QINGCLOUD青云

青云存储实践

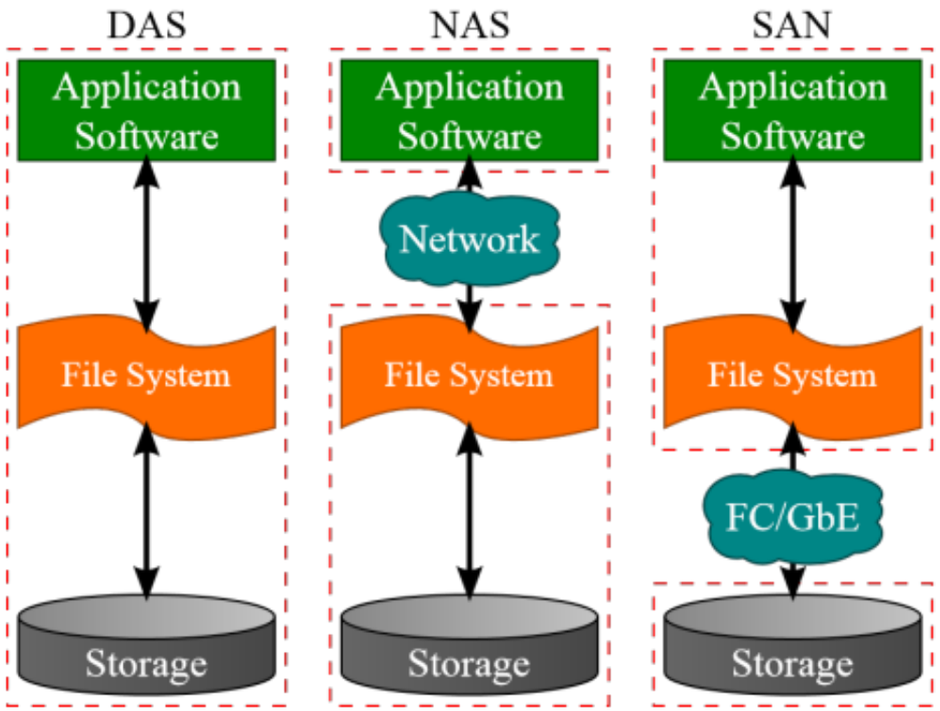
文: 甘泉（Reno）
图: 吕婷（Samantha）

Agenda

- 青云存储分类
- 青云存储架构
- 与其它存储方案的比较

Traditional Enterprise Storage

DAS (Direct Attached Storage)
SAN (Storage Area Network)
NAS (Network Attached Storage)



青云存储分类

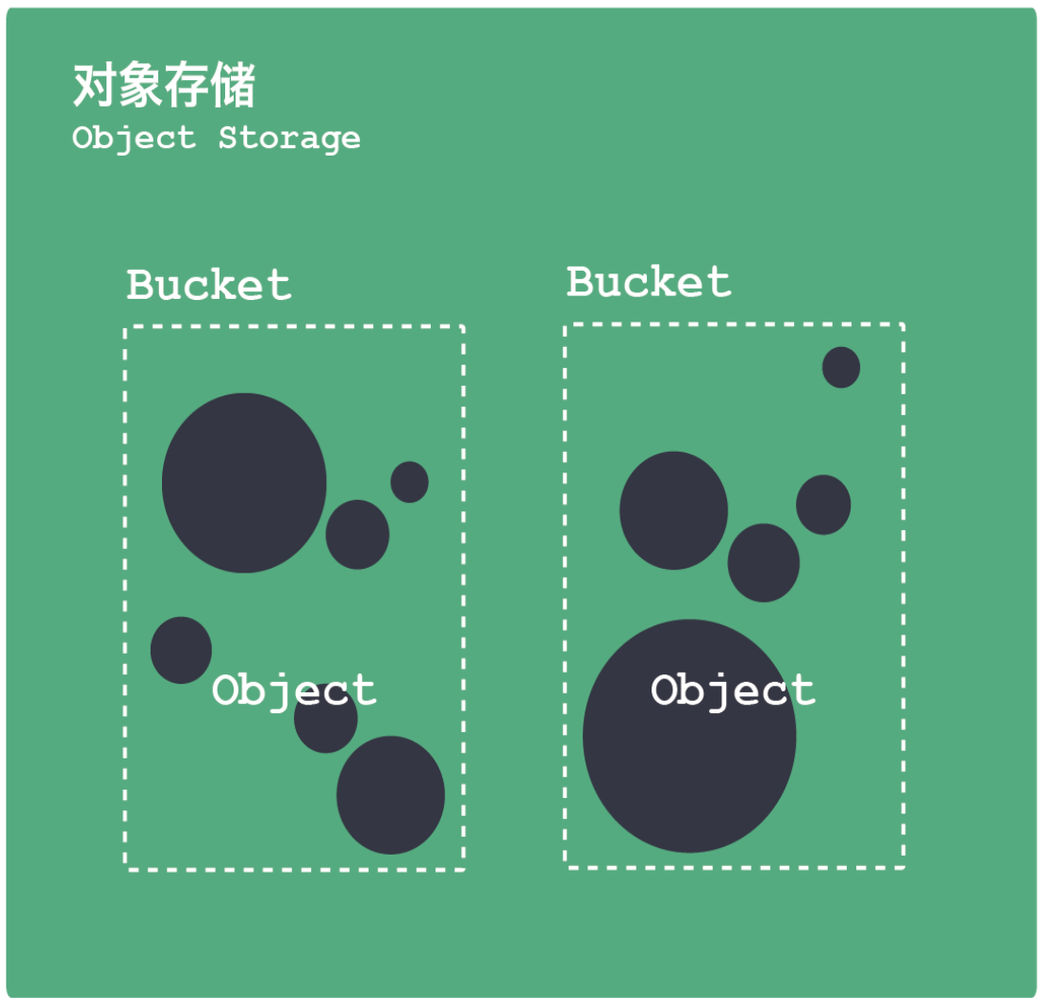
块存储
Block Storage

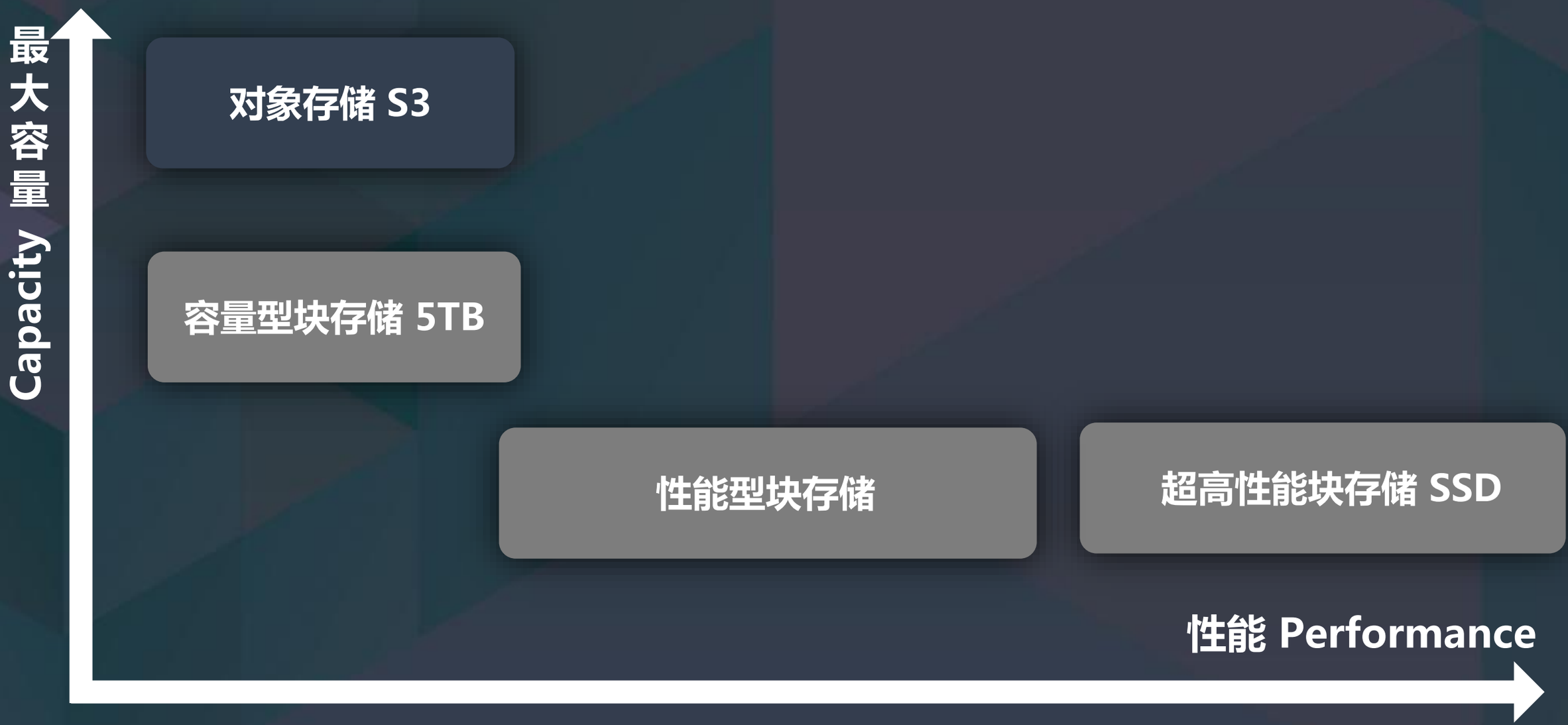
- 高性能 High Performance (Image / Swap)
- 大容量 Large Capacity (Snapshot)

共享存储
Shared Storage

- Virtual SAN
- NAS (nfs/SMB/CIFS)

对象存储
Object Storage




最大容量
Capacity
对象存储 S3
容量型块存储 5TB
性能型块存储
超高性能块存储 SSD
性能 Performance

多维块存储解决方案

对象存储

超高性能块存储

容量型块存储

性能型块存储

配合传统企业解决方案

Virtual SAN / NAS

数据备份解决方案

备份 Snapshot

青云存储设计原则

- 存储域与计算域融合还是分离？
- 运维优先
- 性能、规模、成本的平衡

存储域与计算域融合还是分离?

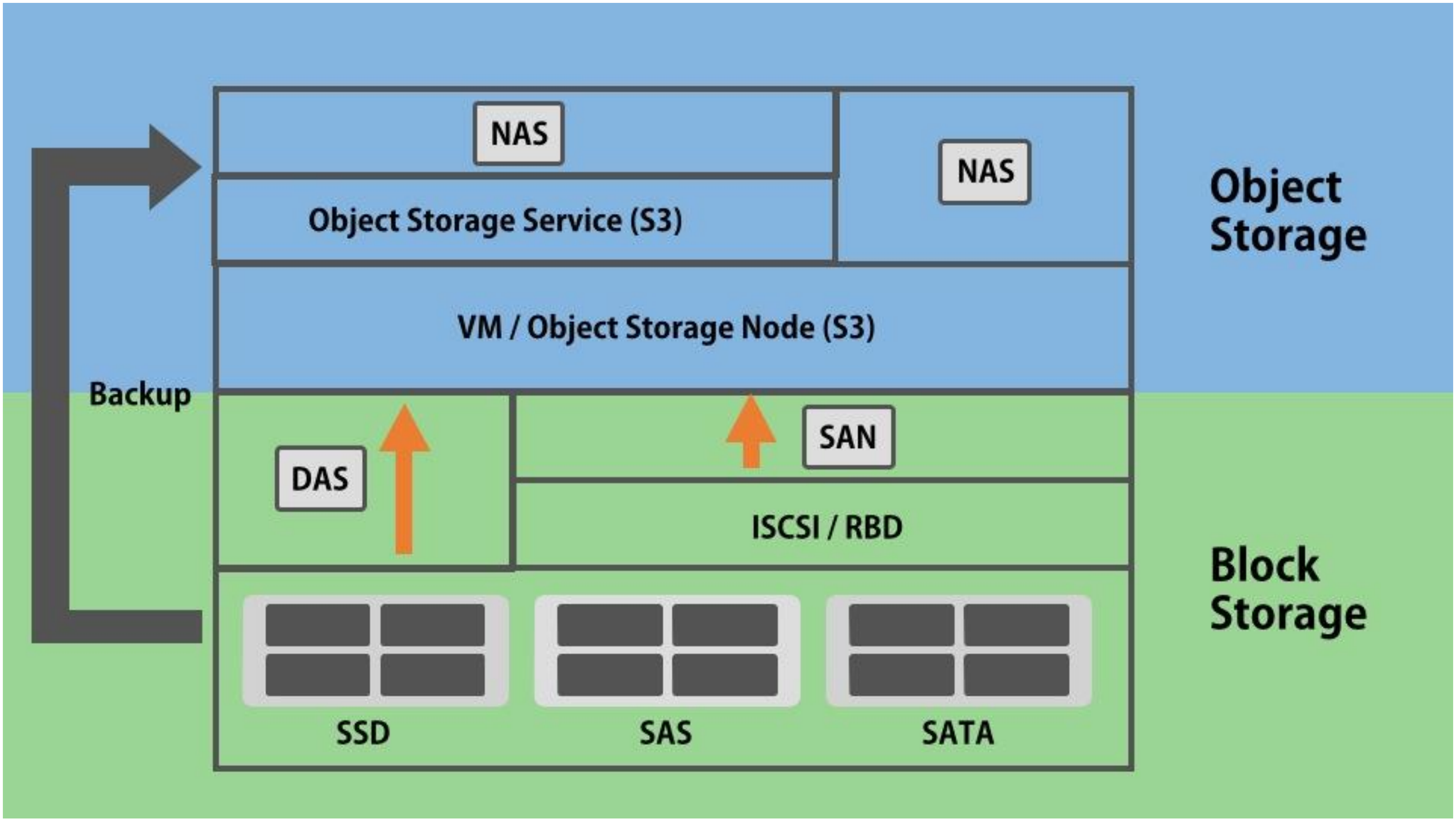
分离的好处：

- **架构简洁**
- **容易迁移**
- **各个域解决自己的问题**
- **有现成的解决方案**

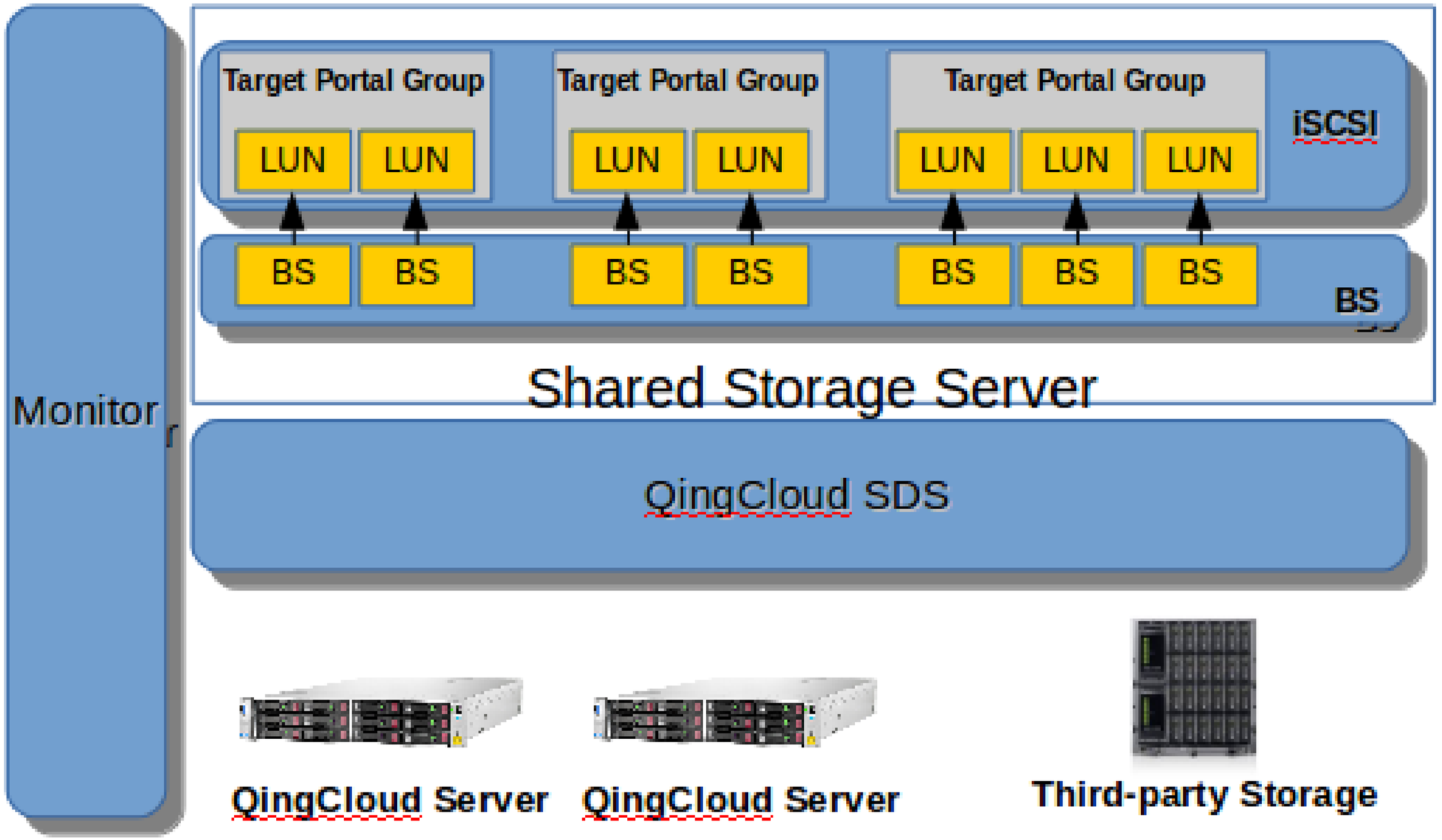
融合的好处：

- **IO路径短**
- **对网络依赖少**
- **物理拓扑简单**

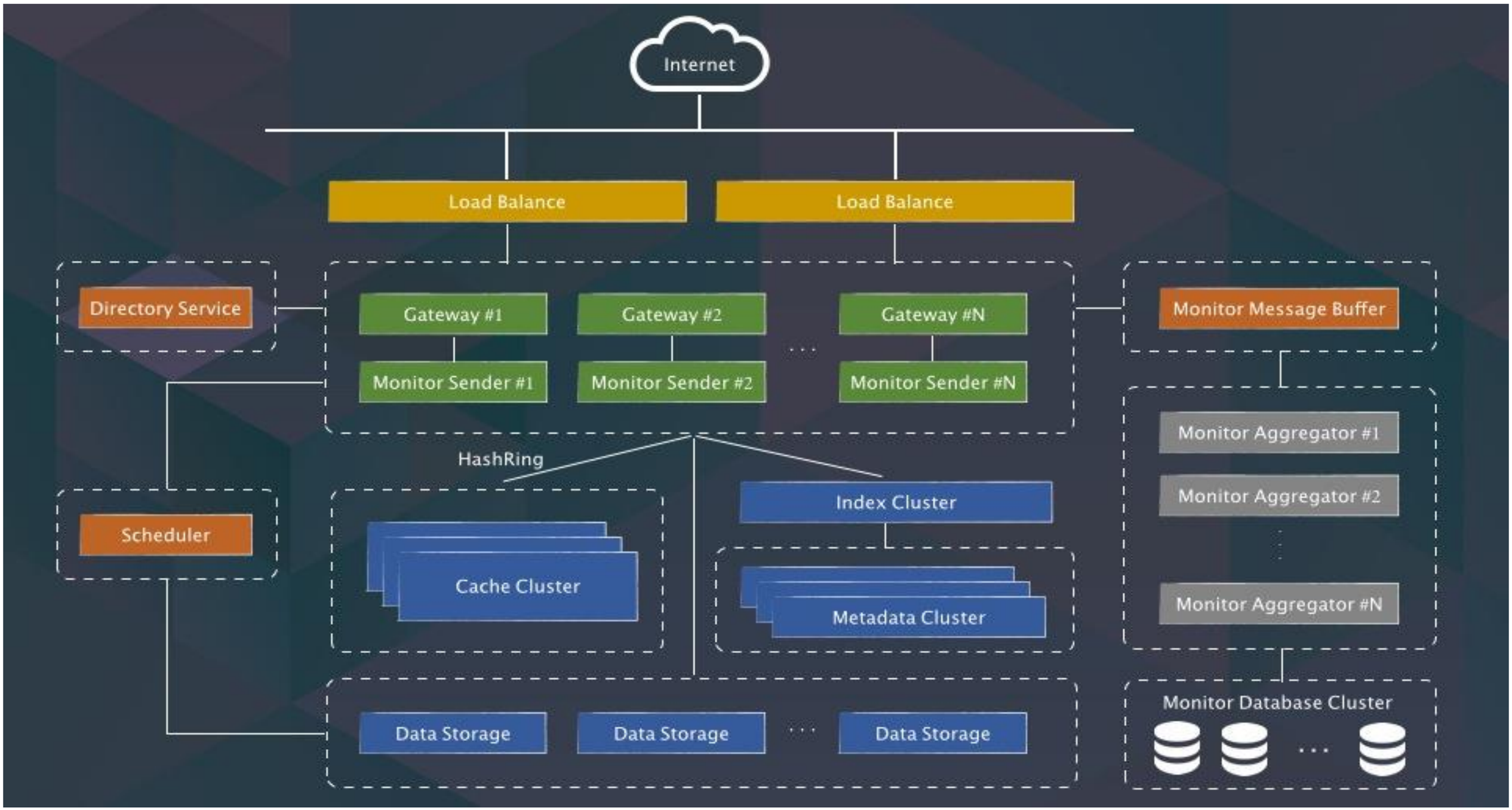
青云存储整体架构


NAS
NAS
Object Storage Service (S3)
Object Storage
VM / Object Storage Node (S3)
Backup
DAS
SAN
ISCSI / RBD
Block Storage
SSD
SAS
SATA

共享存储架构


Target Portal Group
Target Portal Group
Target Portal Group
iSCSI
LUN
LUN
LUN
LUN
LUN
LUN
LUN
BS
BS
BS
BS
BS
BS
BS
BS
Shared Storage Server
Monitor
QingCloud SDS
QingCloud Server
QingCloud Server
Third-party Storage

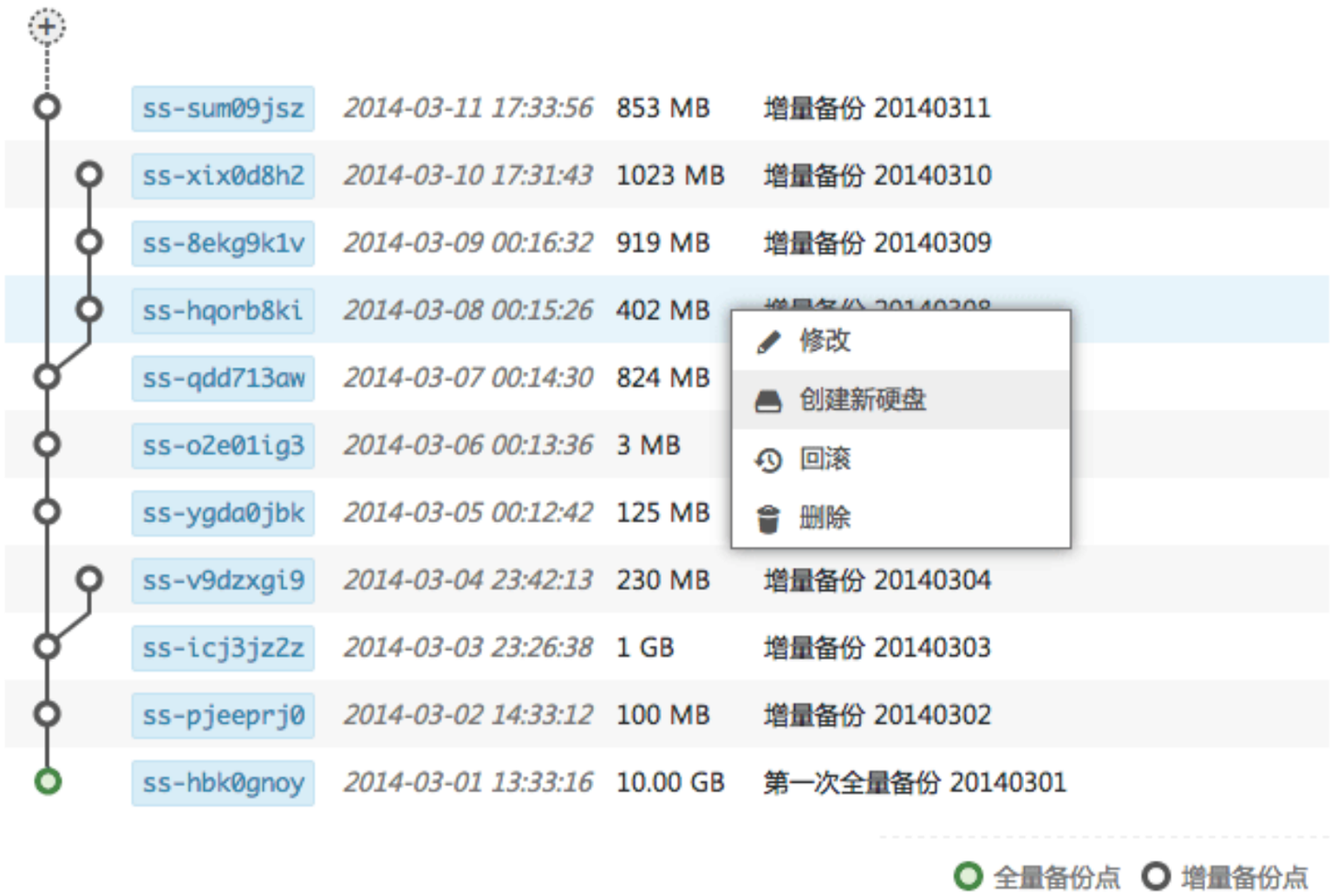
对象存储架构


Internet
Load Balance
Load Balance
Directory Service
Gateway #1
Gateway #2
Gateway #N
Monitor Sender #1
Monitor Sender #2
Monitor Sender #N
Monitor Message Buffer
Monitor Aggregator #1
Monitor Aggregator #2
Monitor Aggregator #N
Scheduler
HashRing
Cache Cluster
Index Cluster
Metadata Cluster
Data Storage
Data Storage
Data Storage
Monitor Database Cluster

备份与恢复

Snapshot

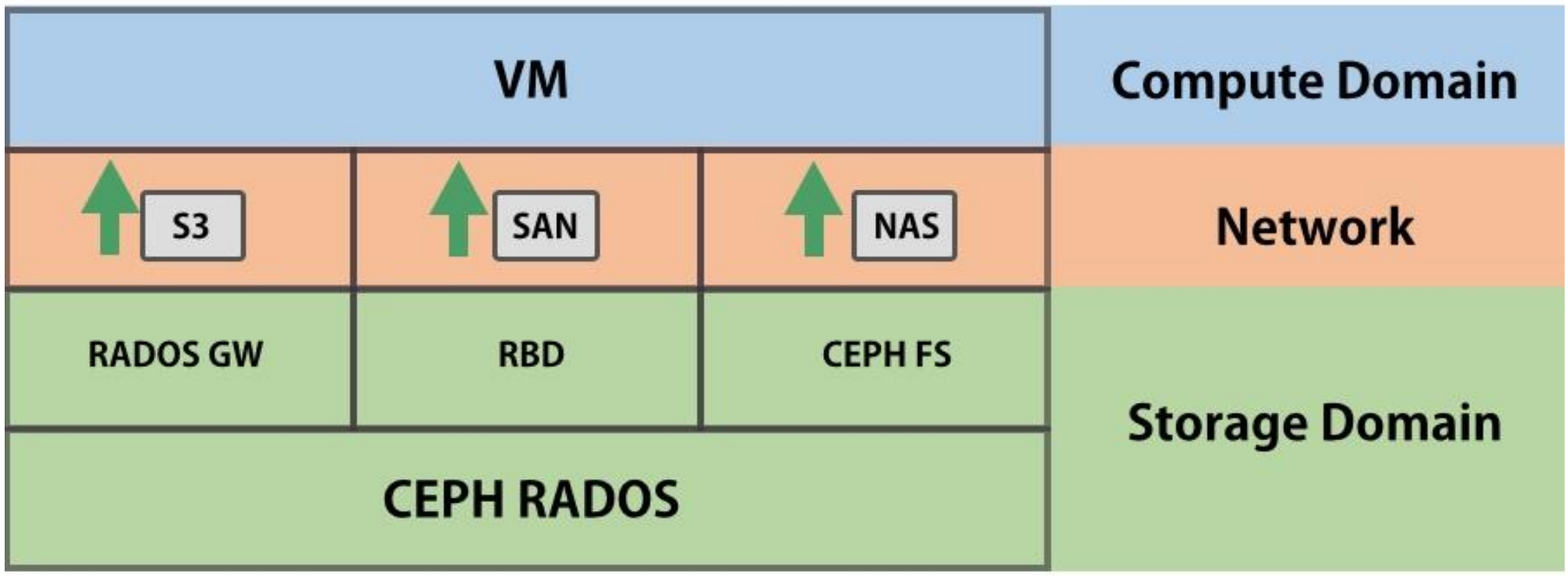
- 与实时副本的区别
- 全量 & 增量
- Offline & Online
- 多张盘并行备份
- 备份链
- 备份回滚
- 备份导出
- 定时备份



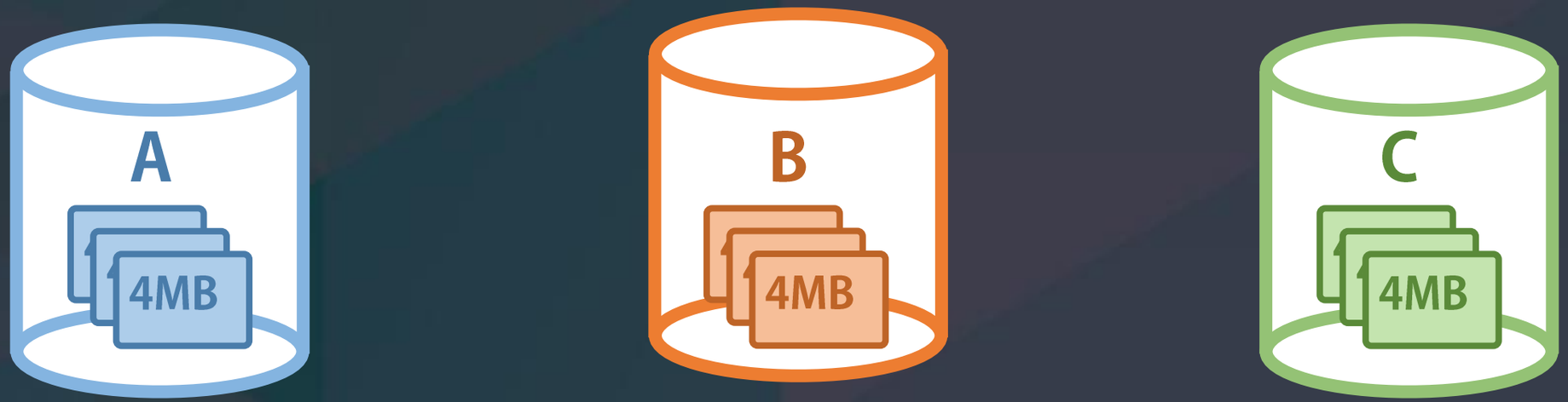
About CEPH

- 原理
- 风险扩散
- 性能低下
- 弥补性能损失 —> 高成本
- S3 架构问题
- No RAID?

About CEPH


VM
Compute Domain
S3
SAN
NAS
Network
RADOS GW
RBD
CEPH FS
Storage Domain
CEPH RADOS

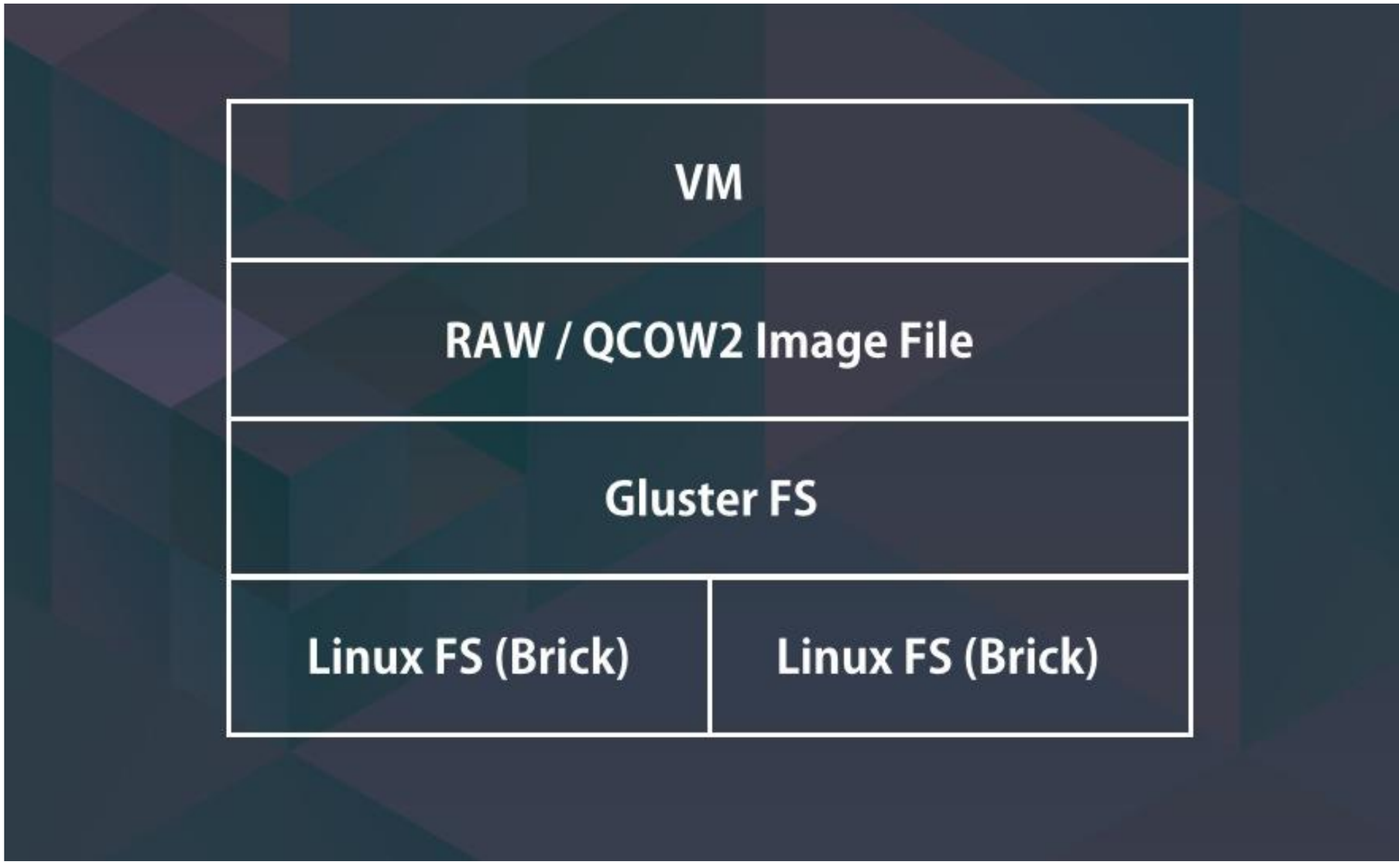
About CEPH



Box1 Box2 Box3

About GlusterFS

- 基于文件系统
- No Stripe
- Split Brain
- 集群规模限制



About EBS

- **存储容量与性能是否有必然关系?**

for workloads where data is accessed infrequently, and scenarios where the lowest storage is important. **Magnetic volumes provide approximately 100 IOPS on average, with an ability to burst to hundreds of IOPS.**

development and test environments, and boot volumes. General Purpose (SSD) volumes are designed to offer single digit millisecond latencies, **deliver a consistent baseline performance of 3 IOPS/GB to a maximum of 10,000 IOPS**,
and provide up to 160 MBps of

QINGCLOUD青云

Thank you.

reno@yunify.com