云文件存储

徐立
mars.xul@alibaba-inc.com

Agenda

阿里云 aliyun.com

- NAS架构和发展趋势
- 典型应用场景
- 业界分布式文件系统的比较
- 阿里云NAS

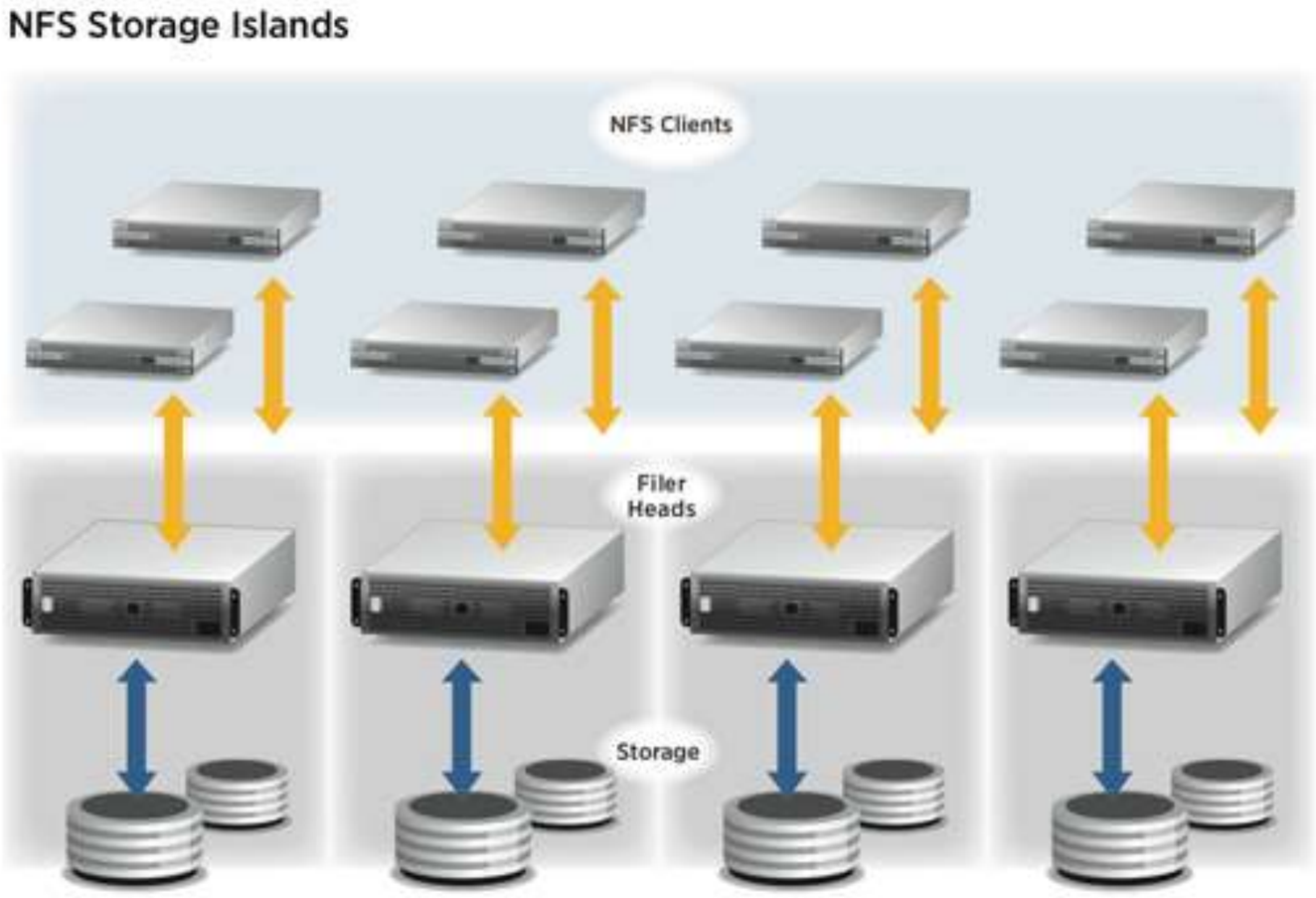
上海交通大学 SHANGHAI JIAO TONG UNIVERSITY

NAS架构

- 传统的纵向扩展(scale-up)架构
- 集群NAS架构 – 三种主流技术架构
  - –基于SAN的共享存储架构
  - –集群文件系统架构
  - –并行NAS架构（即pNFS/NFSv4.1架构）
- Cloud NAS架构

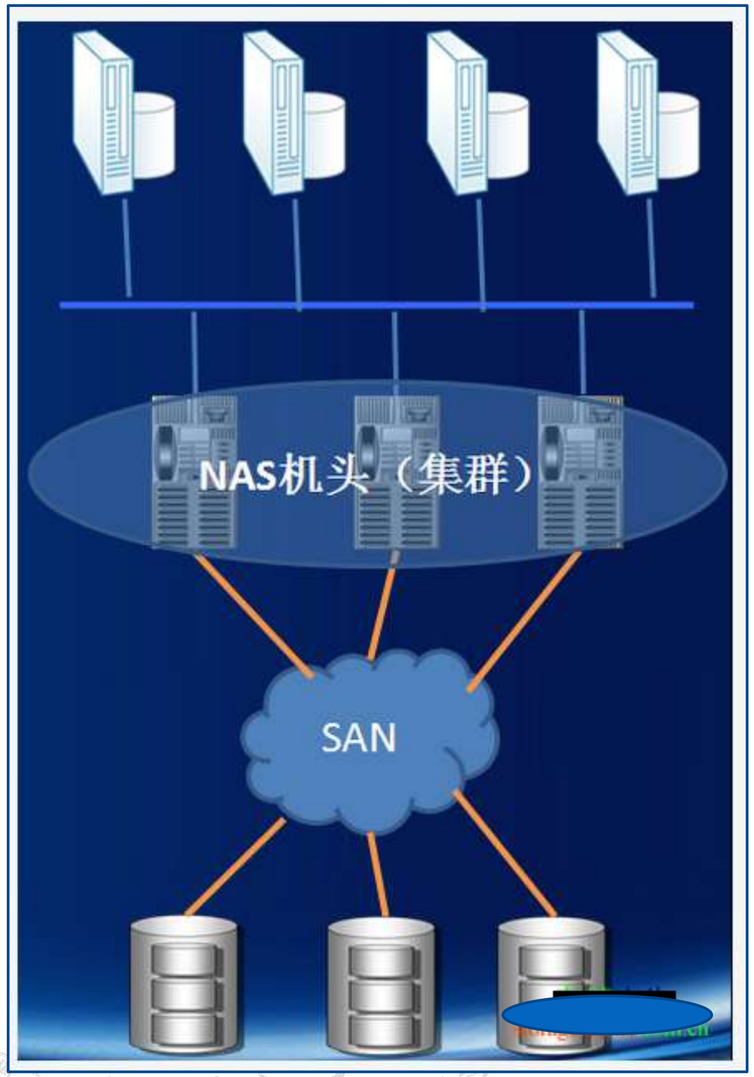
传统的纵向扩展(scale-up)架构

Scale Up



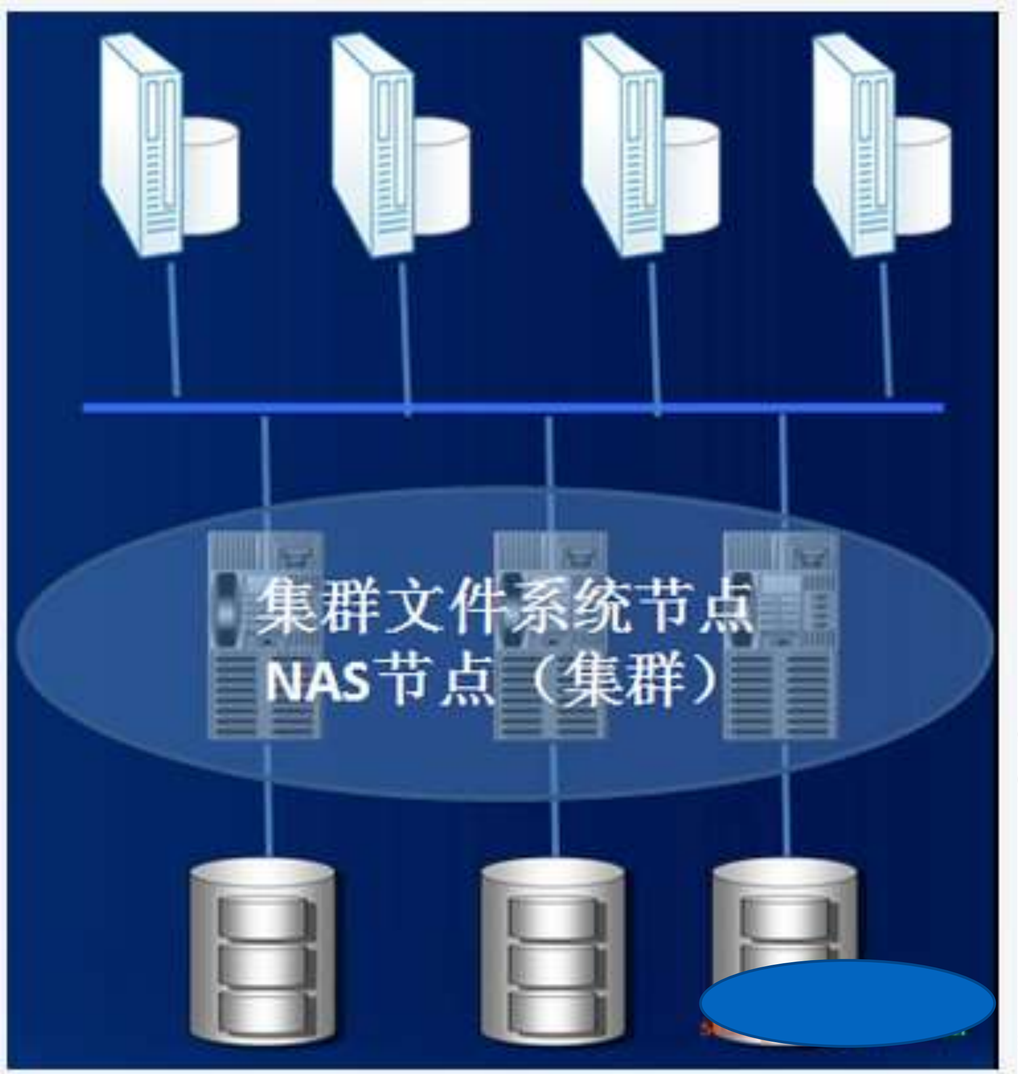
Island

基于SAN的共享存储架构



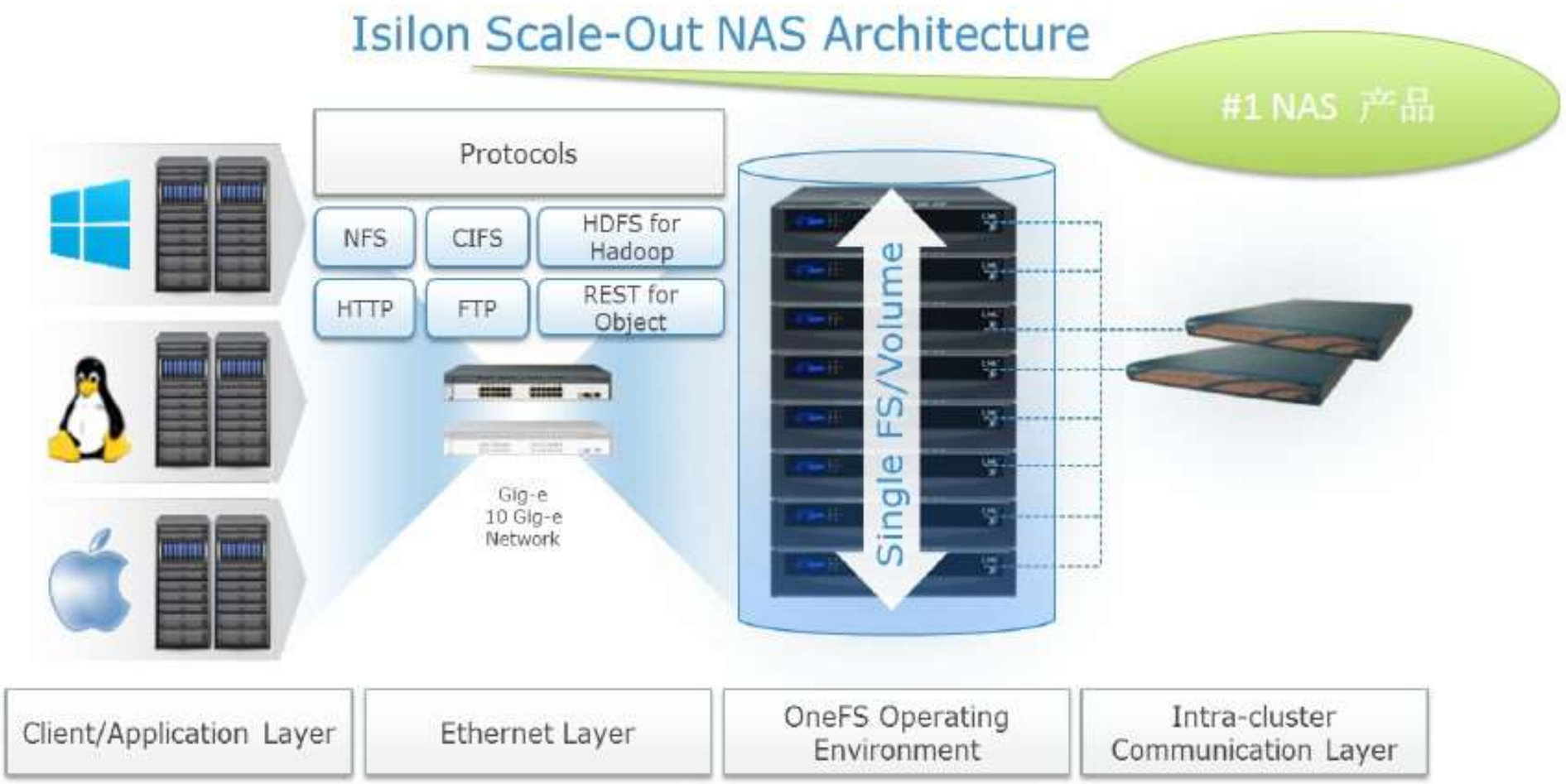
- 后端存储使用SAN
- NAS机头集群结点通过光纤连接到SAN
- DNS或LVS实现负载平衡和高可用
- 稳定的高带宽和IOPS性能
- 成本高，管理复杂
- 扩展规模有限

集群文件系统NAS架构

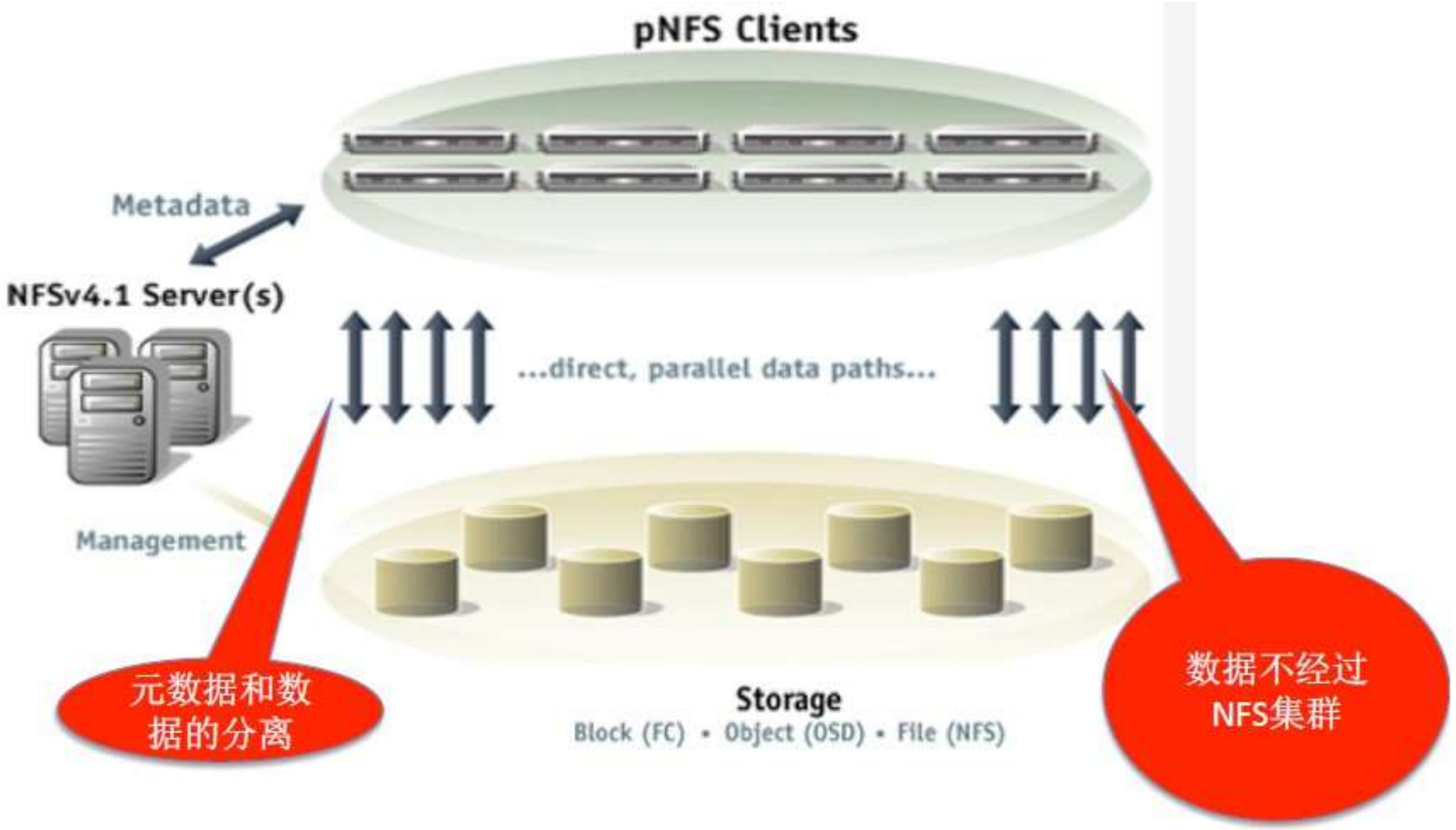


- 存储由普通的服务器加本地存储组成
- 集群文件系统管理存储空间并提供单一的名字空间
- NAS集群，元数据集群和存储集群一般共享物理机器

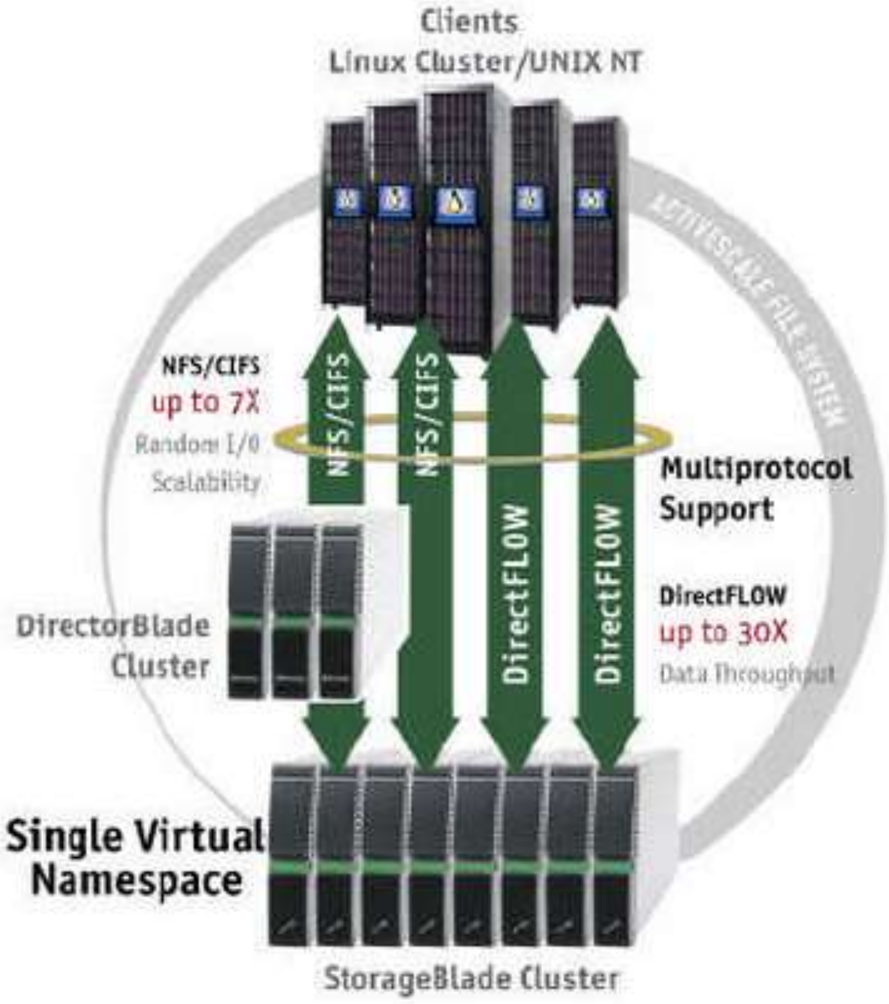
EMC Isilon -- 全对称的集群文件系统



并行NAS架构 -- pNFS/NFSv4.1协议



Panasas PanFS



- 单一全局名字空间
- 存储刀片和客户端直接的和高度并行的数据访问
- 指挥刀片负责文件管理，对象映射元数据管理
- 指挥刀片软件Quorum based的高可用

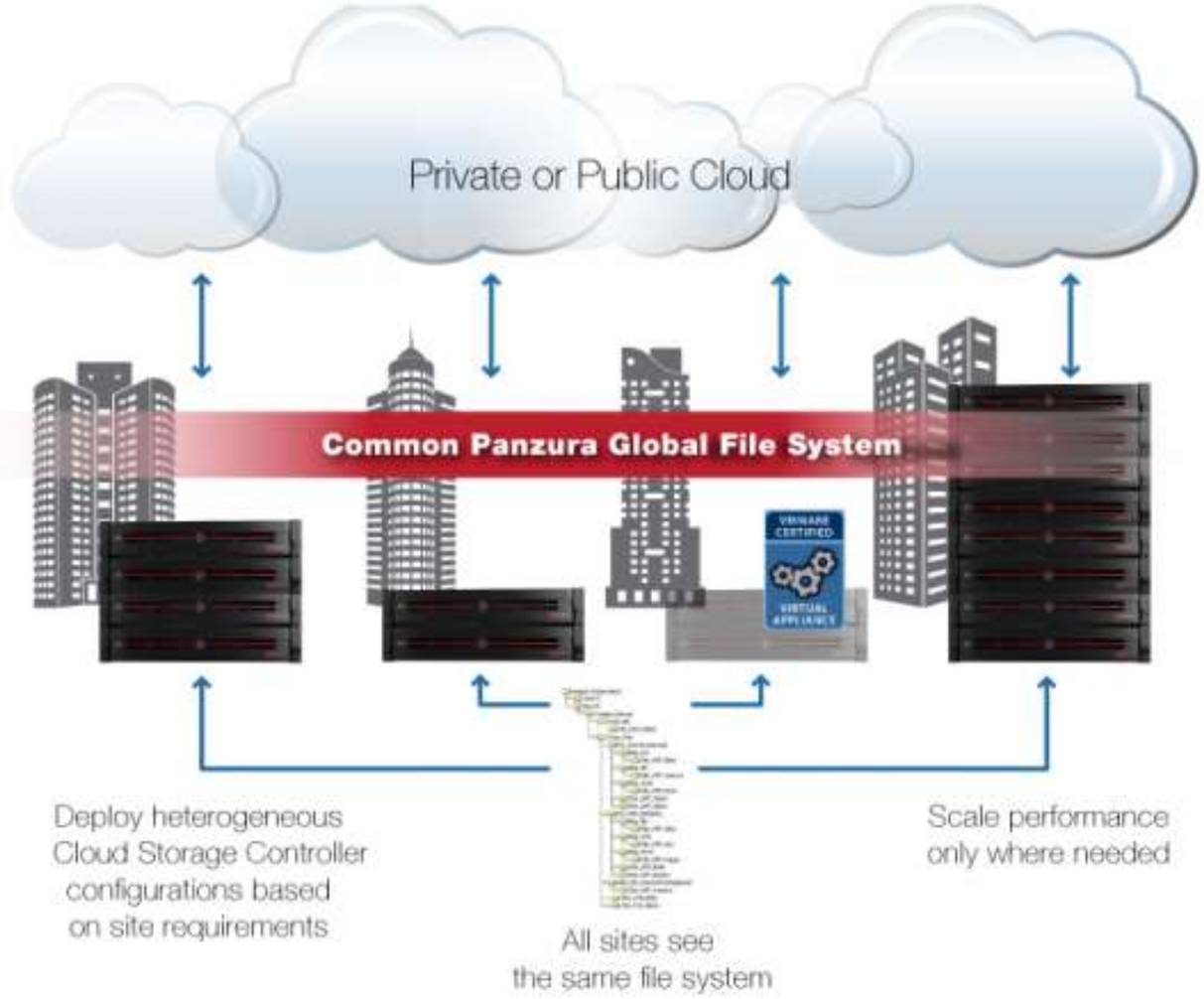
Cloud NAS架构

- Cloud NAS Gateway架构
- Cloud NAS分布式文件系统
- Azure File Storage
- AWS Elastic File Service

Cloud NAS Gateway架构

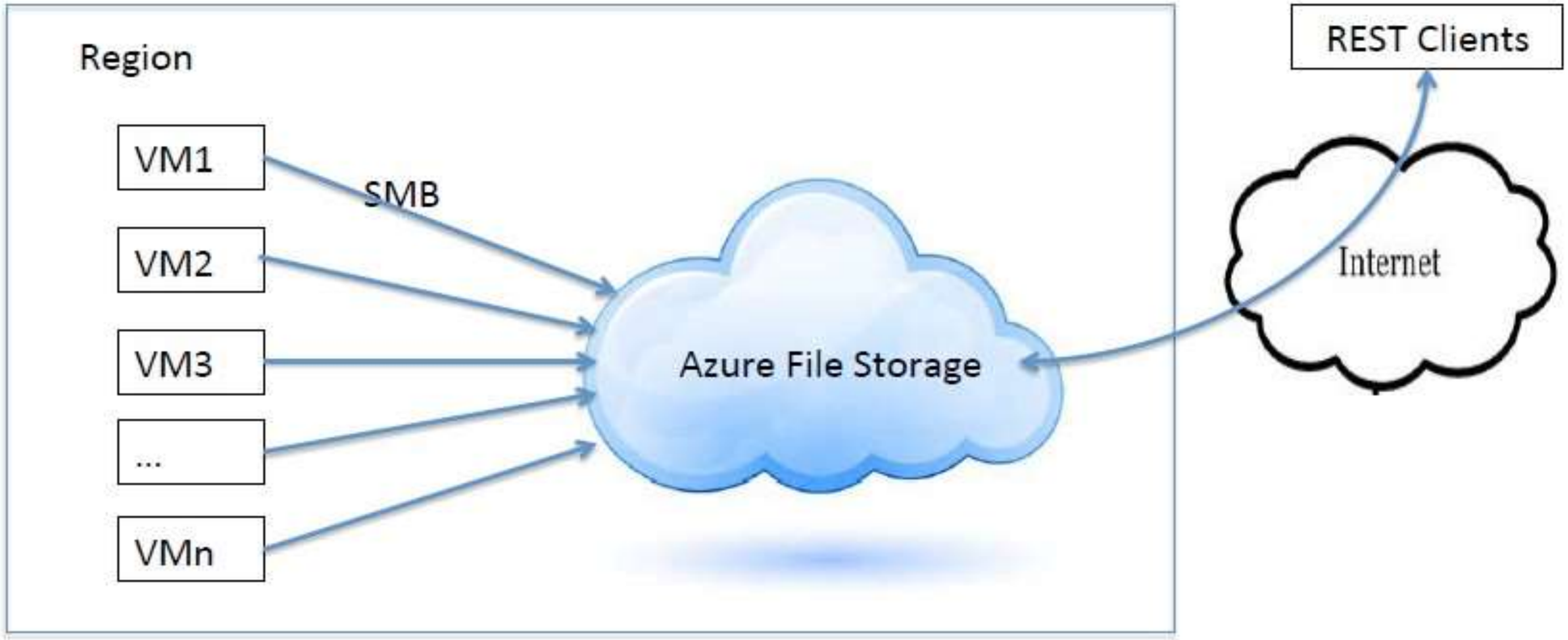
- Azure StorSimple
- Nasuni
- Panzura
- 以Azure Blob或者AWS S3为存储建立一个NAS的gateway
- 实现全局命名空间
- 锁管理
- 权限管理
- Cache提速
- 去重

Cloud NAS分布式文文件系统



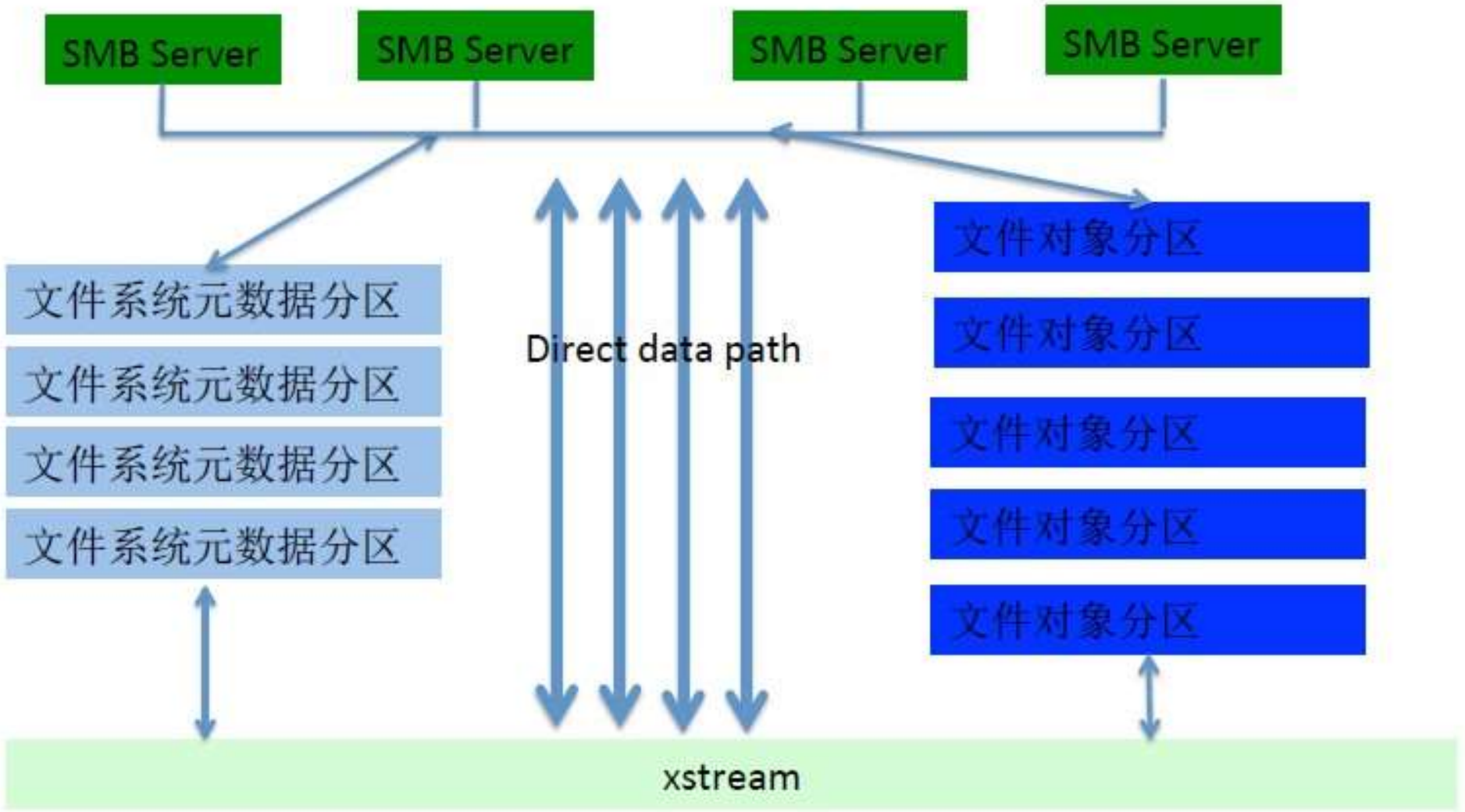
- Panzura，Nasuni
- 全局命名空间
- 各地的控制器交换元数据形成分布式文件系统
- 控制器同云端通过VPN连接

Azure File Storage

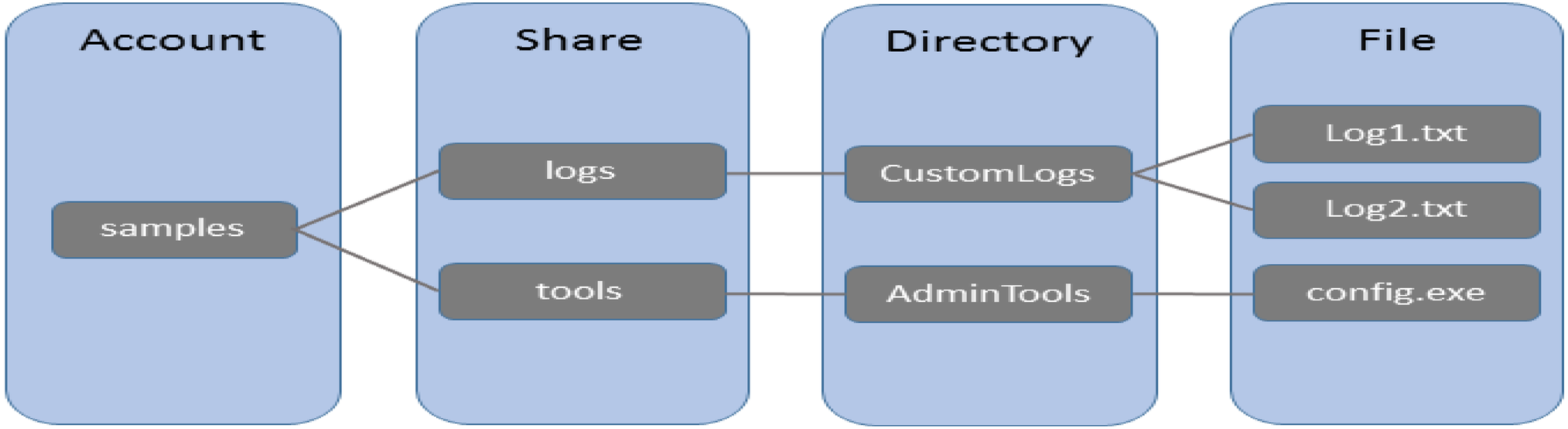


- 双接口：支持SMB2.1/SMB3.0协议和REST的访问
- File Share的访问仅支持同一Region内的VM的访问

Azure File Storage架构



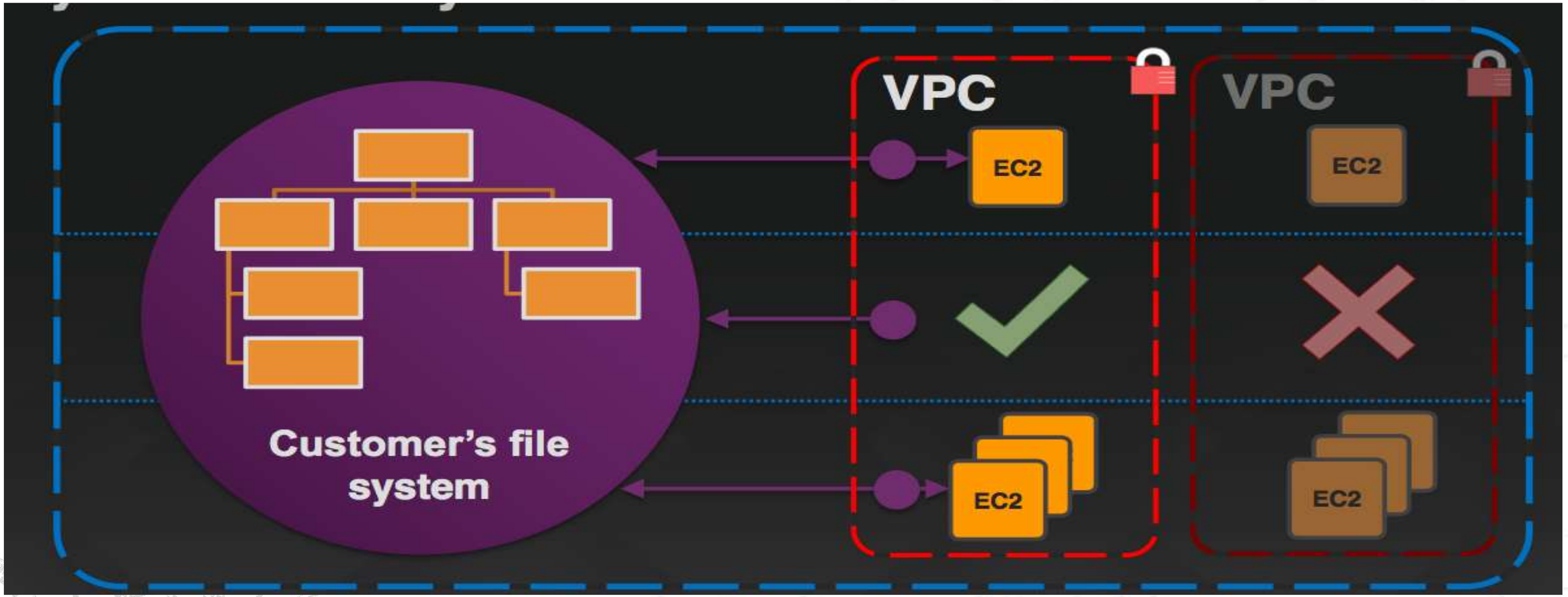
Azure File的使用方式



- 只能从同一Region中的VM来访问
- 一个用户可以开多个Share
- 一个share就相当于一个挂载的目标
- 单一Share可以被多个VM共享

AWS EFS

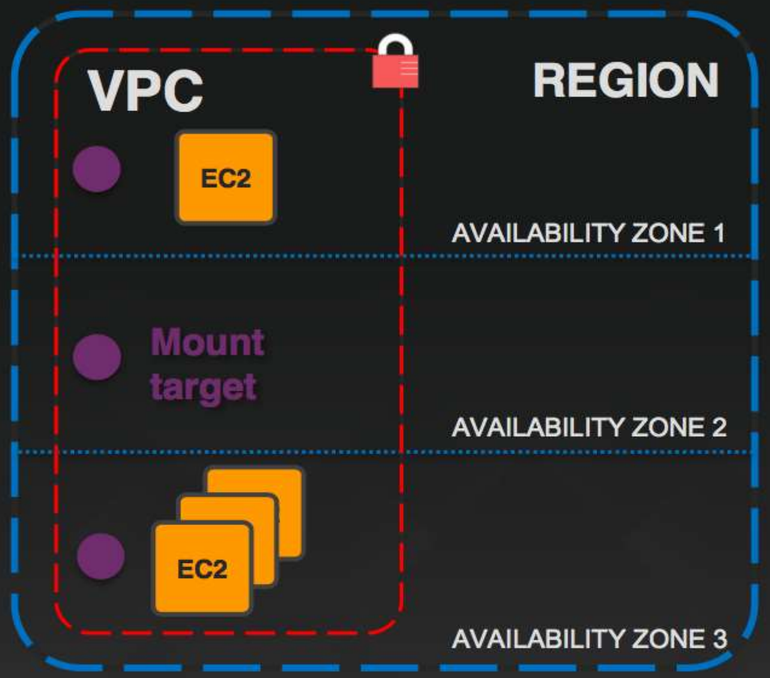
- NFSv4协议
- Access EFS in VPC



EFS的挂载目标

What is a mount target?

- To access your file system from instances in a VPC, you create *mount targets* in the VPC
- A mount target is an NFSv4 endpoint in your VPC
- A mount target has an IP address and a DNS name you use in your mount command



amazon web services

```
mount –t nfs4
    [file system DNS name]:/
    /[user's target directory]
```

Scale-out NAS的发展趋势（一）

- 共享，thin provisioning
- 可扩展性at scale
  - 50PB，不中断服务的水平扩展
  - 性能的线性扩展，支持1.8M IOPS,150GB/s
- 多协议支持，操作灵活
  - NFS，CIFS，HTTP，FTP，HDFS，RESTAPI（Swift,S3等）
  - 全局命名空间
- 层级数据存储和自动数据移动管理
  - Memory，NVME，SSD，HDD，NL HDD
  - 不同层级之间的移动

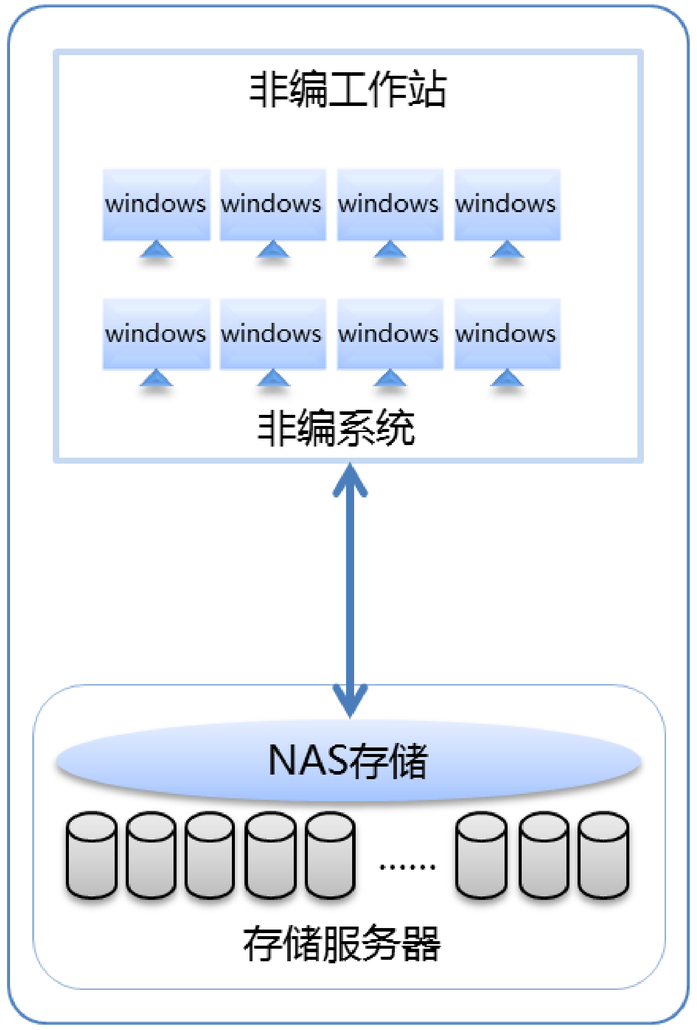
Scale-out NAS的发展趋势（二）

- 支持容量型和IOPS型的混合负载
- 企业级数据保护：快照，备份，恢复和容灾
- 企业级安全选项：ACL，分离，加密等
- 与Cloud存储协议的兼容
  - HDFS接口
  - RESTful接口
  - Swift/S3接口的兼容

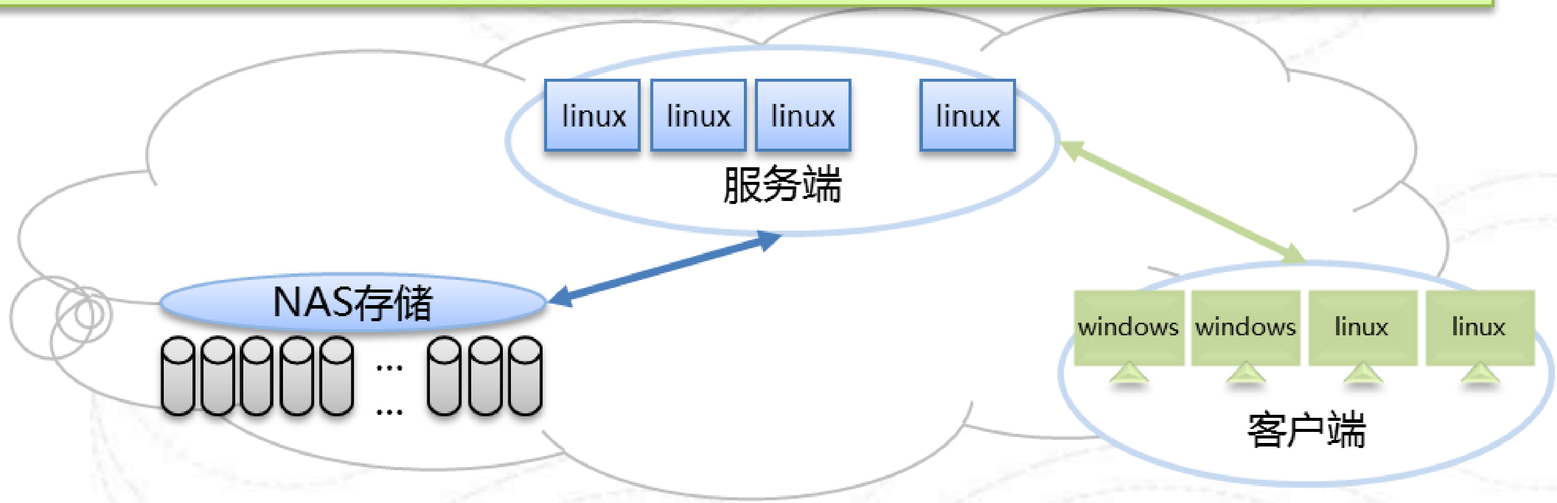
典型应用场景

- 媒资行业：桌面图形工作站
- 高性能计算：数据分析和处理
- 金融行业：传统金融软件访问文件系统

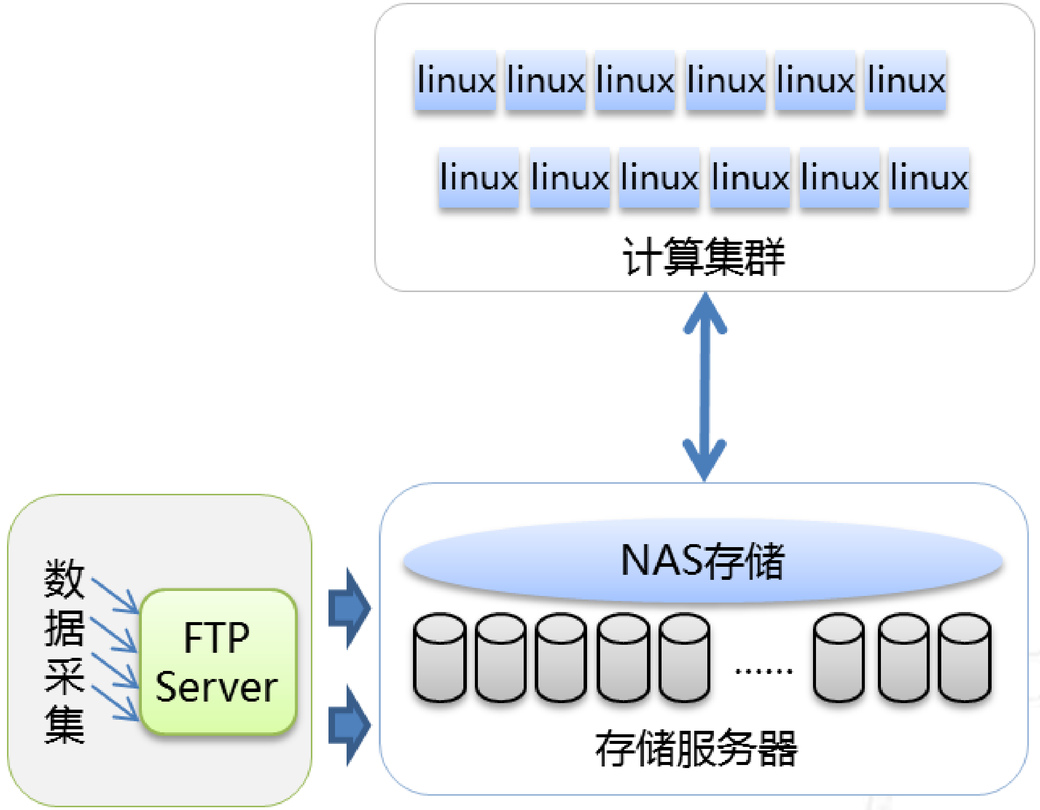
媒资行业应用场景



- 码率：每个工作站4~8路码流、每路130mbps吞吐量
- 图形工作站：单个新闻非编系统超过40+图形工作站，要求访问同一个存储；
- 单系统的吞吐量：8*130Mbps/8*40=5200MBps
- 时延：≤10ms
- 容量：≥500TB
- 编辑客户端：windows 7专业版（64位）、windows xp专业版(32位)
- 配额：不同图形工作站有不同的存储容量要求与限制

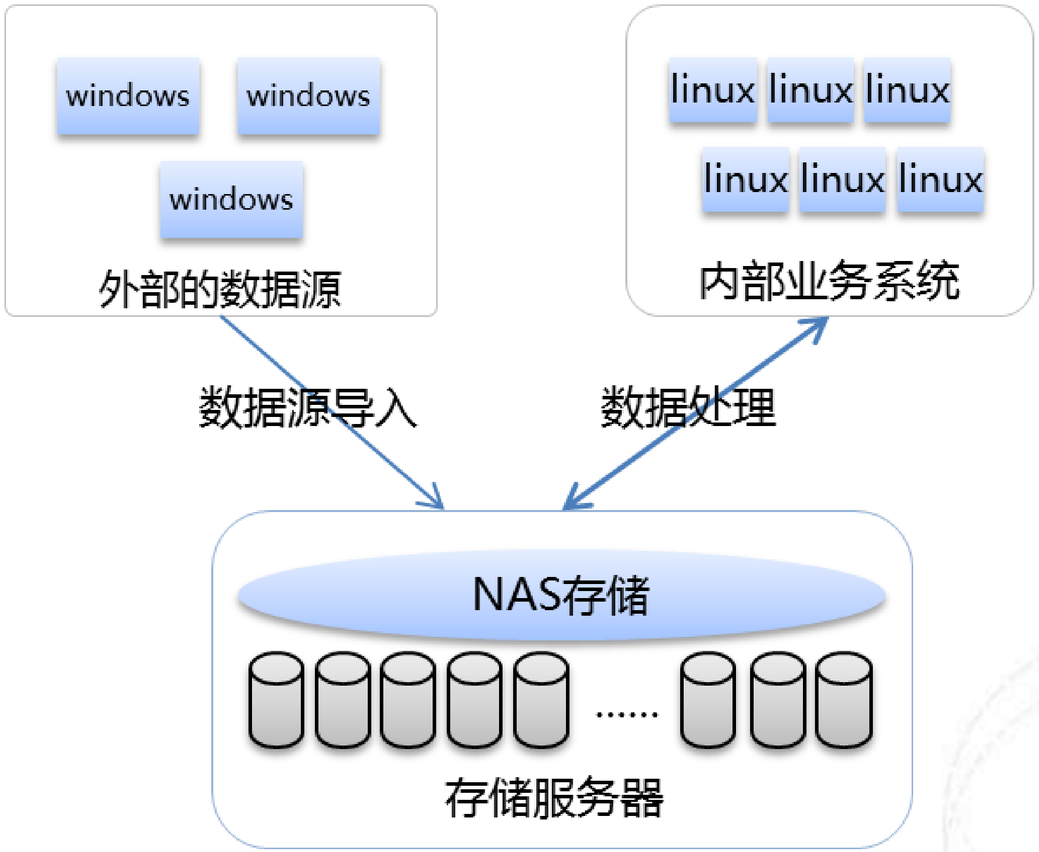


高性能计算应用场景



- 数据类型：非结构化数据（采集的数据源、过程及产品数据）
- 数据量：5PB~6PB
- 吞吐量：≥2GBps
- 时延：≤10ms
- 计算集群：linux计算集群为主，少量的windows 2003/2008桌面应用

金融行业典型应用场景



- 数据类型：非结构化数据
- 数据量：50TB
- 文件数：3亿+
- 时延：≤10ms
- 高可用，主备切换
- 支持同城和异地容灾
- 从nas迁出的方案，不能让应用改代码
- cifs和nfs同时访问一个文件系统
- 每天至少要有一份快照备份

业界商业化分布式文件系统的对比

阿里云 aliyun.com

产品	云服务	支持协议	性能	其他功能集	成本
EMC Isilon	专有云	NFS/CIFS/S3/HDFS/RESTful/iSCSI	容量可用提供20PB，吞吐106GB/s，读写3M IOPS，1.6M/s CIFS file operation	容量线性扩展，快照，异地备份，多级数据保护，支持EC	一体机售卖
华为OceanStor9000	专有云	NFS，CIFS，NDMP，FTP，HDFS	单一文件系统容量高达60PB；单节点吞吐可达1.6GB/s，整系统可达400GB/s；	同EMC；统一存储（文件和对象存储）	一体机售卖
Amazon EFS	公共云	NFSv4.0	可以支持 PB 级的文件系统，还能支持数千个并行 NFS 连接，性能和容量线性扩展，可以burst up性能	全SSD，VPC，跨可用区冗余存储，无强大的企业级功能	每月每 GB 0.30 USD
Azure File Storage	公共云	支持smb2.1/smb3.0，提供smb和REST两种接口	一个share最大空间5TB，单个文件最大1TB，每个share最大throughput是60MB/s，最大1000 iops(8KB)	混合存储，无强大的企业级功能	空间：$0.08/GB/月 流量：读$0.015/100K IOPS；写$0.15/100K IOPS

EMC Isilon分析

Isilon 横向扩展 NAS 的价值

简便性和易用性
单个文件系统，单个卷，全局命名空间

巨大的可扩展性
在单个文件系统中扩展到 20PB 以上

创世界记录的性能
超过 100 GB/s 的吞吐量，每秒 160 万次 SPECsfs 操作

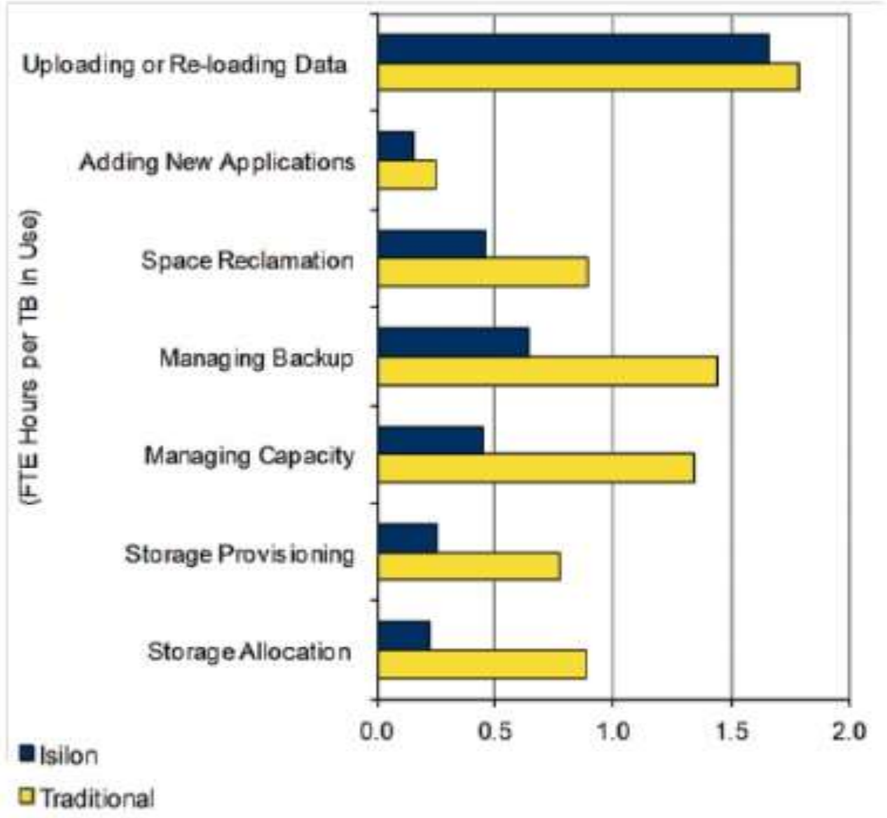
Unmatched efficiency
超过 80% 的存储利用率，自动存储分层

企业数据保护
高效的备份和恢复，可靠的灾难恢复，以及 N+1 到 N+4 冗余

强大的安全选项
基于角色的管理；身份验证分区；符合 SEC 17a-4 法规要求的 WORM 数据安全性

操作灵活性
对包括 NFS、SMB、HTTP、FTP 和 HDFS 在内的行业标准协议的集成支持；平台 REST API VMware 集成

EMC Isilon
Ease of Use And Management

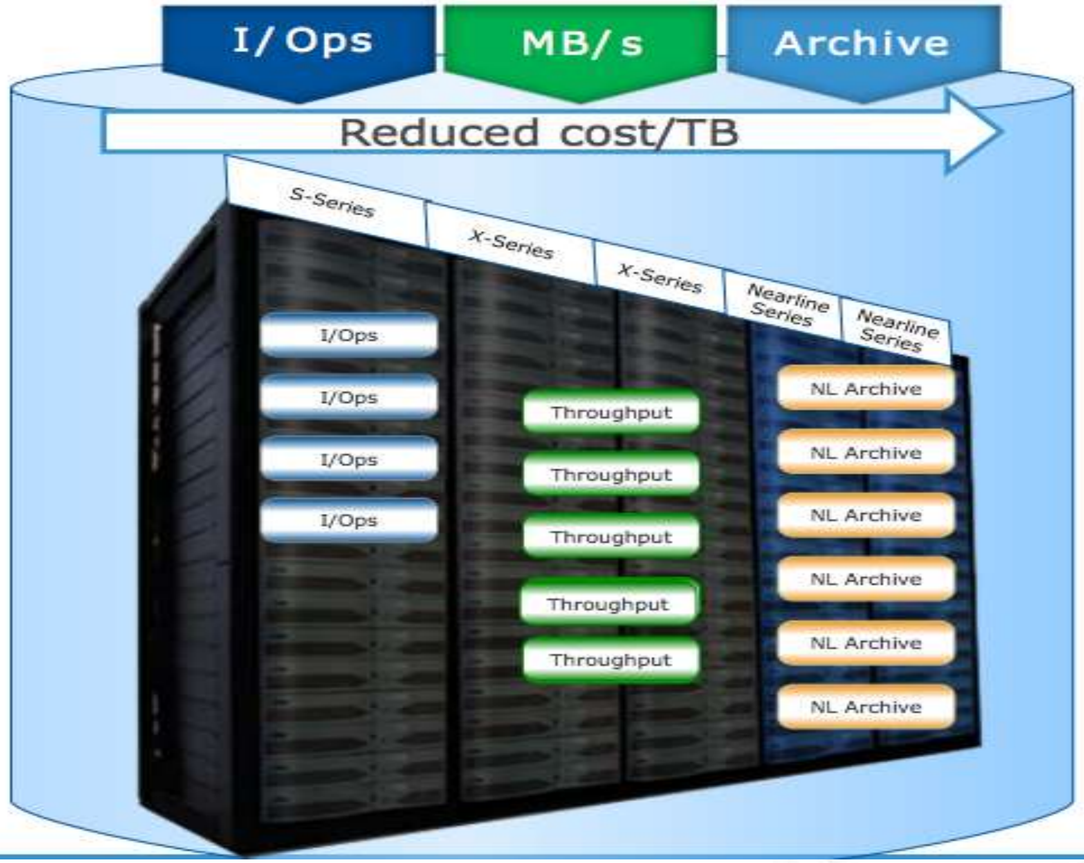


Isilon的三种产品形态

IOPS随机读写型　　高并发高带宽型　　低成本大容量型

Isilon层级存储和自自动数据移动

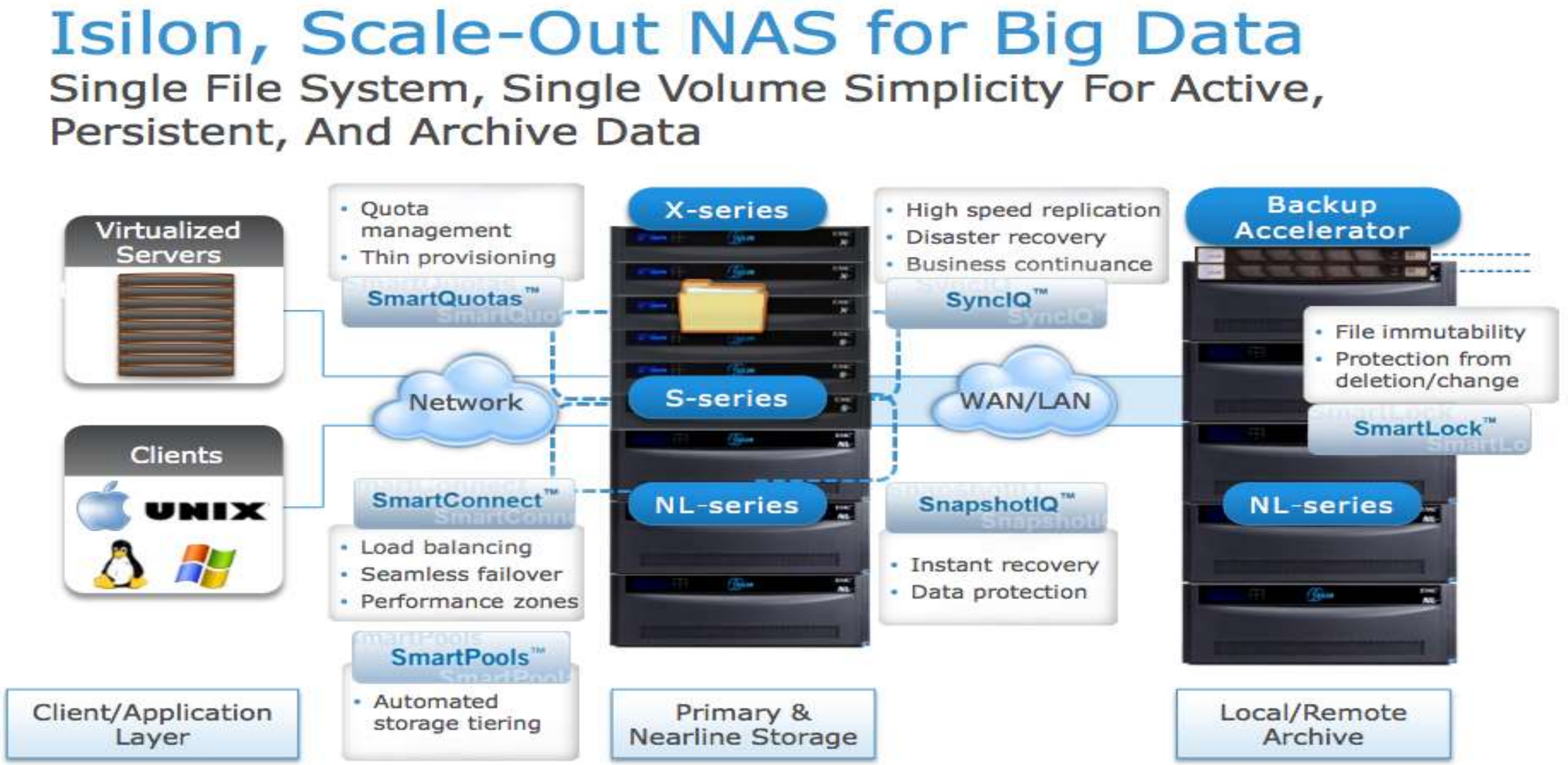
Optimize Resources With SmartPools Automated Tiering



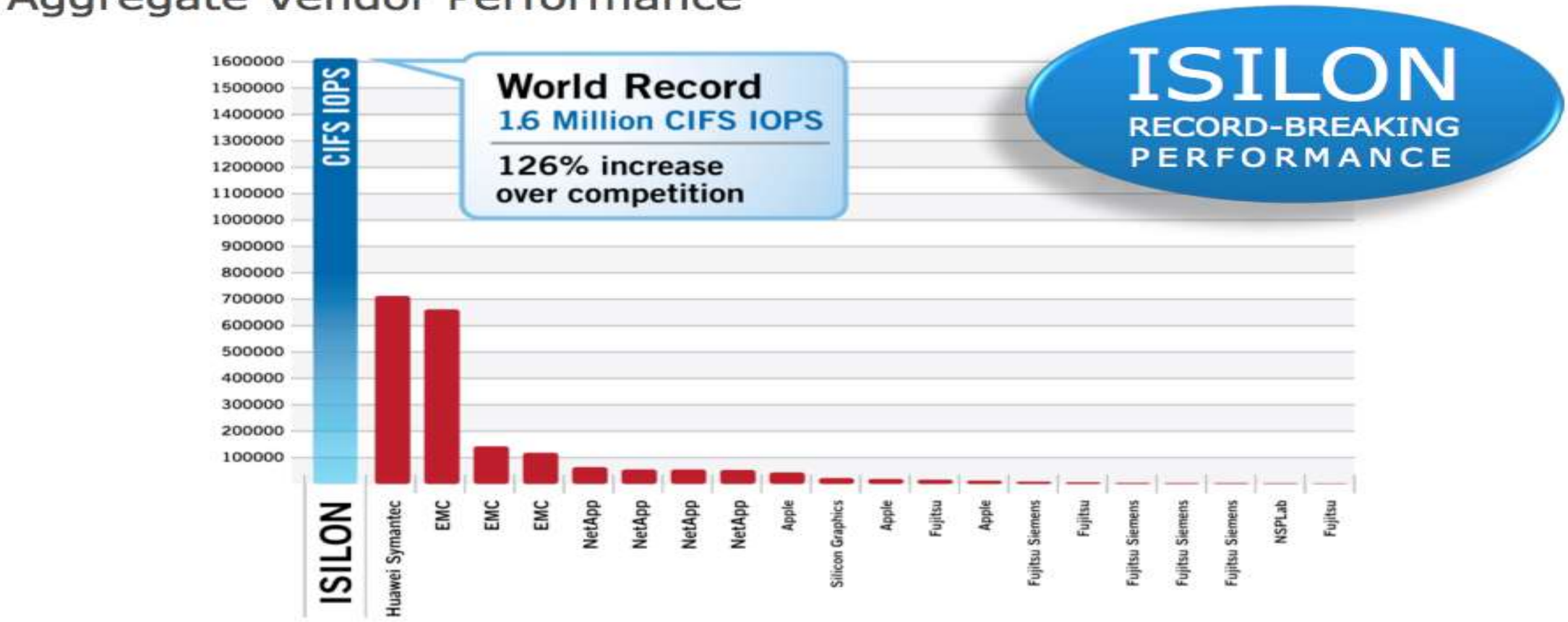
SmartPools™ Innovation

- Single point of management
  - Single file system
  - Single volume
  - Tiers of performance
- Automatic data movement
  - Policy-based movement
  - Transparent reallocation
  - NO application-changes
- Investment protection
  - Eliminate data migration
  - Scale any application
  - Completely transparent
  - Pay as you grow

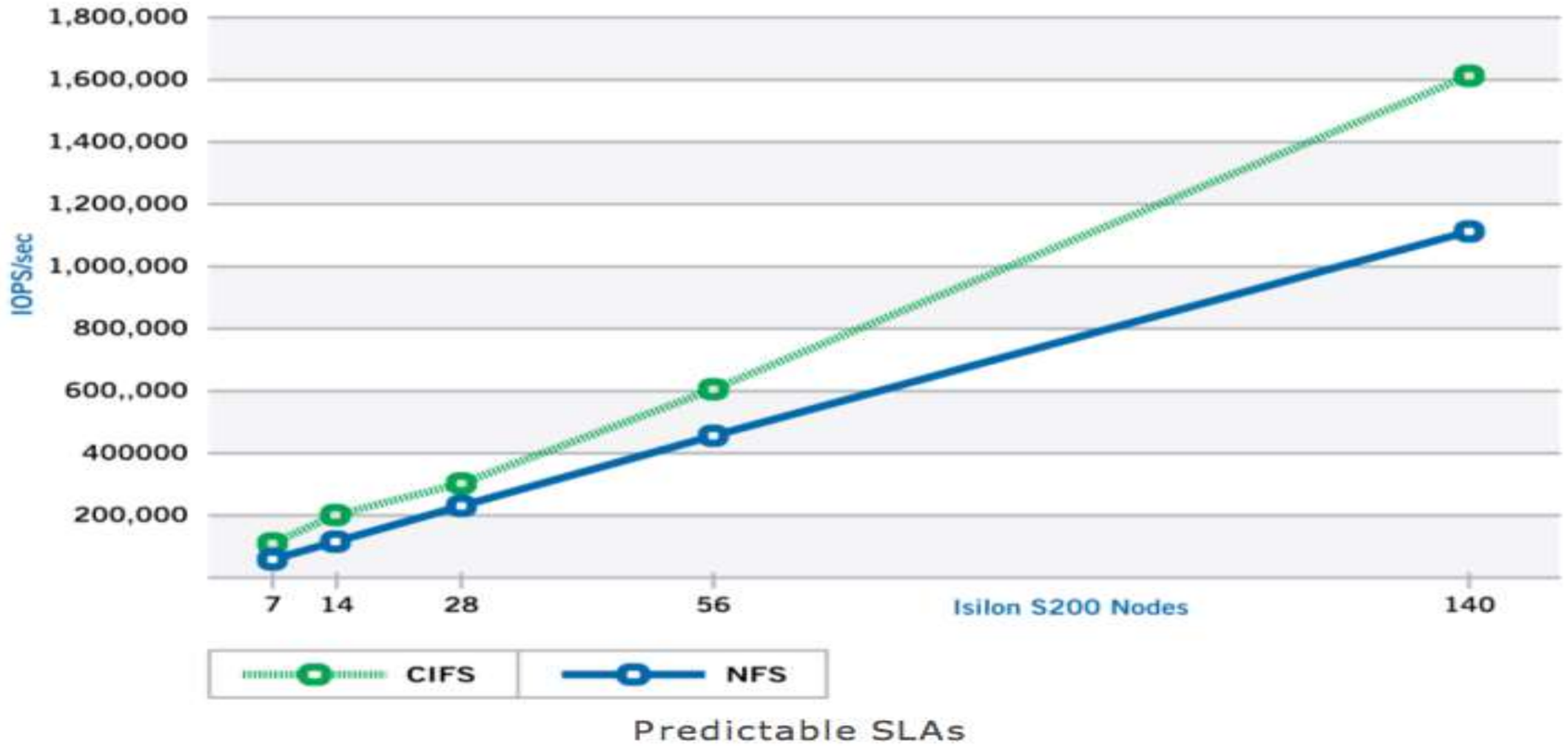
Isilon Big Data的组合拳方案



Isilon CIFS(SMB)IOPS性能



Isilon线性扩展的性能



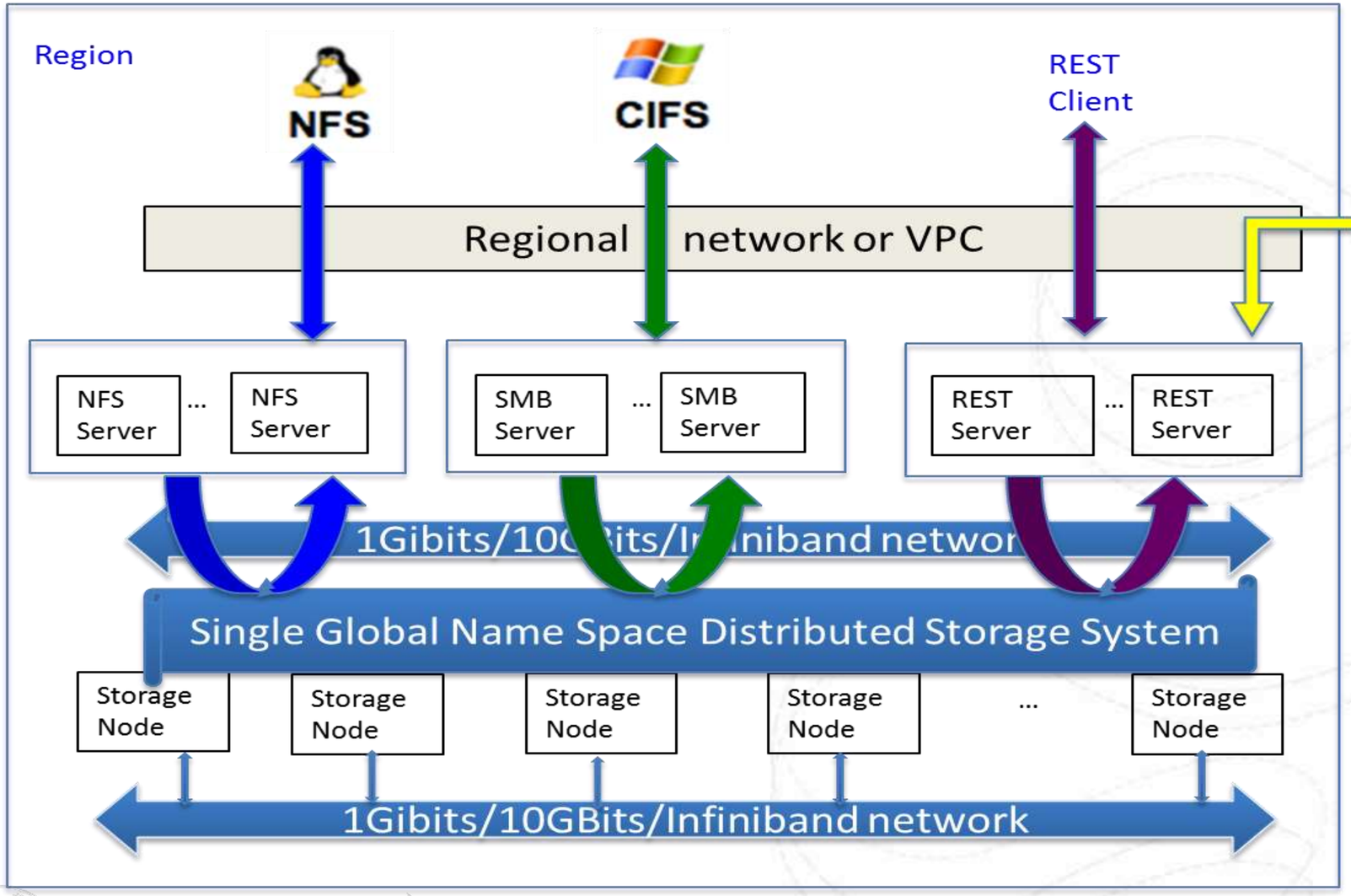
Scale out NAS的主要产品

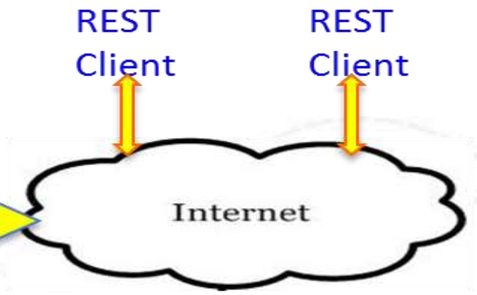
原生态横向扩展存储	独立存储方案供应商	综合方案供应商	特殊领域供应商
• EMC Isilon	• NetApp (FAS Series - Clustered Data ONTAP) • HDS (HNAS)	• IBM (SONAS) • HP (Ibrix) • DELL(FS7500) • Huawei (N8000)	• DDN (ExaScaler) • Panasas (ActiveStor) • Quantum (StorNext)

阿里云NAS产品简介

阿里云文件存储是提供标准的NAS文件存储接口，无限容量，单一命名空间，共享，安全，高可用，高可靠，高性能的分布式文件存储服务。

阿里云NAS架构





- 支持多种协议
- NFS，CIFS，REST，FTP，HDFS，S3等
- NFS and CIFS 访问限制在区域网络中
- 支持VPC的访问
- 同一个文件可以被多个协议访问
- 高可扩展，高可靠，高可用

阿里云NAS产品特性

完整的文件存储产品
- 控制台、售卖、计费、监控、运维

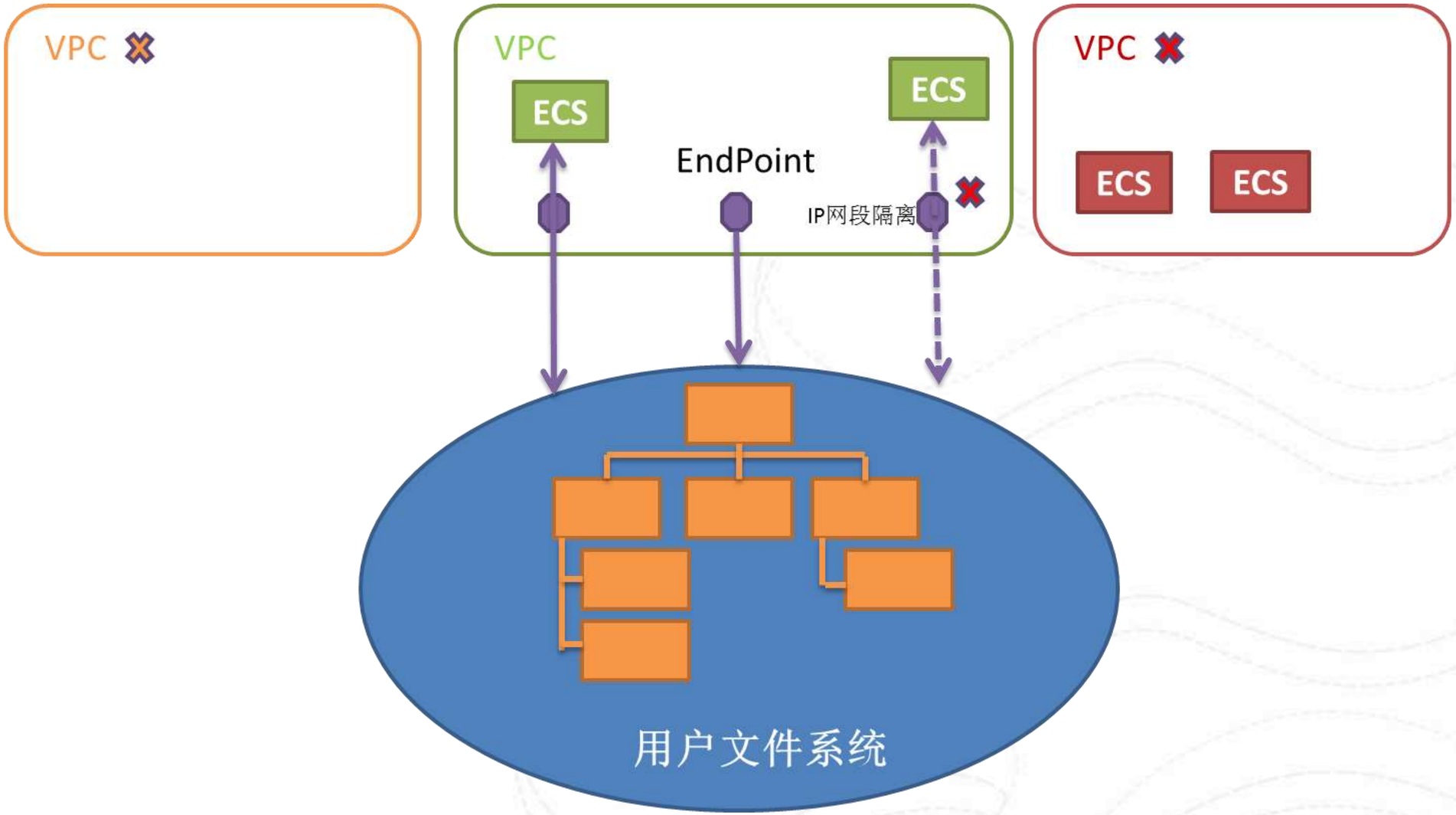
协议：支持NFS3.0/NFS4.0

安全
- 网络层控制：VPC、安全组ACL
- 控制文件和目录访问：标准的目录/文件级权限
- 控制台访问：RAM

高可用
- SLB
- Clustered NFS Server
- 基于飞天分布式系统

水平扩展
- 分布式元数据管理，分布式文件数据管理
- 盘古分布式存储
- 多个NFS Server的扩展

安全的访问控制（接入VPC）



应用案例（ECS共享存储）

使用SLB+多台ECS（通常web服务器）部署业务，多台ECS需要访问同一个存储空间，以便多台ECS能共享数据。

日志共享：多台ECS应用，需要将日志写到同一个存储空间，以方便做集中的数据处理与分析。

办公文件共享：公共的文件需要共享给多组业务使用，需要集中的共享存储来存放。

数据备份：用户在线下机房的数据希望备份到云上，同时要求云上的存储服务兼容标准的文件接口。

阿里云 aliyun.com

Thanks!