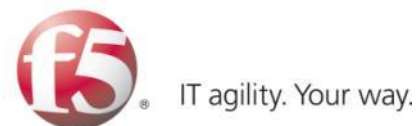


文件存储虚拟化

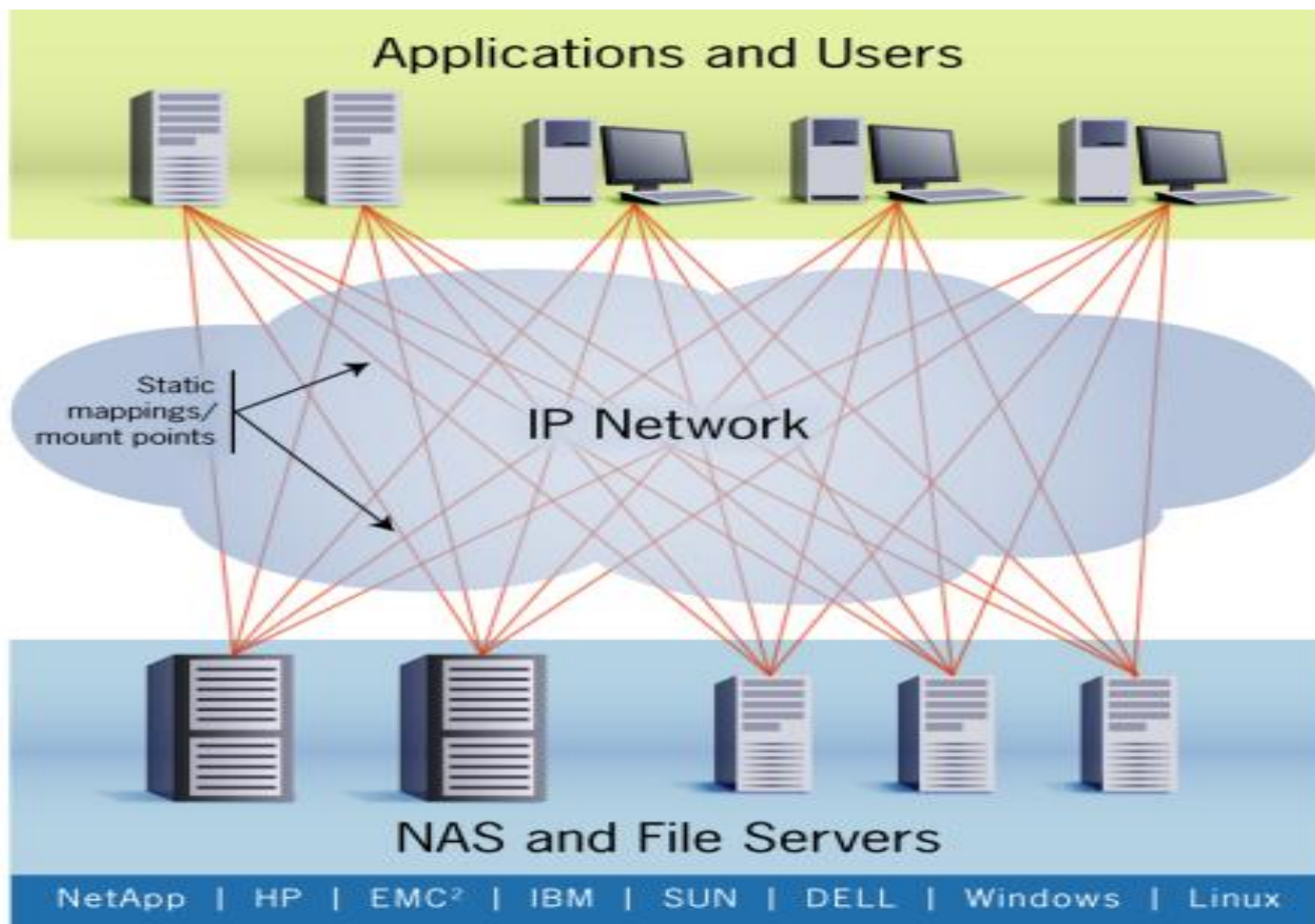
杨明非
F5北方区技术经理



Agenda

- 今天的文件存储系统存在的问题
- 如何解决问题，基于软件的分布式文件系统 and 硬件文件存储虚拟化的区别
- 构建高性能稳定可靠的虚拟化文件存储系统及案例分析

存储的架构混乱



NAS存储已经成为系统的瓶颈

- 1984年制定的NFS协议标准已经不能满足今天的性能要求
- 即使在一个10Gbps的网络上，现有的NFS 3.0运行起来也就和在一个1Gbps网络上的性能差不多
- NFS 4.1(pNFS)还是在一个未来中
- 当NFS Mount客户端点超过16个的时候，性能将急剧下降
- 高端的NAS服务器每秒钟能执行10万次I/O操作，但其价格已经很难接受

用磁带存储还是用磁盘存储

- HP LTO4 Ultrium磁带(C7974A) 800GB 500元
 - IBM LTO4(46C5359) 800GB 680元
 - 希捷 (Seagate) 1TB ST31000525SV 7200转32M SATAII 监控级硬盘 719元
 - 西部数据 (Western Digital) 1TB WD1000FYPS 7200转16M SATAII绿色环保企业级盒装硬盘7*24 880元
 - NetGear的ReadyNAS, 24TB, 10000美元
-
- 磁带存储的优点是容量大，缺点是存取速度慢
 - 硬盘存储的优点是速度快，缺点是。。。

文件存储空间需求的快速增长

- 大多数网站的文件存储空间是以倍数以上的每年增长
- 新的富媒体文件如图片、视频等导致存储空间需求急剧增大
- 法规遵从使企业必须保留多年的数据随时备查
- Web 2.0的兴起导致了文件存储空间需求呈指数方式上升

如何管理不同厂商的产品



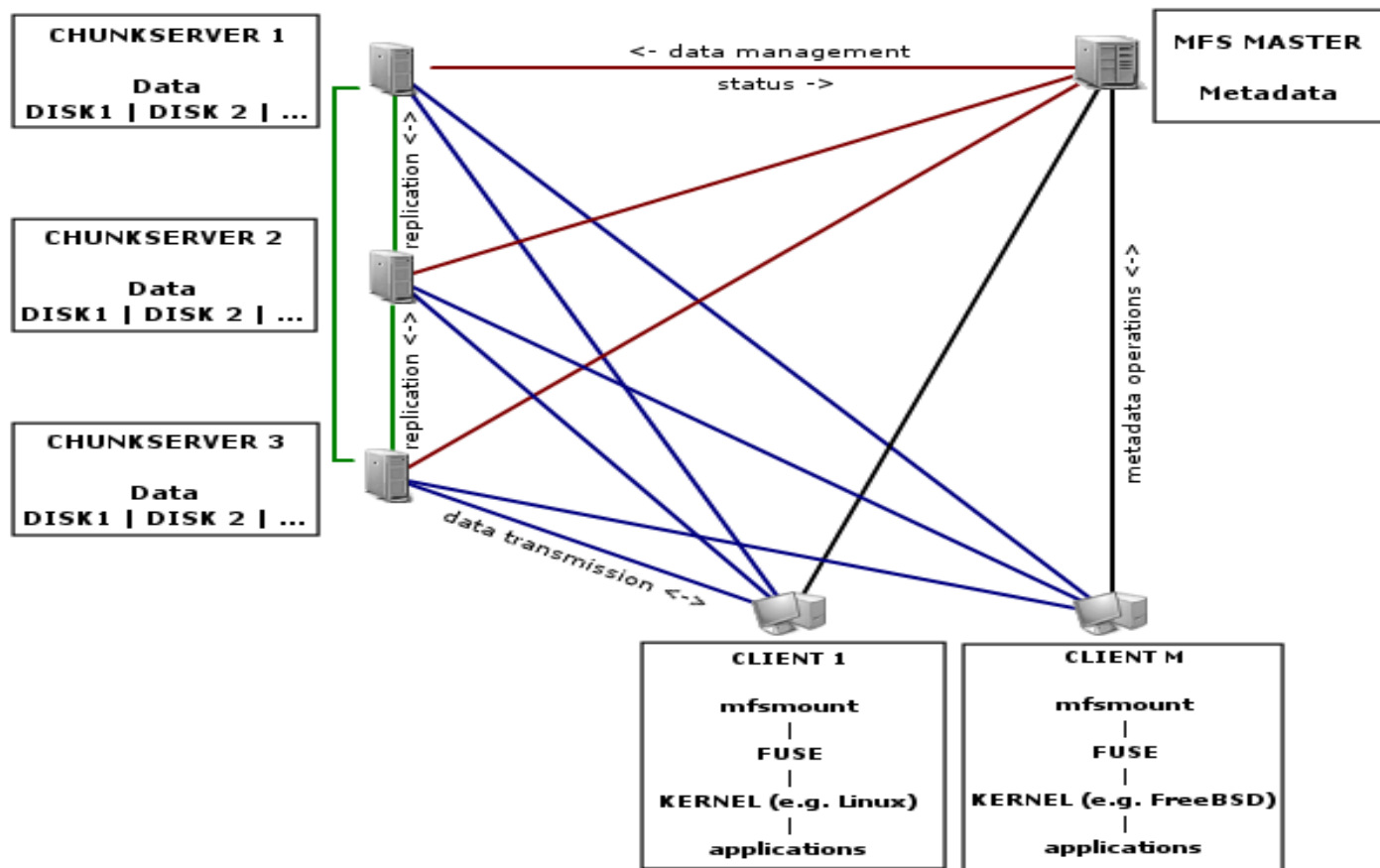
- 企业在发展的过程中，必然的会采购有不同厂家的存储产品
- 存储产品的特性导致每个厂家基本上都采用封闭的解决方案
- 一旦采用某一厂家的产品，将来的扩容、备份、迁移、复制等职能采用同一厂家的产品

如何省钱！！！！

- 经济危机大环境下，必然承受的压力
- 如何有效的使用现有的存储资源
- 如何避免存储设计的30%法则，提高资源的利用率？
- 在省钱的时候不要忘记了高可用和稳定

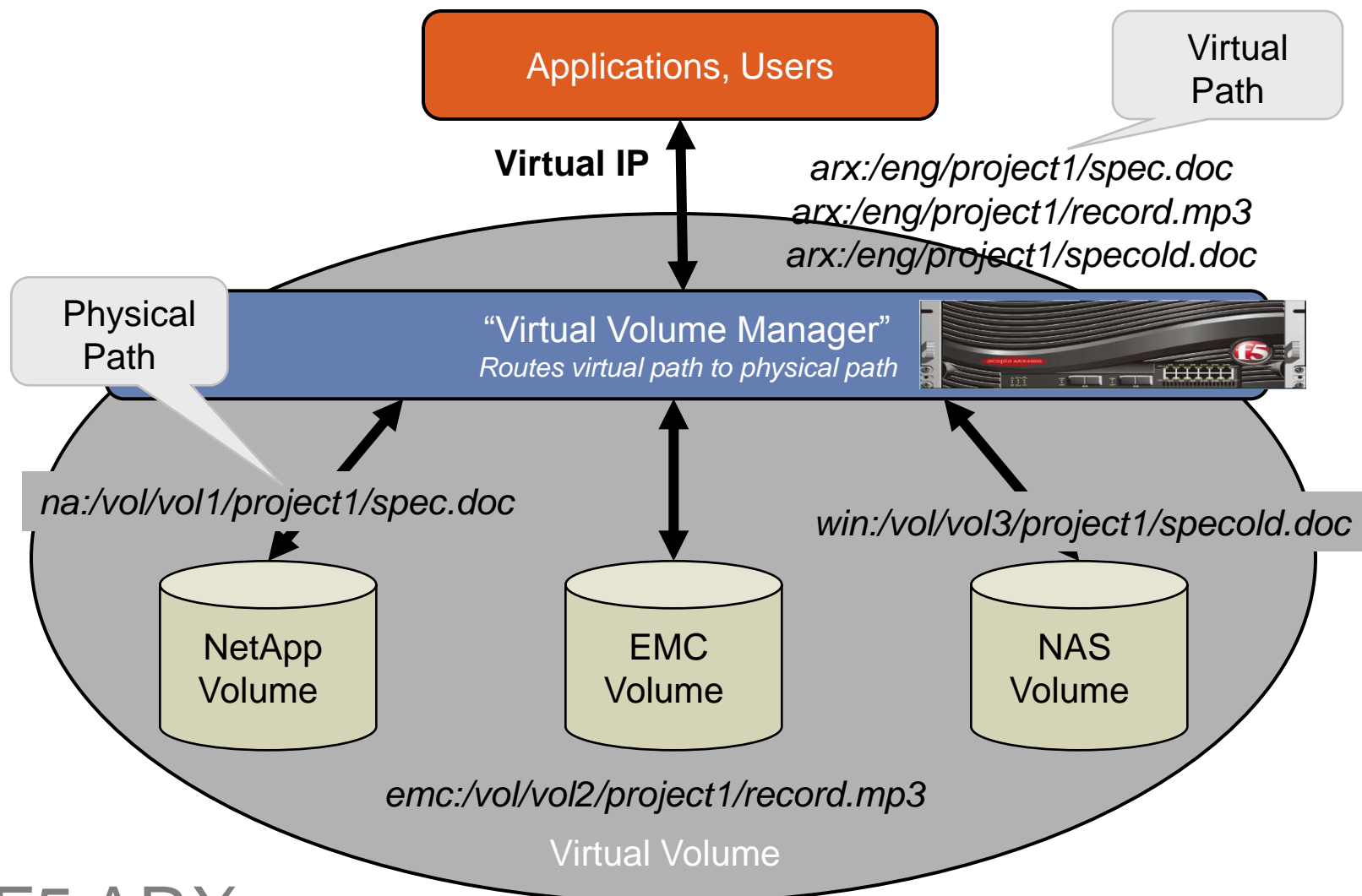
如何解决问题？

基于软件的分布式文件虚拟化存储



- lustre, hadoop, pNFS, Moosefs, GFS, DFS,

基于硬件的文件虚拟化存储

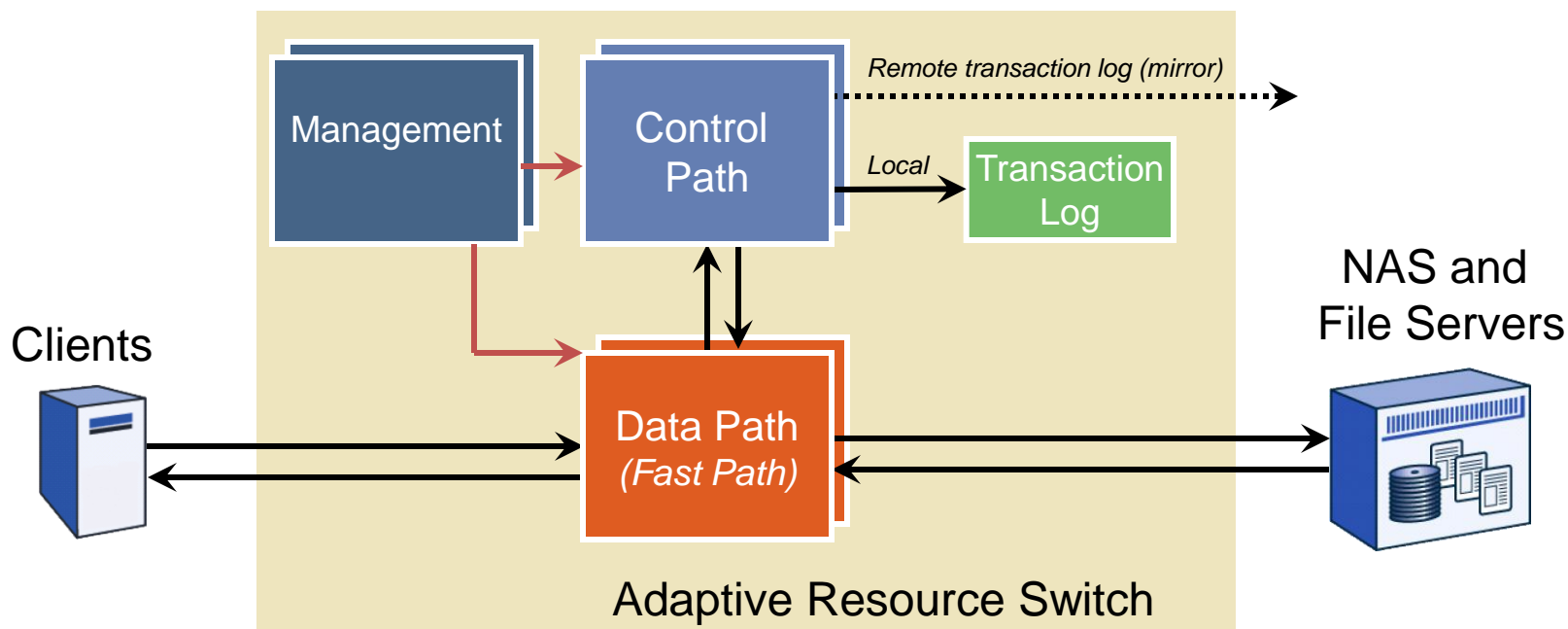


两种方式的比较

	软件分布式存储	硬件虚拟化存储
数据完整性	不完整，通常采用link方式或者文件分块方式存放	完整，对于单个文件完整存放，不采用link
数据可恢复性	一旦管理服务服务器崩溃，很难恢复数据	只需要重新扫描文件系统
性能	高，但受限于客户端代理的性能和管理服务器性能	高，但受限于硬件芯片的吞吐能力
可靠性	相对低，核心系统的可靠性取决于操作系统和服务端硬件	高，专用设计的硬件芯片稳定、可靠
技术支持	开源软件没有技术支持	厂商提供专业技术支持
功能扩展性	单一，只针对分布式存储	丰富，可实现分级存储、动态迁移、远程复制等多项功能
客户端Agent	必须通过客户端Agent截获操作系统的文件读写	不需要安装客户端软件
支持协议	只支持NFS协议	同时支持NFS和CIFS协议

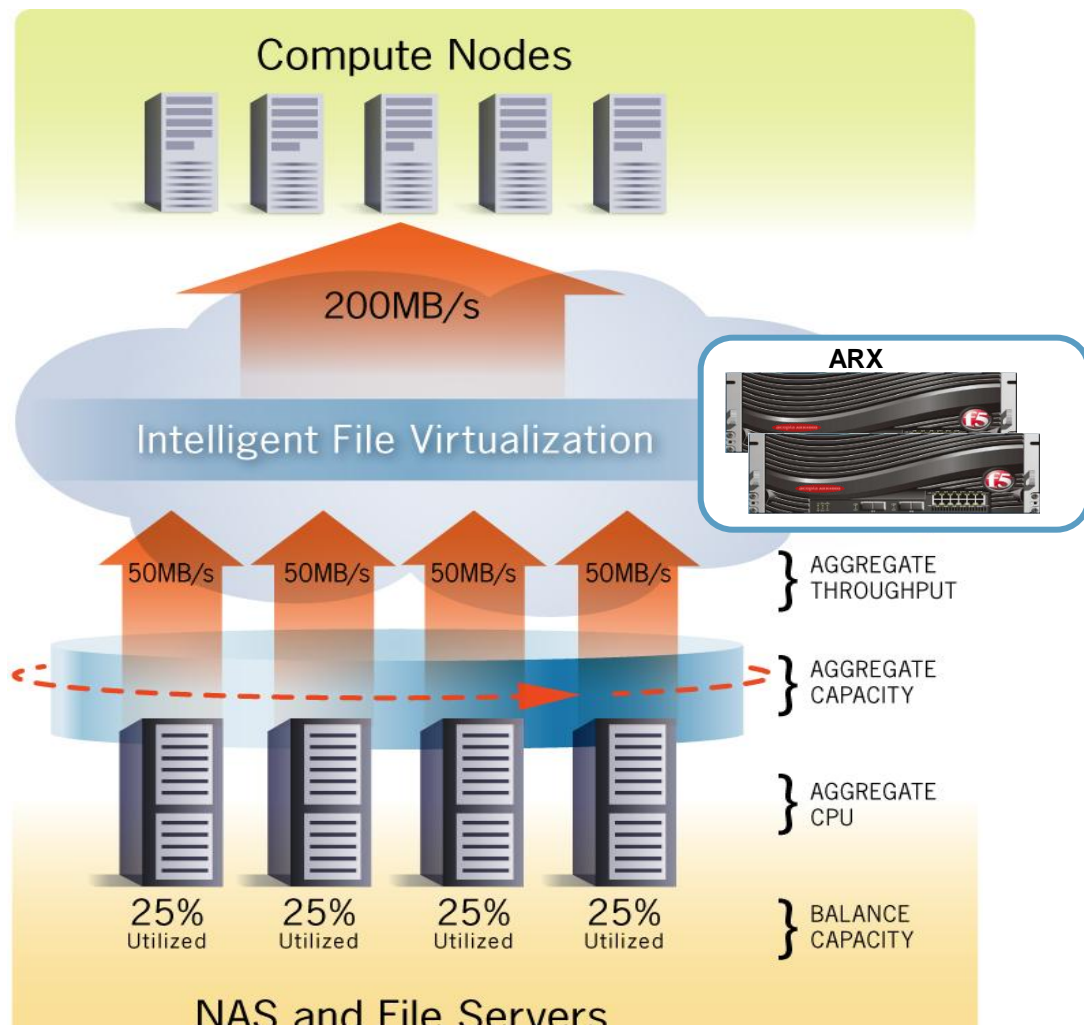
构建高性能稳定可靠的虚拟化文件存储系统

F5硬件文件存储交换机ARX架构



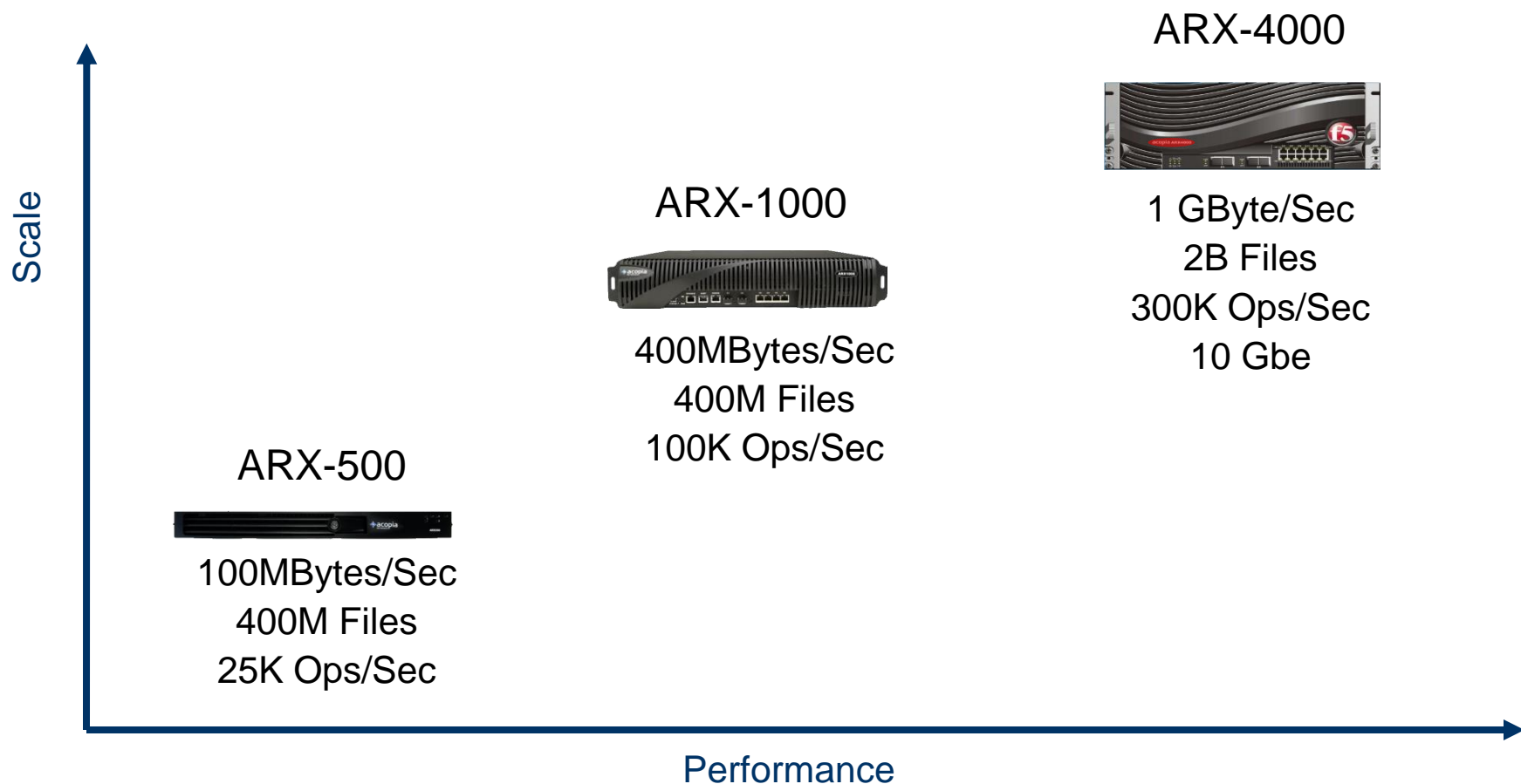
- 采用专用的硬件FPGA芯片处理数据流量
- 采用高性能多核CPU管理MetaData信息

实现负载均衡的NAS集群



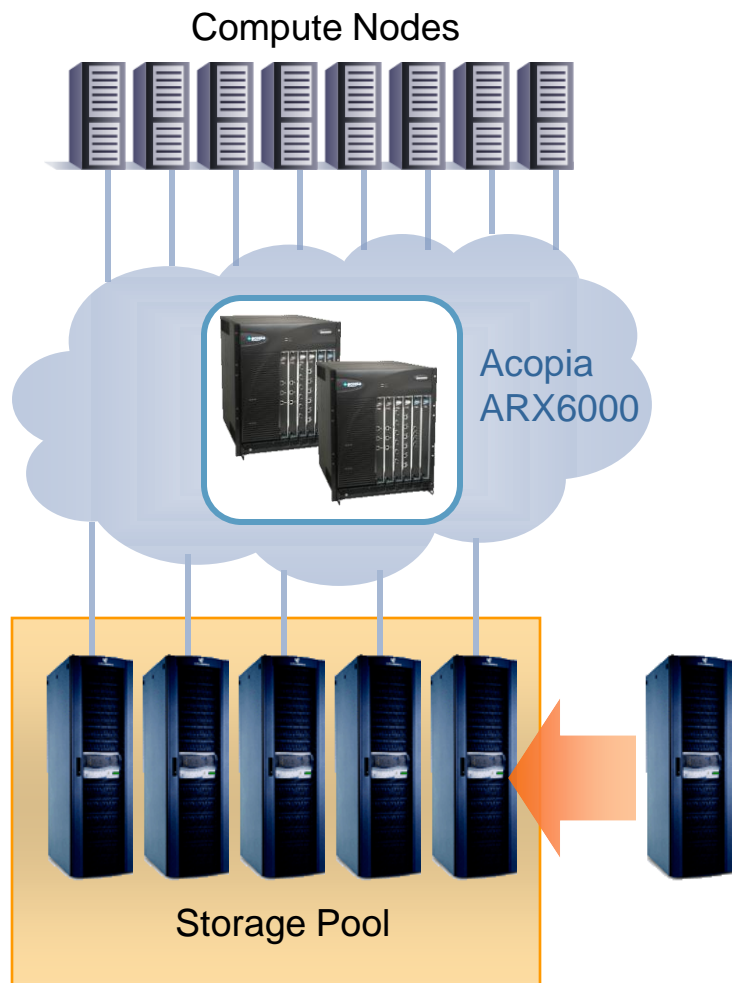
- 通过ARX，可以把多台NAS集合起来，整体对外提供更高的I/O读写性能和吞吐率
- 可以实现基于容量、带宽、NAS CPU占用率和延迟的负载均衡算法

关于性能



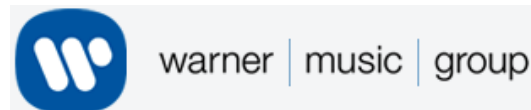
- 对比于业界目前能查到的最高性能NAS：并行I/O的主要好处是它可以提供很高的应用性能和巨大的可调整性。“BlueArc的Titan 3200的速度性能可以达到1.6GBps，I/O性能可达到380KIOPS。

文件存储动态扩展案例

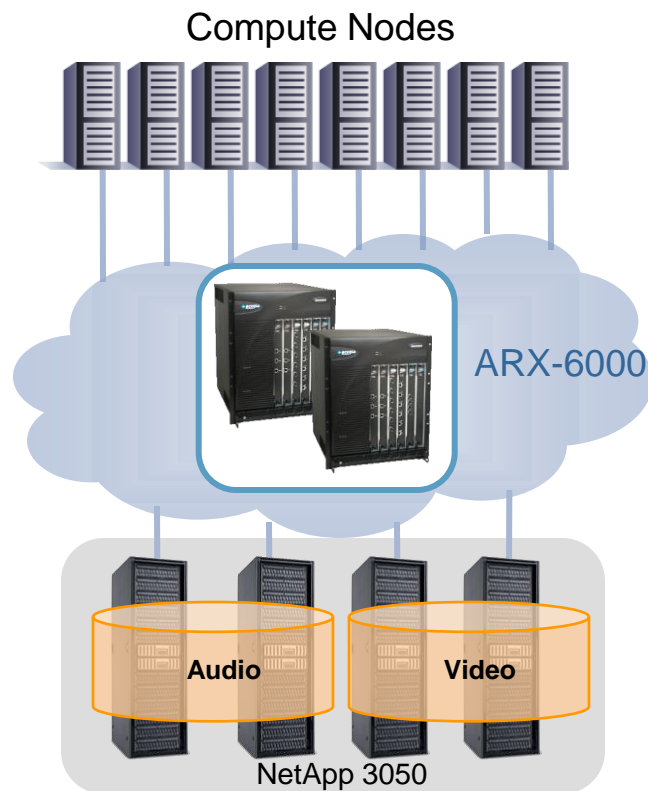


- 挑战
 - 存储容量的高速增长
 - 当增加新的存储的时候需要不停机
 - 停机成本很高
- 解决方案
 - ARX6000 集群
- 好处
 - 无间断的扩容
 - 消除了停机成本
 - 通过动态扩容和在线迁移减小运营成本OPEX

存储负载均衡案例



- 环境:** NetApp 3050s; ~150TB
- 存在的严重问题:** 架构设计限制了数字化歌曲的速度
- 原因:** 文件系统的限制了必须使用手工方式来回收空间和扩展控件
- 需求:** 巨大的存储资源池, 并且可以随时扩展、高性能, 简化管理
- 解决方案:** 使用ARX-6000 集群, 分开Audio和Video存储, 形成两个存储资源池
- Result:**
- ❑ 生产效率提高560%
 - ❑ 管理难度降低20%



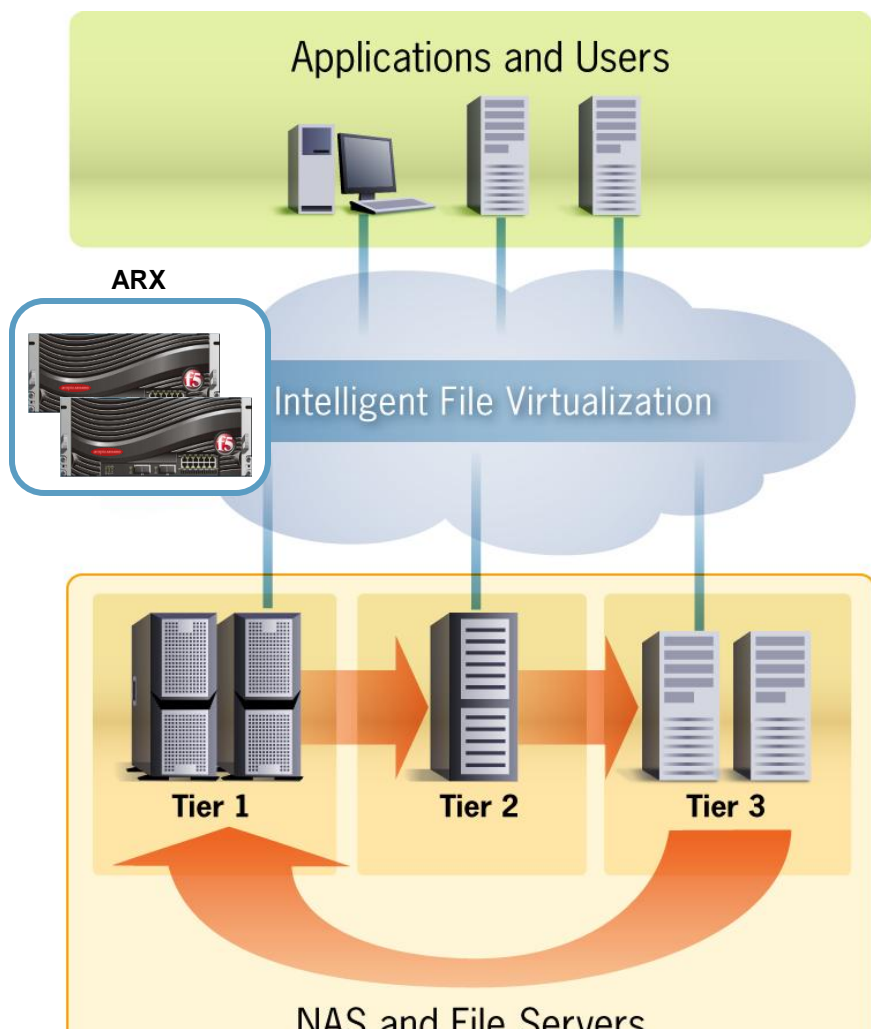
"F5's products increased our business workflow by 560% and reduced the overhead associated with storage management by 20%."

VMG

关于负载均衡后的文件一致性和备份

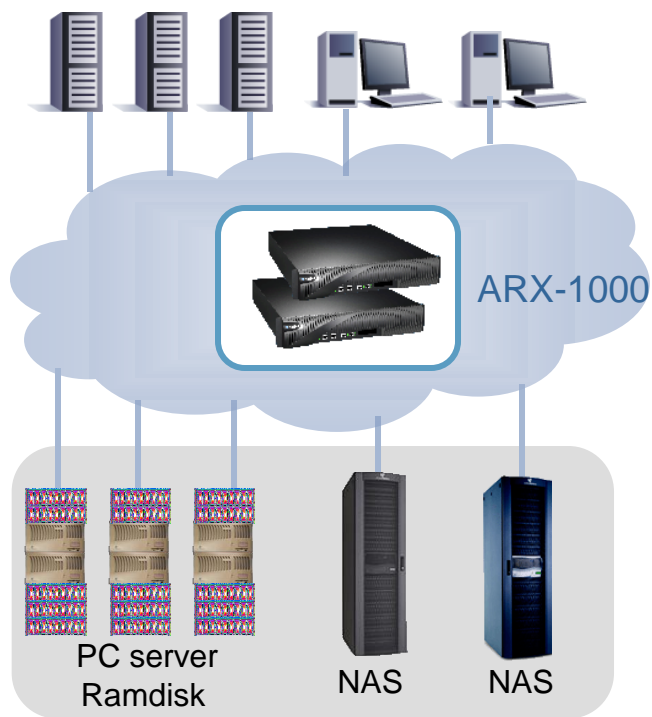
- 单个文件完整的保存在一个NAS上
- 可以基于目录进行负载均衡处理，保证一个目录下的文件都在一台NAS上
- ARX是通过写文件的负载均衡来提高读文件的访问速度
- 后台支持不同厂家的NAS设备

分级文件存储/归档管理



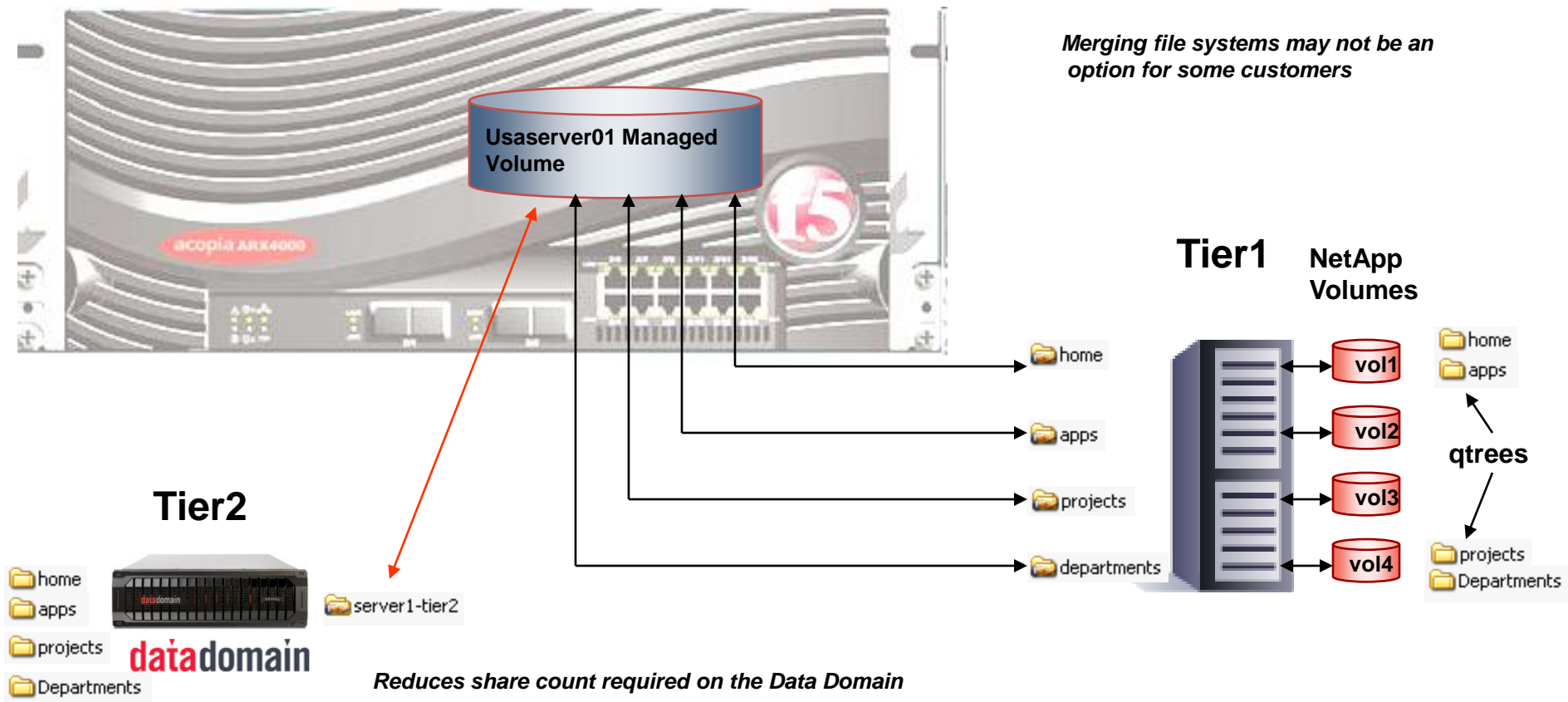
- 通过**ARX**，可以轻松的实现文件归类存储、分级存储、归档管理
- 使用大容量的廉价存储空间，使近线存储的成本降低
- 通过使用**Data DeDup** 存储设备，可以进一步降低二级存储成本
- 对于客户端而言，原有的存储结构没有任何变化

通过分级存储构建高速NAS



- 通过使用大内存的PC server用虚拟磁盘技术，将内存export为NFS server，得到超高速的NAS文件存储性能
- 在F5的一个特例用户中，构建了10TB的高速NAS系统，来满足动画渲染的速度需求

使用Data Dedup设备作为近线存储



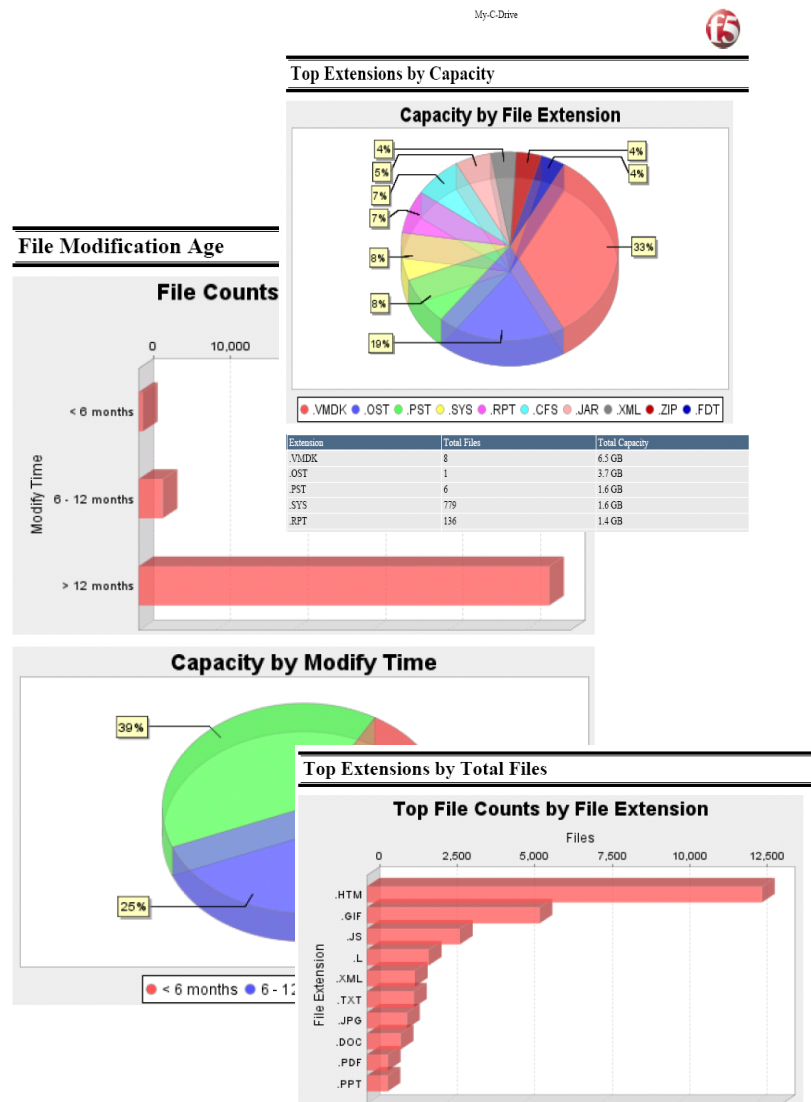
F5 DataManager的分析报告

❑ Data Manager 扫描文件系统后，主要生成报表：

- 文件读取时间
- 文件修改时间

❑ 其他报表包括

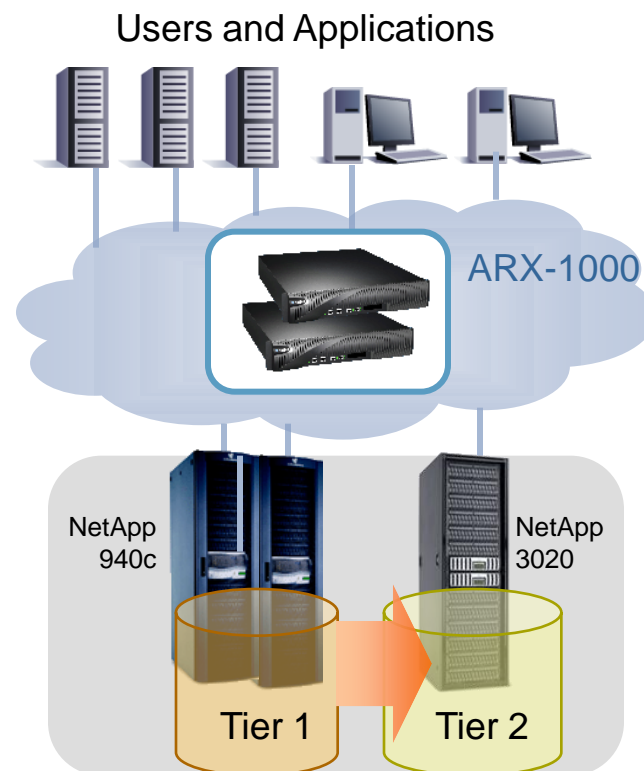
- Disk Capacity
- CIFS Information
- File System Structure
- File Capacity
- Top File Extensions
- Top Extensions by Capacity
- Top Users by Total Files
- Top Users by Capacity
- Top File Paths by Total Files
- Top File Paths by Capacity



案例：分级存储节省存储成本



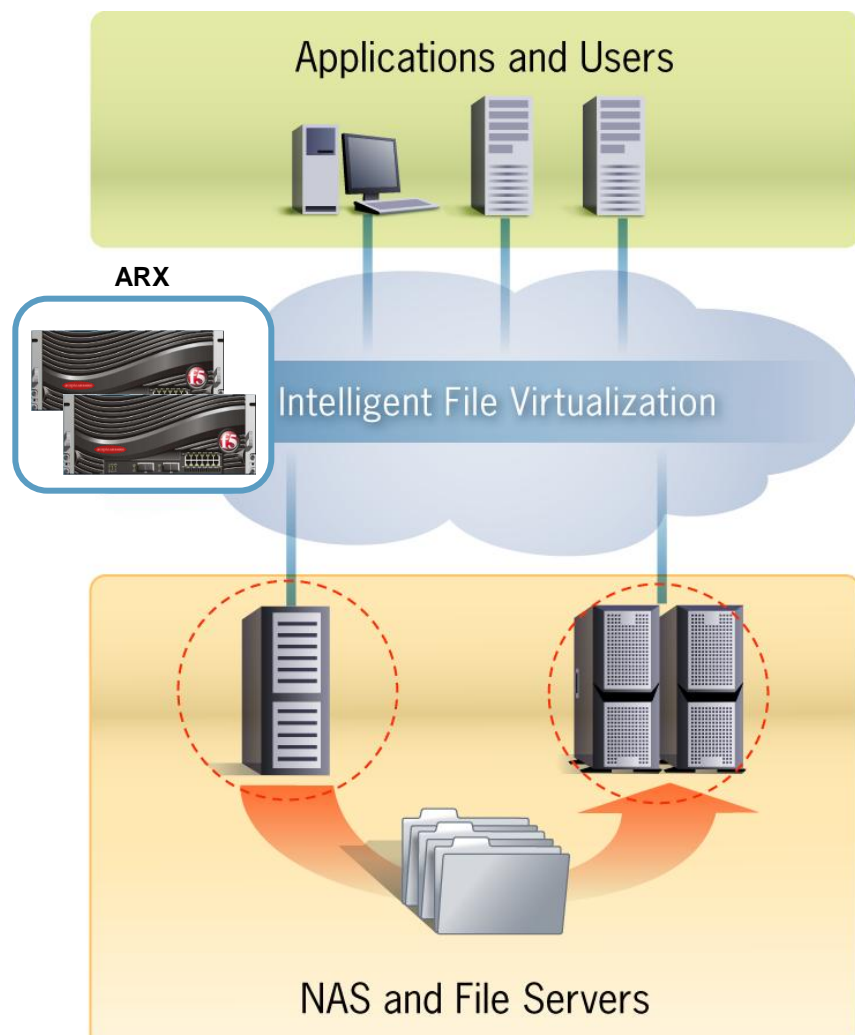
- 环境:** NetApp FAS940c, FAS3020
- 问题:** 存储空间增长带来的成本压力
- 原因:** 快速增长; 以前所有的数据都存放在单一的昂贵的存储设备上
- 需求s:** 自动化, 透明的将数据移动到不是很昂贵的二级存储上
- 解决方案:** ARX-1000 cluster
- 结果:**
- ❑ 减小了50%的磁盘购买成本
 - ❑ 备份窗口从14小时降到3小时



“Based upon these savings, we estimate that we will enjoy a return on our Acopia investment in well under a year.”

orf

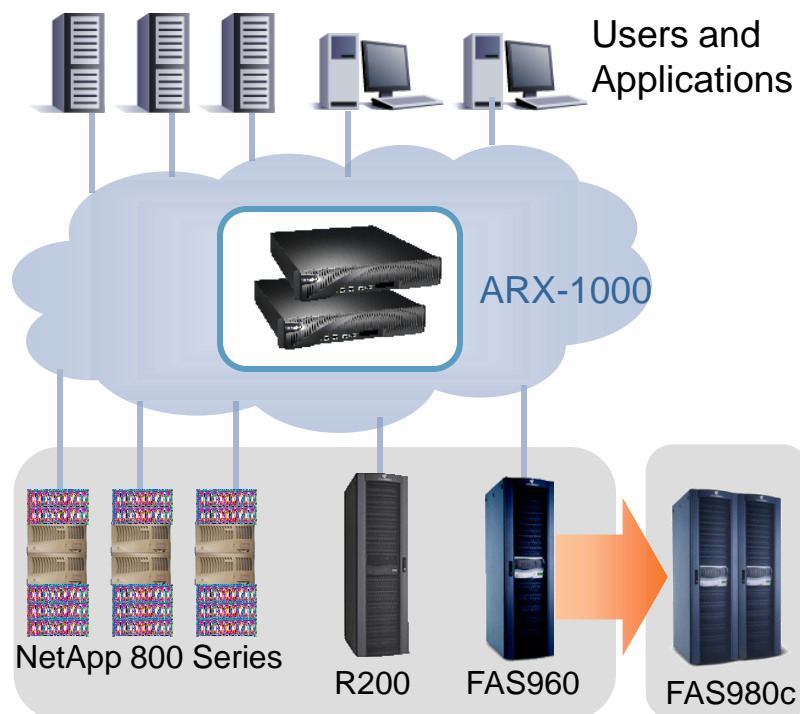
关于数据迁移



- 通过ARX文件虚拟化存储，可以实现在线的数据迁移
- 无需进行冗长的迁移规划
- 无需进行多次增量拷贝

案例：数据迁移

TOSHIBA



“F5’s products allowed us to perform a multi-terabyte data migration without disrupting our users”

Director of IS,
Toshiba America

挑战

- 需要在线的将数据从老的存储设备迁移到新的存储设备
- 需要升级到NetApp ONTAP 7G 而无需中断业务

解决方案

- ARX-1000 集群

好处

- 在无间断的状态下实现很多个T级的数据迁移
- NAS软件升级和老系统下线无需中断
- 通过自动化数据迁移和减小停机实现节省运营成本

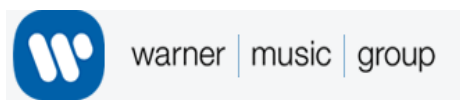
如何省钱？

- 存储设备节省：二级存储可以采用廉价的大容量存储方案
- 备份的费用节省：可以为二级存储设定更长时间的备份窗口
- 系统空间利用率提高：可以突破30%的设计界限，使系统可以使用更高的存储空间利用率
- 电力成本节省：将二级存储建立在更省电，单盘容量更大，转速更低的整列上，减小电力消耗
- 扩展的成本节省：不再受限于单一厂家的扩容方案
- 功能扩展成本节省：不用单独购买归档软件、数据迁移软件、远程复制软件
- 管理运维成本节省：在线数据迁移无需中断业务，在线扩容、在线移动文件

F5 ARX客户



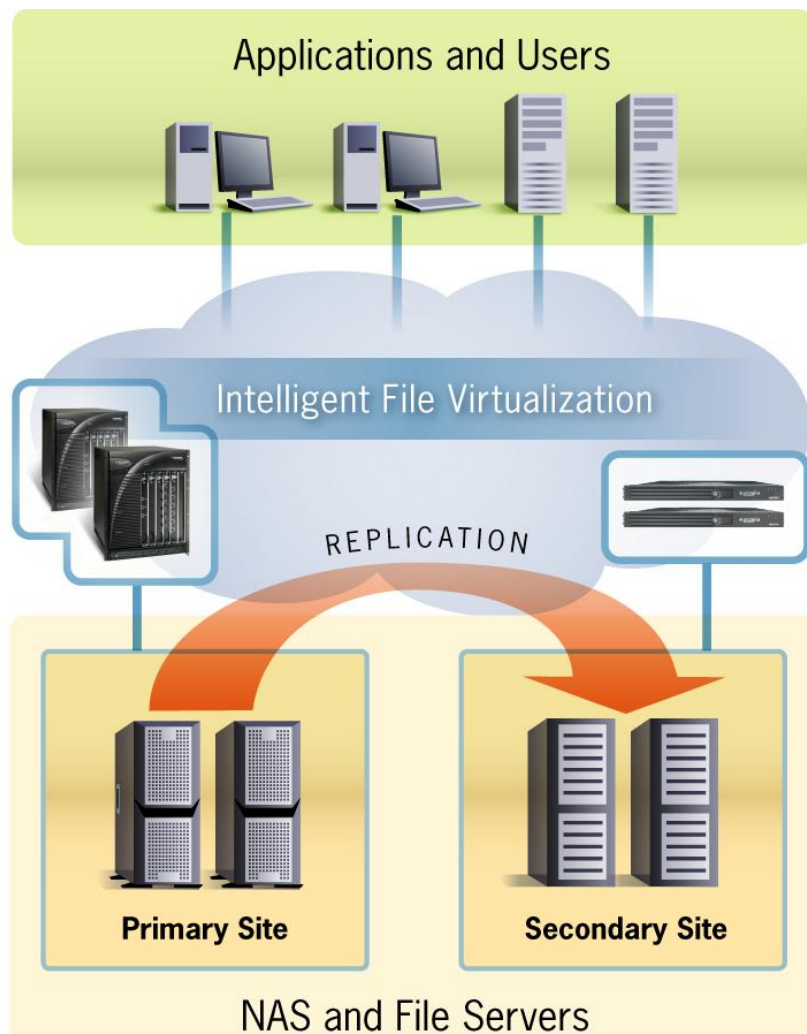
Foster and Partners





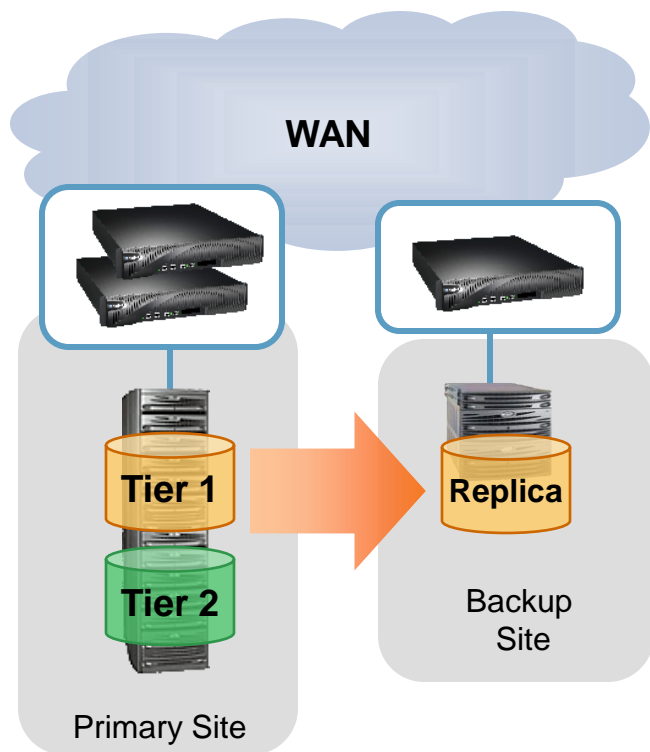
IT agility. Your way.

关于灾备站点复制



- 当在生产中心和备份中心都部署ARX的时候，可以通过ARX实现自动数据复制
- 可支持不同品牌的后端NAS设备
- 支持NFS和CIFS
- 支持文件增量复制

案例：远程复制



“Acopia has reduced our total backup and replication times by about 70%.”

Senior Systems Administrator, United Rentals

- World’s largest equipment rental company
- Challenges
 - Upgrade NAS platform
 - Introduce lower cost ATA disk
 - File-based Disaster Recovery solution
- Solution
 - ARX-1000 cluster at primary data center
 - ARX-1000 at disaster recovery facility
- Benefits
 - NAS upgrade with no impact to users
 - 50% savings through use of ATA disk
 - Cost effective offsite backup solution
 - Dramatic reduction in backup and replication times