

存储阵列技术与应用

www.huawei.com





目 录

1. 存储阵列系统组成

1.1 存储阵列系统

1.2 存储阵列硬件组成

1.3 存储阵列软件组成

存储阵列系统



-
- 容灾解决方案
 - 备份解决方案
 -
-

- 存储管理软件 (ISM)
 - 快照、镜像软件
 - 备份软件
 - 多路径软件
-

• 外置存储系统

- ▣ 磁盘阵列
- ▣ NAS
- ▣ 磁带库
- ▣ 虚拟磁带库

• 存储连接设备

- ▣ FC HBA卡
- ▣ FC交换机
- ▣ 以太网交换机
- ▣ 连接线缆



目 录

1. 存储阵列系统组成

1.1 存储阵列系统

1.2 存储阵列硬件组成

1.3 存储阵列软件组成

存储阵列硬件组成

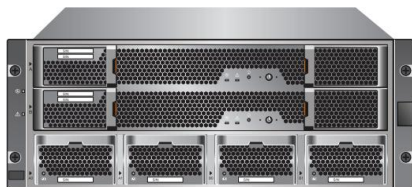
- 存储阵列硬件系统由二部分组成：

- 控制框

控制框用于处理各种存储业务，并管理级联在控制框下面的硬盘框。

- 硬盘框

硬盘框主要用于容纳各种硬盘，为应用服务器提供充足的存储空间。

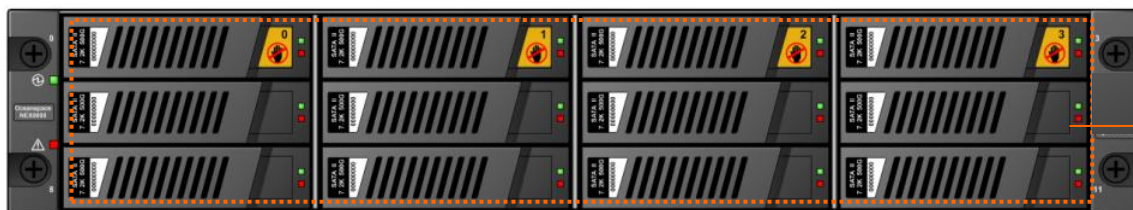


存储阵列硬件组成 — 控制框（续）



2.5inch硬盘插框

- 24 × 2.5inch硬盘。
- 支持主流SAS、NL-SAS、SSD盘。
- 前四个盘为保险箱盘，2*（1+1）冗余。
- 每个硬盘可独立上下电。

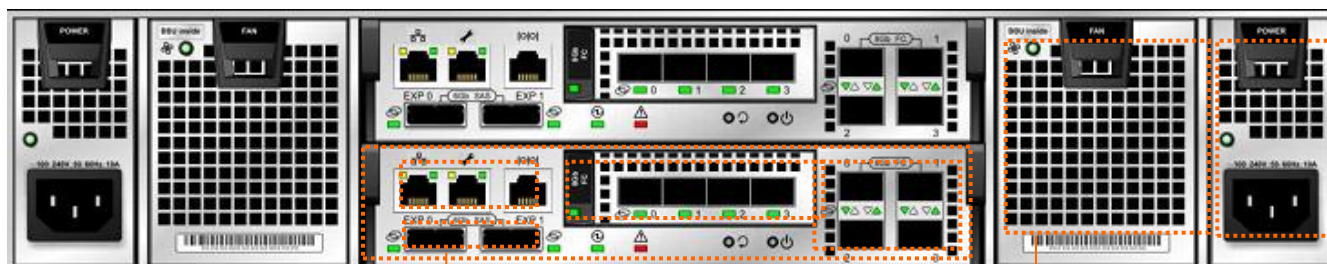


3.5inch硬盘插框

- 12 × 3.5inch硬盘。
- 支持主流SAS、NL-SAS、SSD盘。
- 前四个盘为保险箱盘，2*（1+1）冗余。
- 每个硬盘可独立上下电。

注：以S5500T为例，控制框一前面板图

存储阵列硬件组成 — 控制框（续）



电源模块

- 1+1冗余。
- 转换效率最高达92%。

盘控一体控制模块

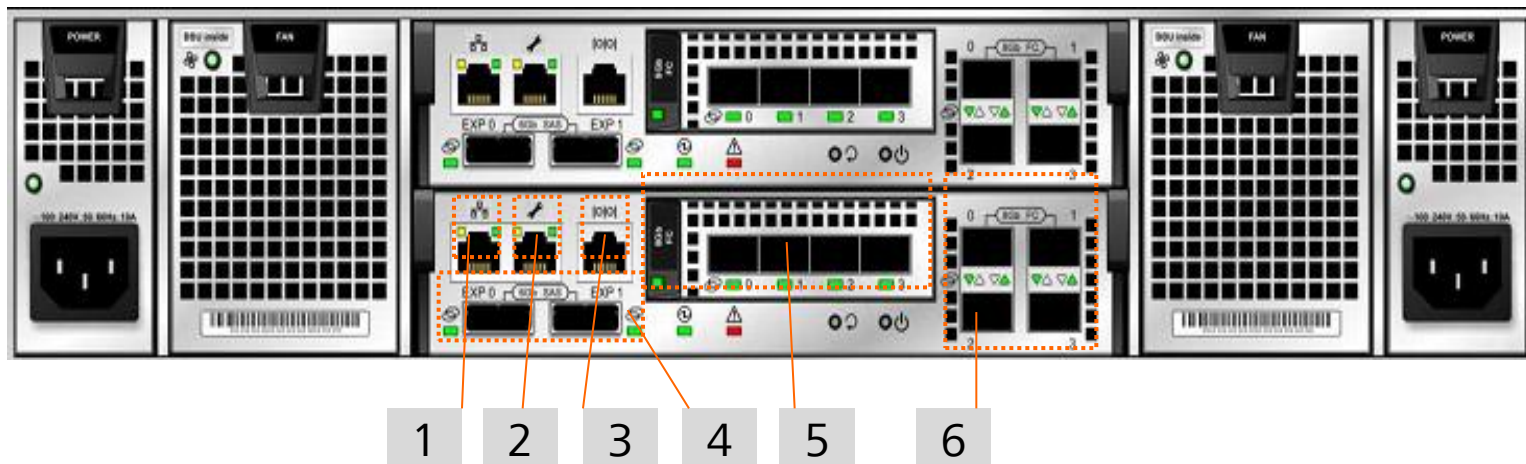
- 主流桌面CPU平台。
- 提供管理网口、调试网口和串口。
- 板载2个6Gb SAS接口。
- 支持1个接口模块。
- 板载4个8G FC接口。

风扇/BBU模块

- 风扇3+1冗余。
- 智能精细化调速。
- 内置BBU，支持掉电数据保护，1+1冗余

注：以S5500T为例，控制框一后面板图

存储阵列硬件组成 — 控制框（续）



1 管理网口

2 维护网口

3 串口

4 mini SAS级联端口

5 8Gb FC主机端口

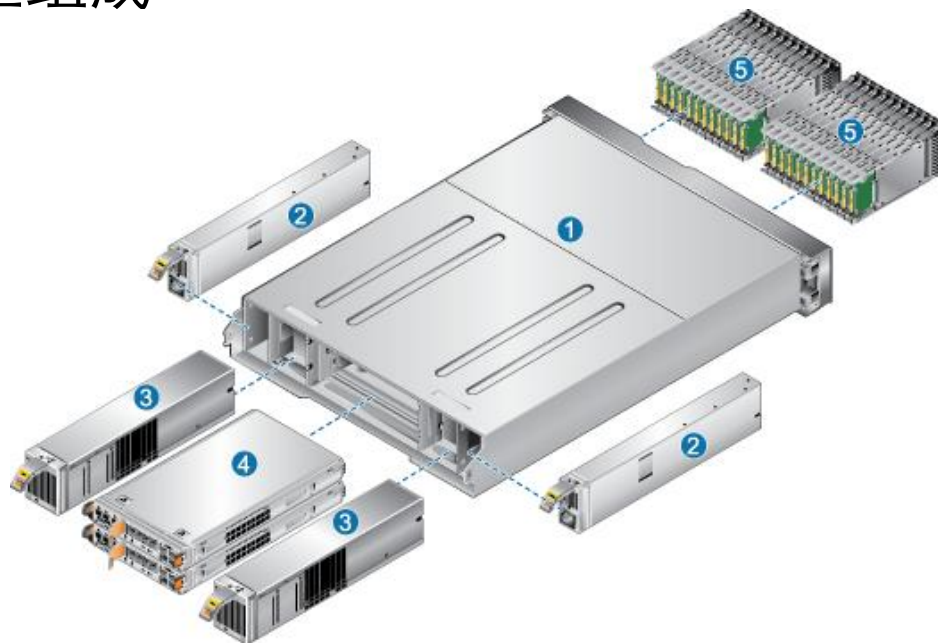
6 8Gb FC主机端口

注：以S5500T为例，控制框一后面板接口图

存储阵列硬件组成 — 硬盘框

- 存储阵列系统硬盘框组成：

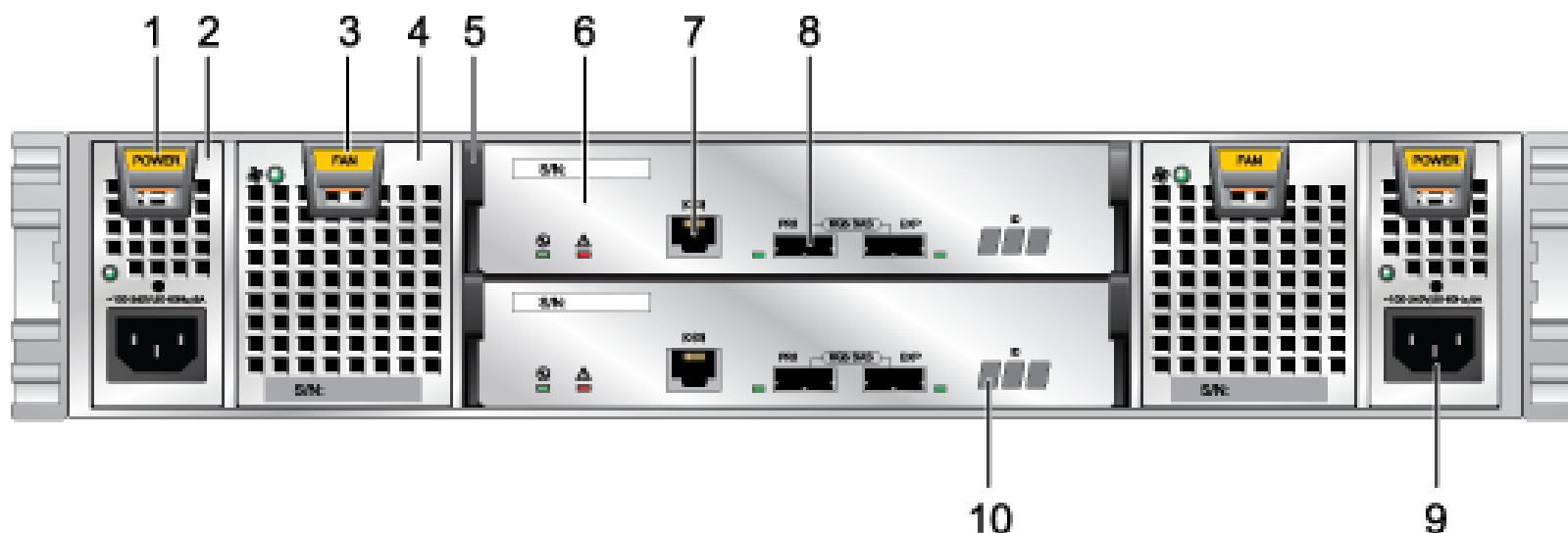
- 系统插框
- 电源模块
- 风扇模块
- 级联模块
- 硬盘模块



1	系统插框	2	电源模块
3	风扇模块	4	级联模块
5	硬盘模块		

存储阵列硬件组成 — 硬盘框（续）

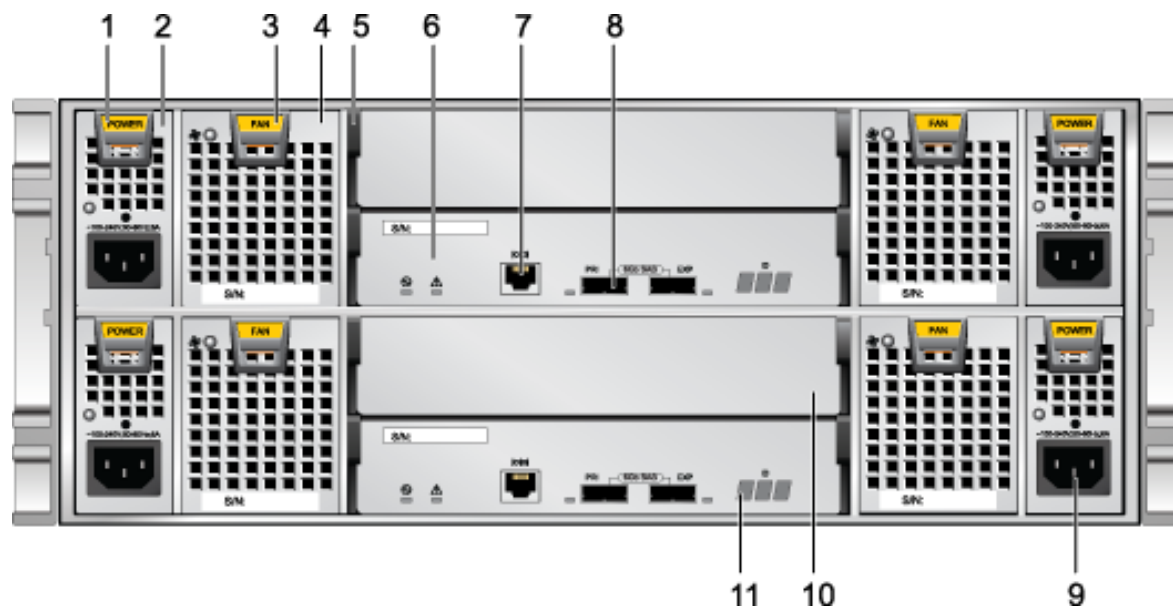
- T系列存储硬盘框——2U SAS硬盘框



1	电源模块拉手	2	电源模块	3	风扇模块拉手
4	风扇模块	5	级联模块拉手	6	级联模块
7	串口	8	mini SAS级联端口	9	电源插座
10	硬盘框ID显示器				

存储阵列硬件组成 — 硬盘框（续）

- T系列存储硬盘框——4U SAS硬盘框4U SAS硬盘框后视图



1	电源模块拉手	2	电源模块	3	风扇模块拉手
4	风扇模块	5	级联模块拉手	6	级联模块
7	串口	8	mini SAS级联端口	9	电源插座
10	假面板	11	硬盘框ID显示器		



目 录

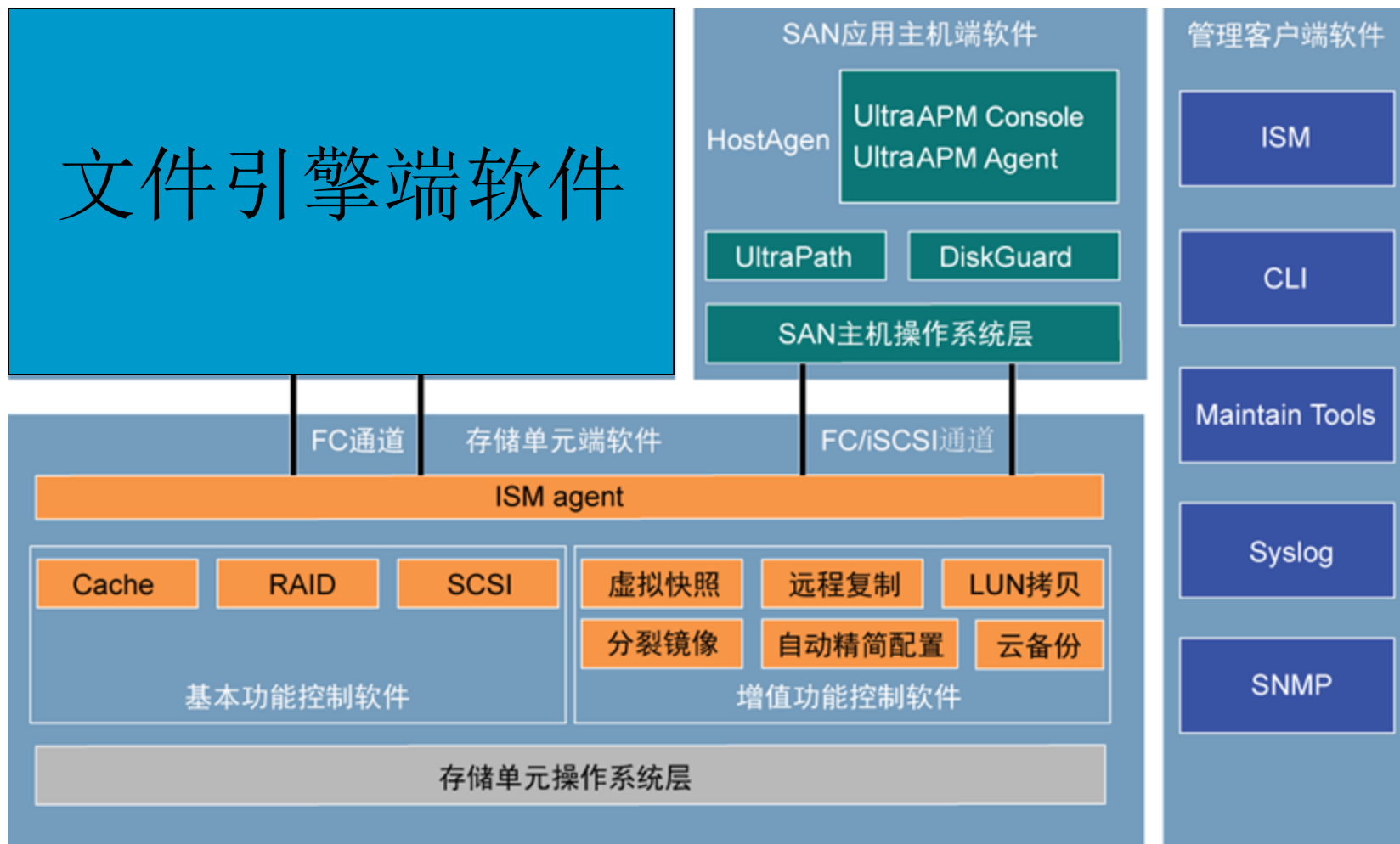
1. 存储阵列系统组成

1.1 存储阵列系统

1.2 存储阵列硬件组成

1.3 存储阵列软件组成

存储阵列软件组成





2. 华为存储阵列技术

2.1 华为存储阵列技术总览

2.2 华为存储阵列技术

华为存储阵列技术总览

高可靠性

- 器件冗余（多控）
- 硬盘坏道检测/修复
- 磁盘健康分析
- 多路径技术
- BBU掉电保护
- RAID重构
- 缓存镜像技术
- 磁盘保险箱技术
- 磁盘预拷贝技术
- 快照/克隆技术
- LUN拷贝
- 远程复制
-

可扩展性

- Scale-out
- iSCSI
- FC技术
- SAS技术
- PCI-e
- FcOE

高性能

- 块虚拟化技术
- cache回写
- cache预取技术
- 12GbSAS
- 16Gb FC

高可用性

- SmartTier技术
- SmartQos技术
- SmartThin技术

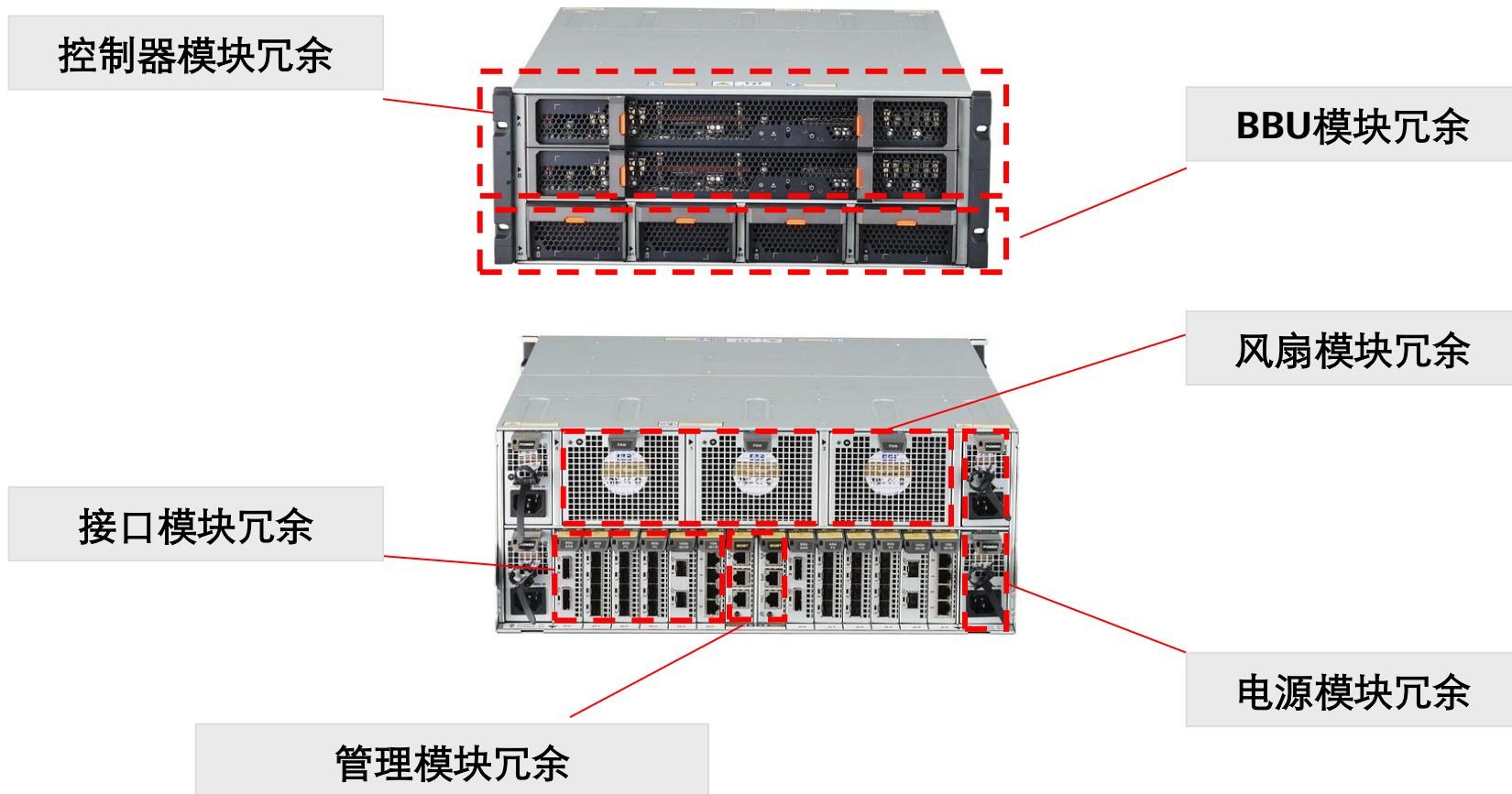


2. 华为存储阵列技术

2.1 华为存储阵列技术总览

2.2 华为存储阵列技术

华为存储阵列技术 — 器件全冗余

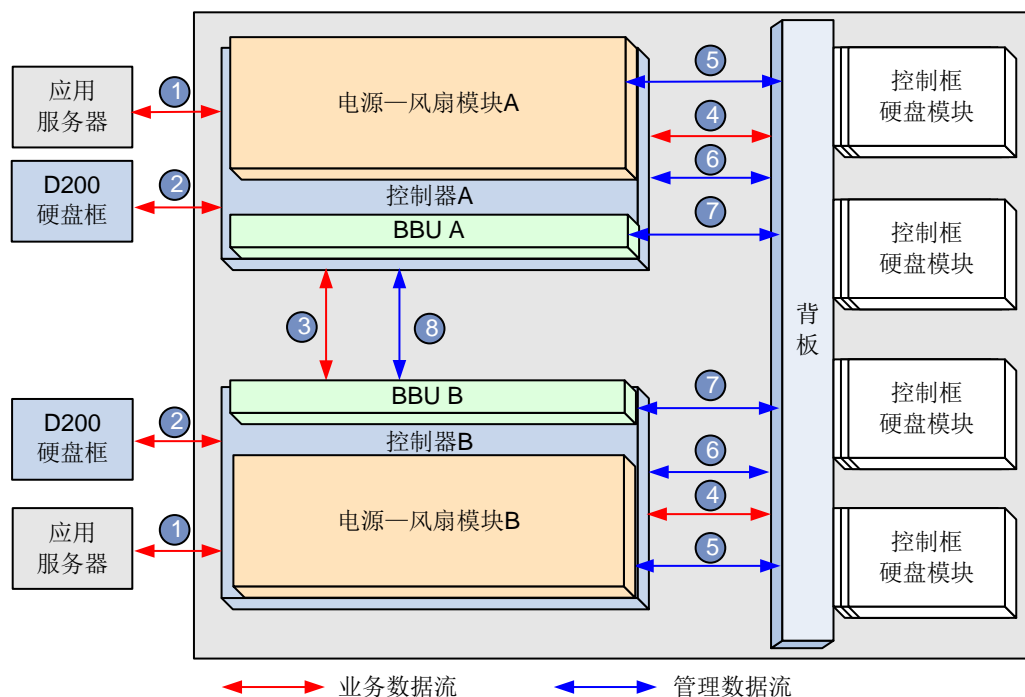


华为存储阵列技术 — 多控技术

- 双控制器系统的工作模式:

- 主备模式 (AP)

- 双活模式 (AA) ———> 华为存储都采用AA模式



华为存储阵列技术 — 硬盘坏道检测技术



✓ 面对磁盘坏道，被动应对还是主动出击？

读写失败自动分析

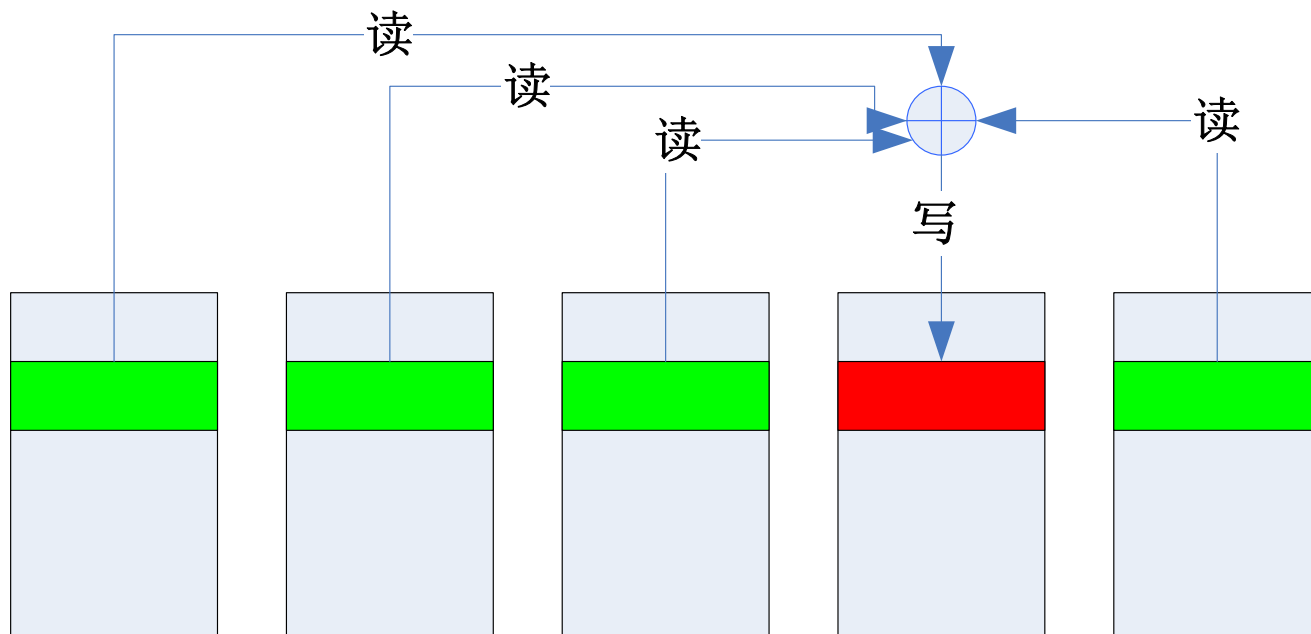
- 读写失败有多重原因
- 根据系统当前状态、硬盘当前状态、IO失败信息等进行分析

硬盘介质自动扫描

- 直接使用硬盘的内建介质扫描功能
- 避免了硬盘扫描对后端带宽的占用
- 将对系统性能的影响降到最低

华为存储阵列技术 — 磁盘坏道修复技术

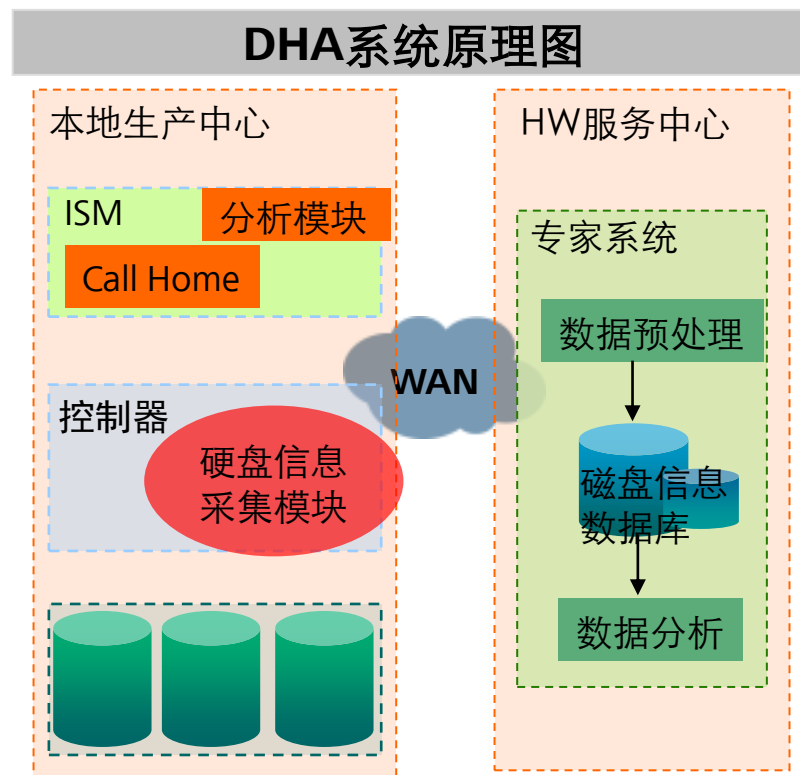
- RAID 5 坏道修复示意图（红色色表示坏道）



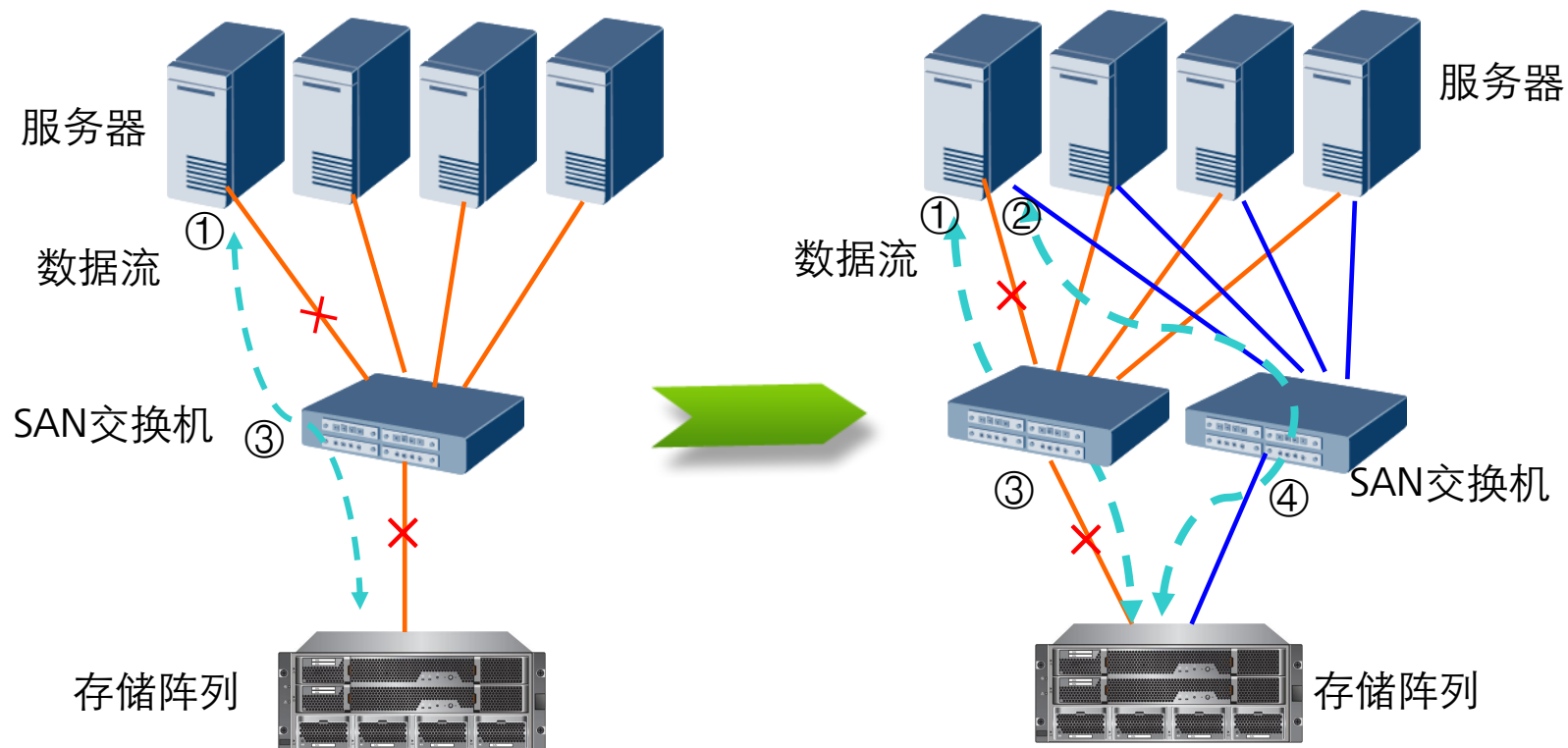
华为存储阵列技术 — 磁盘健康分析DHA

- **DHA (Disk Health Analyzer) 系统包括:**

- 硬盘信息采集模块
- 分析模块
- Call Home模块
- 数据预处理
- 硬盘信息数据库
- 数据分析

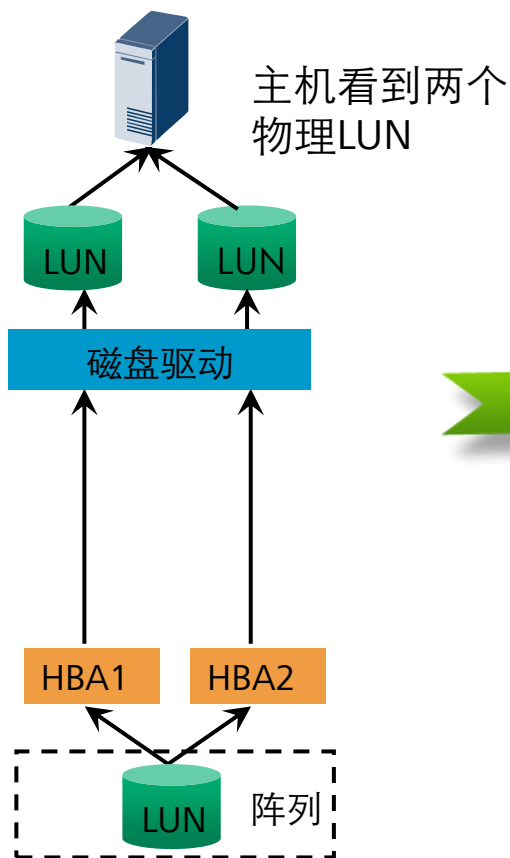


华为存储阵列技术 — 多路径技术

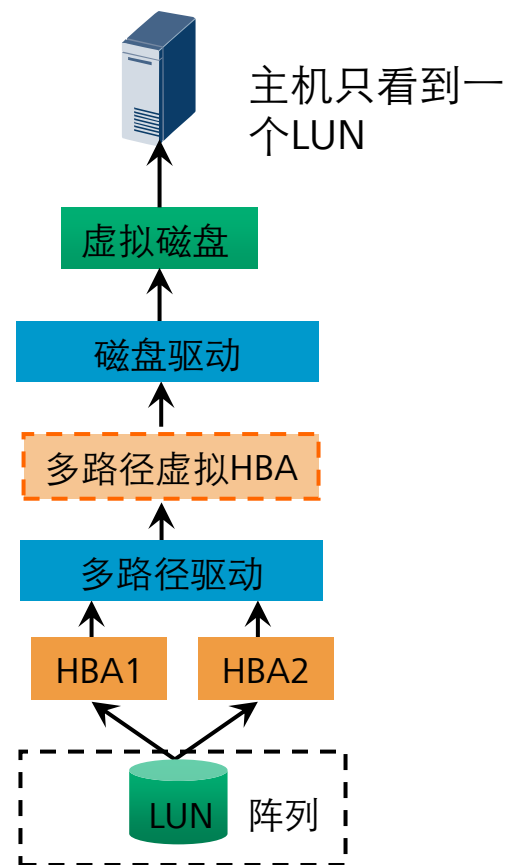


华为存储阵列技术 — 多路径技术（续）

未安装多路径

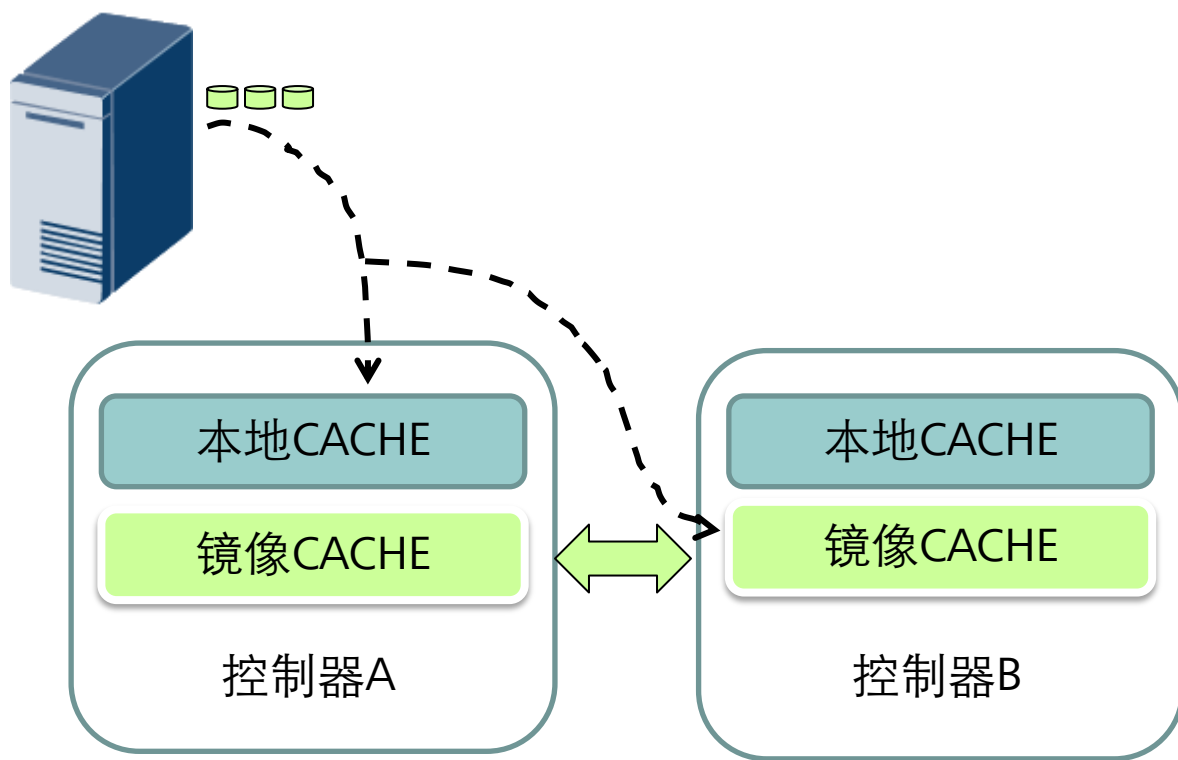


安装了多路径



华为存储阵列技术 — 镜像技术

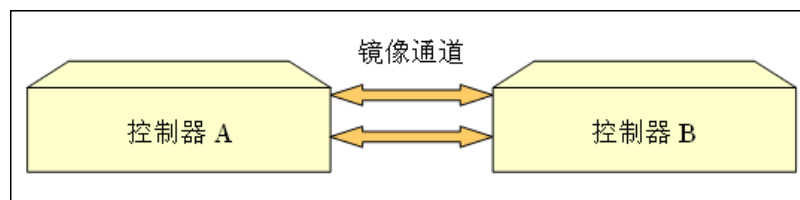
- 两个控制器的写Cache数据通过相互镜像实现备份，确保数据的安全和完整，提高了系统的可靠性。



华为存储阵列技术 — 镜像技术（续）

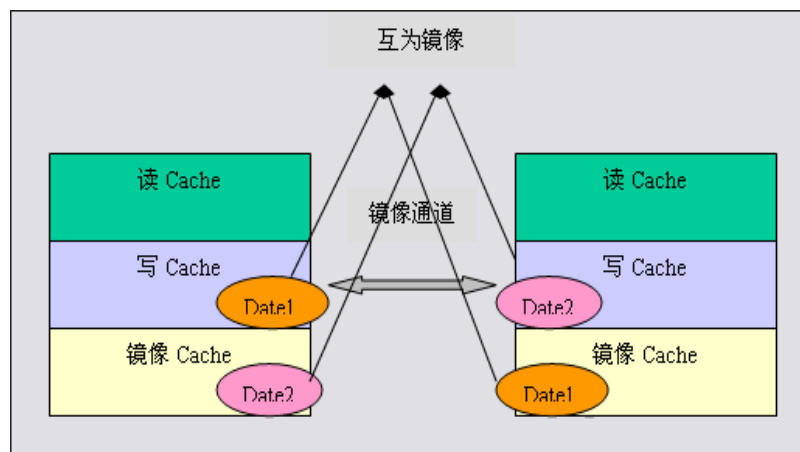
- 镜像通道

- SAS
- PCI-E
- FC



- 实现机制

- 读Cache
- 写Cache
- 镜像Cache



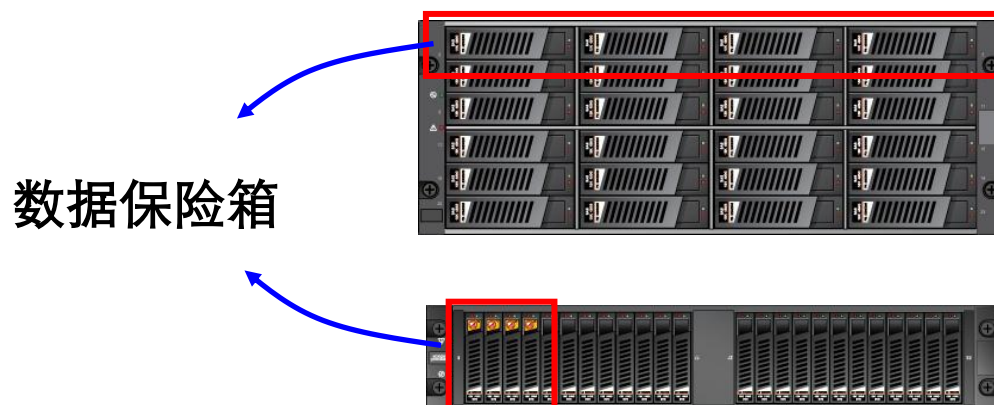
华为存储阵列技术 — 镜像技术（续）

读策略	<ul style="list-style-type: none">• 智能预取：对主机读请求进行连续性判断。如果是连续请求，则将当前读请求后面的一段数据从硬盘预取到Cache中，提高读Cache命中率。如果是随机读，则不预取，只从硬盘读取需要的部分。• 固定预取：Cache读取硬盘数据时，每次从硬盘中读取固定的长度（用户配置）。• 可变预取：按照主机I/O请求中读取长度的倍数将数据预取到Cache中。
写策略	<ul style="list-style-type: none">• 透写：应用下发写数据请求时，既将数据写入Cache，同时也将数据写入硬盘。• 回写/镜像：应用下发写数据请求时，将数据写入本地Cache，同时也将数据写入对端Cache。• 强制回写/镜像：当存储系统发生故障（例如高温故障或BBU供电不足）时，强制将数据写入本地Cache，同时也将数据写到对端Cache。• 强制回写/不镜像：当存储系统发生故障（例如高温故障或BBU供电不足）时，强制将数据写入本地Cache。

华为存储阵列技术 — 数据保险箱技术

- 数据保险箱技术:

用于保存Cache数据，避免因系统意外断电时数据丢失。内置BBU电池可保证在系统意外断电时，对Cache和系统保险箱硬盘同时供电，让Cache中的数据写到数据保险箱中，实现Cache数据永久保存。



华为存储阵列技术 — 硬盘预拷贝技术



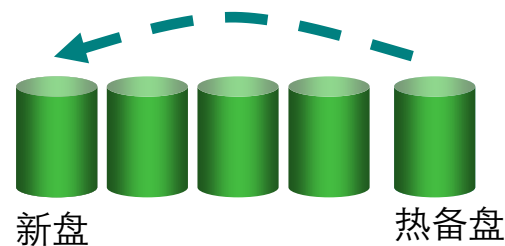
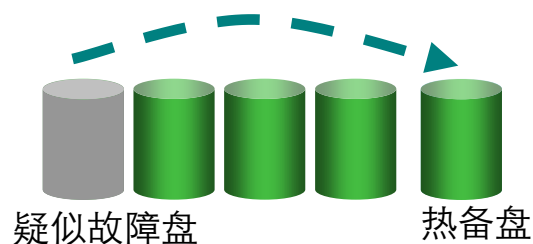
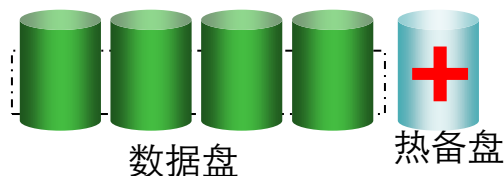
正常状态时，实时监控硬盘状态



当某个硬盘出现疑似故障时：将该盘上的迁移数据到热备盘



迁移完成后，用新盘替换掉故障盘，数据会Copy back到新更换的硬盘上

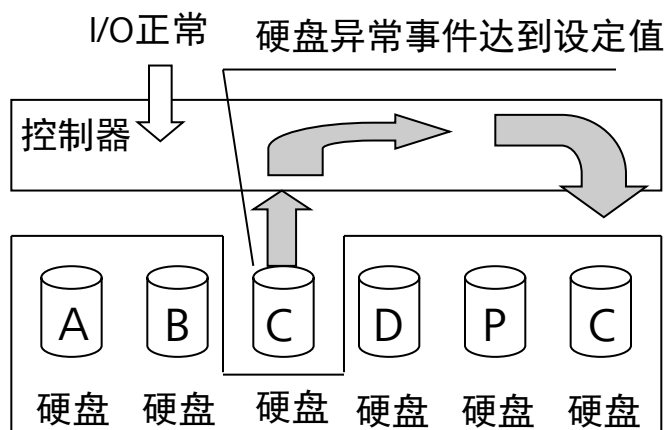


华为存储阵列技术 — 预拷贝与重构

- 数据重构的影响

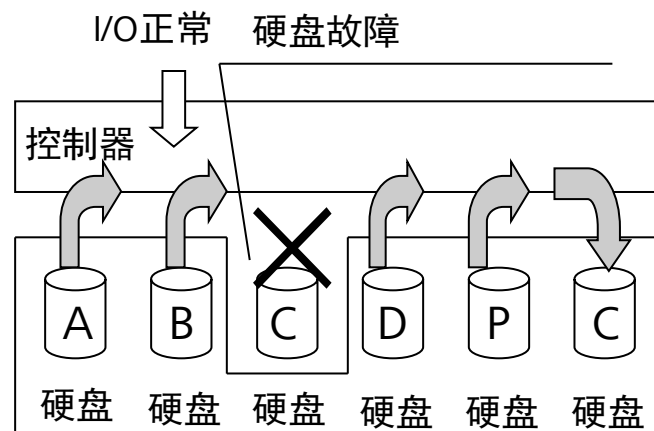
- 系统性能的大幅降低
- 大量数据读写易导致硬盘损坏
- 会导致业务中断

磁盘预拷贝技术示意图



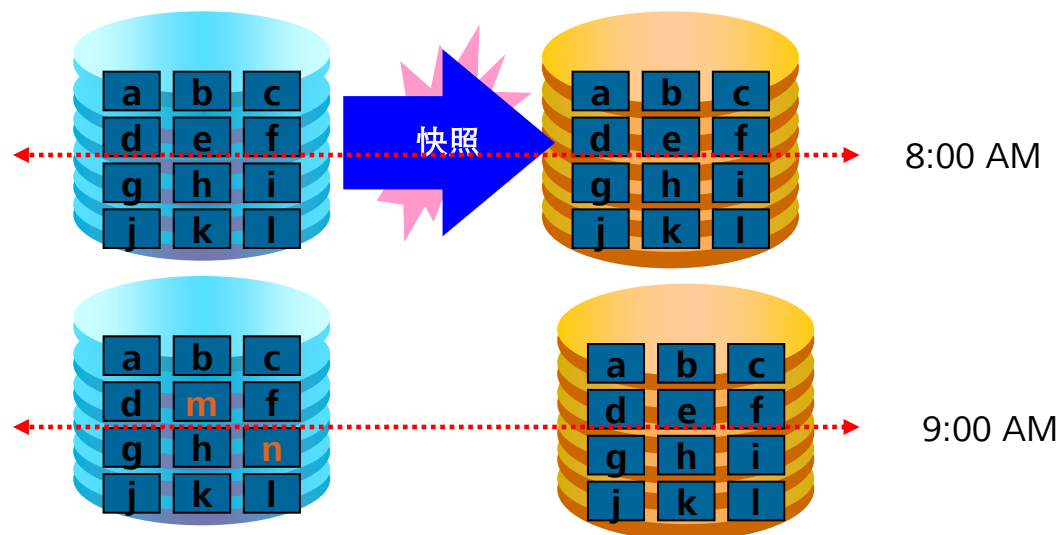
VS

数据重构技术示意图



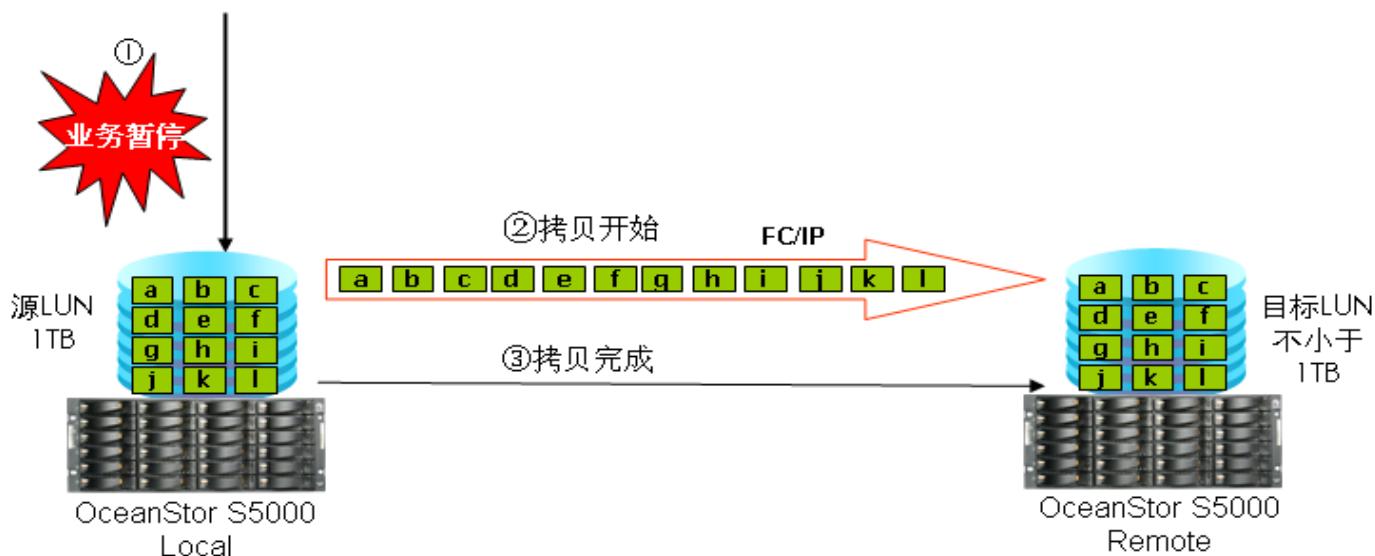
华为存储阵列技术 — 快照技术

- 快照为一个数据对象产生完全可用的副本，它包含该数据对象在某一时间点的映像。
- 数据对象：对存储阵列来说就是可映射给主机的LUN。
 - 完全可用：可以正常读写。
 - 时间点：数据具有一致性。



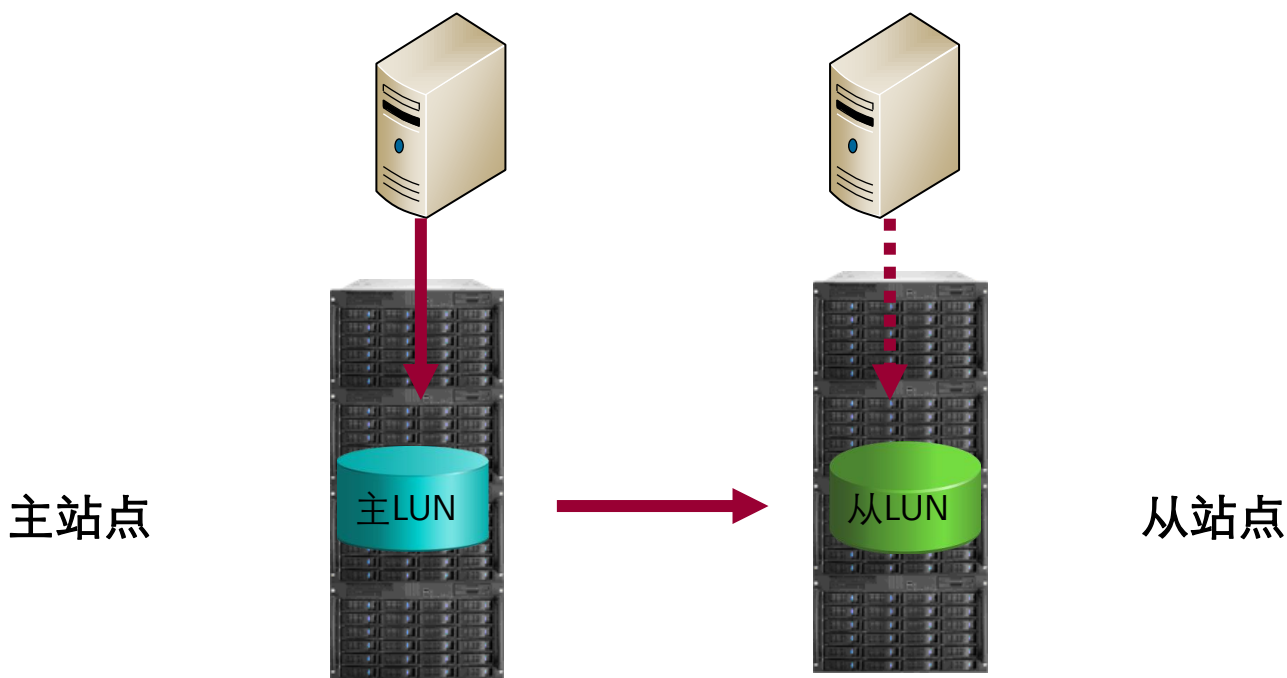
华为存储阵列技术 — LUN拷贝技术

- **定义：**一种基于块的将源LUN的数据复制到目标LUN的技术。
- **应用：**通过LUN拷贝，实现分级存储、系统升级、异地备份等应用需求。



华为存储阵列技术 — 远程复制技术

- **远程复制（HyperMirror）**：提供不同区域间数据的同步/异步镜像。保护用户数据，避免灾难性事件带来的损失。

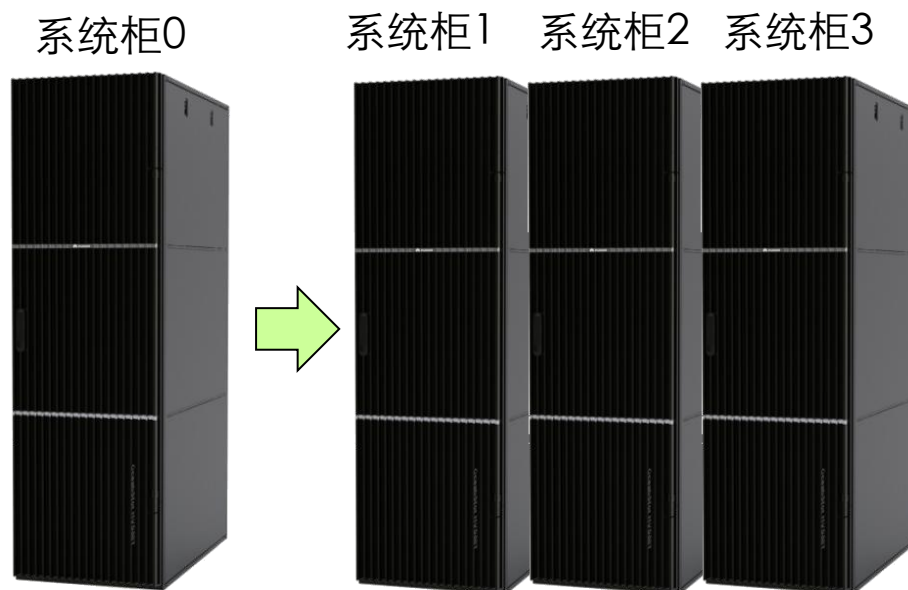


华为存储阵列技术 — 对比

对比项目	远程复制	LUN拷贝
兼容性	只能在同类型存储系统之间运行。	不但支持同类型存储系统，而且支持经过认证的第三方的存储系统。
数据下发	每个主LUN只能向1个（异步模式）或2个（同步模式，分别位于不同存储系统上）从LUN复制数据。	每个源LUN可以向数十个或者更多目标LUN复制数据。
数据备份	用于持续的数据保护。从LUN可读，但始终不可写。	用于数据备份。数据拷贝完成后，主机即可访问目标LUN。

华为存储阵列技术 — 可扩展性技术

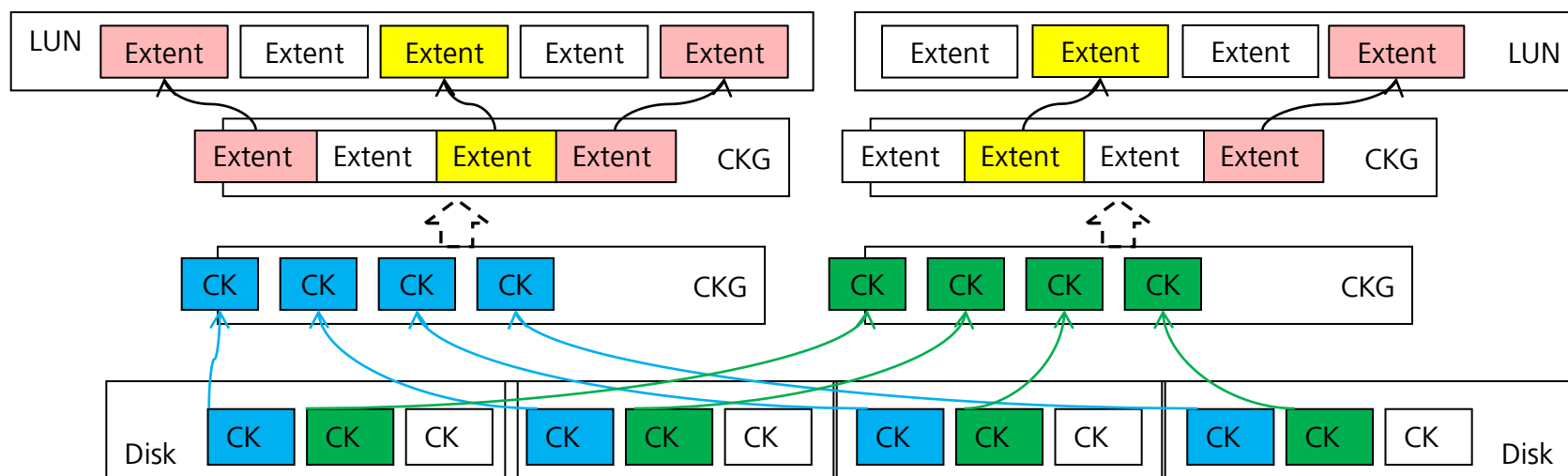
- Scale Out
- iSCSI技术
- FC技术
- SAS技术
- PCI-e



扩展接口协议	优点	缺点
SCSI	成熟稳定	只适用于直连，扩展能力差
iSCSI	组网方便，管理简单，不受距离限制	数据传输效率低，安全性差
FC	吞吐量大，可靠性高，低时滞，安全性高，数据传输效率高	需存储专网，成本高
SAS	性价比高，发展空间大，技术新	连接距离短，只适用于直连

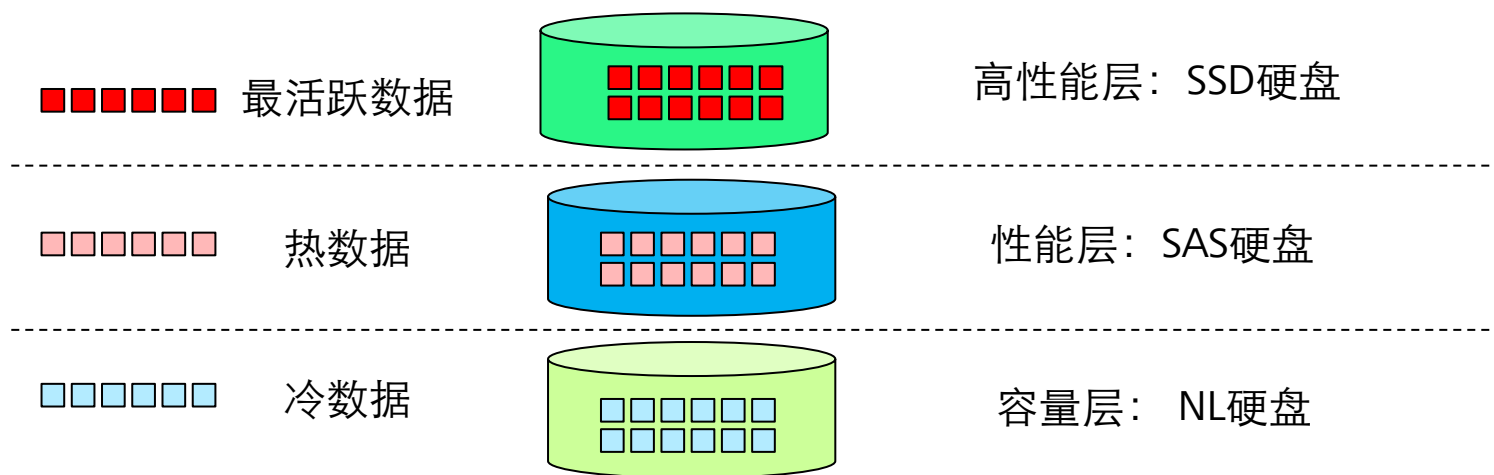
华为存储阵列技术 — 块级虚拟化技术

- **块级虚拟化技术**：一种新型RAID技术，该技术将硬盘划分成若干固定大小的块（chunk），然后将其组合成若干个小RAID组。



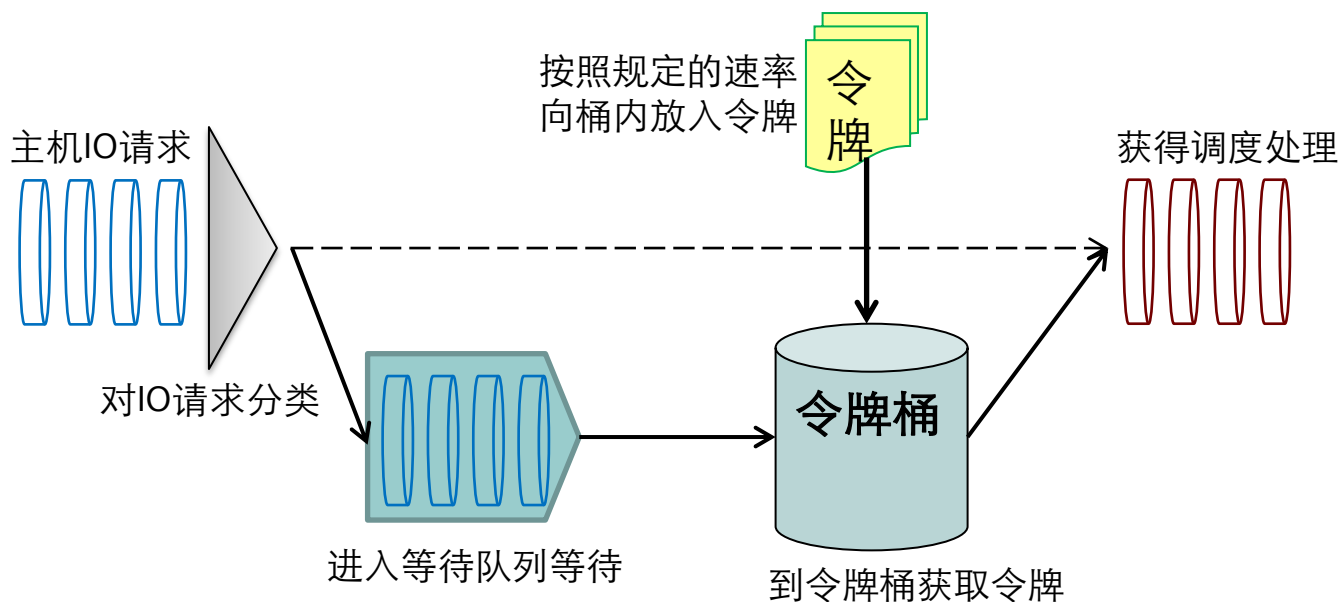
华为存储阵列技术 — SmartTier技术

- 动态分级存储技术（SmartTier）：自动将不同活跃度的数据和不同特点的存储介质动态匹配，提高存储系统性能并降低用户成本。



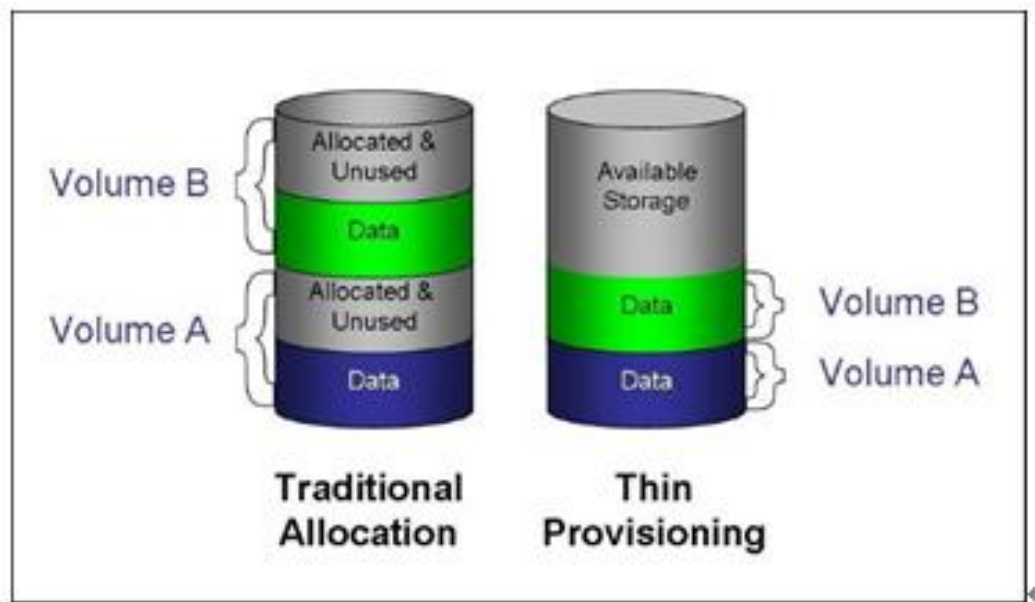
华为存储阵列技术 — SmartQoS技术

- **SmartQoS**：是一种性能特性，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。



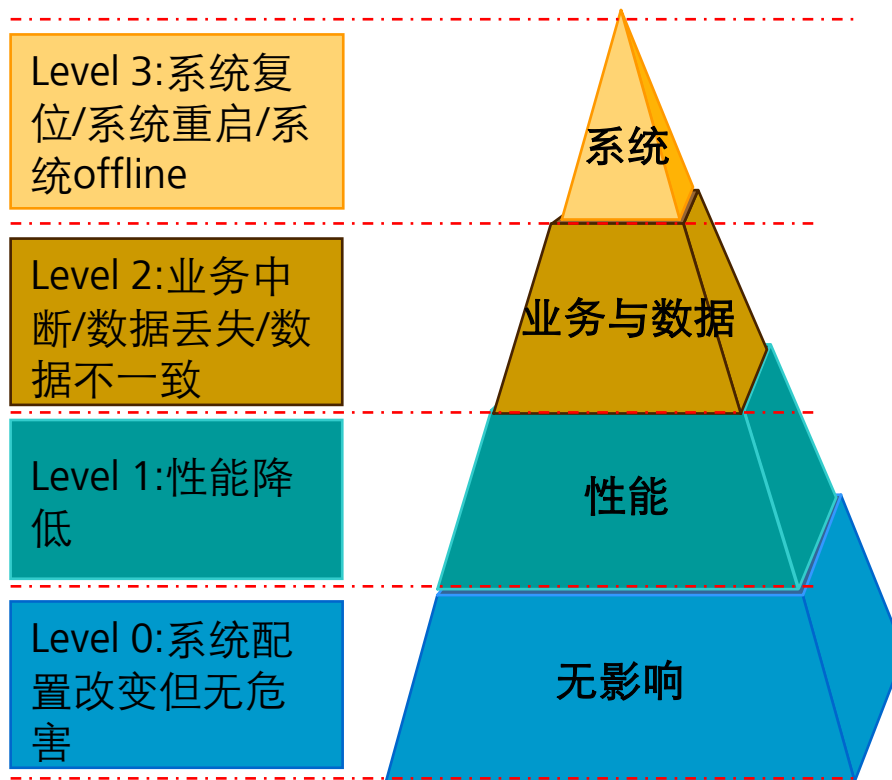
华为存储阵列技术 — Smart Thin技术

- **SmartThin**: 能够实现按需分配存储空间。在存储空间配额范围内，应用服务器用到多少空间，存储系统才给它分配多少空间，从而节省了宝贵的存储资源。



华为存储阵列技术 — 操作分级管理

- Level 0: 无影响，不进行处理
- Level 1: 提示
- Level 2: 警告
- Level 3: 危险



谢谢

www.huawei.com