



# SanDisk®

## 软件定义全闪存存储

赵建宁(Jimmy Zhao)

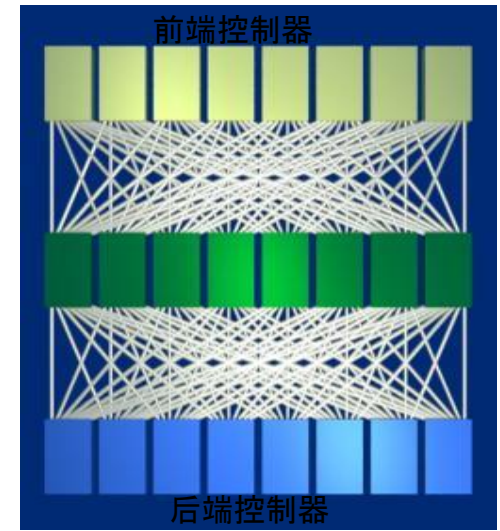
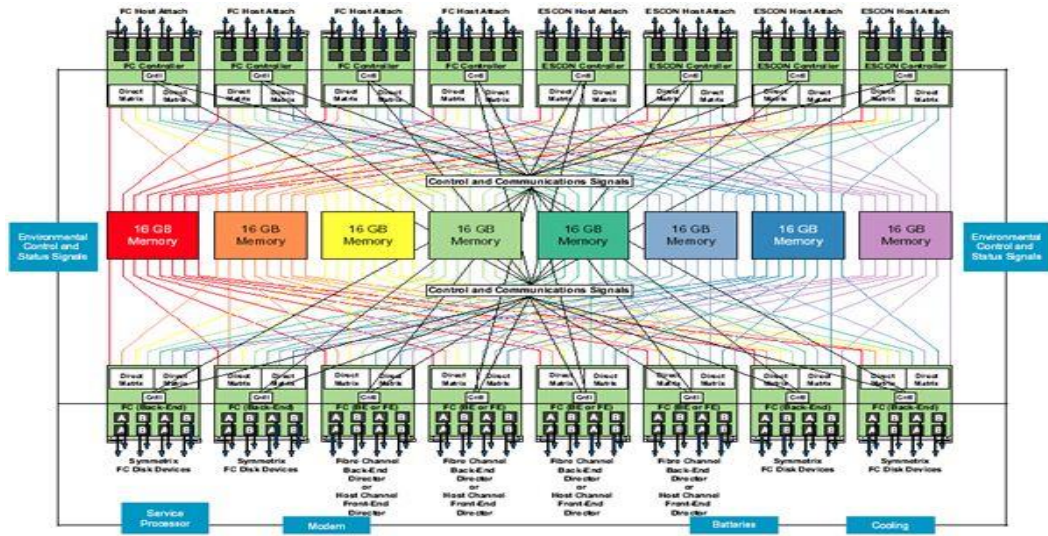
资深技术顾问

2016-01-09

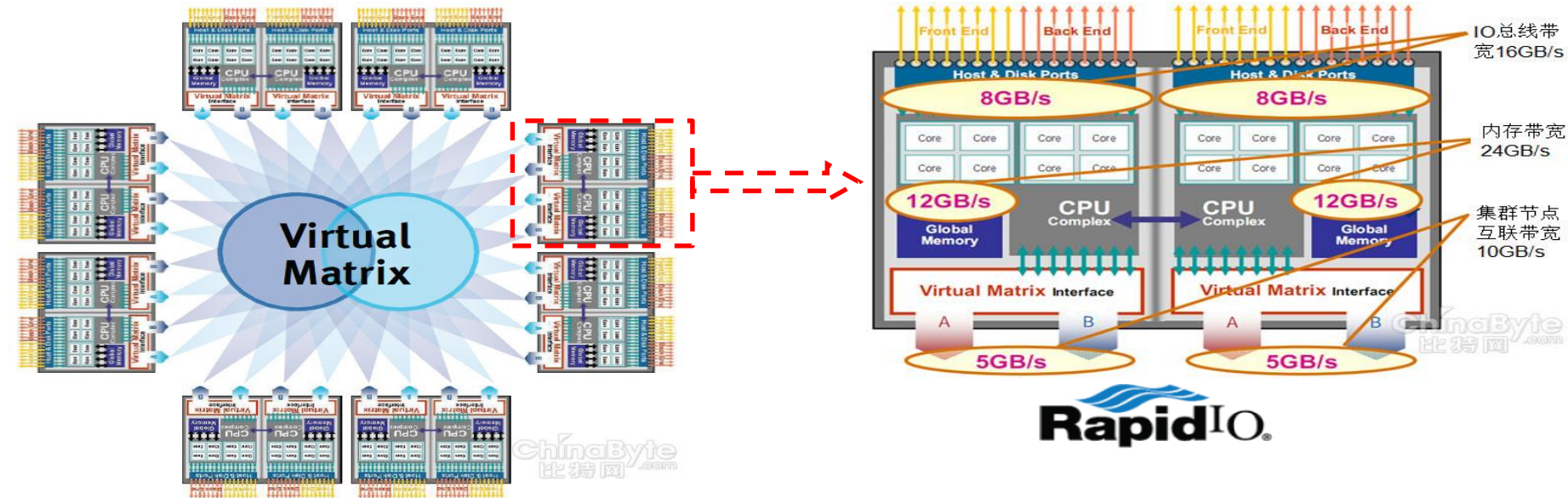
# 内容

- 传统存储硬件架构
- Server SAN 架构 (SDS)
- SANDisk全闪存存储

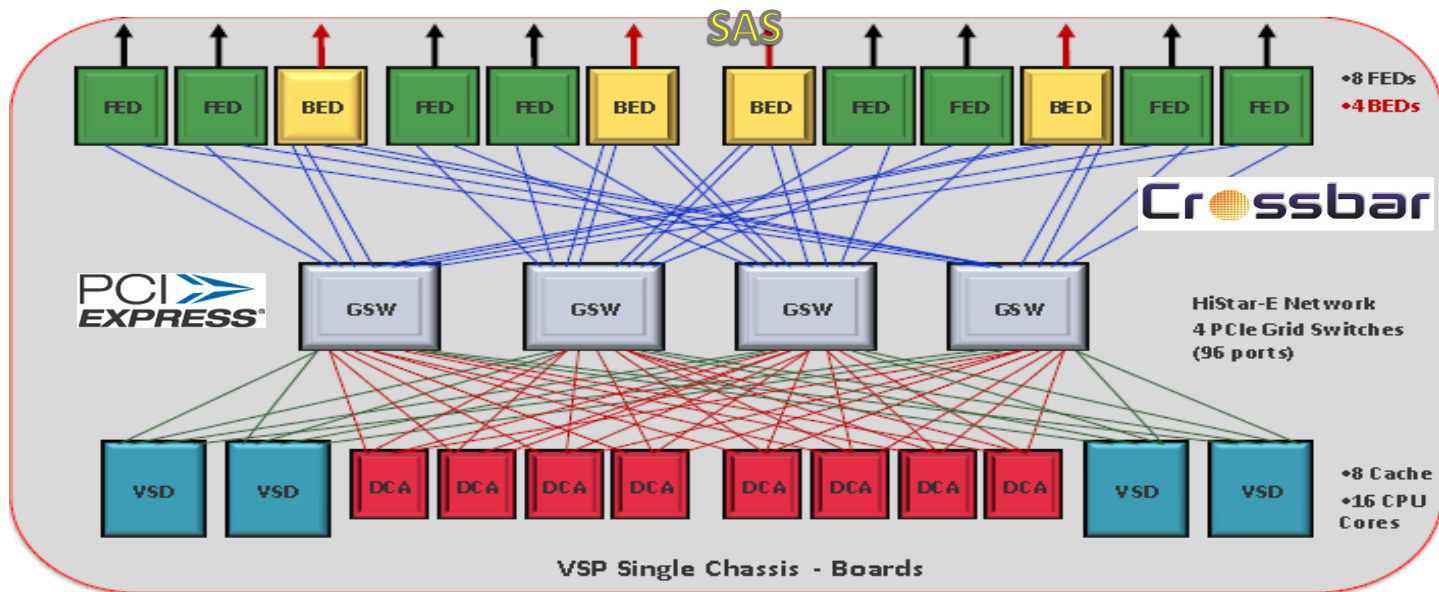
# EMC DMX4



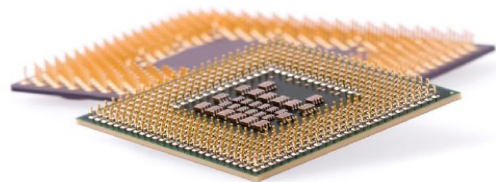
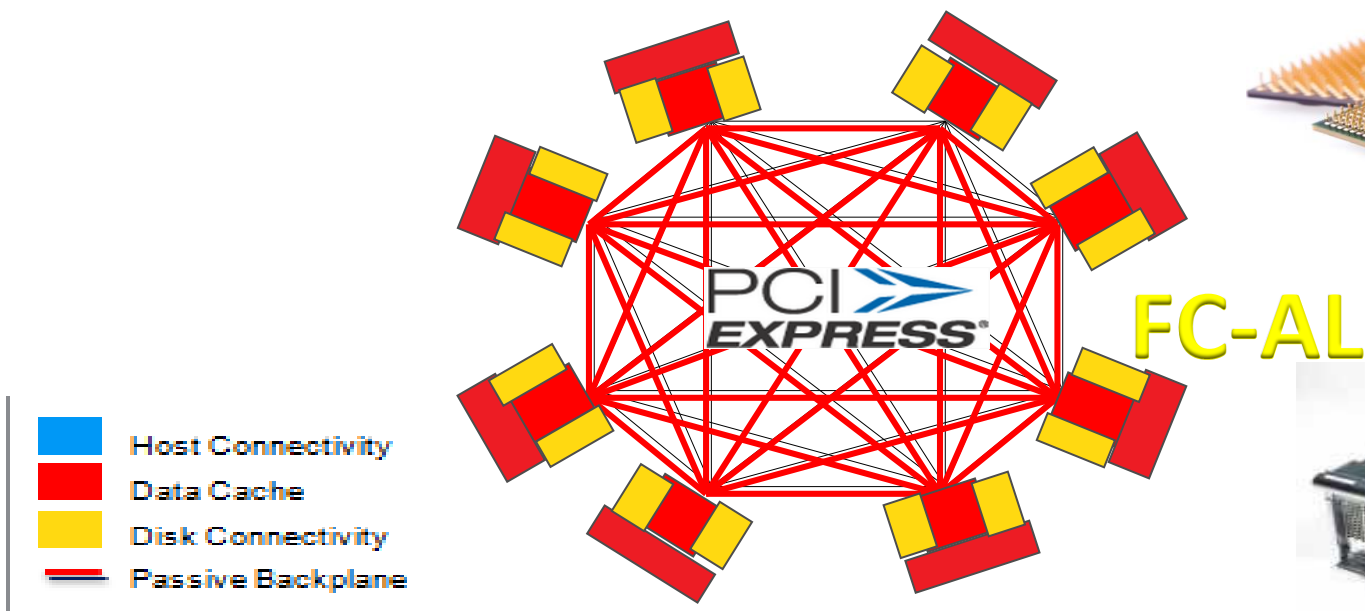
# Virtual Matrix(VMAX)架构



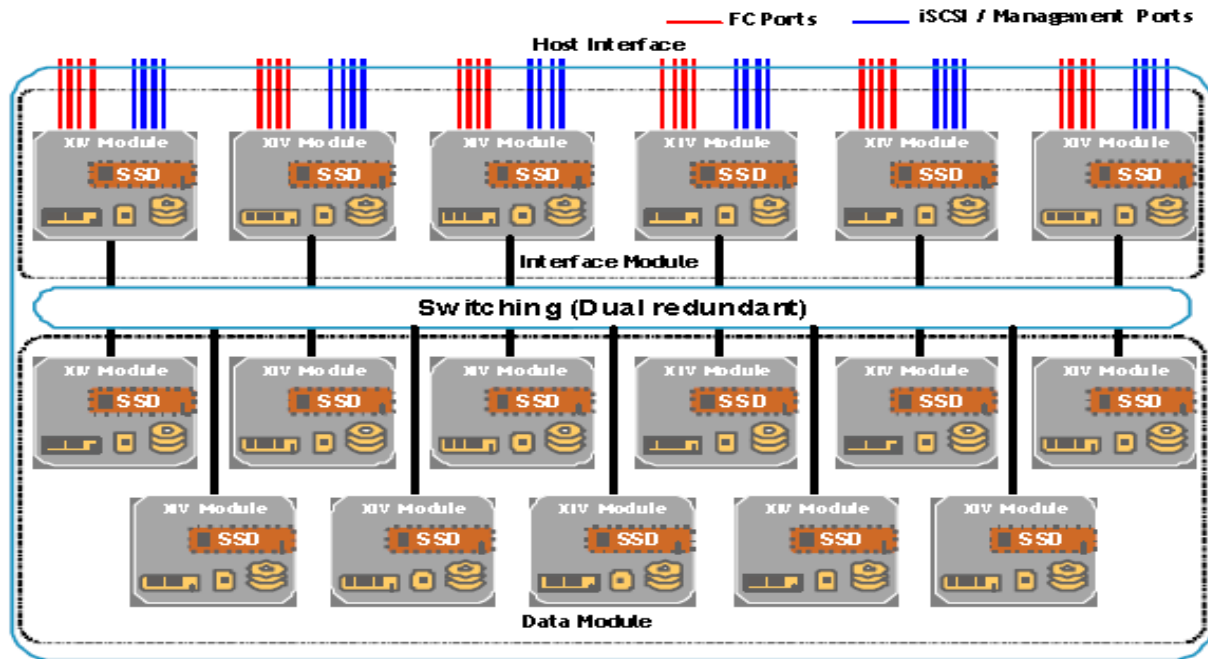
# HDS VSP最大配置



# HP 3PAR 硬件拓扑



# IBM XIV 硬件结构



*Interface and data modules are connected to each other through an internal switching network*



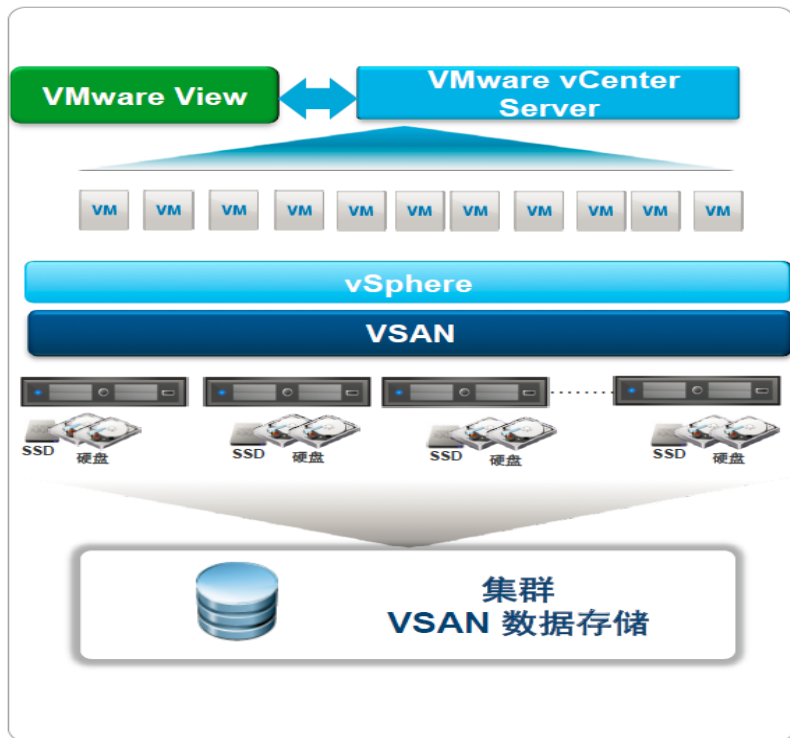




# 内容

- 传统存储硬件架构
- **Server SAN 架构（SDS）**
- SANDisk全闪存存储

# VMWare Virtual SAN



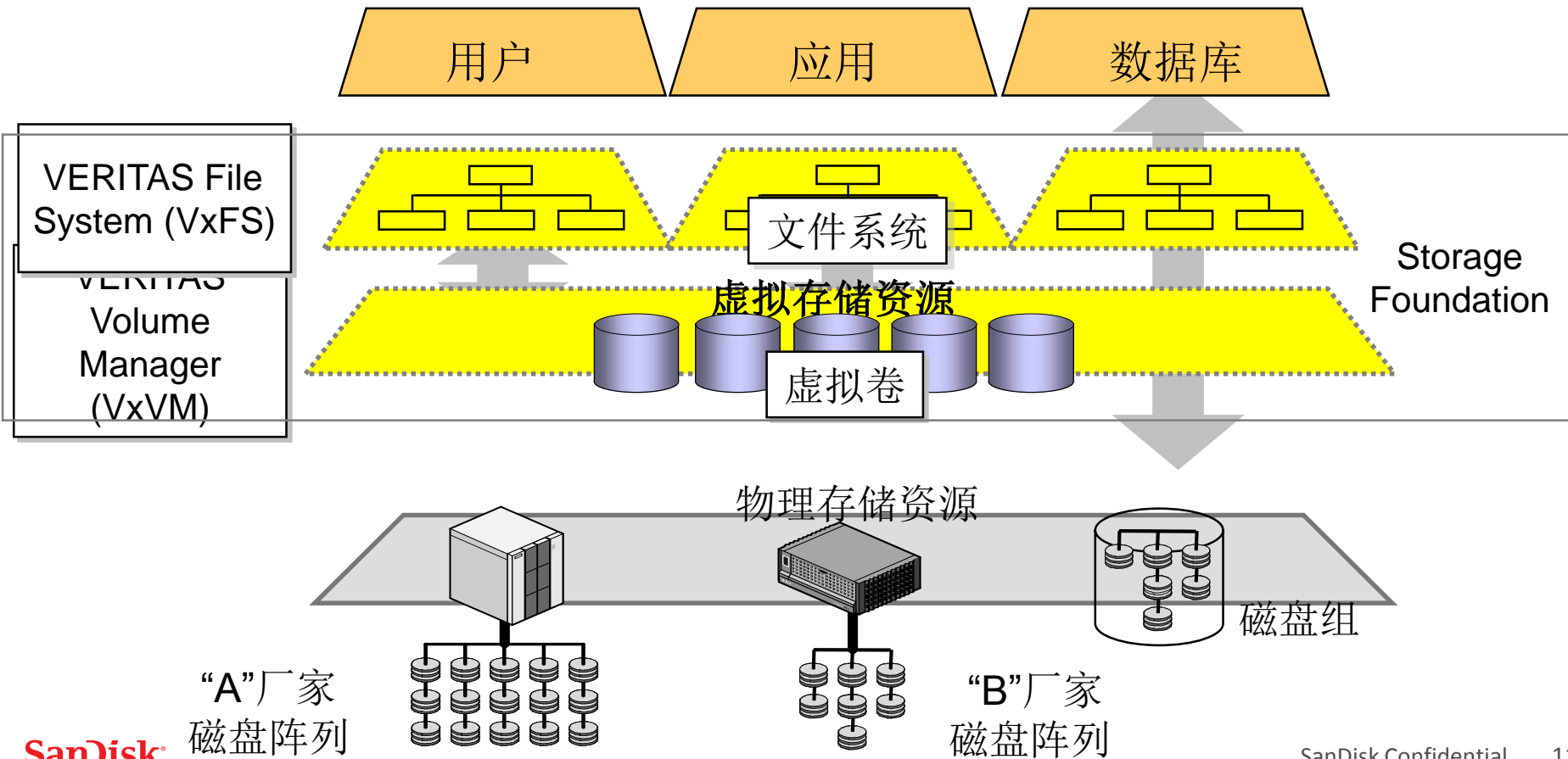
## 概述

- Virtual SAN 将来自多个服务器的固态硬盘和硬盘组成集群，以创建共享存储
- 重新定义虚拟化管理程序以建立计算和存储集群
- 针对自我调节、以虚拟机为中心的存储进行基于策略的管理
- 内置固态硬盘缓存的横向扩展体系结构

## 优势

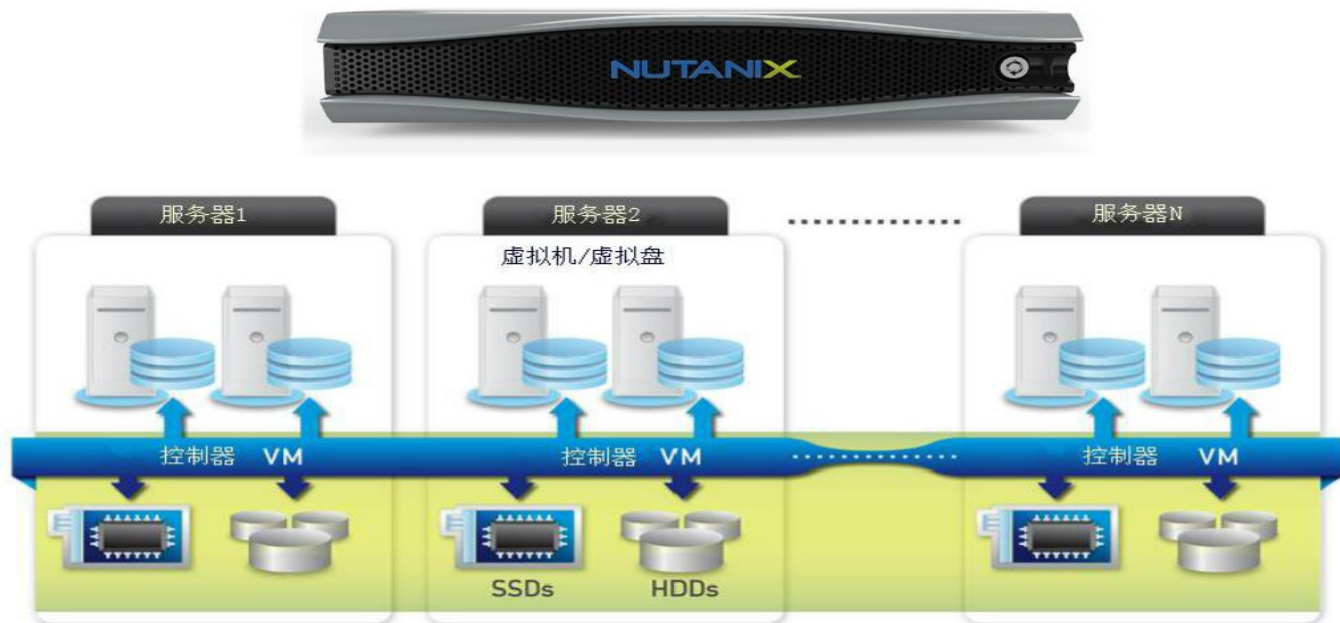
- 简单 – 专为虚拟机设计的存储
- 速度快、恢复能力强、可动态变化
- 能在提高相同用户体验的前提下显著降低 TCO
- 可以从小规模开始，线性扩展性能、容量和成本

# Symantec Storage Foundation



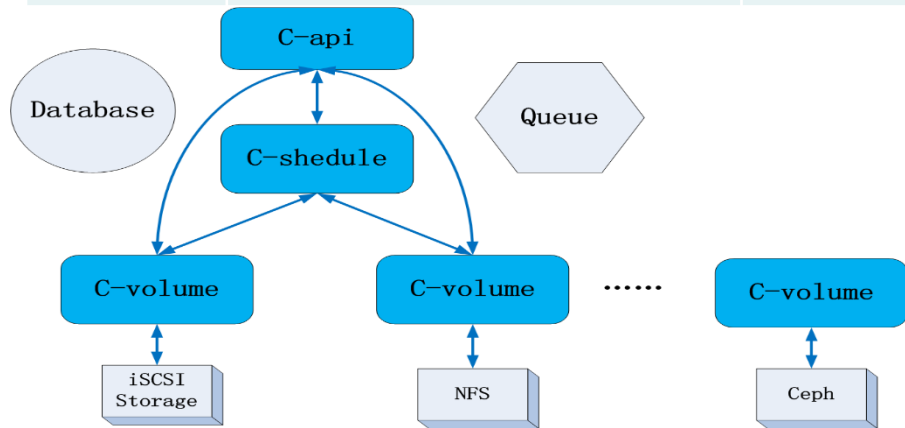
融合 2.0: 虚拟运算平台  
存储就在主机里, 简单, 高效, 灵活!

NUTANIX

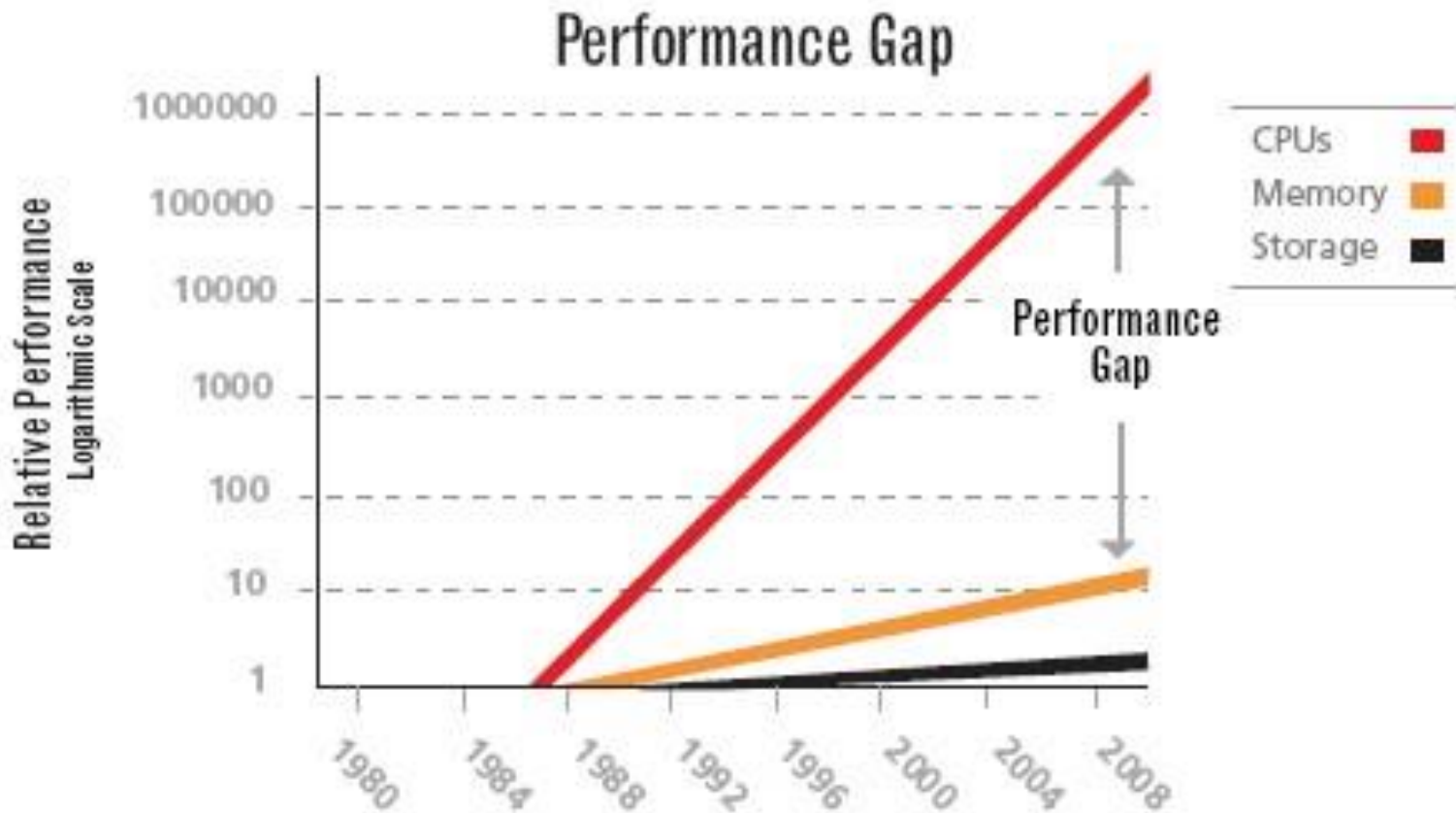


# OpenStack架构 > 组件Cinder

	Ephemeral Storage	Block Storage	Object Storage
用于	运行OS	给虚拟机增加额外的持久存储	存储各种类型的文件
访问方式	文件系统(Qemu支持的接口)	块设备 ( 比如 /dev/sdc )	REST API
如何访问	VM	VM	anywhere
管理方式	OpenStack Compute(Nova)	OpenStack Block Storage(Cinder)	OpenStack Object Storage(Swift)
数据持久性	VM被删除	volume被用户删除	对象被用户删除
大小	由管理员配置flavors	由用户指定	由管理员设置
用例	10GB root disk, 30GB second disk	1TB disk	10GB网盘



- iSCSI
- Fibre Channel
- SAN
- EMC VNX/VMAX
- NetAPP
- IBM Storwize/XIV
- HUAWEI
- HP 3PAR





# 内容

- 传统存储硬件架构
- Server SAN 架构（SDS）
- **SANDisk全闪存存储**

# SANDisk软件定义全闪存存储方案

1. 软件定义全闪存存储1（OpenStack Ceph+IF100）
2. 软件定义全闪存存储2（集成软件+IF100）

# IF100(JBOF)

**容量** 512TB raw

全闪存配置的3U高设备

64x 8TB 闪存卡

## 性能

100万IOPS,

8\*6GB/s 吞吐量

99.999% 可靠性

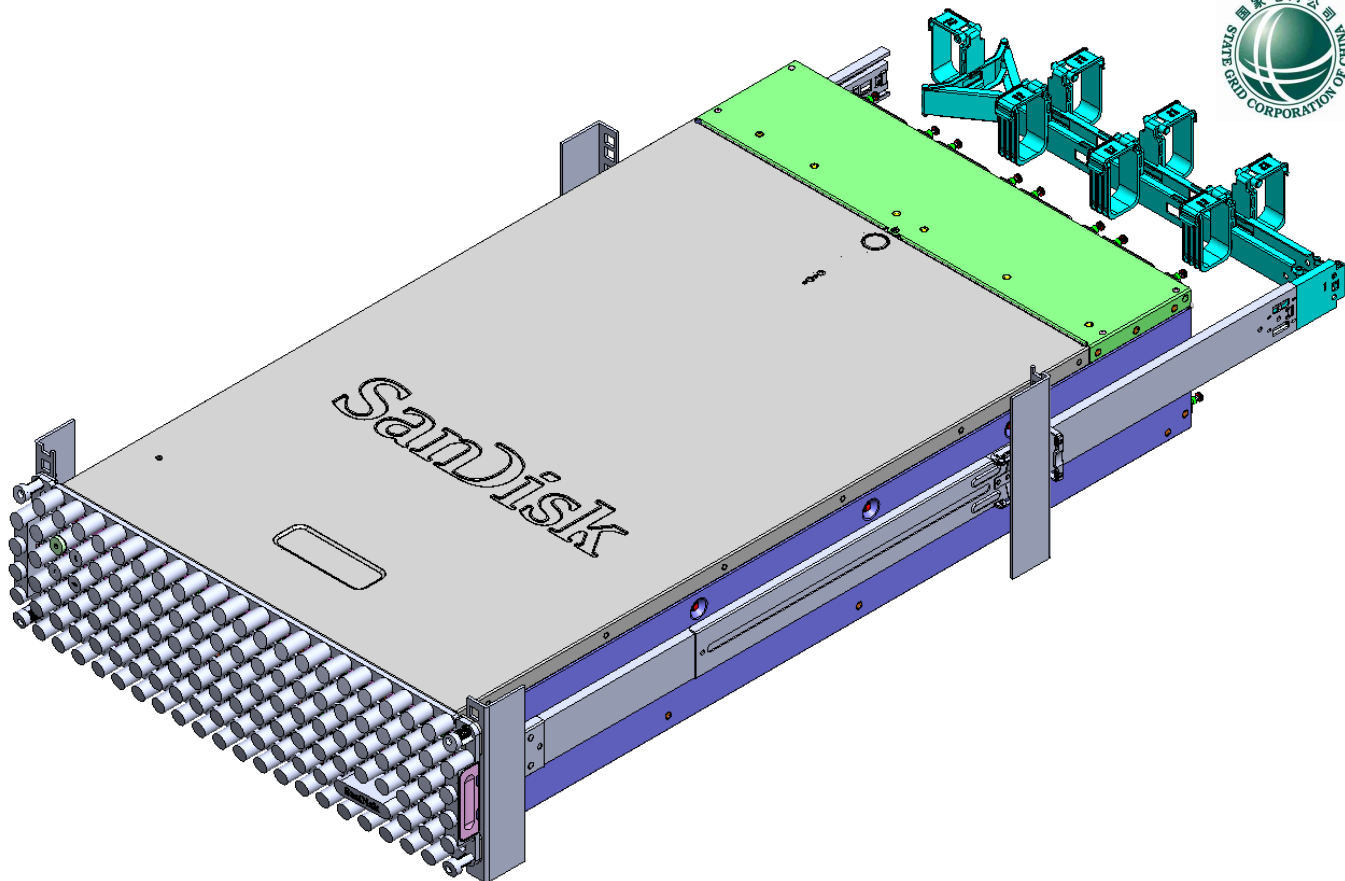
延迟 < 4 ms

**运行效率& 使用灵活性**

全热插拔设备

低功耗30W(休眠) - 750W(活动)

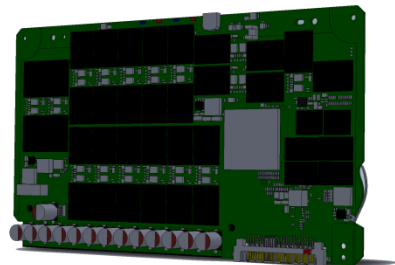
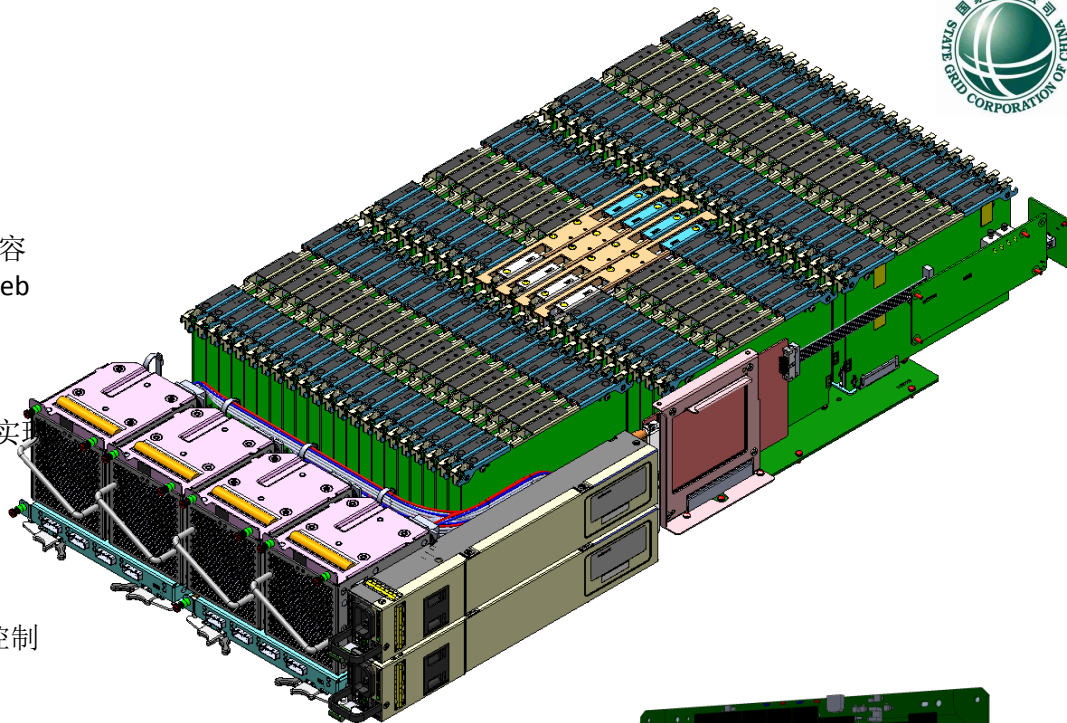
MTBF 超过150万小时



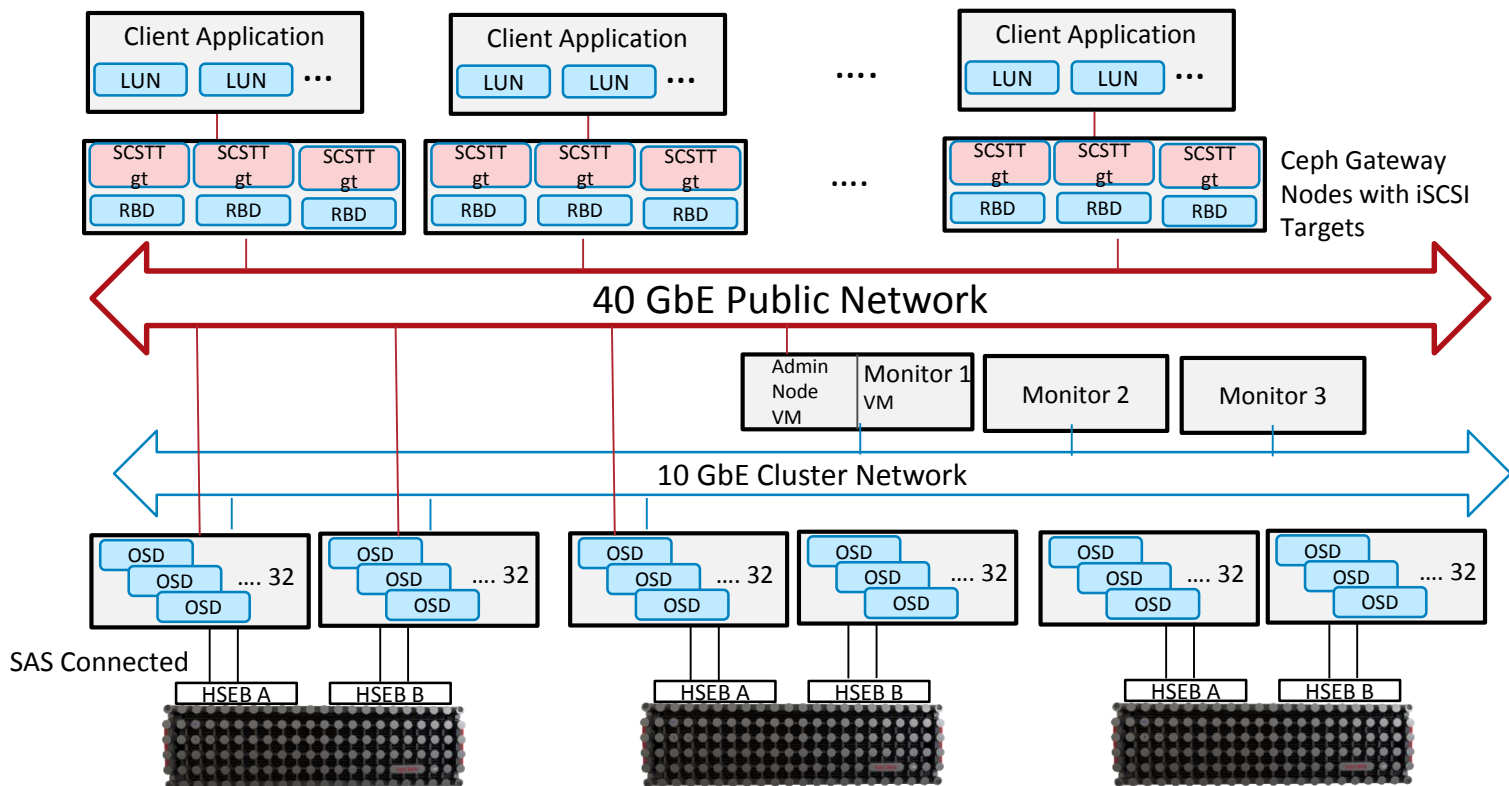
# 具体组成

## 硬件

- 高密度，低成本，3Ux30" 空间配置512TB到1PB的全闪存容量，满足多种大数据的高负载应用 (Archive, Analytics & Web 2.0 Content Depots)
- 64 SCSI 目标, 通过SAS扩展的JBOF
- 8x SAS 6Gb/s主机连接(4-lane SFF-8088), 到2015年下半年实现主机接口12Gb/s
- 通过外部服务实现风扇和驱动器卡的电源控制
- 冗余的扩展模块，电源以及热插拔的闪存卡
- 可调节的速度，耐久度和比特误码率以达到最低的成本控制
- 驱动器闪存卡(Ice Chip)
  - 8TB 1Y MLC
  - 读吞吐量: >100MB/s 和读IOPS: > 20K IOPS
  - 随机读/写 (4K块) 90/10 > 15K IOPS



# 方案1: OpenStack Ceph+IF100



## 网关节点

- 2 per Ceph Node, 3 RBDs per GW node
  - 1 Socket x 10 cores
  - 32GB RAM
- 可以与应用服务器公用一台服务器

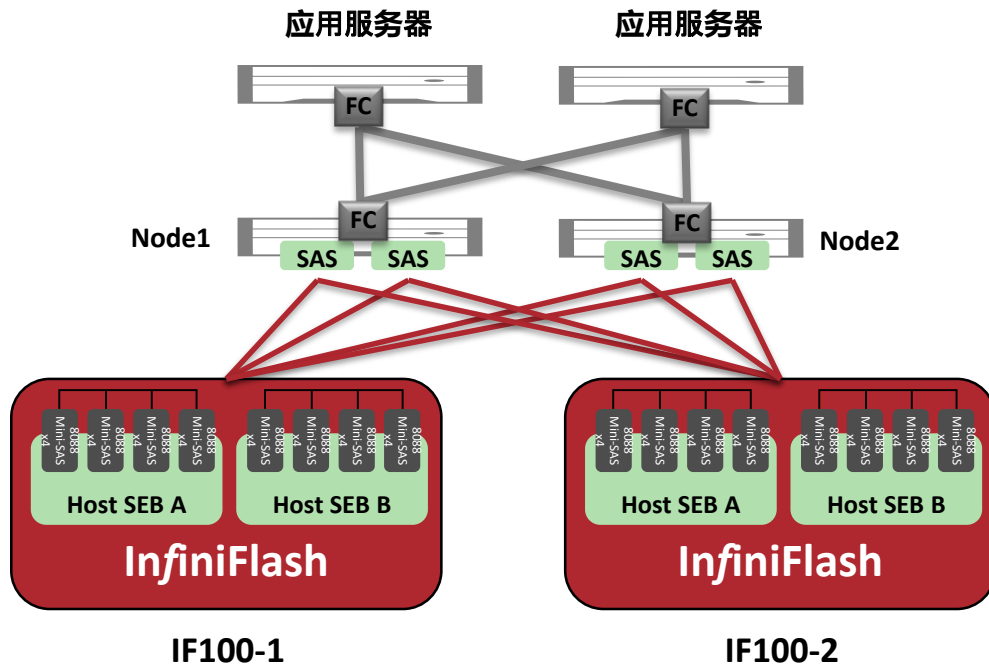
## 管理节点

- 3 Monitors (min)
- 1 Socket x 3 cores
- 32GB RAM

## InfiniFlash Ceph节点

- 2 per InfiniFlash
- 2 Sockets x 10 cores
- 64GB RAM

# 方案2:集成软件+IF100





# 方案2: 集成软件+IF100

- **NAS 和 SAN**
  - NFS, CIFS, FTP, WebDAV, FC, iSCSI
- **可扩展的容量和性能**
  - 1TB 数据缓存 (Cache)
  - 性能基准测试, 单节点吞吐量超过2.5GB/sec
- **无限量的快照&克隆数量**
- **HA 高可用集群**
  - 消除单点风险
  - 活动/活动 or 活动/被动 模式
- **块和文件复制**
  - 简单的DR配置
  - 同步or 异步, 同构或者异构平台
- **软件 RAID, 可支持3份校验数据的保护**
- **VM 集成**
  - VMware, XEN, & Hyper-V
- **在线的压缩和去重**
- **端到端的数据一致性保护**
- **瘦供给**
- **无中断升级**
  - 软件微码升级 (with HA Cluster)
  - 在线容量扩展
- **简单的用户界面**
- **可编程的 API**
- **Cloud-storage ready**

# 方案2: 集成软件+IF100

## ▪ 统一存储服务

- 文件: 10GbE NFSv3 / NFS v4 / CIFS / SMB 2.1
- 块: 8Gbps FC / 10GbE iSCSI

## ▪ 高可用& 容灾

- 2 节点 active / active 集群
- 远距离定期异步数据复制

## ▪ 数据保护

- 基于软件的 RAID 1, RAID N+1, RAID N+2 and RAID N+3
- 所有的RAID操作建立在实际使用的数据块, 缩短重构时间

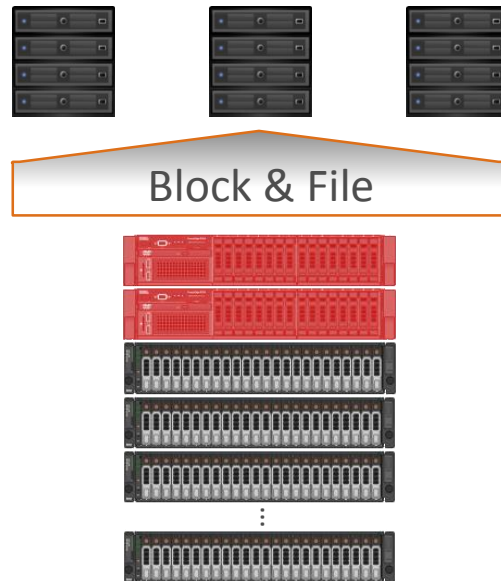
## ▪ 容量优化

- 瘦供给
- 在线数据除重

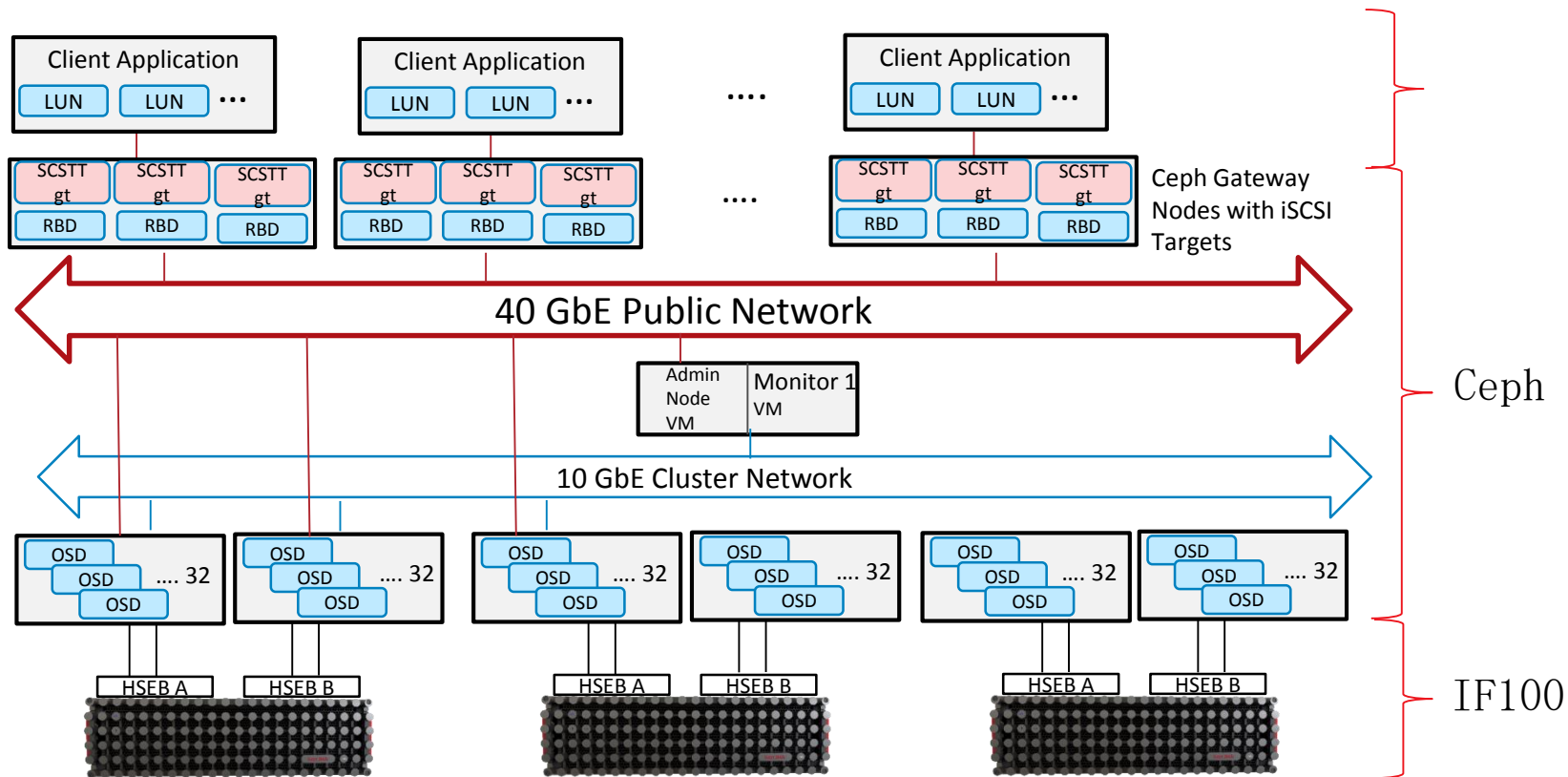
## ▪ 扩展性

- 从10TB到2PB (裸容量)
- 无限的文件系统容量

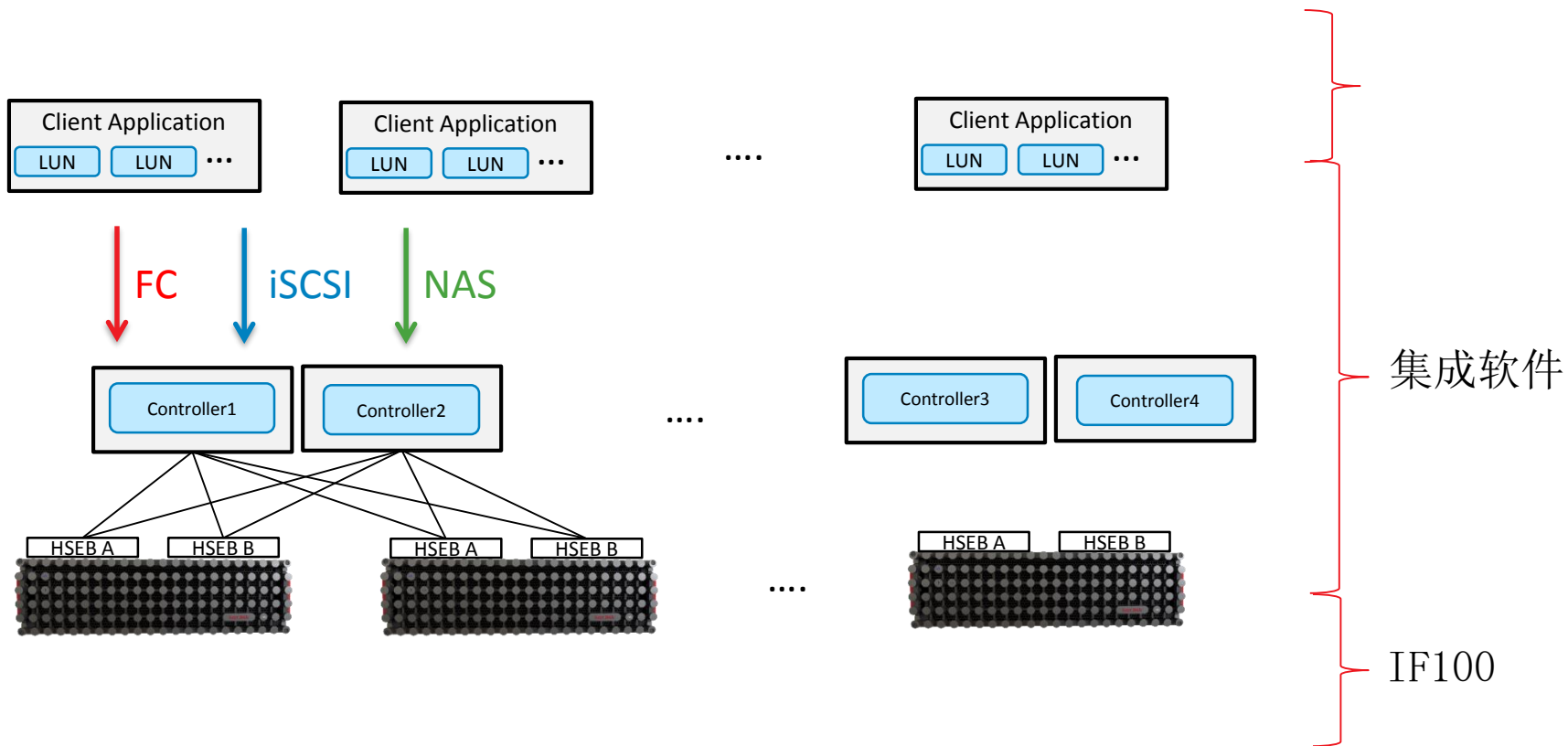
无限的快照和克隆



# 目前使用方式1: Ceph+IF100(横向扩展)



# 目前使用方式2：集成软件+IF100



# SANDisk全闪存存储优势

## ■ 功能完善，扩展性好

- 扩展性好 （ 1PB ） 满足大部分工作负载以及性能要求
- 全协议支持 （ FC + iSCSI + NFS + CIFS 同时提供 ）
- 企业级功能 （ 无限量快照、克隆；在线压缩除重；异步数据复制（同构，异构存储）
- 网关特性，整合原有小旧存储
- 横向扩展

## ■ X86服务器平台

- 最优秀的硬件平台（ 高性能、完善管理功能、符合行业标准）
- 对客户企业提供企业级支持

## ■ 客户定义的存储

- 数据中心的未来
- 灵活性、可扩展性、标准化、去硬件化

# SanDisk的优势

## 垂直整合: 从头到尾的完整生态链



### 技术领先

- 11年Nand flash技术积淀
- 业界第一的 X3 19nm
- 专业的芯片和固件针对闪存管理



### 300mm NAND Flash 制造

- 三家与Toshiba合资的Fab工厂
- 第二代19nm(1Y)产品的生产



### 组装和分销

- 世界级的工厂@上海
- 3百万片die 每天
- 缩短产品的生产周期



# Thank You

A hand holding a white marker is shown on the right side of the image, drawing a long, slightly curved horizontal line across the red background. The line starts under the 'T' of 'Thank' and extends past the 'u' of 'You'.

©2014 SanDisk Corporation. SanDisk is a trademark of SanDisk Corporation, registered in the United States and other Countries. Other brand names mentioned herein are for identification purposes only and may be the trademarks of their respective holder(s).

\*1GB=1,000,000,000 bytes. Actual user capacity less.

# SanDisk®