清华同方
TSINGHUA TONGFANG

密　　级：　外部公开

文档版本：　V1.2

产品技术白皮书

同方超强 RS6800 系列存储系统

版　　本	V1.2
发 布 日 期	2022 年 05 月 27 日
生 效 日 期	2022 年 05 月 27 日
拟　　制	张鹏

商标声明

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受同方股份有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，同方股份有限公司对本文档内容不做任何明示或默示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

同方股份有限公司

地址：北京市海淀区王庄路 1 号清华同方科技大厦 邮编：100083

网址：http://www.thtf.com.cn/

客户服务电话：400 922 0586

版本更新说明

版本号	生效日期	创建/变更人	变更说明
V1.0	2021 年 4 月 20 日	付君	首发
V1.1	2022 年 05 月 27 日	张鹏	修订
V1.2	2022 年 09 月 27 日	张鹏	修订

目 录

1 产品概述

同方超强 RS6800 系列高端智能全闪存存储系统是面向企业关键应用的存储产品，为企业核心业务提供卓越的数据服务。

同方超强 RS6800 系列高端全闪存，继承同方超强高端存储丰富的企业级特性，并凭借全互联架构、SAN 与 NAS 一体化双活等更先进的可靠性技术，全面保障用户关键业务连续。同时通过高效能的硬件设计、端到端的闪存提速、以及智能的管理，为企业提供最高水平的数据存储服务，满足大型数据库 OLTP/OLAP、云计算等各种应用的数据存储需求，是企业核心应用的信赖之选，广泛适用于金融、政府、运营商、制造、交通等行业。

除此之外，同方超强 RS6800 系列存储系统还提供易于使用的管理方式和方便快捷的本地/远程维护方式，大大降低了设备管理和维护的成本。

2 产品特点

2.1 深度融合：融会贯通，提升数据服务效率

- SAN 与 NAS 的融合：同方超强 RS6800 系列存储可提供 SAN 和 NAS 两种服务，满足业务弹性发展需求，提升存储资源利用率，并有效地降低 TCO。SAN 服务与 NAS 服务由底层存储资源池直接提供，缩短了存储资源的访问路径，从而保证两种服务（SAN 与 NAS）的性能、功能业界领先。
- 存储资源池的融合：通过内置异构虚拟化功能，同方超强 RS6800 系列存储能高效接管其它主流厂商存储阵列，并整合成统一的资源池，消除数据孤岛，资源可统一管理，自动化&服务编排；同时，还可以实现第三方设备迁移“0”中断，迁移操作工具化自动完成，耗时平均缩短 60%。
- 多数据中心的融合：同方超强 RS6800 系列存储提供免网关 SAN 与 NAS 一体化双活方案实现跨数据中心的业务永续；支持从双活数据中心平滑升级到 3DC，提供两地三中心级别的业务连续性；可实现 64:1 的多级 DC，提供数据集中容灾与保障。
- 云的融合：提供存储混合云解决方案，通过云上云下资源协同和数据流动，实现私有云和公有云间的数据容灾，助力企业存储服务平滑向云化转型。

2.2 卓越性能：满足企业业务的性能弹性增长需求

- 面向闪存的存储架构：同方超强 RS6800 系列存储采用面向闪存的系统架构，基于闪存融合技术在智能 CPU 多核优化、资源调度算法、cache 自适应算法以及 IO 智能调度等闪存优化设计，确保存储系统处理大量业务访问时依然能够提供稳定低于 1ms 的 I/O 快速响应，保证用户关键应用的卓越性能体验。
- 匹配闪存设计的领先规格：同方超强 RS6800 系列存储同一控制框内控制器之间采用 PCI-E 高速互联，并可通过 RDMA 级联多个控制框，支持 32Gbps FC/100Gbps Ethernet 等主机接口；后端采用 NVMe over Fabric（100Gbps）/SAS 3.0（12Gbps）高速接口，满足高性能、高带宽应用场

景所需。

- 灵活的扩展性：同方超强 RS6800 系列存储可线性扩展系统资源，能够平滑扩展至最大 32 个控制器、32TB 缓存，6400 块 SSD，使得性能及规格全面领先，实现 600 万 IOPS@1ms 性能值，满足用户未来业务高速增长的数据需求，帮助用户提升收益。同时可利用多个控制器并发加速同一主机业务，消除单控制器性能瓶颈，实现性能加倍。
- 全面效率提升方案：同方超强 RS6800 系列存储提供 TFCQ 系列增值功能，包括 SSD 缓存，数据分级（包括块和文件级）、服务质量控制、缓存分区、自动精简配置、文件配额等多样的效率提升方案，应用效率大幅度提升。

2.3 稳定可靠：从产品到方案实现 99.9999%高可用

- 多控制器负载均衡：同方超强 RS6800 系列存储提供 Active-Active 双活架构实现多个控制器间负载均衡技术，消除单点故障，实现系统高可用，保护业务稳定在线。
- 独有的数据快速恢复技术：采用创新的块级虚拟化技术，1TB 数据重构时间从 10 个小时降低到 30 分钟，与传统存储相比，因硬盘故障引起的数据失效风险降低 95%。
- 丰富的数据保护方案：同方超强 RS6800 系列存储系统数据保护特性包含快照、克隆、一体化备份、远程复制等数据保护技术，可以实现用户系统内、本地、异地以及多地的数据保护方案，实现 99.9999%的可用性，最大程度保障用户业务连续性和数据可用性。
- 领先的 SAN 与 NAS 一体化双活保护：同方超强 RS6800 系列存储支持 SAN 与 NAS 一体化双活，确保数据库与文件业务同时高可用。凭借 Hyper Metro A-A 双活，存储系统间可实现负载均衡的双活镜像以及无中断的跨站点接管，保障用户关键应用数据零丢失，业务零中断，让用户的核心应用系统不受宕机困扰。同时采用免网关设计可有效降低用户购置成本和部署复杂度，单套设备可平滑升级到双活，并能进一步扩展至两地三中心解决方案。

3 硬件架构

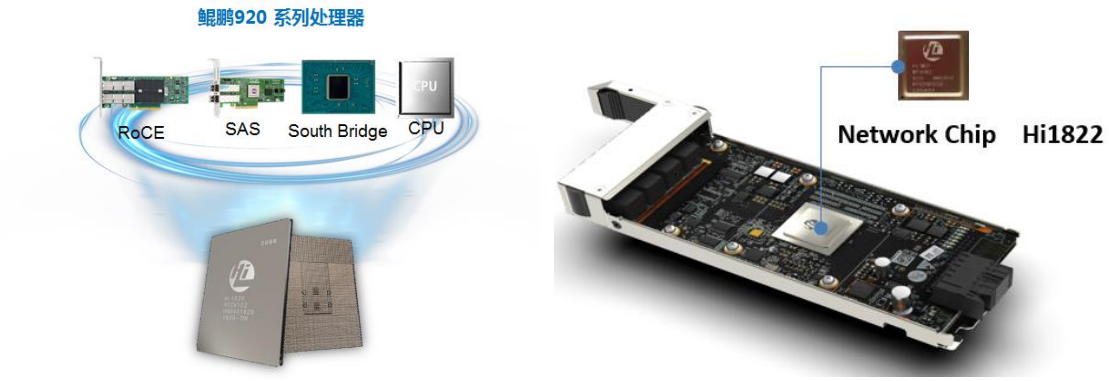
同方超强 RS6800 系列存储系统采用智能矩阵式多控架构，以控制框为单位横向扩展，达到性能和容量的线性增长。单个控制框采用双控冗余架构，双控间采用 100Gbps RDMA 高速总线实现双控缓存镜像通道，多控制框之间通过 25Gbps RDMA 直连或 100Gbps RDMA 无损交换机实现 Scale-out。控制框或硬盘框内的硬盘通过双端口连接到两个控制器。通过 BBU（Backup Battery Unit），在系统掉电时把 Cache 中的缓存数据持久化到数据保险箱上实现缓存数据的保护。

3.1 关键芯片设计

同方超强 RS6800 系列存储产品使用国产鲲鹏 CPU 处理器 96 核心，最高主频 2.6GHz。除了中央处理器的功能，还集成了 100G RoCE 网络芯片、SAS 启动器芯片、南桥芯片的能力，一颗处理器可以提供 100G RDMA 连接智能硬盘框、SAS 协议连接 SAS 硬盘框以及南桥连接存储管理口、串口等硬件能力，使同方超强 RS6800 系列的硬件设计极简，功耗降低。

CPU 处理器支持 RAID、DIF、重删算法(SHA-256)、压缩算法(Gzip/ZLib)、加密算法(AES256)、中国国产密码算法(SM3/SM4)的硬核算法实现，在存储使用这些算法时不需要软件参与，可直接由处理器进行硬件卸载。

图1-1 鲲鹏 920 系列处理器 与 网络芯片



同方超强 RS6800 系列存储产品全系列支持 IO 融合接口卡，其采用融合网络处理芯片，基于该款高性能芯片可以支持最大 32G 速率的 FC 接口或者最大 100GE 速率的以太网接口。

IO 融合接口卡的 FC/以太网接口支持 FastWrite 模式，可以将传统传输的 4 次握手交互，缩减为 2 次，从而在远距离数据传输时可以大幅缩短传输时延，推荐在部署远距离双活、复制特性时开启 FastWrite 模式。

IO 接口卡的以太网接口支持 TCP/IP Offload Engine(TOE)功能，可以由接口卡芯片处理 TCP/IP 协议的解析而不需要主 CPU 参与，通过释放一定 CPU 处理能力提升存储整体性能。

图1-2 SSD 芯片

同方超强 RS6800 系列“芯”系列存储产品使用的 NVMe SSD 及 SAS SSD，同时支持 NVMe 协议及 SAS 协议接口，并实现 SSD FTL 的硬件处理缩短 IO 处理时延。

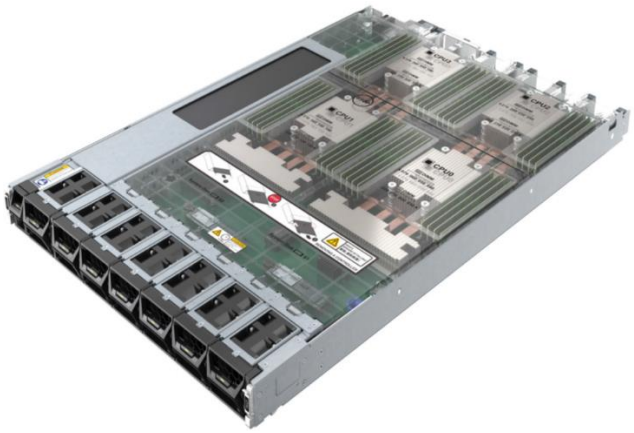
3.2 控制框

同方超强 RS6800 系列高端存储系统采用智能矩阵式多控架构，以控制框为单位横向扩展，达到性能和容量的线性增长。单个控制框采用四控冗余，前后端全互联共享架构，主机通过 FC 前端全共享接口卡支持四控共享访问，通过 SAS 3.0（12Gbps）或 RDMA（100Gbps）全共享接口卡支持后端四控共享访问，通过 Scale-Out 全共享接口卡支持引擎间免交换机直连共享访问。前端接口模块、控制器模块、后端接口模块、电源模块、BBU 模块、风扇模块、硬盘单元等所有 FRU 不存在任何单点故障，支持 2 个或者 2 个以上 FRU 冗余。各种 FRU 均支持在线热插拔，支持在线可更换。

同方超强 RS6800 系列高端存储系统采用 4U Acitve-Active 四控冗余高密架构设计，每个控制框支持 4 个控制器和 2 个控制器 2 种方式，每个控制器最大支持 4 个处理器，每个控制框最大支持 16 个处理器，控制器的整体结构如图 1-

3 所示。

图1-3 控制框整体结构



控制框前视图如图 1-4 所示。

图1-4 控制框前视图

控制框后视图如图 1-5 所示。

图1-5 控制框后视图

同方超强 RS6800 系列高端存储系统控制框采用盘控分离形态，每个控制框支持 28 张四控全互联共享访问的可热插拔 IO 模块，其中：

- 前端主机 IO 模块包括 4 端口 8G/16G/32G FC 接口模块，4 端口 10GE/25GE ETH 接口模块，2 端口 40GE/100GE 接口模块；
- Scale-Out 接口模块包括 2 端口 100Gb RDMA 接口模块；
- 后端接口模块包括 4 端口 12Gb SAS 接口模块（接入 SAS 硬盘框），2 端口 100Gb RDMA 接口模块（接入智能 SAS/NVMe 硬盘框）；

同方超强 RS6800 系列高端存储系统支持 3 种类型硬盘框分别如下：

- SAS 硬盘框；
- 智能 SAS 硬盘框；
- 智能 NVMe 硬盘框；

同方超强 RS6800 系列高端存储系统每个控制框有 2 个管理模块，每个管理模块有 3 个 GE（维护/管理网）口，1 个 USB 口，1 个串口。

3.3 接口模块

本章对接口模块的功能、外观和指示灯状态信息进行了描述。各个产品型号具体支持的可热插拔接口模块类型请参见硬件规格。

3.3.1. GE 电接口模块

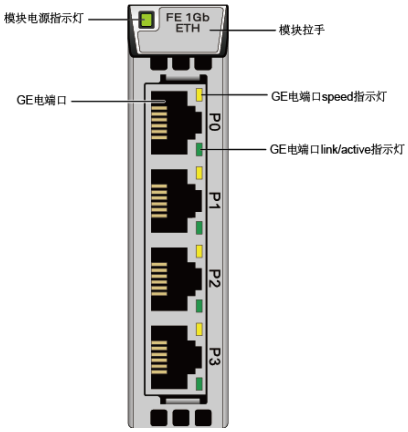
1. 功能

GE 电接口模块提供 4 个传输速率为 1Gbit/s 的电端口。不支持降速至 100Mbit/s 和 10Mbit/s。

2. 接口

GE 电接口模块如图 1-6 所示。标签中的 FE 表示 Front-end。

图1-6 GE 电接口模块



3. 指示灯

存储设备上电后，GE 电接口模块指示灯说明如表 1-1 所示。

表1-1 GE 电接口模块指示灯说明

指示灯	状态和说明
模块电源指示灯	• 绿色，亮：模块正常运行。 • 绿色，闪烁（2Hz）：模块有热插拔请求。 • 黄色，亮：模块故障。 • 灭：模块未上电或可热插拔。
端口 link/active 指示灯	• 绿色，亮：与应用服务器连接正常。 • 绿色，闪烁（2Hz）：正在传输数据。 • 灭：与应用服务器连接异常。
端口 speed 指示灯	• 黄色，常亮：速率为最高速率。 • 灭：速率为非最高速率。

3.3.2. 25Gb RDMA 接口模块

25Gb RDMA 接口模块主要用于直连组网中控制框之间的连接。

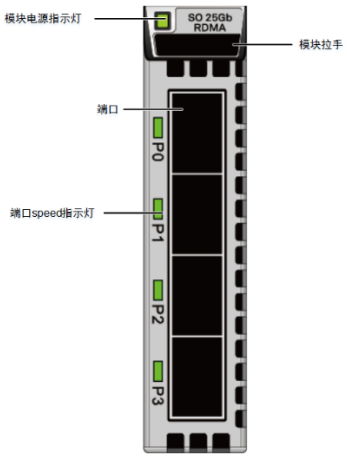
1. 功能

25Gb RDMA 接口模块提供 4 个传输速率为 25Gbit/s 的光口。

2. 接口

25Gb RDMA 接口模块如图 1-7 所示，其中标签中的 SO 表示 Scale-Out。

图1-7 25Gb RDMA 接口模块



3. 指示灯

存储设备上电后，25Gb RDMA 接口模块指示灯说明如表 1-2 所示。

表1-2 25Gb RDMA 接口模块指示灯说明

指示灯	状态和说明
模块电源指示灯	• 绿色，常亮：模块正常运行。 • 绿色，闪烁：模块有热插拔请求。 • 黄色，常亮：模块故障。 • 灭：模块未上电或可热插拔。
端口 speed 指示灯	• 蓝色，常亮：速率为最高速率。 • 蓝色，闪烁（2Hz）：速率为最高速率，正在传输数据。 • 绿色，常亮：速率为非最高速率。 • 绿色，闪烁（2Hz）：速率为非最高速率，正在传输数据。 • 黄色，常亮：端口光模块、线缆故障或者插入端口不支持的光模块、线缆。 • 灭：端口链路无连接。

3.3.3. 40GE 接口模块

40GE 接口模块主要用于存储设备与应用服务器之间的连接。

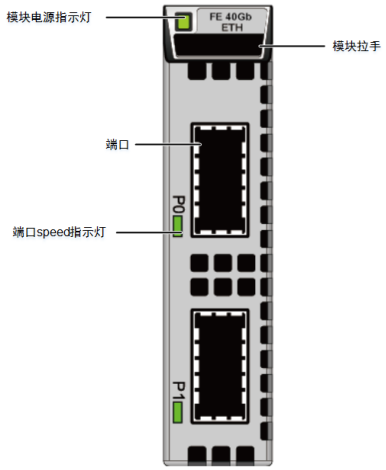
1. 功能

40GE 接口模块提供 2 个传输速率为 40Gbit/s 的光口。

2. 接口

40GE 接口模块如图 1-8 所示，其中标签中的 FE 表示 Front-end。

图1-8 40GE 接口模块



3. 指示灯

存储设备上电后，40GE 接口模块指示灯说明如表 1-3 所示。

表1-3 40GE 接口模块指示灯说明

指示灯	状态和说明
模块电源指示灯	• 绿色，常亮：模块正常运行。 • 绿色，闪烁：模块有热插拔请求。 • 黄色，常亮：模块故障。 • 灭：模块未上电或可热插拔。
端口 speed 指示灯	• 蓝色，常亮：速率为最高速率。 • 蓝色，闪烁（2Hz）：速率为最高速率，正在传输数据。 • 绿色，常亮：速率为非最高速率。 • 绿色，闪烁（2Hz）：速率为非最高速率，正在传输数据。 • 黄色，常亮：端口光模块、线缆故障或者插入端口不支持的光模块、线缆。 • 灭：端口链路无连接。

3.3.4. 100GE 接口模块

100GE 接口模块主要用于存储设备与应用服务器之间的连接。

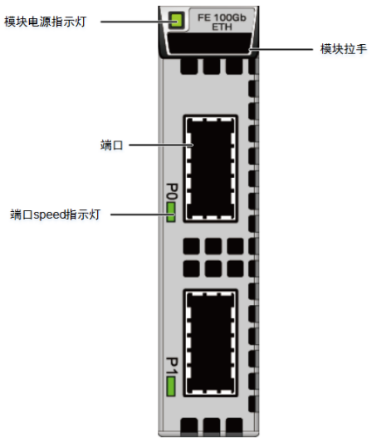
1. 功能

100GE 接口模块提供 2 个传输速率为 100Gbit/s 的光口。

2. 接口

100GE 接口模块如图 1-9 所示，其中标签中的 FE 表示 Front-end。

图1-9 100GE 接口模块



3. 指示灯

存储设备上电后，100GE 接口模块指示灯说明如表 1-4 所示。

表1-4 100GE 接口模块指示灯说明

指示灯	状态和说明
模块电源指示灯	• 绿色，常亮：模块正常运行。 • 绿色，闪烁：模块有热插拔请求。 • 黄色，常亮：模块故障。 • 灭：模块未上电或可热插拔。
端口 speed 指示灯	• 蓝色，常亮：速率为最高速率。 • 蓝色，闪烁（2Hz）：速率为最高速率，正在传输数据。 • 绿色，常亮：速率为非最高速率。 • 绿色，闪烁（2Hz）：速率为非最高速率，正在传输数据。 • 黄色，常亮：端口光模块、线缆故障或者插入端口不支持的光模块、线缆。 • 灭：端口链路无连接。

3.3.5. 100Gb RDMA 接口模块

100Gb RDMA 接口模块主要是多控制器间 Scale-out，通过交换机相连或直接连接。或用于智能 SAS 硬盘框组网连接。

1. 功能

100Gb RDMA 接口模块提供 2 个传输速率为 100Gbit/s 的光口。

2. 接口

100Gb RDMA 接口模块如图 1-10 和图 1-11 所示，其中标签中的 SO 表示 Scale-Out，BE 表示 Back-end。

图1-10 SO 100Gb RDMA 接口模块

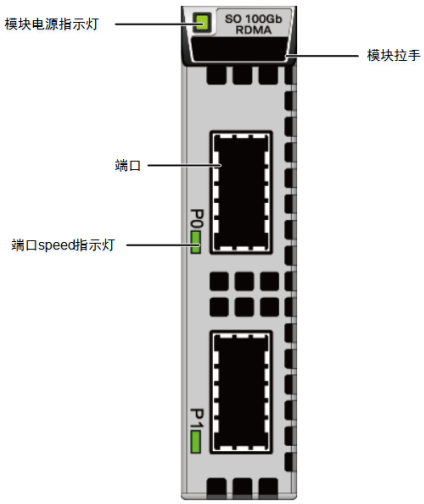
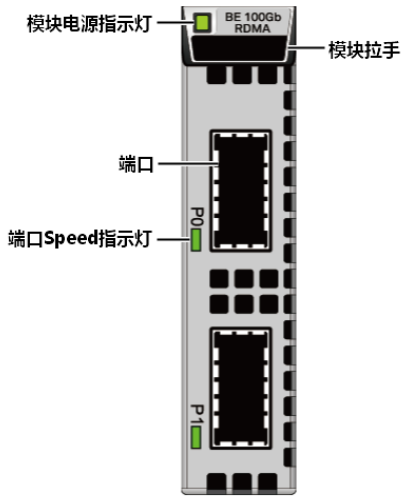


图1-11 BE 100Gb RDMA 接口模块



3. 指示灯

存储设备上电后，100Gb RDMA 接口模块指示灯说明如表 1-5 所示。

表1-5 100GE 接口模块指示灯说明

指示灯	状态和说明
模块电源指示灯	• 绿色，常亮：模块正常运行。 • 绿色，闪烁：模块有热插拔请求。 • 黄色，常亮：模块故障。 • 灭：模块未上电或可热插拔。
端口 speed 指示灯	• 蓝色，常亮：速率为最高速率。 • 蓝色，闪烁（2Hz）：速率为最高速率，正在传输数据。 • 绿色，常亮：速率为非最高速率。 • 绿色，闪烁（2Hz）：速率为非最高速率，正在传输数据。 • 黄色，常亮：端口光模块、线缆故障或者插入端口不支持的光模块、线缆。 • 灭：端口链路无连接。

3.3.6. TFCQ IO 接口模块

1. 功能

TFCQ IO 接口模块支持 8Gbit/s，10Gbit/s，16Gbit/s，25Gbit/s，32Gbit/s 五种速率的光模块。

2. 接口

速率为 8Gbit/s、10Gbit/s、16Gbit/s、25Gbit/s、32Gbit/s 的 TFCQ IO 接口模块如图 1-12、图 1-13、图 1-14、图 1-15、图 1-16 所示，其中标签中的 FE 表示 Front-end。

图1-12 速率为 8Gbit/s 的 TFCQ IO 接口模块

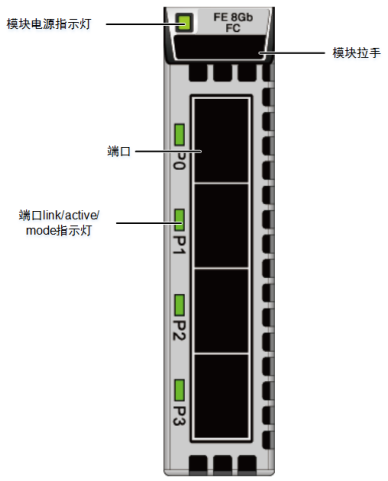


图1-13 速率为 10Gbit/s 的 TFCQ IO 接口模块

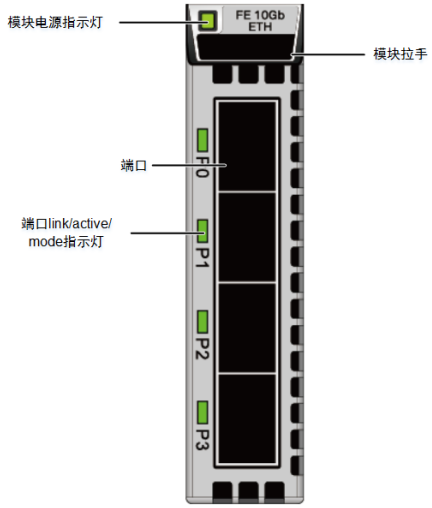


图1-14 速率为 16Gbit/s 的 TFCQ IO 接口模块

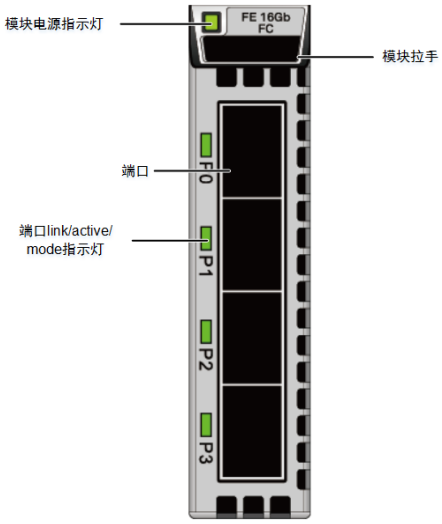


图1-15 速率为 25Gbit/s 的 TFCQ IO 接口模块

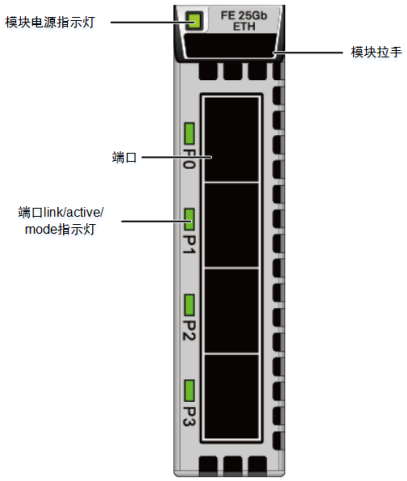
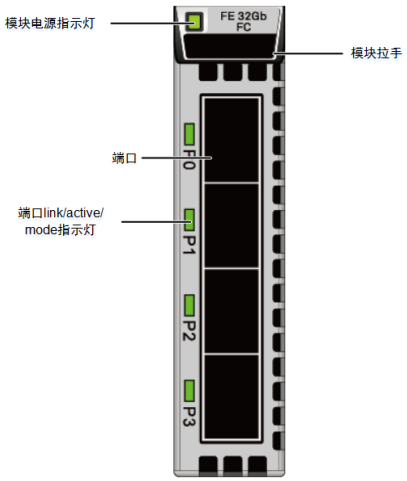


图1-16 速率为 32Gbit/s 的 TFCQ IO 接口模块



3. 指示灯

存储设备上电后 TFCQ IO 接口模块指示灯说明如表 1-6 所示。

表1-6 TFCQ IO 接口模块指示灯说明

指示灯	状态和说明
模块电源指示灯	• 绿色，常亮：模块正常运行。 • 绿色，闪烁：模块有热插拔请求。 • 黄色，常亮：模块故障。 • 灭：模块未上电或可热插拔。
端口 link/active/mode 指示灯	• 蓝色，常亮：当前工作在 FC 模式，端口链路正常，无数据传输。 • 蓝色，闪烁（2Hz）：当前工作在 FC 模式，正在传输数据。 • 绿色，常亮：当前工作在 ETH 模式，端口链路正常，无数据传输。 • 绿色，闪烁（2Hz）：当前工作在 ETH 模式，正在传输数据。 • 黄色，常亮：端口故障。 • 灭：端口未连接。

3.3.7. 12Gb SAS 级联模块

级联模块通过级联端口连接控制框和 SAS 硬盘框，是控制框和 SAS 硬盘框之间进行数据传输的连接点。

1. 功能

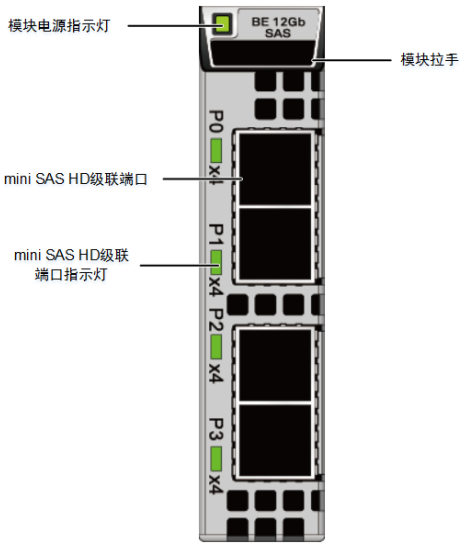
12Gb SAS 级联模块提供 4 个传输速率为 4×12Gbit/s 的 mini SAS HD 级联端口，用于级联硬盘框。SAS 级联模块通过 mini SAS HD 线缆与存储系统的后端

硬盘阵列连接。当连接的设备传输速率低于级联端口速率时，级联端口将自动适应传输速率，以保证数据传输通道的连通性和数据传输速率的一致性。

2. 接口

12Gb SAS 级联模块如图 1-17 所示，其中标签中的 BE 表示 Back-end。

图1-17 12Gb SAS 级联模块



3. 指示灯

存储设备上电后，12Gb SAS 级联模块指示灯说明如表 1-7 所示。

表1-7 12Gb SAS 级联模块指示灯说明

指示灯	状态和说明
模块电源指示灯	• 绿色，亮：模块正常运行。 • 绿色，闪烁：模块有热插拔请求。 • 黄色，亮：模块故障。 • 灭：模块未上电或可热插拔。
mini SAS HD 级联端口指示灯	• 蓝色，亮：端口传输速率为 4×12Gbit/s。 • 绿色，亮：端口传输速率为 4×6Gbit/s 或 4×3Gbit/s。 • 黄色，亮：端口出现故障。 • 灭：端口链路无连接。

3.3.8. 12Gb SAS V2 共享级联模块

级联模块通过级联端口连接控制框和 SAS 硬盘框，是控制框和 SAS 硬盘框之间进行数据传输的连接点。

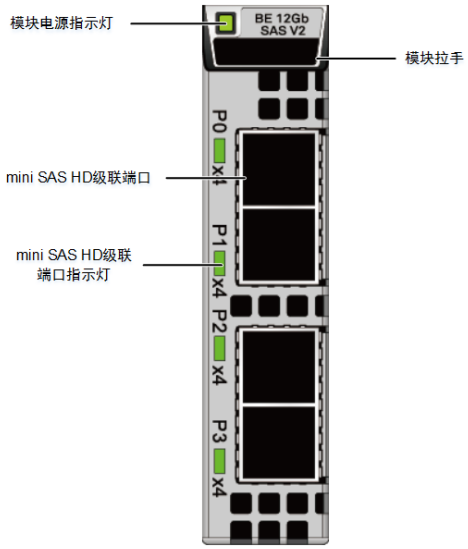
1. 功能

12Gb SAS V2 共享级联模块提供 4 个传输速率为 4×12Gbit/s 的 mini SAS HD 级联端口，用于级联 SAS 硬盘框。SAS 级联模块通过 mini SAS HD 线缆与存储系统的后端硬盘阵列连接。当连接的设备传输速率低于级联端口速率时，级联端口将自动适应传输速率，以保证数据传输通道的连通性和数据传输速率的一致性。

2. 接口

12Gb SAS V2 共享级联模块如图 1-18 所示，其中标签中的 BE 表示 Back-end，V2 主要用于与 12Gb SAS 级联模块进行区分。

图1-18 12Gb SAS V2 共享级联模块



3. 指示灯

存储设备上电后，12Gb SAS V2 共享级联模块指示灯说明如表 1-8 所示。

表1-8 12Gb SAS V2 共享级联模块指示灯说明

指示灯	状态和说明
模块电源指示灯	• 绿色，亮：模块正常运行。 • 绿色，闪烁：模块有热插拔请求。 • 黄色，亮：模块故障。 • 灭：模块未上电或可热插拔。
mini SAS HD 级联端口指示灯	• 蓝色，亮：端口传输速率为 4×12Gbit/s。 • 绿色，亮：端口传输速率为 4×6Gbit/s 或 4×3Gbit/s。

指示灯	状态和说明
	• 黄色，亮：端口出现故障。 • 灭：端口链路无连接。

3.4 SAS 硬盘框（2U，2.5 英寸硬盘）

本章将对硬盘框的硬件结构、各部件功能特性、前后视图及指示灯状态等详细信息进行描述。

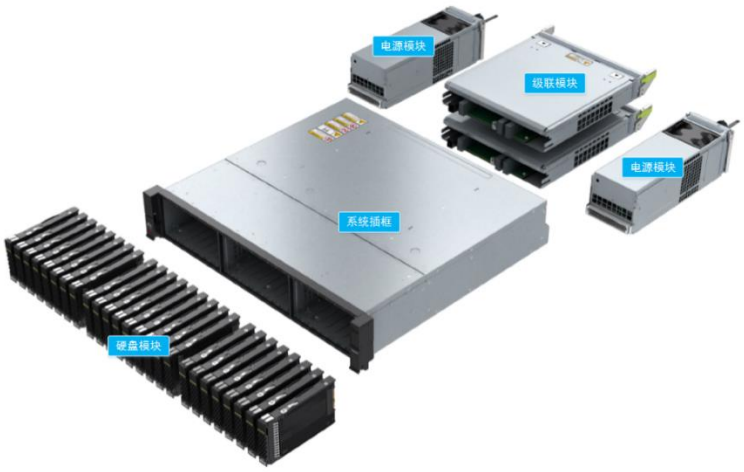
3.4.1. 概述

硬盘框采用部件模块化设计，主要由系统插框、级联模块、电源模块和硬盘模块组成。

1. 整体结构

2U 硬盘框的整体结构如图 1-19 所示。

图1-19 2U 硬盘框整体结构



2. 前视图

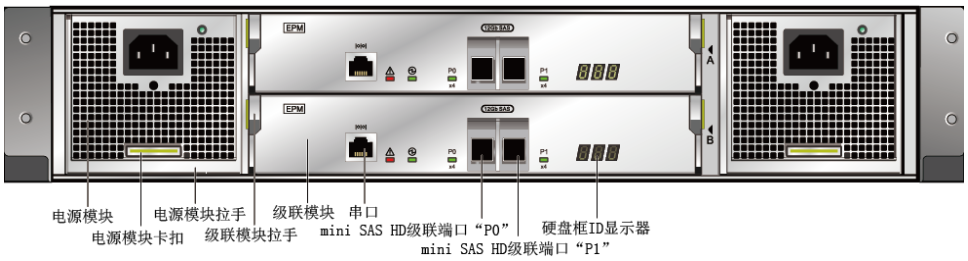
硬盘框前视图如图 1-20 所示。

图1-20 硬盘框前视图

3. 后视图

硬盘框后视图如图 1-21 所示。

图1-21 硬盘框后视图



3.4.2. 部件介绍

本节提供存储设备各部件的外观及接口信息。

1. 系统插框

系统插框通过背板为各种接口模块提供可靠的连接，实现各个模块之间的信号互连与电源互连。系统插框的外观如图 1-22 所示。

图1-22 系统插框

2. 级联模块

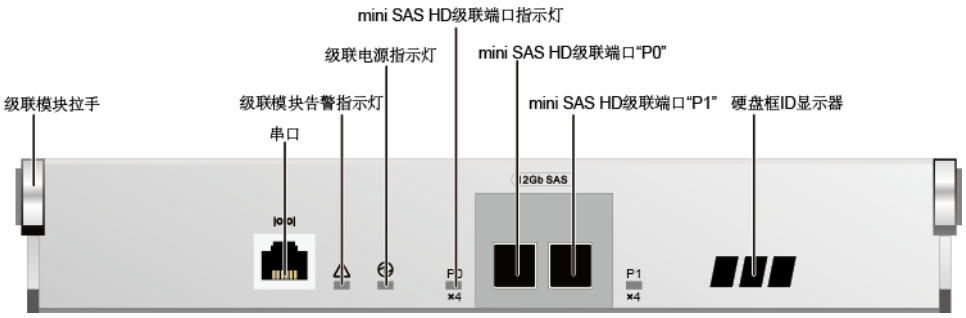
每个级联模块提供 1 个级联端口“P0”和 1 个级联端口“P1”。级联模块通过级联端口来级联控制器或硬盘框，实现与控制器或硬盘框的通信，是控制器与硬盘框之间进行数据传输的连接点。级联模块的外观如图 1-23 所示。

图1-23 级联模块

接口

级联模块的接口信息如图 1-24 所示。

图1-24 级联模块接口信息



指示灯

存储设备上电后，级联模块指示灯说明如表 1-9 所示。

表1-9 级联模块指示灯说明

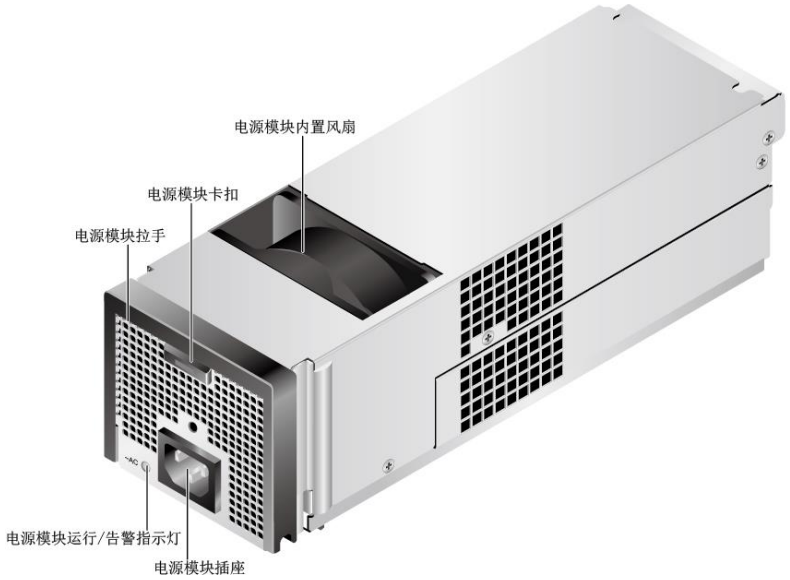
指示灯	状态和说明
级联模块告警指示灯	• 黄色，亮：级联模块出现告警。 • 灭：级联模块正常运行。

指示灯	状态和说明
级联模块电源指示灯	• 绿色，亮：级联模块正常上电。 • 灭：级联模块未上电。
mini SAS HD 级联端口指示灯	• 蓝色，亮：端口传输速率为 4×12Gbit/s。 • 绿色，亮：端口传输速率为 4×6Gbit/s 或 4×3Gbit/s。 • 黄色，亮：端口出现故障。 • 灭：端口链路无连接。

3. 电源模块

电源模块支持交流电源模块，可以支持硬盘框在最大功耗模式下正常运行。交流电源模块的外观如图 1-25 所示。

图1-25 交流电源模块



指示灯

存储设备上电后，电源模块指示灯说明如表 1-10 所示。

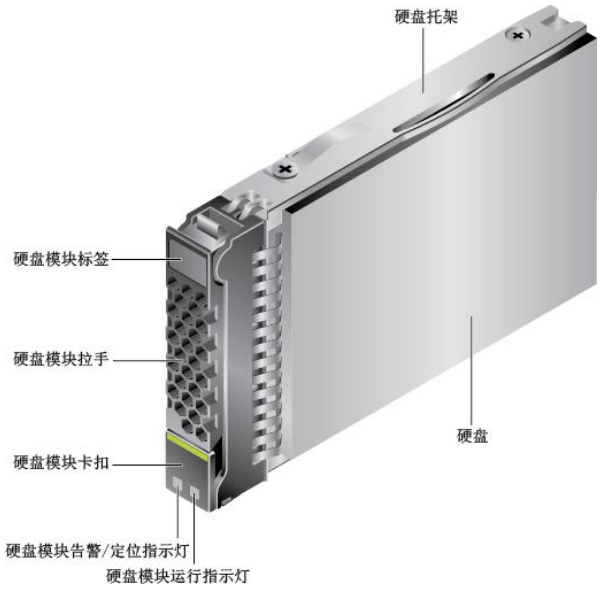
表1-10 电源模块指示灯说明

指示灯	状态和说明
电源/风扇模块运行/告警指示灯	• 绿色，亮：电源正常供电，未发生故障。 • 绿色，闪烁（1Hz）：电源输入正常，设备未上电。 • 绿色，闪烁（4Hz）：电源模块处于在线升级过程中。 • 黄色，亮：电源或风扇模块发生故障。 • 灭：无外部电源输入。

4. 硬盘模块

硬盘模块为存储系统提供存储容量，并且可以实现业务数据、系统数据和缓存数据的存储作用。硬盘模块的外观如图 1-26 所示。

图1-26 硬盘模块



指示灯

存储设备上电后，硬盘模块指示灯说明如表 1-11 所示。

表1-11 硬盘模块指示灯说明

指示灯	状态和说明
硬盘模块运行指示灯	• 绿色，亮：硬盘模块正常运行。 • 绿色，闪烁（4Hz）：硬盘模块正在读写数据。 • 灭：硬盘模块未上电或上电异常。
硬盘模块告警/定位指示灯	• 黄色，亮：硬盘模块出现故障。 • 黄色，闪烁（2Hz）：定位至此硬盘。 • 灭：硬盘模块正常运行或硬盘模块可插拔。

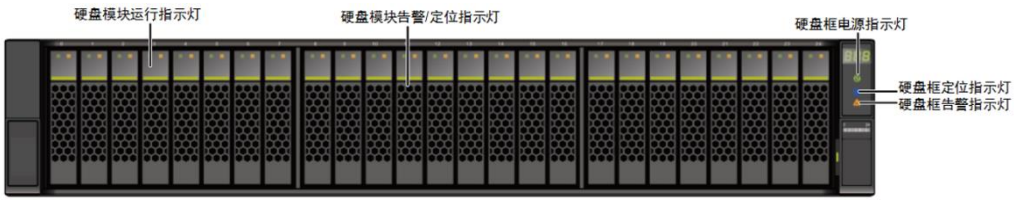
3.4.3. 指示灯介绍

硬盘框上电后，可以通过观察指示灯初步检查硬盘框当前的工作状态。

前面板指示灯

硬盘框前面板指示灯如图 1-27 所示。

图1-27 硬盘框前面板指示灯



硬盘框前面板指示灯说明如表 1-12 所示。

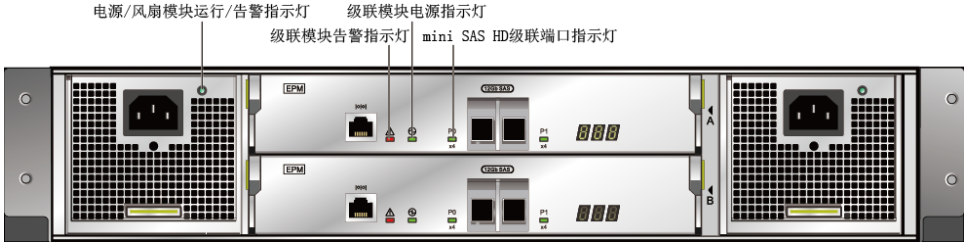
表1-12 硬盘框前面板指示灯说明

模块	指示灯类型	状态和说明
硬盘模块	硬盘模块运行指示灯	• 绿色，亮：硬盘模块正常运行。 • 绿色，闪烁（4Hz）：硬盘模块正在读写数据。 • 灭：硬盘模块未上电或上电异常。
	硬盘模块告警/定位指示灯	• 黄色，亮：硬盘模块出现故障。 • 黄色，闪烁（2Hz）：定位至此硬盘。 • 灭：硬盘模块正常运行或硬盘模块可插拔。
系统插框	硬盘框定位指示灯	• 蓝色，闪烁：硬盘框正在定位。 • 灭：硬盘框未定位。
	硬盘框告警指示灯	• 黄色，亮：硬盘框出现告警。 • 灭：硬盘框正常运行。
	硬盘框电源指示灯	• 绿色，亮：硬盘框已上电。 • 灭：硬盘框未上电。

后面板指示灯

硬盘框后面板指示灯如图 1-28 所示。

图1-28 硬盘框后面板指示灯



硬盘框后面板指示灯说明如表 1-13 所示。

表1-13 硬盘框后面板指示灯说明

模块	指示灯名称	状态和说明
级联模块	级联模块告警指示灯	• 黄色，亮：级联模块出现告警。 • 灭：级联模块正常运行。
	级联模块电源指示灯	• 绿色，亮：级联模块正常上电。 • 灭：级联模块未上电。
	mini SAS HD 级联端口指示灯	• 蓝色，亮：端口传输速率为 4×12Gbit/s。 • 绿色，亮：端口传输速率为 4×6Gbit/s 或 4×3Gbit/s。 • 黄色，亮：端口出现故障。 • 灭：端口链路无连接。
电源模块	电源/风扇模块运行/告警指示灯	• 绿色，亮：电源正常供电，未发生故障。 • 绿色，闪烁（1Hz）：电源输入正常，设备未上电。 • 绿色，闪烁（4Hz）：电源模块处于在线升级过程中。 • 黄色，亮：电源或风扇模块发生故障。 • 灭：无外部电源输入。

3.5 硬件全冗余

同方超强 RS6800 系列存储系统所有组件与通道均为全冗余设计，无单点故障，各组件与通道均可独立完成故障检测、修复和隔离，确保系统稳定运行。

表1-14 硬件部件全冗余

位置	系统部件	冗余情况	故障影响
控制框	控制器	1+1 冗余	性能按比例下降
	电源模块	1+1 冗余	无影响
	风扇模块	有冗余（不同产品型号冗余程度不同）	无影响
	BBU 模块	有冗余（不同产品型号冗余程度不同）	无影响
	接口卡	1+1 冗余	无影响
	管理板	1+1 冗余	无影响
硬盘框	级联板	1+1 冗余	无影响
	电源模块	1+1 冗余	无影响
	风扇模块	1+1 冗余	无影响

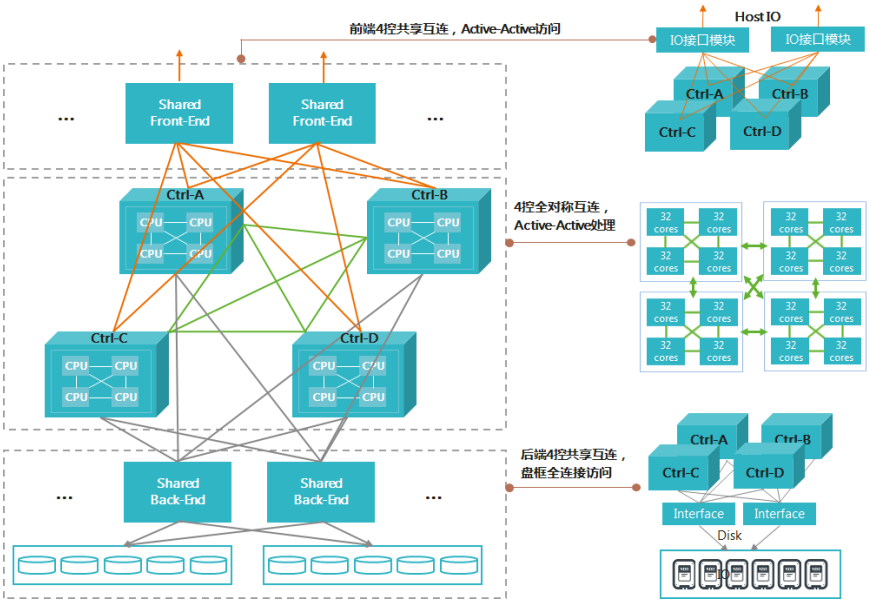
3.6 全互联均衡架构

同方超强 RS6800 系列高端存储系统采用全互联均衡架构，该架构采用了高速、矩阵式全互联无源背板，可以同时接入 4 个控制器节点以及 28 张接口模块，所有接口模块采用全共享方式接入背板，允许主机从任意端口接入，直达任意控制器进行处理，完全避免了传统高端存储在系统升级或者控制器故障时的单点运行状态，保证了关键应用的业务连续性。

3.6.1. 全互联的控制器

同方超强 RS6800 系列高端存储系统一个控制框内包含 4 个控制器，每个控制器是一个独立可热插拔服务处理单元，每个控制器出 3 对 RDMA 高速链路，连接到无源背板连接，与其他 3 个控制器全交叉连接。一个控制框内包含 28 张接口卡，每张接口卡提供 4 条 PCIe 链路同时连接到 4 个控制器节点，前端共享卡、控制器、后端共享卡三层通过无源背板全互联，控制器之间的数据流动能够在不经过第三方中转的情况下实现一次 RDMA 直达，从而达到均衡、快速、高效的访问效果。由于 4 个控制器位于同一个引擎内，不需要外部线缆和交换机，节省了组网步骤和线缆，部署更简单，不易出现人因差错。无源背板由于只使用了无源线路，可靠性极高。

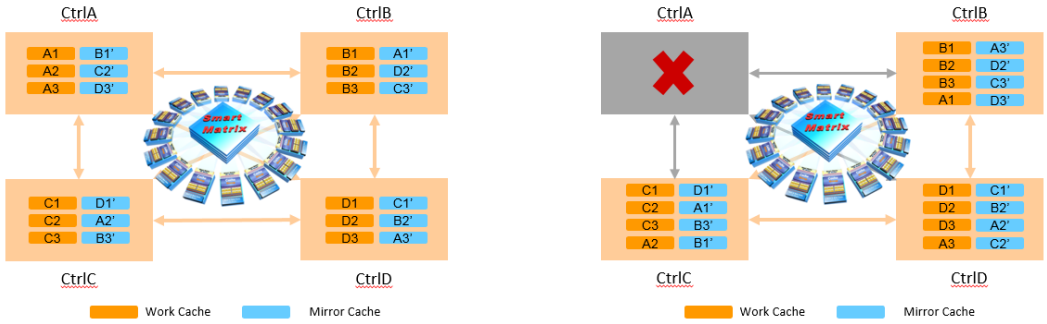
图1-29 全互联架构



同方超强 RS6800 系列高端存储系统基于全互联架构提供了 4 控负载均衡的

持续镜像功能，如下图所示：每个控制器的 Cache 数据会均衡的向其他 3 个控制器做缓存镜像，控制器 A 控故障，Cache 数据在 B 控、C 控、D 控间均衡的做缓存镜像；当控制器 D 再故障，Cache 数据在 B 控和 C 控间均衡的做缓存镜像，从而实现了 4 坏 3 的业务高可用能力。这样的设计保证了即使单控制器故障，客户未及时更换，再次出现控制器故障，也不会导致数据丢失和业务中断，最大限度的保证了业务连续性。

图1-30 负载均衡持续镜像



在业务连续性，通过高精度的设备健康状态监控和亚健康识别能力，快速识别故障点并通过冗余接管技术进行故障隔离和故障自愈修复，修复成功则继续接入到系统中提供服务，修复失败在通过设备告警提示人工介入进行故障部件更换。基于系统强大的数据可靠性和业务连续性的保障技术，实现存储控制器 8 坏 7 故障容错业务无中断，软件升级业务不中断主机无感知。

3.6.2. 全互联的前端共享卡

同方超强 RS6800 系列高端存储系统支持前端共享卡（front-end interconnect I/O module，简称 FIM），每张卡通过 4 个 PCIE 3.0 X4 链路分别连接到 4 个控制器，共享卡的每一个 FC 端口均允许主机接入并同时访问 4 个控制器。同方超强 RS6800 系列高端存储系统当前支持 8G/16G/32G FC 前端共享卡。

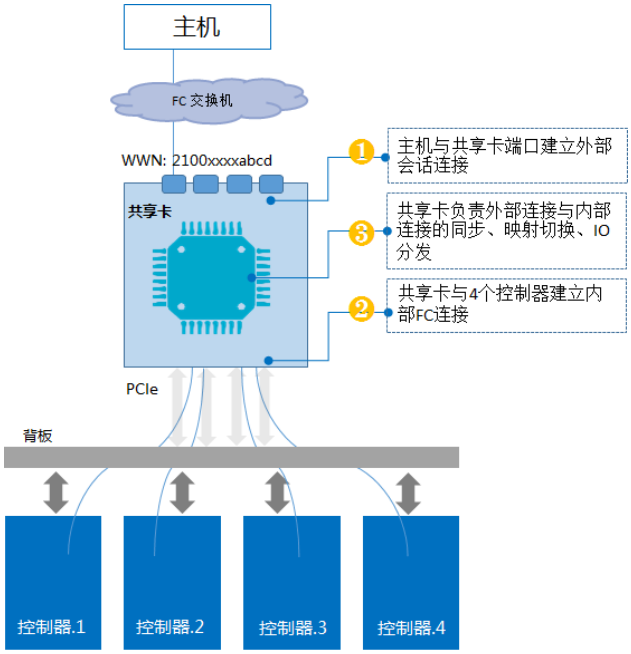
前端共享卡支持对主机的 IO 进行智能识别处理，并按特定规则分发，使得主机 IO 无需控制器预处理就直接发送给最佳处理控制器，实现了主机 IO 直通，避免主机 IO 在控制器之间转发。

由于共享卡的 PCIe 侧每个物理链路在内部都与 4 个控制器有独立通道，因此在一种极端情况下，如主机与阵列只连接了一条 FC 线缆时，主机的 IO 依然可以不经转发直达最佳控制器，因此即使对控制器进行在线升级或某些控制器故障，

业务也不受影响。

前端 FC 共享卡的实现原理如下图所示。

图1-31 前端共享卡原理



FC 共享卡有 4 个 FC 物理端口，每个端口有 1 个 WWN，主机与共享卡端口建立 1 个外部会话连接，同时共享卡与内部控制器分别建立 4 条连接，从主机视角来看，主机与存储系统仍然只有一条 FC 链路。共享卡内部完成 FC 协议与连接处理，主机的 IO 请求在共享卡内部按照智能分发算法分发到四条链路上。从控制器的视角看，每个控制器都与主机各建立了一条 FC 链路。

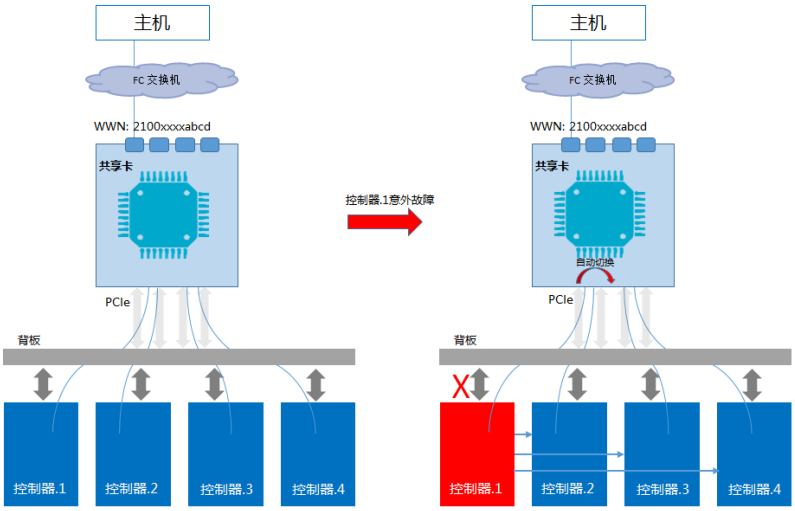
全共享的前端设计可以使系统组网更简单。在普通接口模块情况下，主机需要与每个控制器建立物理链路，一个四控的系统就需要最少 4 条线（链路）。在使用共享接口模块情况下，只需要 2 条线（出于冗余路径考虑），共享卡通过内部全共享实现了每个控制器与主机之间的全连接。

使用前端共享卡可以在控制器意外故障的场景下，主机与前端端口的连接不会中断，主机感知不到控制器故障，达到高可用目的。当控制器意外故障时，共享卡端口芯片会感知到与控制器之间的 PCIe 链路断开，配合控制器内的业务倒换，接口模块把主机的请求重新分发到其他控制器，实现了控制器故障秒级切换，主机不感知。相比在传统接口模块下的控制器故障场景需要主机多路径进行链路切换，倒换通常需要 10~30 秒，而共享卡切换时间更短，可靠性更高。

如下图所示，假设控制器.1 故障，控制器.1 上的业务被快速倒换到其他控

制器，共享卡智能算法根据控制器故障情况，把 IO 分发到新的控制器处理。整个快速完成，共享卡与主机的 FC 链路无闪断，主机无感知。

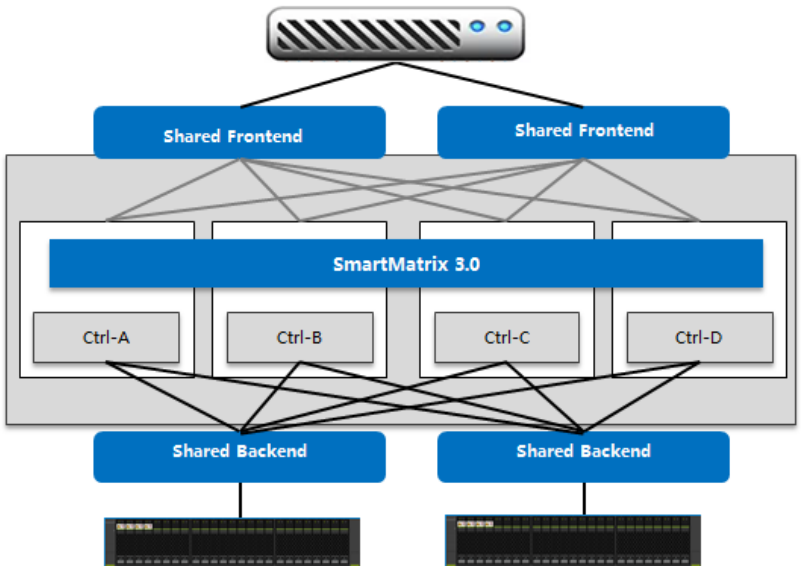
图1-32 控制器故障情况下主机业务切换



3.6.3. 全互联的后端共享卡

同方超强 RS6800 系列高端存储系统后端支持 SAS 硬盘框、智能 SAS 硬盘框、智能 NVMe 硬盘框三种磁盘框。通过 SAS3.0 共享接口卡或者 100Gb RDMA 共享接口卡进行后端扩展。SAS 共享卡连接 SAS 硬盘框，100Gb RDMA 共享卡连接智能 SAS/NVMe 硬盘框。共享接口卡插在控制框内，每张卡通过 4 条 PCIE3.0*4 连接到框内四个控制器，四个控制器允许同时访问与共享卡相连的硬盘框内硬盘单元。

图1-33 后端共享卡原理

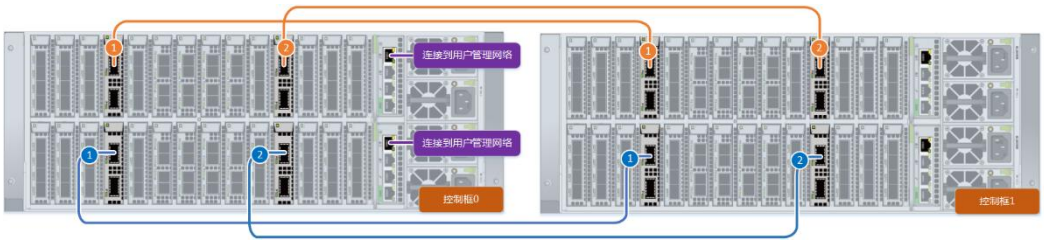


3.6.4. 全互联的 Scale-Out 共享卡

同方超强 RS6800 系列高端存储系统通过控制框内的无源背板实现四控互联，控制框之间可以采用交换机组网和免交换直连组网两种方式，系统最大支持 8 个控制框 32 个控制器。Scale-Out 扩展采用 100Gb RDMA 共享卡，Scale-Out 共享卡与本引擎 4 个控制器直接连接，每个端口可以收发来自其他引擎 4 个控制器的数据，数据能够不经过转发而直接发送给对应控制框的最佳控制器处理，实现了多控控之间的全互联。

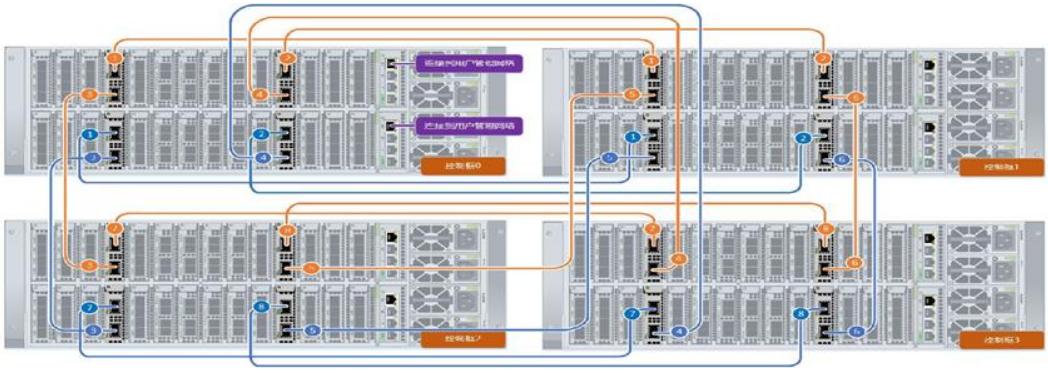
同方超强 RS6800 系列高端存储系统扩展到 8 控时，采用直连组网方式。每个引擎通过 4 张 Scale-Out 共享卡组网，8 控直连组网节省了 2 台交换机，并且减少了一半的线缆，能够有效降低系统组网成本及管理复杂度。

图1-34 8 控免交换机 Scale-Out 组网



同方超强 RS6800 系列高端存储系统直连组网最大可支持扩展到 16 控。交换机组网时时使用 100G DCB 交换机，后续版本支持。

图1-35 16 控免交换机 Scale-Out 组网



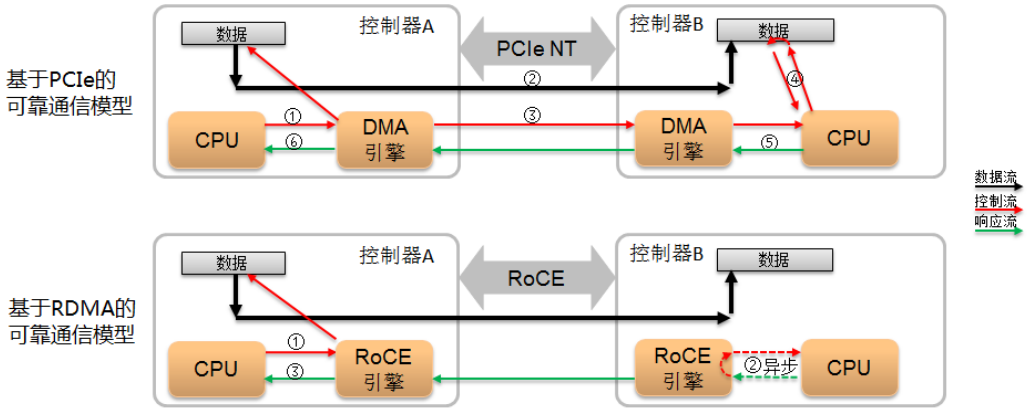
3.6.5. 低时延 RDMA 互联通道

同方超强 RS6800 系列高端存储系统控制器之间均采用 RDMA 组网连接，控制框与智能 NVMe 硬盘框、智能 SAS 硬盘框之间也采用 RDMA 组网连接。数据经过 RDMA 链路进行远程 DMA 数据搬移，搬移工作由接口模块完成，无需两侧 CPU 参

与，数据可以直接 RDMA 一跳到达到对端节点内存，大大提高了数据传输效率，降低了访问时延。

同方超强 RS6800 系列高端存储系统使用基于 RoCE 的 RDMA 技术，相比 PCIE 及 SAS 链路，基于 RoCE 通道的可靠通信时延更低。下图是基于 PCIE 链路和 RoCE 链路的 IO 交互过程对比。通过 RoCE 和 PCIe 进行数据传递包含三个阶段：启动控制命令，传递传输到对端，以及对端接收数据进行验证并回响应消息。在 PCIe 通信模型下，数据从控制器 A 发送到控制器 B 以后，控制器 A 的 CPU 还需要通过控制流通知控制器 B 数据已送达（触发控制器 B 的中断），控制器 B 调用中断处理过程，对消息进行校验并回响应消息。对于 RoCE 链路并无这个过程，当数据发送成功后，控制器 A 无需通知控制器 B 数据已经送达，控制器 B 会轮询并处理达到的数据，并回响应。RoCE 相比 PCIE 就减少了通知数据已经到达的过程，减少了交互次数，时延更低，带宽更高。

图1-36 PCIe 和 RoCE 交互过程对比



可靠通信模型（带对端确认机制）	操作步骤数	RoundTrip 总时延(us)
基于 PCIe 通信模型	6	~50
基于 RDMA 通信模型	3	~30

RDMA 全互联设计在连接距离、扩展灵活性、共享访问上还产生额外收益，下表是 PCIE、RoCE 以及 SAS 通道比较。

表1-15 PCIe/SAS/RDMA(RoCE)可靠通信模型下性能对比

不同组网协议对照		PCIe (PCIe3.0x4 端口)	SAS (12G SAS3.0x4 端口)	RoCE (100Gb RDMA 端口)
性能	性能：带宽 (GB/s)	3.2	4	10

不同组网协议对照		PCIe (PCIe3.0x4 端口)	SAS (12G SAS3.0x4 端口)	RoCE (100Gb RDMA 端口)
	性能：RoundTrip 时延 (us)	~50	~60	~30
扩展性	连接距离(铜缆)	约 0.5 米	3~5 米	5~7 米
	最大支持的 NVMe 盘数(块)	占用总线 ID 与接口卡共用，一般最大支持盘数<100	不支持	无限制
	热插拔	需先按按钮再插拔，暴力插拔有宕机风险	支持	支持
	盘框支持共享访问的控制器数量	2 控	4 控	4 控

4 软件架构

同方超强 RS6800 系列存储系统提供的软件包括存储系统端软件、维护终端软件和应用服务器端软件。这三部分软件相互配合，从而智能、高效、经济地实现各种存储业务、备份业务和容灾业务。

存储系统端软件采用专用操作系统，实现硬件管理和支撑存储业务软件的运行。存储系统通过基本功能控制软件实现基础的数据存储和读写功能；通过增值功能控制软件实现各种备份、容灾和性能调优等高级功能；通过管理功能控制软件实现对（多套）存储系统的管理功能。

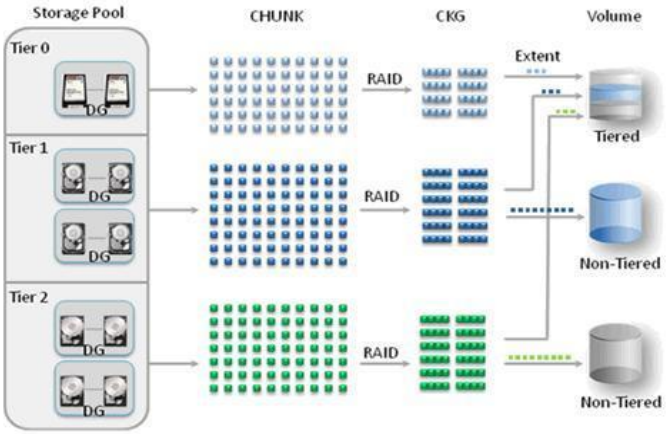
下面从块级虚拟化、SAN/NAS 一体化、负载均衡、数据缓存、端到端数据完整性保护、软件特性等方面进行关键软件架构技术介绍。

4.1 块级虚拟化

4.1.1. 块级虚拟化原理

同方超强 RS6800 系列存储系统采用 RAID2.0+块虚拟化架构。不同于传统 RAID 固定成员盘的做法，RAID2.0+是基于硬盘的块级虚拟化技术，实现动态 RAID 策略。阵列内所有的硬盘被划分为固定大小的 CHUNK，系统自动随机选择多个硬盘的多个 CHUNK 按照 RAID 算法组成 CKG，CKG 直接分配给 Volume 或被划分为固定大小的数据块（Extent）分配给不同的 Volume 使用。Volume 对外体现为 LUN 或文件系统（File System，缩写为 FS）。RAID2.0+ 如下图所示：

图1-37 RAID2.0+块虚拟化



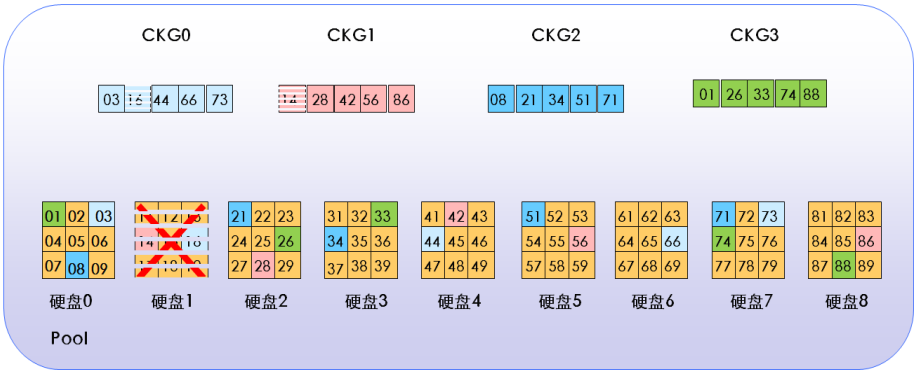
4.1.2. 快速重构

快速重构，每个物理硬盘的 CHUNK 会和多个硬盘的 CHUNK 组成 RAID，单个硬盘故障后参与重构的硬盘比传统方式多很多，可以极大提高重构速度，最快可以达到每 TB 重构 30 分钟完成。

以 9 块硬盘 RAID5 为例。当硬盘 1 损坏，造成 CKG0 和 CKG1 的数据损坏。系统随机选择 CHUNK 进行重构。

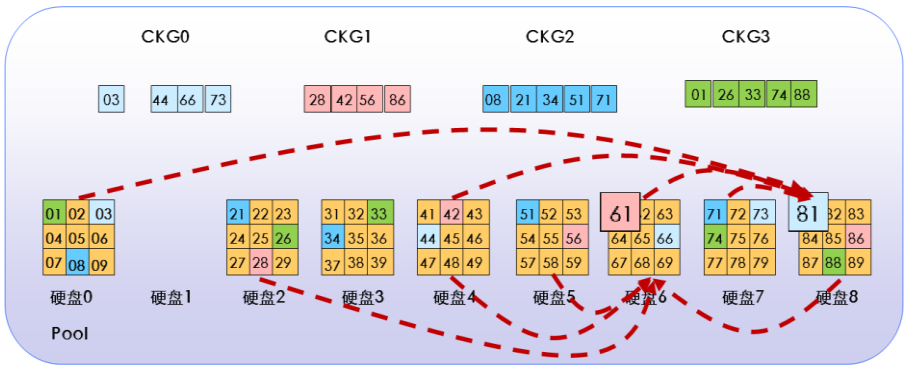
如下图，14 和 16 两个 CHUNK 损坏，将随机选择 POOL 中的空闲 CHUNK 进行重构（如图 1-38 黄色方块），随机选择的 CHUNK 将保证尽量分布在不同的硬盘上。

图1-38 RAID2.0+快速重构示意图（一）



如下图，随机选择硬盘 6 的 61 号 CHUNK 和硬盘 8 的 81 号 CHUNK，数据将从其他成员盘重构到这两个 CHUNK。

图1-39 RAID2.0+快速重构示意图（二）



传统硬盘重构的瓶颈主要在目标盘（热备盘），因为所有成员盘将所有数据读出后重构数据会全部写入到目标盘，其写带宽就成了整个重构速度的关键，比如一块传统 2T 大容量硬盘，重构时间就是 2T 除以 30M/S，也就是 18 个小时。

而经过 RAID2.0+块虚拟化后将有两个方面的提升：

1. 多块目标盘，如上例子就是两块目标盘，重构时间就将缩短为 9 小时，当 CHUNK 数量和成员盘增加时，目标盘可以达到成员盘个数，所以重构速度将极大提升。
2. 按 CHUNK 重构，当故障盘上分配的 CHUNK 较少时，需重构的数据将大幅降低，重构速度将进一步提升。

RAID2.0+在正常工作状态下（不停机）最快可以达到每 TB 重构 30 分钟完成，重构时间的缩短，将大大降低双盘失效的概率。

4.1.3. 硬盘负载均衡

RAID2.0+技术将使硬盘自动负载均衡，Volume 的数据被均匀分布到阵列内所有的硬盘上，可以防止局部硬盘过热，提升可靠性。在参与业务读写过程中，阵列内硬盘参与度高，有效提升系统性能。

4.1.4. 最大化硬盘资源利用率

最大化硬盘资源利用率：

1) 性能：在 RAID2.0+环境中，LUN/文件系统基于资源池（Pool）创建，不再受限于 RAID 组磁盘数量，实现在线（不停机）微码升级，单个 LUN/文件系统的性能可得到大大提升；
2) 容量：由于资源池中的磁盘数量不受限于 RAID 级别，免除传统卷管理技术环境下有些 RAID 组空间利用率高而有些 RAID 组空间利用率低的状况，并借助 LUN/文件系统动态扩容，从而提升磁盘的容量利用率。

4.1.5. 提升存储管理效率

1) 易规划：无需花费过多的时间做存储预规划，只需简单地将多个硬盘组合成存储池，设置存储池的分层策略，从存储池划分空间(卷)即可；
2) 存储池易扩容：当需要扩容存储池，只需插入新的硬盘，系统会自动的调整数据分布，让数据均衡的分布到各个硬盘上；
3) 卷易扩容：当需要扩容卷时，只需输入想要扩容的卷大小，系统会自动从存储池中划分所需的空间，并自动调整卷的数据分布，使得卷数据更加均衡的分布到所有的硬盘上。

4.2 SAN/NAS 一体化

同方超强 RS6800 系列存储系统采用 SAN/NAS 一体化设计，不再需要 NAS 网关设备，一套软硬件同时支持 SAN 和 NAS，支持 NFS、CIFS、FTP、HTTP 等文件访问协议，以及 NDMP 文件备份协议。NAS 与 SAN 一样，同样支持 16 控的 Scale-out，主机可以从任意 1 个控制器上的前端主机端口访问任意 1 个 LUN 或文件系统。

统一存储的同方超强 RS6800 系列融合平台架构如下图，文件系统和 LUN 是平行的出在 Space 子系统之上，下面是基于 RAID2.0+的块虚拟化存储池子系统。在这个架构中，文件系统和 LUN 都直接与底层的 Space 子系统交互。文件系统架构是基于对象的，每个文件或文件夹是一个对象，每个文件系统是由对象组成的对象集。对于 LUN 来说，LUN 分 Thin LUN 和传统的 Thick LUN。两种 LUN 也都来自于 Pool 和 Space 系统，并没有建立在文件系统之上。这样简化的软件栈带来的存储效率比传统的统一存储架构效率要高，同时 LUN 和文件系统各自保持独立，互不影响。

4.3 负载均衡

4.3.1. SAN 负载均衡

默认情况下，同方超强 RS6800 系列存储系统自动将不同 LUN 均衡分配到不同控制器，LUN 空间通过 RAID2.0+块级虚拟化技术均衡打散到系统内所有硬盘。

如果主机到存储阵列的每个控制器都有 I/O 路径，多路径软件会优选 LUN 归属控制器的路径下发；如果没有优选路径，则 IO 下发到阵列后，系统会自动判断对应 LUN 业务应由哪个控制器处理，通过 TFCQ Matrix 智能矩阵将 IO 转发到对应控制器进行处理。

通过将 LUN 均衡分配给不同控制器及 LUN 空间在全局范围均衡分布，使得不同控制器业务、硬盘压力相对均衡，配合多路径选择最优路径下发 IO，使系统性能达到了最优。

4.3.2. NAS 负载均衡

默认情况下，同方超强 RS6800 系列存储系统自动将不同文件系统均衡分配到不同控制器，文件系统空间通过 RAID2.0+块级虚拟化技术均衡打散到系统内所有硬盘。

同方超强 RS6800 系列存储系统也同时提供内置 DNS 负载均衡特性，可根据每个控制器的业务负载智能地将主机 NFS/CIFS/FTP 客户端连接分发给配置在不同节点、不同端口上的业务 IP 进行处理，从而提升系统的性能和可靠性。

DNS 负载均衡特性是指主机通过域名访问存储阵列的 NAS 业务时，先发送 DNS 解析请求到阵列内置 DNS 服务器，根据域名获取 IP 地址。域名下包含多个 IP 时，内置 DNS 服务器会智能计算每个控制器的 CPU 利用率、端口带宽利用率、所在控制器的 NAS 连接数等，选择负载较轻的控制器 IP 作为 DNS 响应返回给主机。主机收到 DNS 响应后，向目标 IP 发起业务请求。

DNS 负载均衡特性支持的负载均衡策略有轮循方式、按节点 CPU 利用率、按节点连接数、按节点带宽利用率、按节点综合负载。

4.4 数据缓存

- **缓存分布：**

同方超强 RS6800 系列存储系统的物理内存的使用分布情况为：

物理内存 = 操作系统等占用缓存 + 读缓存 + 本地写缓存 + 镜像写缓存 + 业务特性占用缓存

- **缓存类型：**

同方超强 RS6800 系列存储系统缓存分为读缓存、写缓存。

读缓存：将已读取的数据保存在内存空间中（读缓存），当下次再次读取同一数据时就不必重新从磁盘上读取，从而提高速率。

写缓存：将要写入磁盘的数据先保存在内存空间中（写缓存），当保存到写缓存中的数据达到一个阈值时，便将数据保存到硬盘中。通过读写缓存可以减少实际的磁盘操作，提升系统读写性能，同时有效的保护磁盘免于重复的读写操作而导致的损坏。

写缓存没有使用时，系统所有缓存都可以用作读缓存。系统对读缓存有最小

容量预留，以保证在写业务压力很大时，仍能保证读业务缓存资源可以使用。

- **缓存预取：**

同方超强 RS6800 系列存储系统实现了多路顺序流识别算法，即在大量乱序和随机的 IO 中识别出顺序 IO 流，对顺序的读写 IO 流采用预取和合并算法，能优化多种应用场景的系统性能。

同时，同方超强 RS6800 系列存储系统的预取算法实现了智能预取、固定预取、倍数预取等算法。智能预取能自动识别 IO 特征，根据 IO 特征决定是否预取、预取多大长度，确保产品性能能满足不同应用场景。

系统默认采用智能预取算法，在某些 I/O 模型非常明确的应用场景，用户也可以配置固定预取或倍数预取算法，这两种算法支持由用户自行配置预取数据长度。

- **缓存淘汰：**

当系统缓存占用率达到阈值时，淘汰算法根据历史访问频率和当前的访问频率，计算数据块的热度，结合多路顺序流识别算法，选择合适的数据进行淘汰。另外根据用户的具体需要，可配置 Volume 的缓存优先级，还可以对具体业务调整每个 IO 的优先级。低优先级的数据，优先淘汰；高优先级的数据缓存更多，保证数据命中率。

4.5 端到端数据完整性保护

ANSI T10 PI（Protection Information）标准提供了一种方法来校验访问存储系统过程中的数据完整性。这种检查通过 T10 标准中定义的 PI 字段来实现。该标准通过在每个扇区数据后加上 8 字节的 PI 字段来实现数据完整性检查。T10 PI 通常用来保证存储系统内部的数据完整性。

DIX（Data Integrity Extensions）进一步延伸了 T10 PI 的保护范围，实现了从应用到主机 HBA 的数据完整性保护，因此，DIX+T10 PI 可以实现从应用到硬盘的完整的端到端数据保护。

同方超强 RS6800 系列存储系统不但支持 T10 PI 来保证存储系统内部的数据完整性保护，而且支持从应用到硬盘的 DIX+T10 PI 端到端数据完整性保护。阵列对数据 PI 字段进行实时校验并下发，如果主机侧不支持 PI，则阵列会在主

机接口增加 PI 字段并下发。在存储系统中，PI 跟随用户数据一起参与各种转发、传输并最终存储到磁盘介质中。数据被主机应用重新读出前，系统会通过数据 PI 检查数据的正确性和完整性，保证用户数据的可靠性。

4.6 面向闪存的系统优化

SSD 盘优势在于随机 IO 性能好，时延低，劣势在与擦写次数有限；而 HDD 优势在于顺序 IO 性能好，无擦写次数限制。同方超强 RS6800 系列存储系统对 SSD 盘与 SSD/HDD 混合存储进行了针对性优化，以达到更好的性能和可靠性。

1. 系统与 SSD Firmware 无缝联动

SSD 由于采用 Flash 的原因，盘片内部会存在擦除操作，当盘片内部正在擦除时，与擦除相同通道的其他数据不能读写，因此会造成大约 1～2ms 的时延，导致性能波动。

存储系统和盘片配合，协调多个硬盘轮流执行擦除操作，系统不选择从正在擦除的盘读取数据，而是通过 RAID 冗余从其他盘上读取数据，从而保证稳定的时延。

2. Cache 针对 SSD 的智能“感知”

针对 SSD 和 HDD，存储系统采取不同的脏数据刷盘策略，充分发挥出各自优势：当经过认证的硬盘接入时，系统自动识别介质类型。针对 SSD 硬盘，系统按照 LRU 算法刷盘，降低算法计算复杂度（也降低了时延），延迟活跃数据的刷盘时间，减少下盘次数，减小写放大，提升系统性能，同时也延长了 SSD 寿命。

3. 多核性能优化

在多核调度机制方面，针对 NUMA 架构进行性能优化，例如把单个 IO 的消息调度在一个 CPU 核上进行处理，减少多 CPU 间访问开销，提升 CPU 缓存命中率。

在多线程运行效率上，通过数据结构的合理设计，避免多线程并发访问 CPU L1 Cache 一个缓存单位（Cacheline）上的数据，消除 CPU L1 Cache 伪共享的问题，极大提升 CPU L1 Cache 利用效率，减小数据内存访问的 CPU 开销。

4. 盘内磨损均衡

磨损均衡是指 SSD 控制器通过对 NAND Flash 中 Block 的 P/E 次数进行监控，通过一定的软件算法使所有 Block 的 P/E 次数比较平均，防止单个

Block 因过度 擦写而导致失效，延长 NAND FLASH 整体的使用寿命。

超强 RS 存储系统 SSD 采用的磨损均衡分为动态磨损均衡和静态磨损均衡。动态磨损均衡是指在主机数据写入的时候，优先挑选磨损较小的 Block 使用，这样保证 P/E 消耗平均分布；静态磨损均衡是指盘片定期在整个盘片的范围内寻找 P/E 消耗较少的 Block 并回收其上的有效数据，从而使得保存冷数据的 Block 也参与到磨损均衡的循环中。超强 RS 存储系统 SSD 通过这 2 种方案的结合来保证全盘磨损均衡。

4.7 丰富软件特性

同方超强 RS6800 系列存储系统提供了用于系统效率提升的 TFCQ 软件系列和用于数据保护的 Hyper 系列软件：

效率提升系列（TFCQ 系列）：在线重删（TFCQ Dedupe）、在线压缩（TFCQ Compression）、智能精简配置（TFCQ Thin）、异构虚拟化（TFCQ Virtualization）、智能数据迁移（TFCQ Motion）、智能数据迁移（TFCQ Migration）、智能数据分级（TFCQ Tier）、智能服务质量控制（TFCQ QoS）、智能缓存分区（TFCQ Partition）、数据销毁（TFCQ Erase）、多租户（TFCQ Multi-Tenant）、SSD 智能缓存（TFCQ Cache）、智能配额（TFCQ Quota），主要为用户提供存储效率提升方面的功能，降低用户的 TCO。

数据保护系列（Hyper 系列）：快照（HyperSnap）、克隆（HyperClone）、远程复制（HyperReplication）、双活（HyperMetro）、一体化备份（HyperVault）、LUN 拷贝（HyperCopy）、卷镜像（HyperMirror）、WORM（HyperLock），主要为用户提供数据容灾备份相关的功能。同时支持丰富的两地三中心（3DC）解决方案。

5 精简高效特性软件功能说明

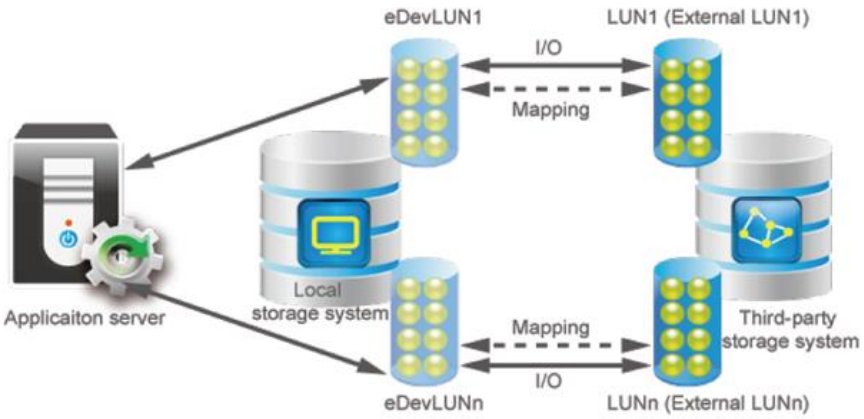
5.1 异构虚拟化（TFCQ Virtualiztaion）

异构虚拟化特性 TFCQ Virtualization 用于接管其他存储系统（包括其他存储系统和第三方厂商的存储系统），保护现有投资。使用 TFCQ Virtualization 后，本端存储系统能够将异构存储系统提供的存储资源当作本地存储资源进行使用并对其进行集中管理，无需关注存储系统间软件架构和硬件架构的差异。同时，结合 TFCQ Migration 特性还可以实现对异构存储系统中的数据进行在线迁移，帮助客户完成新老设备的更新换代和数据搬迁。

1. 工作原理

通过把异构阵列映射到本端阵列，把异构阵列的存储空间通过 eDevLUN（External Device LUN）的方式管理和利用起来。eDevLUN 包括元数据卷（Meta Volume）和数据卷（Data Volume）。元数据卷用于对 eDevLUN 的数据存储位置进行管理，其所需要的物理空间由本端存储系统提供。数据卷是对外部 LUN 数据的逻辑抽象，其所需的物理空间由异构存储系统提供，不占用本端存储系统空间。本端存储系统上创建的 eDevLUN 与异构存储系统上的外部 LUN 是一一对应的关系。应用服务器可以通过对 eDevLUN 的读写操作实现对外部 LUN 的数据访问。

图1-40 异构虚拟化技术示意图



TFCQ Virtualization 通过 LUN 伪装技术，将同方超强 RS6800 系列存储系统的 eDevLUN 的 WWN 和 Host LUN ID 设置成与异构存储系统上的 LUN 的信息一致，在数据迁移完成后，通过主机多路径软件实现在线 LUN 的无缝切换，从而在

主机不中断业务的情况下完成数据迁移。

5.2 数据重删压缩（TFCQ Dedupe&TFCQ Compression）

数据重删压缩功能为文件系统和 Thin LUN 提供数据精简的服务。可以为客户节约空间的同时也减少了企业 IT 架构的 TCO（Total Cost Ownership）。

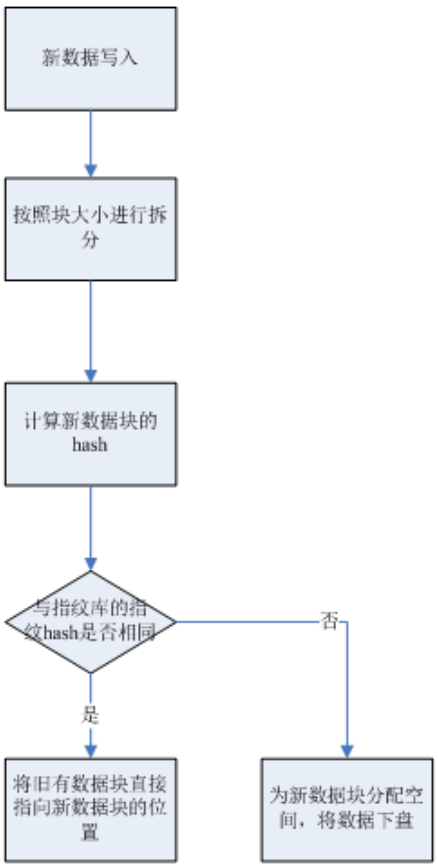
5.2.1. 在线重删（TFCQ Dedupe）

在线重删特性基于在线处理的方式实现了文件系统和 Thin LUN 的数据重删功能。

在系统中，重删功能的粒度和文件系统或者 ThinLUN 的最小读写单元 Grain 保持一致。同时，由于用户在创建文件系统或者 ThinLUN 时可以指定 Grain 的大小（4KB~64KB），同方超强 RS6800 系列存储系统也即实现了基于不同粒度的数据重删功能。

进行重复数据删除处理的流程如图 1-41 所示。

图1-41 重复数据删除处理流程

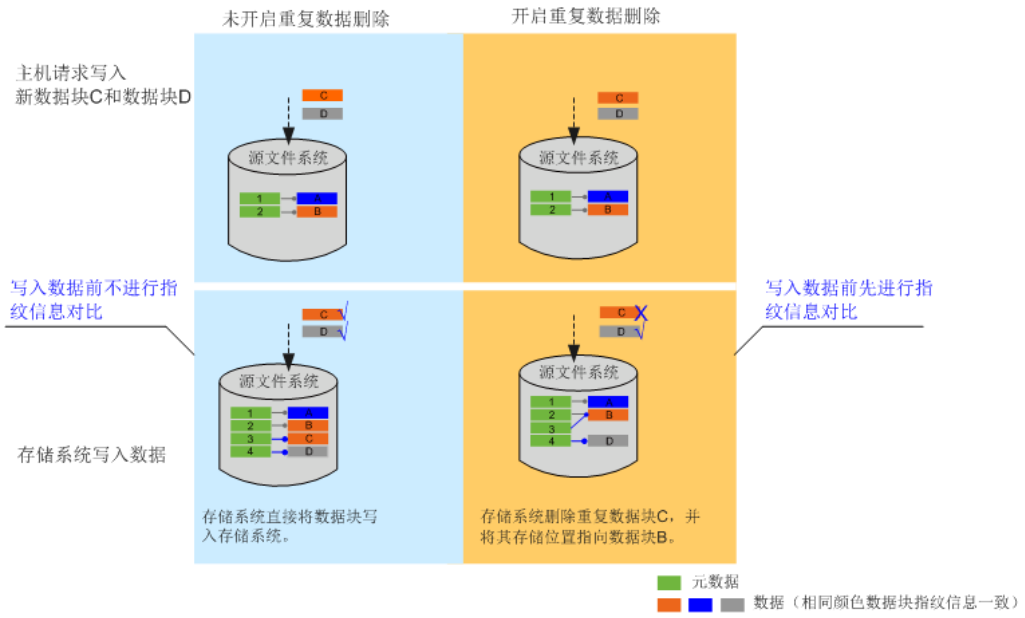


1. 存储系统按照块大小进行拆分。
2. 存储系统会将新写入的数据块与旧的数据块通过指纹库进行对比，如果指纹不同，当做是新数据块，写入。如果指纹不同：

- 逐字节比较功能关闭（默认），存储系统会将旧有数据块直接指向新写入的数据块存储位置，而不分配空间。
- 逐字节比较功能开启时，将之前写入的数据与当前的数据内容进行字节级比较，如果完全相同，则认为是重复数据块。如果不同，当做是新数据块。

例如，文件系统中原有数据块为数据块 A 和数据块 B。应用服务器写入数据块 C 和数据块 D，数据块 C 与数据块 B 指纹信息一致，数据块 D 与与原有数据块 A、B 的指纹信息均不一致。采用不同的重复数据删除策略时，数据重删处理结果如示意图所示。

图1-42 数据重删处理结果



5.2.2. 在线压缩（TFCQ Compression）

业界一般的压缩做法有在线压缩以及后压缩。存储系统实现了在线压缩，对新写入的数据在写盘前先进行压缩处理，再将压缩后的数据写盘，能有效的节省用户的空间。和后压缩（在数据下盘后再执行压缩）相比，在线压缩有以下优点：

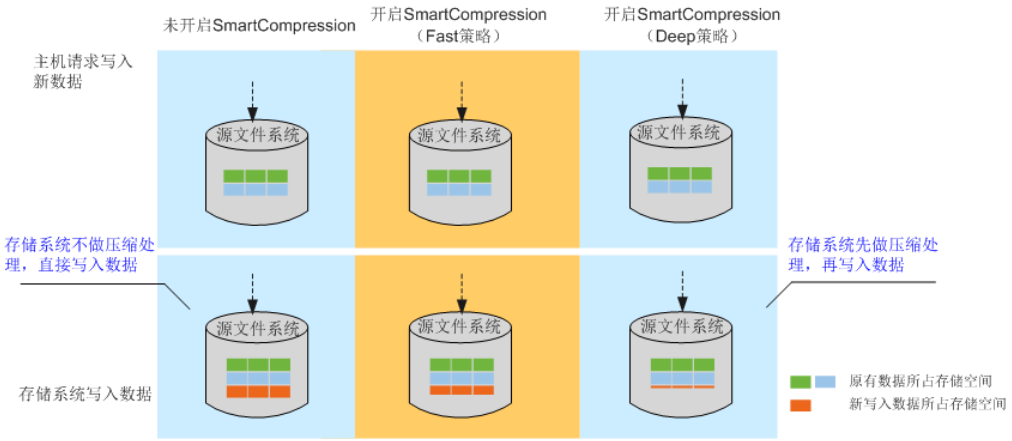
- 更小的初始存储空间，降低客户初始投资。
- 更少的 I/O，适合有读写寿命限制的 SSD 磁盘。
- 在线压缩是在执行压缩后再打快照，能做到最大限度的节省存储空间。

存储系统在进行数据压缩处理时，会根据用户设定的压缩策略进行不同程度的压缩。存储系统支持如下两种压缩策略：

- Fast 策略：Fast 策略是系统默认使用的压缩算法。该算法压缩速度快，但与 Deep 策略相比压缩后空间节省效率低一些。
- Deep 策略：Deep 策略可以获得空间节省效率的明显提升，但压缩和解压需要花费更长的时间。

数据压缩处理过程如图 1-43 所示。

图1-43 数据压缩处理结果



5.2.3. 重删、压缩效果可叠加

TFCQ Dedupe 和 TFCQ Compression 功能支持同时开启。当同时开启时，数据先经过重删再执行压缩，可以实现缩减效果的叠加，可以为用户节省更多存储空间。

由于采用在线处理的方式实现数据的重删和压缩，当用户开启重删压缩功能时，只对后续写入的数据执行重删压缩；当关闭重删压缩功能时，之前重删过的数据不会被恢复成原来的格式。

5.3 智能数据分级（TFCQ Tier）

5.3.1. 块数据分级（TFCQ Tier for Block）

TFCQ Tier 是针对块存储的智能数据分级特性。

TFCQ Tier 按照应用对性能的需求，将 SSD，SAS 和 NL-SAS 三种类型磁盘分别对应高性能存储层，性能存储层和容量存储层。每个存储层可以单独使用，也可以根据需要两两组合，或者三者组合在一起组成存储池提供数据存储空间。

TFCQ Tier 进行 LUN 级别的智能化数据存放管理，以“extent”为单位，默认 4MB （512KB ~ 64MB 可配）来统计和分析数据的活跃度，将不同活跃度的数据和不同特点的存储介质动态匹配，并通过数据迁移将活跃度高的“繁忙”数据迁移至具有更高性能的存储介质（如 SSD 硬盘），将活跃度低的“空闲”数据迁移至具有更高容量且更低容量成本的存储介质（如 NL-SAS 硬盘）。

TFCQ Tier 经历的数据监控、排布分析、数据迁移三个阶段，如下图所示：

图1-44 TFCQ Tier 处理过程

其中，数据监控、排布分析阶段由存储系统自动完成，数据迁移阶段通过用户手动触发或根据用户配置的定时策略触发。

TFCQ Tier 提高存储系统性能并降低用户成本，满足企业对性能和容量的双重需求，避免历史数据占用昂贵的存储介质，保证企业有效投入，消除无用容量带来的能耗开销，降低企业 TCO，得到最优性价比。

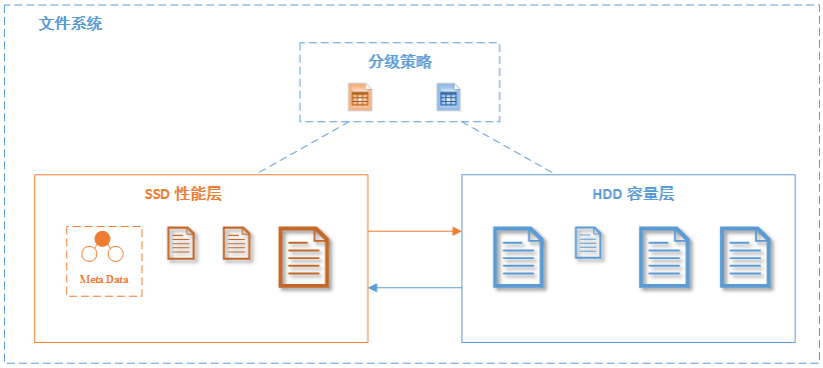
5.3.2. 文件数据分级（TFCQ Tier for File）

TFCQ Tier 是为了满足客户简化数据生命周期管理，提升介质利用率，降低客户成本而推出的面向文件系统的分级特性。它是基于用户自定义策略，以文件为粒度在不同介质中进行动态迁移的技术。

企业存储的 Storage Pool 可以由多种介质（例如 SSD、HDD）混合组成，TFCQ

Tier 根据用户指定的策略（例如：文件名、文件大小、文件类型、文件创建时间、SSD 使用率等），使数据可以在多种介质间流动，比如从高性能介质（例如 SSD）自动迁移到大容量介质（例如 HDD，包括 SAS 或 NL-SAS）。原理如图 1-45 所示。

图1-45 TFCQ Tier 原理



TFCQ Tier 有以下技术特点：

1） 自定义策略

支持灵活的用户自定义策略（例如：文件名、文件大小、文件类型、文件创建时间、SSD 使用率等），策略可以通过条件组合满足不同的应用场景。

2） 访问加速

文件系统的元数据默认保存在 SSD 中，这样可以对文件、目录等进行快速的定位，从而加速文件访问。

3） 智能流控

在文件迁移过程中 CPU 和硬盘的负载会有一定增加，系统会根据业务压力对迁移任务进行智能流控，降低迁移任务对业务性能影响。

4） 成本优势

采用 SSD 和 HDD 分级存储，相对于全闪存存储，即能保障性能又能节省客户购置成本。

5） 简化管理

支持在一个文件系统内进行分级，冷数据自动迁移至 HDD 中，简化了用户对数据生命周期的管理，无需借助其他特性或者应用进行数据迁移归档，迁移过程客户端无感知。

5.4 智能精简配置（TFCQ Thin）

智能精简配置以一种按需分配的方式来管理存储设备。智能精简配置不会预先分配所有的空间，而是将大于物理存储空间的容量形态呈现给用户，使用户看到的存储空间远远大于系统实际分配的空间。用户对这部分空间的使用实行按需分配的原则。如果用户的存储空间不足，可通过扩充后端存储资源池的方式来进行系统扩容，整个扩容过程无需业务系统停机，对用户完全透明。

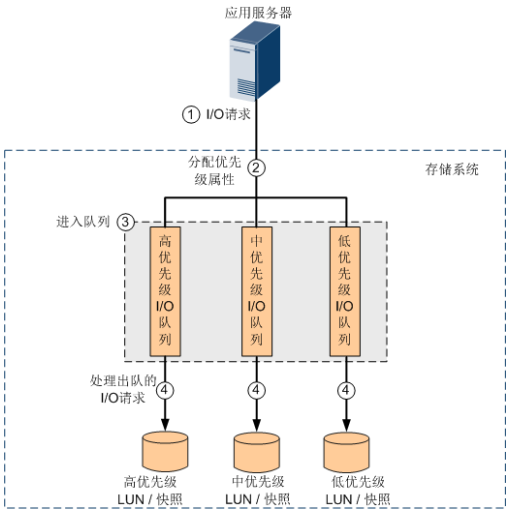
5.5 智能服务质量控制（TFCQ QoS）

TFCQ QoS 特性又叫智能服务质量控制特性，可以通过动态地分配存储系统的资源来满足某些应用程序的特定性能目标。TFCQ QoS 特性允许用户根据应用程序数据的一系列特征（IOPS、占用带宽）对特定应用程序设置特定的上限目标。存储系统根据设定的上限目标，准确限制应用程序的性能，避免非关键应用程序抢占过多存储系统资源，影响关键应用程序的性能。

TFCQ QoS 采用基于 LUN、FS 或快照的 I/O 优先级调度技术和 I/O 流量控制技术两种方式来保证数据业务的服务质量：

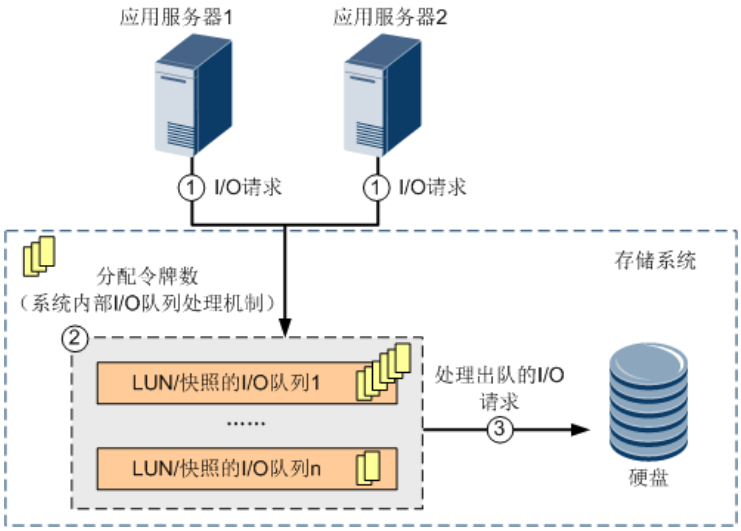
I/O 优先级调度技术：通过为业务设置优先级来区分不同业务的重要性。在存储系统为不同业务分配存储系统资源时，优先保证高优先级业务的资源分配请求。在存储系统资源紧张的情况下，为高优先级的业务分配较多的资源，以此尽可能保证高优先级业务的服务质量。

图1-46 基于 I/O 优先级调度技术



I/O 流量控制技术：TFCQ QoS 特性的 IO 流量控制技术是对应用程序 I/O 的队列管理，通过对 I/O 请求处理、令牌分发和出队控制三部分实现。针对用户设置的性能控制目标（IOPS、带宽）进行流量限制，通过 I/O 流控机制，限制某些业务由于流量过大而影响其它业务。

图1-47 基于 LUN 或快照的 I/O 流量控制队列管理



5.6 智能缓存分区（TFCQ Partition）

智能缓存分区特性的核心思想是通过对系统核心资源的分区，保证关键应用的性能。管理员可以针对不同的应用配置不同大小的缓存分区，系统将保证该分区中的缓存资源被该应用独占。

缓存是影响存储系统性能的最主要因素：

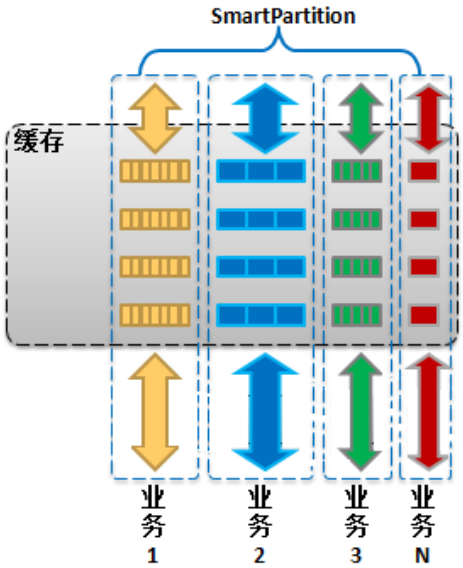
- 对写业务来说，更多的缓存意味着更高的写合并率、写命中率（同一块数据在缓存中被再次写中的比率）；
- 对读业务来说，更多的缓存通常意味着更高的读命中率。
- 同时，不同类型的业务对缓存的需求也有很大不同：
- 对顺序类业务来说，缓存不需要很大，只需要满足 I/O 合并要求即可；
- 对随机类业务来说，更大的缓存通常意味着更好的聚合度，从而带来性能的提升。
- TFCQ Partition 还可以与其他 QoS 技术（如智能服务质量控制 TFCQ QoS）相配合，从而达到更好的服务质量保证效果。

工作原理：

- Cache 分区技术通过隔离不同的业务所需要的缓存资源，保证某些关键业务的服务质量。
- 缓存是影响存储系统性能的最主要因素：
- 对写业务来说，更多的缓存意味着更高的写合并率、写命中率（同一块数据在缓存中被再次写中的比率）；
- 对读业务来说，更多的缓存通常意味着更高的读命中率。
- 同时，不同类型的业务对缓存的需求也有很大不同：
- 对顺序类业务来说，缓存不需要很大，只需要满足 I/O 合并要求即可；
- 对随机类业务来说，更大的缓存通常意味着更好的聚合度，从而带来性能的提升。
- 合理分配缓存资源是提高存储系统服务质量的主要手段。

TFCQ Partition 可以针对不同的业务（实际控制对象为 LUN 和文件系统），根据分区大小，分配不同大小的缓存分区资源，从而保证关键业务的服务质量。

图1-48 缓存分区示意图



技术特点：

TFCQ Partition 的技术优势主要体现在以下几个方面：

智能的分区控制：

TFCQ Partition 根据用户配置的缓存大小自动调配系统缓存资源以及与其他 QoS 策略相互配合，达到系统服务质量的最大化和分区质量保证达标。

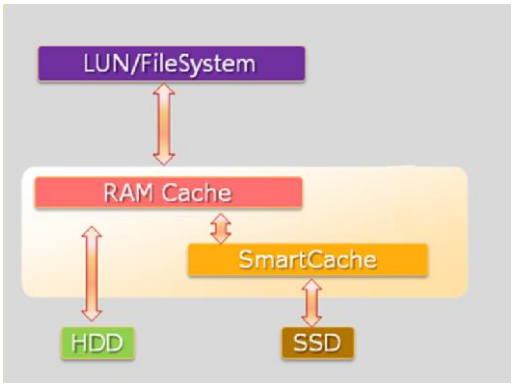
简便易用：

TFCQ Partition 配置简单，用户界面操作友好，所有操作立即生效，不用重启系统，分区的调整均不需要用户参与，从而大大提升了分区功能的易用性。

5.7 SSD 智能缓存（TFCQ Cache）

TFCQ Cache 作为存储系统中的一个读缓存服务模块，采用 SSD 盘作为存储系统中 RAM Cache 的扩充，为 RAM Cache 缓存其存放不下的干净热点数据，数据的交互关系见图 1-49。

图1-49 TFCQ Cache 逻辑架构



TFCQ Cache 对热点数据性能的提升是以 LUN/文件系统为单位的.

对一个 LUN 或文件系统开启 TFCQ Cache 功能后，RAM Cache 会下发热点数据给 TFCQ Cache；

TFCQ Cache 在内存中建立数据与 SSD 盘之间的映射关系后将数据保存到 SSD 盘；

之后在新的 IO 再次读取存储设备时会优先在 RAM Cache 中尝试命中；

如果不命中再去 TFCQ Cache 中尝试命中；

如果命中则从 SSD 中读取对应数据返回给主机。

当 TFCQ Cache 中缓存的数据量达到容量上限时，会按照 LRU（Least Recently Used）算法选取最近最少访问到的缓存块，清除查找表中的映射项淘汰掉数据，数据写入与数据淘汰处理是个循环反复的过程，这样就保证 TFCQ Cache 保存的是热度相对较高的数据。

5.8 数据销毁（TFCQ Erase）

磁盘的磁头每次读/写数据时，不可能绝对精确地定位在同一个点上，写入新数据的位置不会刚好覆盖在原来的数据上。原有数据总是会留下一些痕迹，利用专用的设备可以分析出原有数据的副本——称为影子数据。当然，如果我们反复执行覆盖操作，原有数据的痕迹也会越来越弱。

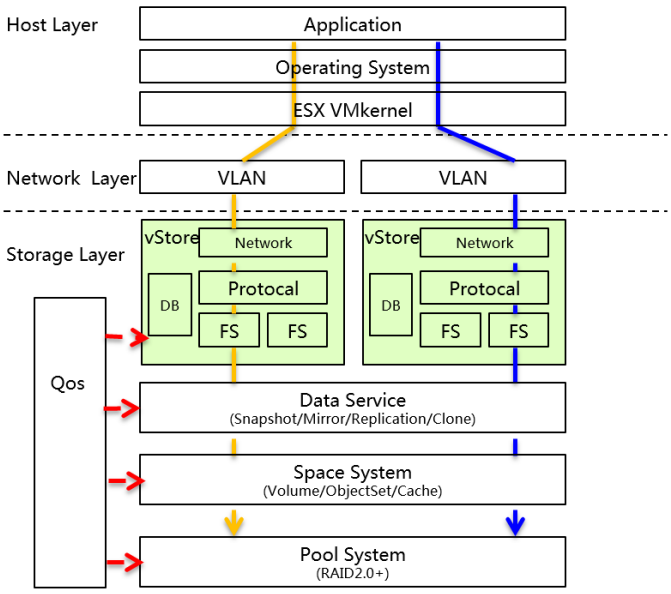
同方超强 RS6800 系列存储系统采用数据覆盖写来对 LUN 数据进行销毁，为客户提供两种数据销毁方式，用户可以根据效率和安全性灵活地选择不同的销毁方式或者配置不同的销毁参数。

5.9 多租户（TFCQ Multi-Tenant）

多租户特性又称 TFCQ Multi-Tenant，实现了在一套物理存储系统中创建多个虚拟存储系统，让租户在多协议统一存储架构中既能共享相同的存储硬件资源，又不影响相互的数据安全性和隐私。

多租户特性主要解决租户之间的隔离问题，包括管理隔离、业务隔离、网络隔离。租户之间不能相互访问数据，以此来达到安全隔离的效果。多租户特性逻辑架构如下图。

图1-50 多租户逻辑架构



管理隔离：每个租户都有自己的租户管理员。租户管理员只能通过 GUI 或

RESTful API 来配置和管理自己的存储资源。

租户管理用户支持基于角色的权限控制，创建租户管理用户时，必须选择需要的权限的对应的角色。

业务隔离：每个租户都有自己的文件系统，用户/用户组和共享/导出。用户只能通过租户 LIF 访问租户本身文件系统。

多租户的业务隔离，主要体现为用户的业务数据（主要是文件系统以及配额和快照等）、业务访问和业务配置（主要是 NAS 协议配置）隔离。

1） 业务数据隔离

系统管理员分配不同的文件系统给不同的租户，以此达到租户文件系统的隔离. 同理基于文件系统的配额、快照也是隔离的。

2） 业务访问隔离

每个租户都具有独立的 NAS 协议实例，包括 SMB 服务、NFS 服务、NDMP 服务。

3） 业务配置隔离

每个租户可以有自己独立的用户、用户组、用户映射规则、安全策略、SMB 共享、NFS 共享、AD 域、DNS 服务、LDAP 服务以及 NIS 服务。

网络隔离：租户的网络由 VLAN 和 LIF 隔离，以防止非法主机访问租户的存储资源。

租户使用逻辑端口 LIF 配置业务，一个 LIF 只能归属一个租户，做到逻辑上的端口隔离。可以选择从 GE/10GE/25GE/40GE/100GE、绑定口、VLAN 口创建 LIF。

5.10 智能配额（TFCQ Quota）

在 NAS 文件服务环境中，通常以共享目录的方式将资源提供给使用的部门、组织或个人。而每个部门或个人，都有其独特的资源需求或限制。因此，系统需要基于共享目录，因地制宜地对各个使用者，进行资源分配和限制。

文件系统配额特性称为 TFCQ Quota，正是用于满足此需求的技术，该技术可以针对目录、用户、用户组这三类资源的使用者分别进行资源控制。TFCQ Quota 可配置的配额选项有容量软配额、容量硬配额、文件软配额、文件硬配额。

容量软配额（space soft quota）：配额对象上用于空间容量告警的配置值。

当配额对象已用空间超过所设置的容量软配额时，向系统告警提示空间资源紧张，提醒用户删除不用的文件或扩大配额。此时用户仍然可以继续写入数据。

容量硬配额（space hard quota）：配额对象上用于限制最大可用容量的配置值。当配额对象已用空间到达所设置的硬配额时，如果用户继续写入新数据，向用户返回空间不足的错误。

文件软配额（file soft quota）：配额对象上用于文件数告警的配置值。当配额对象已用文件数超过所设置的文件软配额时，向系统告警提示文件资源紧张，提醒用户删除不用的文件或扩大配额。

文件硬配额（file hard quota）：配额对象上用于限制最大可用文件数的配置值。与容量硬配额一样，当配额对象的已用文件数到达所设置的硬配额时，向用户返回空间不足的错误，保证使用文件数不得超出该值。

TFCQ Quota 使用硬配额（包括容量硬配额和文件数硬配额）来限制每个使用者最多可以使用的资源。关键流程如下：

1) 在每次写 I/O 操作时，将配额的已用容量和文件数，累加本次操作增加的容量和文件数，检查其和值是否超出硬配额。
2) 若和值未超出硬配额，则允许操作向下执行；
3) 否则，写 I/O 操作失败。
4) 在检查到写 I/O 操作被允许之后，需将增量的容量和文件数，累加到之前的容量和文件数上。
5) 然后，将配额更新（即容量和文件数的最新和值）和 I/O 数据一起写入文件系统。
6) 整个 I/O 操作及配额更新，要么全部成功，要么完全失败。这样保证了已用容量在每次写 I/O 检查时，都是准确无误的。

5.11 智能数据迁移（TFCQ Motion）

在当今 IT 领域，企业和管理部门遇到的数据存储挑战往往是容量、性能和价格的要求。一方面企业购买存储系统时对后续业务性能增长需求难以准确估计；另一方面，随着业务量增加，已有系统增加硬盘后对现有业务的调整操作非常困难。

如何解决上述问题，是企业发展过程中必须要思考的问题，尤其是在 IT 系统建设初期，要做好性能需求的长期规划。

TFCQ Motion 特性可以在系统扩容时，通过自动对数据进行迁移，将数据均衡的存放在所有硬盘上。客户购买时不需要对性能需求做长远估计，只需要满足近期的性能需求即可，有效的降低了初始购买成本，从而降低 TCO。并且业务量增加造成对性能的需求也大幅增长后，只需要往系统中增加硬盘即可，TFCQ Motion 能够通过数据迁移将原有业务的数据均衡分布到所有硬盘上，从而增大原有业务的性能。

RAID2.0+技术中，硬盘域中所有硬盘被切分成相同大小的块（CHUNK），当需要分配 CKG 时，通过伪随机算法选择出需要的硬盘，再从选出的硬盘上分配 CHUNK 按照 RAID 算法组成 CKG。通过伪随机算法随机选择后，所有 CHUNK 会均匀地分布在每个硬盘上。

TFCQ Motion 基于 RAID2.0+技术实现。当硬盘域中增加硬盘后，系统自动启动 TFCQ Motion，其工作过程如下：

1) 选择第一个未做均衡处理的 CKG；
2) 通过伪随机算法对 CKG 重新选择一次硬盘；
3) 如果选择的硬盘和已有的硬盘列表完全一致，则跳过该 CKG 回到步骤 1；
4) 比较 CKG 的原盘和新选出的盘列表，通过列表的差异计算出需要迁移的硬盘对应关系，选择出 CKG 需要迁移的原硬盘和对应的目标硬盘；
5) 遍历 CKG 中每个待迁移的硬盘，从目标硬盘分配新的 CHUNK，将数据从原盘迁移到新硬盘并释放原硬盘的 CHUNK。
6) 如果已遍历系统中所有 CKG，则完成 TFCQ Motion，否则回到步骤 1 继续处理下一个 CKG。

当 TFCQ Motion 完成后，所有的 CKG 都通过伪随机算法重新选择新硬盘并完成了必要的数据迁移。此时，所有 CHUNK 已经均衡的分布在所有的硬盘上（包括新扩容的硬盘）。

6 数据保护特性软件功能说明

6.1 快照（HyperSnap）

6.1.1. LUN 快照（HyperSnap For Block）

LUN 快照(HyperSnap For Block)特性可以生成源 LUN 在某个时间点上的一致性映像，在不中断正常业务的前提下，快速得到一份与源 LUN 一致的数据副本。副本生成之后立即可用，并且对副本的读写操作不再影响源数据。因此通过快照技术就可以解决如在线的备份、数据分析、应用测试等难题。LUN 快照采用了映射表和写前拷贝（copy-on-write）相结合的技术方式来实现。

LUN 快照（HyperSnap For Block）有如下技术特点：

1） 零备份窗口

传统的备份会导致应用主机的性能下降，甚至导致用户业务中断，所以传统的备份作业必须在应用停机或业务量较小的时候进行。备份窗口是指应用所能容忍的完成数据备份作业时间，实际上就是应用所能容许的停机时间。在采用快照从事备份业务时，可以在线进行，备份窗口基本为零，无需业务停机。

2） 节省硬盘空间

采用快照获取源 LUN 在快照时间点的一致性副本时，通过 COW 卷保存源 LUN 在快照时间点后首次更新的数据即可，COW 卷的大小与源 LUN 没有关系，仅由快照时间点后源 LUN 数据的变化量决定。在源 LUN 数据量变化不大的情况下，快照通过很少的硬盘空间获得了源 LUN 的一致性副本，供其他的测试业务使用，非常节省硬盘空间。

3） 快速的数据恢复

对于传统的离线备份，备份数据无法直接在线读取，必须经过较长时间的数据恢复过程才能够获得原数据在备份时间点的可用副本，从而才能实现数据的还原。快照可通过直接读取快照卷的方式获得快照时间点的原数据，当源 LUN 数据受到意外的破坏时，可以直接从快照卷中恢复出快照时间点的数据，从而实现了很方便的数据回滚。

4） 快照一致性激活

在 OLTP 应用中，通常需要对多份源 LUN 数据创建同一时间点的快照，才能将该应用分布在不同 LUN 中的关联数据保持在同一时间点。如果不能保证同一时间点创建快照，在通过快照进行数据恢复时，应用可能无法正常使用。比如数据库应用中，管理数据、业务数据、日志信息通常会分布在不同的源 LUN 中，在进行快照时，必然要对 3 个部分的源 LUN 在同一时间点进行快照，才能实现在数据恢复时保持 3 个部分的数据恢复到同一时间点，否则造成了 3 个部分的数据无法恢复到同一时间点而失去数据相关性，数据的恢复也失去了意义。存储阵列的快照一致性激活很好的解决了这个问题，它在快照点同时冻结住一致性组中多个源 LUN 正在处理的 I/O，然后得到这些源 LUN 在同一时间点上一致的快照。

5） 数据持续保护

存储阵列对同一源 LUN 支持多个时间点的快照功能，结合主机侧的软件 BCManager 可以实现定时创建、删除快照功能，定时间隔在分钟级别；同样结合 BCManager 还可以设定策略为定时自动激活快照和停止快照操作。当多个时间点的快照采用循环的方式沿时间轴向前推进自动操作时，就非常方便且低成本的近似实现了持续数据保护的功能。

6） 快照副本

快照副本是对快照激活时刻的数据进行备份的一种技术，它不包含源快照激活后快照的私有数据。快照副本和源快照共享源 LUN 的 COW 卷空间，但私有空间是完全独立的，可以理解为快照副本就是一个可写快照，与源快照完全独立。对快照副本进行读写和普通快照的读写流程完全一致。

通过快照副本技术，可以获取相同快照的多份数据拷贝；创建多份快照副本可以用于不同的数据用途。

6.1.2. FS 快照（HyperSnap For File）

文件系统快照(HyperSnap For File)特性可以生成源文件系统在某个时间点上的一致性映像，在不中断正常业务的前提下，快速得到一份与源文件系统一致的数据副本。副本生成之后立即可用，并且对副本数据的读写操作不再影响源文件系统中的数据。因此通过文件系统快照技术就可以解决如在线备份、数据分析、应用测试等难题。用户可以通过多种方法使用文件系统快照。例如，它们可用于：

- 创建文件系统快照并将快照数据备份到磁带。
- 创建文件系统快照之后，在意外删除或破坏情况下，最终用户可以从快照恢复自己的文件。
- 远程复制、一体化备份等特性需要使用到文件系统快照，能将快照数据复制或备份到远端。

文件系统快照是基于 ROW 型（Redirect On Write，写时重定向）文件系统技术来实现的。所谓 ROW 型文件系统，是指向文件系统新写入或者修改写入数据时，新数据不会覆盖掉原来的旧数据，而是在存储介质上新分配空间来写入数据，此种方式保证了数据的高可靠性和文件系统的高扩展性。基于 ROW 技术的文件系统快照，可实现快速创建（秒级），并且除非原始文件被删除或者更改，快照数据并不占用额外的磁盘空间。

FS 快照（HyperSnap For File）有如下技术特点：

1） 零备份窗口

备份窗口是指应用所能容忍的完成数据备份的作业时间，实际上就是应用所能容许的停机时间。而传统的备份会导致文件系统的性能下降，甚至导致用户业务中断，所以传统的备份作业必须在应用停机或业务量较小的时候进行。而采用文件系统快照从事备份业务时，可以在线进行，备份窗口基本为零，无需业务停机。

2） 秒级快照

文件系统快照创建就是树根的拷贝，创建时间短，实现秒级快照。

3） 低性能损耗

文件系统的快照创建实现原理简单，下盘数据量极少，几乎不会对系统的性能产生影响。快照创建以后，文件系统的 IO 流程仅需在数据空间被释放之前，加入是否受快照保护的检查，并记录被快照保护而被文件系统删除的数据块空间，对文件系统性能影响几乎可以忽略。仅当快照删除后，数据的后台回收会跟文件系统业务竞争一些 CPU 和内存资源，但性能损耗也同样在一个低水位上。

4） 节省磁盘空间

采用文件系统快照获取源文件系统在快照时间点的一致性副本时，快照独占的文件系统空间由快照时间点后源文件系统的数据变化量决定，并且永远不会超

过快照创建时间点时的文件系统大小。在源文件系统数据量变化不大的情况下，文件系统快照通过很少的存储空间获得了源文件系统的一致性副本，非常节省硬盘空间。

5） 快照数据快速访问

文件系统的快照作为一个单独的目录呈现在文件系统的根目录中，用户可以通过访问快照对应的目录，快速读取访问快照的数据。在不需要快照回滚的场景下，可以方便的访问到快照时间点的数据，并且在当前文件系统的文件数据被破坏的情况下，通过文件/目录拷贝的方式进行数据修复。

在 windows 客户端下访问通过 CIFS 共享的文件系统，还支持针对某个文件或者目录进行还原，可以将某个文件或目录还原到某个时间点下快照的内容。只需要对要还原的目录或文件点击右键，选择以前的版本，可以看到包含此文件或目录的快照的所有时间点，可以选择其中一个时间点的数据进行还原。

6） 文件系统快速回滚

对于传统的离线备份，备份数据无法直接在线读取，必须经过较长时间的数据恢复过程才能够获得原数据在备份时间点的可用副本，从而实现数据的还原。而统一存储的文件系统快照可以直接将文件系统的树根替换成指定快照的树根，并清掉缓存数据，以实现文件系统快速回滚到指定的快照时间点。

用户需要小心使用回滚命令，因为在完成文件系统回滚之后，会自动删除回滚时间点之后的快照。

7） 定时快照实现持续数据保护

文件系统快照支持用户配置策略定时的自动进行快照创建操作，包括支持用户指定时间点创建快照和指定时间间隔创建快照。

文件系统支持的最大定时快照的个数视具体的产品型号而定，超过规格后，自动删除时间点最早的快照，而不需要用户进行介入。文件系统也支持用户主动删除定时创建的快照。

这样通过时间轴向前推进的多个时间点快照，就非常方便且低成本的实现了近似持续数据保护的功能。需要注意的是，采用快照实现的持续数据保护不能做到真正意义上 CDP（Continuous data protection），两个快照点之间的最小时间间隔决定了数据持续保护的粒度。

6.2 克隆（HyperClone）

6.2.1. LUN 克隆（HyperClone For Block）

HyperClone 在不中断业务的前提下，为存储系统的 LUN 建立一份某时刻的完整物理拷贝，并且在分裂后对物理拷贝的读写操作不会影响源 LUN 上的数据。

1. 技术原理：

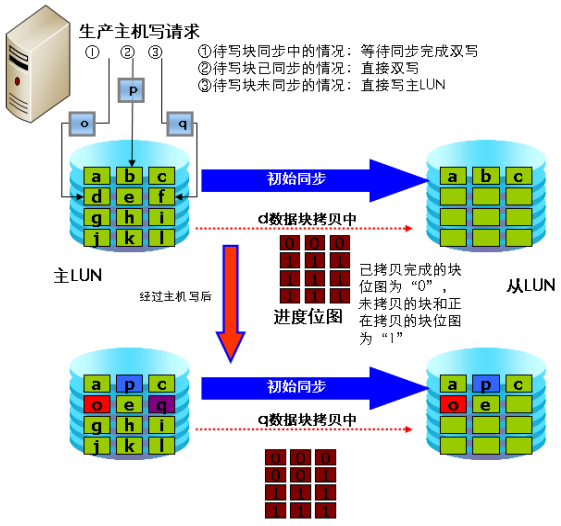
HyperClone 采用了位图和写前复制（copy-on-write）、位图和双写（并行写从 LUN 和主 LUN）相结合的技术方式来实现，其实现原理如下：

在克隆组中添加从 LUN 后，默认是需要经过一次从主 LUN 到从 LUN 的完全同步，通过进度位图显示拷贝进程，在初始同步时主 LUN 收到生产主机写请求，如下图所示，需要检查同步进度。

- 若要写入位置的数据块尚未拷贝到从 LUN 只需要写主 LUN 即可返回主机成功，稍后利用同步任务将整个数据块同步到从 LUN；
- 若要写入位置的数据块已经拷贝，需要分别写入主 LUN 和从 LUN；
- 若要写入位置的数据块正在拷贝，需要等待该数据块拷贝完成后分别写入主 LUN 和从 LUN；

初始同步完成以后可以将主、从 LUN 分裂，这时主、从 LUN 都可以用于独立的数据分析及测试，主、从 LUN 数据的变化互不影响，只是用进度位图记录主、从 LUN 上对应数据块的变化。

图1-51 克隆技术原理示意图



2. **技术特点：**

- **支持 1 对 16 模式：**

一个 LUN 最多可添加 16 个克隆从 LUN，利用 1 对多模式的克隆，可以同时备份出多份源数据，应用于不同方式的数据分析。

- **零备份窗口：**

在采用克隆从事备份业务时，无需用户的应用停机，备份窗口为零。

- **支持动态改变拷贝速度：**

支持动态手动改变拷贝速度来避免拷贝任务和生产业务的冲突。当存储阵列监测到系统业务繁忙时，在不影响业务的情况下，用户可以手动降低拷贝的速度，让阵列的系统资源为业务使用；当业务空闲时，动态提升拷贝的速度，加快拷贝进度，减少和业务高峰期的冲突。

- **支持反向同步：**

当主 LUN 的数据不完整或是被损坏需要还原时，可以通过从 LUN 到主 LUN 的增量反向同步来恢复原有的业务数据。

- **支持断开后自动恢复：**

在遇到一些故障时会进入断开状态，这些故障包括：主 LUN、从 LUN 失效等。当这些故障排除时，克隆会根据恢复策略进行恢复：如果恢复策略为自动恢复，克隆会自动进入“同步”状态，将有差异的数据增量同步到从 LUN；如果恢复策略为手动恢复，克隆会进入待恢复状态，等待用户手动启动同步。由于断开后的同步采用的是增量同步，可以大大地减少故障或灾难恢复的时间。

- **克隆一致性分裂：**

在 OLTP 应用中，通常需要对多个主 LUN 数据创建同一分裂时间点的一致性拷贝，才能将该应用分布在不同 LUN 中的关联数据保持在同一分裂时间点。克隆同时分裂多个从 LUN 很好的解决了这个问题，它在分裂时间点上同时冻结住多个主 LUN 数据，并且得到这些主 LUN 在同一分裂时间点上一致性的拷贝。

3. **技术效果：**

- **数据备份分析：**

可以通过对一个主卷生成多个物理副本，允许用户在同一时间运行多种业务对数据进行访问。

- **数据恢复和数据保护：**

克隆能够保护原有业务数据，当主 LUN 数据出现病毒入侵、人为损坏或物理的损坏时，可以选择合适的时间点数据副本，将从 LUN 的数据通过反向同步拷贝到主 LUN 上，实现数据的恢复。

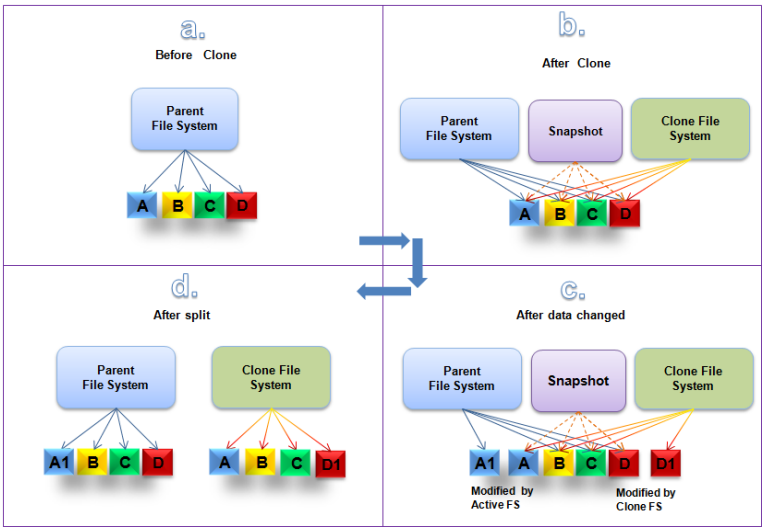
6.2.2. FS 克隆（HyperClone For File）

文件系统克隆特性是父文件系统某个时间点的副本，可以独立共享给客户端读写，从而满足快速部署、应用测试、容灾演练等场景。

1. 技术原理：

文件系统克隆是基于 ROW 技术的文件系统快照基础上实现的某个时间点的可读可写副本。

图1-52 文件系统克隆原理示意图



如图 a 所示，文件系统是 ROW 方式，数据写入不会覆盖原有数据，而是分配新磁盘空间写入；数据每次写入都会记录一个时间点信息，表明写入的时序。时间点实际是一个依次递增的序列号。

如图 b 所示，创建克隆文件系统步骤：

- 基于新建快照方式会在父文件系统中创建只读快照；
- 拷贝快照的根节点生成克隆文件系统的根节点；
- 在克隆文件系统中创建初始快照。

与创建只读快照类似，整个过程中不需要拷贝任何用户数据，因此整个过程

耗时极少，通常在一两秒内完成。并且在数据被修改之前，克隆文件系统与父文件系统共享数据。

如图 c 所示，从父文件系统修改 A 数据块时，会新分配一个 A1 数据块，并且由于有快照保护，A 数据块不会释放，因此修改父文件系统数据不会影响克隆文件系统；从克隆文件系统修改 D 数据块时，也会分配一个 D1 数据块，D 数据块写入时间点小于克隆文件系统初始时间点，D 数据块也不会释放，因此修改克隆文件系统数据也不会影响父文件系统；

如图 d 所示，分裂克隆文件系统步骤：

- 删除克隆文件系统中所有只读快照；
- 遍历克隆文件系统中所有对象的数据块，通过覆盖写触发共享数据在克隆文件系统中新分配数据块，从而达到共享数据分裂的目的；
- 删除父文件系统中关联快照。

分裂完成后克隆文件系统和父文件系统完全独立，没有依赖。分裂克隆文件系统时间根据共享数据大小而定。

2. 技术特点：

- **快速创建**

对于绝大部分场景，创建克隆文件系统秒级完成，克隆完成后克隆文件系统就可以独立共享给客户端读写；

- **节省存储空间**

克隆文件系统与父文件系统共享数据，在克隆文件系统对共享数据修改之前，这些共享数据不会占用额外的存储空间，因此创建克隆文件系统只会从 POOL 中消耗较少的空间。

- **低性能损耗**

克隆文件系统是基于父文件系统的快照生成的，因此创建克隆文件系统对父文件系统的性能影响几乎可以忽略。

- **克隆文件系统分裂**

克隆分裂将共享数据分开，分裂完成后克隆文件系统和父文件系统完全独立，没有依赖。

6.3 远程复制（HyperReplication）

基于统一存储软件平台开发，使得同方不同代次、不同档次(高中低端)存储产品的复制协议完全兼容，支持在同方及以后产品的高中低端的不同型号产品间创建远程复制，构建高度灵活的容灾解决方案。远程复制特性包括：LUN 同步远程复制、LUN 异步远程复制、FS 异步远程复制。

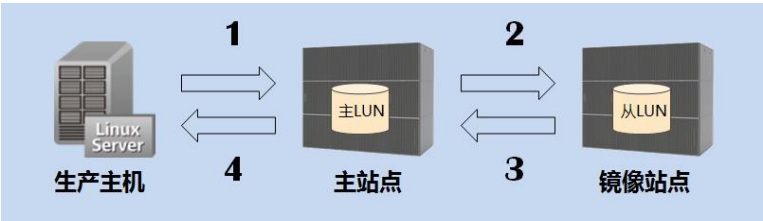
6.3.1. LUN 同步远程复制（HyperReplication/S For Block）

1. 技术原理：

同步远程复制，利用日志原理实现主、从 LUN 的数据一致性，首先当主站点的主 LUN 和远端复制站点的从 LUN 建立同步远程复制关系以后，会启动一个初始同步，也就是将主 LUN 数据全量拷贝到从 LUN，初始同步完成后，主 LUN 收到生产主机写请求，按照下面的流程进行 I/O 处理：

- 主站点接收生产主机写请求，记录这个 I/O 对应数据块的差异日志值为“有差异”；
- 同时把写请求的数据写入主 LUN 和从 LUN，写从 LUN 时需要利用配置好的链路将数据发送到远端复制站点；
- 判断写主 LUN 和写从 LUN 的执行结果，如果都成功，则将差异日志改为“无差异”，否则保留“有差异”，在下一次启动同步时重新拷贝这一个数据块；
- 主 LUN 返回生产主机写请求完成。

图1-53 同步复制技术原理示意图



2. 技术特点：

- **零数据丢失：**

对主、从 LUN 同时进行紧密的数据更新，保证 RPO 为 0。

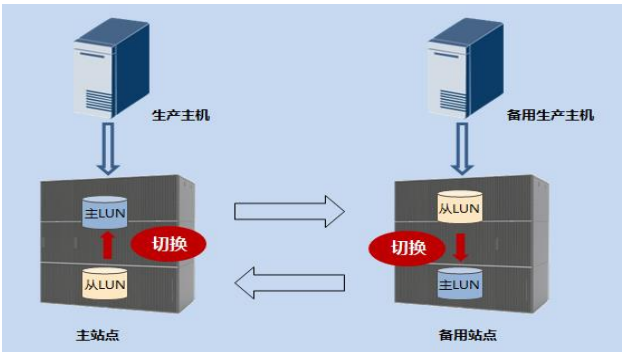
- **支持分裂模式：**

支持分裂模式，在分裂状态下，生产主机对主 LUN 的写请求只会写到主 LUN，满足用户的一些需求：如暂时性的链路维修、网络带宽扩容、需要从 LUN 保存某一个时间点的数据等等。

- **支持复制的主从切换：**

支持用户进行主从切换操作（见下图），主站点的主 LUN 在切换后变成了新的从 LUN，而复制站点的从 LUN 在切换后变成了新的主 LUN。经过一些在主机侧的简单操作以后（主要是将新主 LUN 映射给备用生产主机，也可提前映射），复制站点的备用生产主机接管业务并对新主 LUN 下发读写请求。

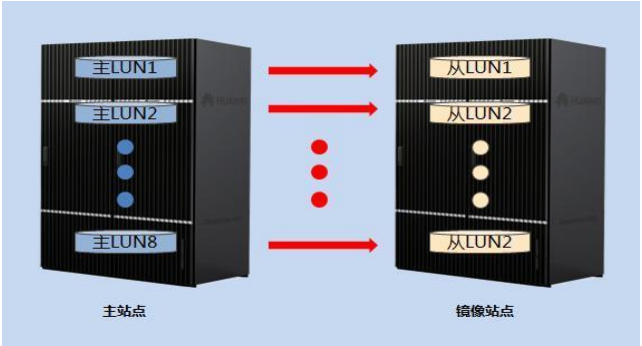
图1-54 同步复制主从切换示意图



- **支持一致性组：**

同步远程复制提供一致性组功能来保证多个 LUN 之间复制数据时间一致性。用户创建一致性组以后，可以将远程复制对添加到一致性组中，如下图。一致性组也可以进行分裂、同步和主从切换等操作，在进行这些操作时，一致性组的所有成员复制对同时进行分裂、同步、主从切换。此外，当遇到故障时，一致性组的所有复制对都会一起进入断开状态。

图1-55 同步复制一致性组示意图



3. **技术效果：**

同城的数据容灾备份：对于同步远程复制而言，每一个写请求都需要同时写到主站点和远端复制站点以后才会返回生产主机写完成，在主站点和复制站点相距较远的情况下，存储系统对前台应用程序的写延迟较高，不利于用户正常业务的运行。因此，同步远程复制 HyperReplication/S 主要应用于主站点和复制站点相距较近的容灾场景，如同城灾备。

6.3.2. LUN 异步远程复制（HyperReplication/A For Block）

1. **技术原理：**

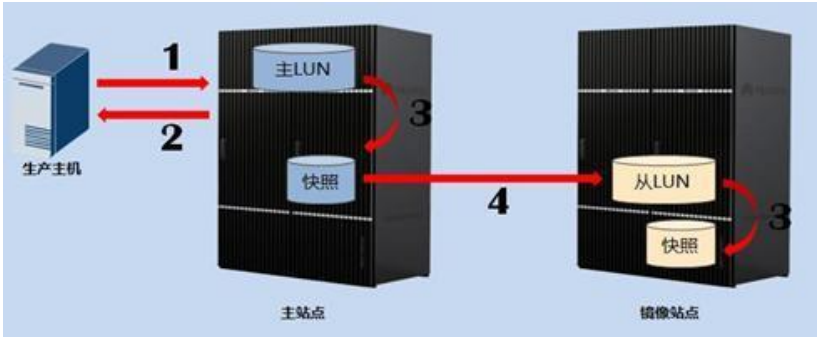
异步远程复制，与同步远程复制类似，当主站点的主 LUN 和远端复制站点的从 LUN 建立异步远程复制关系以后，也会启动一个初始同步，初始同步完成后，从 LUN 数据状态变为已同步或一致，然后按照下面流程处理（见下图）：

- **主 LUN 接收生产主机的写请求；**
- **写请求数据写入主 LUN 后，立即响应主机写完成；**

每当间隔一个同步周期（由用户设定，范围为 1-1440 分钟）以后，会自动启动一个将主 LUN 数据增量同步到从 LUN 的同步过程（如果同步类型为手动，则需要用户来触发同步）。在同步开始以前，先对主 LUN 和从 LUN 分别生成快照：主 LUN 的快照可以保证同步过程中读取到的主 LUN 数据是具备一致性的；从 LUN 的快照用于备份从 LUN 在同步开始前的数据，避免同步过程发生异常导致从 LUN 的数据不可用；

主 LUN 向从 LUN 同步数据时，读取主 LUN 快照的数据，复制到从 LUN。同步数据完成后，分别取消主 LUN 和从 LUN 的快照，然后等待下一个同步的到来。

图1-56 异步复制技术原理示意图



2. 技术特点：

- **支持数据压缩和数据加密**

针对 iSCSI 链路支持数据加密，加密算法为 AES256 算法；针对 iSCSI 链路支持数据压缩，数据压缩比按业务数据类型的不同区别较大，数据库业务最大能做到 4：1。

- **快速响应主机请求**

主机对主 LUN 的写请求在主站点完成后即可响应主机写完成，不必等待数据写到从 LUN。并且，数据由主 LUN 到从 LUN 同步过程是在后台进行的，不会影响主机对主 LUN 的正常访问。由于异步远程复制主 LUN 上的数据更新不是立即同步到从 LUN 的，所以数据遗失量取决于用户设置的同步周期，用户可以根据应用场景设置不同的同步周期（范围是 3 秒钟-1440 分钟，默认 30 秒钟）。

- **支持镜像分裂、主从切换和故障快速恢复**

异步远程复制与同步远程复制类似，同样拥有分裂、同步、主从切换和断开后恢复的功能。

- **支持一致性组**

支持一致性组的相关功能，包括一致性组的创建、删除、添加成员、删除成员、分裂、同步、主从切换等等。

3. 技术效果：

异地的数据容灾备份：对于异步远程复制而言，存储系统对前台应用程序的写延迟与主站点和复制站点的距离无关，所以异步远程复制 HyperReplication/A 适用于长距离或网络带宽有限情况下的容灾场景。

6.3.3. FS 异步远程复制（HyperReplication/A For File）

HyperReplication 文件系统异步远程复制提供对文件系统的远距离数据容灾功能。它将生产端的文件系统的全部内容复制到灾备端的文件系统中，适用于需要在跨异地的数据中心间进行容灾，同时降低对生产业务的性能影响的场景。它也支持阵列内两个文件系统进行异步复制，适用于本地数据容灾、数据备份、数据迁移等应用场景。

HyperReplication 文件系统异步远程复制基于文件系统对象层，周期性的

同步主、从 FS 的数据，上一次同步以来主 FS 上发生的所有变化会在下一次同步时写到从 FS 上。

1. 技术原理：

1） 基于对象层的复制

HyperReplication 文件系统异步远程复制采用基于对象层的方式进行数据复制。文件系统的所有内容，比如文件、目录、文件属性，都是由对象构成。基于对象层的复制直接将对象从主文件系统复制到从文件系统，不需要关心复杂的文件层的信息，比如文件与目录间的依赖关系、各种文件操作，从而使复制变得更加简单高效。

2） 基于 ROW 快照的周期性复制

HyperReplication 文件系统异步远程复制采用基于 ROW 快照的周期性的方式进行数据复制。

周期性复制可以提高复制效率和带宽利用效率。在一个周期中，如果主机重复写入相同地址的数据（比如对同一文件相同地址的重复修改），只需要将最后一次写入的数据进行复制。

文件系统及其快照都是采用 ROW 方式处理数据写入，不管文件系统是否带有快照，数据都是写入新分配的地址空间，创建快照后几乎不会带来性能影响。因此，文件系统异步远程复制对生产业务的性能影响也很小。

写入的数据在后台周期性地复制到从 FS。复制周期由用户设定，每个周期内数据的变化会记录增量信息，增量信息记录数据变化的地址，不会记录数据内容。每次周期复制过程中，当增量数据没有传输完成时，从 FS 还不能构成完整的文件系统，因此每次周期复制完成时，从 FS 形成数据一致性点后，会创建从 FS 的快照，如果下一次周期复制过程中断（生产端发生故障、链路发生故障等原因），当用户需要使用从 FS 时，文件系统异步远程复制可以将从 FS 回滚到上个周期完成时的快照点，获得一致性数据。

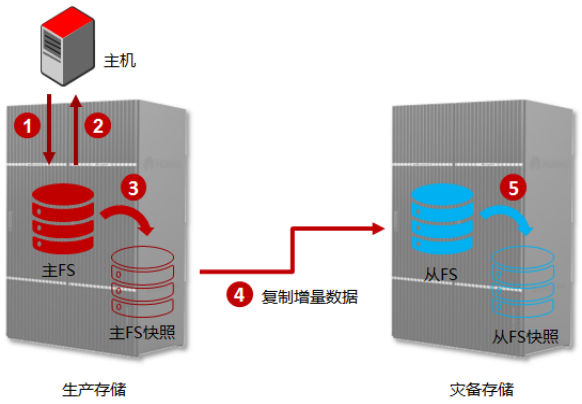
3） 支持阵列内文件系统异步复制

阵列内文件系统异步复制支持支持阵列内同一个租户的两个文件系统之前容灾、备份、迁移等功能。通过建立阵列内异步复制，无需购买远端设备和网络资源，可以快速构建阵列内备份功能，降低成本提升效率。

4） 支持主端文件系统快照同步到从端

支持同步主端文件系统的用户快照/定时快照同步到从端，用户可以设置从端快照保留个数。在主从切换后，新主端依然具有历史时间点的备份快照。可以设置从端快照保留个数，允许在从端保留更多的快照，降低本端文件系统容量压力；允许用户在从端手动创建快照、删除快照、修改快照属性。

图1-57 文件系统异步远程复制原理



- 主机下发写 I/O；
- 主机数据写入主 FS 即可返回；
- 每个同步周期开始时，文件系统异步远程复制创建主 FS（主文件系统）的快照；
- 根据上一周期复制完成到本周期开始这段时间内的增量信息，读取快照的数据复制到从 FS；
- 增量复制完成后，从 FS 的内容与主 FS 的快照内容相同，从 FS 形成数据一致性点。

2. 技术特点：

- 分裂和增量重同步

当用户希望暂停复制时，可以对远程复制进行分裂。分裂将停止从主 FS 到从 FS 的数据复制。

分裂后主机写入的数据会记录增量信息。分裂后可以再次重同步，在重同步时，已经复制过的数据不会再复制，会根据增量信息只复制主从之间有差异的数据。

分裂常用于计划性的设备维护的场景，比如存储阵列升级、复制链路变更。

在这类维护场景下，通常降低各类并发处理的任务会使系统更加可靠，等到维护结束后，再重启或继续任务。

- **故障断开和自动恢复**

当因为某种故障（比如链路断开）而导致远程复制无法再继续从主 FS 到从 FS 的数据复制时，远程复制会进入异常断开状态。在异常断开状态下，主机写入的数据会记录增量信息。当故障排除后，远程复制会自动恢复，进行增量重同步，不需要人为干预。

- **从 FS 可读写和增量 failback**

通常情况下，从 FS 可读、不可写。从 FS 可读时，读的是上一次复制完成时的快照上的数据，当下一次复制完成时，会自动切换到读最新快照上的数据。

从 FS 可读常用于在复制过程中需要读取从端数据的场景。

- **在满足下列条件时，可以将从 FS 设置为可读写：**

已初始同步完成。对于异步远程复制，完成初始同步后从端数据状态为完整。

远程复制处于分裂、或异常断开状态。

设置从 FS 可读写时，如果从 FS 处于复制过程未完成的阶段，数据不一致，远程复制会将从 FS 回滚到上一次复制完成时的快照点。

设置为可读写后，主机对从 FS 写入数据时，远程复制会记录增量信息，后续用于增量重同步。恢复复制时，可以选择由主到从复制，也可以选择由从到主复制（需做主从切换，然后再启动同步）。复制启动前，远程复制会先将目标端回滚到一个快照点上，该快照点与源端的过去的某个快照点数据相同，然后根据源端的从该快照点到当前的增量信息进行增量重同步。

设置从 FS 可读写常用于灾备场景。

- **主从切换**

FS 异步复制支持在分裂和异常断开状态下将主 FS 与从 FS 的角色互换，原来的主 FS 成为从 FS，原来的从 FS 成为主 FS。主、从角色决定了数据复制方向，数据会由主 FS 向从 FS 同步。

主从切换常用于灾备场景中 failback 过程。

- **快速响应主机 IO**

文件系统异步远程复制的所有复制增加的 IO 处理都是在后台进行。主机数

据写入 Cache 后，即可返回，没有额外的处理。Cache 在下刷数据时，才会记录增量信息，以及进行快照处理。因此，可以快速响应主机 IO。

6.4 阵列双活（HyperMetro）

HyperMetro 是阵列级的 A/A 双活技术，部署双活的两套存储系统可以放在同一个机房、同一个城市或者相距 300 公里的两地。

同方超强 RS6800 系列“芯”系列存储系统同时支持 SAN 双活（HyperMetro For Block）和 NAS 双活（HyperMetro For File）。

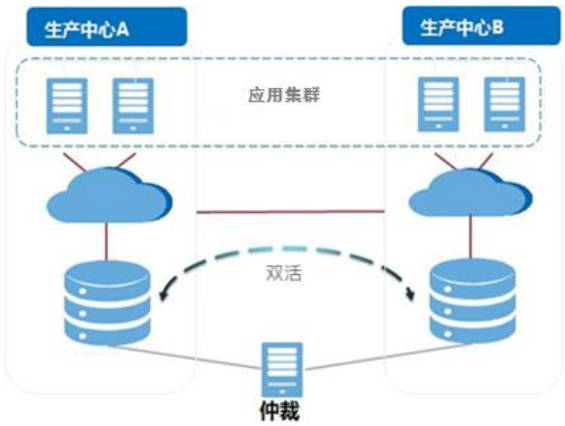
同方超强 RS6800 系列“芯”系列存储系统支持与同方超强 RS6800 系列存储系统组成跨代次双活解决方案。

6.4.1. 阵列双活（HyperMetro For Block）

HyperMetro 使来自两套存储阵列的两个 LUN 数据实时同步，且都能提供主机读写访问。当任何一端磁盘阵列整体故障的情况下主机将切换访问路径到正常的一端继续业务访问；当磁盘阵列间链路故障时只有一端继续提供主机读写访问，具体由那端提供服务将取决于仲裁的结果。仲裁服务器部署在第三方站点，用于两套存储阵列之间链路中断时，仲裁由哪边的存储继续提供业务访问。

HyperMetro 既支持 FC 组网（8G/16G/32G），也支持 IP 组网（10GE/25GE/40GE/100GE）。

图1-58 阵列双活示意图



阵列双活示意图

- HyperMetro 主要技术特点：

免网关双活方案：组网简单，容易部署；减少一个故障点，可靠性更好；避免了网关设备额外引入的 0.5~1ms 的时延，性能更好。

Active-Active 双活：真正 A/A 双活，两个数据中心的存储都支持业务读写，上层应用系统可以充分应用该业务能力实现业务负载分担部署，实现跨数据中心的业务负载均衡。

地域访问优化：多路径软件，针对双活场景做了优化，能够识别地域位置信息，减少跨站点访问，从而减少时延提供性能。主机多路径虽然能够从本地或异地存储读取数据，但在本地磁盘阵列正常运行的情况下，多路径软件优先读写本地磁盘阵列，避免主机跨数据中心读写数据。

FastWrite 特性：正常的 SCSI 写流程中，写请求有“写分配（Write Alloc）”和“写数据（Write Data）”这两次交互，一个写请求需要在站点间往返两次才能完成。Fastwrite 特性优化存储传输协议，提前在目标端预留接收写请求的缓存空间，省掉“写分配”环节，变为只要 1 次交互。该特性将阵列之间数据同步时延缩短一半，提升了整体双活方案性能。

按业务粒度仲裁：HyperMetro 可以实现按业务为粒度仲裁，即站点间链路故障后，可以按照配置，有些业务跑在 A 数据中心，有些业务跑在 B 数据中心。相比传统仲裁只有单边设备存活的方案，可以减少主机和存储资源预留，使业务负载更均衡。仲裁粒度支持按 LUN 或按一致性组。

链路质量自适应：如果两个数据中心间存在多条链路，HyperMetro 特性会根据各条链路质量，自动在链路之间均衡负载。系统会动态监控链路质量，动态调整两条链路的负载分担比例，以尽量降低重传率，提升网络性能表现。

现有特性兼容：HyperMetro 支持与 TFCQ Thin、TFCQ Tier、TFCQ QoS、TFCQ Cache 等现有特性同时使用，支持对通过 TFCQ Virtualization 特性接入的异构 LUN 配置双活，可以与 HyperSnap、HyperClone、HyperMirror、HyperReplication 等特性一起组合成更复杂的高级数据保护方案（如本地双活+异地复制的两地三中心容灾方案）。

支持双仲裁：HyperMetro 支持两个仲裁服务器，其中一个故障之后，无缝切换到另外一个仲裁服务器，降低单点故障风险，提升双活可靠性。

6.4.2. 阵列双活（HyperMetro For File）

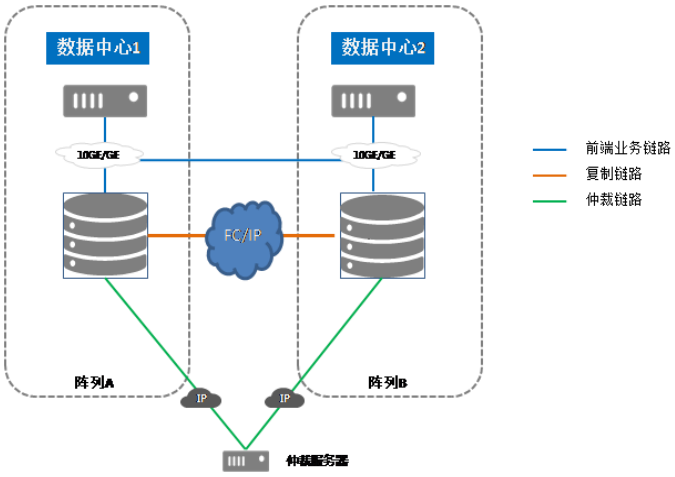
HyperMetro 使主机能够将两个存储系统的文件系统视为单个存储系统上的单个文件系统，并且使两个文件系统上的数据相同。NAS 双活由主端提供数据读写服务，数据实时同步至从端；当主站点发生故障时，以租户为粒度进行双活切换，从站点将自动接管服务，而不会对应用程序造成任何数据丢失或中断。

NAS 双活为客户提供以下的收益：

- 跨站点的高可用持续保护
- 简易的管理
- 避免数据丢失的风险，减少系统宕机时间以及快速的灾难恢复
- 对应用和用户不感知故障处理

NAS 双活既支持 FC 组网（8G/16G/32G），也支持 IP 组网（GE/10GE/25GE/40GE/100GE）。

图1-59 NAS 双活逻辑组网图



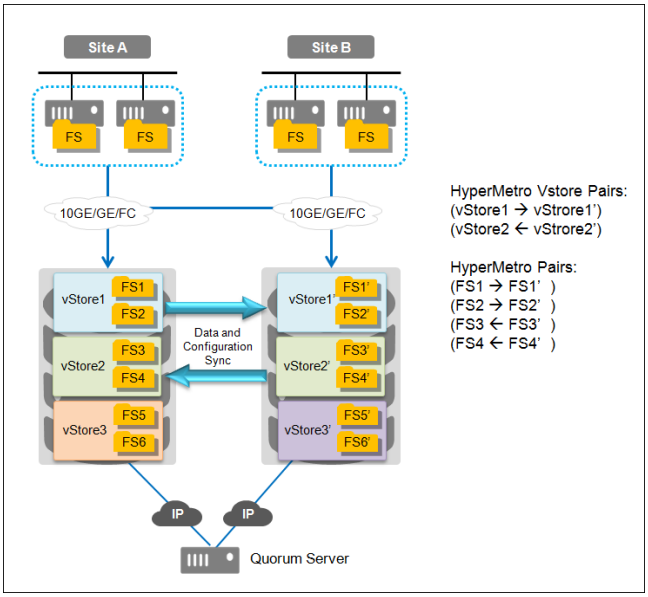
- NAS 双活主要技术特点：

免网关双活方案：免网关设计使主机 I/O 请求，无需经过存储网关转发，避免了网关转发引起 I/O 时延；同时减少网关故障点，提高方案可靠性；显著降低双活组网复杂度，便于维护。

组网简单：整合双活数据复制链路、配置同步链路和心跳链路到一个物理网络，简化了两个数据中心的组网，两套存储系统间的复制链路可以使用 IP 或 FC 链路。考虑到前端业务链路一定是 IP 网络，所以 HyperMetro 能工作在全 IP 网络环境下，以降低构建成本。

基于 vStore 的 NAS 双活：传统的 NAS 双活，主要是通过将集群的节点分别部署在两个数据中心，从而实现数据中心间的双活，不能进行灵活资源配置和调优。而 NAS 双活是通过将部署在两个数据中心的 vStore 来实现双活关系，实现了 vStore 粒度的数据和配置的实时镜像，每个 vStore 双活都有自己独立的仲裁结果，提供了真正的 vStore 层面的跨站点高可用能力，这使得客户可以更灵活的部署业务，实现更好的负载均衡，上层应用更高效。一个 vStore Pair 包括两个互为主备的 vStore，它们组成了跨站点的高可靠关系。当一个存储系统发生故障时，或两个存储系统间的连接断开时，双活仲裁以 vStore Pair 为单位发起仲裁申请。两个 vStore 中的资源互为冗余，为客户提供服务，从而实现故障时，业务不中断。

图1-60 基于 vStore 的 HyperMetro 架构



自动恢复:如果站点 A 发生故障导致 vStore Pair 工作站点切换到站点 B，在 A 故障恢复后，可以自动发起重同步，而无需人工接入。在重同步完成后，双活变为正常状态，该状态下站点 B 再发生故障，双活可以继续切换到站点 A，业务不中断。

易于升级扩展:当用户需要为原有业务配置双活时，只需要购买双活 license 并升级到最新的软件版本，就可以和另一台阵列配置双活，而不需要额外的数据迁移过程。所以用户可以选择购买设备时就初始配置双活还是后续扩展到双活。

FastWrite 特性：正常的 SCSI 写流程中，写请求有“写分配（Write Alloc）”和“写数据（Write Data）”这两次交互，一个写请求需要在站点间往返两次才能完成。Fastwrite 特性优化存储传输协议，提前在目标端预留接收写请求的缓存空间，省掉“写分配”环节，变为只要 1 次交互。该特性将阵列之间数据同步时延缩短一半，提升了整体双活方案性能。

链路质量自适应：如果两个数据中心间存在多条链路，HyperMetro 特性会根据各条链路质量，自动在链路之间均衡负载。系统会动态监控链路质量，动态调整两条链路的负载分担比例，以尽量降低重传率，提升网络性能表现。

现有特性兼容：HyperMetro 支持与 TFCQ Thin、TFCQ QoS、TFCQ Cache 等现有特性同时使用，可以与 HyperSnap、HyperReplication、HyperVault 等特性一起组合成更复杂的高级数据保护方案（如本地双活+异地复制的两地三中心容灾方案）。

支持双仲裁：HyperMetro 支持两个仲裁服务器，其中一个故障之后，无缝切换到另外一个仲裁服务器，降低单点故障风险，提升双活可靠性。

6.5 一体化备份（HyperVault）

一体化备份（HyperVault）特性可以实现系统内和系统间的文件系统数据备份和恢复。

HyperVault 可以工作在以下两种模式：

- 本地备份：

存储系统内部的备份，基于文件系统的快照机制，对需要备份的文件系统按照一定的定时策略进行备份，生成备份副本，同时对生成的备份副本按照策略保留一定的数量，默认保留 5 份。

- 异地备份：

存储系统之间的备份，基于文件系统的远程复制技术，对需要备份的文件系统按照定时策略进行备份：在主存储端创建一个备份快照，然后获得和上一次异地备份时的备份快照的之间的差异数据，将差异数据拷贝到备份存储端文件系统，备份完成后，在备份存储端文件系统创建一个快照，并对生成的备份快照按照策略保留一定的数量，默认保留 35 份。

技术特点：

1） 成本更节约

一体化备份功能完美融入主存储，用户通过主存储自带管理软件，配置灵活的备份策略，完成备份功能，不依赖于商业的备份软件。

2） 备份效率更高

一体化备份的本地备份采用快照技术进行备份，可以实现秒级备份；异地备份除初始备份为全备外，后续只对增量数据块进行备份，相比以文件单位的备份软件效率更高。

3） 恢复效率更高

一体化备份本地恢复利用阵列的快照回滚技术，不需要额外的数据解析，实现秒级恢复；本地恢复不能满足恢复要求时，可采用异地恢复，异地备份采用增量的方式进行恢复。每一份备份数据从逻辑上来看都为业务数据的一次全备，备份数据以原有格式存放，可以被立即访问。

4） 管理更简单

采用两台设备融合主存和备份的功能，不需要采用主存+备份软件+备份介质的复杂组合，只需使用存储自带管理软件即可，管理简单易懂。

6.6 LUN 拷贝（HyperCopy）

LUN 拷贝(HyperCopy)特性可以将磁盘阵列中的源 LUN 数据 Copy 到目标 LUN 中，支持阵列内和阵列间的拷贝功能。

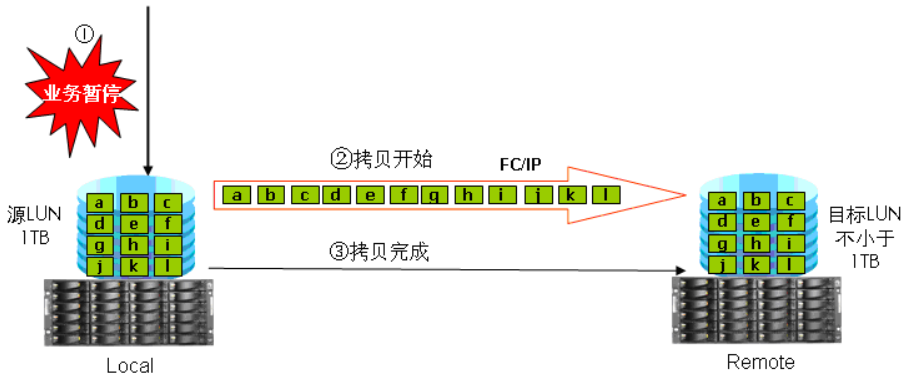
LUN 拷贝采用全量拷贝方式，即把源 LUN 数据从头到尾进行一次完整的至目标 LUN 的拷贝，其工作原理如下图所示：

暂停业务：全量 LUN 拷贝须在业务离线的条件下进行，以避免启动拷贝后主机业务被中断；

完整拷贝开始：数据拷贝可通过 FC 链路或者 IP 链路进行拷贝，但必须保证目标 LUN 容量不小于源 LUN，否则拷贝将失败；

拷贝过程中，源 LUN 有拷贝进度显示，直至拷贝完成。

图1-61 LUN 拷贝工作原理



LUN 拷贝也支持读取快照卷的数据进行全量拷贝，无需对源 LUN 停止业务，零备份窗口。

- 技术特点：

1） 支持多种拷贝方式

LUN 拷贝不仅支持阵列间的拷贝，也支持在同一台阵列内的 LUN 拷贝。不仅支持本阵列向目标阵列拷贝数据，也支持目标阵列向本阵列拷贝数据。支持一对多 LUN 拷贝，为一份数据同时备份多份数据。

2） 支持动态改变拷贝速度

LUN 拷贝支持动态改变拷贝速度来避免和生产业务的冲突。当存储系统监测到系统业务繁忙时，可以降低 LUN 拷贝的速度，让存储系统的资源为业务使用；当业务空闲时，可以提升 LUN 拷贝的速度，加快拷贝进度，减少和业务高峰期的冲突。

3） 支持第三方存储的 LUN 拷贝

LUN 拷贝可以在存储系统内或者与经认证的第三方存储系统的系统间进行，其中 LUN 拷贝的支持方式如表 3-1 所示。

表1-16 LUN 拷贝的支持方式

源 LUN 位于的存储系统	目标 LUN 位于同方存储系统	目标 LUN 位于经同方认证的第三方存储系统
同方存储系统	支持	支持
经认证的第三方存储系统	支持	不涉及

4） 支持基于 IP 网络的 LUN 拷贝

对于阵列间的 LUN 拷贝，HyperCopy 不仅支持传统的 FC 链路的 LUN 拷贝，

还支持基于 IP 网络的 LUN 拷贝，为客户提供了更加灵活的选择。由于 IP 网络的大量普及，基于 IP 网络的 LUN 拷贝成本更低，更加容易部署，维护也更加简单方便。

6.7 卷镜像（HyperMirror）

卷镜像(HyperMirror)特性可以使一个 LUN 可以拥有 2 个物理副本。每个副本的空间可以来源于本地存储池，也可以来源于外部 LUN。每个副本都具有与镜像 LUN 相同的虚拟容量。当服务器对镜像 LUN 执行写操作时，系统会将数据同时写入每个副本。当服务器对镜像 LUN 执行读操作时，系统会选取其中一个副本进行读取。如果其中一个镜像副本暂时不可用（例如，由于提供存储池的存储系统不可用），那么服务器仍然可以访问 LUN。系统会记住执行写操作的 LUN 区域，并会在镜像副本恢复后，对这些区域进行再同步。

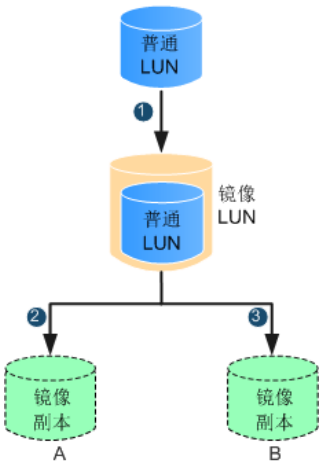
1. 技术原理：

卷镜像的实现过程分为三个阶段：创建镜像 LUN、同步和分裂。

- 创建镜像 LUN

镜像 LUN 的创建过程如 6.7 卷镜像（HyperMirror）所示。

图1-62 卷镜像创建过程



对一个普通 LUN（本地 LUN 或外部 LUN）执行创建镜像 LUN 操作，此时镜像 LUN 完全继承普通 LUN 的存储空间；同时继承普通 LUN 的基本属性和业务，主机侧不中断业务。

创建镜像 LUN 过程中会在本地自动生成一个镜像副本 A，普通 LUN 变为镜像

LUN，并将数据存储空间交换到镜像副本 A，镜像 LUN 从镜像副本 A 中同步数据。

此后需再给镜像 LUN 添加一个镜像副本 B，创建之初从镜像副本 A 同步数据。此时普通 LUN 具有空间镜像功能，同时拥有镜像副本 A 和镜像副本 B 两份镜像数据。

镜像 LUN 创建完成后，主机下发 I/O 的情况：

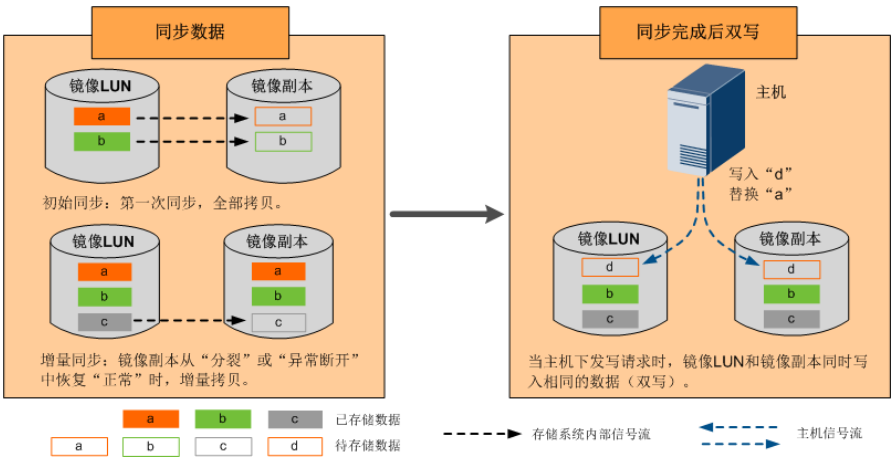
当主机对镜像 LUN 下发读请求时，存储系统会以轮询方式在镜像 LUN 和镜像副本之间进行读操作。当镜像 LUN 或者某个镜像副本故障时，主机侧业务不受影响。

当主机对镜像 LUN 下发写请求时，存储系统会以双写方式对镜像 LUN 和镜像副本进行写操作。

- 同步

同步过程的原理如下图所示。当一个副本从故障恢复或从数据不完整恢复到数据完整的过程中，增量从完整的镜像副本同步数据，最终达到镜像副本间的数据完全一致。

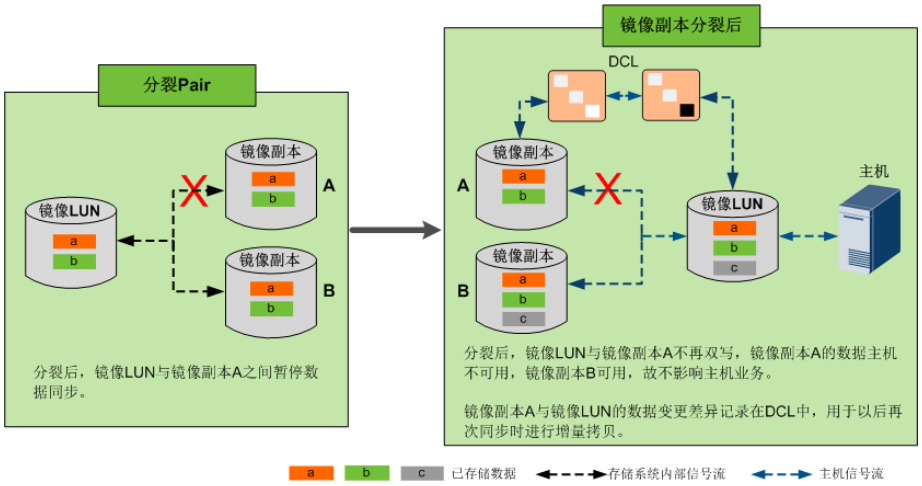
图1-63 卷镜像同步过程



- 分裂

在业务需要隔离一个副本时，可执行分裂操作，将副本从镜像 LUN 中隔离，此时镜像 LUN 不具备数据镜像功能，同时镜像副本记录此后的数据差异，当需要重新建立镜像关系时，根据差异增量同步差异数据。

图1-64 卷镜像分裂过程



2. **技术特点：**

- 存储系统内数据可靠性

通过对业务 LUN 创建卷镜像，使其同时拥有两分完全独立的数据副本，其中一份数据副本故障，不影响主机业务的运行，大大提高的数据的可靠性。

- 异构阵列的可靠性

通过 TFCQ Virtualization 特性，将异构阵列的 LUN 接管，对其创建镜像 LUN，添加一个本地镜像副本，防止由于异构阵列不稳定或链路异常导致的业务中断。

- 主机性能影响小

镜像数据副本位于业务 LUN 的 CACHE 之下实现，且镜像空间之间实现并发写和轮询读技术，不影响主机业务性能。

- 主机业务连续性

通过在线将正在运行业务的 LUN 创建镜像副本，主机业务不感知 LUN 数据空间的变化。

6.8 WORM（HyperLock）

随着科学技术的进步和社会发展，信息呈爆炸式增长，数据的安全访问和应用的问题逐渐受到人们的重视，例如法院案件、医疗病例、金融证券等，这些重要的数据按照法律规定在指定的时间周期内只能读不能写。因此需要对此类数据进行防篡改保护。WORM（Write Once Read Many）特性提供一次写入多次读取技

术，是存储业界常用的数据安全访问和归档的方法，旨在防止数据被篡改，实现数据的备案和归档。

HyperLock 特性支持文件被写入完成后通过去掉文件的写权限，使其进入只读状态。在该状态下文件只能被读取，无法被删除、修改或重命名。通过配置 WORM 特性对存储数据进行保护后，可以防止其被意外篡改，满足企业或组织对重要业务数据安全存储的需求。

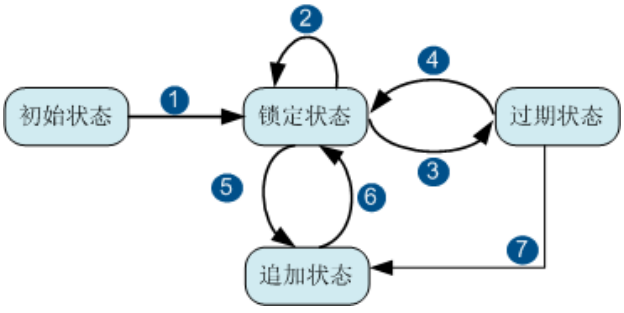
具有 WORM 特性的文件系统（以下简称 WORM 文件系统）只能由管理员进行设置。根据管理员权限不同，WORM 文件系统可分为法规遵从模式（Regulatory Compliance WORM，简称 WORM-C）和企业遵从模式（Enterprise WORM，简称 WORM-E）。法规遵从模式主要应用于遵从法规施行数据保护机制的归档场景，而企业遵从模式主要应用于企业内部管理。

- WORM 原理

WORM 技术使文件只能写入一次数据，不能重复写入且不允许被修改、删除或重命名。WORM 特性是在普通文件系统的基础上增加了 WORM 属性，使 WORM 文件系统内的文件在保护期内只能被读取。创建 WORM 文件系统后，通过 NFS 或者 CIFS 协议映射给应用服务器。

通过使用 WORM 特性，存在于 WORM 文件系统中的文件可以在初始状态、锁定状态、追加状态以及过期状态之间进行转换，从而防止重要数据在指定周期内被意外或恶意篡改。各状态间的转换关系如图 1-65 所示。

图1-65 文件状态的转换



初始状态 -> 锁定状态：文件可以通过两种方式从初始状态转换至锁定状态。

在自动锁定模式打开的情况下，文件结束修改后超过“锁定等待时长”自动进入锁定状态。

手动将文件设置成锁定状态，在设置锁定状态前可以明确给出文件的保护期，也可以使用系统默认的保护期。

锁定状态 -> 锁定状态：当文件处于锁定状态时，可以手动延长文件的保护期。保护时间只能延长不能缩短。

锁定状态 -> 过期状态：在 WORM 文件系统的法规时钟超过文件过期时间之后，文件就会由锁定状态转换至过期状态。

过期状态 -> 锁定状态：通过延长文件的保护期，可以实现文件从过期状态转换至锁定状态。

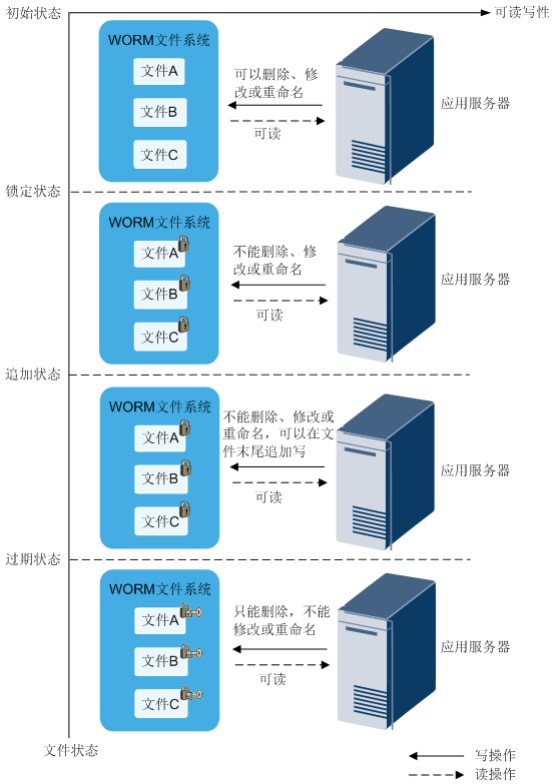
锁定状态 -> 追加状态：通过去掉文件的只读权限，将处于锁定状态的文件设置成追加状态。

追加状态 -> 锁定状态：通过设置文件为只读状态，将处于追加状态的文件设置为锁定状态，以保证文件不再被修改。

过期状态 -> 追加状态：可手动将处于过期状态的文件设置成追加状态。

用户根据业务需求将需要保存的文件放入 WORM 文件系统中，并设置文件的 WORM 属性使其进入保护状态。WORM 文件系统中文件在各状态的读写过程如图 1-66 所示。

图1-66 WORM 文件系统中文件的读写



6.9 两地三中心（3DC）

同方超强 RS6800 系列“芯”系列存储系统支持丰富的 SAN 3DC 解决方案，包括：

- 双活+同步复制 级联、并联组网
- 双活+异步复制 级联、并联组网
- 双活+异步复制 环形组网
- 同步复制+异步复制 级联、并联组网
- 异步复制+异步复制 级联、并联组网
- 同步复制+异步复制 环形组网

同方超强 RS6800 系列“芯”系列存储系统支持的 NAS 3DC 解决方案包括：

- 双活+异步复制 级联、并联组网
- 双活+一体化备份 级联、并联组网

3DC 解决方案可以在不增加外部网关的情况下，从单站点平滑扩展到三站点保护。

6.10 轻松接云——数据灵活兼容无缝衔接

同方超强 RS6800 产品不仅支持现代的 OpenStack 云原生应用，同时能够支持多云环境，并具备良好的扩展性、易用性、灵活部署、敏捷弹性等特性。它支持应用程序/数据在边缘、核心、容器、云中透明部署、移动和管理，确保数据在合适的时间，存放在合适的地方，并满足工作负载不同服务级别的要求。并可提供化繁为简的异构存储管理功能，为企业现有存储资产提供支持。

7 产品规格

7.1 硬件规格

硬件规格包括硬件配置、端口规格、硬盘规格、尺寸和重量、电气规格和可靠性规格。表 1-17 描述了硬件规格各个分类的详细信息，可快速查找需要的规格信息。

表1-17 硬件规格分类说明

分类名称	详细说明
硬件配置	介绍存储系统内存、硬盘和端口等主要硬件的配置情况。
端口规格	介绍各个接口模块包含的端口数量、每控制器支持最大接口模块数量等与端口相关的详细规格。
硬盘规格	介绍硬盘尺寸、转速、容量和重量。
尺寸和重量	介绍控制框、硬盘框的尺寸和重量。
电气规格	介绍控制框、硬盘框的电气规格。
可靠性规格	介绍存储系统的可靠性规格。

7.1.1. RS6800 硬件规格

1. 硬件配置

配置项目	RS6800
处理器物理核数（单控制器）	96
Cache 容量（单控制器）	• 256GB • 512GB • 1TB
单框最大控制器数量	2
最大控制器数量（直连组网）	8
最大控制器数量（交换机组网）	16*
最大硬盘配置数量（整系统）	3200
最大硬盘配置数量（单控制框）	2400
控制框配置	4U 控制框，不能容纳硬盘。
支持的硬盘框类型	• 2U SAS 硬盘框，可容纳 25 个 2.5 寸硬盘。

配置项目	RS6800
	• 4U SAS 硬盘框，可容纳 24 个 3.5 寸硬盘。 • 2U 智能 SAS 硬盘框，可容纳 25 个 2.5 寸硬盘。 • 2U 智能 SAS 硬盘框，可容纳 12 个 3.5 寸硬盘。 • 2U 智能 NVMe 硬盘框，可容纳 36 个 Palm 硬盘。
最大硬盘框数量	• 2U SAS 硬盘框：100。 • 4U SAS 硬盘框：100。 • 2U 智能 SAS 硬盘框（25 盘位/12 盘位）：80。 • 2U 智能 NVMe 硬盘框：80。
后端通道（端口）最大级联硬盘框数	• 每对 SAS 端口，最多级联 4 个 SAS 硬盘框，推荐 2 个； • 每对 RDMA 端口，最多级联 2 个智能 SAS 硬盘框，推荐 2 个；
支持的硬盘类型	SSD、SAS、NL-SAS、Palm 规格的 NVMe SSD
支持的可热插拔主机接口模块类型	• 8Gbit/s FC • 16Git/s FC • 32Git/s FC • GE • 10GE • 25GE • 40GE • 100GE
支持的可热插拔后端级联模块类型	• 12Gbit/s SAS V2 • 100Gb RDMA
每个控制器支持的可热插拔 I/O 接口模块数量	28（四控共享）
板载 I/O 端口（每控制器）	无
级联 SAS 线缆长度	• 缆：1m、2m、3m、5m • 光缆：15m
后端 RDMA 线缆长度	• 电缆：1m、2m • 光缆：10m
支持的 Scale-out 线缆长度	光缆：10m
支持的前端线缆和类型	DLC to DLC 多模（8G/10G/16G/25G/32G）：3m、10m、20m、50m、100m MPO to MPO 多模（100G/40G）：3m、10m、20m、50m、100m LC to LC 单模（16G FC/10GE）:10km（提供光模块，光纤客户自配）

配置项目	RS6800
主要部件冗余配置	• BBU：1（每控制器） • 电源：2+2 • 风扇：6+1（每控制器）
*：如有*号标注处规格需求，请联系华为销售人员。	

2. 端口规格

接口模块的最大端口数量（单接口模块）	RS6800
TFCQ IO 接口模块	用作前端，每个模块 4 个光端口，支持 8Gbit/s FC、16Gbit/s FC、32Gbit/s FC、10GE、25GE（不支持协商到 GE）。
GE 接口模块	用作前端，每个模块 4 个 1Gbit/s ETH 电端口
40GE/100GE 接口模块	用作前端，每个模块 2 个 40Gbit/s 或 100Gbit/s ETH 光口。
25Gb RDMA 接口模块	用作多控间直连组网，每个模块 4 个光端口，支持 25Gbit/s RoCE。
SO 100Gb RDMA 接口模块	用作多控间交换机组网，每个模块 2 个光端口，支持 100Gbit/s RoCE。
BE 100Gb RDMA 接口模块	用作后端智能硬盘框组网，每个模块 2 个光端口，支持 100Gbit/s RDMA。
12Gbit/s SAS 级联模块	用作后端，每个模块 4 个端口。

最大端口数量	RS6800
8Gbit/s FC 端口	80
16Gbit/s FC 端口	80
32Gbit/s FC 端口	80
GE 端口	20
10GE 端口	20
40GE 端口	10
100GE 端口	10
25GE 端口	20
12Gbit/s SAS 端口	48（双控/单控制框）
100Gb RDMA	48（双控/单控制框）

3. 硬盘规格

硬盘类型 [a]	尺寸	转速	重量	容量
SAS	3.5 寸	7.2K		• 4TB • 6TB • 8TB • 10TB • 14TB
NL-SAS	2.5 寸	10K		• 1.2TB • 1.8TB • 2.4TB
SSD	2.5 寸	-	0.25kg	• 960GB • 1.92TB • 3.84TB • 7.68TB • 15.36TB
	Palm 盘 说明 长：160mm 宽：79.8mm 厚：9.5mm		0.25kg	• 3.84TB • 7.68TB
a：由于 SSD 硬盘存储原理的限制，不能在下电状态下长期保存。在下电状态下，且环境温度在 40℃以下时，未存储数据的 SSD 盘最长放置时间不能超过 12 个月，已存储数据的 SSD 盘最长放置时间不能超过 3 个月，否则有可能导致数据丢失或者 SSD 盘故障。				

4. 尺寸和重量

配置模块	参数	RS6800
控制框	尺寸	• 长：865mm • 宽：447mm • 高：175mm
	重量（不含硬盘）	• 71.9Kg（双控） • 88.2Kg（四控）
2U SAS 硬盘框	尺寸	• 长：410mm • 宽：447mm • 高：86.1mm
	重量（不含硬盘）	13.4kg
4U SAS 硬盘框	尺寸	• 长：488mm • 宽：447mm

配置模块	参数	RS6800
		• 高：175mm
	重量（不含硬盘）	23.95kg
2U 智能 SAS 硬盘框（可容纳 25 个 2.5 寸硬盘）	尺寸	• 长：520mm • 宽：447mm • 高：86.1mm
	重量（不含硬盘）	22.45kg
2U 智能 SAS 硬盘框（可容纳 12 个 3.5 寸硬盘）	尺寸	• 长：600mm • 宽：447mm • 高：86.1mm
	重量（不含硬盘）	22.8kg
2U 智能 NVMe 硬盘框	尺寸	• 长：620mm • 宽：447mm • 高：86.1mm
	重量（不含硬盘）	24.95kg
机柜	尺寸（建议使用）	1200mm

5. 电气规格

数据块组件		RS6800
功耗	控制框	双控： • 最大功耗：2256W • 典型功耗：1649W • 最小功耗：1587W 四控： • 最大功耗：3597W • 典型功耗：2881W • 最小功耗：2657W
	2U SAS 硬盘框	• 最大功耗：323W • 典型功耗：209W • 最小功耗：138W
	4U SAS 硬盘框	• 最大功耗：582W • 典型功耗：406W • 最小功耗：354W
	2U 智能 SAS 硬盘框（可容纳 25	• 最大功耗：640W

数据块组件		RS6800
	个 2.5 寸硬盘）	• 典型功耗：436W • 最小功耗：402W
	2U 智能 SAS 硬盘框（可容纳 12 个 3.5 寸硬盘）	• 最大功耗：630W • 典型功耗：414W • 最小功耗：373W
	2U 智能 NVMe 硬盘框	• 最大功耗：879W • 典型功耗：582W • 最小功耗：512W
电源电压、额定电流	控制框	• 交流：200V～240V、AC±10%、10A、单相、50/60Hz，并支持 110V 双火线输入(2W+PE) • 240V 高压直流：240V、DC±20%、10A
	硬盘框	智能 SAS 硬盘框和智能 NVMe 硬盘框： • 交流：200V～240V、AC±10%、10A、单相、50/60Hz，并支持 110V 双火线输入(2W+PE) • 240V 高压直流：240V、DC±20%、10A SAS 硬盘框： • 交流：100V～240V、AC±10%、10A、单相、50/60Hz，并支持 110V 双火线输入(2W+PE) • 240V 高压直流：240V、DC±20%、10A
	交流电源输入类型（插座类型）	• 交流：IEC60320-C14 • 240V 高压直流：IEC60320-C14
每个 BBU 容量		30Wh

6. 可靠性规格

可靠性项目	数值
方案级可靠性	99.9999%
MTBF[a]	1,000,000 小时
MTTR[b]	2 小时
a：Mean Time Between Failures，平均无故障时间	

可靠性项目	数值
b：Mean Time To Repair，平均修复时间	

7.2 软件规格

存储系统的软件规格包括基本规格、特性规格。基本规格，介绍存储系统的基本软件规格，包括最大应用服务器连接数量、最大 LUN 数量和最大映射视图数量等；特性规格，介绍存储系统的高级特性的软件规格，包括快照、远程复制和 LUN 拷贝等。

7.2.1. RS6800 软件规格

1. 基本规格

规格名称	RS6800
最大应用服务器连接数量	• 配置 FC 端口：8192 • 配置 iSCSI 端口：2048
单个主机组中最大主机数量	64
最大 LUN 数量	65536 说明 克隆文件系统、文件系统、LUN 个数、LUN 可写快照总数、PE LUN、VVol LUN 最大总和数量不超过 65536 个。
最大 LUN 组数量	8192
LUN 映射给单个主机的最大数量	511
单个 LUN 最小容量	512KB
单个 LUN 最大容量	256TB
FC 启动器（WWN 类启动器）	8192
iSCSI 启动器	2048
最大映射视图数量	8191
最大 iSCSI 连接数（每控制器）	4096
最大 iSCSI 连接数（每端口）	512
最大 FC 连接数（每控制器）	8192
最大 FC 连接数（每端口）	2048
最大 PE LUN 数量	64

规格名称	RS6800
最大 VVol LUN 数量	65536 **说明** 克隆文件系统、文件系统、LUN 个数、LUN 可写快照总数、PE LUN、VVol LUN 最大总和数量不超过 65536 个。
最大硬盘域数量	128
单个硬盘域中最大成员盘数量	3200
单个硬盘域中最小成员盘数量（单个引擎）	4
最大存储池数量	128
单个存储池中最大 LUN 数量	65536
RAID 级别	0、1、3、5、6、10、50、RAID-TF（容忍 3 盘同时失效）
最大文件系统数量	• 克隆文件系统数量和文件系统数量总和不超过 4096。 • 克隆文件系统数量、文件系统数量、LUN 数量和 LUN 可写快照数量四者总和不超过 65536。
单文件系统最小容量	1GB
单文件系统最大容量	16PB
最大文件数（每文件系统）	20 亿
单文件最大容量	256TB
最大子目录数和文件数（每目录）	3000 万
最大 SMB 共享数	12000
最大 NFS 共享数	10000
CIFS 和 NFS 最大连接数（每控制器）	31000
FTP 最大连接数（每控制器）	192
HTTP 最大连接数（每控制器）	256
NDMP 最大流数（每控制器）	32
最大本地用户数	4000
最大本地用户组数	70000
最大组成员数	250000
最长文件全路径	4096byte

规格名称	RS6800
最长单个文件/目录名称	1024byte
最大文件系统目录深度	256
最大同时打开文件数（每控制器）	400000
最大逻辑端口数（每控制器）	256
最大 VLAN 数（每控制器）	256

2. 特性规格

特性名称	特性参数	RS6800
快照	最大源 LUN 数量	16384
	每个源 LUN 最大支持的快照数量	1024
	最大 LUN 快照数量	32768
	一致性激活包含的最大 LUN 数量	8192
	最大文件系统只读快照个数	128000
	每个源文件系统最大支持的只读快照数量	2048
	文件系统定时快照最小间隔时间	1 分钟
	文件系统快照恢复时间	<10 秒
LUN 拷贝	最大 LUN 拷贝数量	256
	最大目标 LUN 数量（每个源 LUN）	128
LUN 克隆	最大主 LUN 数量	1024
	最大从 LUN 数量	2048
	每个克隆组中最大从 LUN 数量	8
	一致性分裂的最大 Pair 数量	512
文件系统克隆	最大克隆文件系统数量	• 克隆文件系统数量和文件系统数量总和不超过 4096。 • 克隆文件系统数量、文件系统数量、LUN 数量和 LUN 可写快照数量四者总和不超过 65536。

特性名称	特性参数	RS6800
	最大级联克隆深度	8
远程复制	最大远程复制 Pair 数量（异步+同步）	2048 说明 • LUN 远程复制 Pair、文件系统远程复制 Pair、SAN 双活 Pair、NAS 双活 Pair、一体化备份 Pair、2 倍文件系统迁移任务数量的最大总和数量不超过 2048 个。 • 支持阵列内异步复制，一个阵列内异步复制 Pair 会占用两个远程复制 Pair 的规格，并且显示为两个远程复制 Pair。
	每个 Pair 中的最大从 LUN 数量	• 同步：1 • 异步：2
	每个 Pair 中的最大从文件系统数量	异步：1
	最大支持可连接的远端存储设备数量	64
	远程复制一致性组数量最大值	512（同步+异步）
	远程复制一致性组最大 Pair 数量	512
	最大远程复制租户 Pair 数量	255
	每控制器最大物理链路数	256
TFCQQoS	最大 TFCQQoS 策略数量	4096
	每个策略支持的最大 LUN 数量	64
	优先级数	3
TFCQPartition	Cache 分区数量（每双控）	8
	最小 Cache 分区大小	256MB
	最大 Cache 分区大小	20GB
TFCQTier	最大分级层数	• SAN：3（SSD/SAS/NL-SAS） • NAS：2（SSD/HDD，HDD 包含 SAS 和 NL-SAS）

特性名称	特性参数	RS6800
	迁移粒度（可设置）	• SAN：512KB/1MB/2MB/4MB/8MB/16MB/32MB/64MB（默认 4MB） • NAS：文件的大小
	迁移速率配置策略	• SAN：高/中/低 • NAS：自动（不支持配置）
TFCQMotion	颗粒度	64MB
TFCQThin	最大 thin LUN 数量（与普通 LUN 的最大个数相同）	65536
	thin LUN 最大容量	256TB
	thin LUN 与 thick LUN 相互转换	支持（通过 LUN 迁移实现）
	thin 粒度	• 未执行 TFCQDedupe&TFCQCompression：固定 64KB • 执行 TFCQDedupe&TFCQCompression：默认 16KB，CLI 下 4KB/8KB/16KB/32KB/64KB 可调
	空间回收	支持
TFCQMigration（块业务）	每个控制器同时进行迁移的最大 LUN 数量	8
	系统同时配置迁移的最大 LUN 数量	1024
	支持的一致性分裂的迁移数量	1024
TFCQMigration（文件业务）	每个控制器同时进行迁移的最大文件系统数量	64
	系统同时配置迁移的最大文件系统数量	256
TFCQErase	每个控制器同时进行数据销毁的最大 LUN 数量	16
多租户	最大租户数	255
	最大租户管理员数	512
	一个租户最大租户管理员数	32

特性名称	特性参数	RS6800
TFCQVirtualization	最大外部 LUN 数量	4096
	最大外部阵列数量	256
	每个外部 LUN 最大路径数	32
	最大伪装 LUN 数量	8192
	系统最大外部阵列链路数量	8192
	单控最大外部阵列链路数量	2048
卷镜像	最大卷镜像个数	512
	每个卷镜像的镜像副本数	2
TFCQQuota	每文件系统的配额目录树	4096
	用户配额	4000
	用户组配额	70000
TFCQCompression	压缩数据块粒度	• 文件系统：8KB/16KB/32KB/64KB 可调 • LUN：4KB/8KB/16KB/32KB/64KB 可调
TFCQDedupe	重删数据块粒度（可设置）	4KB/8KB/16KB/32KB/64KB 可调
TFCQCache	支持的 SSD Cache 总容量（每控制器）	• 9600GB（每控 256GB 内存） • 16TB（每控 512GB 内存） • 16TB（每控 1TB 内存）
	SSD Cache 分区数量（每两个控制器）	8 个用户分区和一个默认分区
	SSD Cache 数据块粒度	4KB/8KB/16KB/32KB/64KB/128KB 可调
NAS 防病毒	病毒扫描方式	CIFS 共享，文件关闭时扫描
	最大防病毒服务器数量	512
	最大监控文件系统数量	同系统最大文件系统数量
	最大病毒扫描策略数量	1024
	最大防病毒服务器数量（每个租户）	32

特性名称	特性参数	RS6800
HyperMetro（SAN）	最大双活域数量	2 NAS 双活域数量+SAN 双活域数量不超过 2 个。
	最大双活 LUN 一致性组数量	256
	双活最大 Pair 数量	1024
	每个一致性组最大 Pair 数量	1024
	每个双活域最大 Pair 数量	1024
	每个控制器最大物理链路数量	256
	最远距离	小于 300km
	支持的物理链路类型	8G FC/16G FC/32G FC/10GE/25GE/40GE/100GE
	支持的协议类型	iSCSI/FC
	仲裁模式	静态优先级模式 仲裁服务器模式
HyperMetro（NAS）	最大双活域数量	2 NAS 双活域数量加 SAN 双活域数量不超过 2 个。
	最大双活租户 Pair 数量	255
	最大文件系统双活 Pair 数量	2048
	每个控制器最大物理链路数量	256
	最远距离	小于 300km
	支持的物理链路类型	8G FC/16G FC/32G FC/10GE/25GE/40GE/100GE
	支持的协议类型	SMB3.0/NFSv3/NFSv4.0/NFSv4.1
	仲裁模式	静态优先级模式 仲裁服务器模式
仲裁客户端（存储侧）	每个阵列允许接入仲裁服务器数	32
	每个双活域允许接入仲裁服务器数	2
	每个仲裁服务器可添加最大 IP 地址数（服务器端 IP 地址）	2
	每个阵列的每控制器可接入到同一个仲裁服务器的最大链路数	2

特性名称	特性参数	RS6800
仲裁服务器（服务器侧）	单台仲裁服务器支持双活阵列接入数量	8
	单台仲裁服务器支持添加的仲裁 IP 数量	4
HyperVault	最大备份 pair 数量	1024
	最大备份副本数量	8192
	备份速率	快、高、中、低
	备份周期	5 分钟～1 个月
	最大备份策略数（每个 Pair）	本地备份策略：4 异地备份策略：4
	最大备份副本数量（每个 Pair）	本地备份策略：256 异地备份策略：256

7.2.2. License 控制

功能名称	是否需要 License 控制
快照（HyperSnap）	是 说明：快照块业务和文件业务使用的是相同的 License。
克隆（HyperClone）	是 说明：克隆块业务和文件业务使用的是相同的 License。
LUN 拷贝（HyperCopy）	是
远程复制（HyperReplication）	是 说明：远程复制块业务和文件业务使用的是相同的 License。
TFCQ QoS（智能服务质量控制）	是
TFCQ Motion（智能数据迅移）	是
TFCQ Thin（智能精简配置）	是
TFCQ Partition（智能缓存分区）	是
TFCQ Migration（智能数据迁移）	是 说明：TFCQ Migration 块业务和文件业务使用的是相同的 License。
TFCQ Erase（智能数据销毁）	是
TFCQ Multi-Tenant（多租户）	是
TFCQ Virtualization（异构虚拟化）	是
卷镜像（HyperMirror）	是

功能名称	是否需要 License 控制
TFCQ Compression（智能数据压缩-LUN 与文件系统共用）	是
TFCQ Dedupe（智能数据重删-LUN 与文件系统共用）	是
TFCQ Quota（智能配额）	是
CIFS	是
NFS	是
WORM（HyperLock）	是
NDMP	是
SAN 双活（HyperMetro）	是
NAS 双活（HyperMetro）	是
HyperVault（一体化备份）	是

8 环境要求

8.1 温度、湿度和海拔

存储系统工作或存放时对温度、湿度和海拔有一定的要求，如表 1-18 所示。

表1-18 存储系统对温度和湿度的要求

参数	条件	要求
温度	工作温度	• 海拔为-60m～1800m 时，5℃～40℃。 • 海拔为1800m～3000m时，海拔每升高220m，环境温度降低 1℃。
	工作环境温度变化率	• 工作时：20℃/H • 存储和运输时：30℃/H
	非工作环境温度	-40℃～+70℃
	存放温度（带外包装运输、存放状态）	-40℃～+70℃
湿度	工作湿度	10% RH[a]～90% RH
	存放湿度	5% RH～95% RH
	非工作环境湿度	5% RH～95% RH
	最大湿度变化率	10%/H
海拔	硬盘工作海拔	-304.8m～+3048m
	硬盘非工作海拔	-305m～+12192m
a：RH，Relative Humidity，相对湿度。 说明 硬盘运行受环境影响较大，硬盘在超过环境规格要求下使用，会有较高的故障率，请在规定的环境规格下使用。		

8.2 振动和冲击

存储系统工作或存放时对振动和冲击有一定的要求。

存储系统对振动和冲击的要求如表 1-19 所示。

表1-19 存储系统对振动和冲击的要求

参数	要求
工作振动	5～350Hz，PSD：0.0002 g^2/Hz，350～500Hz，-3dB，0.3Grms，3 axes，15min/axis

参数	要求
非工作振动	10～500Hz，1.49Grms，3 axes，15min/axis PSD： • 10HZ@0.1g²/HZ • 20HZ@0.1g²/HZ • 50HZ@0.004g²/HZ • 100HZ@0.001g²/HZ • 500HZ@0.001g²/HZ
非工作冲击	half sine，70Gs/2ms，1 shock/face，total 6 faces

8.3 颗粒污染物

颗粒污染物和其它环境因素（如温度或相对湿度）发生的交互作用可能会对 IT 设备造成腐蚀故障风险。本条款规定了针对颗粒污染物的限制要求，旨在避免此类风险的发生。

数据中心颗粒污染物应满足 IT 设备制造商普遍采用的由美国采暖、制冷与空调工程师学会技术委员会 ASHRAE（American Society of Heating Refrigerating and Airconditioning Engineers）TC（Technical Committee）9.9 编写的《针对数据中心气体与颗粒污染物指南（2011 版）》白皮书要求。

ASHRAE 是 ISO（International Organization for Standardization）国际标准化组织指定的唯一负责采暖、制冷、空调方面的国际标准认证组织。由 ASHRAE TC 9.9 标准组织成员 AMD、Cisco、Cray、Dell、EMC、Hitachi、HP、IBM、Intel、Seagate、SGI 和 Sun 共同起草的《针对数据中心气体与颗粒污染物指南》已得到业界广泛认可和使用。

依据该白皮书要求，数据中心颗粒污染物应满足 ISO 14644-1 Class8 级别定义的洁净度要求：

- 每立方米中颗粒尺寸≥0.5μm 的颗粒不能超过 3,520,000 个。
- 每立方米中颗粒尺寸≥1μm 的颗粒不能超过 832,000 个。
- 每立方米中颗粒尺寸≥5μm 的颗粒不能超过 29,300 个。

建议使用高效过滤器过滤进入数据中心的空气，建议使用过滤系统定期过滤数据中心内的空气。

8.4 散热和噪音

存储设备通过风扇模块自带的风扇散热，能够长期运行。热空气排出存储设备后，需要外部设备将热空气带走，保证空气循环。

8.4.1. 散热

存储系统的散热方式如下：

控制框采用整体前进风，后出风的散热方式。

冷却空气从控制框接口模块的缝隙进入，在为接口模块、控制器和电源模块散热后，被风扇排出。控制框会根据存储系统的工作温度自动调节风扇转速，保证系统的散热。

硬盘框采用整体前进风，后出风的散热方式。

冷却空气从硬盘框的硬盘之间的缝隙进入，流过背板后，进入电源模块和级联模块区域，在为电源模块和级联模块散热后，被风扇排出。硬盘框会根据存储系统的工作温度自动调节风扇转速，保证系统的散热。

为便于维护、通风及散热，将存储系统安装到机柜中时请注意以下事项：

建议机柜与墙壁之间的距离不小于 100cm，机柜与机柜之间的前后距离不小于 120cm，以确保机柜前后通风顺畅。

机柜内部不得形成封闭空间，确保机柜内部和机房内空气有效对流。建议设备的上方和下方均预留 1U（1U＝44.45mm）的空间。

存储系统的进风量如表 1-20 所示。

表1-20 存储系统的进风量参数

系统进风量	RS5300	RS5500	RS5600	RS5800
控制框	• 150CFM[a]（风扇最大转速） • 55CFM（25℃）	2U 控制框，可容纳 25 个 2.5 寸硬盘： • 240CFM（风扇最大转速） • 96CFM（25℃） 2U 控制框，可容纳 36 个 Palm 硬盘： • 260CFM（风扇最大转速） • 104CFM（25℃）		
2U SAS 硬盘框	• 117CFM（风扇最大转速） • 38CFM（25℃）			

系统进风量	RS5300	RS5500	RS5600	RS5800
2U 智能 SAS 硬盘框	• 150CFM（风扇最大转速） • 55CFM（25℃）			
2U 智能 NVMe 硬盘框	不支持	• 135CFM（风扇最大转速） • 45CFM（25℃）		
a：CFM，Cubic Feet per Minute，立方英尺每分钟。				

存储系统的散热量如表 1-21 所示。

表1-21 存储系统的散热量参数

最大散热量（工作时）	RS5300	RS5500	RS5600	RS5800
控制框	3236.09BTU[a]/h	7328.09BTU/h		
2U SAS 硬盘框	1091.2BTU/h			
2U 智能 SAS 硬盘框	3641.9BTU/h			
2U 智能 NVMe 硬盘框	3454.3BTU/h			
a：BTU，British Thermal Unit，英制单位。				

8.4.2. 噪音

存储系统中的硬盘、风扇工作时会发出噪音，而风扇发出的噪音是主要噪音。风扇工作强度与温度有关，温度越高，风扇工作强度越大。风扇工作强度加大后会发出更强的噪音，因此正常情况下存储系统的噪音与环境温度相关。

环境温度为 25℃，存储系统发出的噪音的声功率级如表 1-22 所示。

表1-22 存储系统的噪音参数

声功率	RS5300	RS5500	RS5600	RS5800
控制框	69dB	2U 控制框，可容纳 25 个 2.5 寸硬盘：68.2dB 2U 控制框，可容纳 36 个 Palm 硬盘：70.5dB		
2U SAS 硬盘框	63.7dB			
2U 智能 SAS 硬盘框	69.9dB			
2U 智能 NVMe 硬盘框	不支持	69.6dB		

9 遵循标准

本存储产品遵循的协议标准、安规和 EMC 标准、行业标准说明。

9.1 协议标准

存储系统遵循的协议标准如表 1-23 所示。

表1-23 遵循的协议标准

名称	标准号
SCSI 体系	FC-PH：ANSI X3.230
	FC-PH2：ANSI X3.297
	SCSI-FCP：ANSI X.269
	FC-AL：ANSI X.272
	FC-AL-2：ANSI NCITS 332-1999
	FC-SW：ANSI NCITS 321
	FC-SW-2：ANSI NCITS 355-2001
	FC-GS：ANSI X.288（针对光纤交换机）
	FC-GS2：ANSI NCITS 288（针对光纤交换机）
	SAS Serial Attached SCSI-1.1（SAS-1.1）
	SAS Serial Attached SCSI-2.0（SAS-2.0）
	SAS Serial Attached SCSI-3.0（SAS-3.0）
	T10/1562D Rev.05 Serial Attached SCSI（SAS）
	T10/1601D Rev.07 Serial Attached SCSI Model-1.1（SAS-1.1）
	T10/1601D Rev.07 Serial Attached SCSI Model-1.1（SAS-2.0）
	T10/1601D Rev.07 Serial Attached SCSI Model-1.1（SAS-3.0）
	SFF 8301 form factor of 3.5' disk drive
	SFF 8323 3.5' disk drive form factor with serial connector
	SFF 8482 SAS plug connector
	SCSI 3 SAM-2：ANSI INCITS 366-2003
	SPC-2：ANSI INCITS 351-2001
	SBC：ANSI INCITS 306-1998

名称	标准号
	PICMG3.0 Advanced Telecommunications Computing Architecture
	PICMG3.1 Ethernet/fiber Channel Over PICMG3.0
	iSCSI RFC 3720/7143
TCP/IP 体系	SNMP v1
	SNMP v2c
	SNMP v3
PCIe 体系	PCI Express Base Specification R1.1
	PCI Express to PCI or PCI-X Bridge Specification v1.0
	PCI Express Base Specification v2.0

9.2 接口标准

存储系统遵循的接口标准如表 1-24 所示。

表1-24 遵循的接口标准

名称	描述
VAAI	一种来自 VMware 的 API（Application Programming Interface）框架。该框架将 VMware 服务器上的一些存储相关任务（比如精简配置）转移到存储阵列上进行。
VASA	一组用于 VMware vSphere ESXi 主机与存储设备通信的 API，实现存储阵列与 vCenter 的集成管理功能。
SRA	一个位于 VMware 的 SRM（VMware Site Recovery Manager）与存储系统之间的接口。它能够使 SRM 能够执行发现存储系统、不影响业务的故障切换测试、紧急或者演练的故障切换、反向拷贝数据以及恢复备份操作。
SMI-S	是 SNIA（Storage Networking Industry Association）开发的一种标准管理接口，旨在减轻多厂商 SAN（存储区域网络）环境的管理负担。SMI-S 为各种网络组件提供了一个通用管理接口，减小了 SAN 管理的复杂性。
ODX	ODX（Offloaded Data Transfer）是 Windows Server 2012 的一项特性，该特性将文件卸载到存储阵列进行传输，利用阵列之间的高传输带宽最大限度地缩短数据的传输延迟和提高数据拷贝的速度，同时降低了主机服务器的资源使用率。

9.3 安规和 EMC 标准

存储系统遵循的安规和 EMC 标准如表 1-25 所示。

表1-25 遵循的安规和 EMC 标准

名称	标准号
中国安规标准	GB 4943.1-2011
北美安规标准	UL 60950-1
欧洲安规指令	2014/35/EU
欧洲安规标准	EN 60950-1
中国 EMC 标准	GB/T9254-2008
	GB17625.1-2012
加拿大 EMC 标准	ICES-003
	CAN/CSA-CEI/IEC CISPR 22:02
北美 EMC 标准	FCC, CFR 47 Part 15, Subpart B
欧洲 EMC 指令	2014/30/EU
欧洲 EMC 标准	EN 55032
	EN 55024

行业标准

存储系统遵循的行业标准如表 1-26 所示。

表1-26 遵循的行业标准

名称	标准号
以太网	IEEE 802.3
快速以太网	IEEE 802.3u
千兆以太网	IEEE 802.3z
	IEEE 802.3ab
万兆以太网	IEEE 802.3ae
VLAN	IEEE 802.1q
IEEE 标准测试接口和边界扫描结构	IEEE 1149.1-2001
故障模式影响分析（FMEA）过程	IEC 812
可靠性、维修性和可用性预计标准	IEC 863
ETSI 标准（环境）	ETS 300 019

名称	标准号
ETSI 标准（电源）	ETS 300 132
ETSI 标准（噪音）	ETS 300 753
ETSI 标准（环境）	ETS 300 119
ETSI 标准（接地）	ETS 300 253
ITUT 标准（接地）	ITUT K.27
环保	ECMA TR/70
可靠性	GR-929、Telcordia SR-332
洁净室及相关受控环境	ISO 14664-1 Class8
气体污染物及环境标准	ANSI/ISA-71.04-1985 气体腐蚀等级 G1

10 术语定义

A		
ANSI	American National Standards Institute	美国国家标准协会
B		
BBU	Backup Battery Unit	备份电池单元
C		
CLI	Command Line Interface	命令行界面
E		
ESN	Equipment Serial Number	设备序列号
F		
FC	Fibre Channel	光纤通道
FC-AL	Fibre Channel Arbitrated Loop	光纤通道仲裁环
FCoE	Fibre Channel over Ethernet	以太网光纤通道
G		
GE	Gigabit Ethernet	千兆以太网
GUI	Graphical User Interface	图形用户界面
H		
HBA	Host Bus Adapter	主机总线适配器
HD	High Density	高密度
I		
IP	Internet Protocol	因特网协议
ISA	Instrument Society of America	美国仪器仪表协会
iSCSI	Internet Small Computer Systems Interface	互联网小型计算机系统接口
ISO	International Organization for Standardization	国际标准化组织
L		
LUN	Logical Unit Number	逻辑单元号
M		
MTBF	Mean Time Between Failures	平均无故障时间
MTTR	Mean Time to Repair	平均修复时间

N		
NL-SAS	Near Line Serial Attached SCSI	近线串行连接的 SCSI
P		
PDU	Power Distribution Unit	电源分配单元
U		
USB	Universal Serial Bus	通用串行总线
R		
RAID	Redundant Array of Independent Disks	独立磁盘冗余阵列
RSCN	Registered State Change Notification	注册状态变化通告
S		
SAN	Storage Area Network	存储区域网络
SAS	Serial Attached SCSI	串行连接的 SCSI
SCSI	Small Computer System Interface	小型计算机系统接口
SSD	Solid State Drive	固态硬盘
V		
VLAN	Virtual LAN	虚拟局域网
VPN	Virtual Private Network	虚拟专用网