存储技术引论

信息存储
数据中心环境
数据保护：RAID
智能存储系统

模块 – 1

信息存储简介

模块 1：信息存储简介

学完本模块后，您将能够：

- 定义数据和信息
- 描述数据类型
- 描述存储体系结构的发展历史
- 描述数据中心的核心元素
- 列出数据中心的关键特征
- 概述虚拟化和云计算

信息存储和管理有何意义？

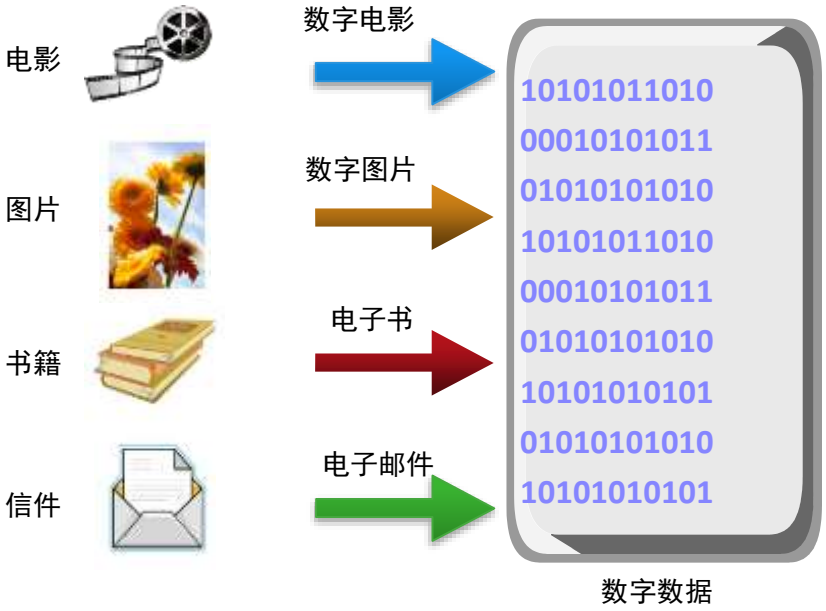
- 信息是从数据中派生出来的知识
- 数字信息的增长导致了信息爆炸
- 我们生活在一个随需而变的世界中
  - 我们随时随地都需要信息
- 对快速、可靠地获取信息的依赖程度日益提高
- 各企业纷纷寻找途径来存储、保护、优化和利用这些信息
  - 获得竞争优势
  - 获得新的业务机会

什么是数据？

数据

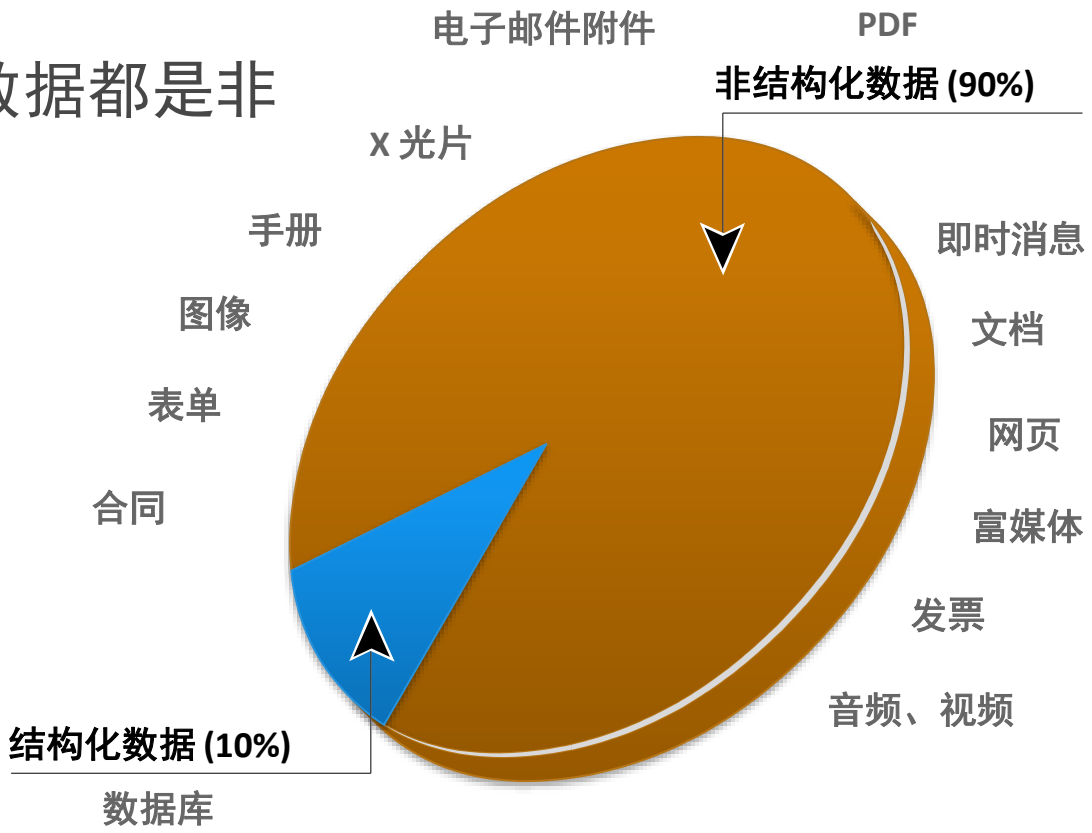
数据是可从中得出结论的未经处理的事实的集合。

- 数据被转换为更便捷的形式 – 数字数据
- 数字数据增长的因素有：
  - 数据处理能力的提高
  - 数字存储成本的降低
  - 价格合理、速度更快的通信技术
  - 应用程序和智能设备的剧增



数据类型

- 数据可分为：
  - 结构化数据
  - 非结构化数据
- 所创建的大部分数据都是非结构化数据



大数据

大数据

它是指其大小超出常用软件工具在可接受时间限制内的捕获、存储、管理和处理能力的数据集。

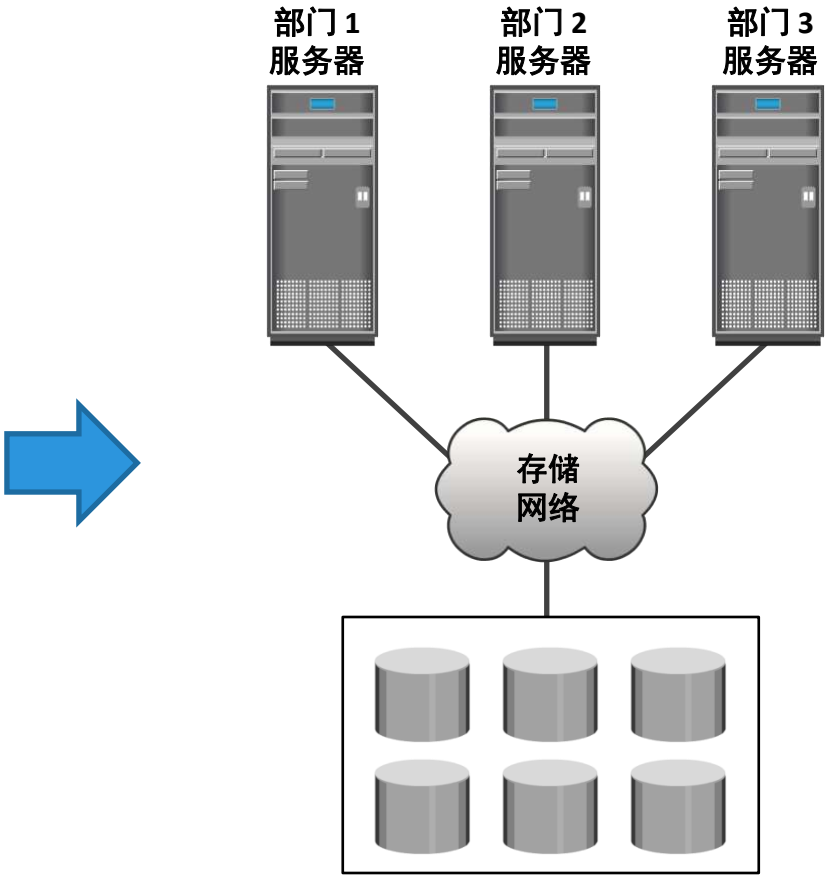
- 包括各种源生成的结构化和非结构化数据
- 大数据实时分析需要提供以下功能的新技术和工具：
  - 高性能
  - 大规模并行处理 (MPP) 数据平台
  - 高级分析
- 通过大数据分析，可以将大量数据转换为正确的决策

存储

- 存储个人和组织创建的数据
  - 提供对数据的访问以备进一步处理
- 存储设备的示例有：
  - 手机或数码相机中的媒体卡
  - DVD、CD-ROM
  - 磁盘驱动器
  - 磁盘阵列
  - 磁带

存储体系结构的发展历史

以服务器为中心的存储体系结构



以信息为中心的存储体系结构

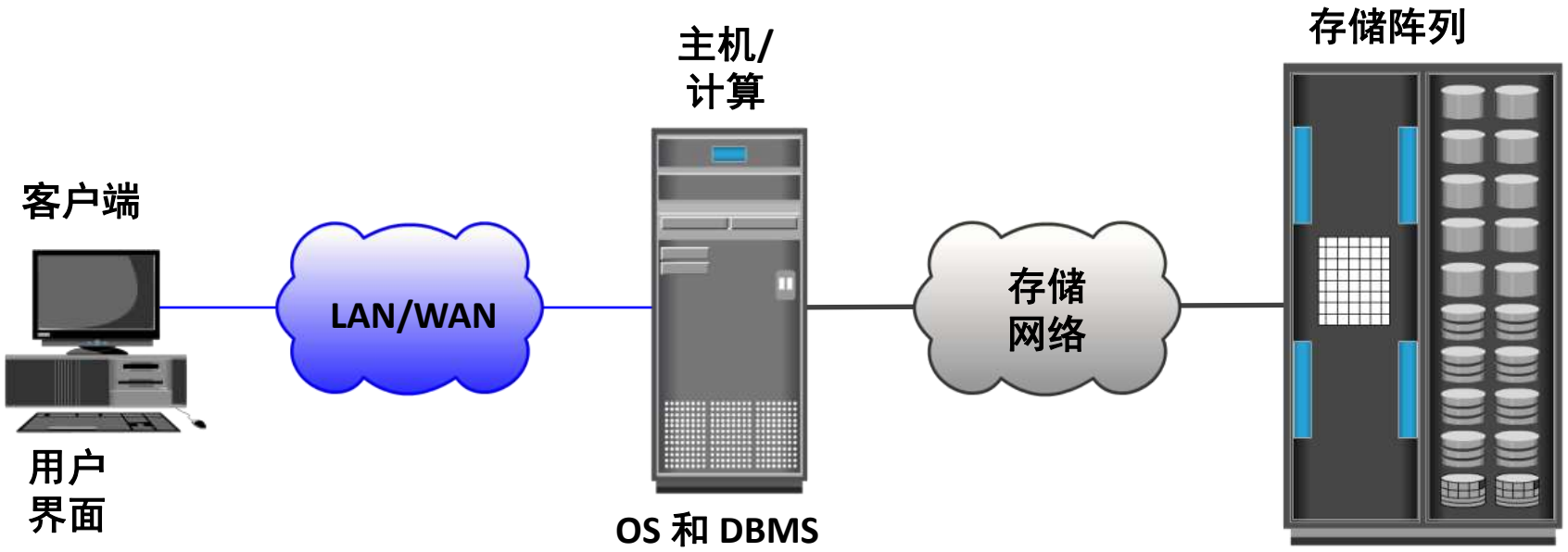
数据中心

数据中心

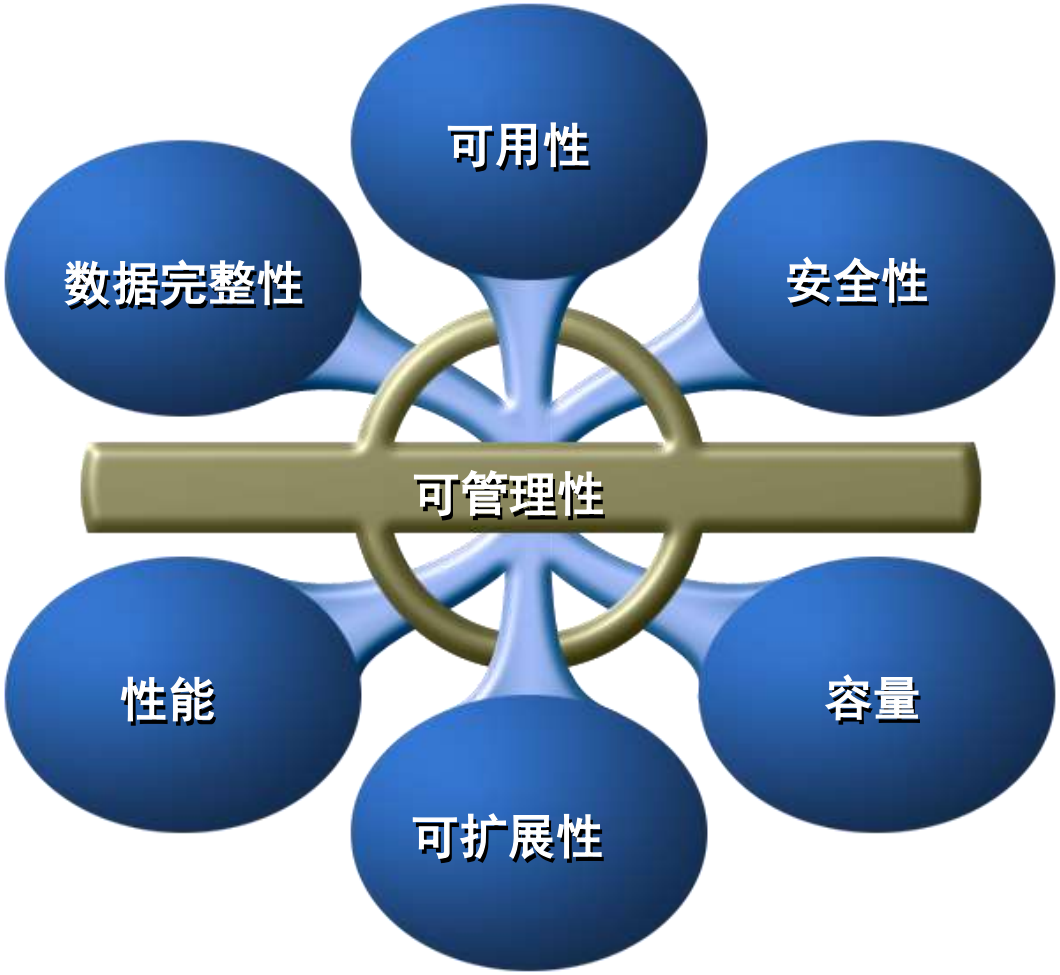
它是包含存储、计算、网络和其他 IT 资源以提供集中式数据处理功能的设备。

- 数据中心的核心元素
  - 应用程序
  - 数据库管理系统 (DBMS)
  - 主机或计算
  - 网络
  - 存储
- 这些核心元素协同工作以满足数据处理要求

数据中心：在线订单交易系统示例



数据中心的关键特征



管理数据中心

- 重要的管理活动包括
  - 监视
    - 收集有关数据中心中运行的各种元素和服务的信息的持续过程
  - 报告
    - 有关资源性能、容量和利用率的详细信息
  - 资源调配
    - 配置和分配资源以可满足容量、可用性、性能和安全要求
- 虚拟化和云计算改变了数据中心基础架构资源的调配和管理方式

虚拟化：概述

- 虚拟化是指抽象化物理资源并让其显示为逻辑资源的技术
  - 例如原始磁盘的分区
- 共用物理资源并提供物理资源功能的聚合视图
- 可根据共用物理资源创建虚拟资源
  - 提高物理 IT 资源的利用率

云计算：概述

- 支持个人和组织通过网络将 IT 资源作为服务使用
- 支持自助请求且自动化请求完成过程
  - 支持用户快速纵向扩展计算资源的使用
- 支持基于消耗量的计量
  - 用户只为他们使用的资源付费
    - 示例：使用的 CPU 小时数、数据传输量和数据存储量 (GB)

模块 1：总结

本模块涵盖以下要点：

- 数据和信息
- 数据类型
- 大数据
- 存储体系结构的发展历史
- 数据中心的核心元素
- 数据中心的关键特征
- 虚拟化和云计算

模块 – 2

数据中心环境

模块 2：数据中心环境

学完本模块后，您将能够：

- 描述数据中心的核心元素
- 描述应用程序和主机层的虚拟化
- 描述磁盘驱动器的组件和性能
- 描述主机通过 DAS 访问存储
- 描述闪存驱动器的工作和优势

模块 2：数据中心环境

第 1 课：应用程序、DBMS 和主机（计算）

本课程将讲述下列主题：

- 应用程序和应用程序虚拟化
- DBMS
- 主机系统的组件
- 计算和内存虚拟化

应用程序

- 为计算操作提供逻辑的软件程序
- 数据中心中通常部署的应用程序
  - 业务应用程序 – 电子邮件、企业资源规划 (ERP)、决策支持系统 (DSS)
  - 管理应用程序 – 资源管理、性能调整、虚拟化
  - 数据保护应用程序 – 备份、复制
  - 安全应用程序 – 身份验证、反病毒
- 应用程序的关键 I/O 特性
  - 读取密集型与写入密集型
  - 按序与随机
  - I/O 大小

应用程序虚拟化

应用程序虚拟化

它是向最终用户提供应用程序而无需任何安装、集成或底层计算平台上的依赖项的技术。

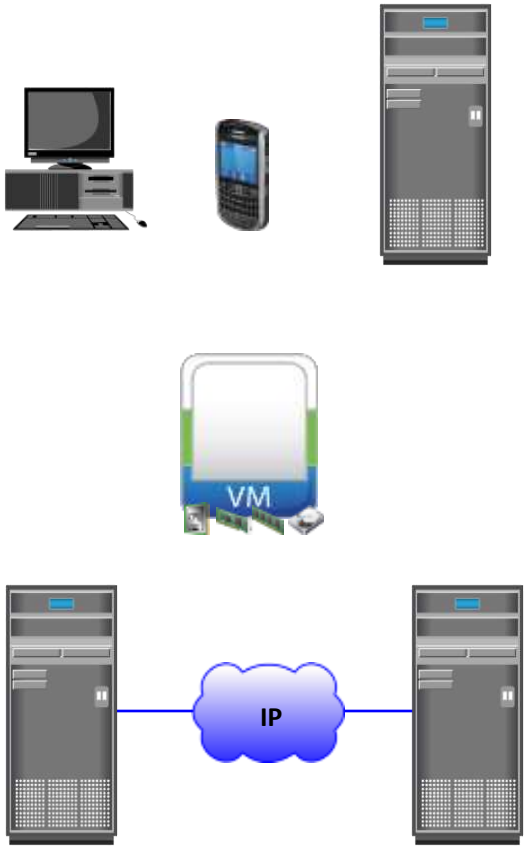
- 允许在单独环境中提供应用程序
  - 将操作系统 (OS) 资源和应用程序聚合到虚拟化容器中
  - 确保操作系统 (OS) 和应用程序的完整性
  - 避免不同应用程序或同一应用程序的不同版本之间发生冲突

数据库管理系统 (DBMS)

- 数据库是一种结构化存储方式，可将数据存储在相互关联并按逻辑组织的多个表中
  - 有助于优化数据的存储和检索
- DBMS 可控制数据库的创建、维护和使用
  - 处理应用程序的数据请求
  - 指示 OS 从存储中检索相应数据
- 常用的 DBMS 示例有 MySQL、Oracle RDBMS、SQL Server 等。

主机（计算）

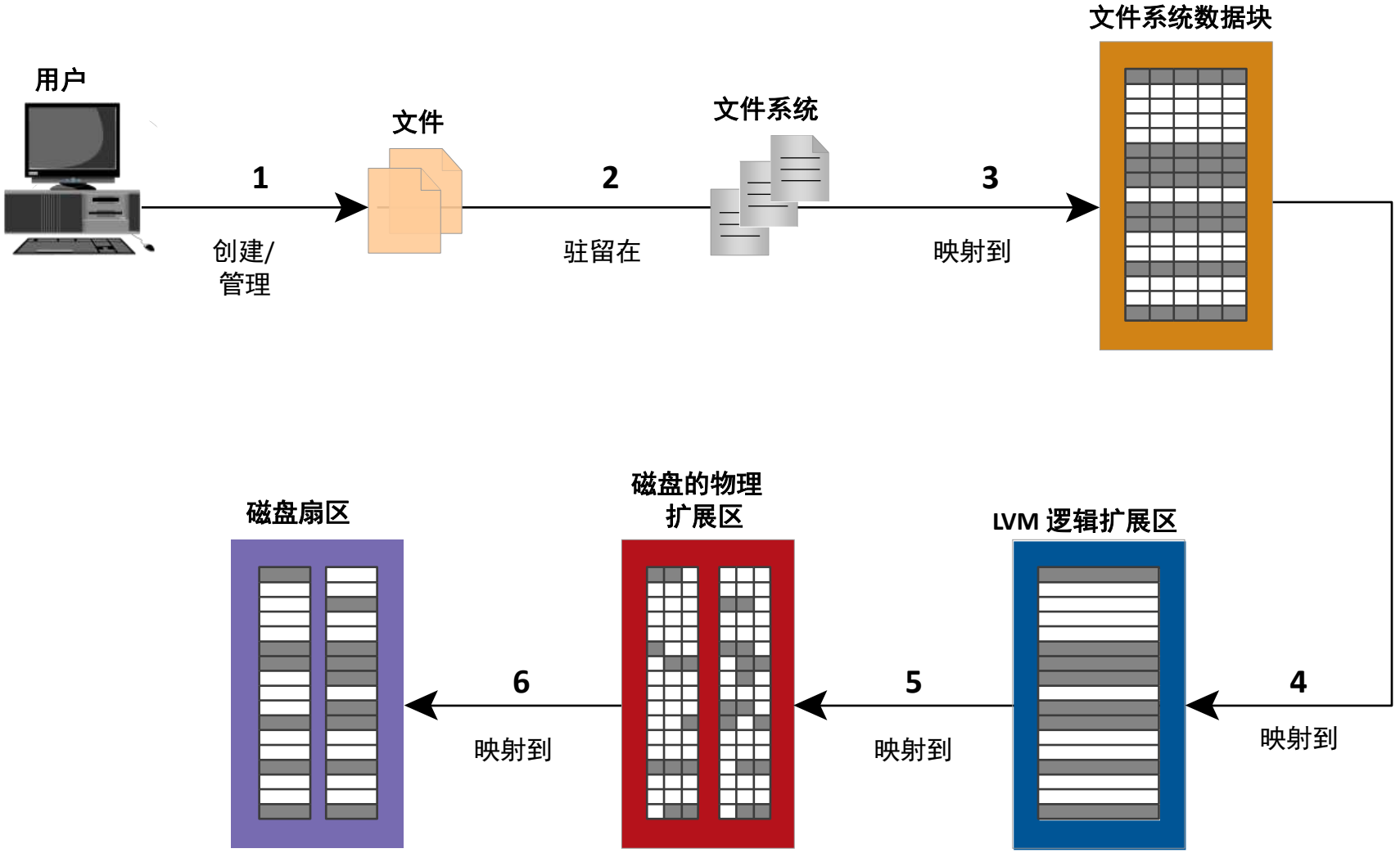
- 借助底层计算组件运行应用程序的资源
  - 示例：服务器、大型机、笔记本电脑、台式机、平板电脑、服务器群集等。
- 包含硬件和软件组件
- 硬件组件
  - 包括 CPU、内存和输入/输出 (I/O) 设备
- 软件组件
  - 包括 OS、设备驱动程序、文件系统、卷管理器等



操作系统和设备驱动程序

- 在传统环境中，OS 位于应用程序和硬件之间
  - 负责控制环境
- 在虚拟化环境中，虚拟化层在 OS 和硬件之间工作
  - 虚拟化层可控制环境
  - OS 作为来宾工作且仅控制应用程序环境
  - 在某些实现中，会修改 OS 以与虚拟化层进行通信
- 设备驱动程序是使 OS 能够识别特定设备的软件

文件系统

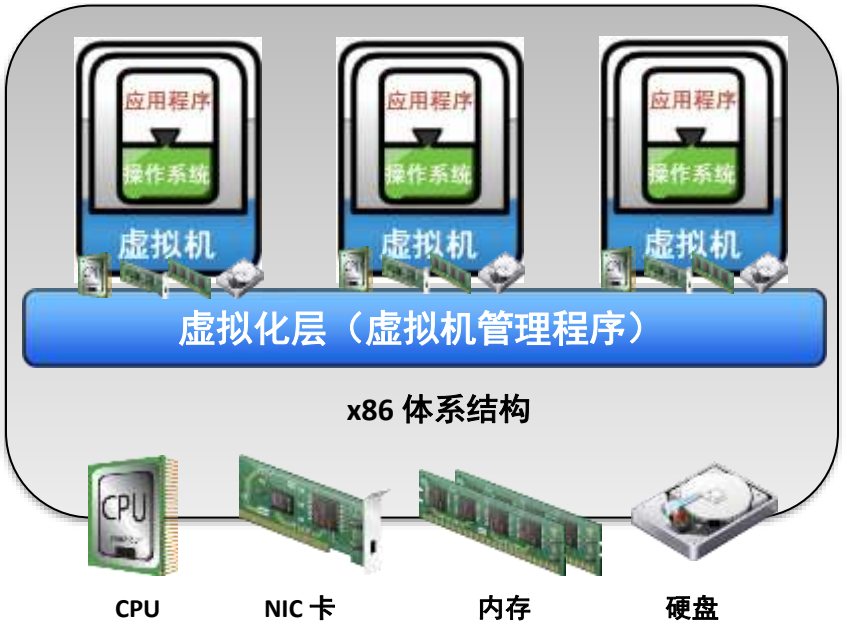

用户
1
创建/
管理
文件
2
驻留在
文件系统
3
映射到
文件系统数据块
4
映射到
LVM 逻辑扩展区
5
映射到
磁盘的物理
扩展区
6
映射到
磁盘扇区

计算虚拟化

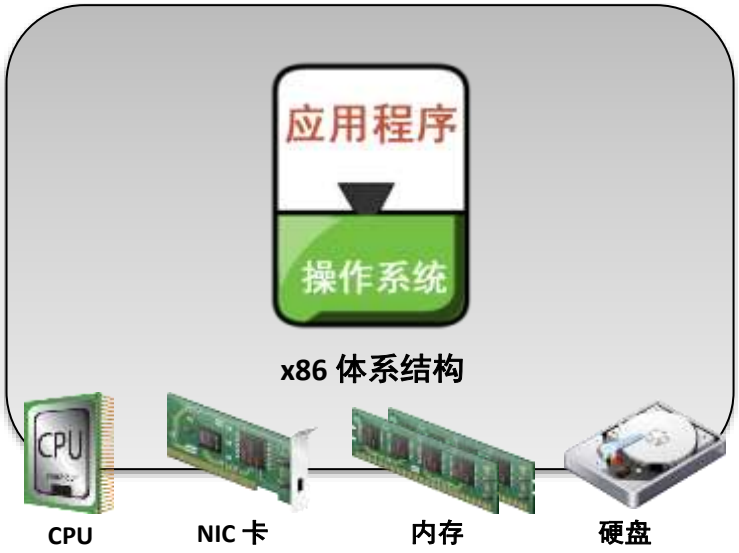
计算虚拟化

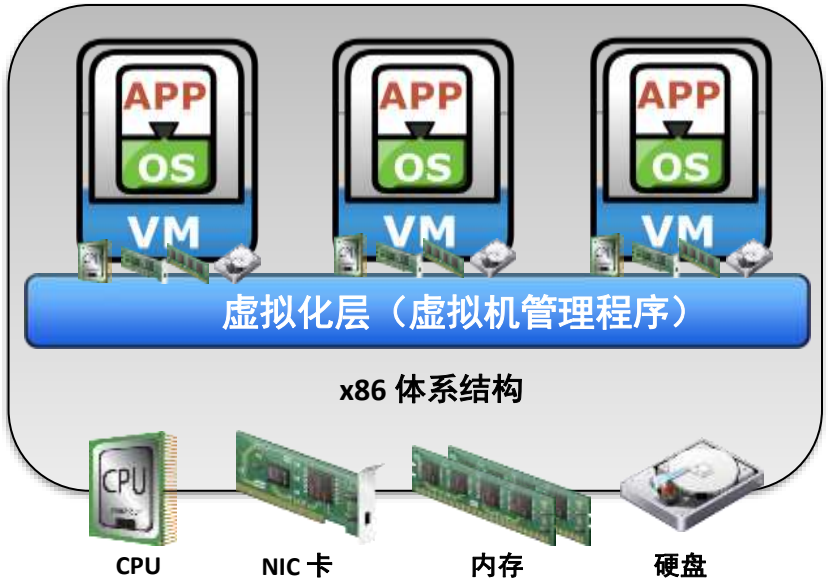
它是一项掩蔽或抽象化物理计算硬件的技术，支持对单个或群集物理机并行运行多个操作系统 (OS)。

- 支持创建多台虚拟机 (VM)，每台都运行一个 OS 和应用程序
  - 虚拟机是外观和行为类似于物理机的逻辑实体
- 虚拟化层位于硬件和虚拟机之间
  - 也称为虚拟机管理程序
- 虚拟机提供有标准化的硬件资源



对计算虚拟化的需求





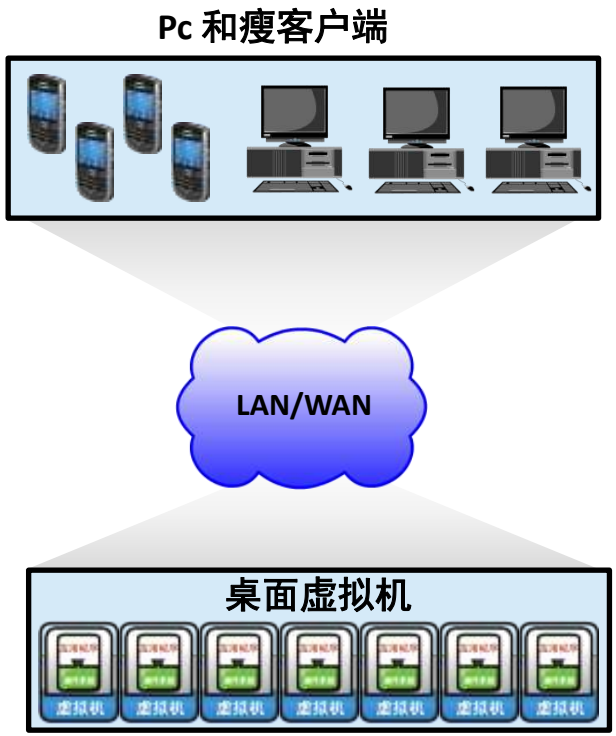
虚拟化之前	虚拟化之后
• 每台计算机一次运行单个操作系统 (OS) • 紧密结合软件和硬件 • 当多个应用程序在相同计算机上运行时可能产生冲突 • 未充分利用资源 • 既不灵活又昂贵	• 每个物理机并行运行多个操作系统 (OS) • 使 OS 与应用程序硬件独立 • 将虚拟机互相隔离，因此，不会有冲突 • 提高了资源利用率 • 以较低成本提供灵活的基础架构

桌面虚拟化

桌面虚拟化

这是一项支持从终端设备断开用户状态、操作系统 (OS) 和应用程序的技术。

- 支持组织托管和集中管理桌面
  - 台式机在数据中心中作为虚拟机运行并通过网络进行访问
- 桌面虚拟化好处
  - 因启用瘦客户端带来的访问灵活性
  - 改进的数据安全性
  - 简化的数据备份和 PC 维护



模块 2：数据中心环境

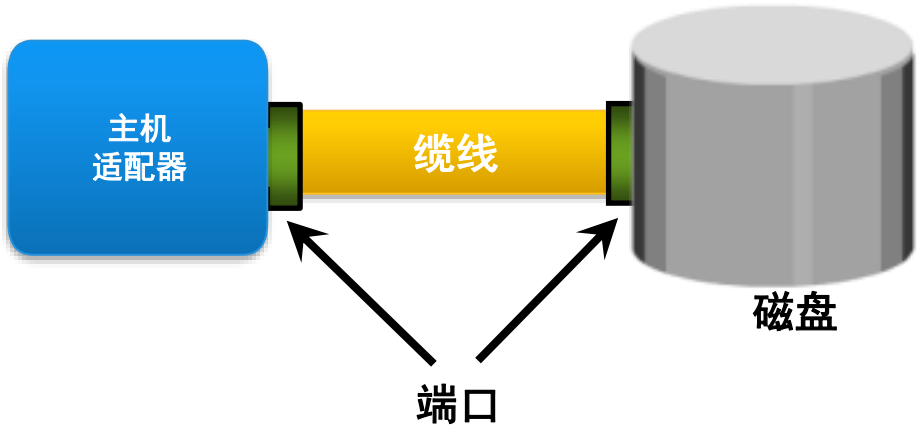
第 2 课：连接

本课程将讲述下列主题：

- 连接的物理组件
- 存储连接协议

连接

- 主机之间或主机与外围设备（如存储）之间的互连
- 连接的物理组件为：
  - 主机接口卡、端口和缆线
- 协议 = 发送设备与接收设备之间定义的通信格式
  - 常用存储接口协议为：IDE/ATA 和 SCSI



IDE/ATA 和串行 ATA

- 集成的设备电子系统 (IDE)/高级技术附件 (ATA)
  - 用于连接硬盘或 CD-ROM 驱动器的常用接口
  - 此协议支持并行传输，因此也称为并行 ATA (PATA) 或简称 ATA。
  - IDE/ATA 具有各种标准和名称。超级 DMA/133 版本的 ATA 支持每秒 133 MB 的吞吐量。
- 串行高级技术附件 (SATA)
  - 已取代并行 ATA 的 IDE/ATA 规范的串行版本
  - 存储互连成本低廉，通常用于内部连接
  - 提供的数据传输速度高达 6 Gb/s（标准 3.0）

SCSI 和 SAS

- 并行小型计算机系统接口 (SCSI)
  - SCSI 已成为高端计算机的首选连接协议。
    - 常用于服务器中的存储连接
  - 与 IDE/ATA 相比，此协议支持并行传输且提供了改进的性能、可扩展性和兼容性，但成本更高，因此在 PC 环境中不常用
  - 可用于众多种类的相关技术和标准
  - 一条总线上最多支持 16 个设备
  - Ultra-640 版本可提供的最大数据传输速度为 640 MB/s
- 串行连接 SCSI (SAS)
  - 取代并行 SCSI 的点到点串行协议
  - 支持的最大数据传输速度为 6 Gb/s (SAS 2.0)

光纤通道和 IP

- 光纤通道 (FC)
  - 用于和存储设备进行高速通信的广泛使用的协议
  - 可提供通过铜线和/或光纤操作的串行数据传输
  - 最新版本的 FC 接口“16FC”允许的最大数据传输速度为 16 Gb/s
- Internet 协议 (IP)
  - 一直以来用于传输主机到主机流量
  - 提供利用现有基于 IP 的网络进行存储通信的机会
    - 示例：iSCSI 和 FCIP 协议

模块 2：数据中心环境

第 3 课：存储

本课程将讲述下列主题：

- 各种存储选项
- 磁盘驱动器的组件、寻址和性能
- 企业级闪存驱动器
- 主机对存储的访问和直连存储

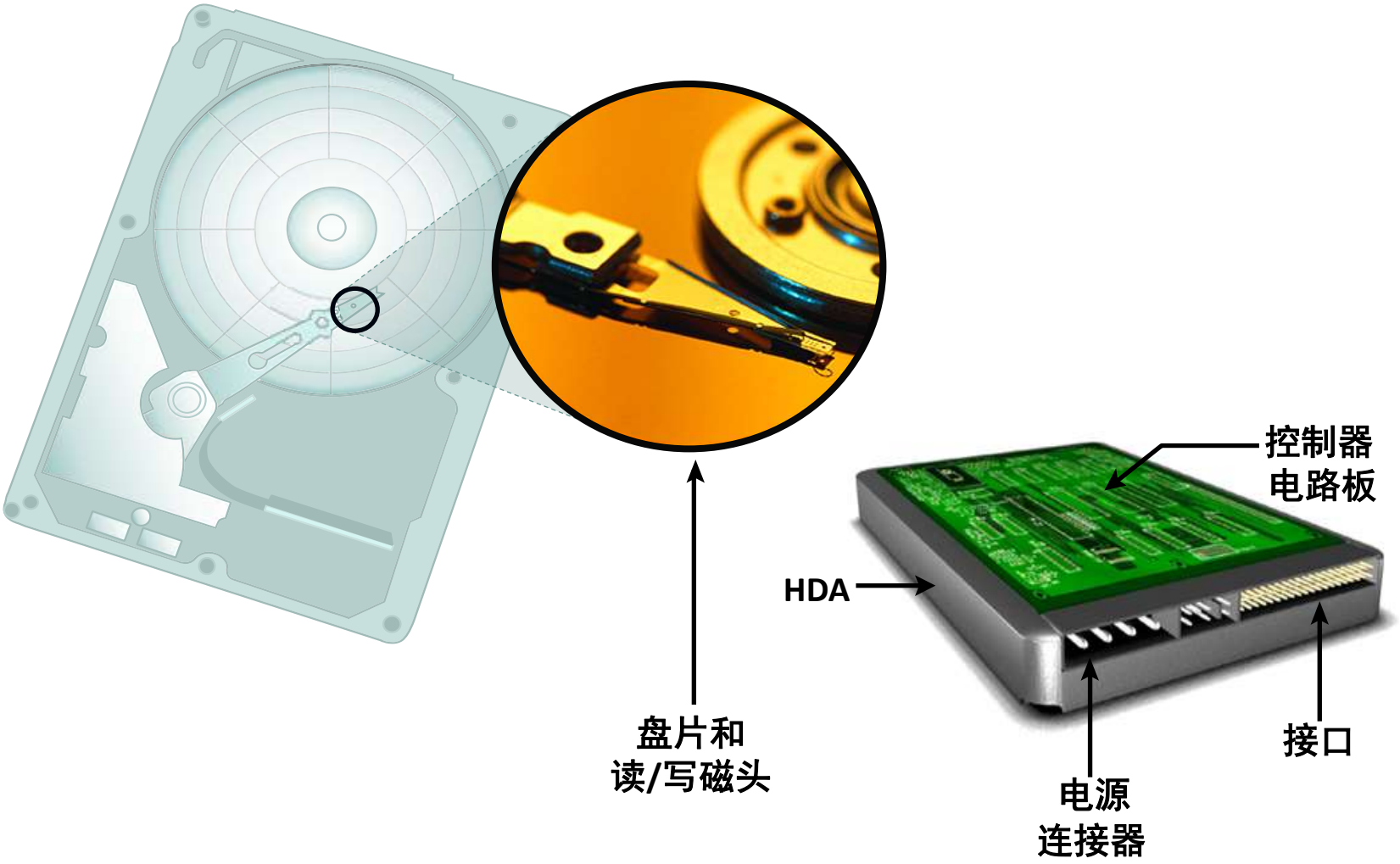
存储选项

- 磁带
  - 低成本的长期数据存储解决方案
    - 过去备份目标的首选选项
  - 局限性
    - 按顺序进行数据访问
    - 一次只能进行一项应用程序访问
    - 存在物理磨损
    - 存储/检索开销

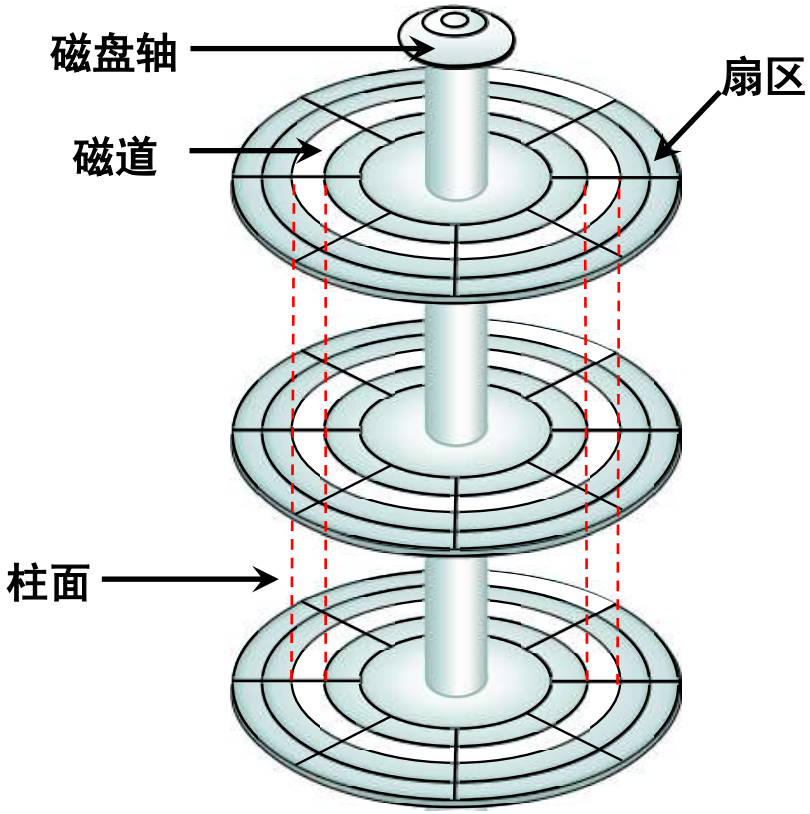
存储选项（续）

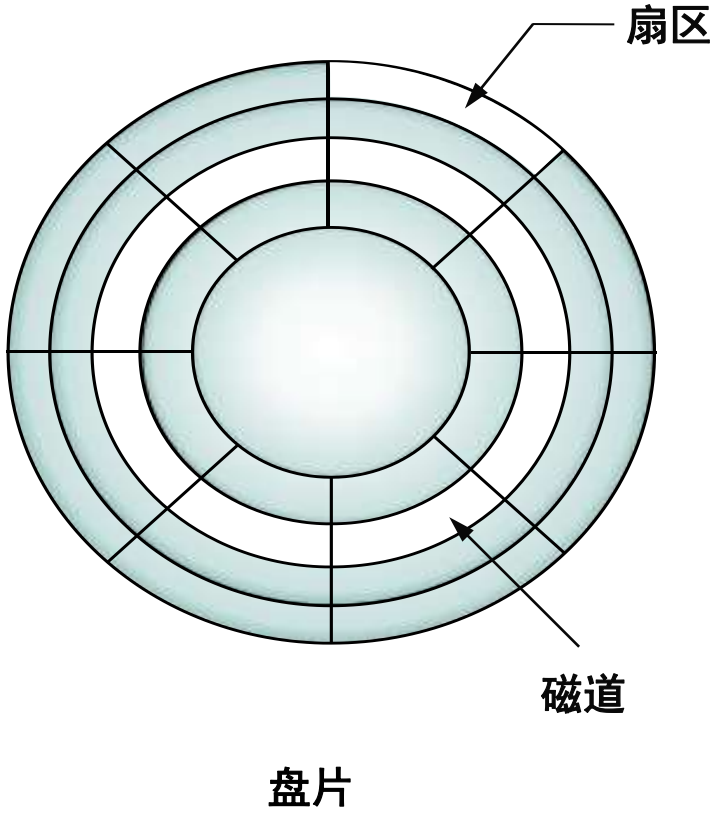
- 光盘
  - 在小型的单用户计算环境中广泛用作分发介质
  - 在容量和速度方面有限
  - 一次写入、多次读取 (WORM)：CD-ROM、DVD-ROM
  - 其他变体：CD-RW、Blu-ray 磁盘
- 磁盘驱动器
  - 最为流行的存储介质
  - 具有很大的存储容量
  - 随机读/写访问
- 闪存驱动器（或固态驱动器 - SSD）
  - 使用半导体介质
  - 提供高性能、低功耗

磁盘驱动器的组件

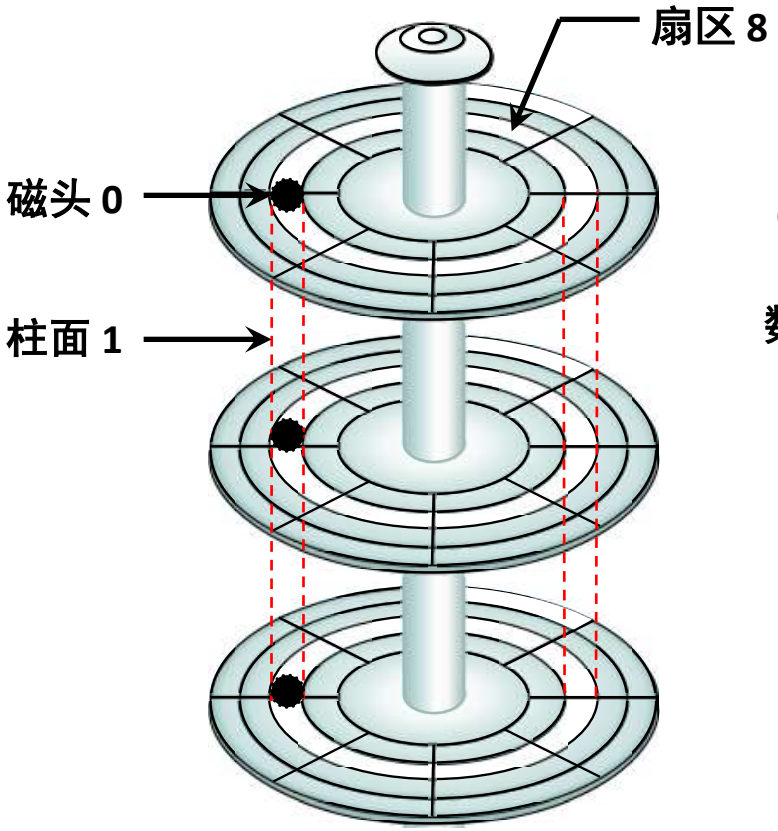


物理磁盘的结构

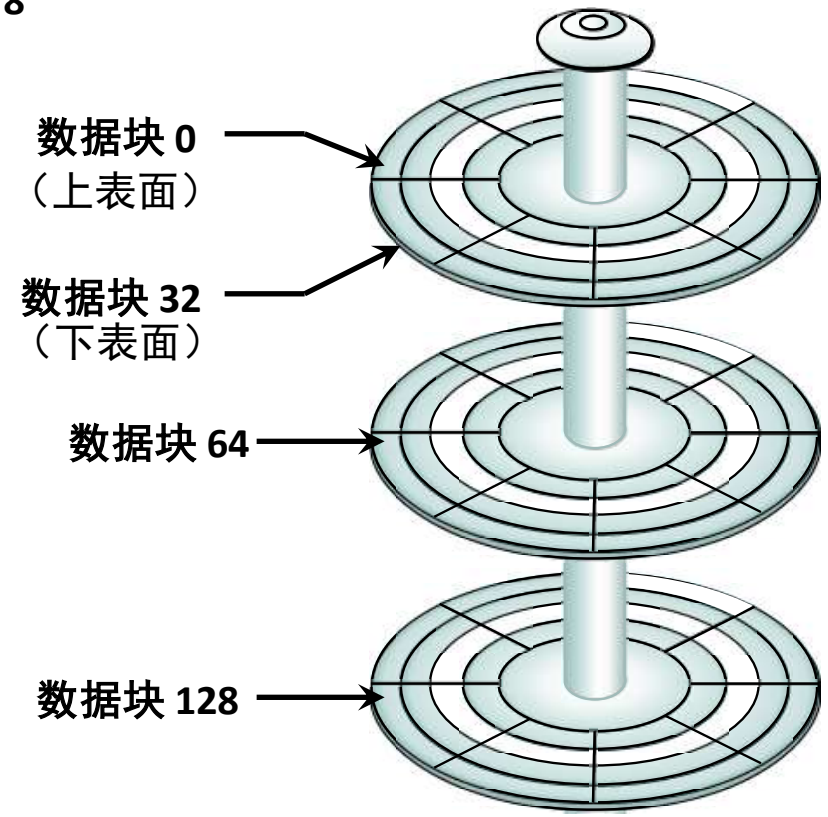




逻辑数据块寻址



物理地址 = CHS



逻辑数据块地址 = 数据块编号

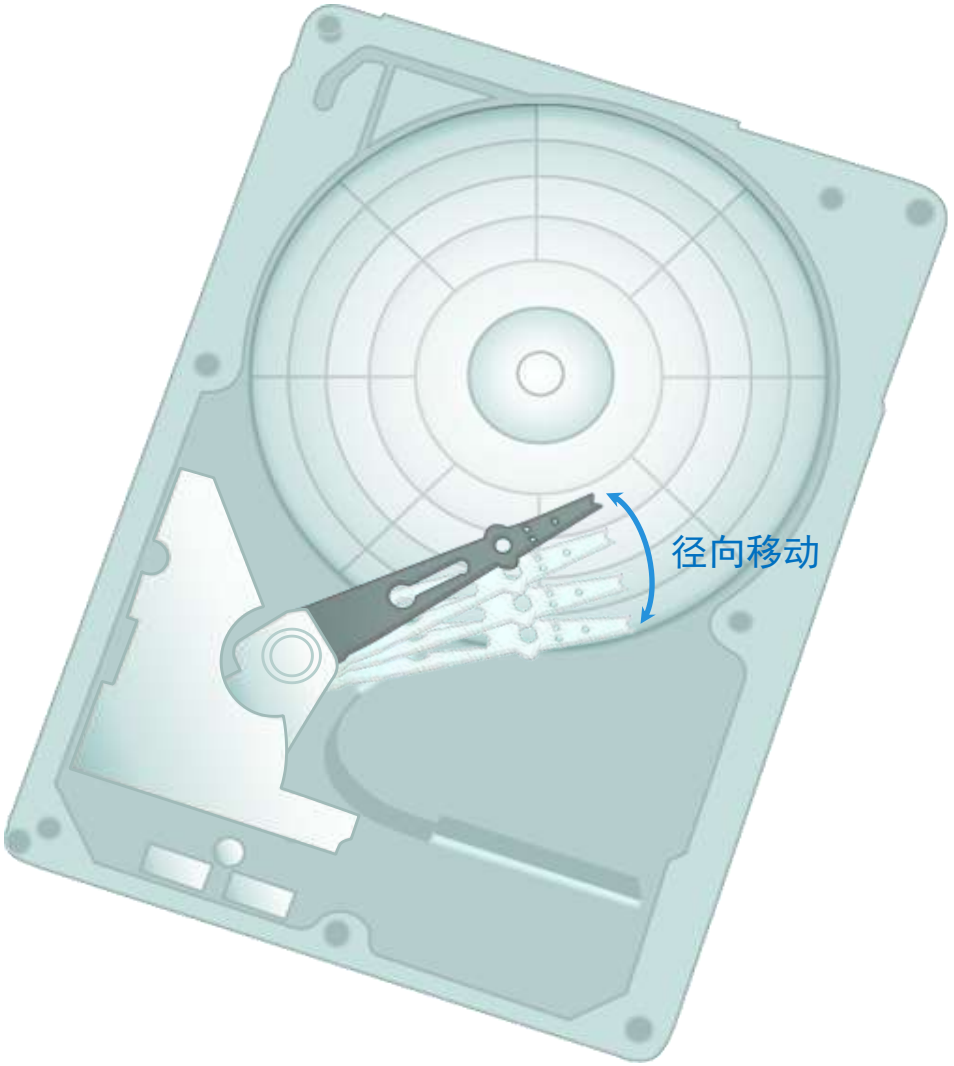
磁盘驱动器的性能

- 电子机械设备
  - 影响存储系统的总体性能
- 磁盘服务时间
  - 磁盘完成 I/O 请求所用的时间，取决于：
    - 寻道时间
    - 旋转延迟
    - 数据传输速度

磁盘服务时间 = 寻道时间 + 旋转延迟 + 数据传输时间

寻道时间

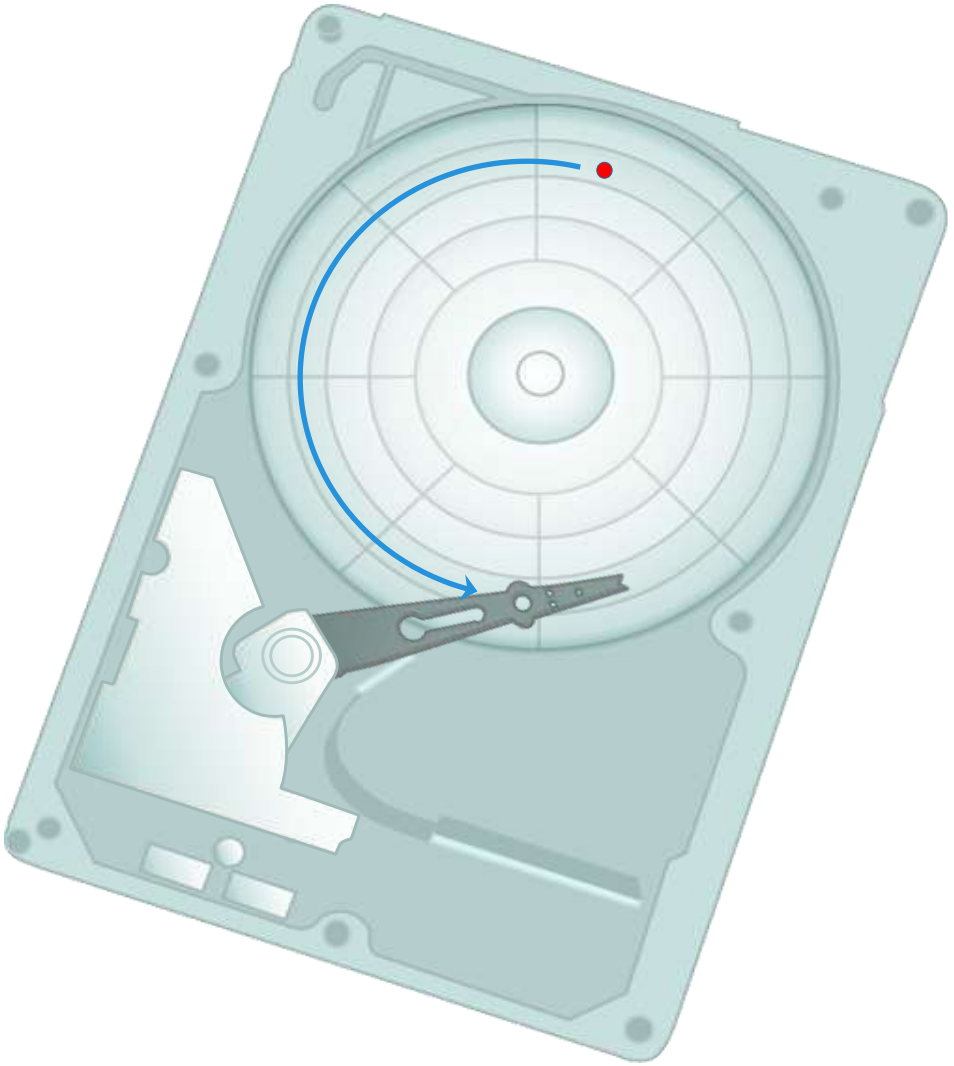
- 定位读/写磁头所用的时间
- 寻道时间越短，I/O 操作越快
- 寻道时间规范包括
  - 全程
  - 平均
  - 道间
- 磁盘的寻道时间由驱动器制造商指定
- 现代磁盘的平均寻道时间通常在 3 到 15 毫秒的范围内。



旋转延迟

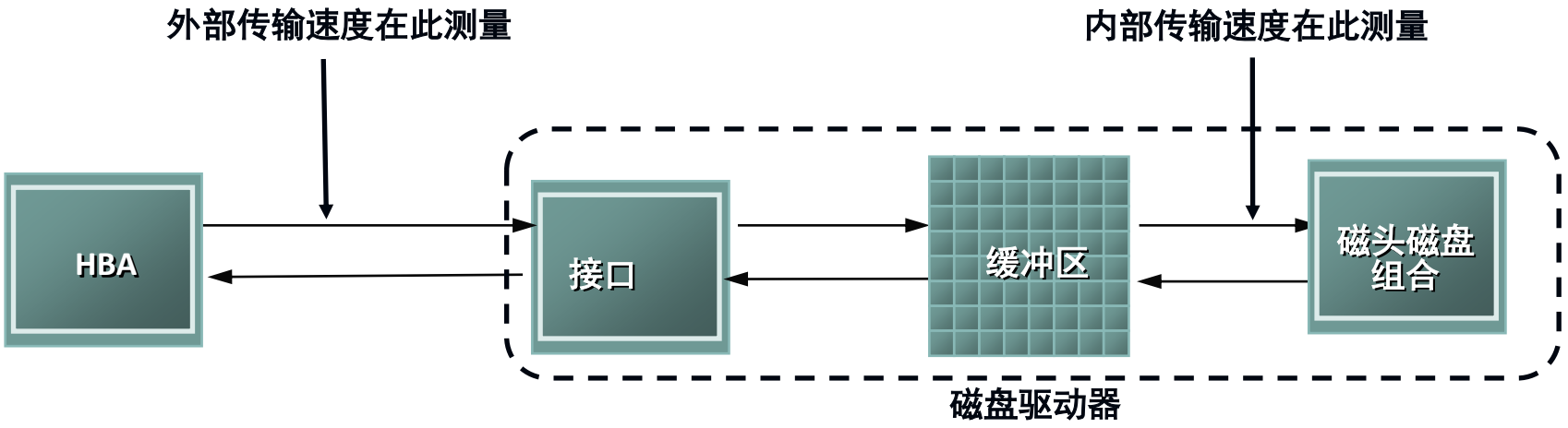
- 盘片通过旋转将数据置于读/写磁头下所用的时间
- 取决于磁盘轴的旋转速度
- 平均旋转延迟
  - 旋转一周所用的时间的一半
  - 对于“*X*”rpm，驱动器延迟以毫秒为单位按以下方式计算：

$$= \frac{1/2}{(X/60)}$$



数据传输速度

- 每单位时间驱动器可以向 HBA 输送的平均数据量
  - 内部传输速度：数据从盘片表面移至磁盘内部缓冲区时的速度
  - 外部传输速度：数据通过接口移至 HBA 时的速度



基于应用程序要求和磁盘驱动器性能的存储设计

- 为满足应用程序容量需求所需的磁盘 (D_C)：

$$D_C = \frac{\text{所需总容量}}{\text{单个磁盘的容量}}$$

- 为满足应用程序性能需求所需的磁盘 (D_P)：

$$D_P = \frac{\text{峰值工作负载时应用程序生成的 IOPS}}{\text{由单个磁盘提供服务的 IOPS}}$$

- 根据磁盘服务时间由磁盘提供服务的 IOPS：

$$T_S = \text{寻道时间} + \frac{0.5}{(\text{磁盘转速}/60)} + \frac{\text{数据块大小}}{\text{数据传输速度}}$$

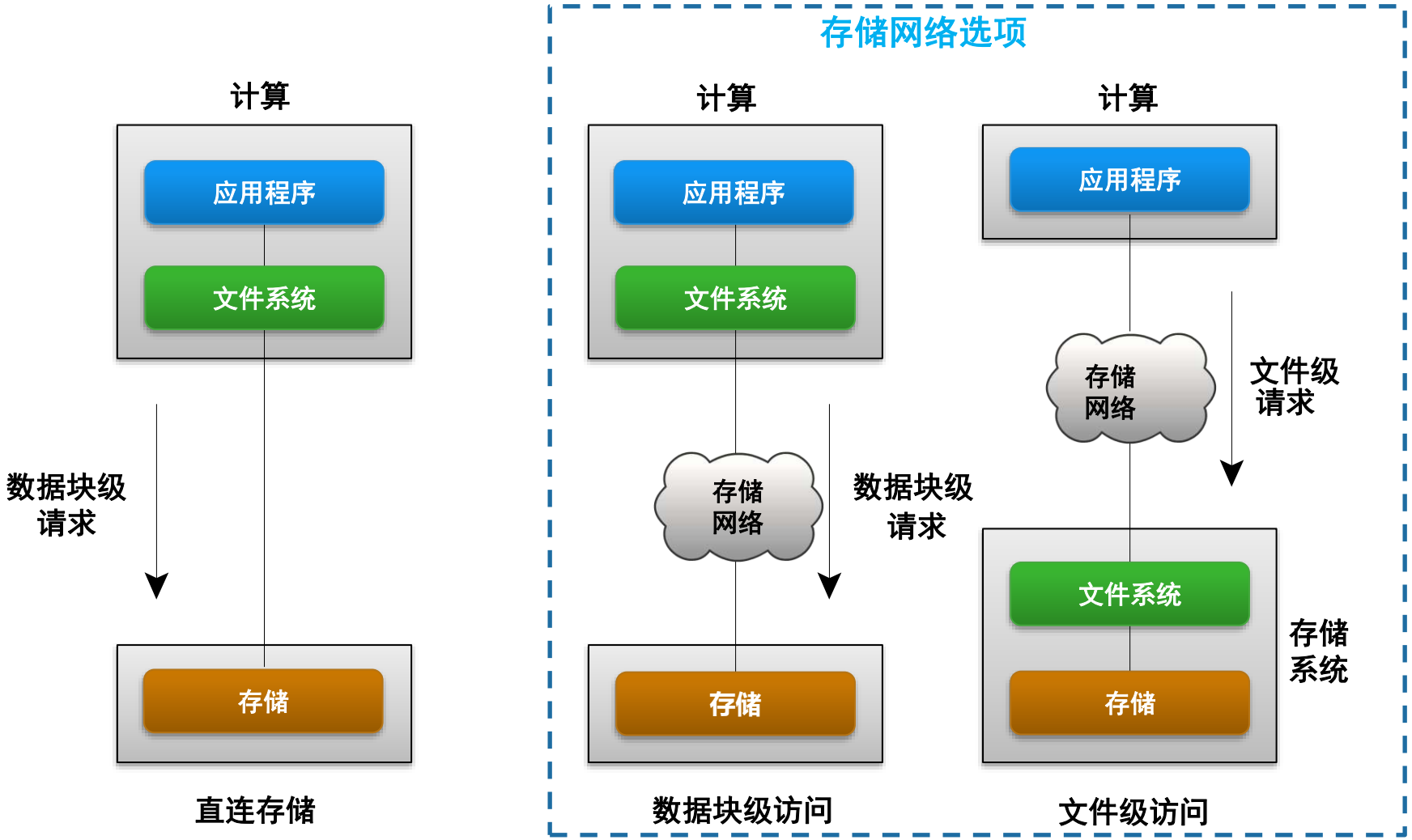
  - T_S 是 I/O 完成所用的时间，因此，由磁盘提供服务的 IOPS 等于 $(1/T_S)$

应用程序所需的磁盘 = max (D_C,D_P)

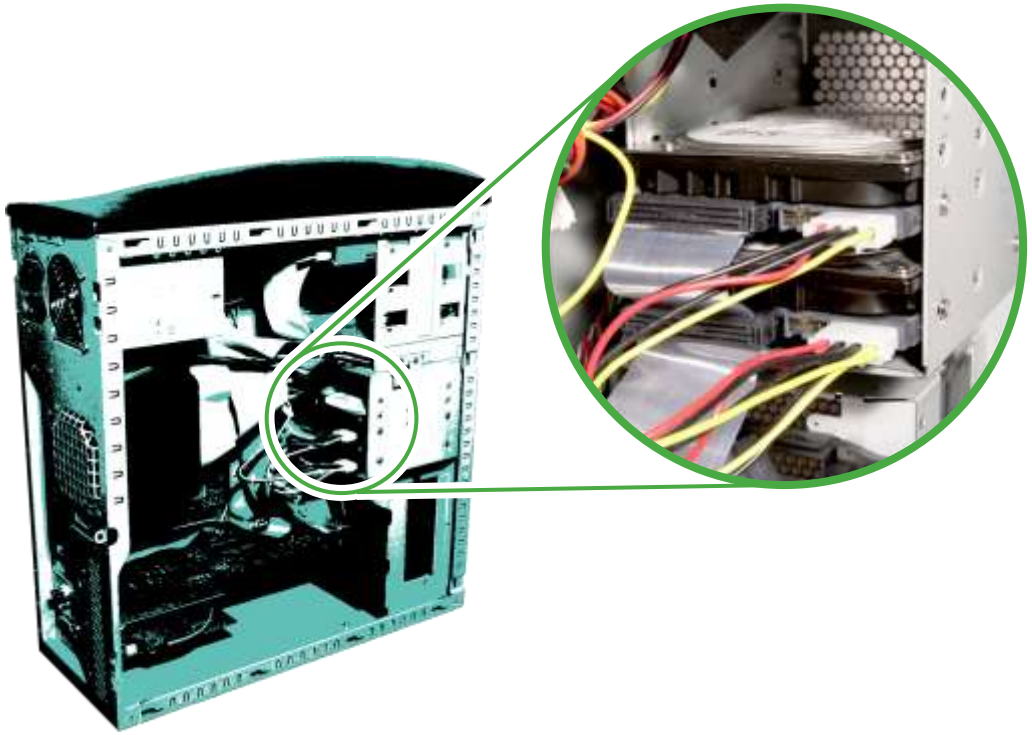
企业级闪存驱动器

传统硬盘驱动器	闪存驱动器
寻道时间和旋转延迟导致机械延迟	由于没有机械运动而使每个驱动器达到最高的吞吐量
性能和 I/O 服务功能有限	每个 I/O 的延迟非常低且 I/O 性能始终如一
由于采用机械操作而增加功耗	高能效 • 每 GB 的电源需求降低 • 每 IOPS 的电源需求降低
平均无故障时间 (MTBF) 缩短	因无运动部件而实现高可靠性
更多数量的磁盘、电源、冷却和管理成本导致 TCO 提高	总体降低了 TCO

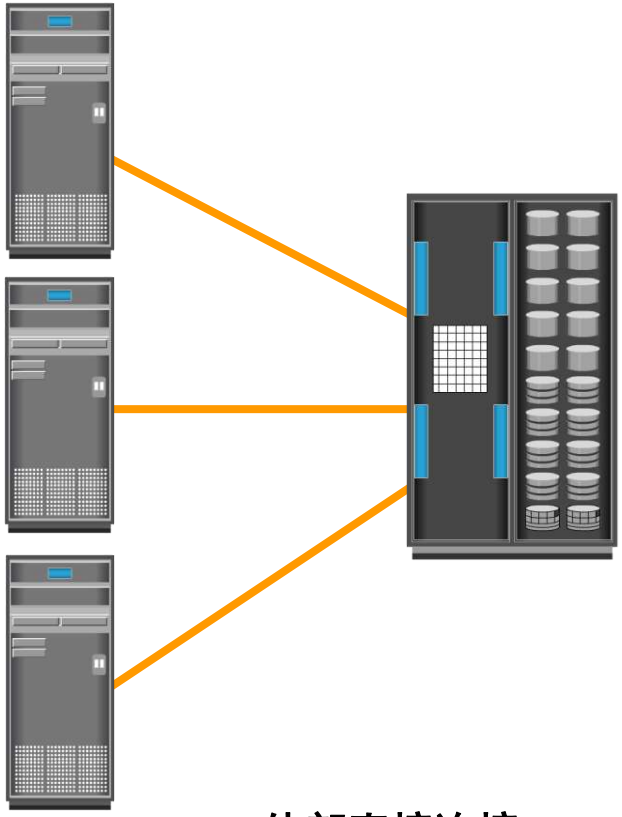
主机对存储的访问


存储网络选项
计算
应用程序
文件系统
数据块级
请求
存储
直连存储
计算
应用程序
文件系统
存储
网络
数据块级
请求
存储
数据块级访问
计算
应用程序
存储
网络
文件级
请求
文件系统
存储
存储
系统
文件级访问

直连存储 (DAS)



内部直接连接



外部直接连接

模块 – 3

数据保护 – RAID

模块 3：数据保护 – RAID

学完本模块后，您将能够：

- 描述 RAID 实现方法
- 描述三种 RAID 技术
- 描述常用 RAID 级别
- 描述 RAID 对性能的影响
- 根据 RAID 级别的成本、性能和保护能力比较各个级别

模块 3：数据保护 – RAID

第 1 课：RAID 概述

本课程将讲述下列主题：

- RAID 实现方法
- RAID 阵列组件
- RAID 技术

为什么选择 RAID？ Redundant Arrays of Independent Disks

RAID

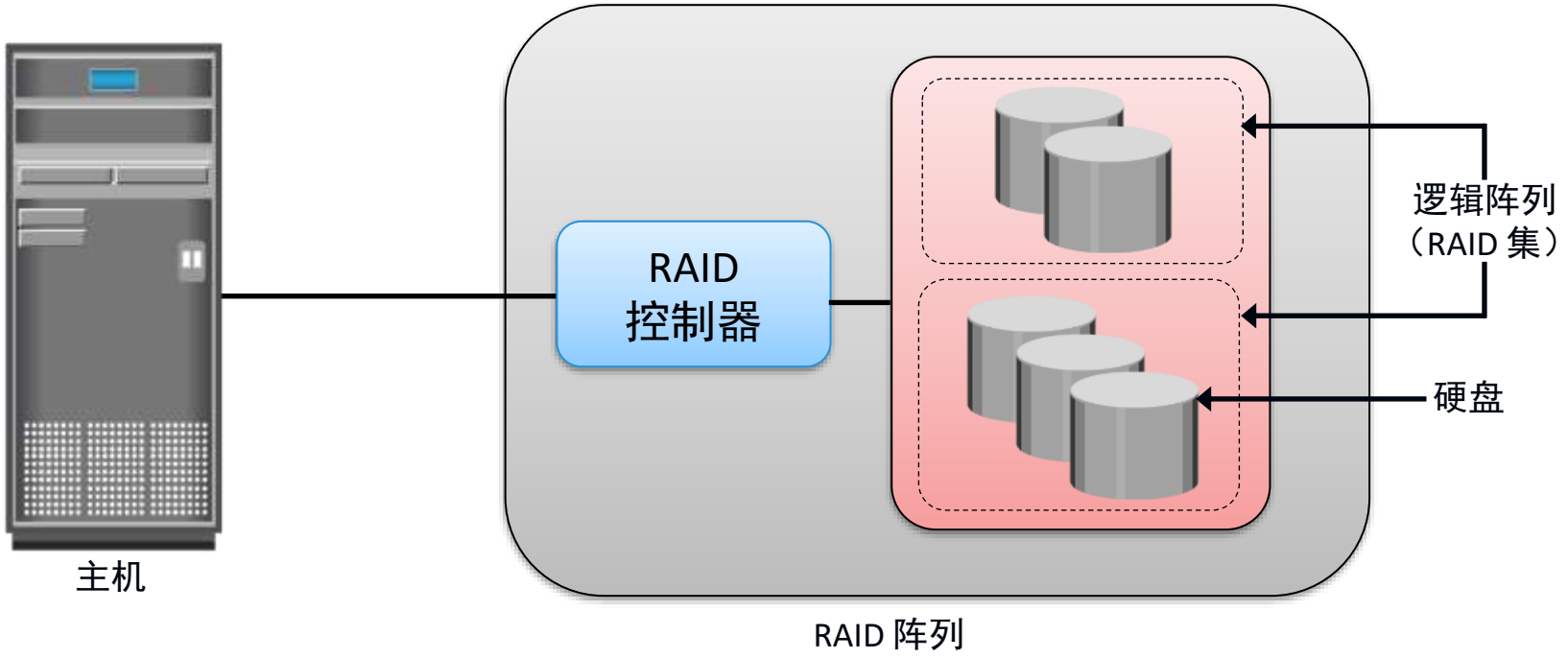
它是一项将多个磁盘驱动器合并到一个逻辑单元（RAID 集）中并提供保护和/或性能的技术。

- 由于磁盘驱动器中包含机械组件，因此它提供的性能有限
- 每个驱动器具有特定的平均预期寿命并以 MTBF 为测量单位：
  - 例如：如果驱动器的 MTBF 为 750,000 小时，而阵列中有 1000 台驱动器，则该阵列的 MTBF 为 750 小时 (750,000/1000)
- 为缓解这些问题而引入了 RAID

RAID 实现方法

- 软件 RAID 实现
  - 使用基于主机的软件提供 RAID 功能
  - 与硬件 RAID 相比，软件 RAID 实现具有成本优势和简单直观的优点，但有较多限制：
    - 使用主机 CPU 周期执行 RAID 计算，从而影响系统整体性能
    - 支持有限的 RAID 级别，仅当 RAID 软件和操作系统兼容时，才可对其进行升级
- 硬件 RAID 实现
  - 可在主机或阵列中实现专用硬件控制器。
  - 主机控制器卡RAID 是基于主机的硬件 RAID 实现，专用 RAID 控制器安装在主机上。该实现在包含大量主机的数据中心环境下不是高效的解决方案。
  - 外部 RAID 控制器是基于阵列的硬件 RAID。它充当主机与磁盘之间的接口。它将存储卷呈现给主机，且主机将这些卷作为物理驱动器进行管理。
  - RAID 控制器的主要功能包括：管理与控制磁盘聚合，转换逻辑磁盘和物理磁盘之间的 I/O 请求，在磁盘出故障时重新生成数据

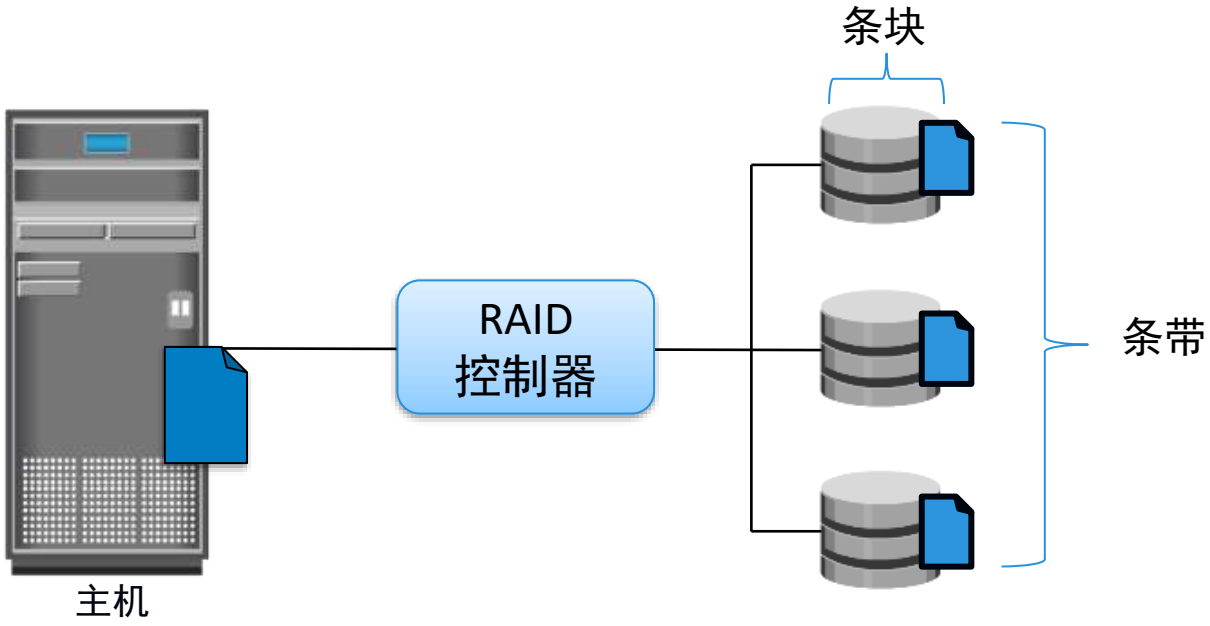
RAID 阵列组件



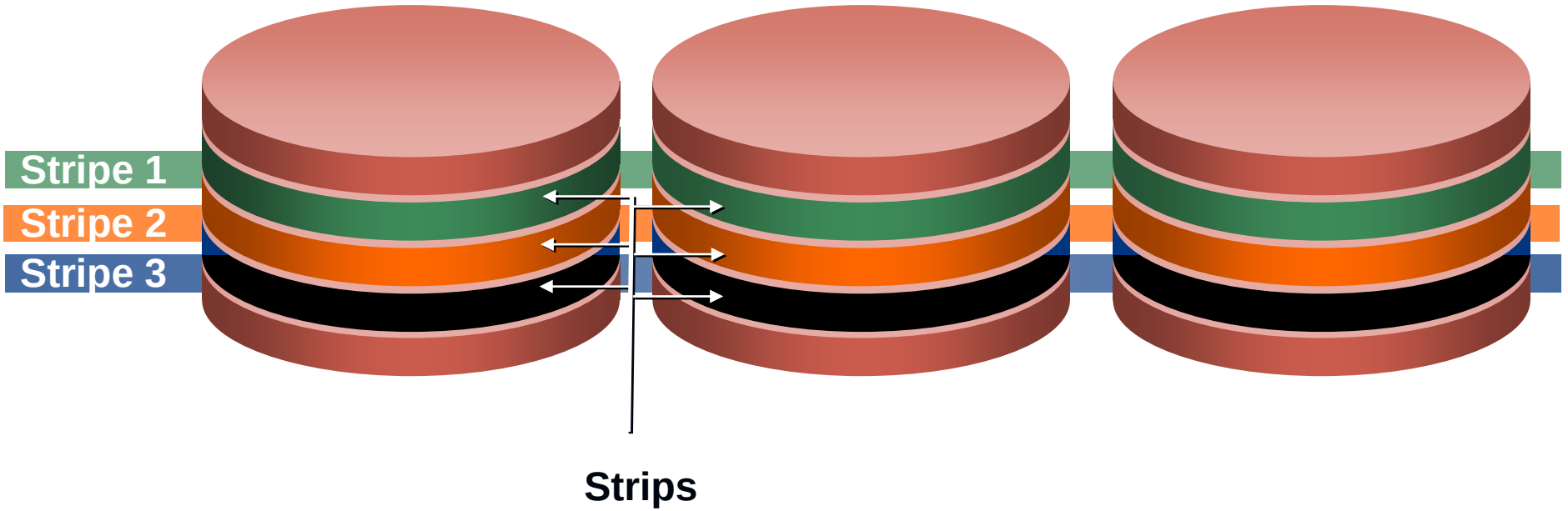
RAID 技术

- 用于 RAID 的三项关键技术是：
  - 分条 (Striping)
  - 镜像 (Mirror)
  - 奇偶校验 (Parity)

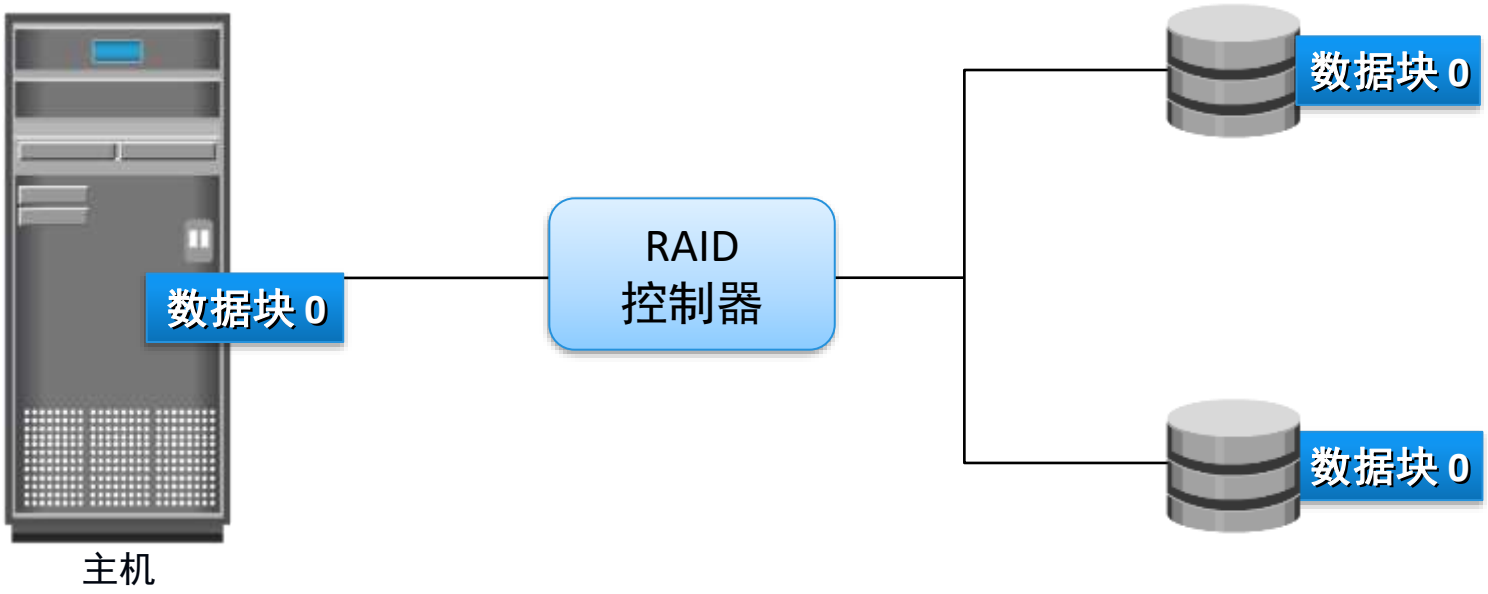
RAID 技术 – 分条



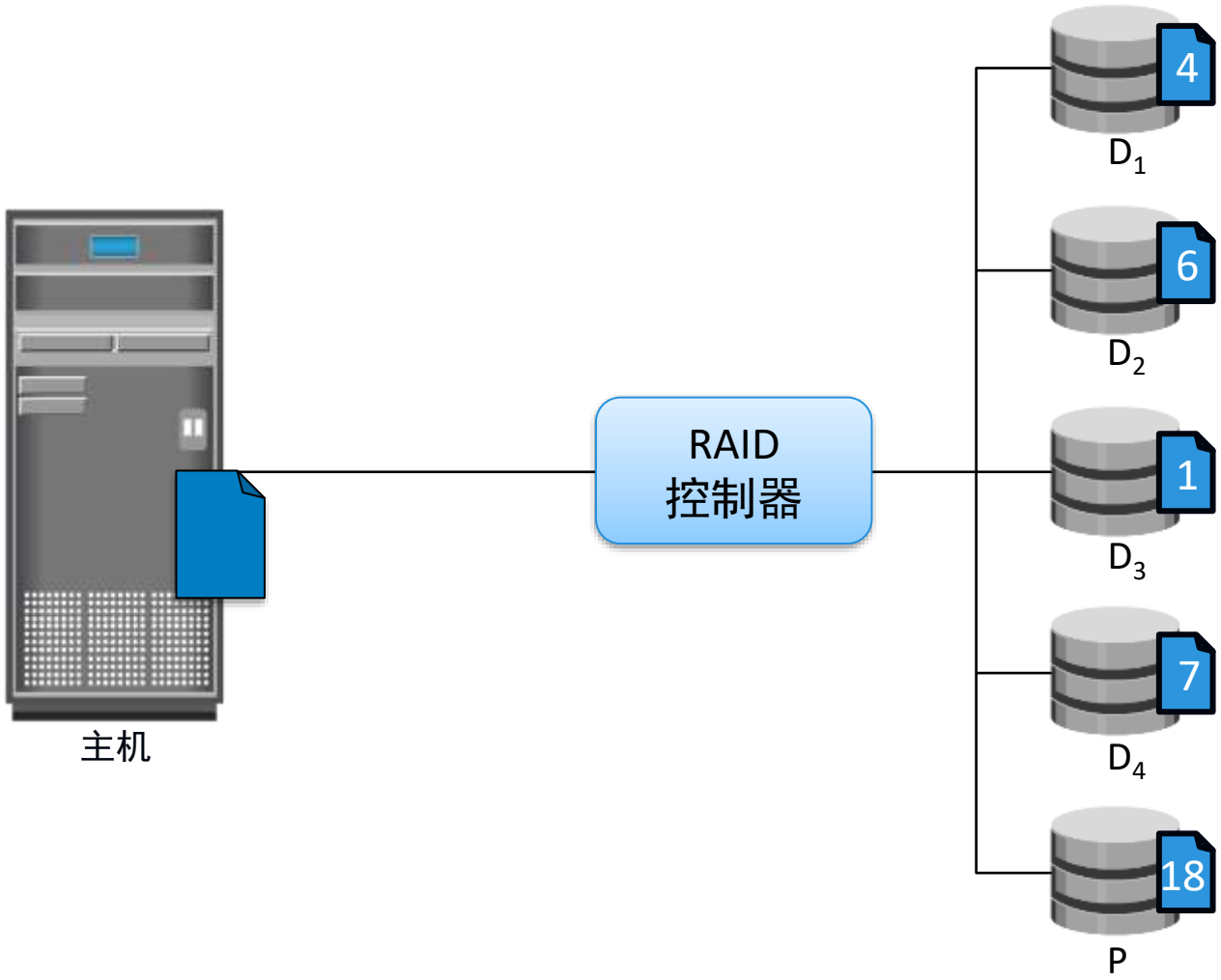
RAID 技术 – 分条- Strips vs Stripe



RAID 技术 – 镜像

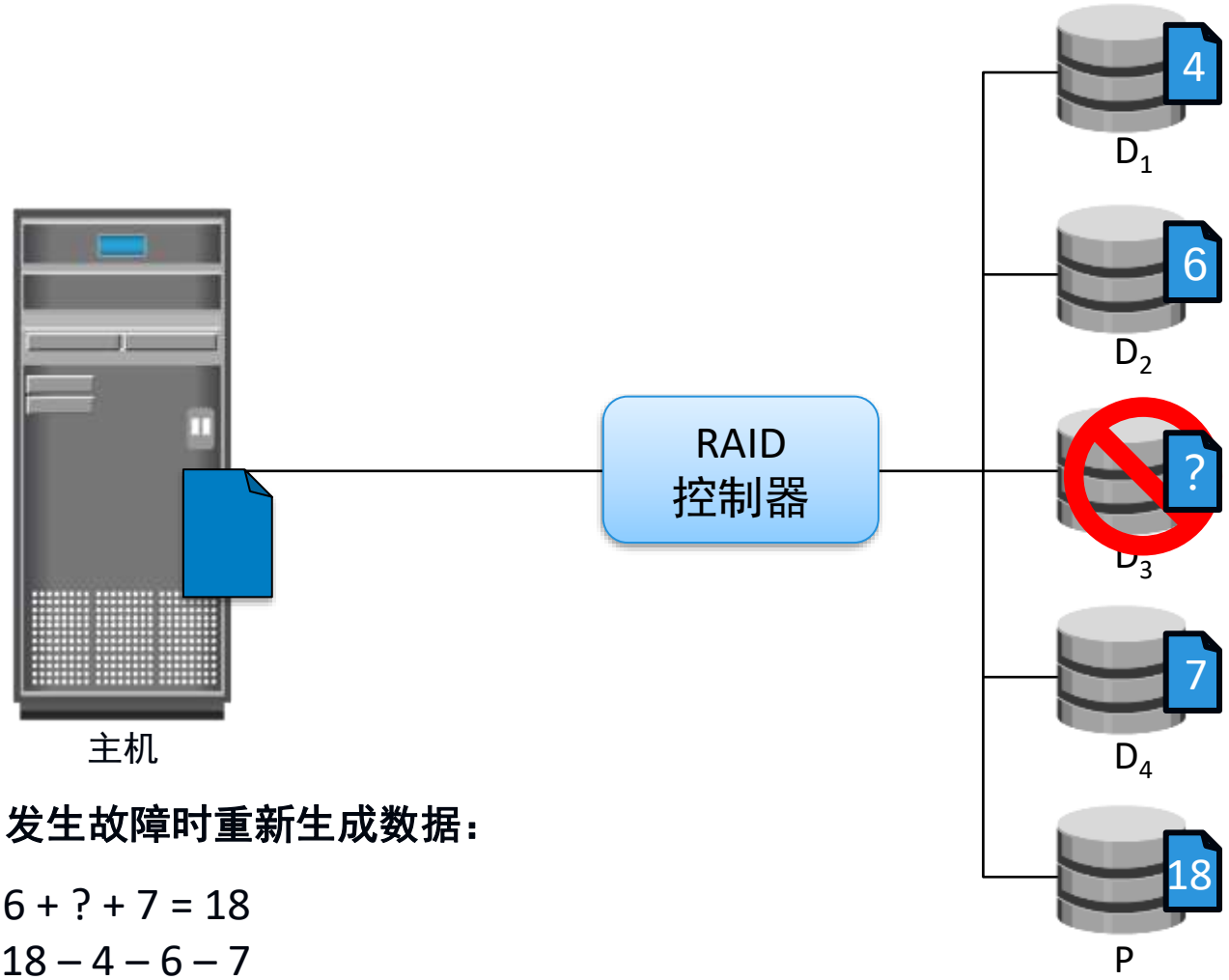


RAID 技术 – 奇偶校验



实际奇偶校验计算是一种 XOR 位运算。

使用奇偶校验技术恢复数据



在驱动器 D_3 发生故障时重新生成数据：

4 + 6 + ? + 7 = 18

? = 18 – 4 – 6 – 7

? = 1

模块 3：数据保护 – RAID

第 2 课：RAID 级别

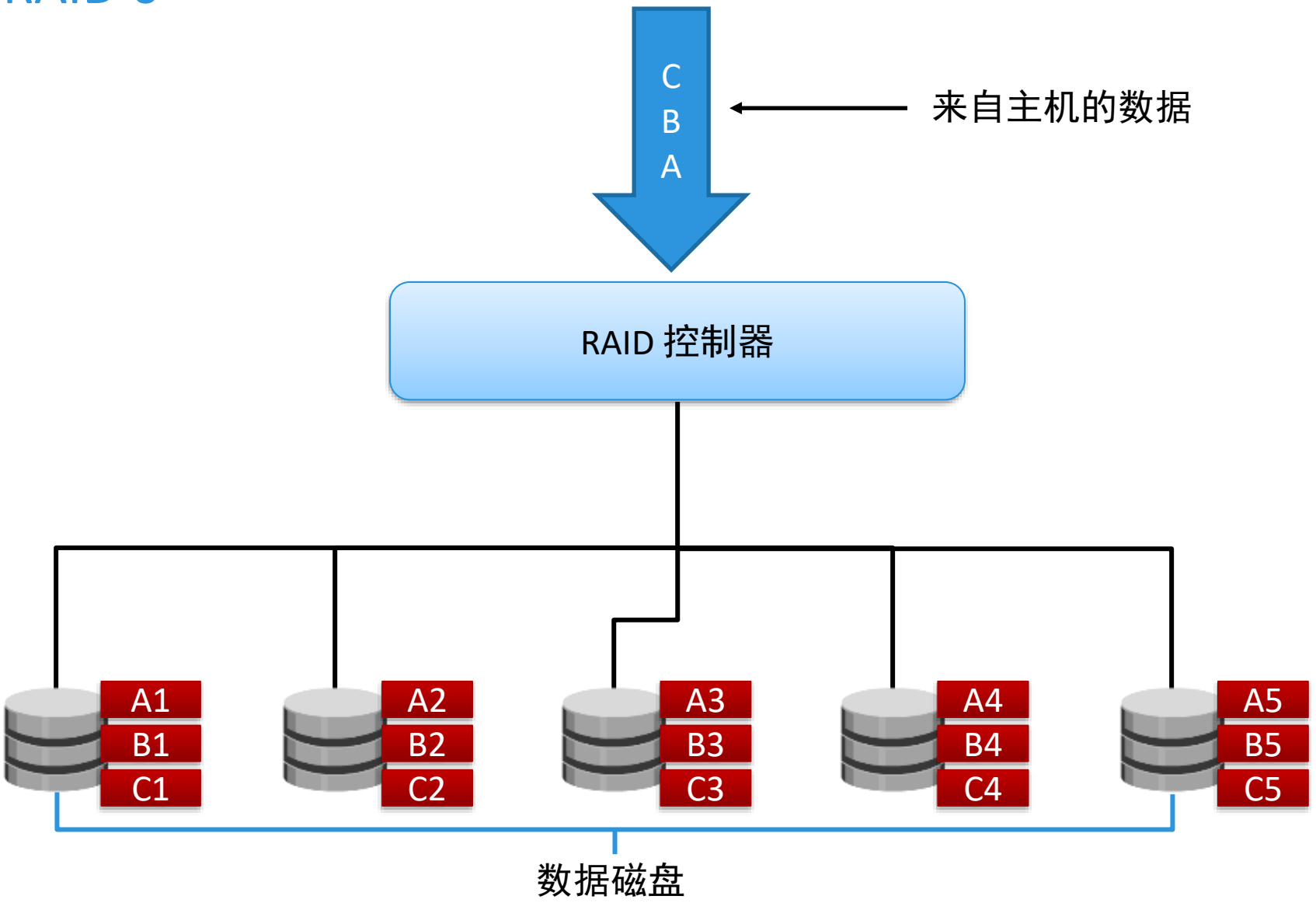
本课程将讲述下列主题：

- 常用 RAID 级别
- RAID 对性能的影响
- RAID 比较
- 热备盘

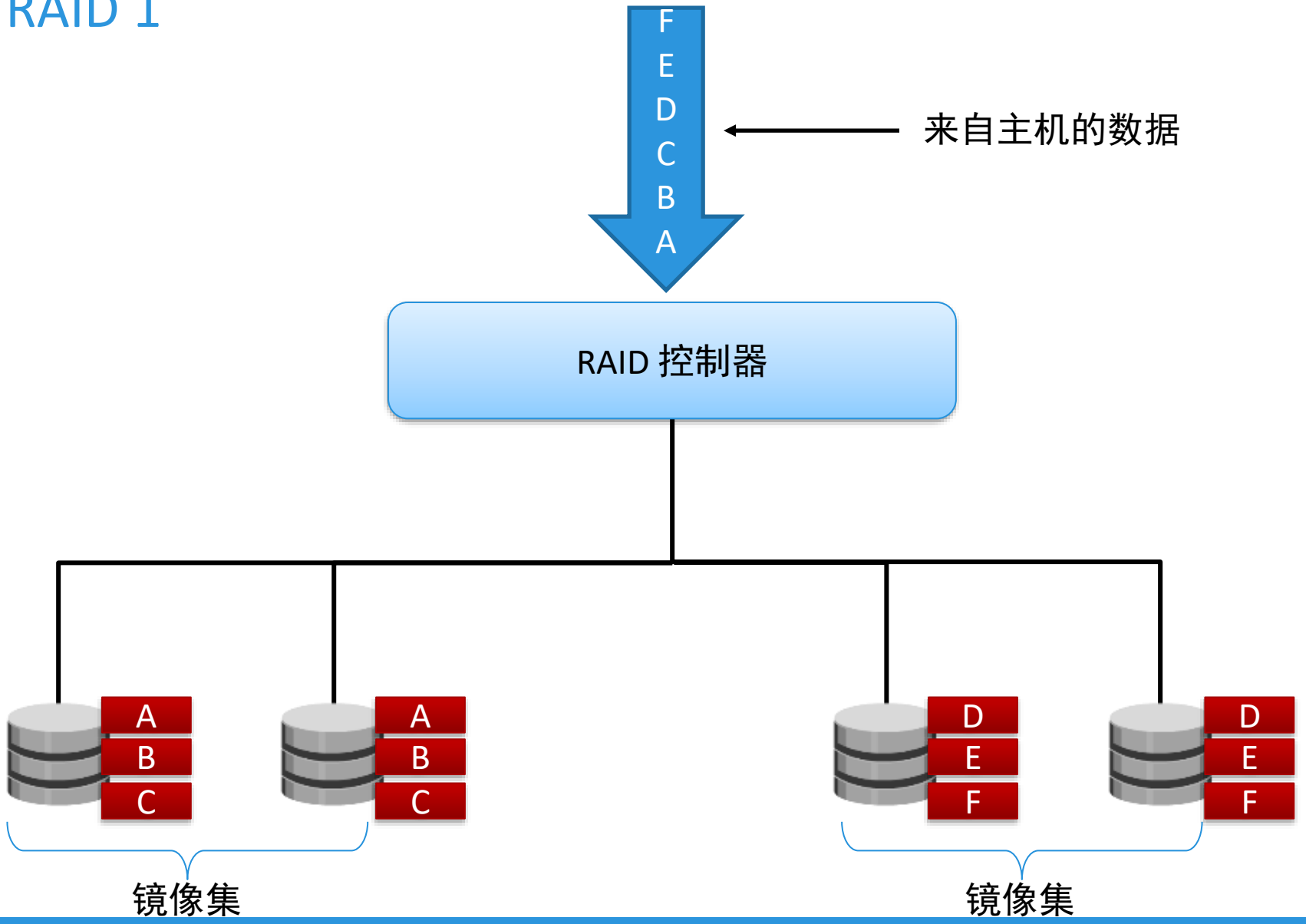
RAID 级别

- 常用 RAID 级别包括：
  - RAID 0 – 无容错能力的分条集
  - RAID 1 – 磁盘镜像
  - RAID 1 + 0 – 嵌套 RAID
  - RAID 3 – 具有并行访问和专用奇偶校验磁盘的分条集
  - RAID 5 – 具有独立磁盘访问和分布式奇偶校验的分条集
  - RAID 6 – 具有独立磁盘访问和双分布式奇偶校验的分条集

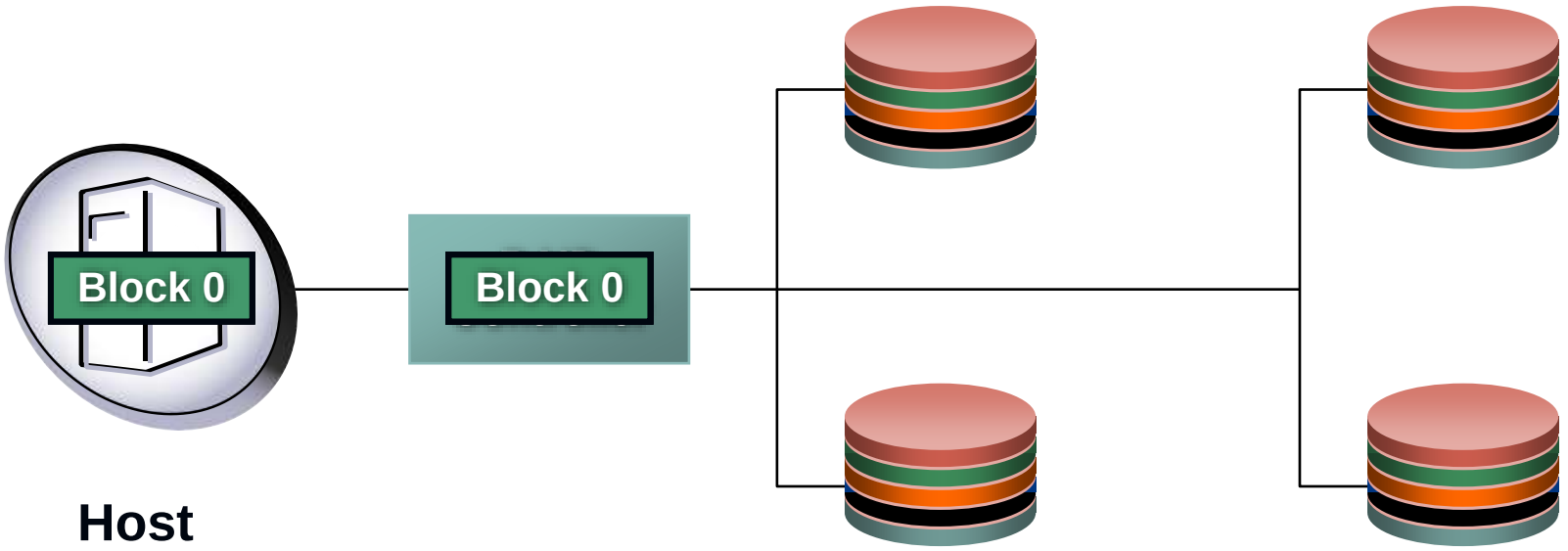
RAID 0


C
B
A
来自主机的数据
RAID 控制器
A1
B1
C1
A2
B2
C2
A3
B3
C3
A4
B4
C4
A5
B5
C5
数据磁盘

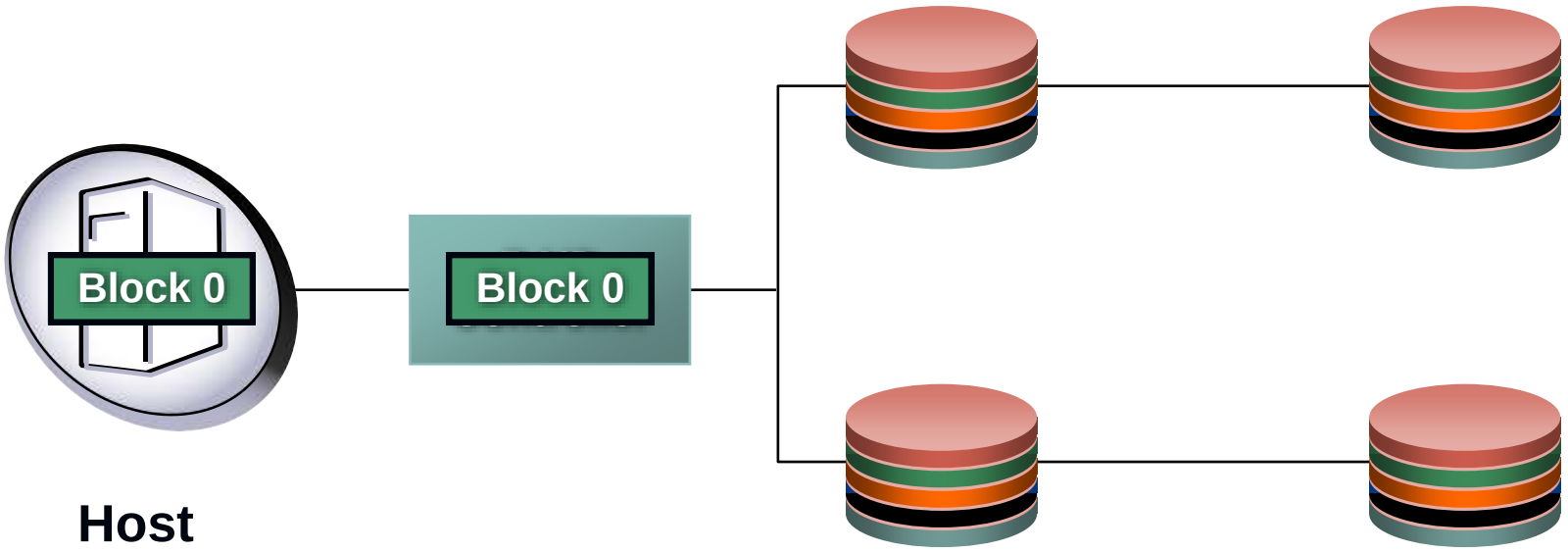
RAID 1



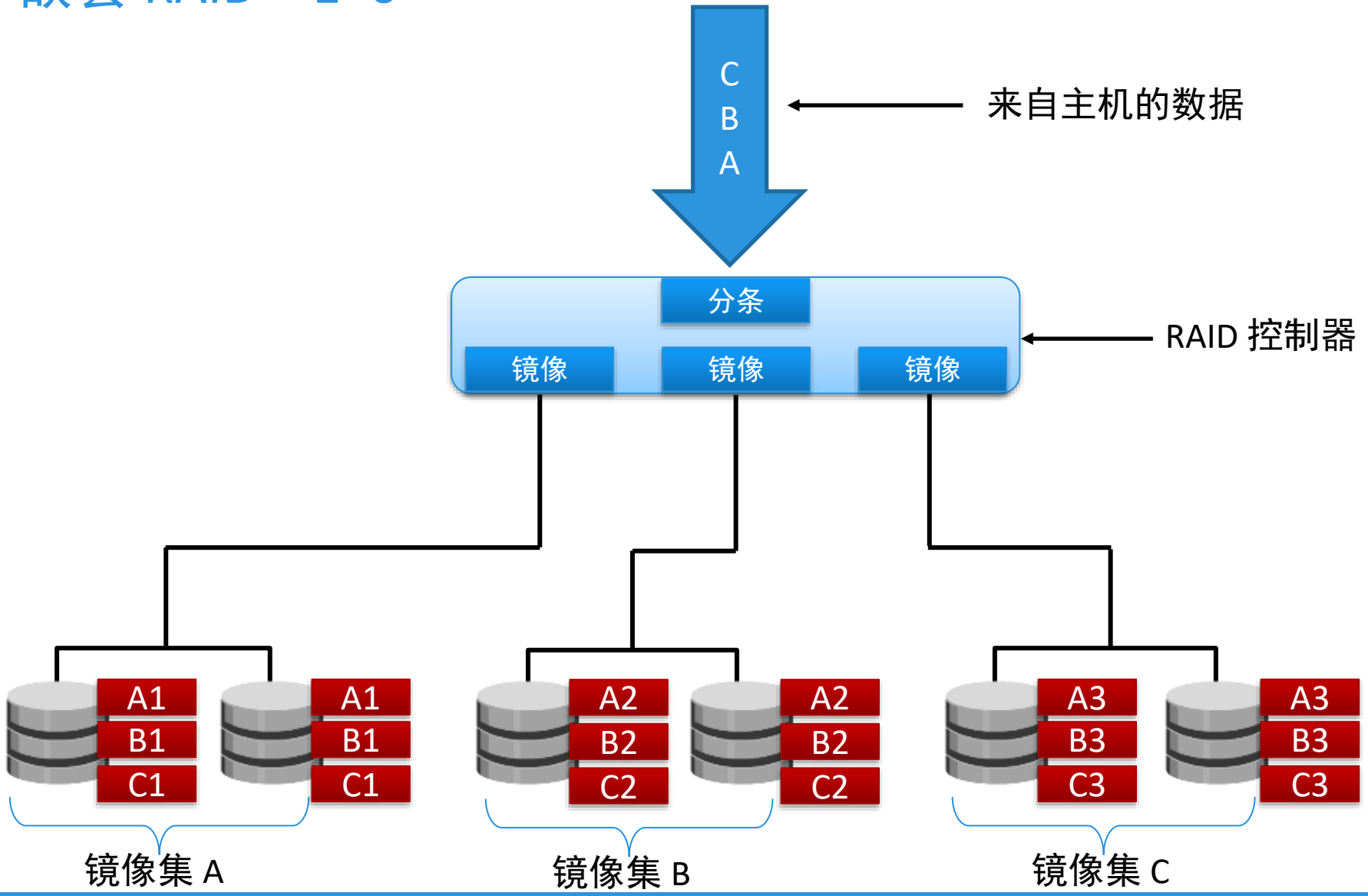
RAID 0+1 –
由条带集（RAID0 Array）组成的镜像集（RAID1Array）



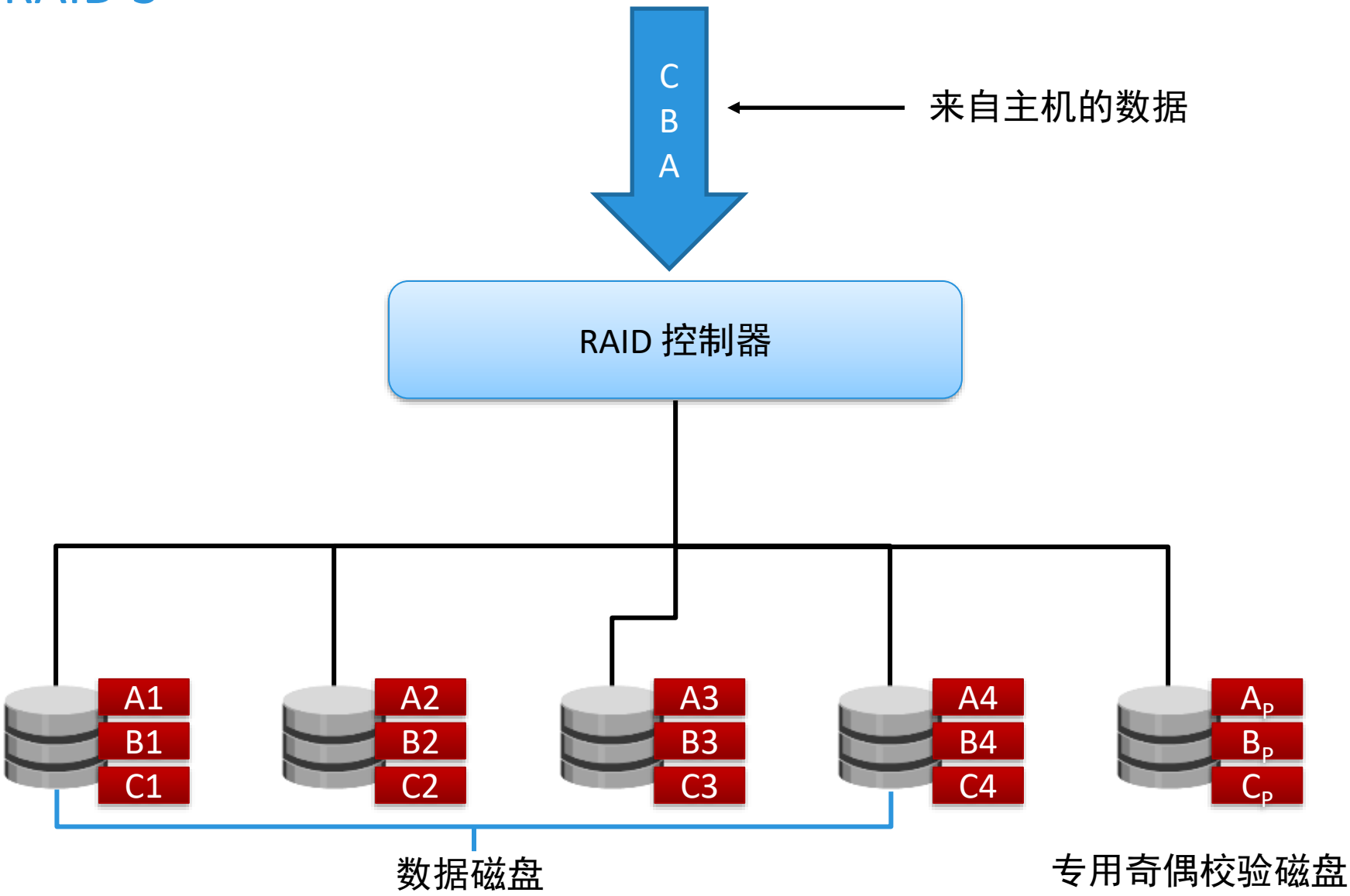
RAID 1+0 –
由镜像集（RAID 1 Array）组成的条带集（striped array ）



嵌套 RAID – 1+0

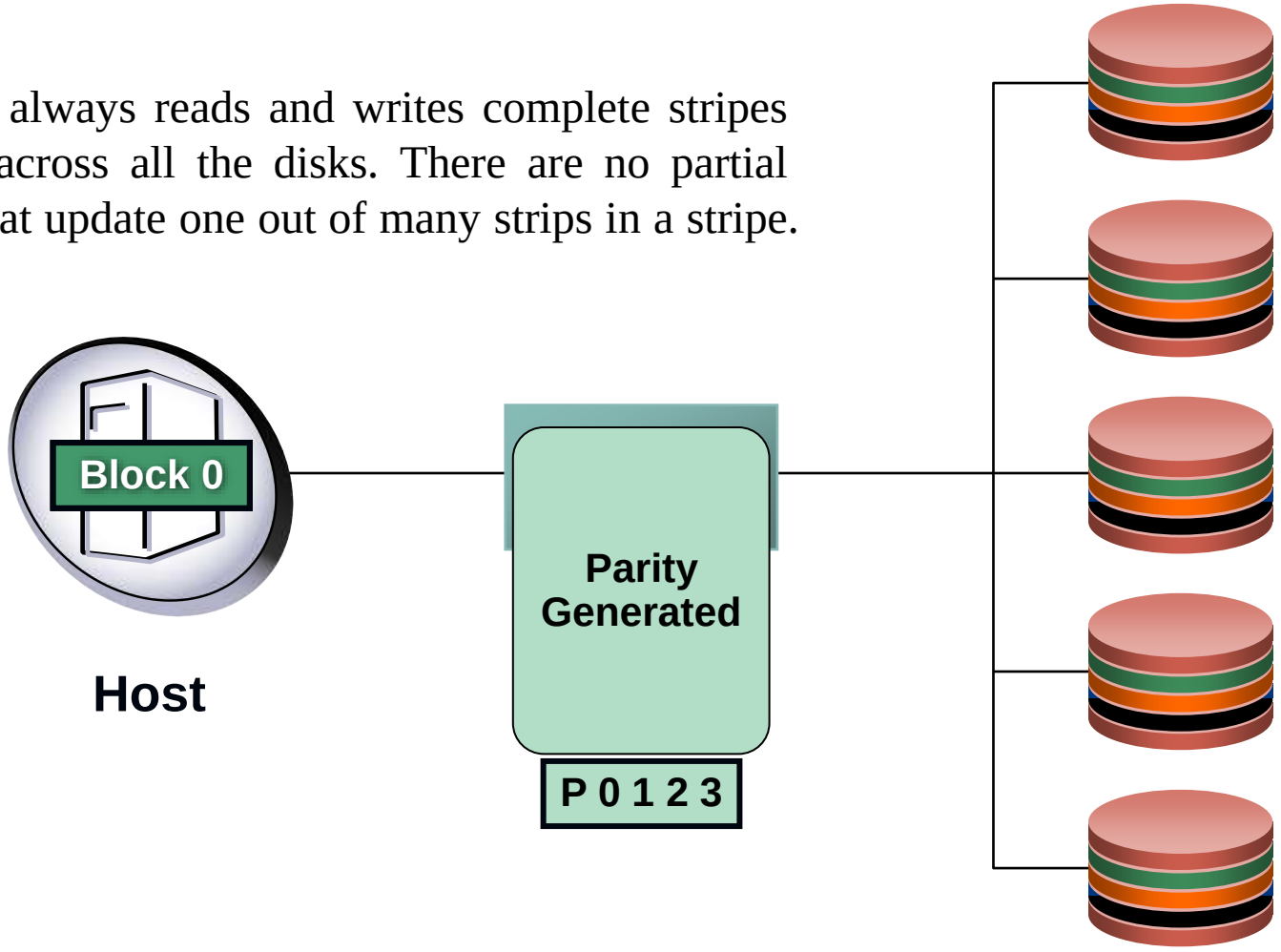

C
B
A
来自主机的数据
分条
镜像
镜像
镜像
RAID 控制器
A1
B1
C1
A1
B1
C1
A2
B2
C2
A2
B2
C2
A3
B3
C3
A3
B3
C3
镜像集 A
镜像集 B
镜像集 C

RAID 3


C
B
A
来自主机的数据
RAID 控制器
A1
B1
C1
A2
B2
C2
A3
B3
C3
A4
B4
C4
A_P
B_P
C_P
数据磁盘
专用奇偶校验磁盘

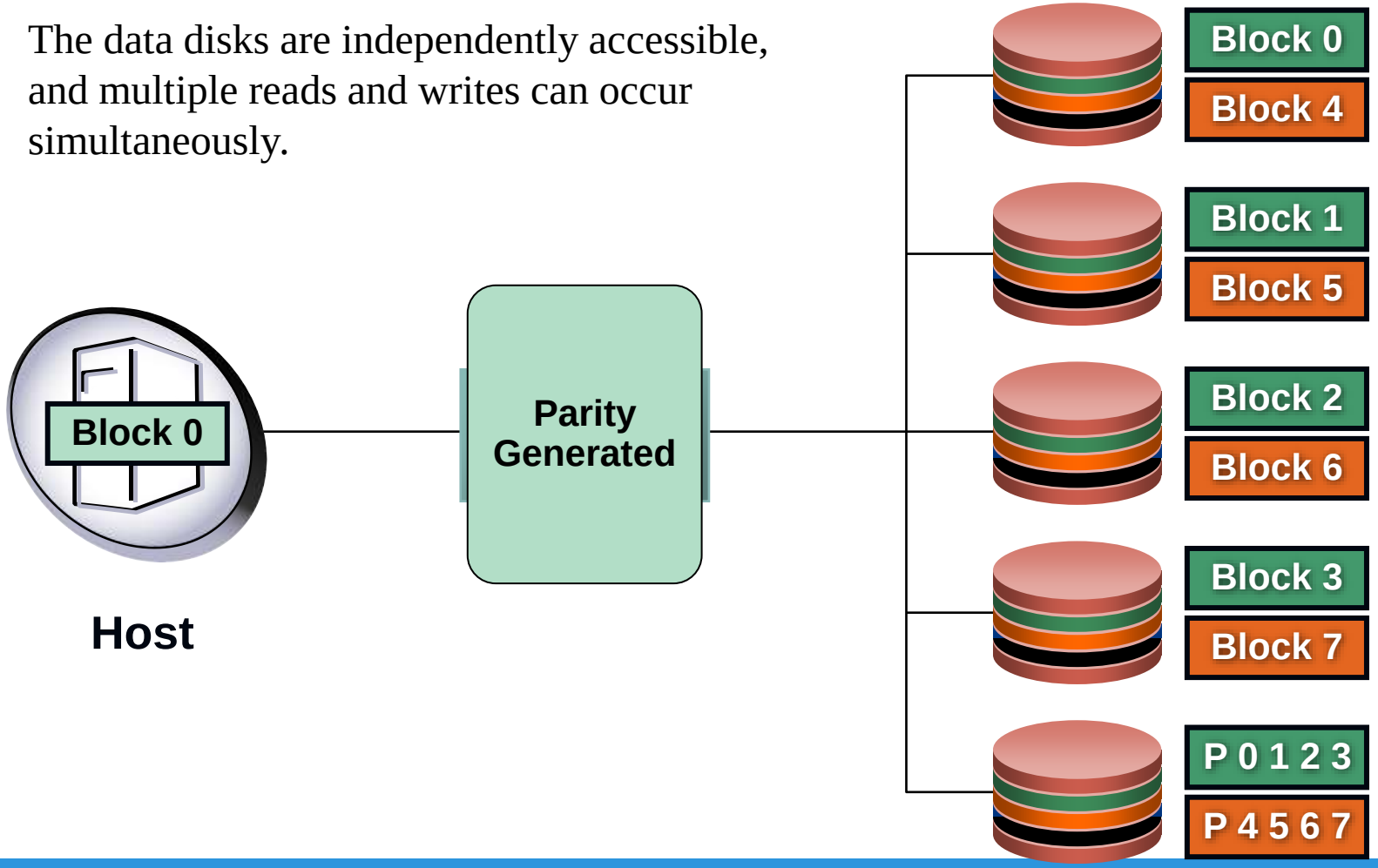
RAID 3 - Parallel Transfer with Dedicated Parity Disk

RAID 3 always reads and writes complete stripes of data across all the disks. There are no partial writes that update one out of many strips in a stripe.

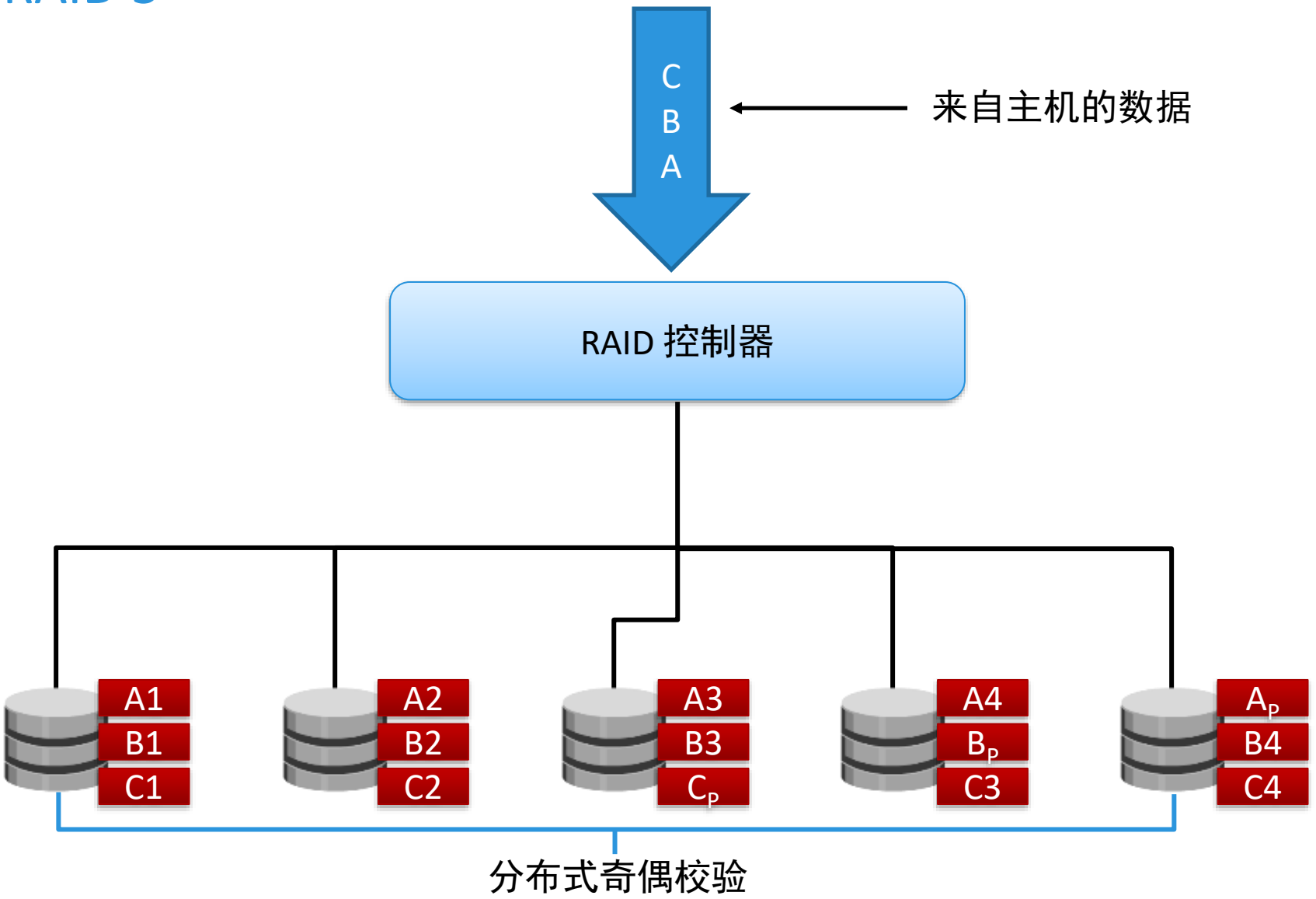


RAID 4 - Striping with Dedicated Parity Disk

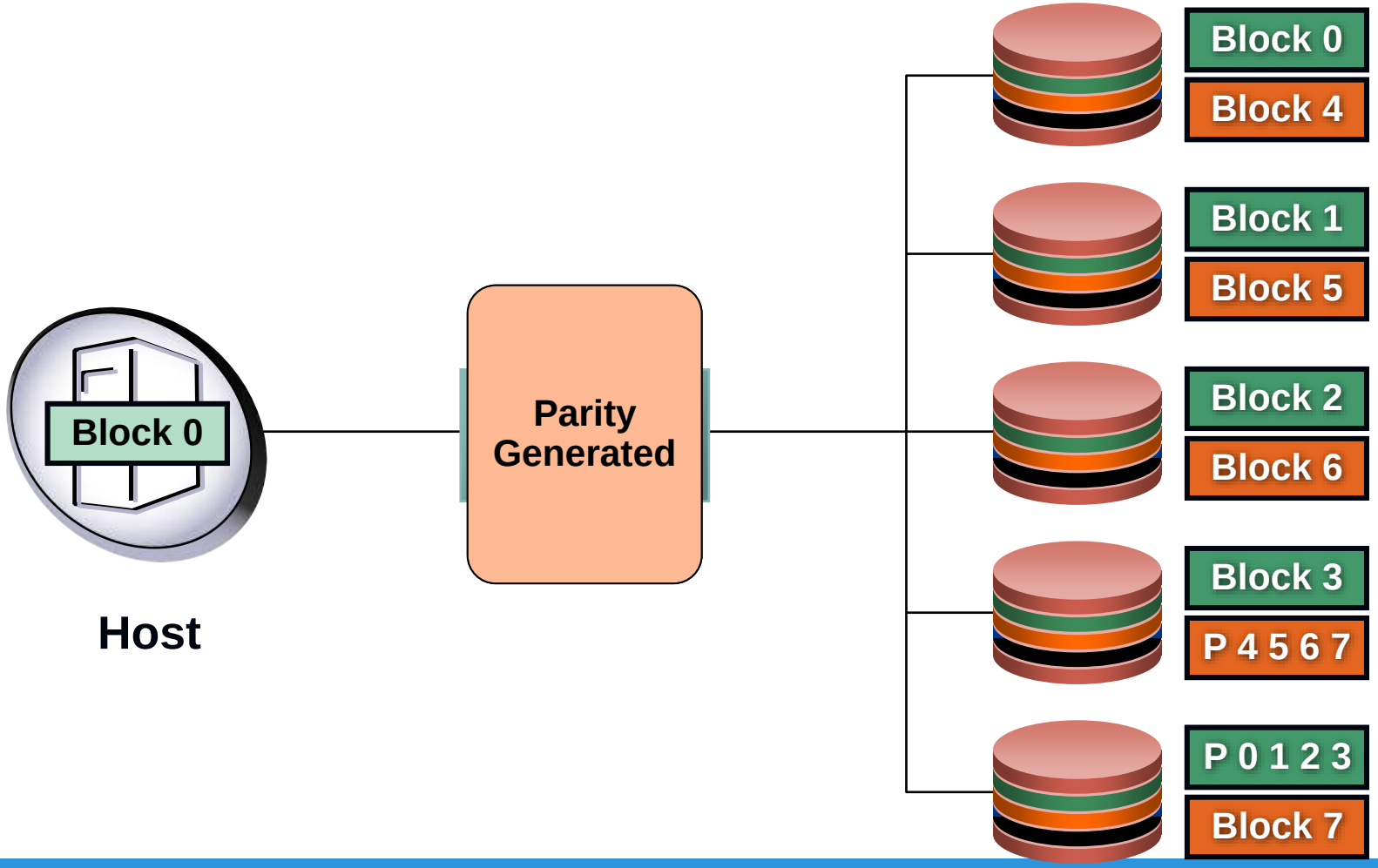
The data disks are independently accessible, and multiple reads and writes can occur simultaneously.



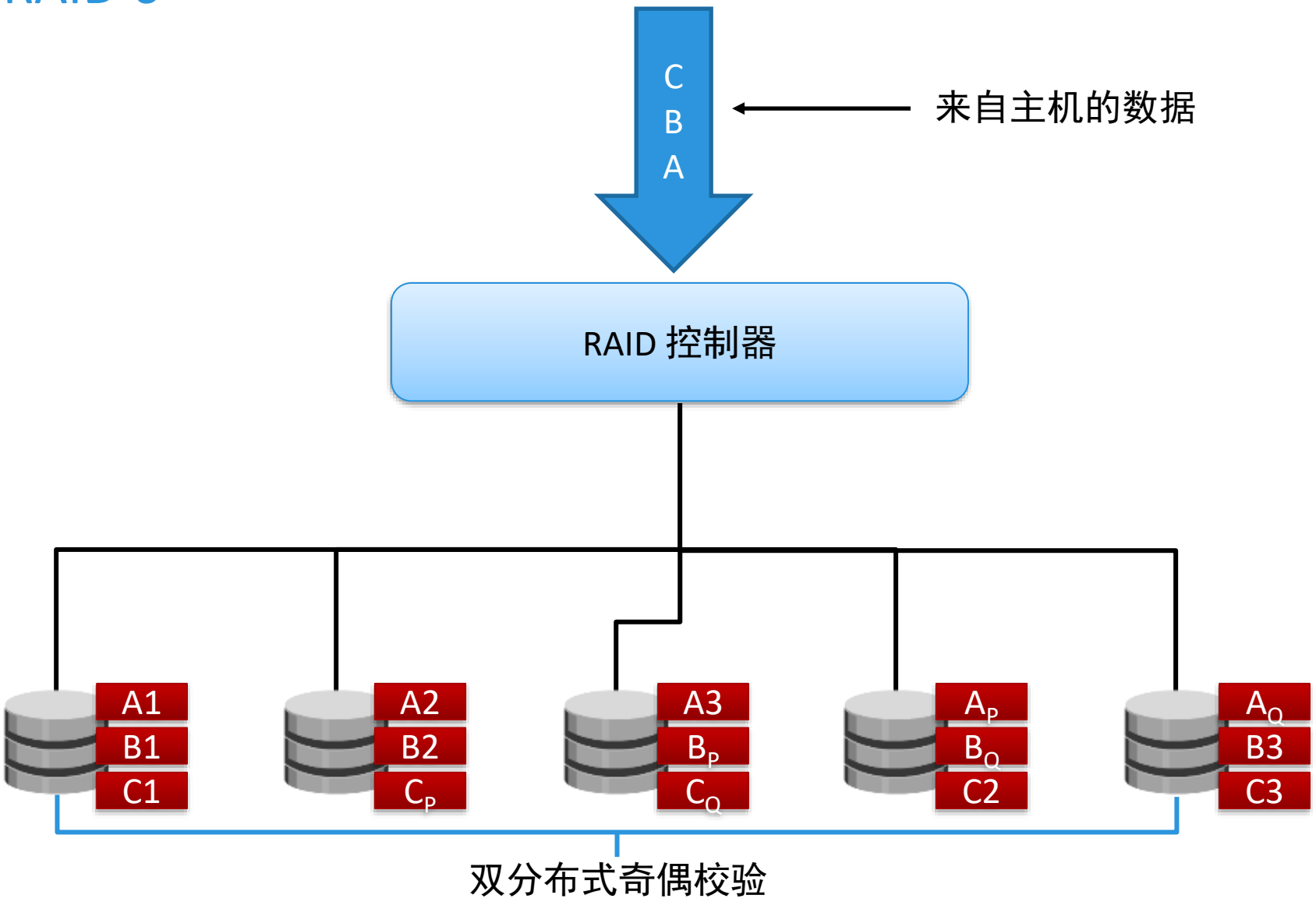
RAID 5


C
B
A
来自主机的数据
RAID 控制器
A1
B1
C1
A2
B2
C2
A3
B3
C_P
A4
B_P
C3
A_P
B4
C4
分布式奇偶校验

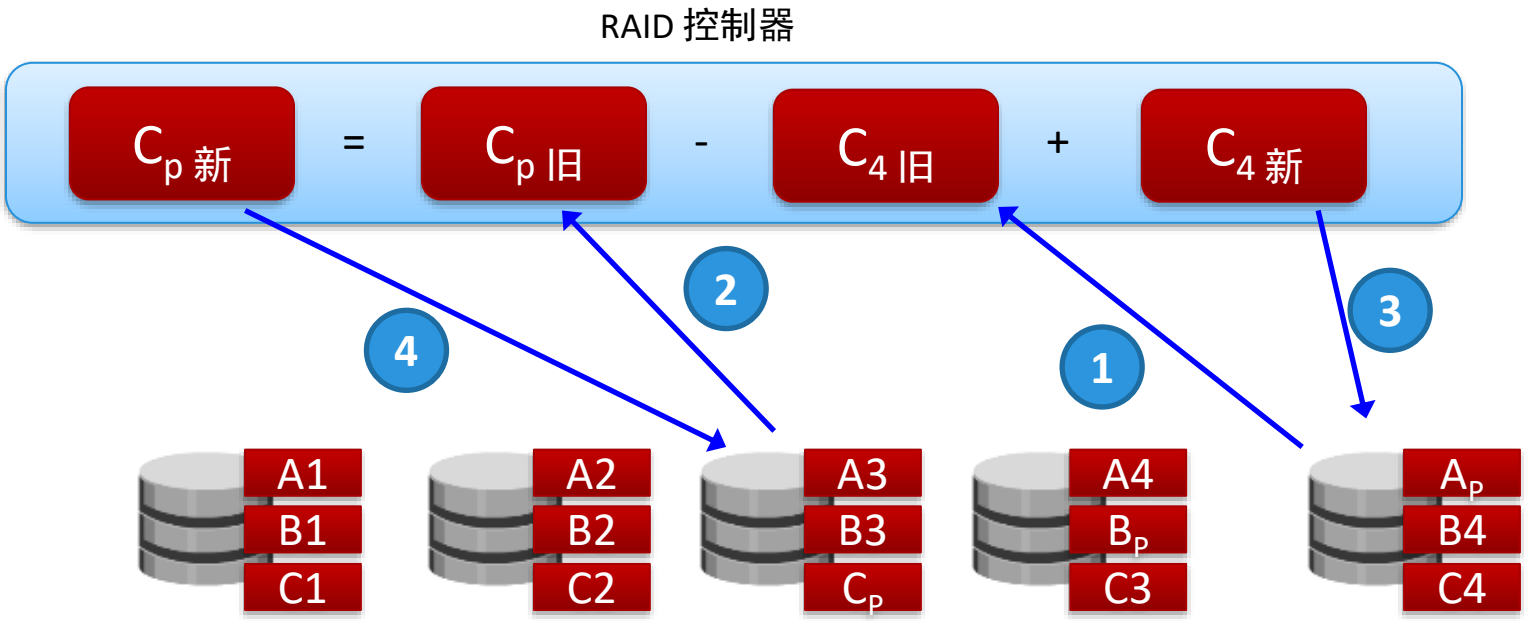
RAID 5 - Independent Disks with Distributed Parity



RAID 6


C
B
A
来自主机的数据
RAID 控制器
A1
B1
C1
A2
B2
C_P
A3
B_P
C_Q
A_P
B_Q
C2
A_Q
B3
C3
双分布式奇偶校验

RAID 对性能的影响



- 在 RAID 5 中，每次磁盘写入（更新）都表现为四次 I/O 操作（2 次磁盘读取和 2 次磁盘写入）
- 在 RAID 6 中，每次磁盘写入（更新）都表现为六次 I/O 操作（3 次磁盘读取和 3 次磁盘写入）
- 在 RAID 1 中，每次写入都表现为两次 I/O 操作（2 次磁盘写入）

RAID 性能损失计算示例

- 高峰工作负载时的 IOPS 为 1200
- 读/写比为 2:1
- 针对以下配置计算高峰活动时的磁盘负载：
  - RAID 1/0
  - RAID 5

解决方案：RAID 性能损失

- 对于 RAID 1/0，磁盘负载（读 + 写）

= (1200 x 2/3) + (1200 x (1/3) x 2)

= 800 + 800

= 1600 IOPS

- 对于 RAID 5，磁盘负载（读 + 写）

= (1200 x 2/3) + (1200 x (1/3) x 4)

= 800 + 1600

= 2400 IOPS

RAID 比较

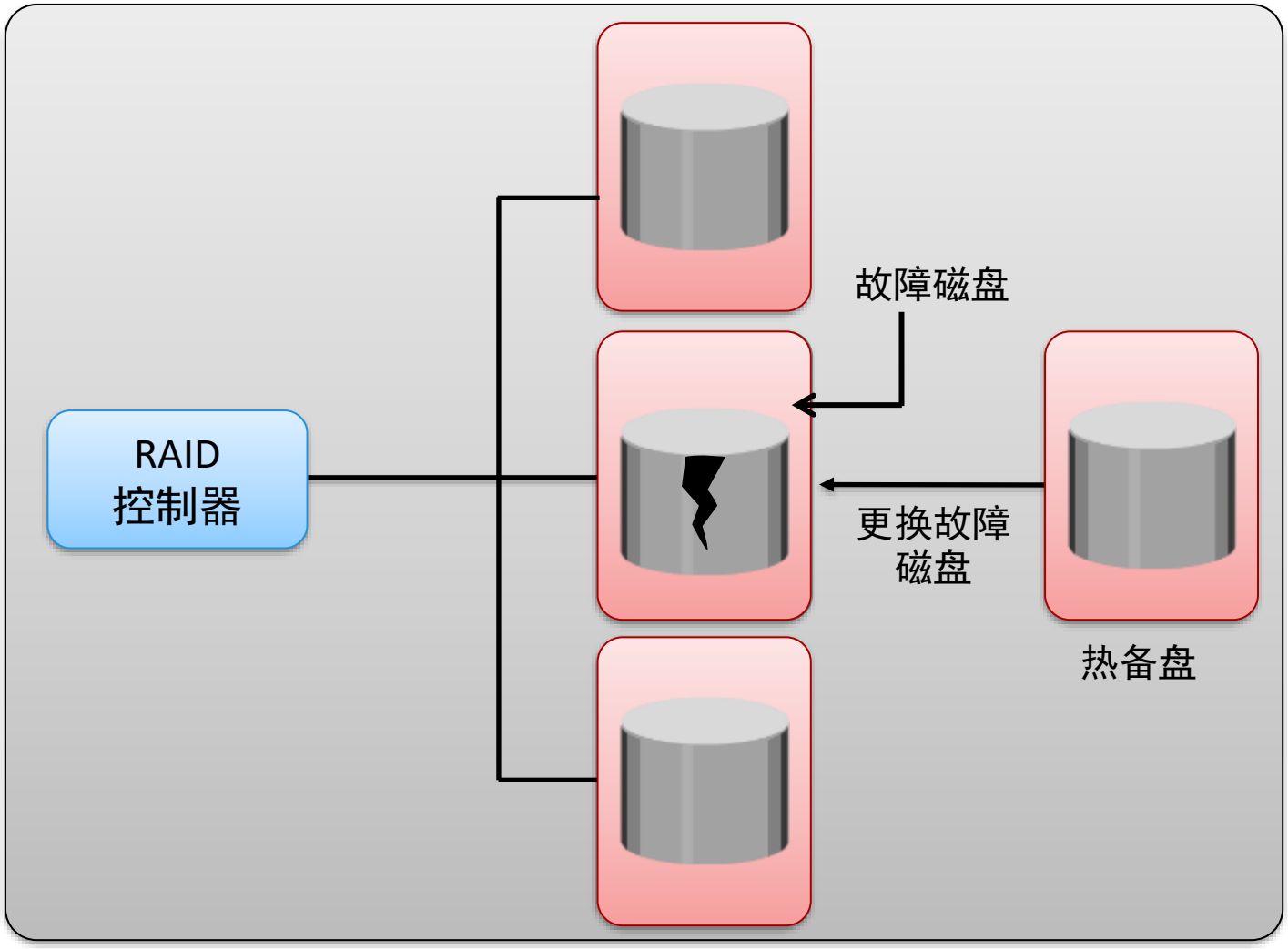
RAID 级别	最少磁盘数	可用存储容量 (%)	读取性能	写入性能	写性能损失	保护
1	2	50	优于单个磁盘	比单个磁盘低，因为必须将每次写入提交至所有磁盘	中等	镜像
1+0	4	50	良好	良好	中等	镜像
3	3	[(n-1)/n]*100	一般（对于随机读取），良好（对于顺序读取）	差到一般（对于小型随机写入） 一般（对于大型顺序写入）	高	奇偶校验（支持单磁盘故障）
5	3	[(n-1)/n]*100	良好（对于随机和顺序读取）	一般（对于随机和顺序写入）	高	奇偶校验（支持单磁盘故障）
6	4	[(n-2)/n]*100	良好（对于随机和顺序读取）	差到一般（对于随机和顺序写入）	非常高	奇偶校验（支持两个磁盘故障）

其中，n = 磁盘数

适用于不同应用程序的 RAID 级别

- RAID 1+0
  - 适合使用小型、随机和写入密集型（写入量通常大于 30%）I/O 配置文件的应用程序
  - 示例：OLTP、RDBMS – 临时空间
- RAID 3
  - 大型、顺序读取和写入
  - 示例：数据备份和多数据流
- RAID 5 and 6
  - 小型、随机工作负载（写入量通常小于 30%）
  - 示例：电子邮件、RDBMS – 数据输入

热备盘



模块 3：总结

本模块涵盖以下要点：

- RAID 实现方法和技术
- 常用 RAID 级别
- RAID 写性能损失
- 根据 RAID 级别的成本和性能比较各个级别

知识测验 – 1

- 关于软件 RAID 实现，以下哪项描述是正确的？
  - A. 操作系统升级不需要验证与 RAID 软件的兼容性
  - B. 其成本高于硬件 RAID 实现
  - C. 支持所有 RAID 级别
  - D. 使用主机 CPU 周期执行 RAID 计算
- 一个应用程序生成 400 个小型随机 IOPS，读写比为 3:1。用于 RAID 5 的磁盘上 RAID 更正的 IOPS 是多少？
  - A. 400
  - B. 500
  - C. 700
  - D. 900

知识测验 – 2

- 用于小型随机 I/O 的 RAID 6 配置中的写性能损失是多少？
  - A. 2
  - B. 3
  - C. 4
  - D. 6
- 以下哪个应用程序可通过使用 RAID 3 获得最大效益？
  - A. 备份
  - B. OLTP
  - C. 电子商务
  - D. 电子邮件

知识测验 – 3

- 一个具有 64 KB 条块大小且包含五个磁盘的奇偶校验 RAID 5 集的条带大小是多少？
  - A. 64 KB
  - B. 128 KB
  - C. 256 KB
  - D. 320 KB

练习 1：RAID

- 某公司计划为其财务应用程序重新配置存储以获得高可用性
  - 当前配置和挑战
    - 应用程序执行 15% 随机写入和 85% 随机读取
    - 当前与包含五个磁盘的 RAID 0 配置一起部署
    - 每个磁盘的已公布格式化容量为 200 GB
    - 财务应用程序数据的总大小为 730 GB，并且可能在未来 6 个月内不会发生更改
    - 财务年度已接近尾声，即使购买一个磁盘也是不可能的
- 任务
  - 为该公司推荐一个可用于重新构造其环境以满足其需要的 RAID 级别
  - 根据成本、性能和可用性来论证您的选择

练习 2：RAID

- 某公司（与练习 1 中所述相同）现计划为其数据库应用程序重新配置存储以获得高可用性
  - 当前配置和挑战
    - 应用程序执行 40% 写入和 60% 读取
    - 当前已部署在包含六个磁盘的 RAID 0 配置中，且每个磁盘已公布的容量为 200 GB
    - 数据库大小为 900 GB 且数据量可能在未来 6 个月内更改 30%
    - 现在是新财年的开始且公司的预算有所增长
- 任务
  - 推荐适合的 RAID 级别以满足公司的需要
  - 估计新解决方案的成本（200 GB 磁盘成本为 $1000）
  - 根据成本、性能和可用性来论证您的选择

模块－4

智能存储系统

模块 4：智能存储系统

学完本模块后，您将能够：

- 介绍智能存储系统的关键组件
- 介绍缓存管理和保护技术
- 介绍两个存储资源调配方法
- 介绍两种类型的智能存储系统

模块 4：智能存储系统

第 1 课：智能存储系统的关键组件

本课程将讲述下列主题：

- 智能存储系统概述
- 智能存储系统的关键组件
- 缓存管理

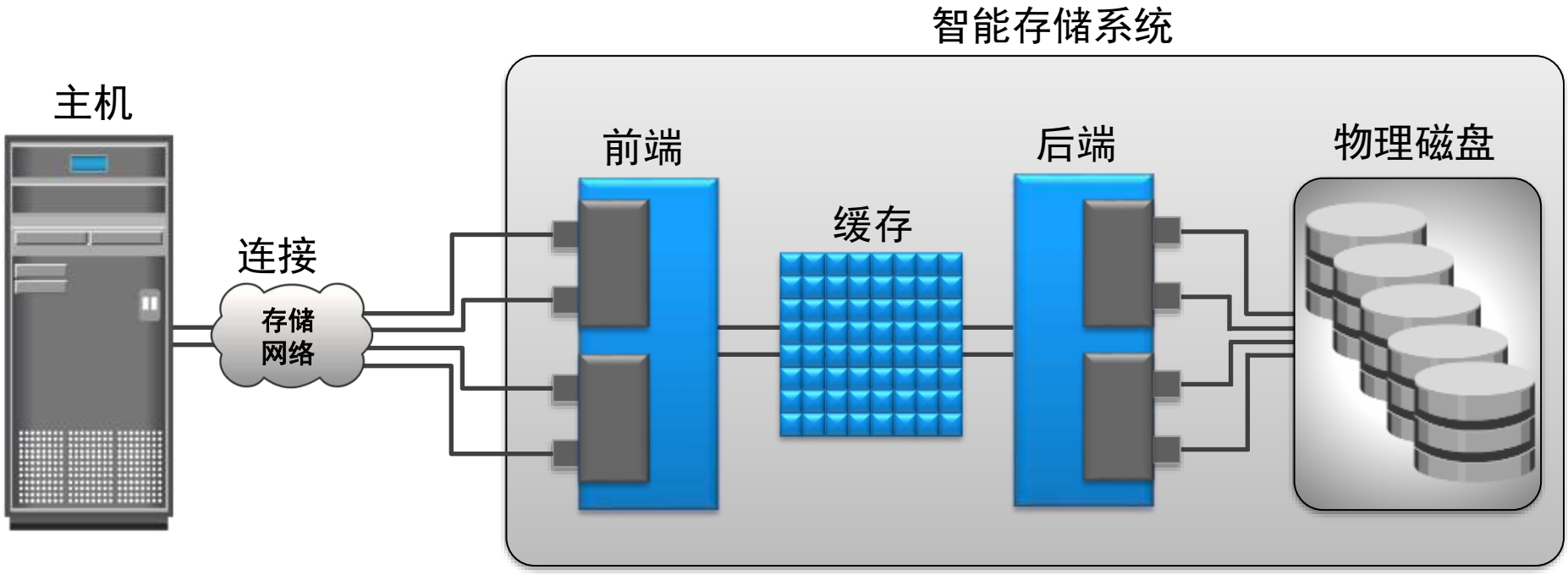
什么是智能存储系统 (ISS)?

智能存储系统

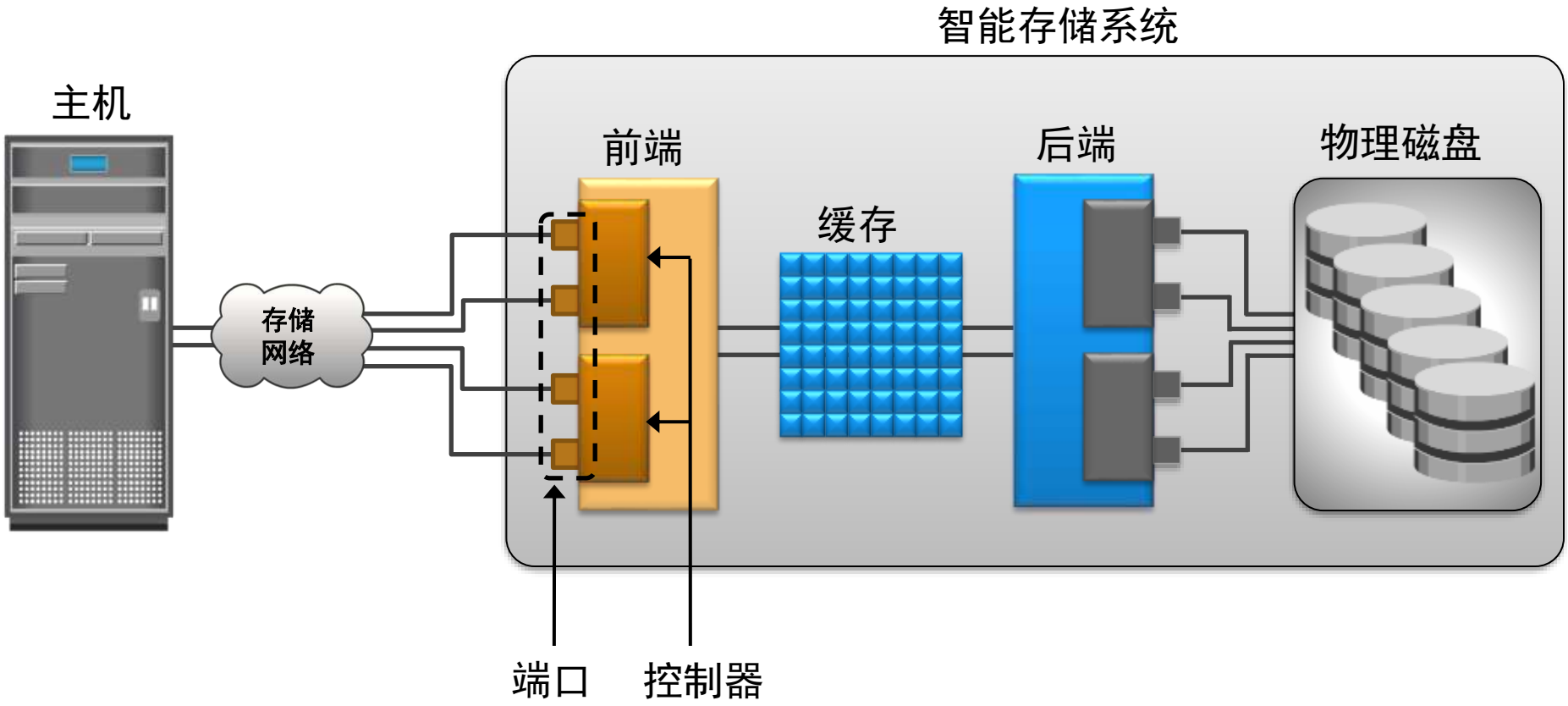
它是功能丰富的 RAID 阵列，提供高度优化的 I/O 处理功能。

- 提供可增强性能的大量缓存和多条 I/O 路径
- 具有提供以下功能的操作环境
  - 智能缓存管理
  - 阵列资源管理
  - 到异构主机的连接
- 支持闪存驱动器、虚拟资源调配和自动存储分层

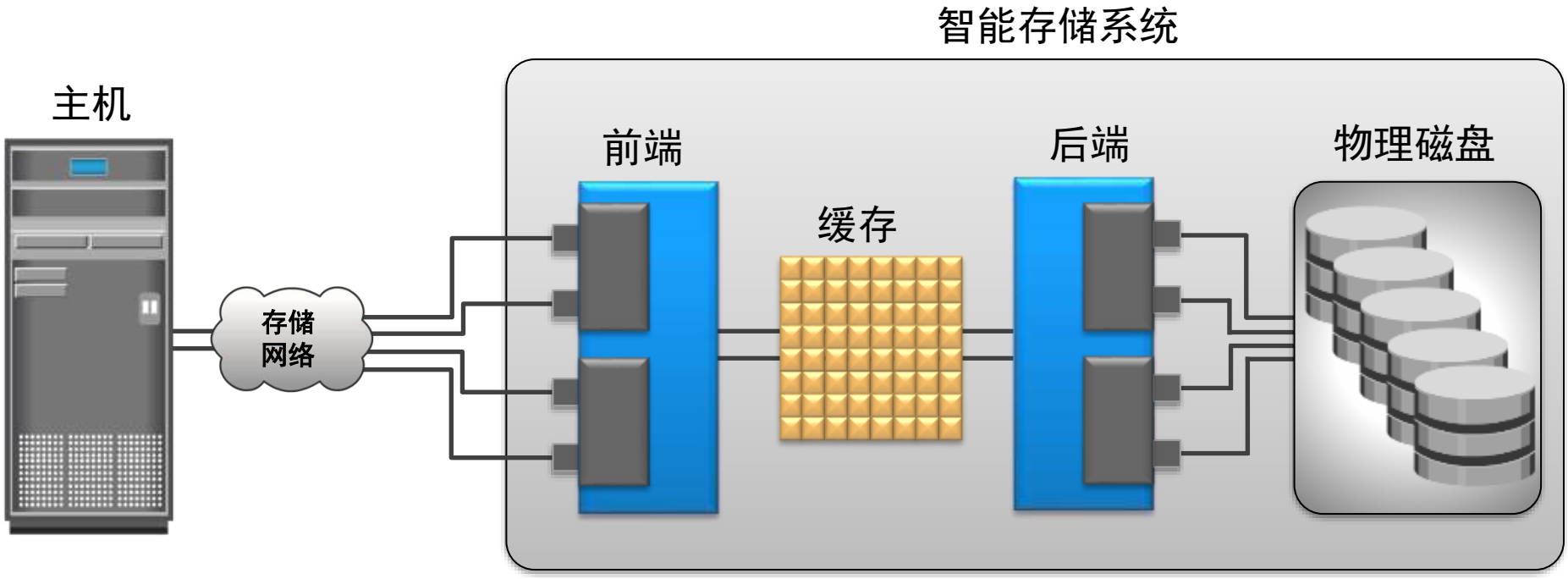
ISS 的关键组件



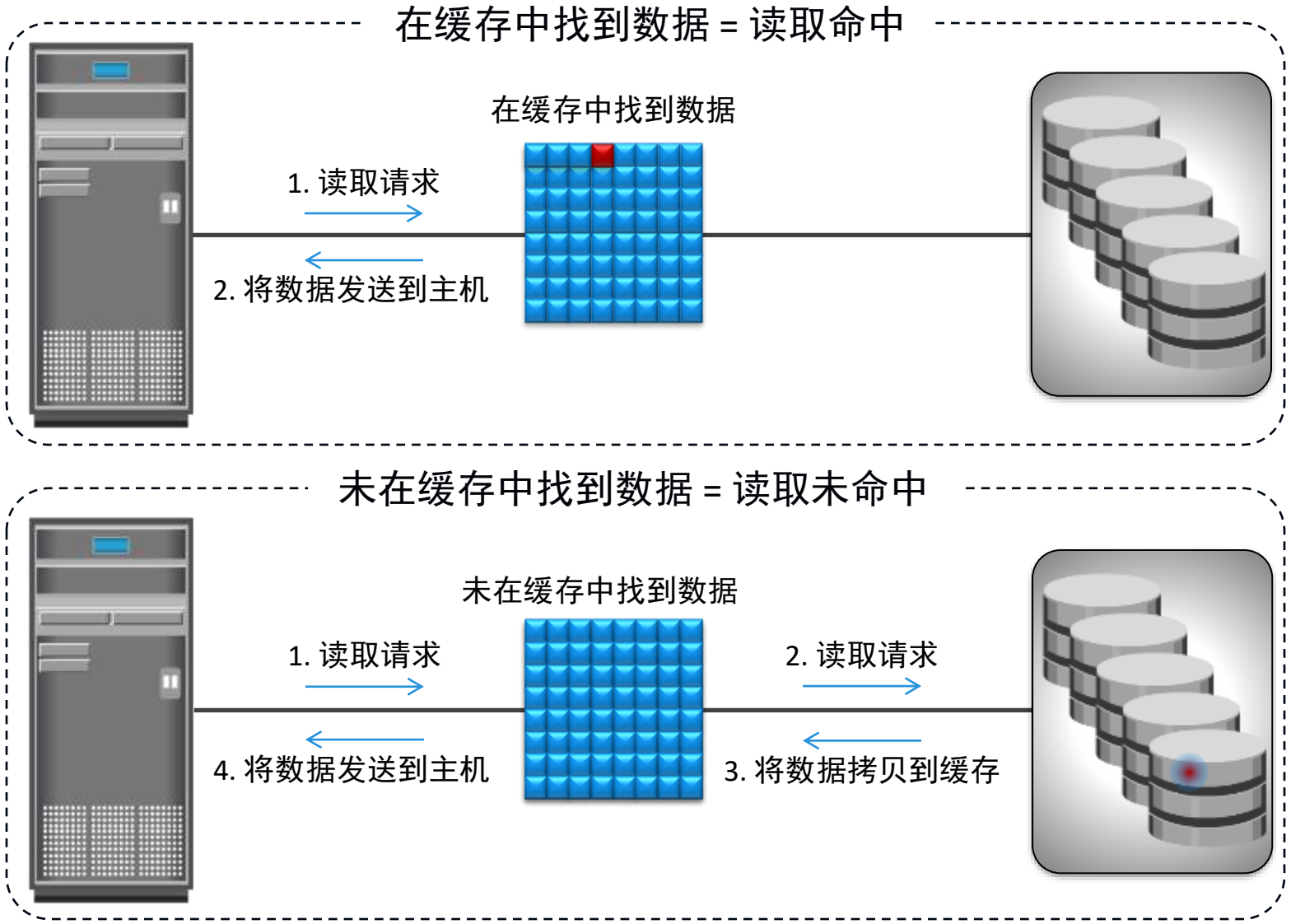
ISS 的关键组件：前端



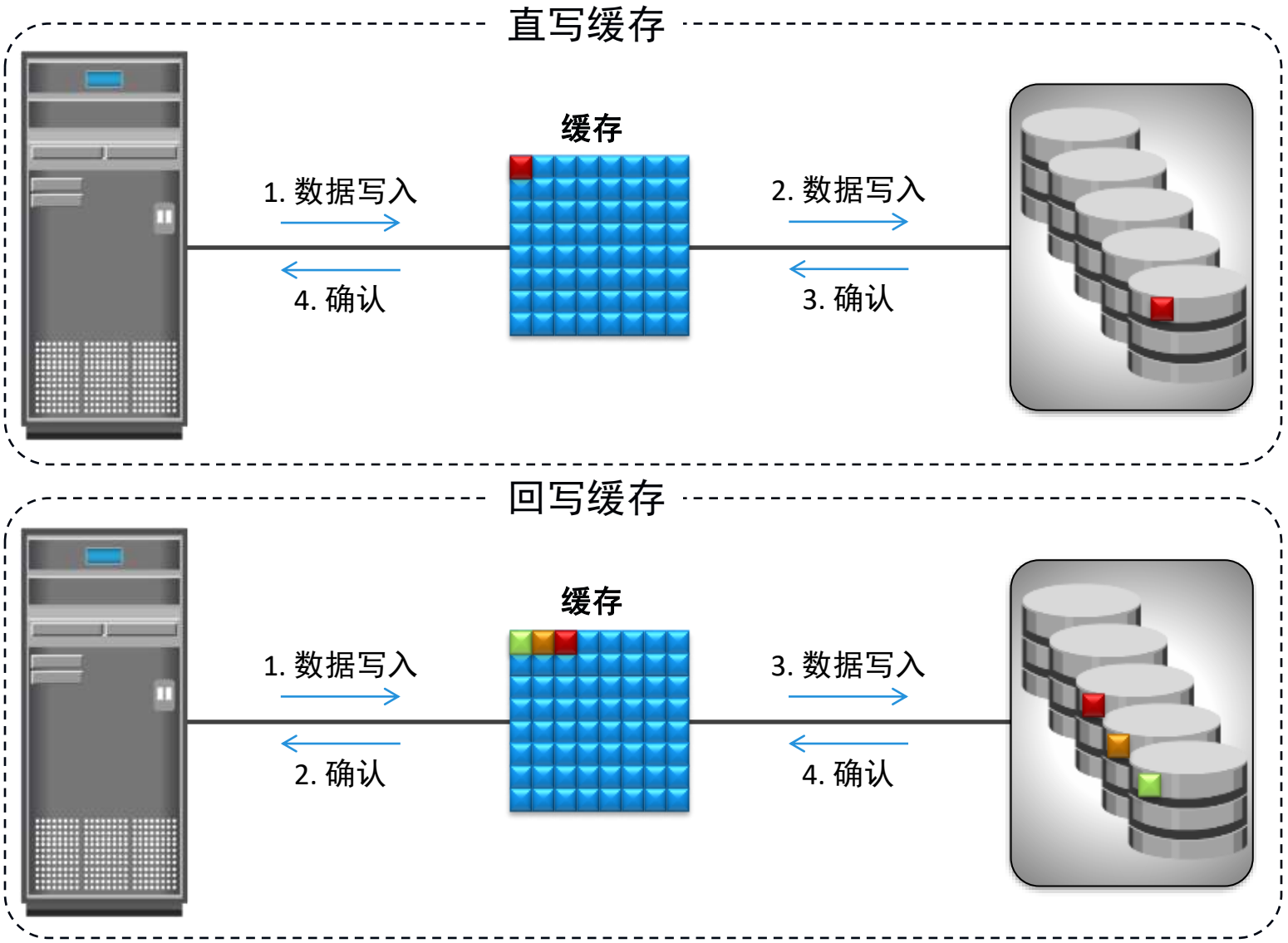
ISS 的关键组件：缓存



使用缓存进行的读取操作

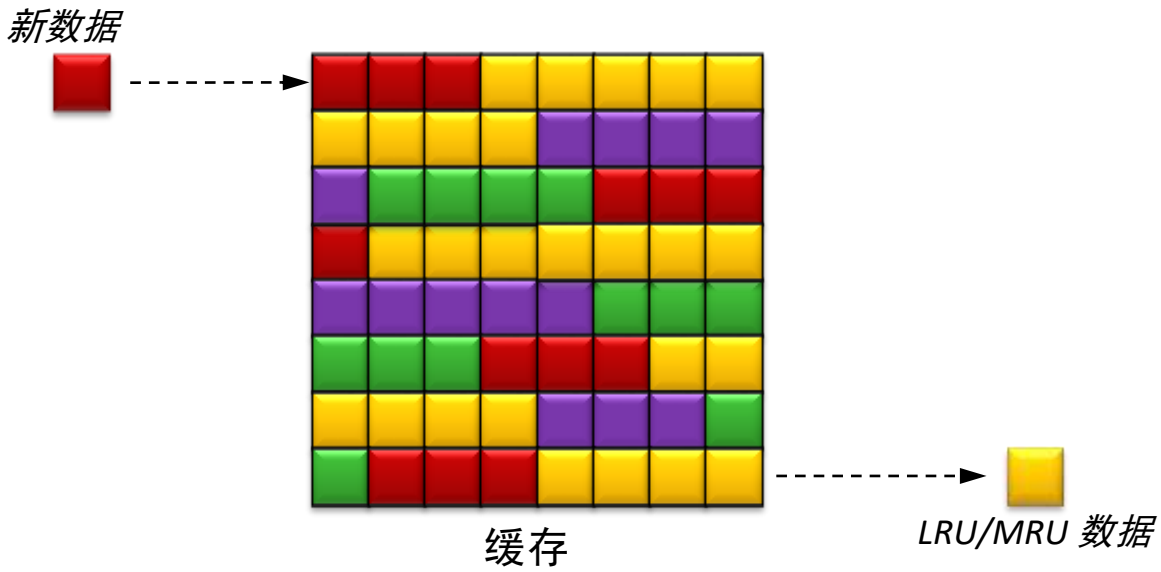


使用缓存进行的写入操作



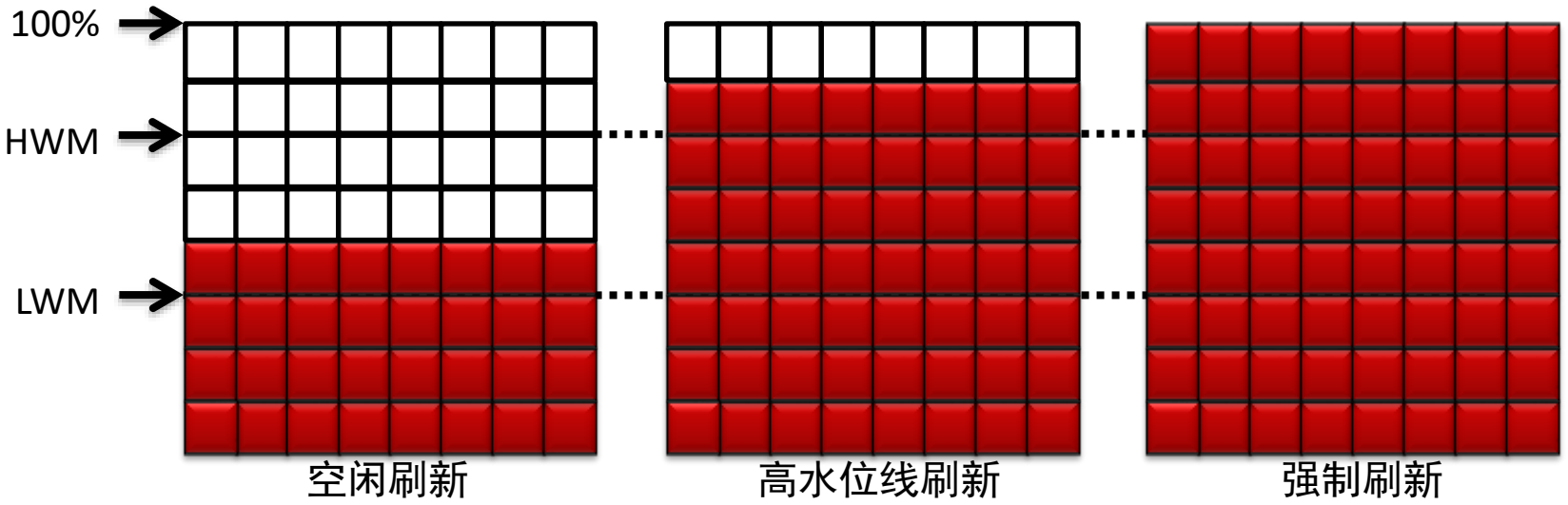
缓存管理：算法

- 最近最少使用 (LRU)
  - 删除很长时间未访问的数据
- 最近最常使用 (MRU)
  - 删除最近最常访问的数据



缓存管理：水位线

- 通过刷新过程管理突发 I/O
  - 刷新是将缓存中的数据提交到磁盘的过程
- 管理缓存利用率的三个刷新模式是：
  - 空闲刷新
  - 高水位线刷新
  - 强制刷新

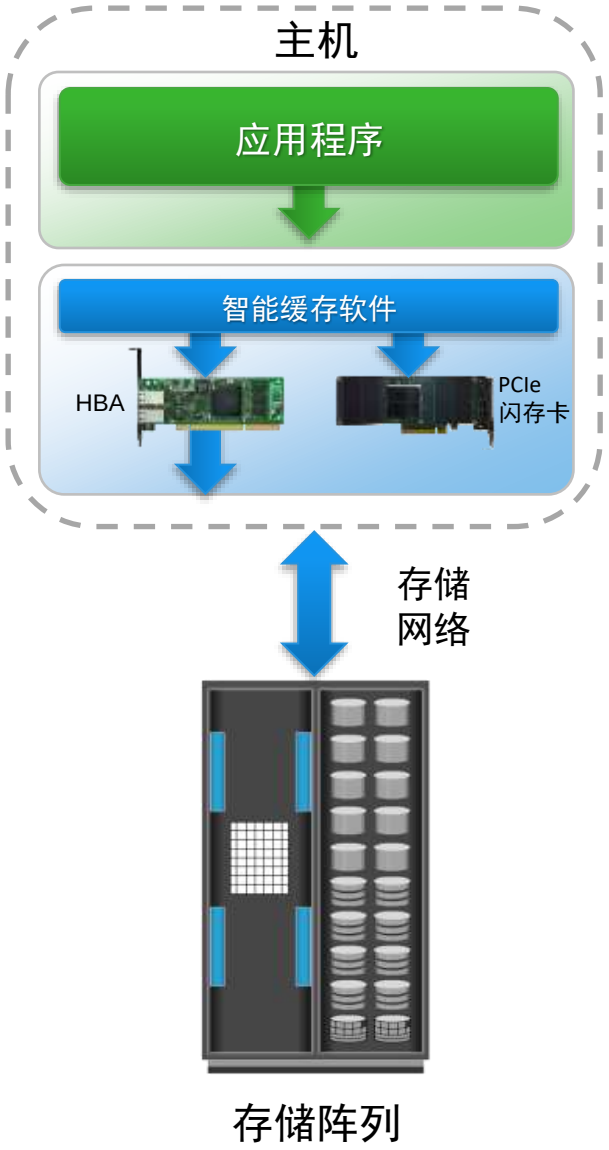


缓存数据保护

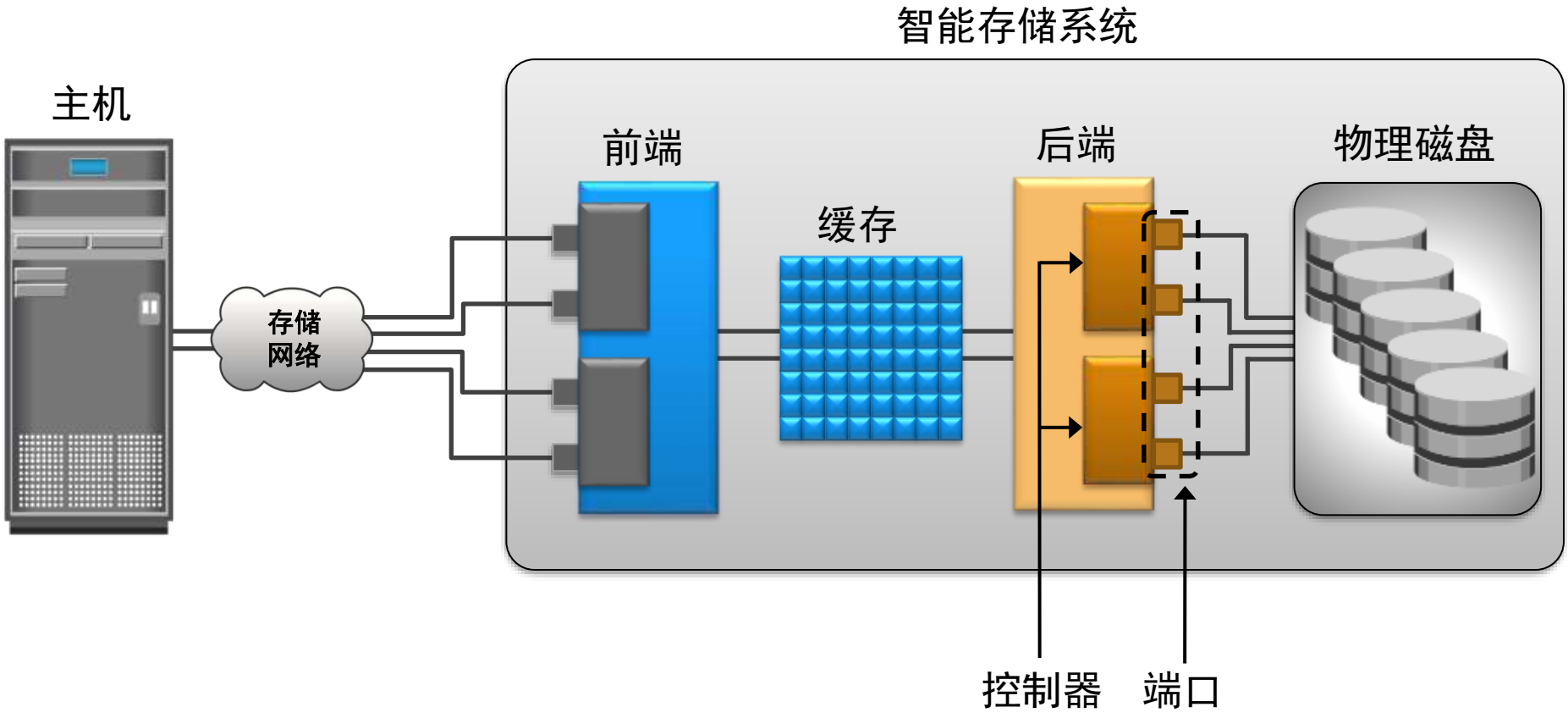
- 防止缓存中的数据受到电源或缓存故障的影响：
  - 缓存镜像
    - 提供防止数据受到缓存故障的影响的保护
    - 每次写入到缓存中的数据都保存在两个独立内存卡中的两个不同内存位置
  - 缓存保险存储
    - 提供防止数据受到电源故障的影响的保护
    - 在出现电源故障时，会将未提交的数据转储到称作“保险存储驱动器”的一组专用驱动器中

服务器闪存缓存技术

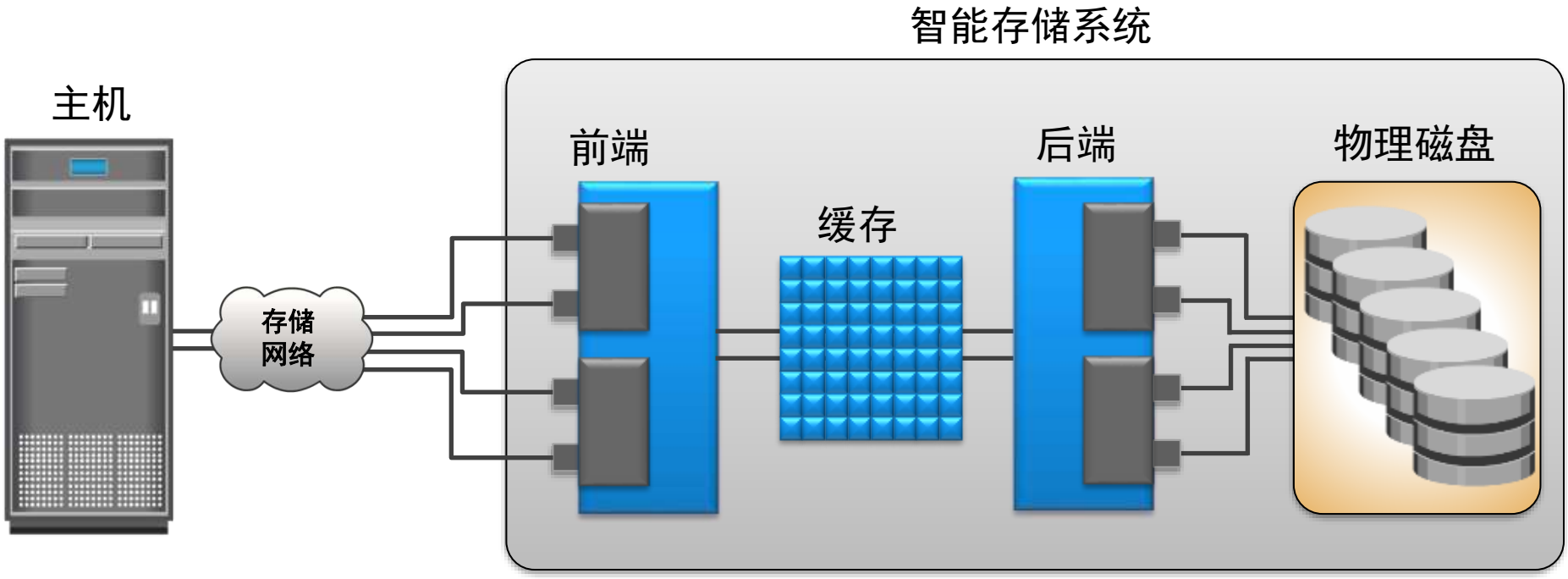
- 使用主机上的智能缓存软件和 PCIe 闪存卡
- 显著提高应用程序性能
  - 为读取密集型工作负载提供性能加速
  - 避免与对存储阵列的 I/O 访问关联的网络延迟
- 通过将数据放在服务器上的 PCIe 闪存中，以智能方式确定将受益的数据
- 使用最少的 CPU 和内存资源
  - 闪存管理减负到 PCIe 卡上



ISS 的关键组件：后端



ISS 的关键组件：物理磁盘



模块 4：智能存储系统

第 2 课：存储资源调配和 ISS 实施

本课程将讲述下列主题：

- 传统存储资源调配
- 虚拟存储资源调配
- ISS 实施

向主机分配存储

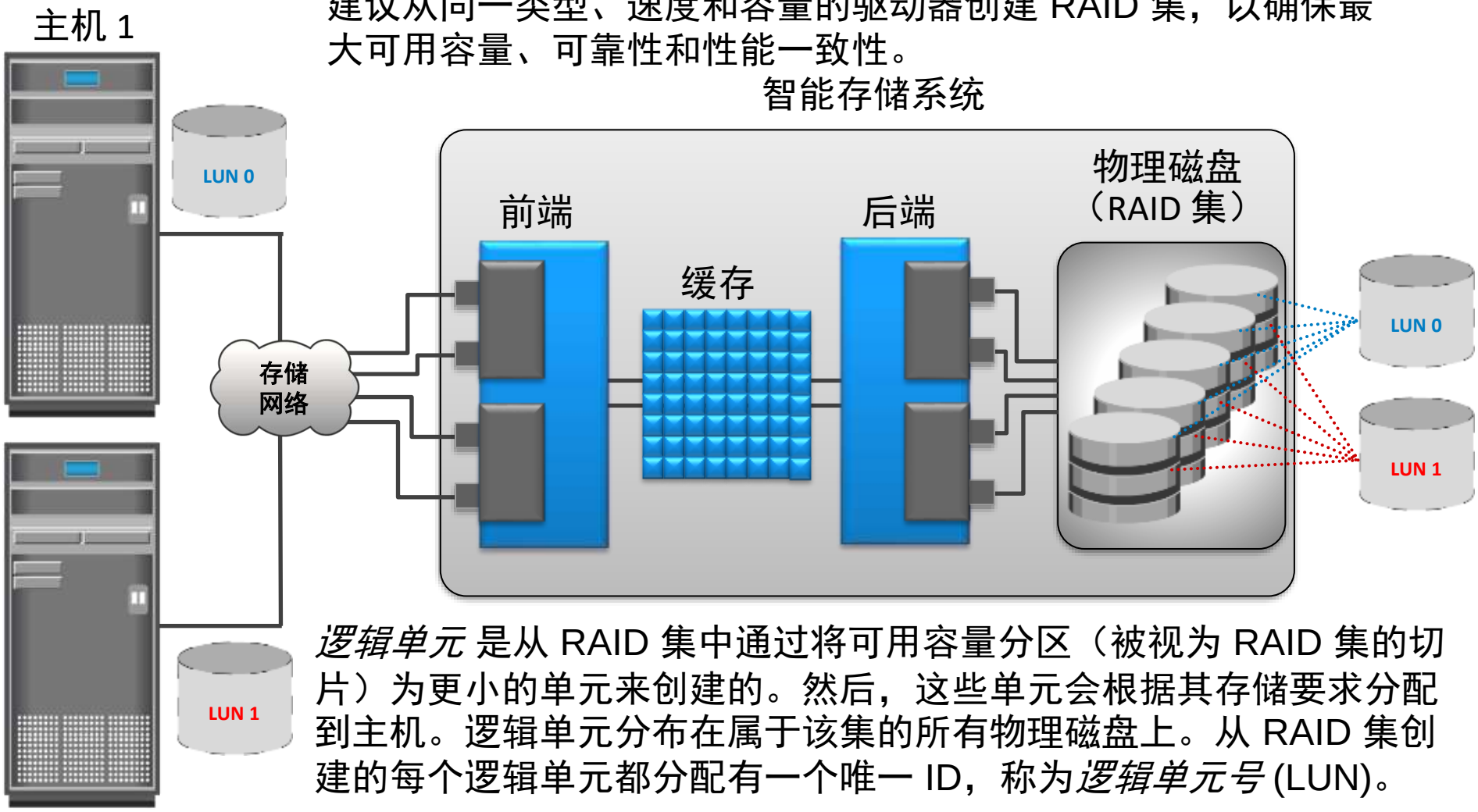
存储资源调配

它是根据主机上运行的应用程序的容量、可用性和性能要求向主机分配存储资源的过程。

- 可以两种方式执行：
  - 传统存储资源调配
  - 虚拟存储资源调配

传统存储资源调配

建议从同一类型、速度和容量的驱动器创建 RAID 集，以确保最大可用容量、可靠性和性能一致性。



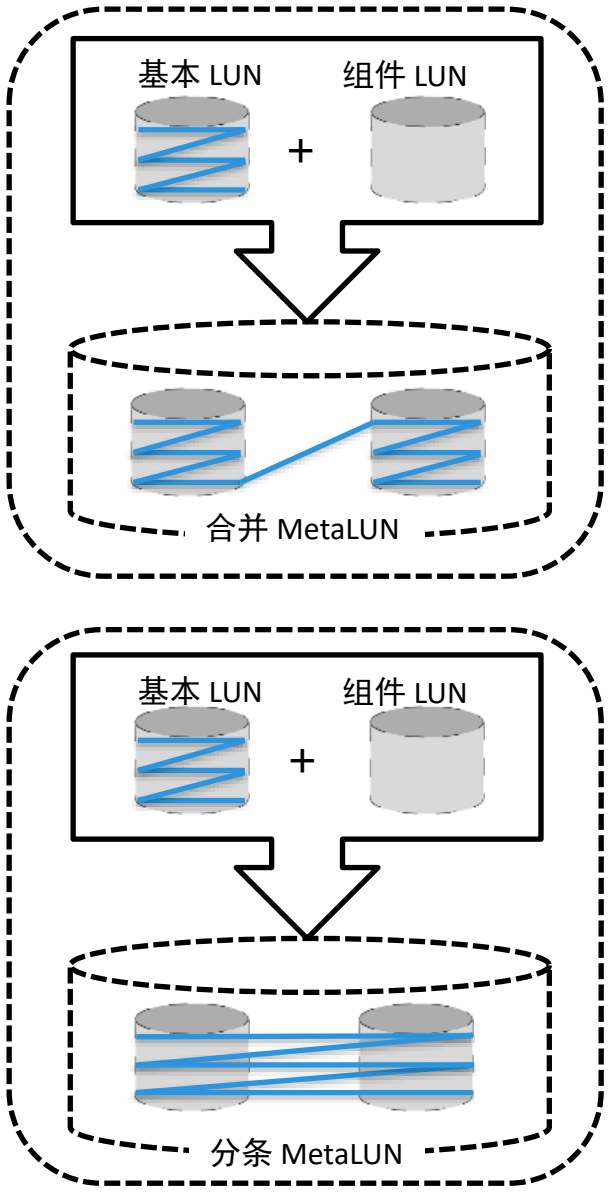
逻辑单元 是从 RAID 集中通过将可用容量分区（被视为 RAID 集的切片）为更小的单元来创建的。然后，这些单元会根据其存储要求分配到主机。逻辑单元分布在属于该集的所有物理磁盘上。从 RAID 集创建的每个逻辑单元都分配有一个唯一 ID，称为*逻辑单元号* (LUN)。

LUN 扩展

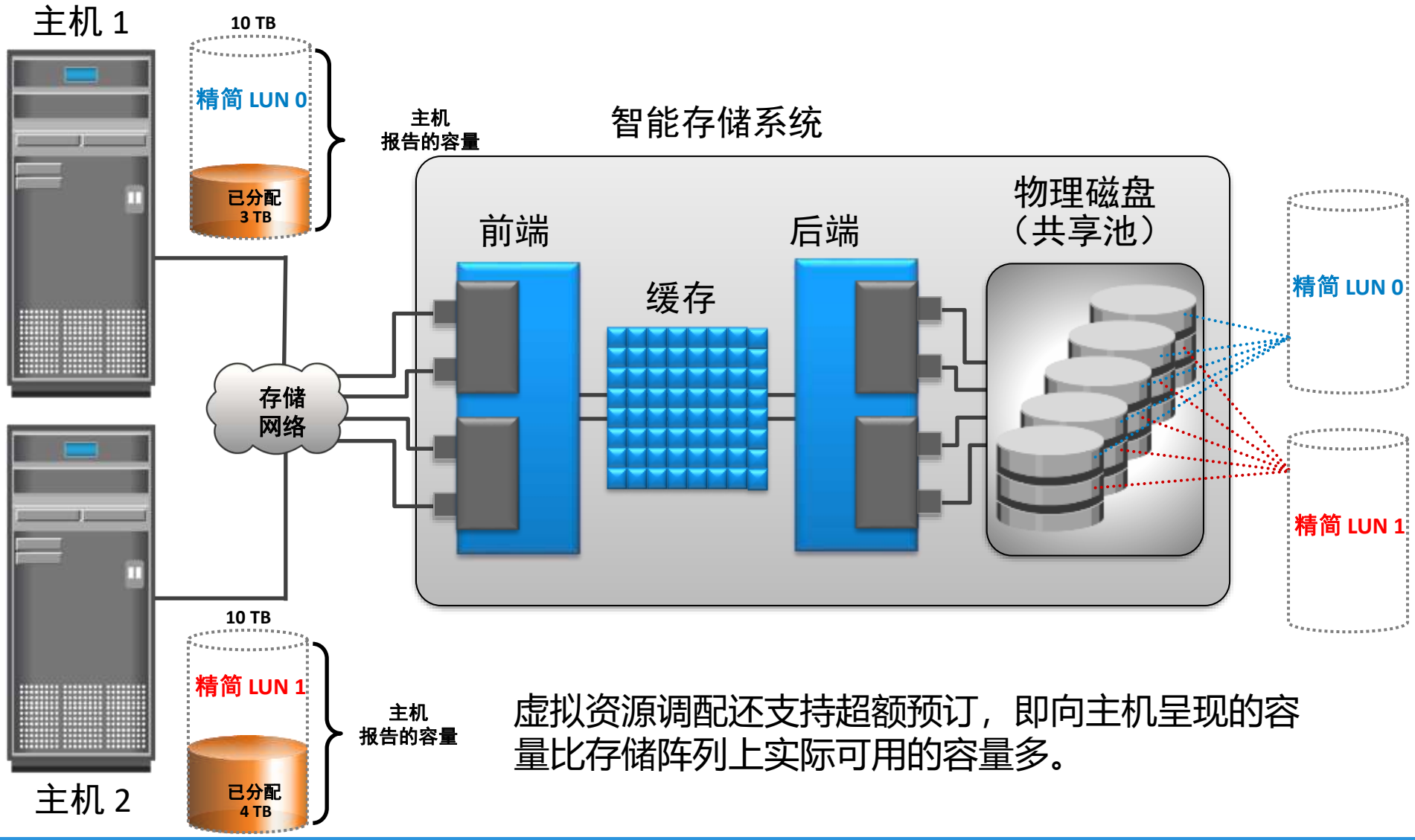
MetaLUN

它是扩展需要附加容量或性能的 LUN 的方法。

- 通过组合两个或更多个 LUN 来创建
- MetaLUN 可以是合并的，也可以是分条的
- 合并 metaLUN
  - 仅提供附加容量，而不提供性能
  - 扩展很快，因为未重新条带化数据
- 分条 metaLUN
  - 提供容量和性能
  - 扩展很慢，因为会重新条带化数据

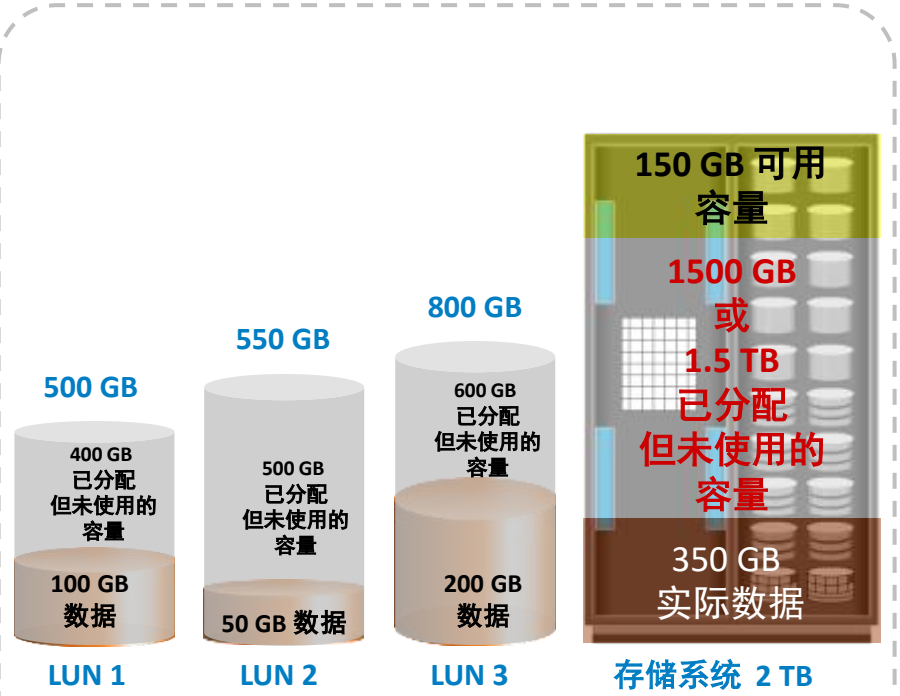


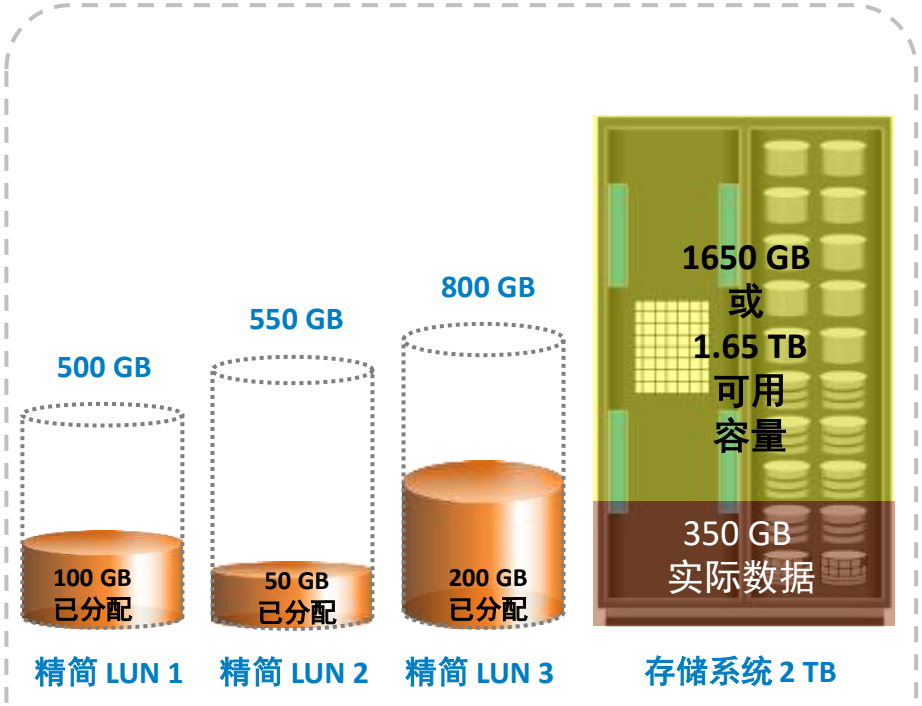
虚拟存储资源调配



虚拟资源调配还支持超额预订，即向主机呈现的容量比存储阵列上实际可用的容量多。

传统资源调配与虚拟资源调配





LUN 掩蔽

LUN 掩蔽

它是通过定义主机可以访问哪些 LUN 来进行数据访问控制的过程。

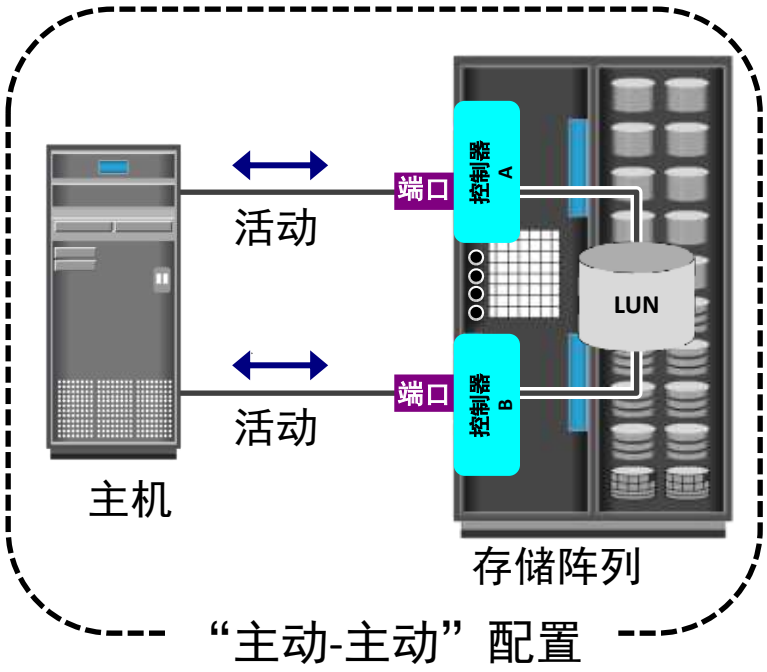
- 在存储阵列上实施
- 防止在共享环境中未授权或意外使用 LUN

例如，假设某一存储阵列具有两个分别存储销售部数据和财务部数据的 LUN。如不进行 LUN 掩蔽，则这两个部门都可以轻松地查看和修改对方的数据，从而给数据完整性和安全性带来很高的风险。

ISS 的类型：高端存储系统

- 称为主动-主动阵列，通常面向大型企业应用程序
  - 通过所有可用路径执行对 LUN 的 I/O

- 这些阵列提供以下功能：
  - 高存储容量和大型缓存
  - 容错体系结构
  - 到大型机和开放系统的连接
  - 多个前端端口和接口协议
  - 能够处理大量并发 I/O
  - 支持本地和远程数据复制



ISS 的类型：中端存储系统

- 称为主动-被动阵列，通常面向中小型企业应用程序
  - 仅通过活动路径执行对 LUN 的 I/O

- 这些阵列通常具有两个控制器，每个控制器都具有缓存、RAID 控制器和磁盘驱动器接口
- 与高端阵列相比，前端端口、存储容量和缓存更少
- 支持本地和远程数据复制