OPENSTACK DAYS
CHINA

SurCloud

降低成本与提高性能

OpenStack存储如何能双赢

金友兵 博士
职务：书生云公司CTO

公司简介

书生云是书生集团旗下，主要从事云存储/云计算/云数据安全相关的技术研究开发及产品服务。

- 全球云技术领导厂商之一
- 在下一代分布式存储和云数据安全都有全球领先的核心技术
- 在中国、美国、日本、欧洲都有业务
- 美国《云计算》杂志“云存储卓越奖”，中国最具价值的安全存储服务解决方案
- Oracle全球合作伙伴，奇虎360战略合作伙伴

SurCloud

目录

SurCloud

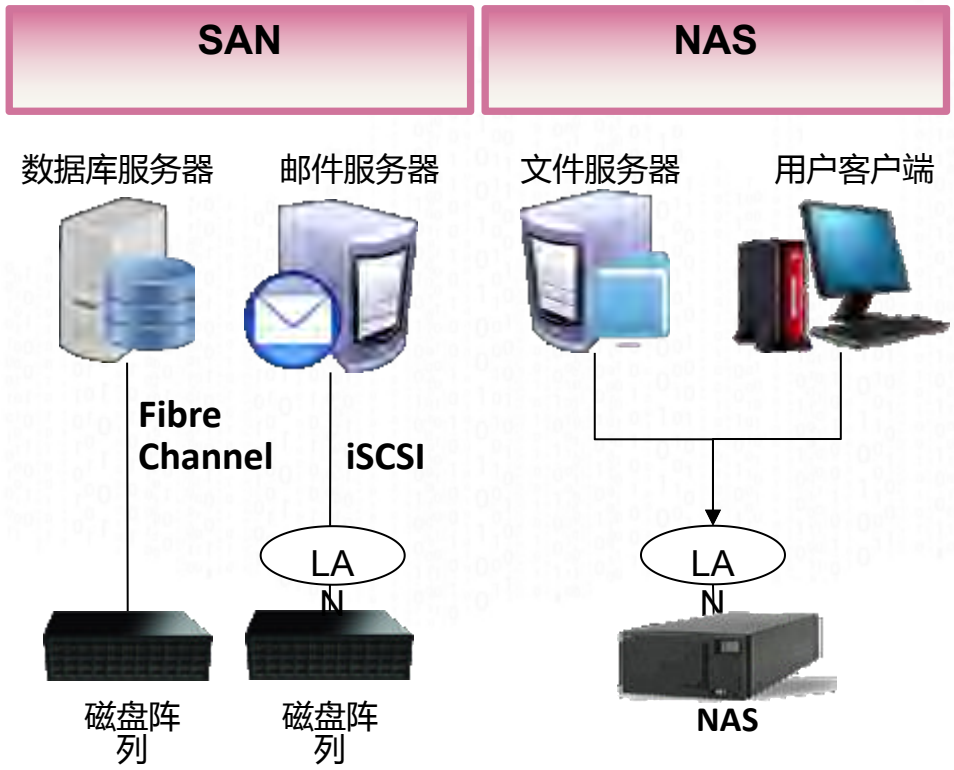
存储需求

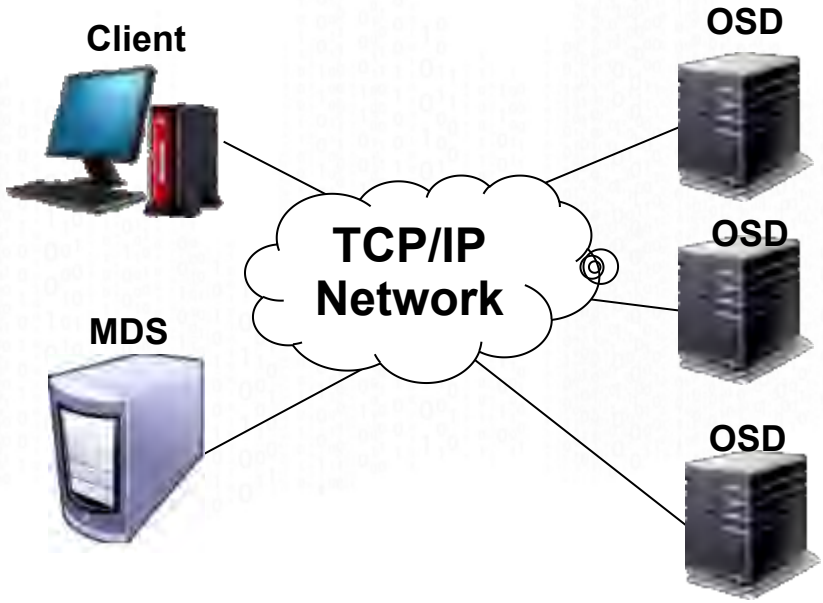
存储的作用

云计算和大数据的发展，对存储系统提出了更高的要求

SurCloud

传统存储逐步向分布式存储转变





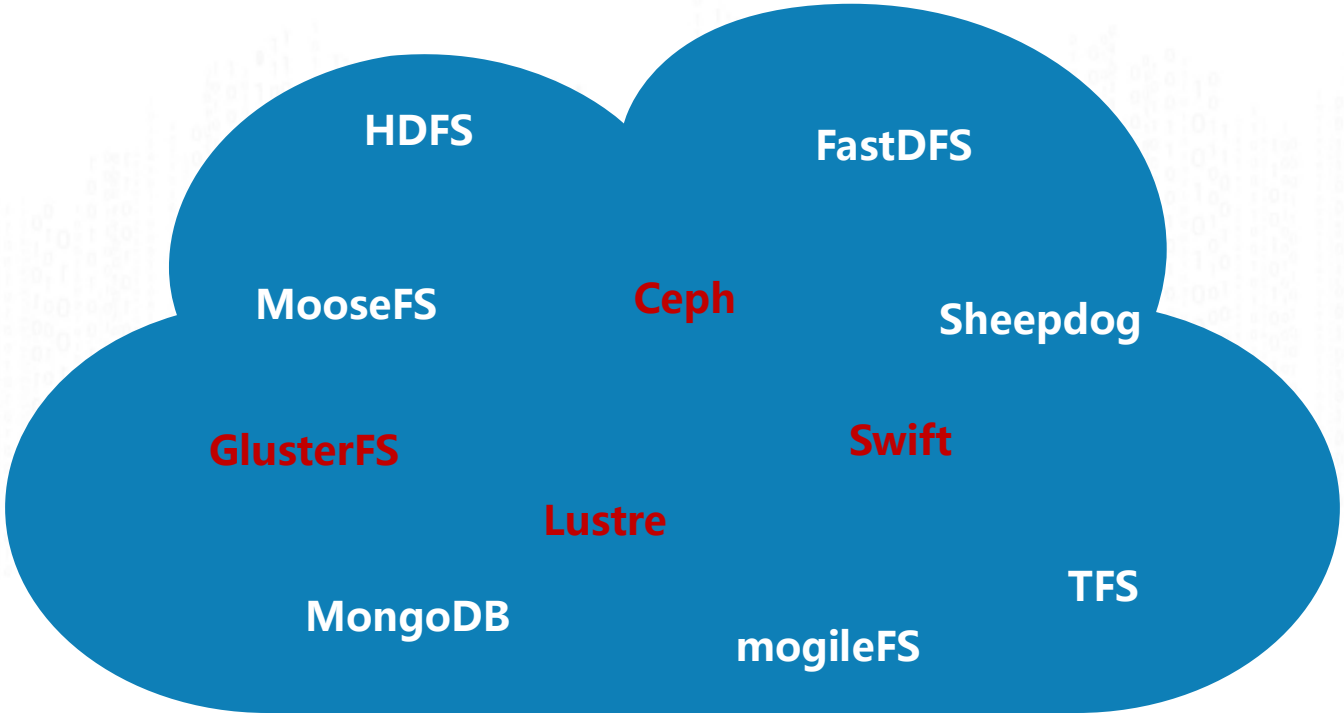
SurCloud

DAS NAS SAN	传统存储特点	分布式存储特点	OSD SSAN 集群NAS
数据路径短，单台服务器性能高		整体的高并发、高性能	
功能成熟，兼容性好		兼容云计算和虚拟化	
高可用HA复杂度高		无接管过程，HA复杂度较低	
开放性差		统一的访问空间，海量的存储	
整体性能低		扩展性、开放性好	
扩容复杂		成本一般	
成本高			

OPENSTACK DAYS CHINA

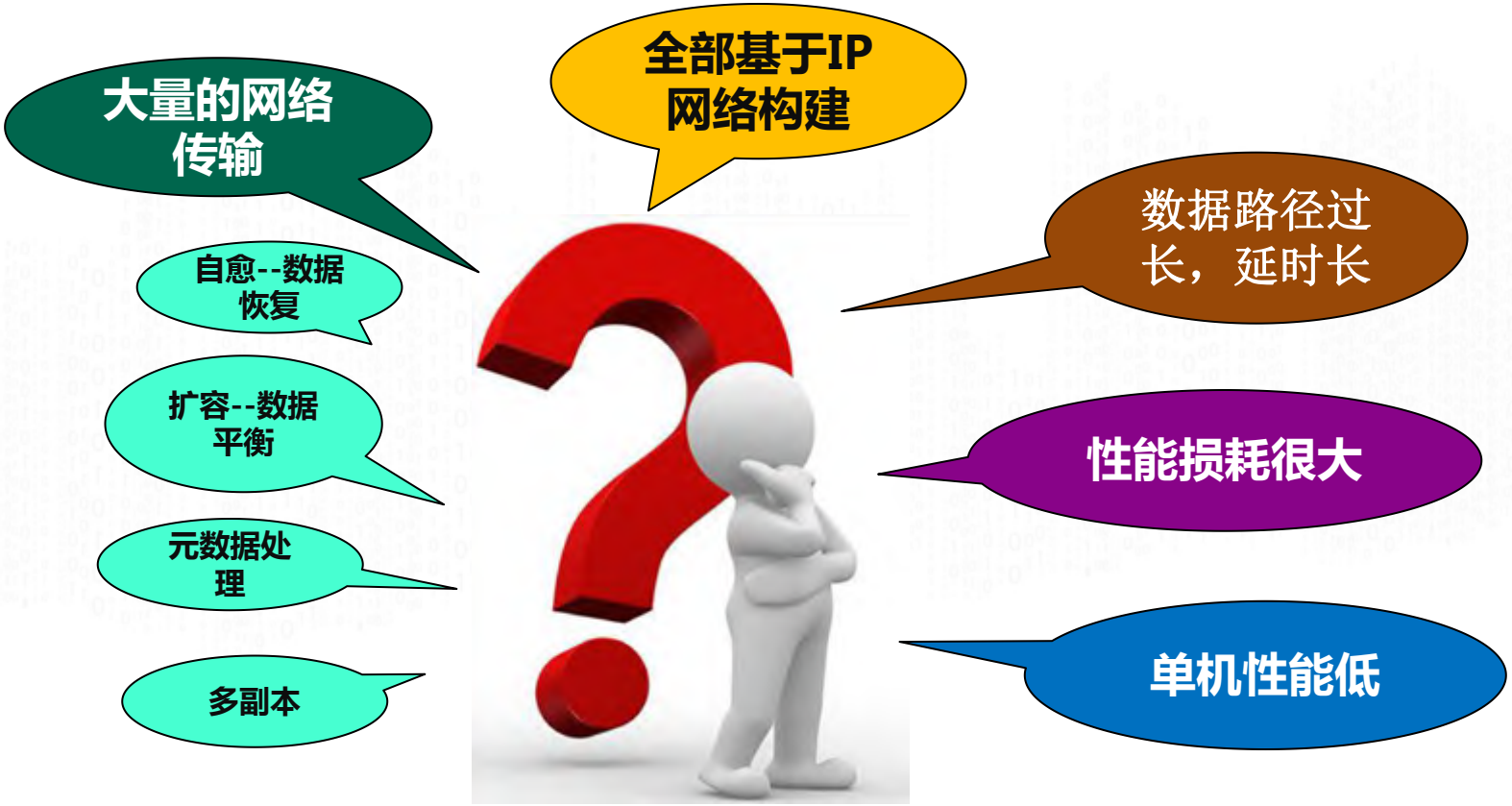
SurCloud

大量的开源分布式存储



SurCloud

常见分布式存储的主要问题



SurCloud

目录

SAS交换网络的存储架构

硬件结构

整合存储架构的数据通道以SAS交换机进行传输

存储控制节点

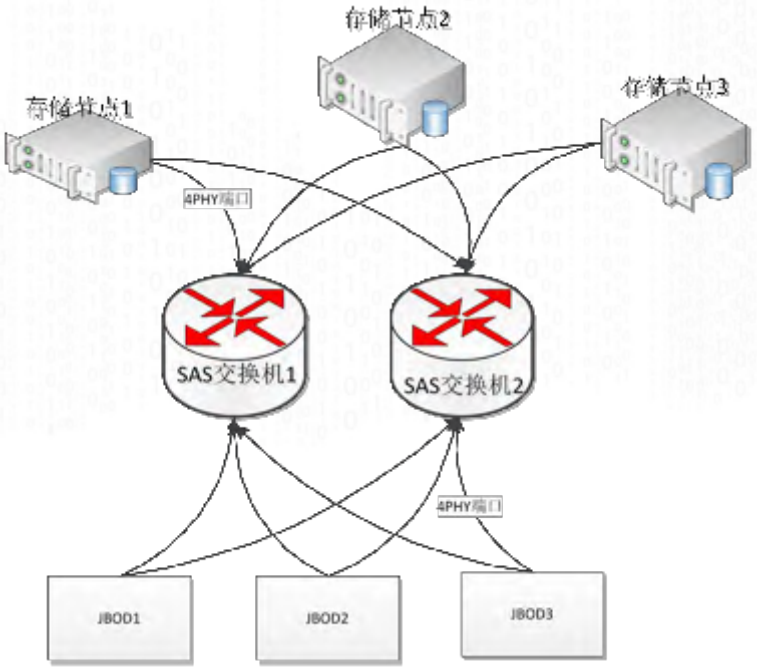
采用x86服务器，对等分布，支持横向扩展，通过HBA卡与SAS交换机连接

SAS交换机

以SAS交换机为核心组成一个SAS存储网络，SAS交换机的带宽24Gb或48Gb

JBOD磁盘柜

独立的扩展柜，可插入45块或更多S机械硬盘或者SSD盘



SurCloud

SAS交换网络的存储架构

核心原理

基于全局存储池

- 一个SAS交换机组成的一个存储集群
- 所有磁盘都是全局可见
- 所有服务器都可以同时访问任意磁盘，形成全局存储池
- 通过服务器上的存储模块协同实现磁盘的并发读写控制
- SurFS本身作为软件模块就是实现一个存储集群中服务器的协作

高性能，低成本，高可靠，
高可用，扩展性好

全局存储池系统的特点

更合理的高可用方式

- 利用多路径技术实现服务高可用
- 利用软Raid或纠删码技术实现数据的高可用
- 实现更简单的虚拟卷接管方式

高性能、低延迟

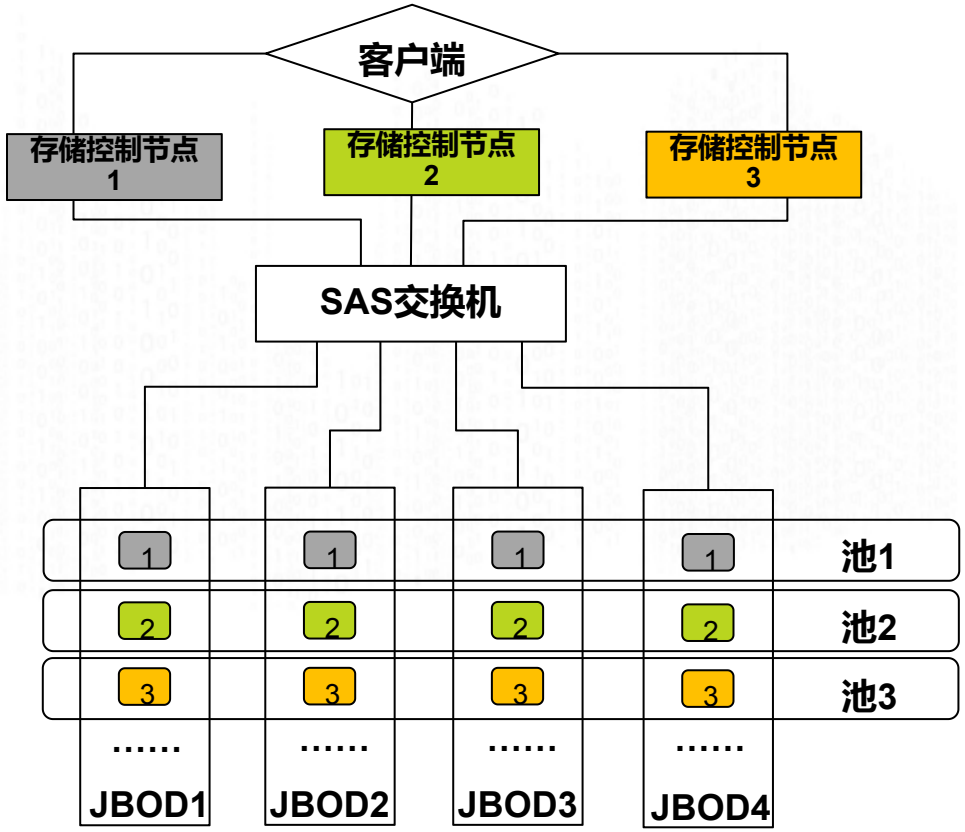
- 数据路径最短：相当于本地存储的数据路径
- 网络带宽最大：24Gb甚至48Gb的SAS网络延迟低：SAS协议的延时远低于TCP/IP延时

大容量，灵活横向扩展

- 一个SAS集群系统可以支撑数PB级存储
- 支持跨SAS网络的集群扩展，支持大规范的存储应用

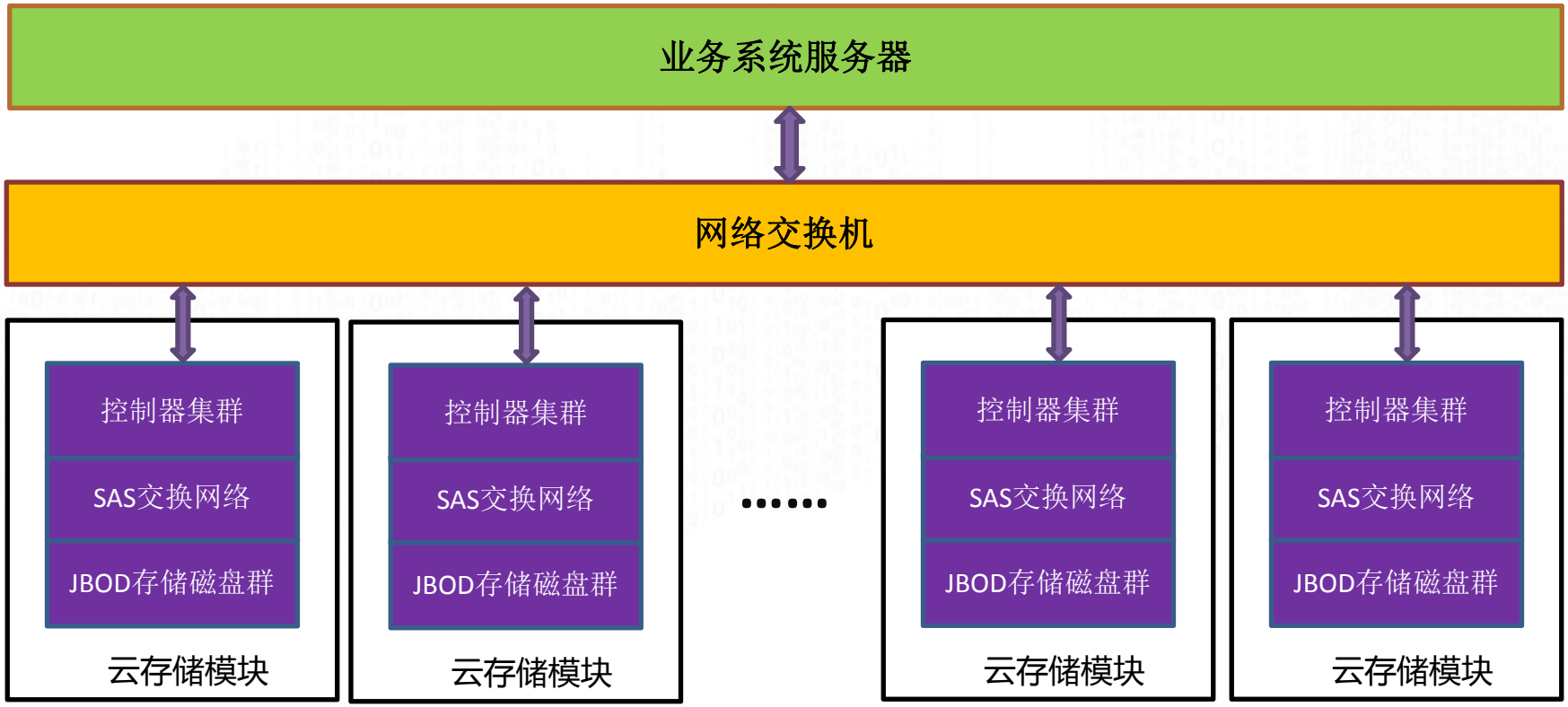
低成本

- 接近千兆网的成本
- 不需要额外的存储设备和组件



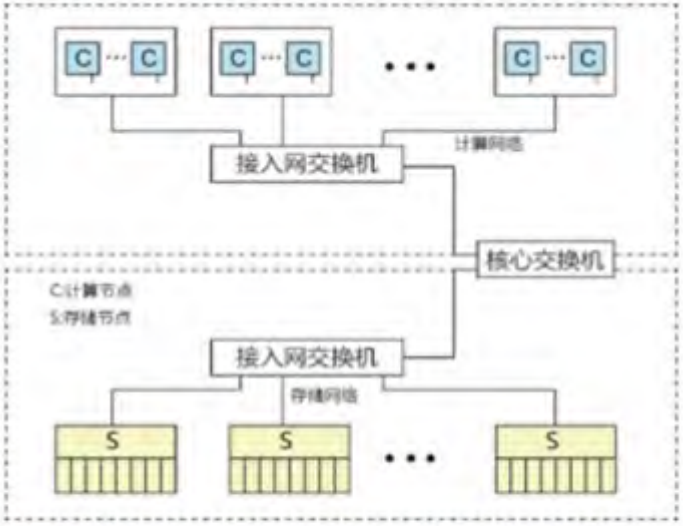
SurCloud

灵活的横向扩展能力



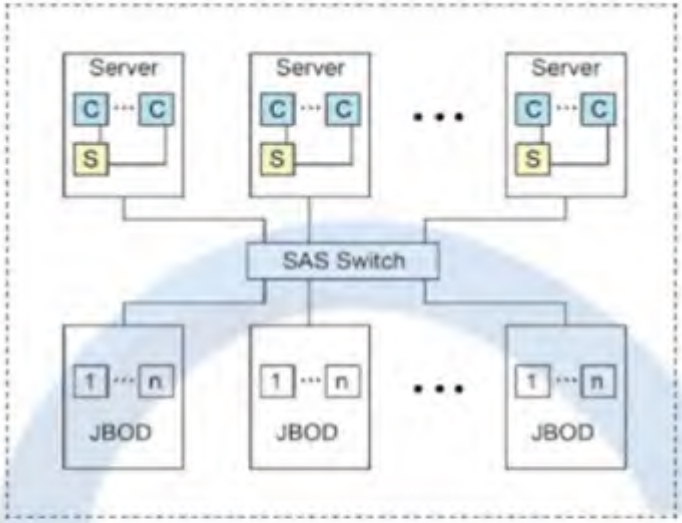
超融合SAS架构与其它分布式存储对比

其它分布式存储系统



第1步：存储介质到存储节点
第2步：存储节点到存储接入网交换机
第3步：存储接入网交换机到核心网交换机
第4步：核心网交换机到计算接入网交换机
第5步：计算接入网交换机到计算节点

SurFS

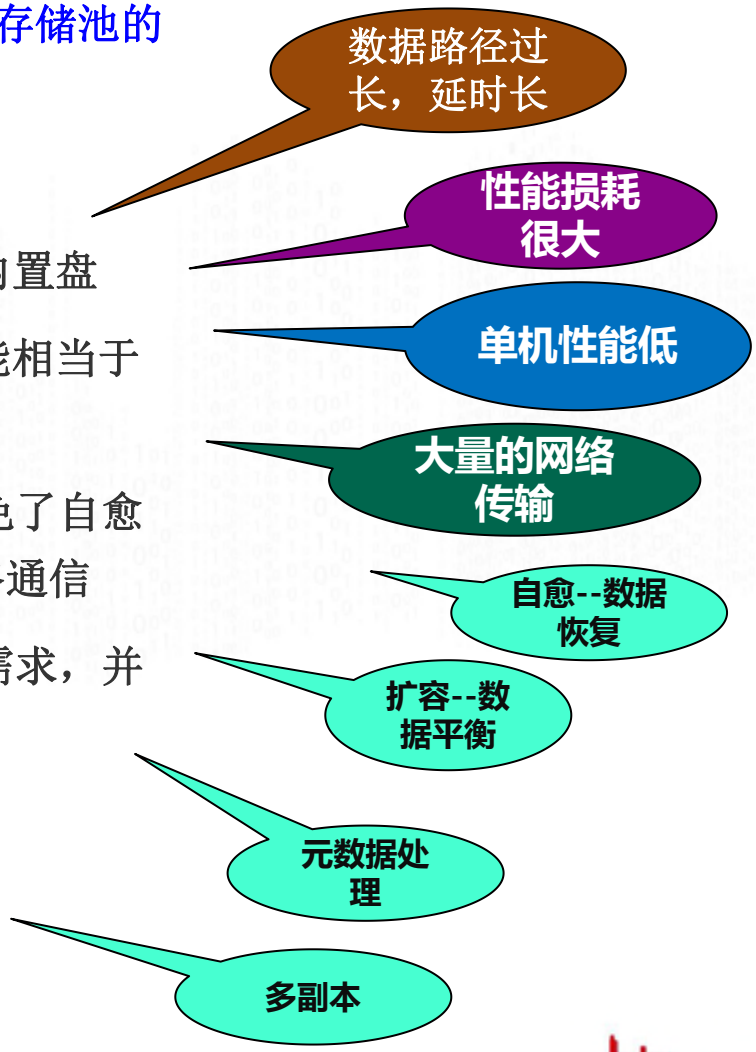


第1步：存储介质到存储节点
第2步：存储节点到聚合在同一服务器的计算节点（CPU总线通道）
总数据路径接近于传统路径的第1步

主要解决的分布式存储问题

SurCloud

- 通过存储控制节点与计算节点超融合，以及全局共享存储池的特性
  - 计算节点上的数据读写直接走SAS网络，延迟低
  - IO路径短，数据路径压缩到极致，几乎相当读写内置盘
  - 避免TCP/IP的打包过程， 性能损耗很小，单机性能相当于传统的商业存储性能
  - 存储节点的宕机，直接由其它节点接管磁盘，避免了自愈和数据平衡过程。不仅性能提高，而且减少了网络通信
  - 采用纠删码技术，大大减少了多副本的网络传输需求，并降低了成本，通过SAS网络又避免了性能下降。
- 成本低：成本节约一半以上
  - 支持纠删码，明显低于多副本的存储容量需求
  - 对IP网络设备的依赖低



SurCloud

基于SAS架构的主要不足

- SAS线的传输距离较短
  - 一个SAS集群内的设备连线不超过10米
  - 一个SAS集群的存储规模一般最多在1-3PB的数据量
- 更大规模的扩展性，需要以太网络通信
  - 一个SAS集群内的扩展性非常好
  - 多个SAS集群之前当前只能走以太网络，降低了超融合能力
  - 需要非常强大的软件控制技术

SurCloud

目录

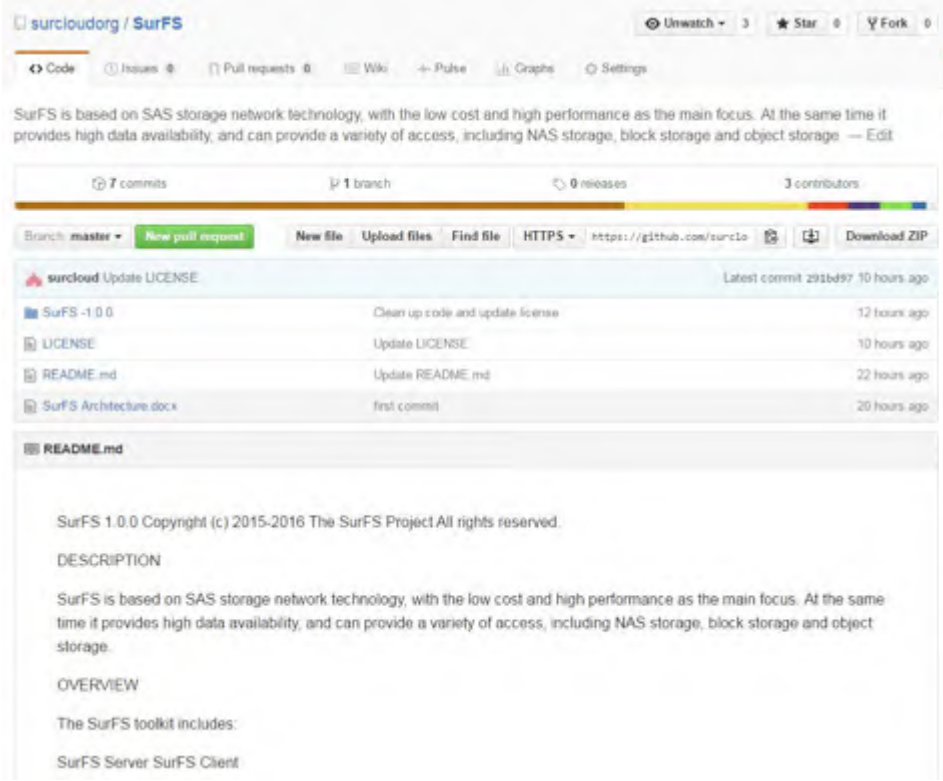
SurFS的整体介绍

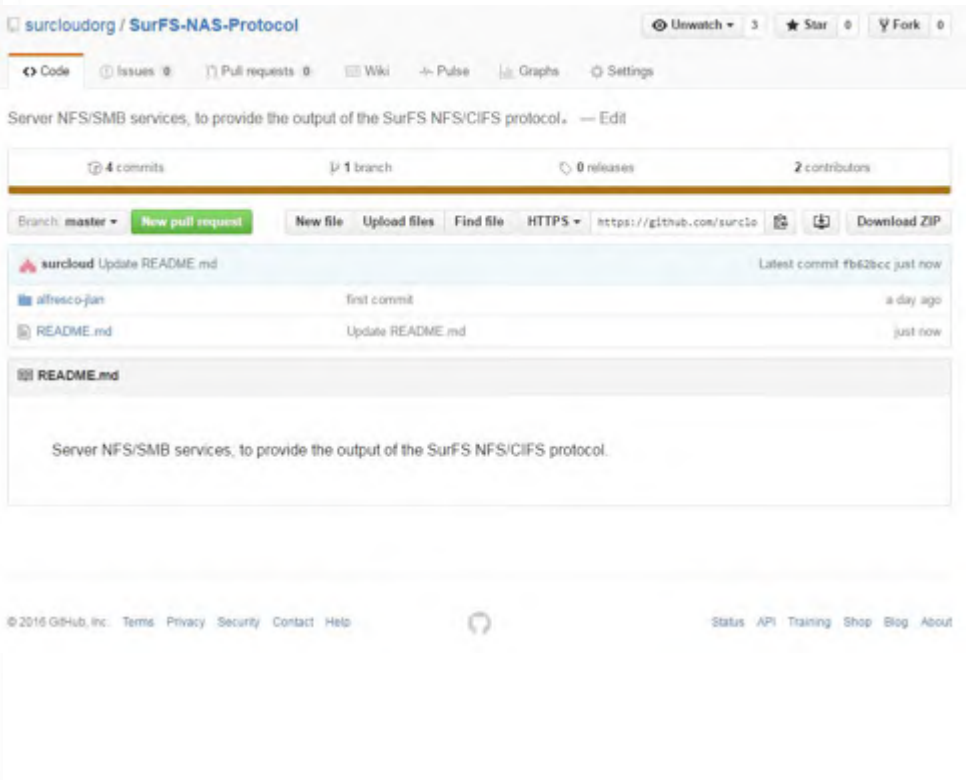
- SurFS是书生云公司首创的基于SAS网络架构构建的分布式存储系统，专为高性能低成本云存储而设计
- SurFS从2012开始开发，历经4年，实现对块存储、NAS存储和对象存储的统一支持
- 主要特点：
  - 基于SAS交换机构建分布式存储的后端，形成24Gb/468Gb的SAS存储网络
  - 通过分离存储控制节点和存储介质，实现全局访问的存储池
  - 利用全局存储池技术，更适合纠删码技术的应用，实现短I/O路径、低延时的高性能存储技术
  - 通过对存储控制节点的软件定制化，能够与OpenStack计算节点实际超融合模式，通过存储与计算的全面融合，为OpenStack社区提供了一种高性能、低成本的存储后端。

SurFS产品开源概况

- SurFS遵守业界标准，开放源码到GitHub。
- SurFS主工程：遵循MPL协议
  - 块存储系统
  - NAS存储系统
  - SAS管理工具
  - 采用ZFS文件系统为底层系统
  - https://github.com/surcloudorg/SurFS
- SurFS-Nas-Protocl工程：遵循GPL协议
  - 基于Alfresco JLAN开源软件，实现对NFS、CIFS协议的支持
  - https://github.com/surcloudorg/SurFS-NAS-Protocol
- 针对OpenStack的驱动模块
  - 提供Cinder Driver，实现OpenStack与SurFS系统的对接
  - 已经提交OpenStack社区

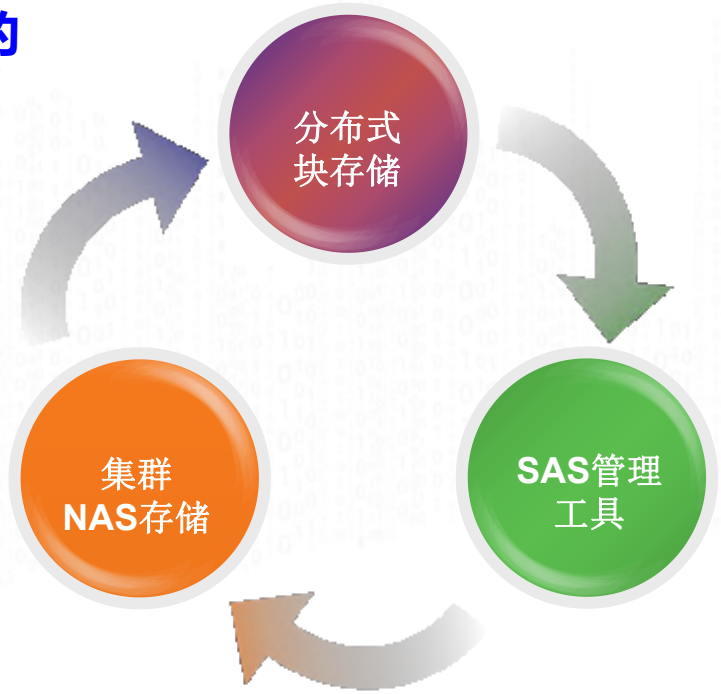
SurFS产品的开源项目截图





SurFS的主要功能

基于SAS架构实现灵活的分布式云存储系统



集群NAS存储

- ✓ 支持负载均衡
- ✓ 支持权限控制
- ✓ 兼容性NFS和CIFS协议
- ✓ 提供分卷管理
- ✓ 针对SAS网络优化

分布式块存储

- ✓ 池管理
- ✓ 卷管理
- ✓ 快照管理
- ✓ 卷导出管理
- ✓ 支持路径的自动优选
- ✓ 支持聚合存储

SAS管理工具

- ✓ 磁盘监控管理
- ✓ 磁盘日志管理
- ✓ 集群监控管理

SurFS的主要优势

- 性能高：逼近内置盘的读写性能
  - 采用SAS网络，延迟低
  - IO路径短，数据路径压缩到极致，几乎相当读写内置盘
- 成本低：成本节约一半以上
  - 支持纠删码
  - 对IP网络设备的依赖低
- SurFS支持横向扩展，实现企业级的高可用和高可靠
  - 更容易的fail-over控制
- 实现与OpenStack的集成：SurFS Driver for OpenStack
  - SurFS支持兼容模式和聚合模式
  - SurFS提供聚合存储API，能够根据VM的地址自动提供最近的存储位置

SurCloud

云存储系统性能对比

	SurFS	Gluster FS	对比
4K随机读	30MB/s	6.5MB/s	461%
4K随机写	32.9MB/s	2.1MB/s	1565%
64K顺序读	364MB/s	38MB/s	957%
64K顺序写	323MB/s	33MB/s	977%
磁盘数	9	18	200%
备注	对比数据来源于OpenStack官方网站http://www.openstack.cn/wp-content/uploads/2014/07/XinLiXun-GlusterFS-VS-Ceph-v3.pdf SurFS数据系采用相同配置自行测试，唯一差别是用EC码使得仅用一半的磁盘就获得同样的可靠度，代价是更高的计算复杂度影响性能，同时磁盘越少对性能也越不利。尽管如此性能上依然有压倒性优势		

SurCloud

云存储成本对比

2PB存储	SurFS		其它云存储	
		成本		成本
硬盘4TB	620块	837000	1500块	2025000
sas交换机	4个	48000	0个	0
45盘位JBOD	14个	210000	0个	0
服务器	4个	312000	150个	1612500
HBA卡	8个	14800	0个	0
万兆网卡	4个(双口)	7400	150个(单口)	195000
24口万兆交换机	1个	15000	7个	105000
万兆网线	4根	724	150根	27150
		￥1,444,924.00		￥3,964,650.00

SurCloud

目录

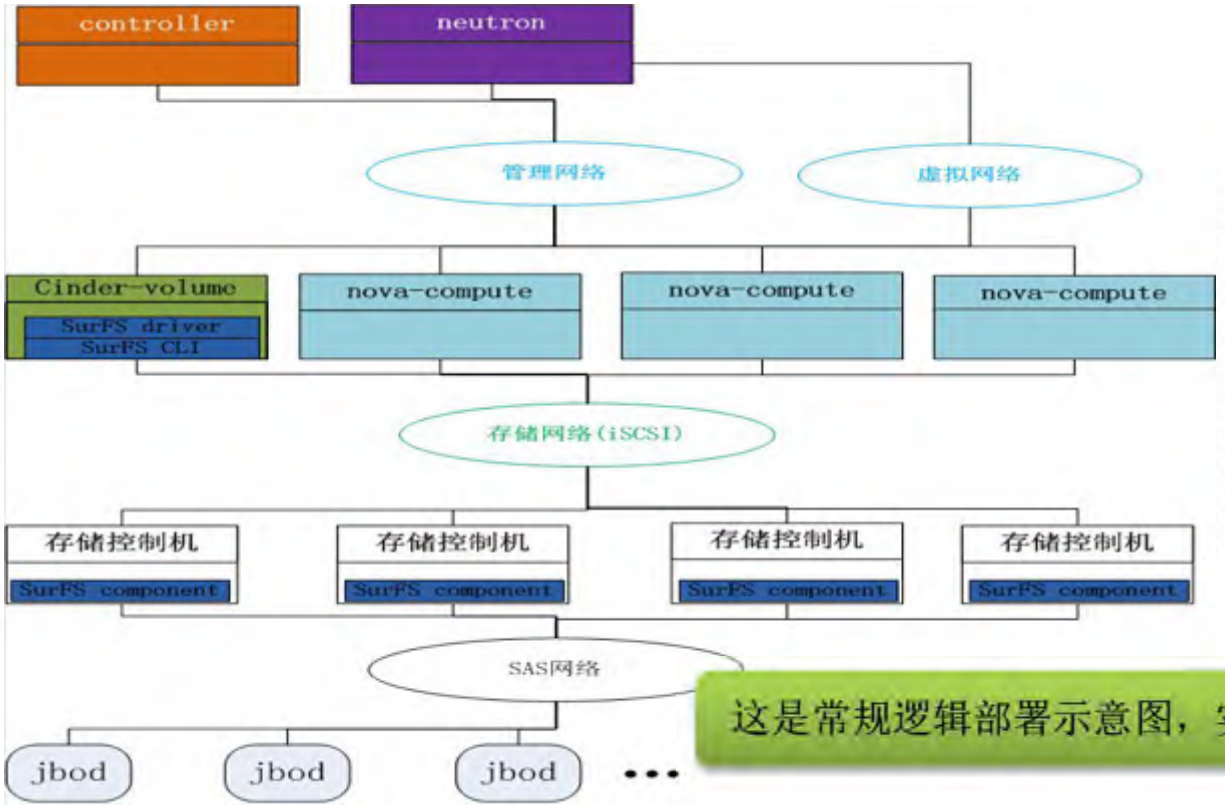
SurFS driver For OpenStack

- SurFS driver 是用来连接openstack 与SurFS存储系统的驱动。
- SurFS driver不像其它驱动一样管理着iSCSI target，这部分功能已经分离出来，交给了SurFS 存储系统本身来管理。
- 通过把控制链路与数据链路分开的方式，让cinder-volume节点更容易实现高可用。

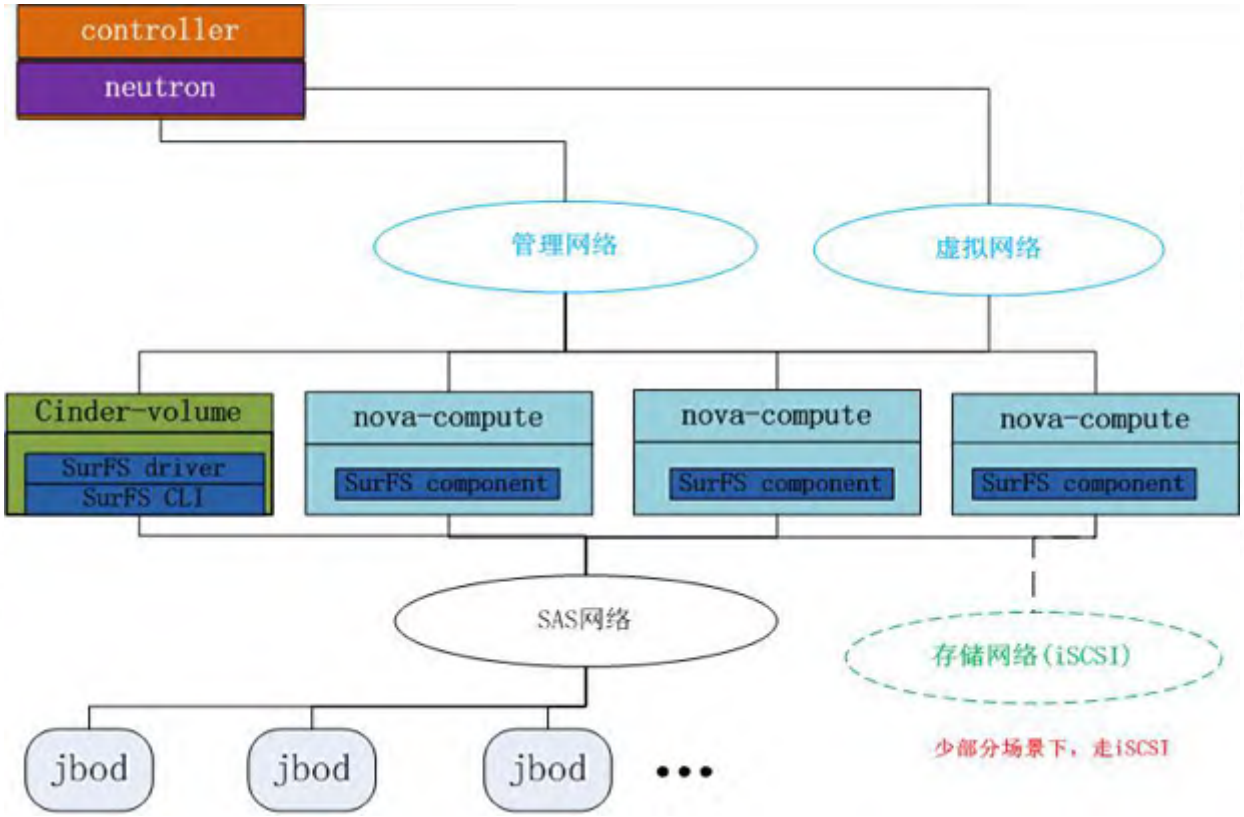
编号	类型	功能
1	卷操作	创建卷
2		从已有卷创建卷（克隆）
3		扩展卷
4		删除卷
5	卷-虚机操作	挂载卷到虚机
6		分离虚机卷
7	卷快照操作	创建卷的快照
8		从快照创建卷（恢复卷）
9		删除快照
10	卷-镜像操作	从镜像创建卷
11		从卷创建镜像

- SurFS driver支持根据VM地址，跳过ISCSI模型，直接支持SAS网络模式访问存储。

SurFS常规部署图(通用模式)



SurFS超融合模式署示意图



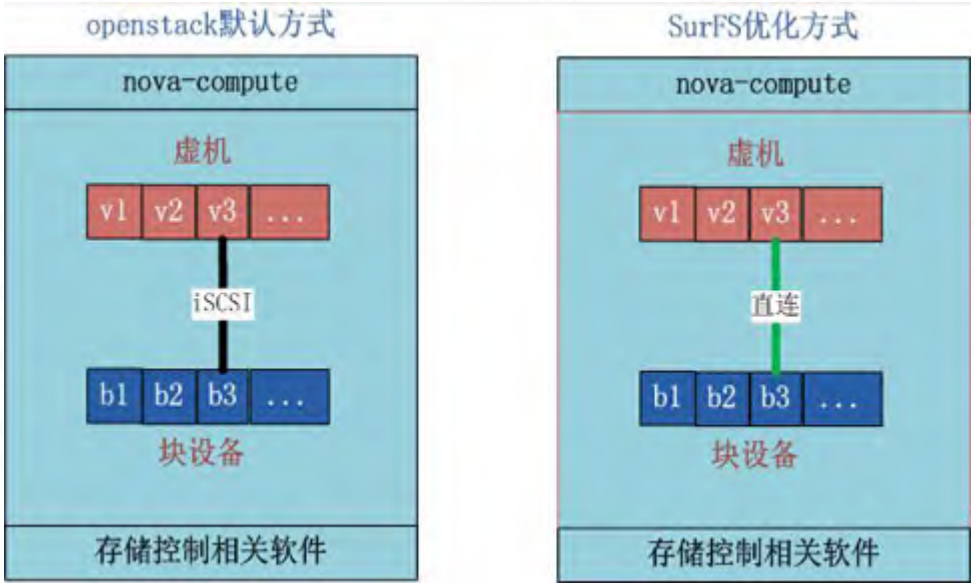
SurFS与openstack 结合所做的优化

优化存储访问路径

1. 挂载的块设备与虚拟机处于同一物理机时，将优化访问路径，绕过**iSCSI**。

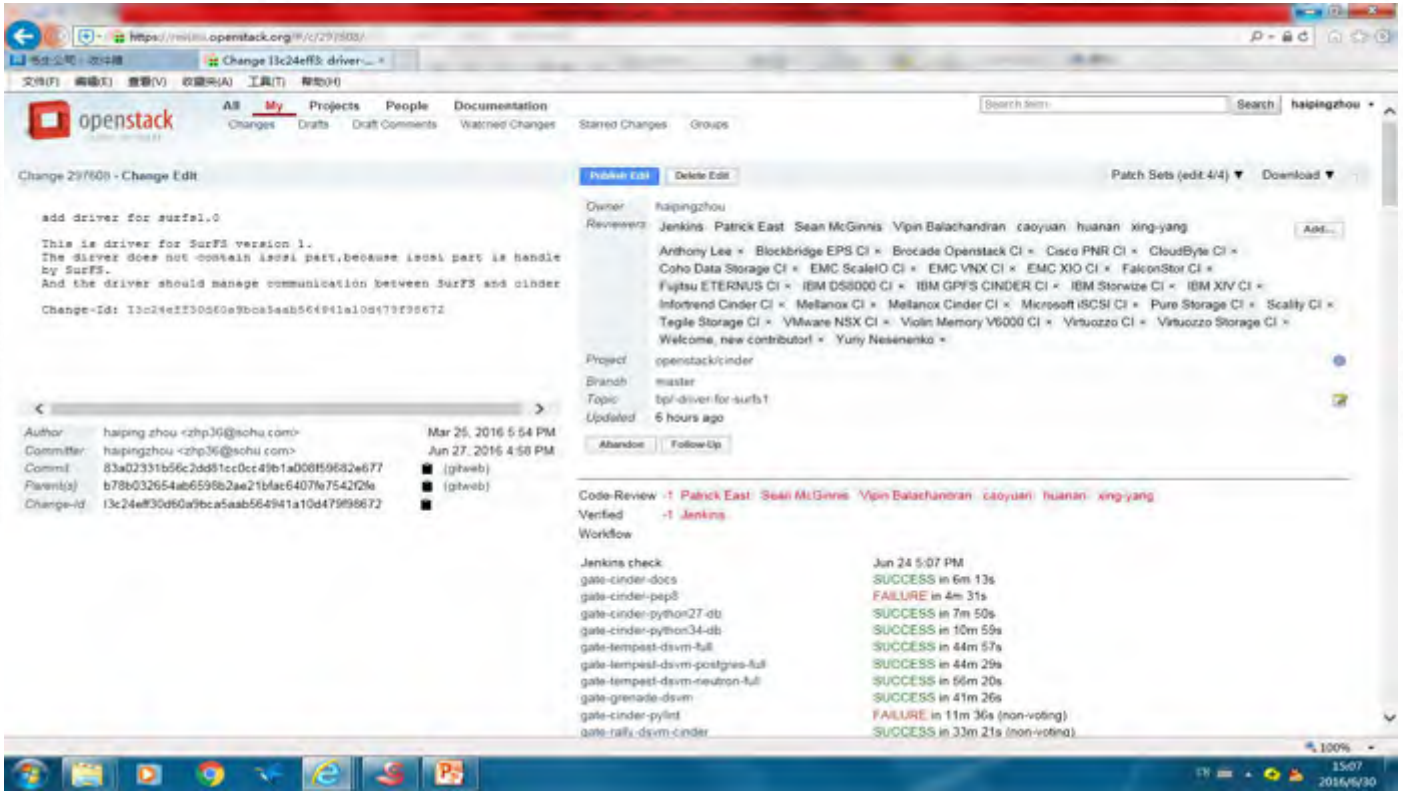
2. 创建块设备时，可以针对指定虚拟机，实现块设备与虚拟机处于同一物理机时。

3. 查看虚拟机与块设备分布情况，执行卷迁移或者虚机迁移。



SurFS driver for cinder

这部分已经提交到openstack社区，正在进行代码review阶段，已经进行到了第三轮review

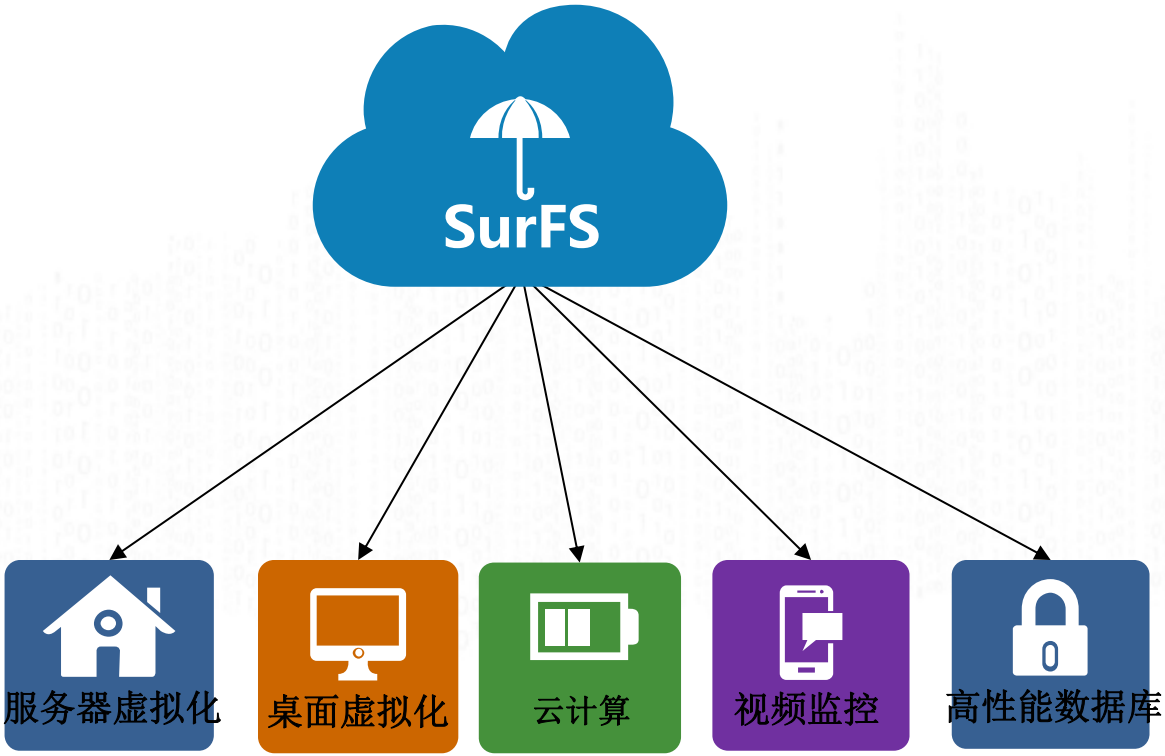


SurCloud

目录

SurCloud

典型应用场景

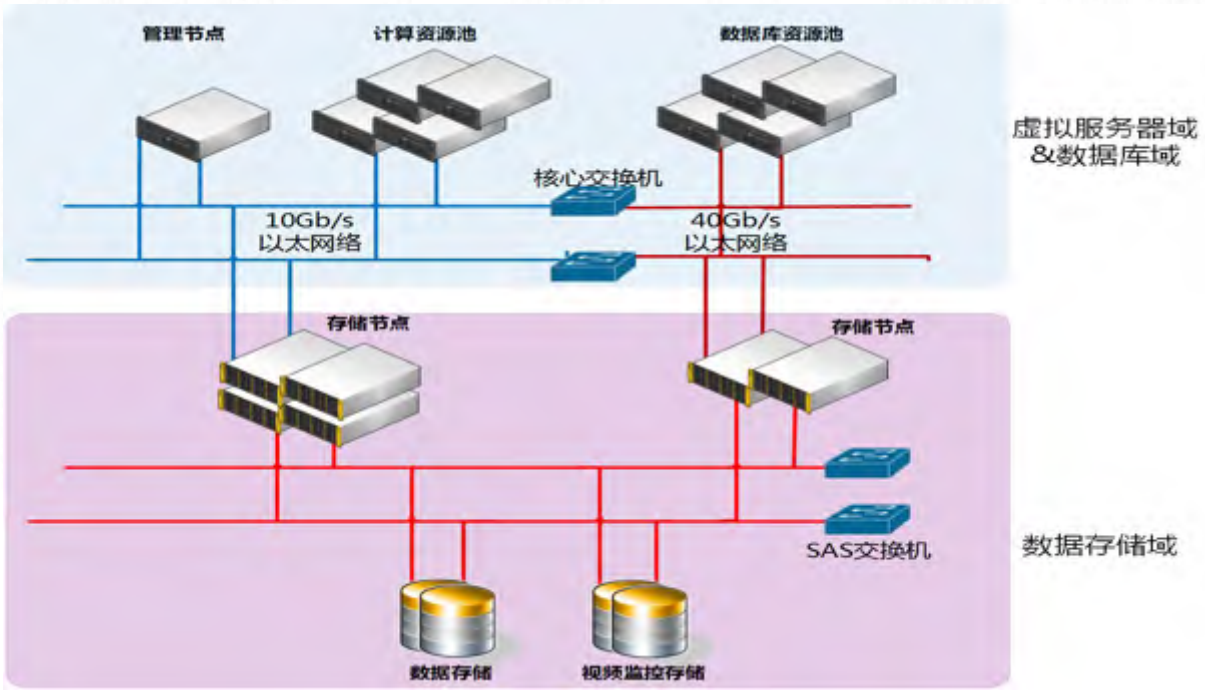


- ✓ 具有分布式存储特征
- ✓ 高性能，低延时
- ✓ 高弹性
- ✓ 计算存储融合
- ✓ 兼容传统存储应用
- ✓ IP网络的带宽依赖低

成功案例-内蒙呼市玉泉区政务云平台

解决的问题

基于SAS架构的SurFS提供了视频监控存储和云平台服务的存储需求，存储容量约2PB

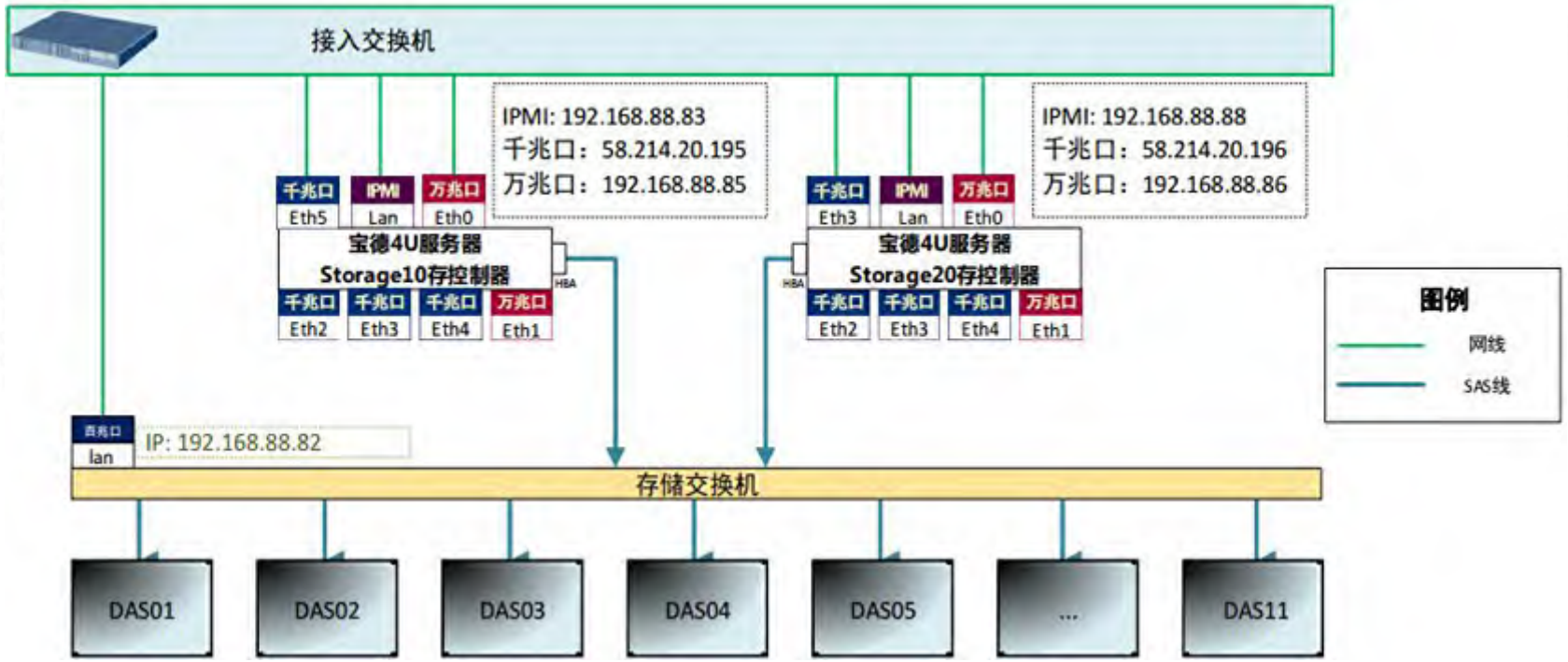


SurCloud

成功案例-无锡媒体云项目

解决的问题

基于SAS架构的SurFS提供了媒体云上桌面虚拟化需求，存储容量约2PB

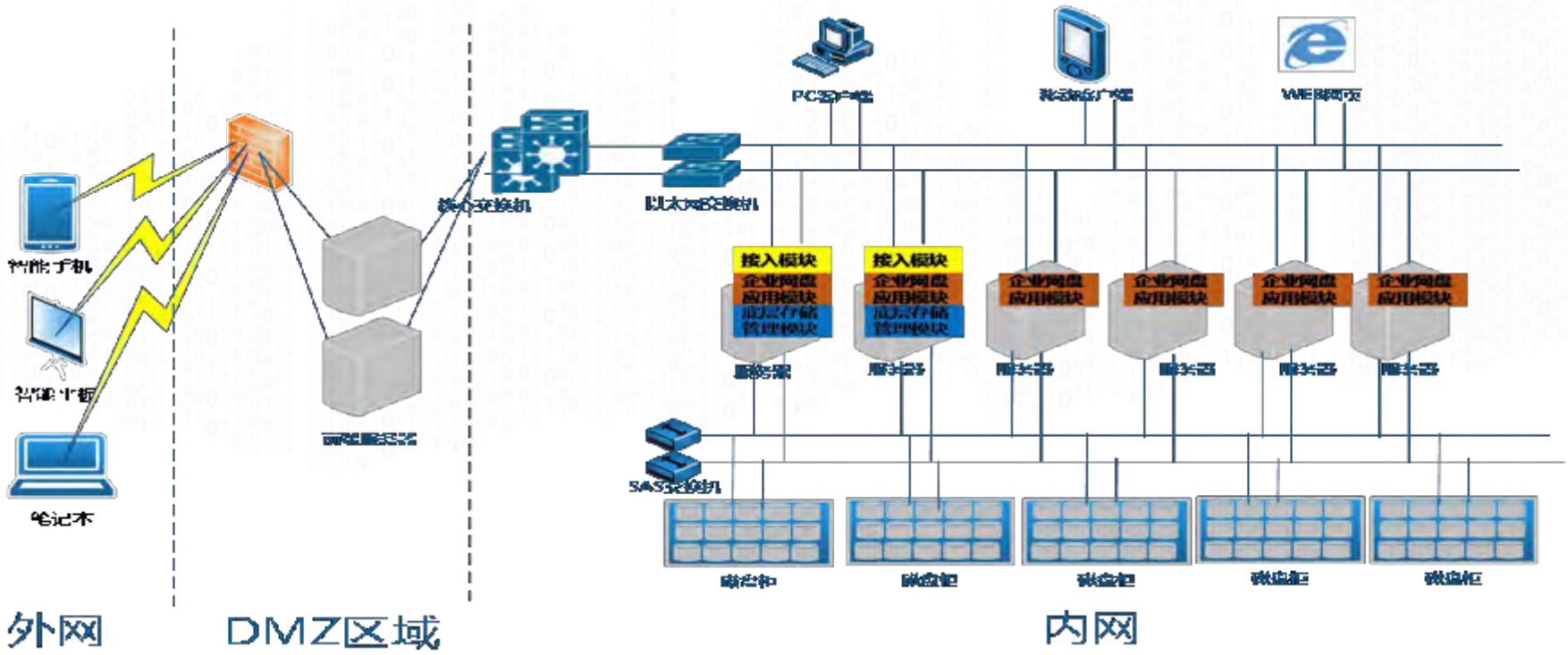


SurCloud

成功案例-东风集团企业云盘

解决的问题

基于SAS架构的SurFS提供了企业网盘和协同办公的存储需求，存储容量约1PB



基于SAS架构的SurFS存储展望

充分利用SAS架构的全局存储池特性

Thanks