



数据驱动 存储创新


2017企业存储技术峰会·北京站

2017 ENTERPRISE STORAGE TECHNOLOGY SUMMIT BEIJING

2017.01.18 · 北京东方美爵酒店

主办方

日知录技术社区



浅谈超融合架构核心技术

张凯 SmartX CTO

Agenda

- IT 基础设施发展过程
- 超融合架构的核心技术

自我介绍

- 毕业于清华大学计算机系高性能计算所
- 2013 年和徐文豪，王弘毅，共同创办 SmartX，任 CTO 职位
- 带领团队完成超融合产品从 0 到 1 的设计和研发
- SmartX 超融合软件已在数千台服务器上稳定运行

IT 基础设施发展历程

服务器
孤岛

集中式
存储

服务器
虚拟化

超融合
架构

虚拟化已成 IT 基础设施“标配”，存储成为瓶颈

业务发展趋势



传统企业互联网+



传统企业物联网转型

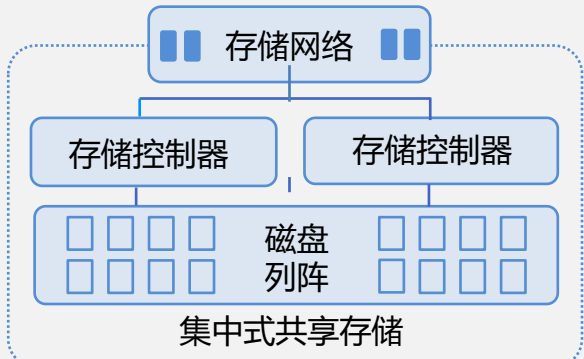


虚拟化、容器技术等普及



业务数据指数级增长

传统架构



多套系统
运维管理复杂



扩展困难、难调配



I/O性能瓶颈
无法发挥SSD特性



专用设备、成本高

企业需要：

更敏捷

更易扩展

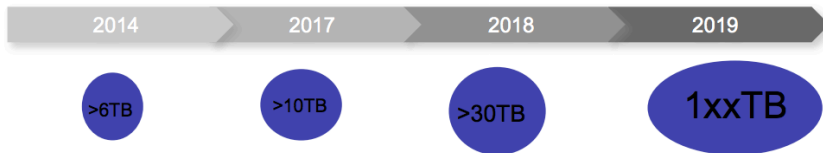
更经济

IT 基础架构

超融合架构--超融合架构的核心技术

- 定位于 **核心业务，核心系统**
- 节省硬件资源
- 可扩展性
- 可靠性
- 生态系统

超融合架构--硬件趋势



- CPU
 - 多核

- SSD
 - 存储密度增大
 - 单位价格降低
 - NVMe 普及

- 网络
 - 200 Gb 以太网

超融合架构--计算资源使用效率

用户态

用户态程序

内核态

任务调度

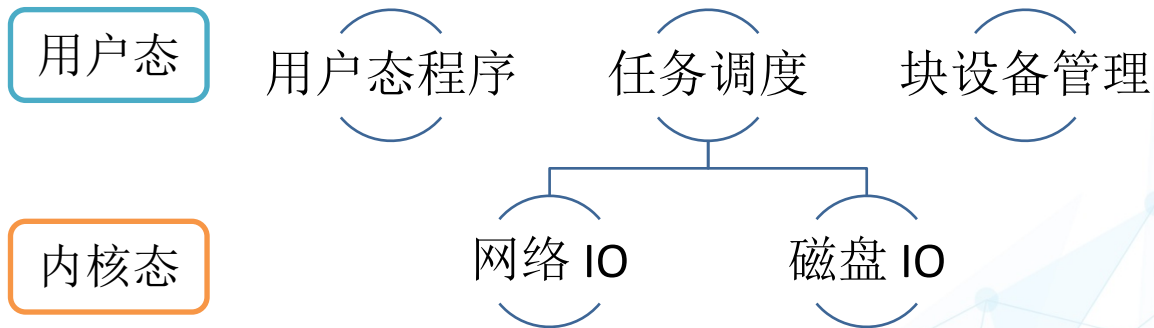
网络 IO

磁盘 IO

文件系统

PageCache

超融合架构--SmartX 如何节省硬件资源



超融合架构--SmartX 如何节省硬件资源

- 协程
 - 轻量级 “线程”
 - 切换只需要 10ns 开销
 - 简化异步程序，用顺序的方式写异步程序
 - 优化内存管理

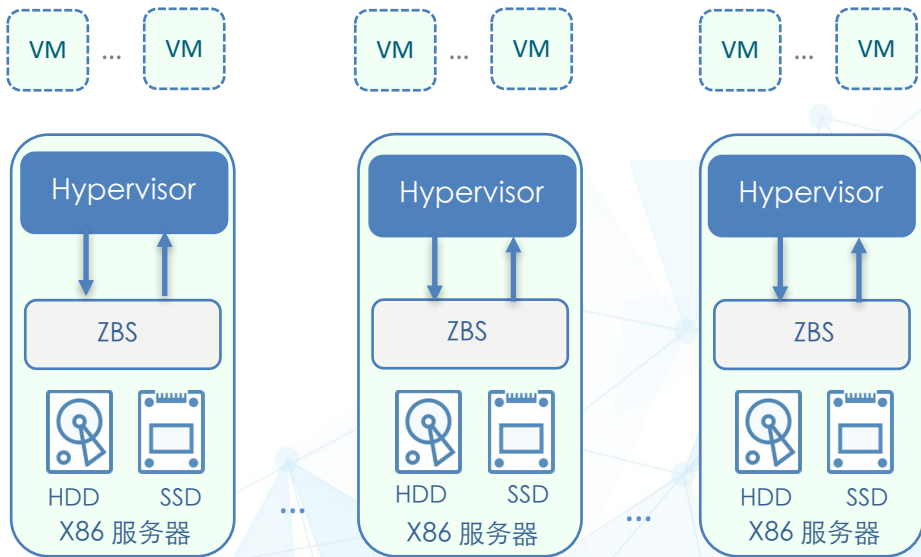
```
Status DoIO(Client* client) {  
    Context ctx;  
    RETURN_IF_ERROR(RecvRequestFromClient(client, &ctx));  
    RETURN_IF_ERROR(GetReplicasFromMetaServer(&ctx));  
    FOREACH_REPLICA(ctx.replicas, replica) {  
        RETURN_IF_ERROR(SendRequestToReplica(&ctx, replica));  
    }  
    return Status::OK;  
}
```

超融合架构--SmartX 如何节省硬件资源

- 用户态直接管理块设备
 - 避免文件系统开销
 - 特殊的缓存系统，利用内存和 SSD 对 HDD 进行加速
 - 通过日志系统，保证数据写入原子性
 - 对于顺序 IO，绕过 SSD，直接写入 HDD

超融合架构--SmartX 如何节省硬件资源

- 本地化 IO 操作，
避免网络 IO

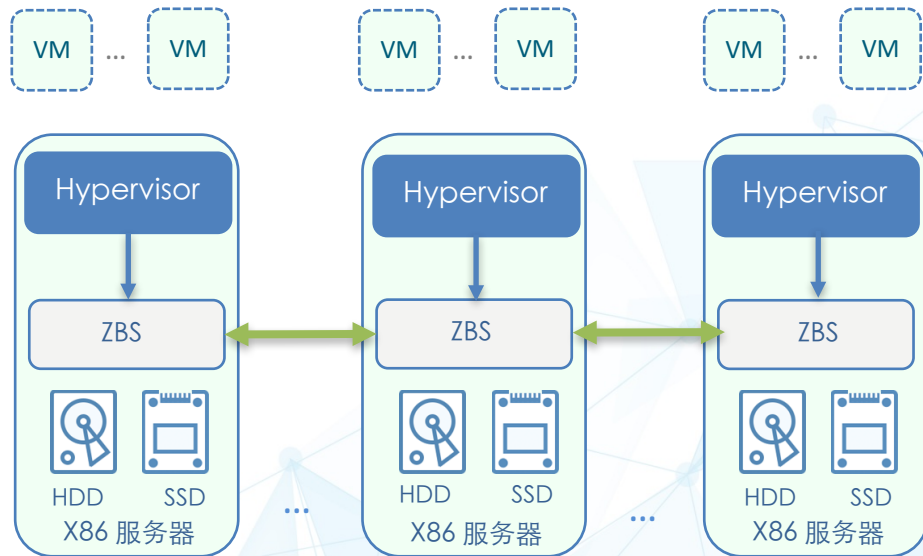


超融合架构--横向扩展性

- 横向扩展 vs 纵向扩展
 - 超融合天然横向扩展
 - 降低对硬件设备需求
 - 单业务对存储性能要求不高

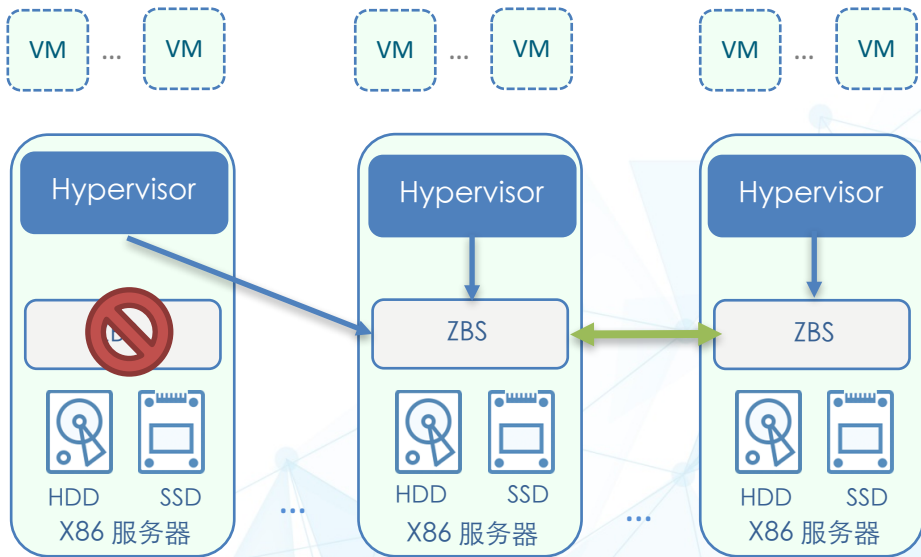
超融合架构--SmartX 如何实现横向扩展

- 无状态多控制器
 - 任何一个节点都可以提供数据访问服务



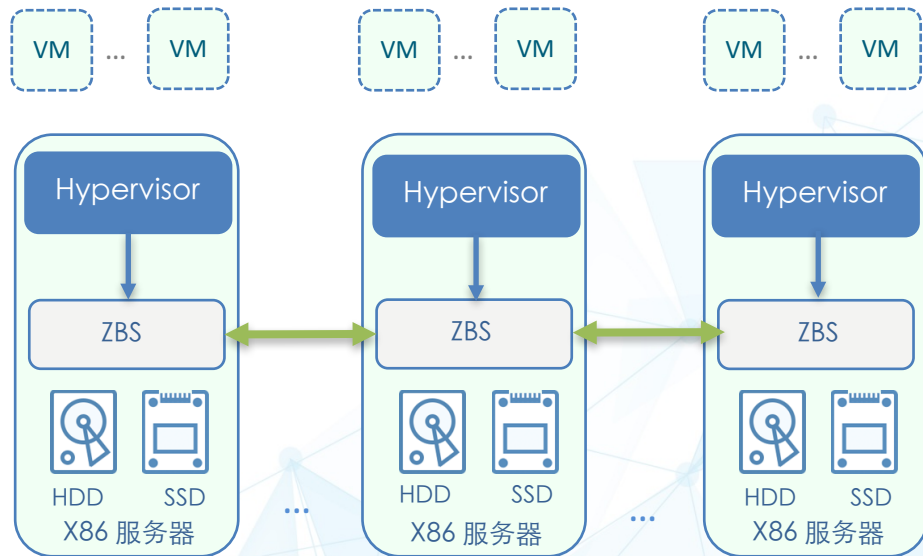
超融合架构--SmartX 如何实现横向扩展

- 无状态多控制器
 - 当遇到服务异常，或网络异常，可以通过其他节点的访问存储服务



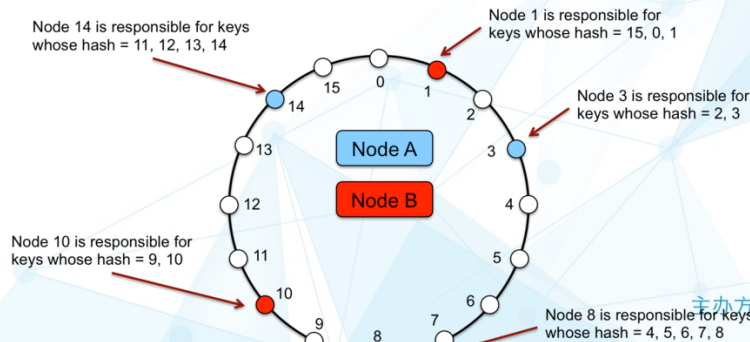
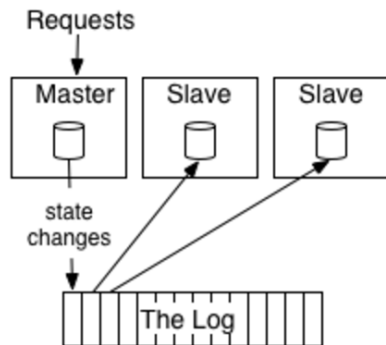
超融合架构--SmartX 如何实现横向扩展

- 无状态多控制器
 - 服务恢复正常，恢复正常访问



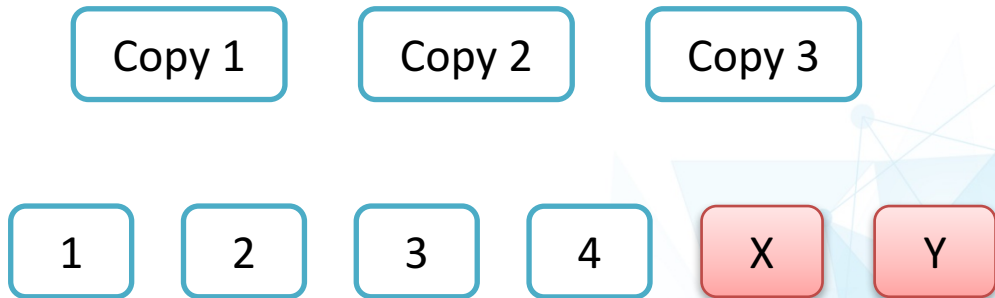
超融合架构--SmartX 如何实现横向扩展

- SmartX 使用集中式元数据管理方式
 - 通过 log replication 保证元数据可靠
- 一致性哈希的问题
 - 无法精确控数据位置
 - 添加删除节点复杂



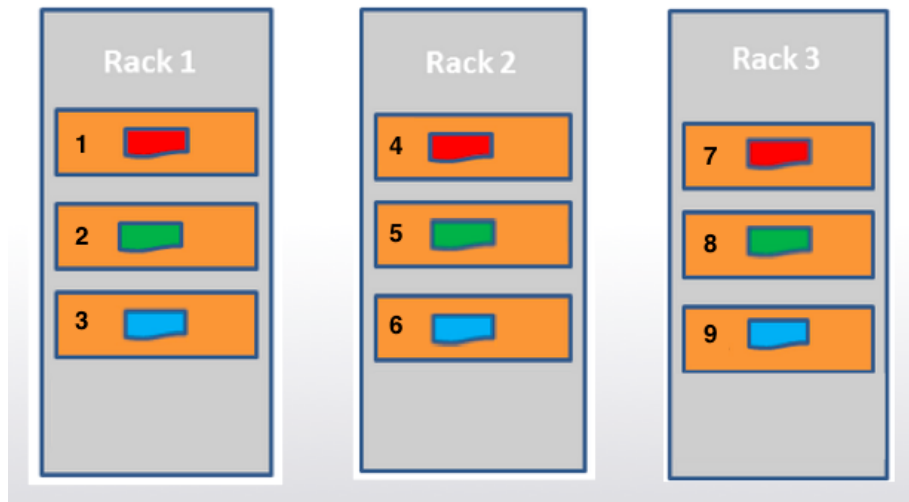
超融合架构--可靠性

- 副本技术 vs EC
 - 副本的写入性能更好
 - 数据恢复代价更小
 - 存储密度不断提高
 - 成本不断降低



超融合架构--可靠性

- 机架感知



超融合架构--可靠性

- 快照
 - 预防软件故障，人为误操作
- checksum 校验
 - 预防磁盘硬件 silence corruption
- 容灾备份
 - 预防站点故障

超融合架构--生态系统

- x86 硬件服务器
 - 兼容主流硬件厂商服务器



管理界面

SmartX Halo 软件

x86 服务器



inspur 浪潮

lenovo

IBM

中科曙光
Sugon

超融合架构--生态系统

- 兼容主流云平台



欢迎大家加入 SmartX



谢谢！

超融合架构--SmartX 如何实现横向扩展

- 无状态多控制器
 - 当遇到服务异常，或网络异常，可以通过其他节点的访问存储服务

