storage *

- 主流备份软件兼容
- 不改变现有架构
- 数据在线，恢复速度快

- 无限的性能和容量扩展能力，
- 满足更高的备份窗口
- 减少存储容量，减少网络传输量

海量文件存储 - SUSE 对象存储



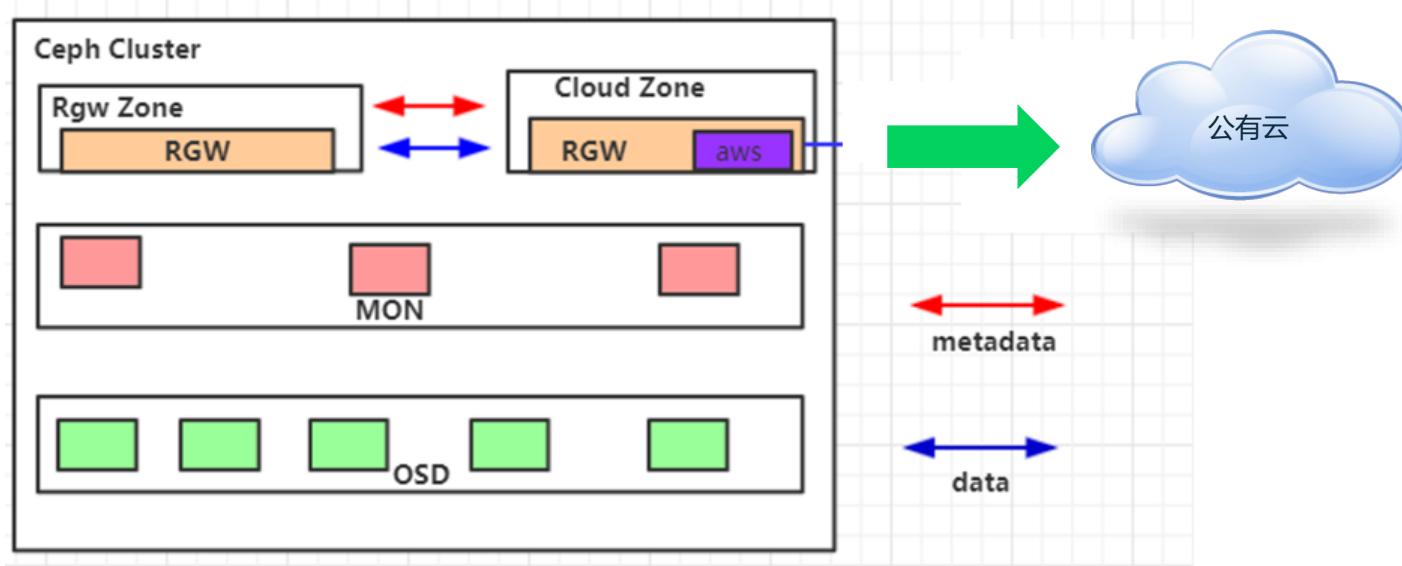
海量文件备份- 混合云备份

面临问题：

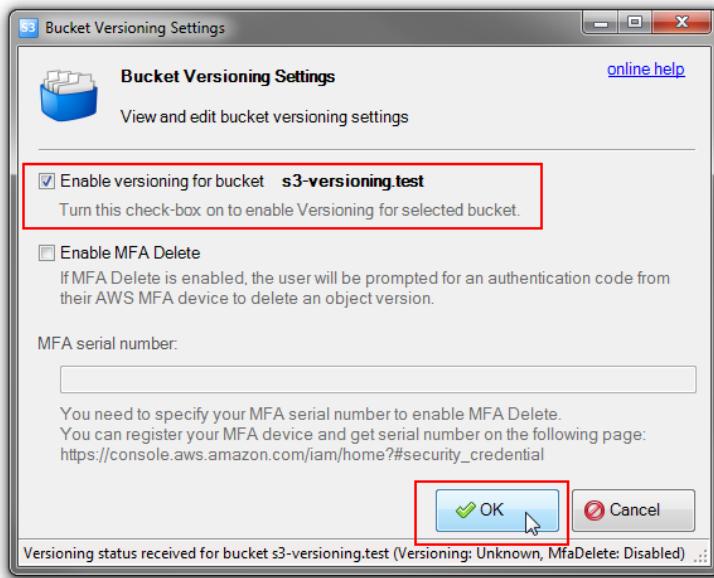
海量文件备份性能低，备份窗口不够
恢复时间长

解决之道：

多站点复制
简单易用且高性价比的在线备份



逻辑故障保护- 多版本功能



The screenshot shows the 'Versions' tab of the AWS S3 console for an object named 'pictures/DSC05100_1200x800.jpg'. It lists five versions, each with its revision number, last modified date, and ETag. The 'revision #. 2' row is highlighted with a red box. Below the table are standard S3 actions: Download, Restore, Delete, and Refresh.

| Key | Last Modified | ETag |
|--------------------------------|---------------------|---------------------------------|
| pictures/DSC05100_1200x800.jpg | | |
| revision #. 4 (current) | 8/5/2012 7:41:01 PM | "c45beb199d4a91591722f34514f2 |
| revision #. 3 | 8/5/2012 7:37:15 PM | "c45beb199d4a91591722f34514f2 |
| revision #. 2 | 8/5/2012 7:35:52 PM | "beeeec3cc1567af6f223ec537f1fa1 |
| revision #. 1 | 8/5/2012 7:33:22 PM | "e650808320602b04dc3bafba61 |

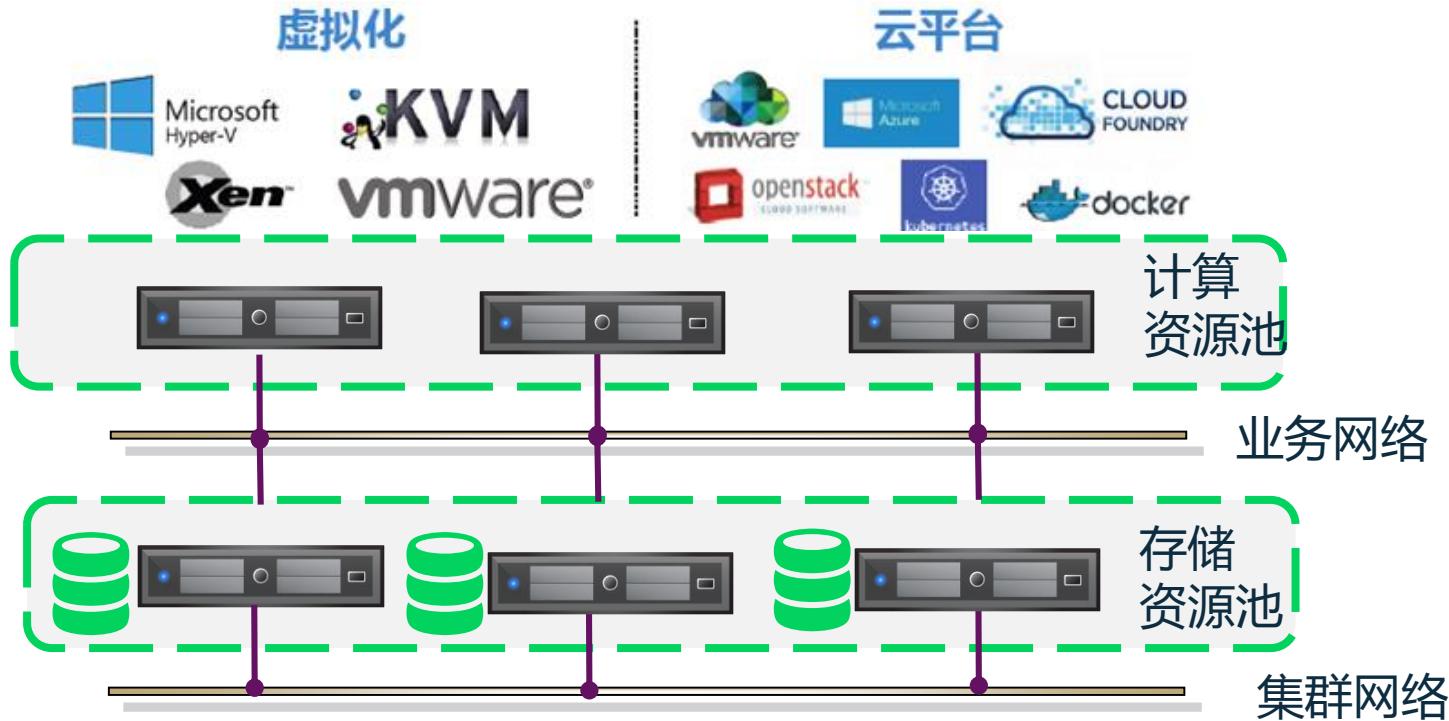
Download **Restore** Delete Refresh

The screenshot shows the same 'Versions' tab as above, but with a context menu open over the second version ('revision #. 3'). The menu options are: Download, Restore, Delete, and Refresh. The 'Delete' option is highlighted with a red box.

| Key | Last Modified | ETag |
|--------------------------------|---------------------|---------------------------------|
| pictures/DSC05100_1200x800.jpg | | |
| revision #. 4 (current) | 8/5/2012 7:41:01 PM | "c45beb199d4a91591722f34514f2 |
| revision #. 3 | 8/5/2012 7:37:15 PM | "c45beb199d4a91591722f34514f2 |
| revision #. 2 | 8/5/2012 7:35:52 PM | "beeeec3cc1567af6f223ec537f1fa1 |
| revision #. 1 | 8/5/2012 7:33:22 PM | "e650808320602b04dc3bafba61 |

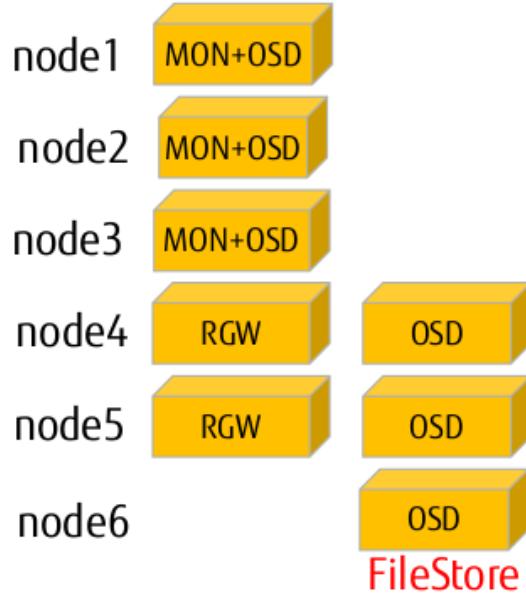
Download **Delete** Refresh

虚拟机和容器存储

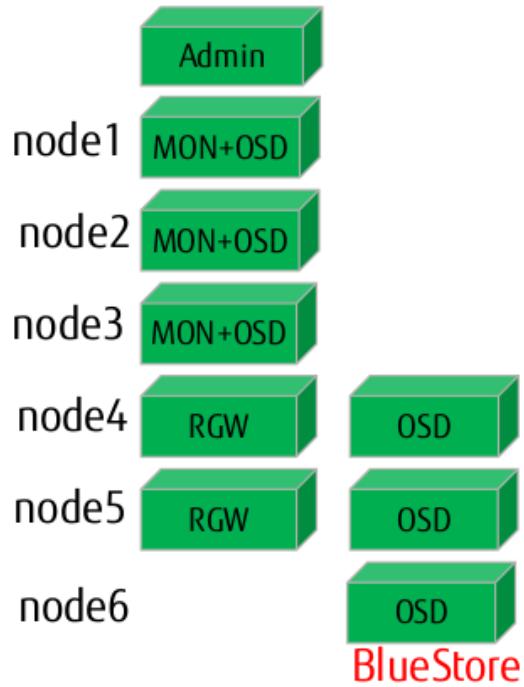


从Ubuntu Ceph 不中断迁移到SUSE Enterprise Storage

Starting point



Goal



OSD : Object Storage Device
MON : Monitor daemon
RGW : RADOS Gateway

Ubuntu 14.04 (Trusty)
Ceph Hammer

SLES12 SP3
SES 5.5 (Ceph Luminous)

of OSD : 6
of MON : 3
of RGW : 2

存储一体机解决方案



SuperPro Storage超融合整体解决方案

产品规格

| 产品系列 | SuperPro Storage超融合分布式存储一体机 | |
|---------|-----------------------------|--|
| OSD节点 | 分布式存储节点 (4U 四节点) | |
| | 应用场景 | 云、虚拟化、视频监控、HPC、大数据、人工智能、VDI、备份归档、文件共享 |
| 数据磁盘 | 最大容量支持 | 192TB (可无限容量扩充) |
| | 高速缓存 | 1.6TB (SAS 400G*4) |
| 数据网络 | 数据网络 | 16 x 10 GbE (每个节点4 x 10GbE) |
| | 处理器类型 | 新一代Intel® 至强® 可扩展处理器 |
| 硬件规格 | 可选网络接口 | 10Gb、25Gb、40Gb 网络接口 |
| | 电源参数 | 1000W ~ 1200W, 1+1冗余, 100-240 V 交流电 50~60Hz |
| | 存储设备节点数 | 无限节点扩充 |
| | 高级软件功能 | 自动精简配置、数据卷离散映射、智能磁盘预测、QoS、分级存储、远程复制、克隆、快照、配额、ACL、SNMP、多租户、断电数据保护、OpenStack支持 |
| 环境规格 | 管理方式 | WEB (网页)、GUI (图形)、CLI (命令) 管理模式 |
| | 环境温度 | 工作温度: 10°C ~ 40°C; 非工作温度: -40°C ~ 60°C |
| 可选服务 | 环境湿度 | 工作湿度: 8%~90%; 非工作湿度: 5%~95% |
| | 设计 | 数据中心设计、验证和引导服务 |
| 机柜整合式服务 | SuperPro机柜整合式服务 | |
| 现场支持 | SuperPro现场服务支持 | |
| 现场整合 | SuperPro现场整合服务 | |



移动集群分布式块存储

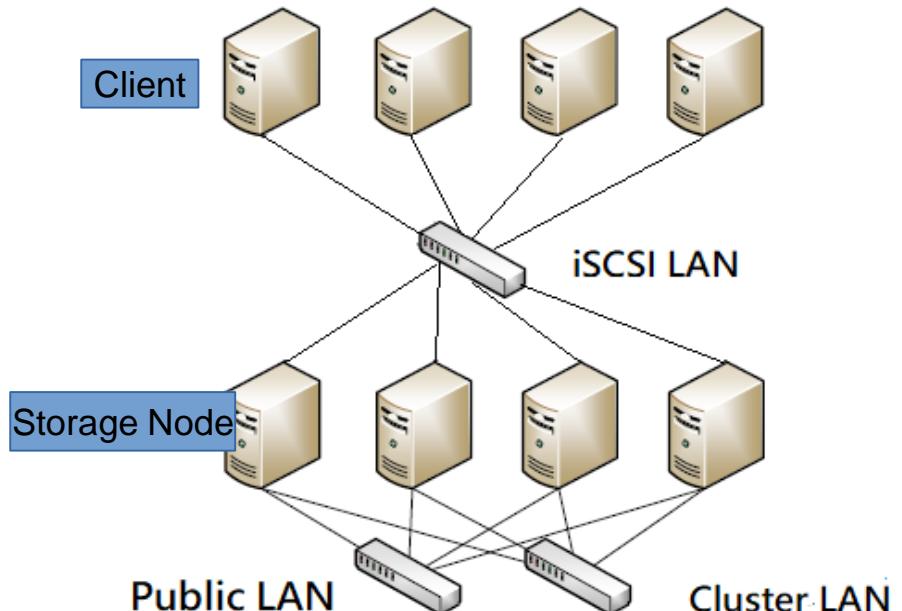
- 两种模型

→ 80TB (可用容量):

- 性能性

→ 200TB (可用容量):

- 通用性



性能要求

| | | Pre-biding | Current |
|-------|----------|--------------|---------------|
| 80TB | OLAP | IOPS >= 8500 | IOPS >= 21000 |
| | OLTP | IOPS >= 6000 | IOPS >= 15000 |
| 200TB | 1024k | >= 2450MB/s | >= 4900MB/s |
| | 1024k-8k | >= 2300MB/s | >= 4600MB/s |
| | 8k | IOPS >= 8000 | IOPS >= 16000 |

| OLTP | | |
|----------------|------------|------------|
| IO Profile | Block size | proportion |
| Random read | 8k | 20.00% |
| Random read | 4k | 45.00% |
| Random write | 8k | 15.00% |
| Sequence read | 64k | 10.00% |
| Sequence write | 64k | 10.00% |

| OLAP | | |
|----------------|------------|------------|
| IO Profile | Block size | proportion |
| Random read | 4k | 15.00% |
| Random write | 4k | 5.00% |
| Sequence read | 64k | 70.00% |
| Sequence write | 64k | 10.00% |

1024k

| IO Profile | Block size | proportion |
|----------------|------------|------------|
| Sequence read | 1024k | 70.00% |
| Sequence write | 1024k | 30.00% |

1024-8k

| IO Profile | Block size | proportion |
|----------------|------------|------------|
| Random read | 8k | 40.00% |
| Random write | 8k | 10.00% |
| Sequence read | 1024k | 40.00% |
| Sequence write | 1024k | 10.00% |

8k

| IO Profile | Block size | proportion |
|--------------|------------|------------|
| Random read | 8k | 70.00% |
| Random write | 8k | 30.00% |

80TB 模型 性能调优

| 80TB Model | 最初要求 | 标前测试要求 | 标后测试结果 |
|--------------|--|--|---|
| 0 SD Servers | 配置要求 (服务器数量: 4) | 配置要求 (服务器数量: 4) | 配置要求 (服务器数量: 4) |
| *CPU: | 2 x Intel V5 4114 CPU (10 cores) | 2 x Intel V5 4114 CPU (10 cores) | 2 x Intel V5 4114 CPU (10 cores) |
| *RAM: | 192G | 192G | 192G |
| *Nic: | public network: 2 X 10Gb Cluster network: 2 X 10Gb | public network: 2 X 10Gb Cluster network: 2 X 10Gb | public network: 2 X 10Gb Cluster network: 2 X 10Gb |
| *Disk: | OS: 2 x 480G SSD Data: 12 x 6TB 7.2K SATA HDD Cache: 3 x 3.84T NVME 1DW PD | OS: 2 x 480G SSD Data: 12 x 6TB 7.2K SATA HDD Cache: 3 x 3.84T NVME 1DW PD | OS: 2 x 480G SSD Data: 19 x 2.4TB 10K SAS HDD Cache: 2 x 3.2T NVME 3DW PD |
| Performance | OLAP : 4000 DPS OLTP : 3500 DPS | OLAP : 9800 DPS OLTP : 7400 DPS | OLAP : 23500 DPS OLTP : 19000 DPS |

200TB模型 性能调优

| 200TB Model | 最初要求 | 标前测试要求 | 标后测试结果 |
|-------------|--|--|--|
| OS Servers | 配置要求 (服务器数量: 4) | 配置要求 (服务器数量: 4) | 配置要求 (服务器数量: 5) |
| *CPU: | 2 x Intel V5 4114 CPU (10 cores) | 2 x Intel V5 4114 CPU (10 cores) | 2 x Intel V5 4114 CPU (10 cores) |
| *RAM: | 192G | 192G | 192G |
| *Nic: | public network: 2 X 10Gb Cluster network: 2 X 10Gb | public network: 2 X 10Gb Cluster network: 2 X 10Gb | public network: 2 X 10Gb Cluster network: 2 X 10Gb SCSI GW Network: 2 X 10Gb |
| *Disk: | OS: 2 x 480G SSD Data: 28 x 6TB 7.2K SATA HDD Cache: 6 x 960GB SSD | OS: 2 x 480G SSD Data: 28 x 6TB 7.2K SATA HDD Cache: 6 x 960GB SSD | OS: 2 x 480G SSD Data: 34 x 4TB 7.2K SATA HDD Cache: 3 x 3.2T NVME 3DWPD |
| Performance | 1024k: 2700 MB/s 1024k-8k: 1500 MS/s 8k: 7000 DPS | 1024k: 3200 MB/s 1024k-8k: 2500 MS/s 8k: 9500 DPS | 1024k: 5400 MB/s 1024k-8k: 5200 MS/s 8k: 20000 DPS |

用例解决方案

合作伙伴生态系统

传统工作负载

向磁盘备份解决方案



灵活配置

认证的参考架构



云原生工作负载

云与应用交付



容器即服务

档同步与共享



HPC存储



设备



定制构建

利用COTS服务器进行
DIY

分析



定制应用



SUSE Enterprise Storage





SUSE LLC未发表的作品。保留所有权利。

本文是未发表的作品，包含SUSE LLC的机密、专有和商业秘密信息。

本文的使用仅限于在工作范围内执行任务需要知道的SUSE员工。未经SUSE提前书面同意，本文任何部分不得实践、执行、复制、分发、修订、修改、翻译、删节、缩写、扩展、收集或者调整。

在未经授权的情况下对本文的使用或处理可能要追究刑事责任。

免责声明

本文不得解释为任何相关企业对于开发、交付或销售产品的承诺。本文不构成提供任何材料、代码或功能的承诺，也不构成采购决策的基础。对于本文的内容，SUSE不作任何陈述或保证，并且特别否认对适销性或者对特定目的的适用性的保证。SUSE产品的特性或功能的开发、发布和时间由SUSE单方决定。另外，SUSE保留随时修订本文并更改本文内容的权利，而无义务将此类修订或更改通知任何人或实体。本文提到的所有SUSE标志是Novell公司在美国和其他国家的商标或注册商标。所有第三方商标是各自所有者的财产。