

软件定义存储产品之---

IBM SVC 助力私有云平台建设

邓毓 某农信 存储架构工程师



目录

Part 1 中小金融企业的需求点及痛点

Part 2 “混合云” ---大势所趋

Part 3 助力私有云建设的“家伙们”

Part 4 何为“软件定义存储”？

Part 5 SVC特性、原理详解

Part 6 SVC如何击破金融企业的“点”？

Part 7 SVC如何助力私有云建设？

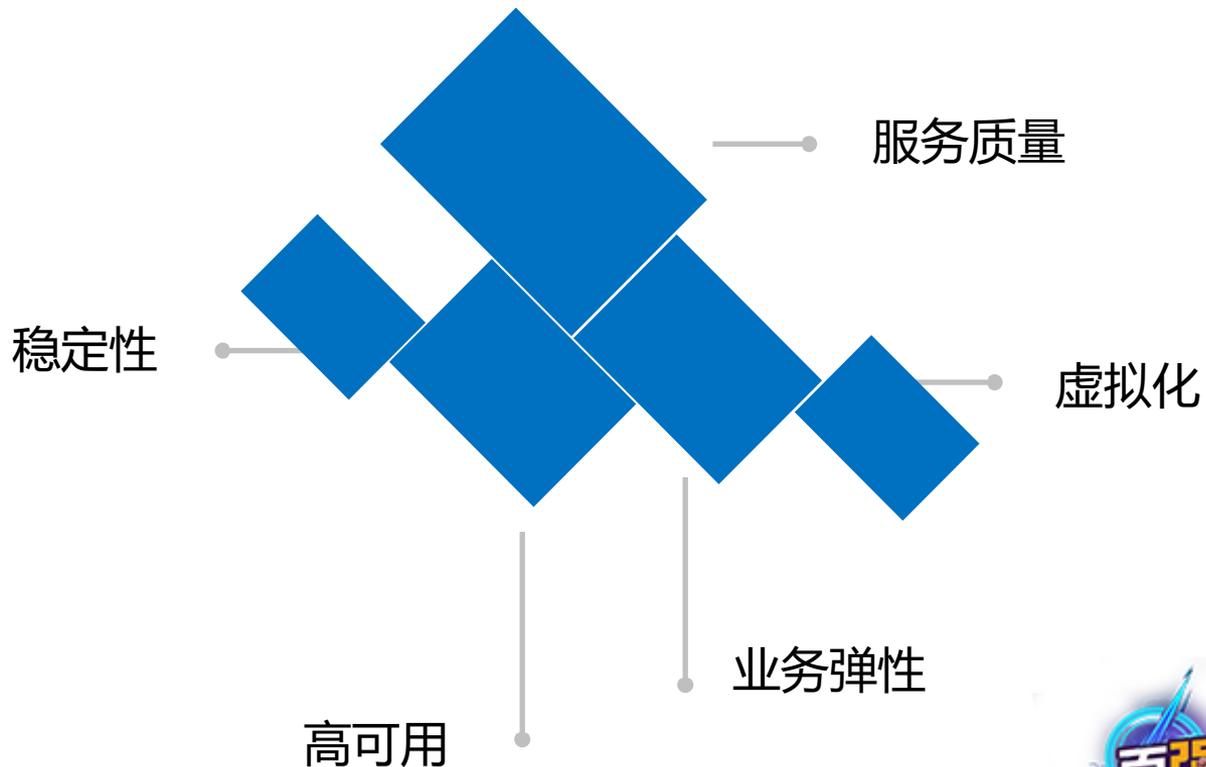
Part 8 SVC方案的收益、经验及总结





中国闪存联盟
China Flash Alliance

Part 1 中小金融企业的需求点及痛点





中国闪存联盟
China Flash Alliance

Part 1 中小金融企业的需求点、痒点及痛点

1. 存储单点
2. 应用单点
3. 容灾难

高可用

01

02

扩展性

1. 资源孤岛
2. 无法灵活扩展
3. 无法敏捷开发
4. 部署效率低下

1. 无统一规范
2. 安全漏洞多
3. 大局观差

规范性

03

04

管理复杂

1. 异构多、繁杂
2. 无服务质量控制
3. 无统一管控平台

1. 担心成本
2. 架构复杂
3. 旧产品多

云建设

05

痛点





Part 2 “混合云” ---大势所趋



混合云 兼具安全性与弹性扩展能力

混合云是融合公有云和私有云，是近年来云计算的主要模式和发展方向。企业用户出于安全考虑，更愿意将数据存放在私有云中，但是同时又希望可以获得公有云的计算资源，在这种情况下混合云被越来越多的采用，它将公有云和私有云进行混合和匹配，以获得最佳的效果，这种个性化的解决方案，达到既省钱又安全的目的”。





中国闪存联盟
China Flash Alliance

Part 2 “混合云” ---大势所趋

特点

1

更完美
私有云的安全性
公有云的计算能力

2

可扩展
突破私有云限制
公有云扩展力
公有云非机密/只读

3

更节省
降低成本
公私兼得
最佳组合





“软件定义”方式

1. 计算资源 --- 虚拟化：
powervm/vmware/kvm/xen
等
2. 网络---OpenFlow
3. 存储---SVC/VPLEX/VSAN等
4. 基础设施---HCI



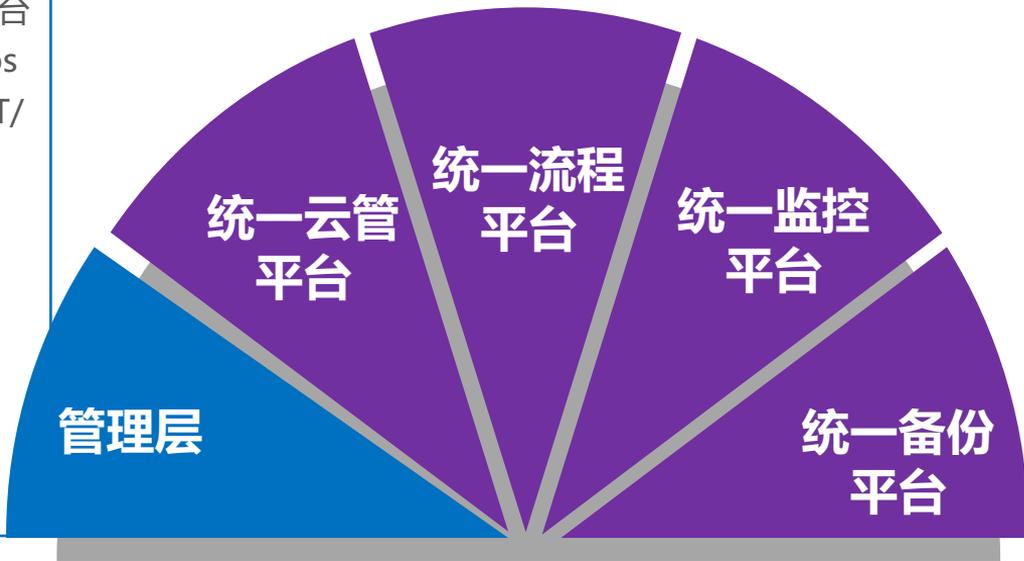


中国闪存联盟
China Flash Alliance

Part 3 助力私有云建设的“家伙们”

“统一平台”

1. 云平台---VC/ICM/ICO/
POWERVC等
2. 流程平台---基于ITIL各类平台
3. 监控---ITM/Zabbix/Nagios
4. 备份---TSM/COMMVAULT/
NBU等





中国闪存联盟
China Flash Alliance

Part 4 何为“软件定义存储”？

概念

软件定义存储（SDS）是一种数据存储方式，所有存储相关的控制工作都仅在相对于物理存储硬件的外部软件中。这个软件不是作为存储设备中的固件，而是在一个服务器上或者作为操作系统（OS）或 hypervisor的一部分。

当今很多企业的数据中心都正在经历从传统架构向云架构的转型。在转型过程中，存储往往是最重要、又是最艰难的一环。用户需要利用软件定义，来解除供应商锁定，消除专有平台孤，降低日益增长的复杂性，从而将物理的存储设备，转化为服务云计算的存储资源。同时，如何在企业自建的数据中心、外包数据中心和不同的云服务商之间，实现存储的统一管理和数据的自由流动，也是云计算落地过程中的重大挑战。

IBM Spectrum Virtualize作为软件定义存储的一种选择，它将虚拟所有的存储硬件资源，将其抽象为软件层以更高效的分配资源。当某个应用负载需要配置具有特定数据保护机制的存储，你只需要从管理界面选择配置，它将随存储自动分配。

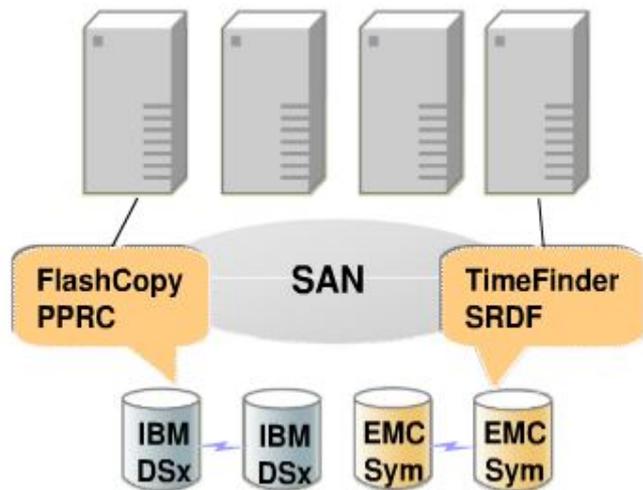
Software Define Storage – Control and Data Planes



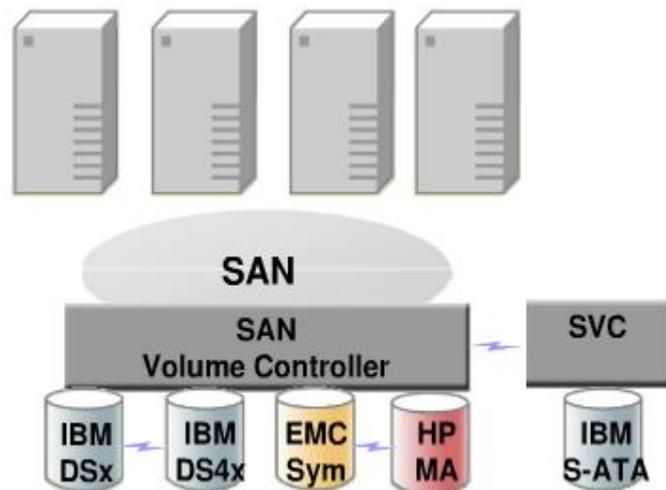


中国闪存联盟
China Flash Alliance

Part 5 SVC特性、原理详解



- 1、存储单点、容灾难
- 2、复制技术乱，不统一，需LICENSE
- 3、复制目的存储需同构
- 4、低端存储无复制功能
- 5、主机多种多路径驱动共存



- 1、VDM解决存储单点
- 2、复制技术统一，无需LICENSE
- 3、复制目的无要求
- 4、低端存储也可容灾
- 5、多路径驱动统一





中国闪存联盟
China Flash Alliance

Part 5 SVC特性、原理详解



特性

1、一套SVC集群最少包含2个节点，最大8个节点，两两节点组成一个I/O group，因此一套SVC集群最大可以包含4个I/O group



2、每组I/O group均承担着主机读写I/O负载，增加I/O group均分主机I/O负载，提升整体性能



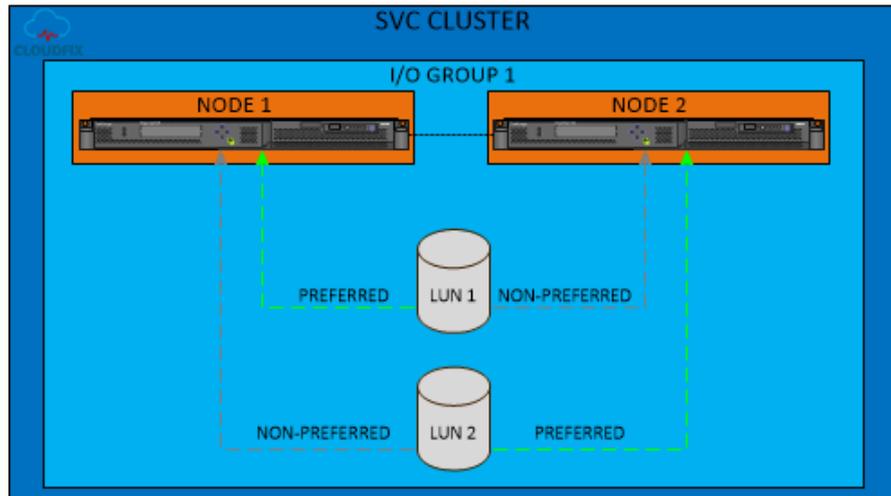


中国闪存联盟
China Flash Alliance

Part 5 SVC特性、原理详解

特性

每一个存储LUN都分配给一个I/O group，你也可以手动转换存储LUN的I/O group，比如说LUN属于I/O group0，但I/O group0负载较多，为了均衡SVC I/O group负载，这时需要手动转换LUN至I/O group1，因为每一个I/O group均由两个SVC节点组成，所以每个存储LUN均分配给一个优先的节点和一个非优先节点



01

TIPS

两个IO Group之间是支持卷在线手动转换的，如果一个IO Group完全不可用时，假如这时该IO Group中的写缓存还没来得及写入存储，该IO Group中的两个节点就同时故障了，这时将丢失一部分写数据，遇到这种情况，需要利用SVC容灾来避免该问题的出现，在灾备搭另一套集群，通过SVC PPRC来实现不同数据中心间不同SVC IO Group节点的缓存同步，避免数据丢失。

02

TIPS

AIX/Win上的SDDPCM/SDDSM多路径软件在识别Volume路径时，自动将来自两个控制器的的path进行优先区分，因此IO在主机端就会被自动发给Preferred node。对于Linux主机而言，可以将Volume的path路径优先等级标记为相同，即一视同仁，那么此时每条路径都会有IO到两个nodes，形成A-A访问模式。但在SVC内部，Non-Preferred Node会转发IO给Preferred Node，消耗一些资源。





中国闪存联盟
China Flash Alliance

Part 5 SVC特性、原理详解

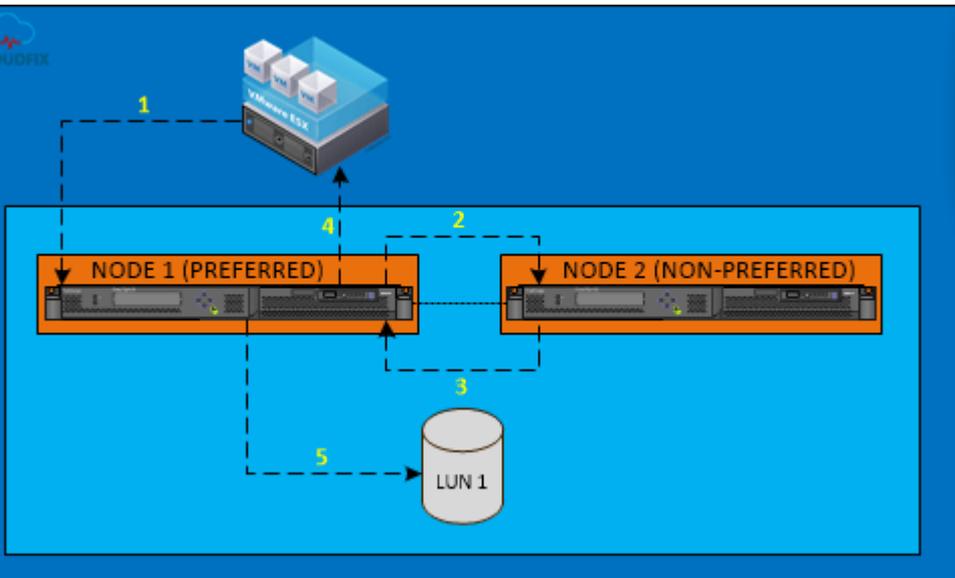


不同的存储LUN的优先节点可以不一样，这是SVC集群系统自动负载和分配的，同一SVC I/O group节点均衡负载着存储LUN，对于读I/O请求来说，来自于优先的SVC节点，而写I/O请求在同一I/O group的两个节点间同步，一次SVC I/O写请求，分解为以下5个步骤：

- 1、主机发送写I/O请求至SVC I/O group
- 2、优先的SVC节点A写入I/O至缓存，并发送I/O至同一I/O group的另一SVC节点B
- 3、节点B将写入I/O至缓存后，并响应至节点A
- 4、节点A收到节点B的响应后，向主机响应
- 5、节点A将缓存写I/O写入LUN当中

有人会说了，第5步最后才将缓存数据写入LUN当中，那么当还没有写LUN时，SVC节点突然断电了怎么办？这种情况SVC也是考虑了，每个SVC节点均配备了UPS电源，可以维持供电，可以保证缓存数据正常写入LUN当中。

同一I/O group的两个SVC节点是实时同步的，当I/O group中的主节点故障时，另一SVC节点将接管，并进入写直通模式，并禁止写缓存，也就是说只有1、5、4三个步骤。





中国闪存联盟
China Flash Alliance

Part 5 SVC特性、原理详解

高可用

VD
M

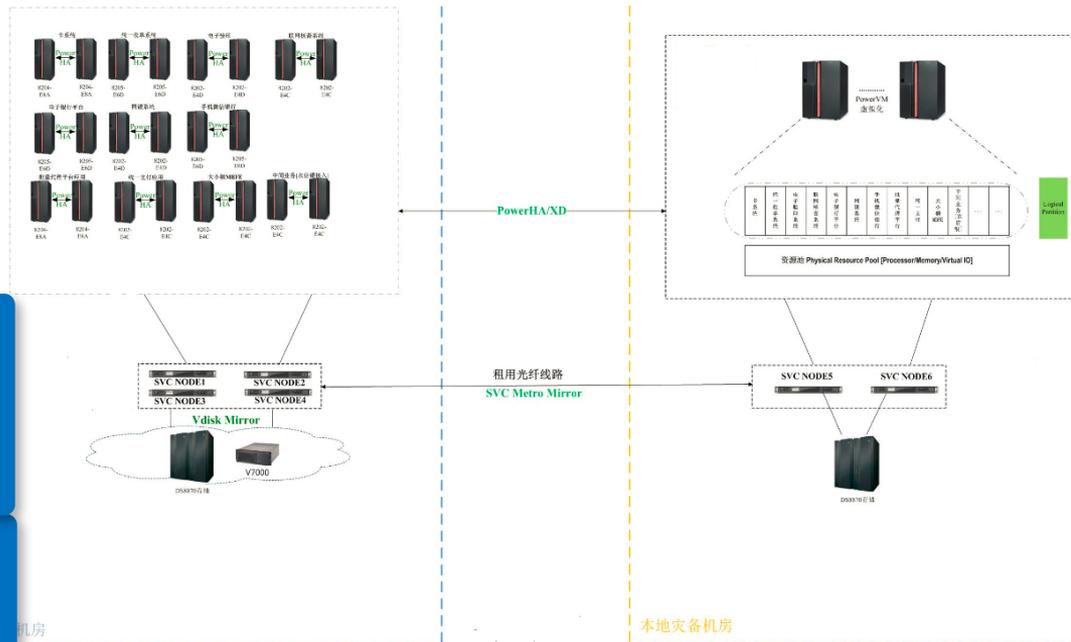
Vdisk-Mirror
本地高可用

MM

Metro-Mirror
同城容灾

GM

Global-Mirror
异地容灾



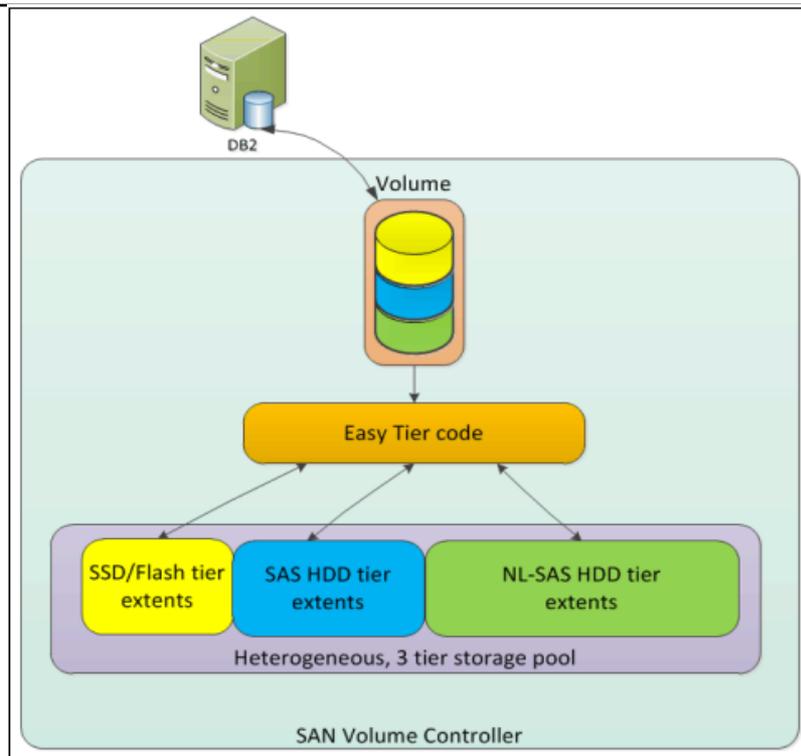


中国闪存联盟
China Flash Alliance



Part 5 SVC特性、原理详解

分层存储，也称为层级存储管理，广义上讲，就是将数据存储在不同层级的介质中，并在不同的介质之间进行自动或者手动的数据迁移，复制等操作。分层存储其实已经不是一个新鲜的概念，而是已经在存储领域应用多年。譬如SVC、V7000的easy tier技术就是从IBM DS8000系列产品中移植过来。SVC的easy tier示意图如下，flash/ssd作为性能最好的tier0层、SAS盘作为性能较好的tier1层、容量大但IO性能稍差的SATA盘作为tier2层。SVC会根据过去24小时收集每个extent访问频繁程度，制定相应的迁移计划，最后将热点数据向性能更好的tier迁移或者将冷数据下移至性能稍差的tier层。需要注意的是为保证Host IO的质量，tier间数据的迁移速率是受到限制的。





中国闪存联盟
China Flash Alliance

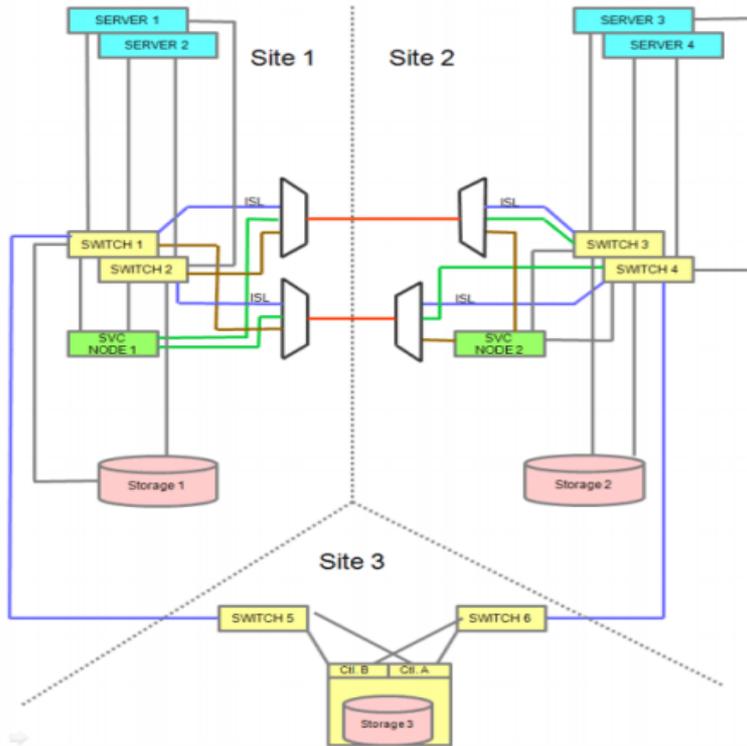
Part 5 SVC特性、原理详解

SVC Stretched Cluster

SVC Stretch Cluster也就是SVC拉伸式集群架构，就是把同一SVC集群中的SVC节点分散在不同的两个数据中心，它们之间通过光纤链路连接，并且需要第三地的仲裁设备。

SVC Stretch Cluster将一套SVC集群下的同一I/O group的两个节点拉开，分别放在两个数据中心，之前连接SVC的一套存储，拉开后各放置一套存储，通过SVC做vdisk镜像同步。说白了还是同一套SVC集群，同一个I/O group，SVC Local VDM(vdisk mirror)拉开成SVC Stretched VDM，映射给主机的volume还是同一个。

SVC版本到了7.2，Enhanced Stretched Cluster出现，在该版本，元素均赋予site属性，如SVC节点、主机、存储等，不同站点的主机对各自站点的SVC节点和存储进行读写操作，对同一volume来说，站点A的主机访问站点A的SVC节点，站点A的存储，站点B的主机访问站点B的SVC节点，站点B的存储，存储访问路径最优，两个站点的SVC节点、主机、存储都同时活动，实现应用级双活（ACTIVE-ACTIVE）





中国闪存联盟
China Flash Alliance

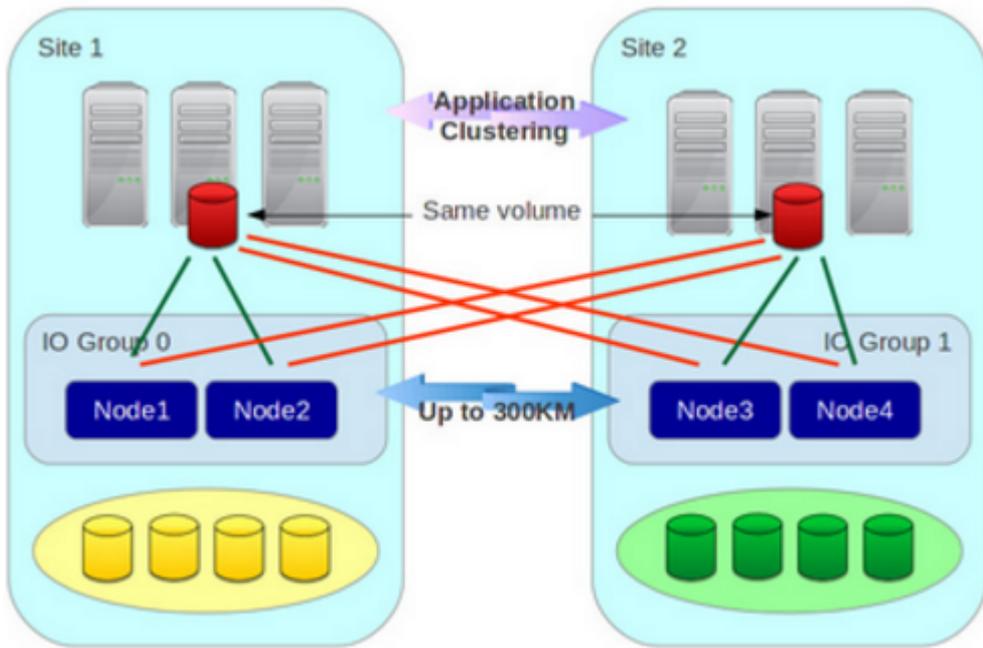
Part 5 SVC特性、原理详解

SVC HyperSwap



HyperSwap技术，在没有SVC存储虚拟化方案之前，HyperSwap技术主要用于IBM高端DS8000系列存储当中，达到应用和存储的高可用和双活需求，但是当时DS8000系列存储成本高昂，适用于核心类或者关键类应用的跨站点双活需求，不利于整体性的双活数据中心规划和建设，更别谈异构存储的跨中心双活建设了。但到了SVC 7.5版本，SVC和V7000都可以支持HyperSwap技术了，中端存储的地位瞬间提升了一个档次，异构的各类中端存储，结合SVC HyperSwap，都可以实现跨中心的双活高可用了。

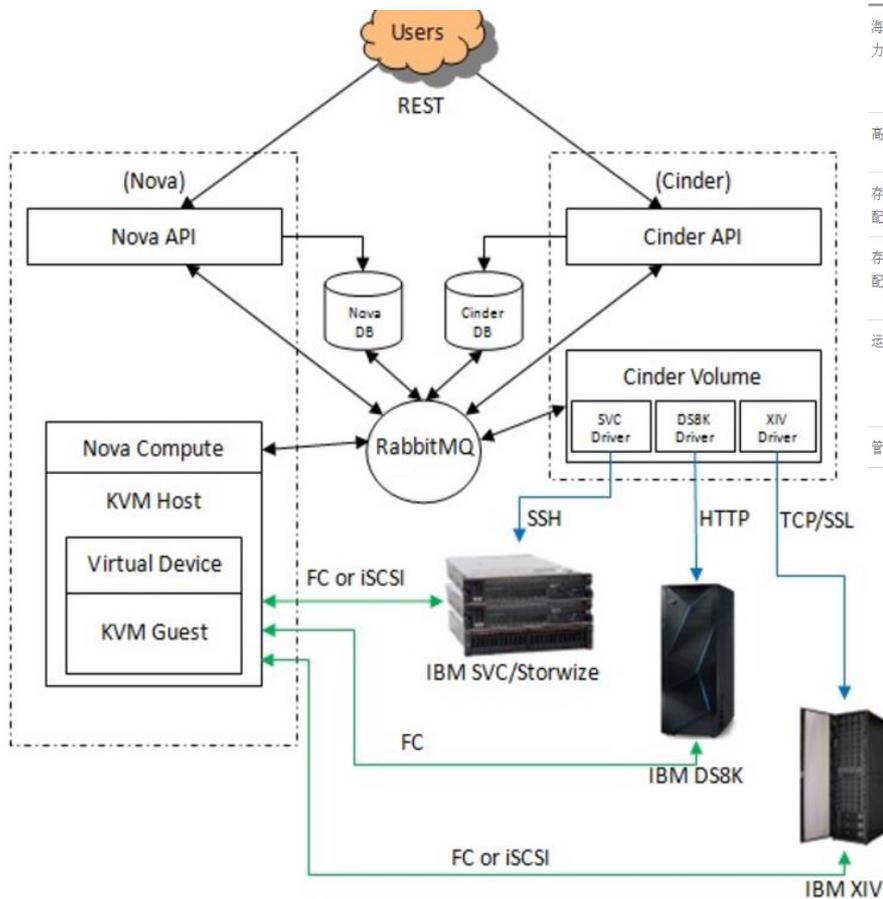
HyperSwap的卷复制ACTIVE-ACTIVE关系中，我们可以看到依然存在MASTER或者AUX的标签，对于主机来说，两个站点的其中一个I/O Group的VDISK是作为PRIMARY，提供读写，所有读写请求必须经过该I/O Group，然而hyperswap会自动决定是本站点的I/O Group的VDISK作为PRIMARY，还是主要承担I/O流量的I/O Group的VDISK作为PRIMARY。在首次创建hyperswap卷和初始化后，被标签为MASTER的VDISK作为PRIMARY，但是如果连续10分钟以上主要I/O流量是被AUX的VDISK承担，那么系统将会转换这种MASTER和AUX的关系，从这点上也可以看出与SVC Stretched cluster的不同，虽然SVC节点一样被赋予站点属性，但SVC hyperswap在另一站点仍然活动时，不局限于只从本地站点读写，它会考量最优存储访问I/O流量，从而保持整个过程中主机存储读写性能。另外需要注意的是主要的I/O流量是指扇区的数量而不是I/O数量，并且需要连续10分钟75%以上，这样可以避免频繁的主从切换。





中国闪存联盟
China Flash Alliance

Part 5 SVC特性、原理详解



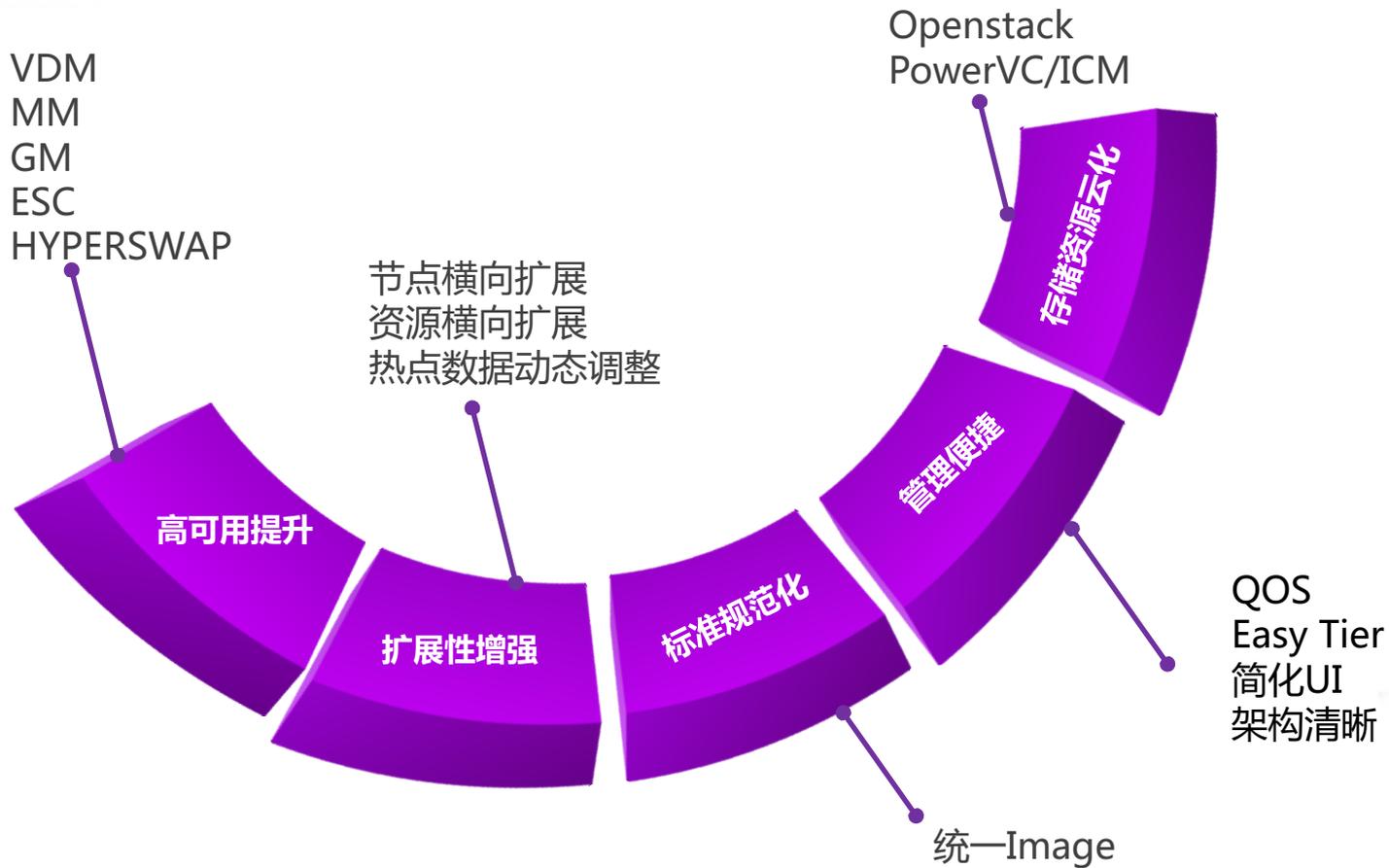
	传统存储架构	openstack 云存储架构
海量数据承载能力	扩展方式是通过增加硬件配置实现, 属于 ScaleUp 方式	存储系统可以达到 PB 级别的扩展空间更适合海量数据的存储、ScaleOut 方式
高可用	昂贵的硬件保证系统的高可用	通过系统自身的机制, 即软件完成的自动化、智能机制来保证系统可用性
存储资源动态调配的能力	存储资源分配给应用后, 难以回收再分配	计算和存储资源虚拟化, 可以按照需求分配, 动态调整
存储资源动态调配的能力	低资源利用率, 高能耗	35%-75%的 TCO 节省, 30%以上的软硬件成本节省, CPU 利用率提升到 60%-80%, 70%-80%运营成本节约
运维效率和成本	运维效率低, 维护成本高, 硬件准备周期长	部署时间缩短到分钟级, 减少硬件准备周期
管理复杂度	高	低





中国闪存联盟
China Flash Alliance

Part 6 SVC如何——击破金融企业各“点”？





中国闪存联盟
China Flash Alliance

Part 7 SVC如何助力私有云建设？

拓展弹性空间
丰富扩展性

STEP 01

提升服务质量
存储服务分层

STEP 02

存储资源虚拟化、云化

STEP 03

轻松融入云平台

STEP 04

SVC高可用和双活
将私有云跨中心
化

STEP 05





中国闪存联盟
China Flash Alliance

Part 8 SVC方案的收益、经验及总结

思路

1、SVC整合异构存储资源，SAN网络和存储架构彻底清晰

2、开展Power AIX/Linux虚拟化、X86虚拟化，分别利用PowerVM、Vmware和KVM技术

3、开始搭建Power虚拟云平台，利用PowerVC接管PowerVM计算资源和SVC存储资源两种虚拟化

4、开始搭建X86虚拟化云平台，利用ICM接管VCenter、KVM计算资源和SVC存储资源两种虚拟化；同时为了统一虚拟化云平台，用ICM接管了PowerVC

5、开始搭建统一私有云平台，利用ICO接管所有ICM，包括管理Power服务器的ICM和管理X86服务器的ICM

6、开始制作各类规范化的镜像
7、将另一数据中心的计算资源和存储资源虚拟化

后续工作：

8、设计ICO的服务编排（各种软件的自动化安装脚本、安全加固脚本、巡检脚本、监控安装配置脚本、备份安装配置脚本等）

9、搭建另一数据中心PowerVC和ICM虚拟云平台

10、利用现有私有云平台ICO/ICM接管另一数据中心PowerVC和ICM，最终形成跨中心统一私有云平台

展望：在双中心资源池吃紧时，部署些许资源至公有云，形成混合云





Part 8 SVC方案的收益、经验及总结

