

服务器RAID技术及应用



范广敏

服务器网络事业部

联想集团有限公司

我们的目标

- 什么是RAID？
- RAID技术是怎样实现数据保护的？
- RAID技术带给用户什么价值？



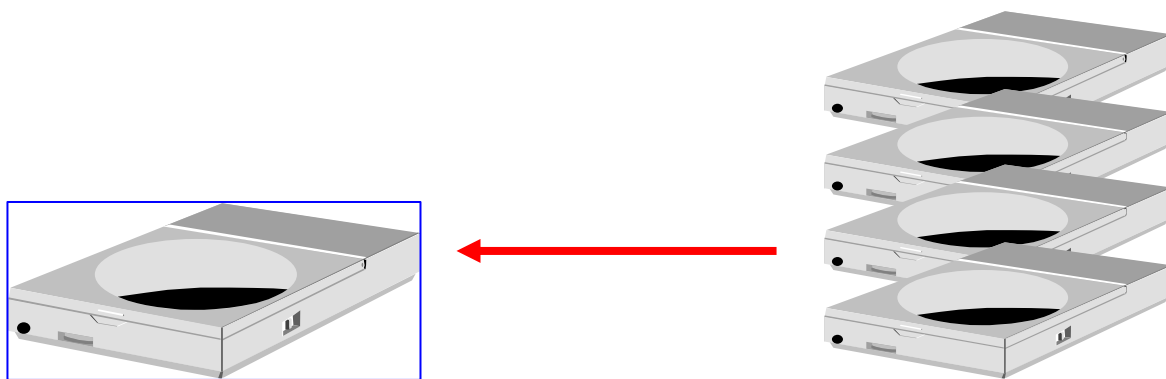
培训内容

- RAID术语
- RAID技术的实现
- IDE RAID与SCSI RAID技术及应用
- 常用RAID卡产品介绍
- 问题与讨论

RAID技术的定义

➤ RAID

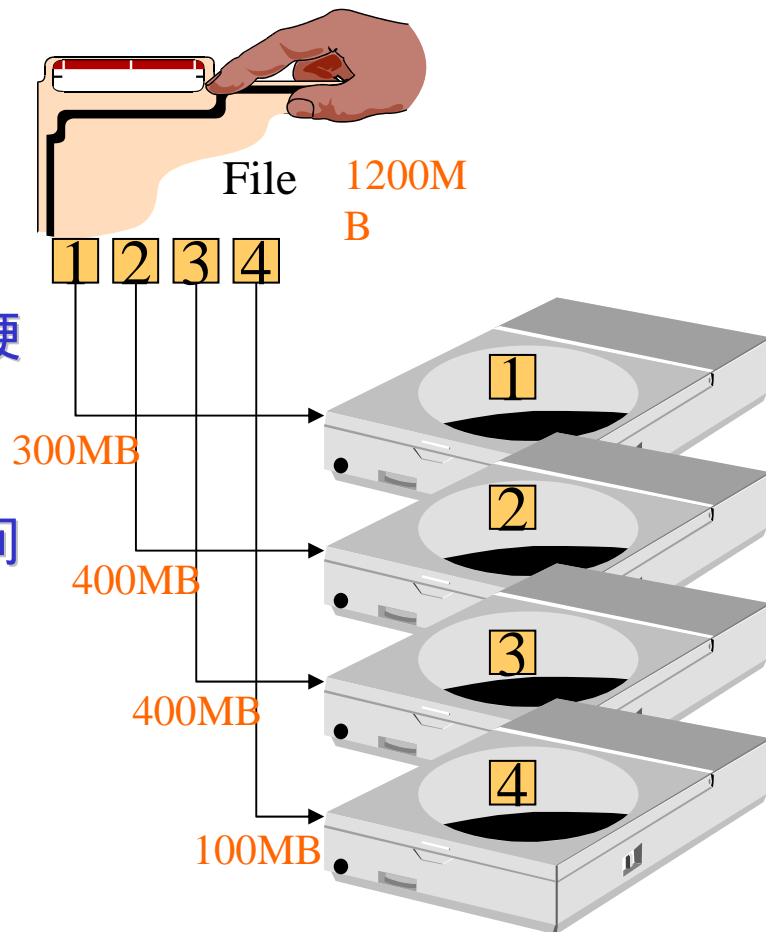
- ❖ Redundant Array of Independent Disks
- ❖ RAID是将同一阵列中的多个磁盘视为单一的虚拟磁盘，数据是以分段的方式顺序存放于磁盘阵列中。



Disk Spanning

➤ Disk Spanning

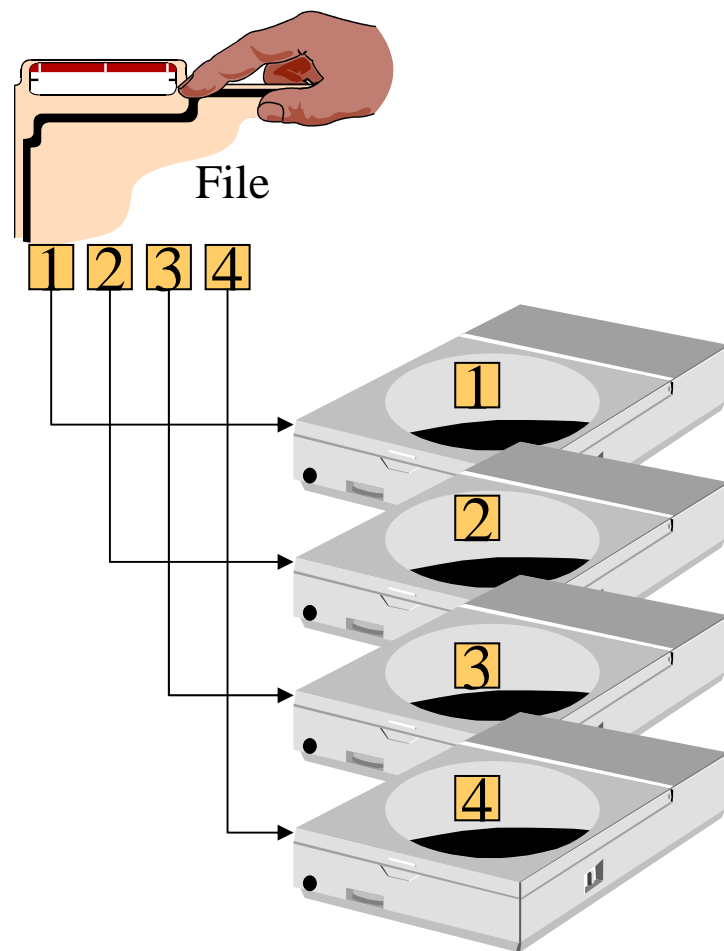
- ❖ 数据跨盘技术使多个硬盘像一个硬盘那样工作
- ❖ 用廉价的资源来突破现有硬盘空间限制
- ❖ 最大限度的利用磁盘空间
- ❖ 不能改善硬盘的可靠性和速度
- ❖ JBOD (Just A Bunch Of Disks)



Disk Striping

➤ Disk Striping

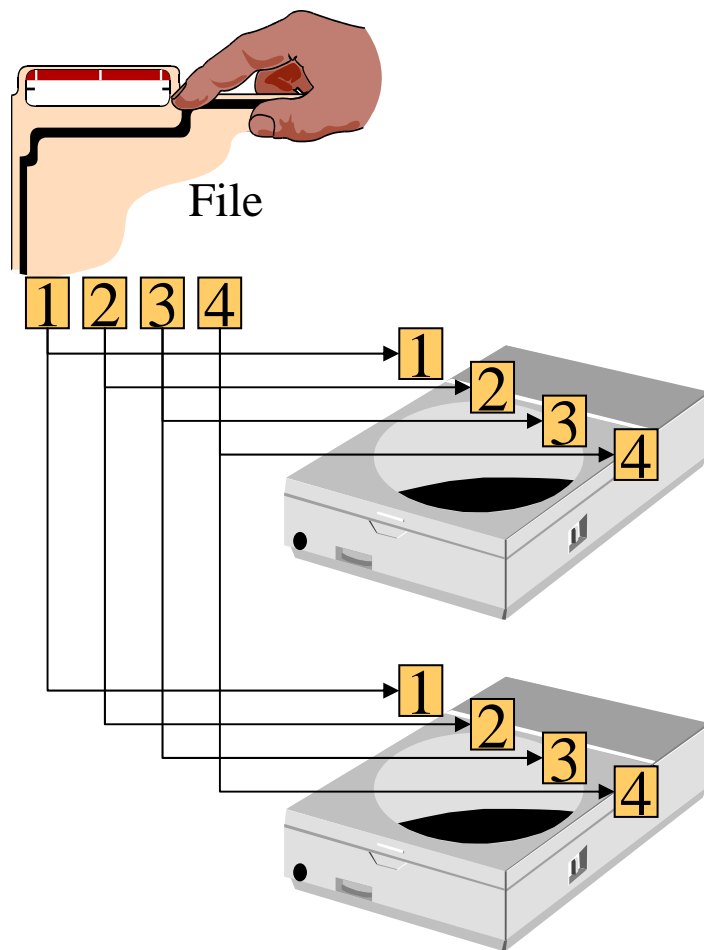
- ❖ 将数据按照一定大小分成多个数据块，这些数据块可以被分别存放在不同的物理盘上
- ❖ 系统在从特定硬盘读取数据时可以通知下个目标盘准备数据
- ❖ 提高系统读写数据的性能



Disk Mirroring

➤ Disk Mirroring

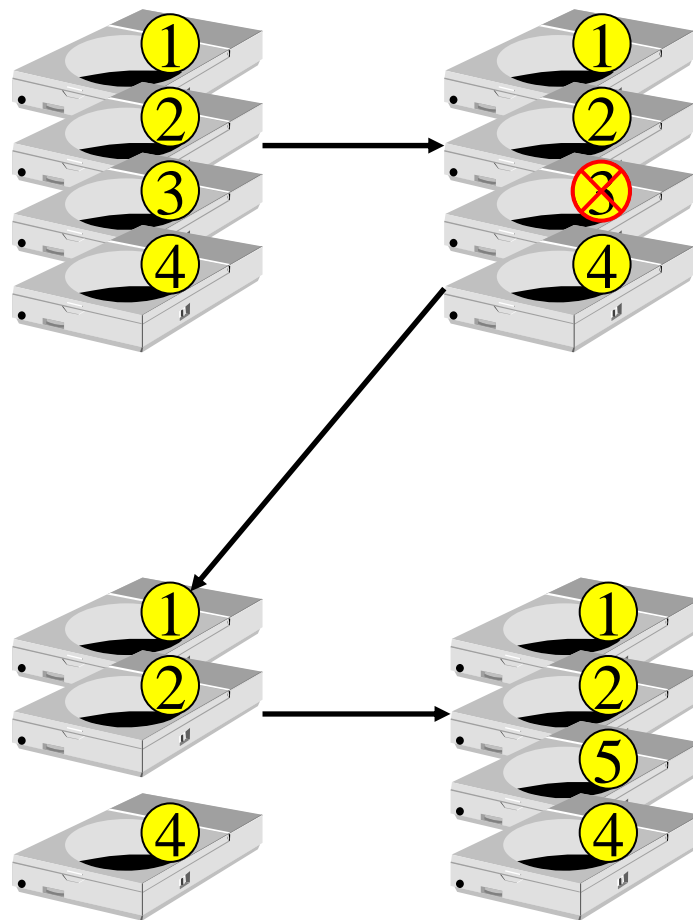
- ❖ 将相同的数据同时写入多个硬盘中
- ❖ 当某个物理硬盘失效时，提供数据资料的保护能力
- ❖ 降低系统写数据的性能



Hot Swap

➤ Hot Swap

- ❖ 处于运行状态的磁盘阵列子系统当出现单个物理盘失效的情况时，采用新硬盘将失效物理盘在线替换，同时保证系统稳定运行
- ❖ 只有RAID级别为1、3、5、10、30、50的阵列才提供该功能



Parity

➤ Parity

- ❖ 来自多个物理磁盘上的数据通过异或(XOR)操作运算产生的冗余奇偶数据
- ❖ 当单个硬盘失效时，这些冗余数据能够通过与其它物理磁盘上的数据进行异或(XOR)操作而恢复由于硬盘失效而丢失的数据
- ❖ 产生的冗余数据可以被存放于一个专作奇偶校验用的硬盘上，也可以将这些奇偶校验数据分散分布在磁盘阵列的全部硬盘中
- ❖ 产生和存储奇偶校验数据需要一些额外的操作，目前产生奇偶校验数据有两种方式：硬件生成和软件计算。

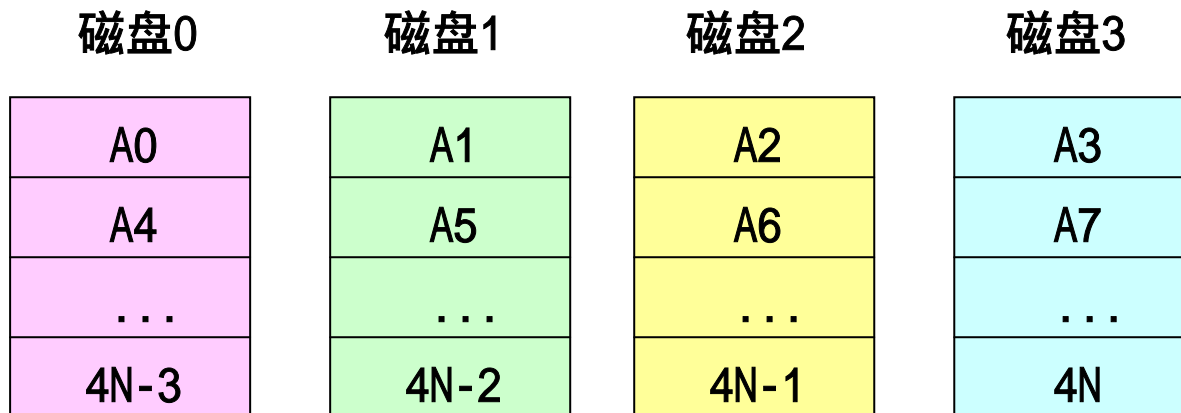
培训内容

- RAID术语
- RAID技术的实现
- IDE RAID与SCSI RAID技术及应用
- 常用RAID卡产品介绍
- 问题与讨论

RAID Levels— RAID 0

➤ RAID0

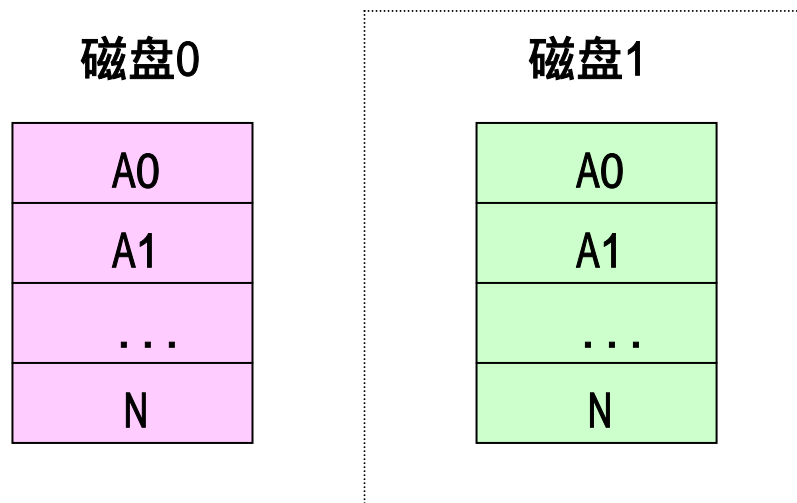
- ❖ Striping
- ❖ 数据以分段（1K - 8M Byte）的方式放在磁盘阵列中，没有校验数据。
- ❖ 没有容错能力，同一通道4个1GB的硬盘做RAID0的效率是1个4GB硬盘的3-3.5倍；



RAID Levels— RAID 1

➤ RAID1

- ❖ Mirroring
- ❖ 使磁盘读取的效率增加，但写入的效率降低
- ❖ 通过数据直接备份具有容错能力



RAID Levels— RAID 3

➤ RAID3

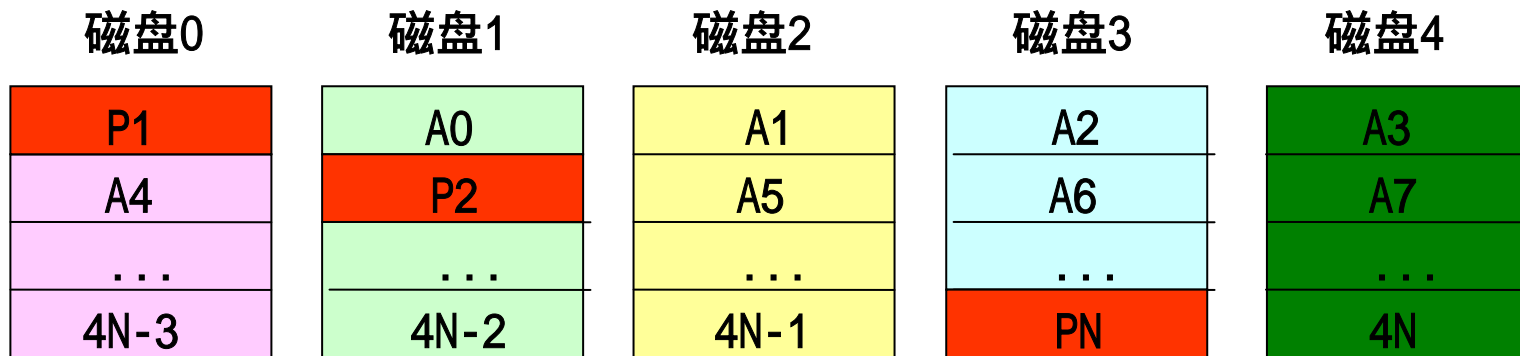
- ❖ Striping with Dedicated Parity Drive
- ❖ 有校验数据，提供数据容错能力
- ❖ 当单个硬盘失效时，会产生奇偶盘I/O瓶颈效应

磁盘0	磁盘1	磁盘2	磁盘3
A0	A1	A2	P1
A3	A4	A5	P2
...
4N-3	4N-2	4N-1	PN

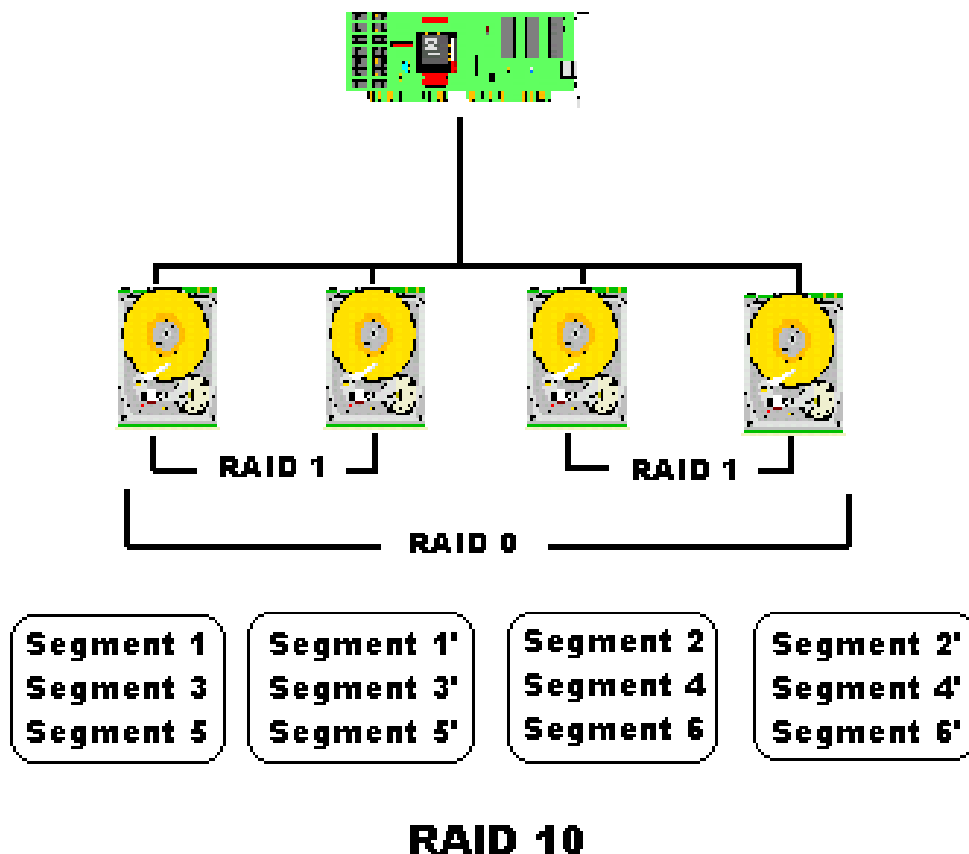
RAID Levels— RAID 5

➤ RAID5

- ❖ Striping with Distributed Parity
- ❖ 有校验数据，提供数据容错能力
- ❖ 校验值分散在各个盘的不同位置，相当程度的分散了负载，故有较好的性能，尤其是对小型数据。RAID5适用于银行和股市的联机交易系统（OLTP）。

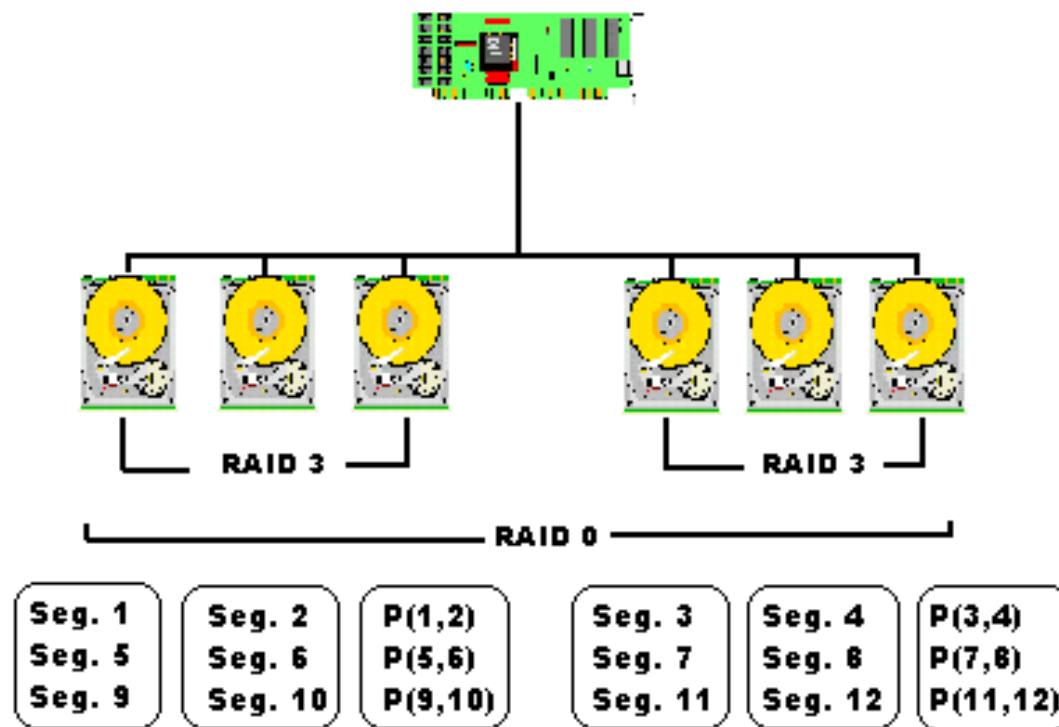


RAID Levels— RAID 10



Two RAID 1's Striped

RAID Levels— RAID 30

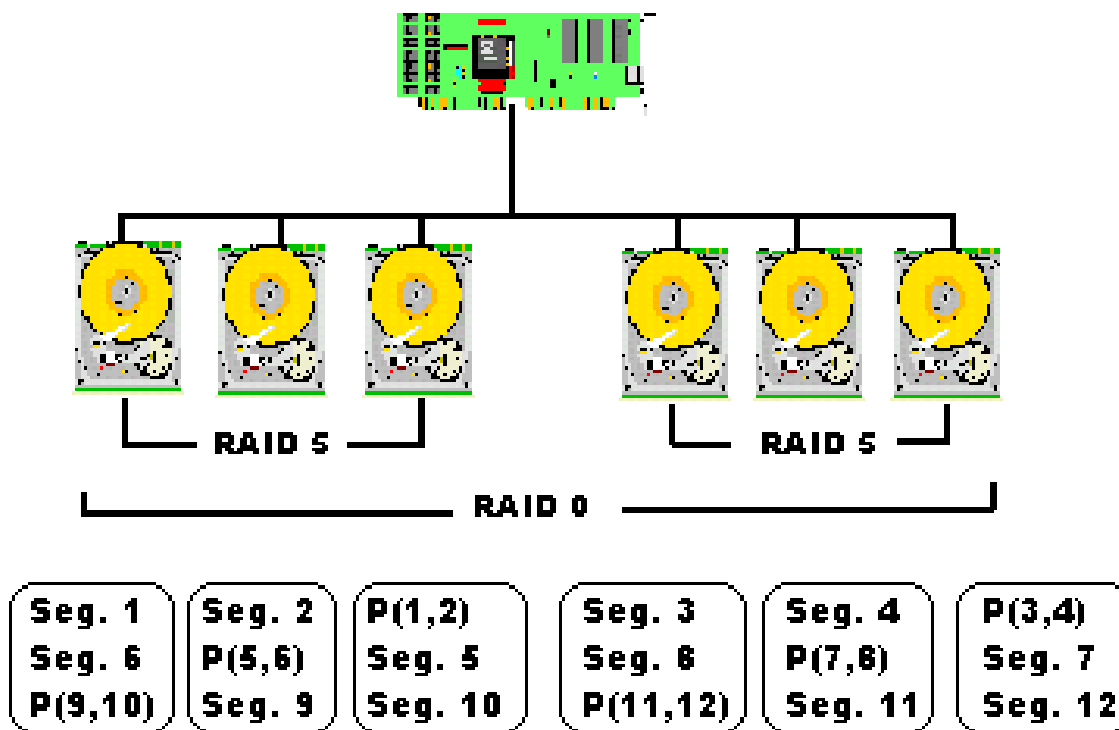


RAID 30

Two RAID 3's Striped



RAID Levels— RAID 50



Two RAID 5's Striped

RAID 50



培训内容

- RAID术语
- RAID技术的实现
- IDE RAID与SCSI RAID技术及应用
- 常用RAID卡产品介绍
- 问题与讨论

SCSI 与 IDE

➤ SCSI(Small Computer System Interface)

- ❖ 定义了一种输入/输出总线和逻辑接口，主要目标是提供一种设备独立的机制用来连接主机和访问设备。

➤ IDE (Intergrated drive electronics)

- ❖ 现在PC机使用的主流硬盘接口

IDE RAID

- ❖ 基于IDE接口的RAID 应用产品称为IDE RAID
 - ❖ 与SCSI RAID相比有及具竞争力的性价比
 - ❖ 为中低端应用的服务器产品提供更强的数据处理和数据保护以及较高的性能价格比
 - ❖ 支持RAID 0 , 1 , 0+1 , 5

SCSI RAID

- 基于SCSI接口的RAID 应用产品称为SCSI RAID
 - ❖ 多个I/O并行操作，设备传输速度快；
 - ❖ 可扩展性好，最多可连接15个外部scsi设备；
 - ❖ 支持RAID级别多，RAID0、RAID1、RAID10、RAID3、RAID5、RAID30、RAID50、JBOD等；
 - ❖ 支持Hotspare,在线热插拔，性能更稳定、可靠性更好；



为什么用RAID卡

➤ 用户对磁盘系统的四大要求

- ❖ 增加磁盘I/O存取速度
- ❖ 数据安全性要求，及容错（Fault tolerance）的能力
- ❖ 有效的利用磁盘空间
- ❖ 分担主机CPU的I/O事务
- ❖ 降低内存及磁盘的性能差异
- ❖ 提高计算机的整体工作性能

IDE RAID 卡

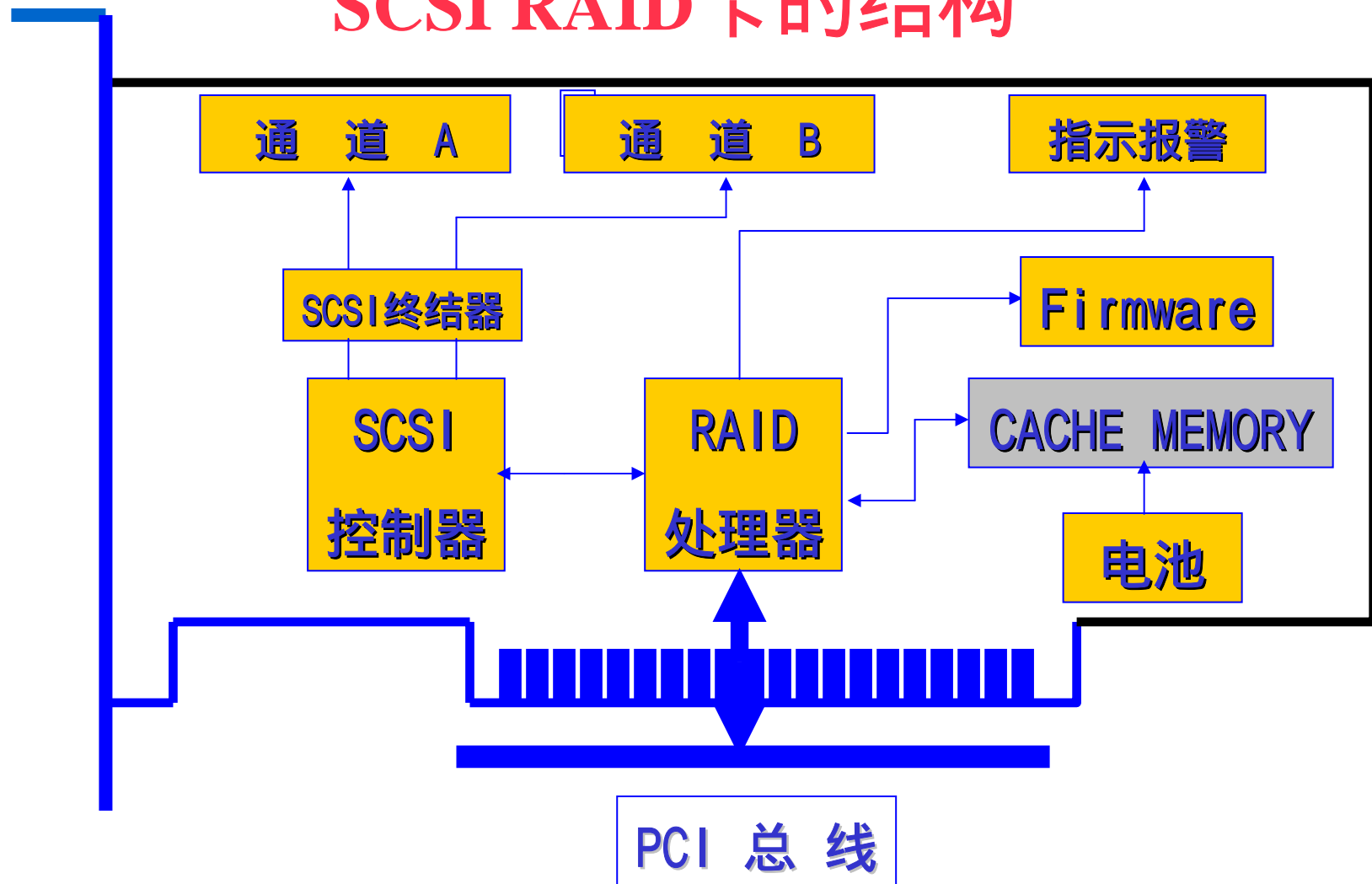


- 将若干IDE硬盘做成一个或几个阵列盘，提高了系统中磁盘I/O的性能，并且使磁盘系统具有数据镜像的安全保护功能。
- 具有可管理性，在操作系统支持下，管理软件可监控阵列的状态，添加和删除阵列，在线数据恢复等；
- 最多支持4个硬盘，可实现RAID 0、1、0+1
- QW1060、1600B应用了IDE RAID卡，2100主板集成IDE RAID 控制器

SCSI RAID卡

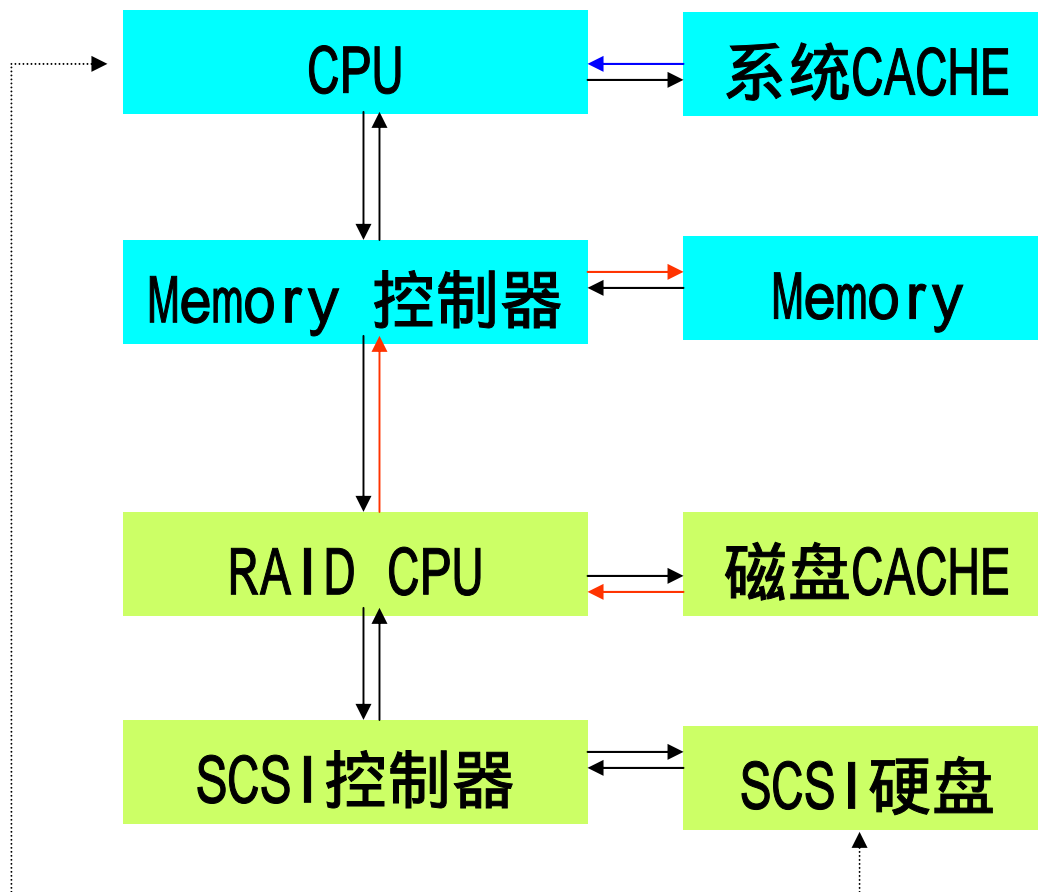
- 将若干硬盘做成一个或几个阵列盘，提高了系统中磁盘I/O的性能，并且使磁盘系统具有容错的安全保护功能。
- 通过磁盘Cache的运用，提高系统的存取速度，
- 结合SAF-TE（SCSI Access Fault Tolerant Enclosure）对磁盘阵列中的磁盘进行监控管理，管理的人机接口有
 - 1) RAID卡本身的指示报警；
 - 2) 具有SAF-TE功能的热插拔背板；
 - 3) 操作系统上安装的监控软件。

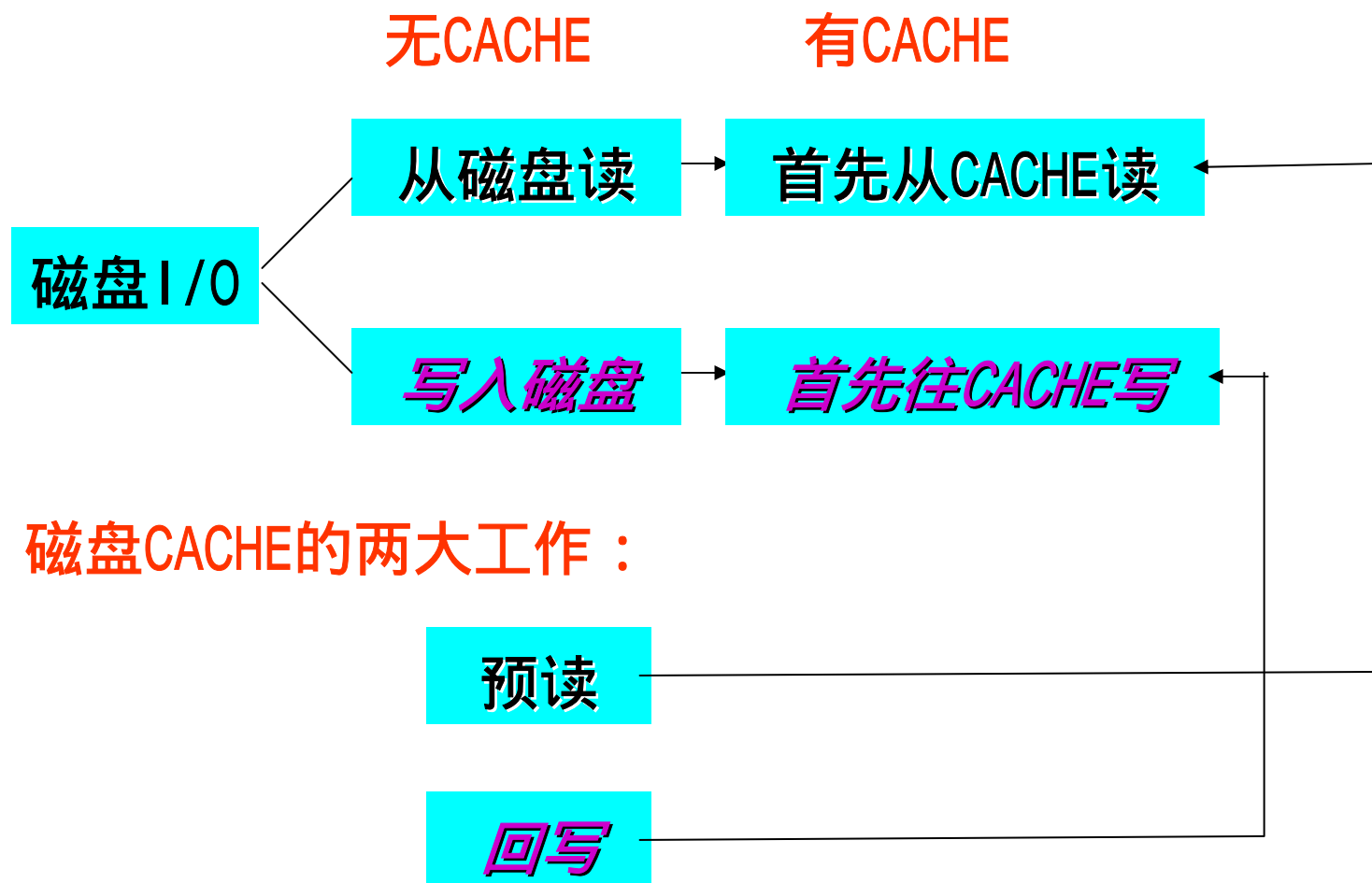
SCSI RAID卡的结构



磁盘CACHE

RAID卡提高磁盘读写性能的另一手段是：磁盘CACHE。

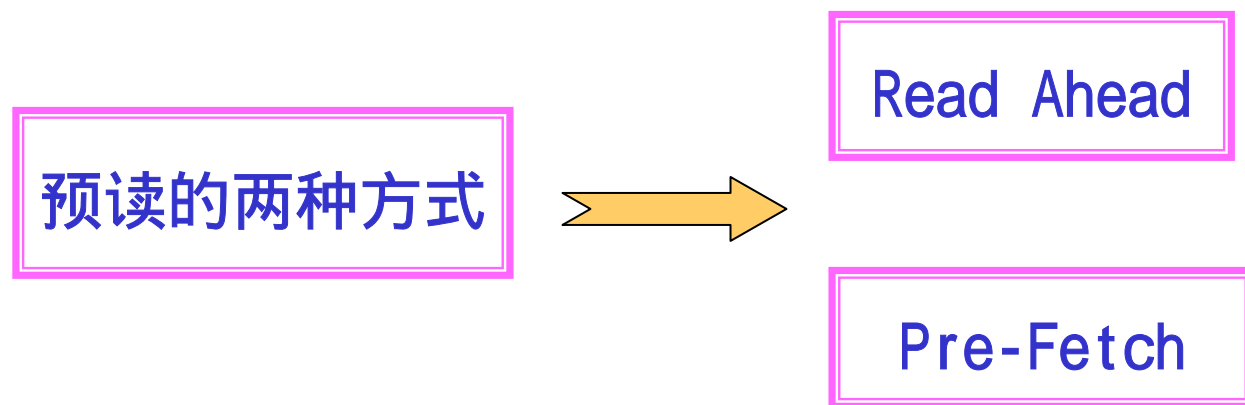




预 读

CACHE预读提高了计算机系统中的硬盘读的功能，尤其是在读取含有大量文件碎片的文件时。

具有良好预读功能的RAID卡能在看起来很随机的读访问中，识别出读取磁盘的规律，通过这个规律提前将系统要读取的数据放在CACHE中。



Read ahead

由于硬盘数据经常是以一族连续的硬盘扇区组织起来的，所以有时候如把系统所请求的扇区随后的一个扇区里的数据同时读进来是有价值的。

— 对于数据文件的读取有利，特别是系统CPU的性能低时。

Pre-Fetch

当RAID卡发现系统要读的是先前已经读过的数据时，在这一次，便将这一个数据块的数据写到CACHE里。

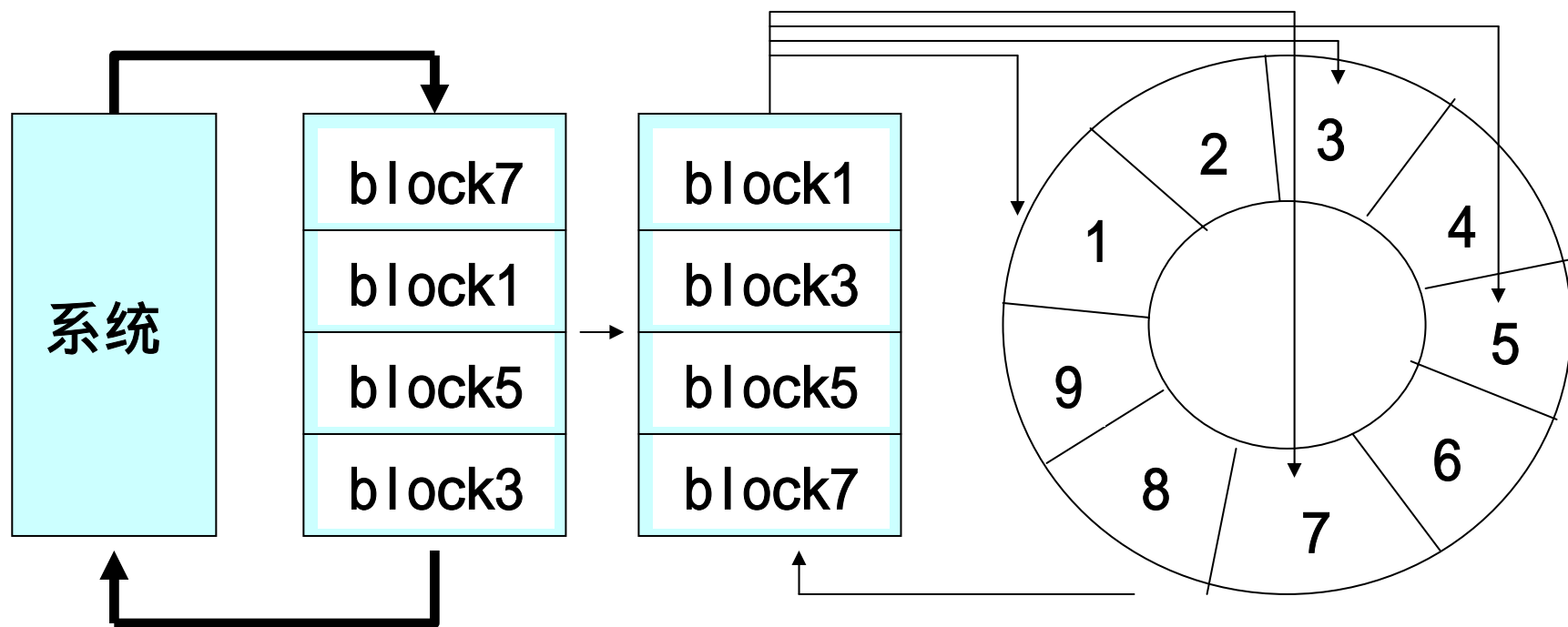
— 对于程序文件的读取有利



回写

回写是通过暂时将数据存在CACHE里，从而推迟将数据写到慢设备（如硬盘、磁带机）的一种工作方式。数据将在随后的时间，硬盘闲置的时候写到硬盘中。写的时候也是统一将CACHE内的尚未写出的数据按照数据块的在硬盘中的BLOCK序号写入，这样可以提高写的效率。



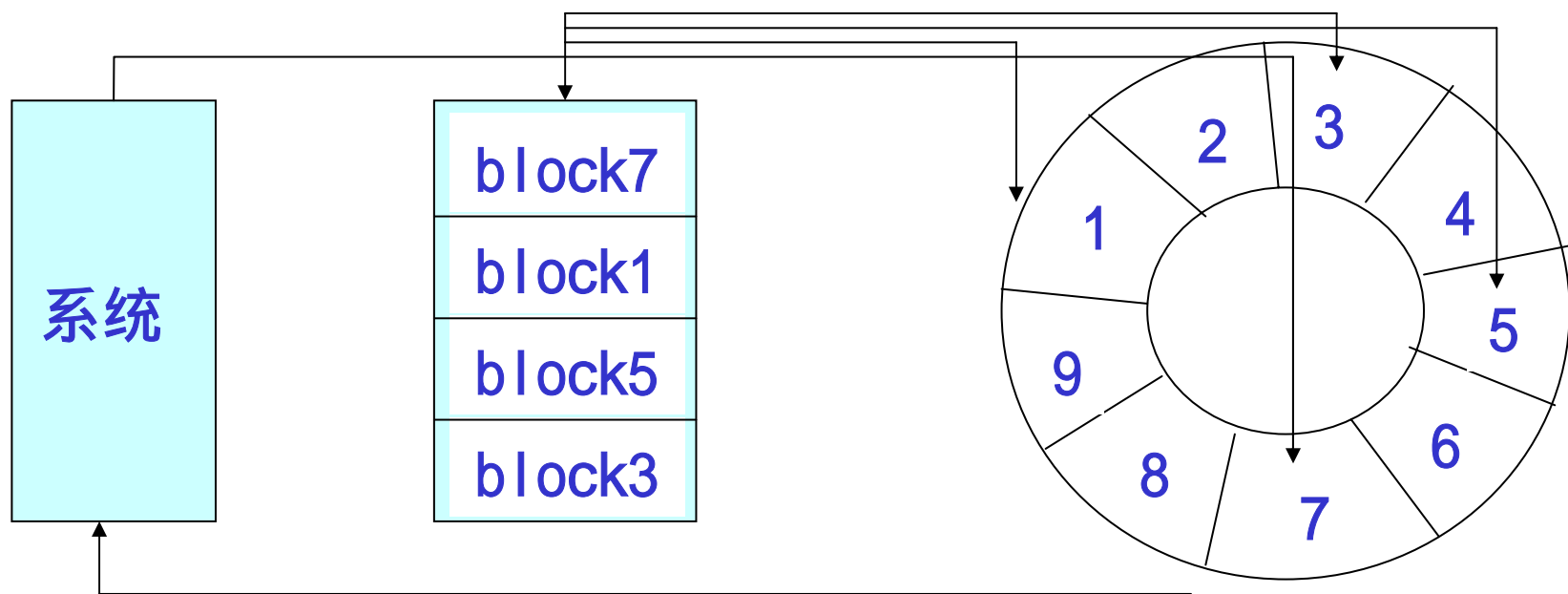


电梯式回写

注：回写需要加电池给CACHE供电，以免数据在写到硬盘之前系统断电导致硬盘数据丢失。

通写

通写模式下，所有数据在以命令完成状态返回到计算机之前，直接写到硬盘。



磁盘Cache性能分析

➤ 容量

- ❖ 增加CACHE大小对于预读来说，为系统提供了更多的来自CACHE的可供读取的记录。对于回写来说，允许控制卡保存更多的记录留待后期写磁盘。特别是对于电梯式回写，使得连续的回写段之间有更近的间隔，降低硬盘写操作的平均访问时间并提高了吞吐率。

大容量cache



- 1、当用户数很多而且不是都访问相同的数据。
- 2、当数据文件很大时。

磁盘Cache性能分析

➤ 磁盘CACHE的类型

EDO：工作频率在33MHz，RAID处理器频率低于100MHz时

SDRAM：工作频率在66MHz，RAID处理器频率高于100MHz时

--- 提高系统I/O的速度

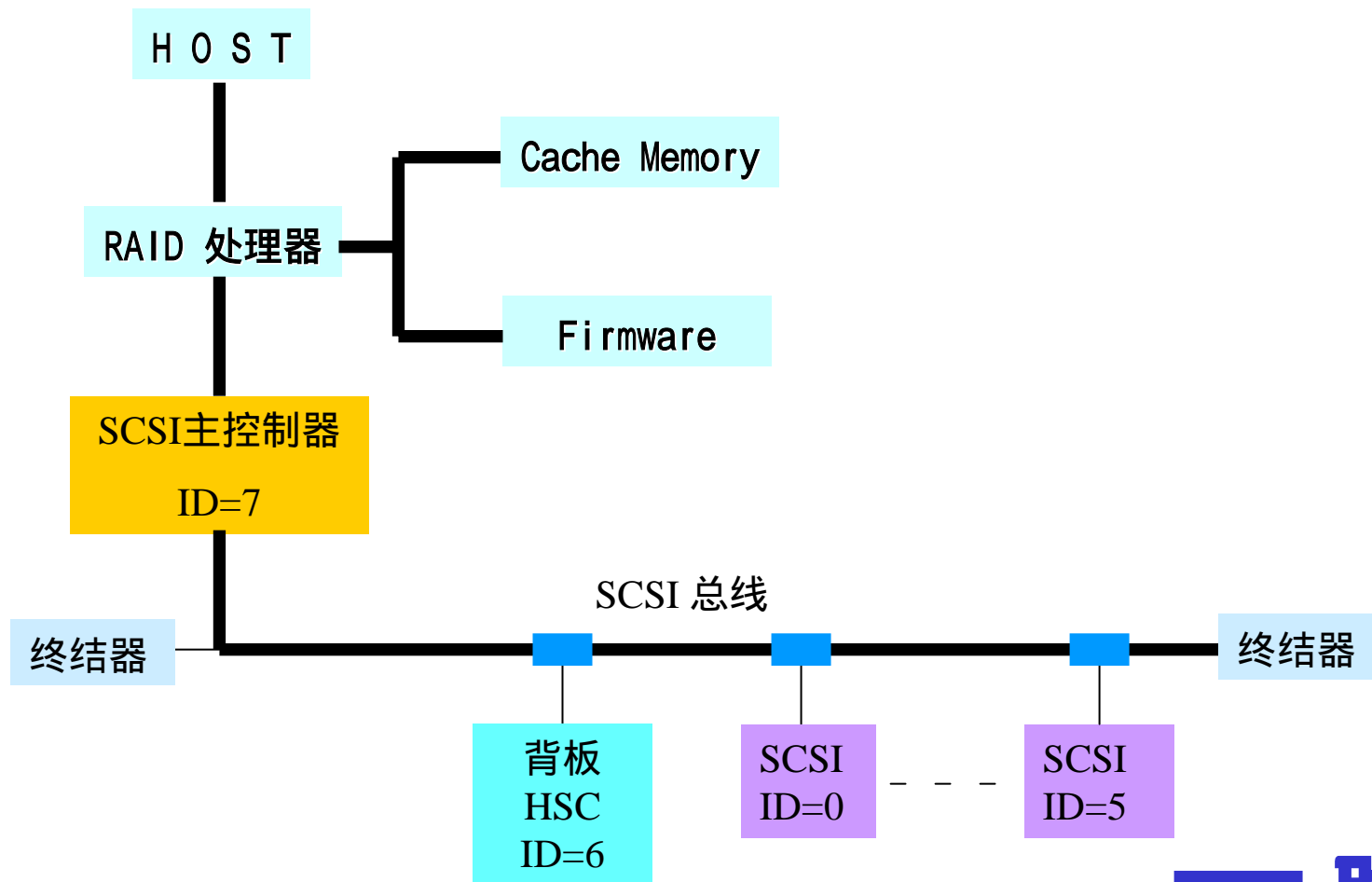
ECC：对CACHE数据进行校验

Non-ECC：不对CACHE数据进行校验

--- 提高系统I/O的可靠性

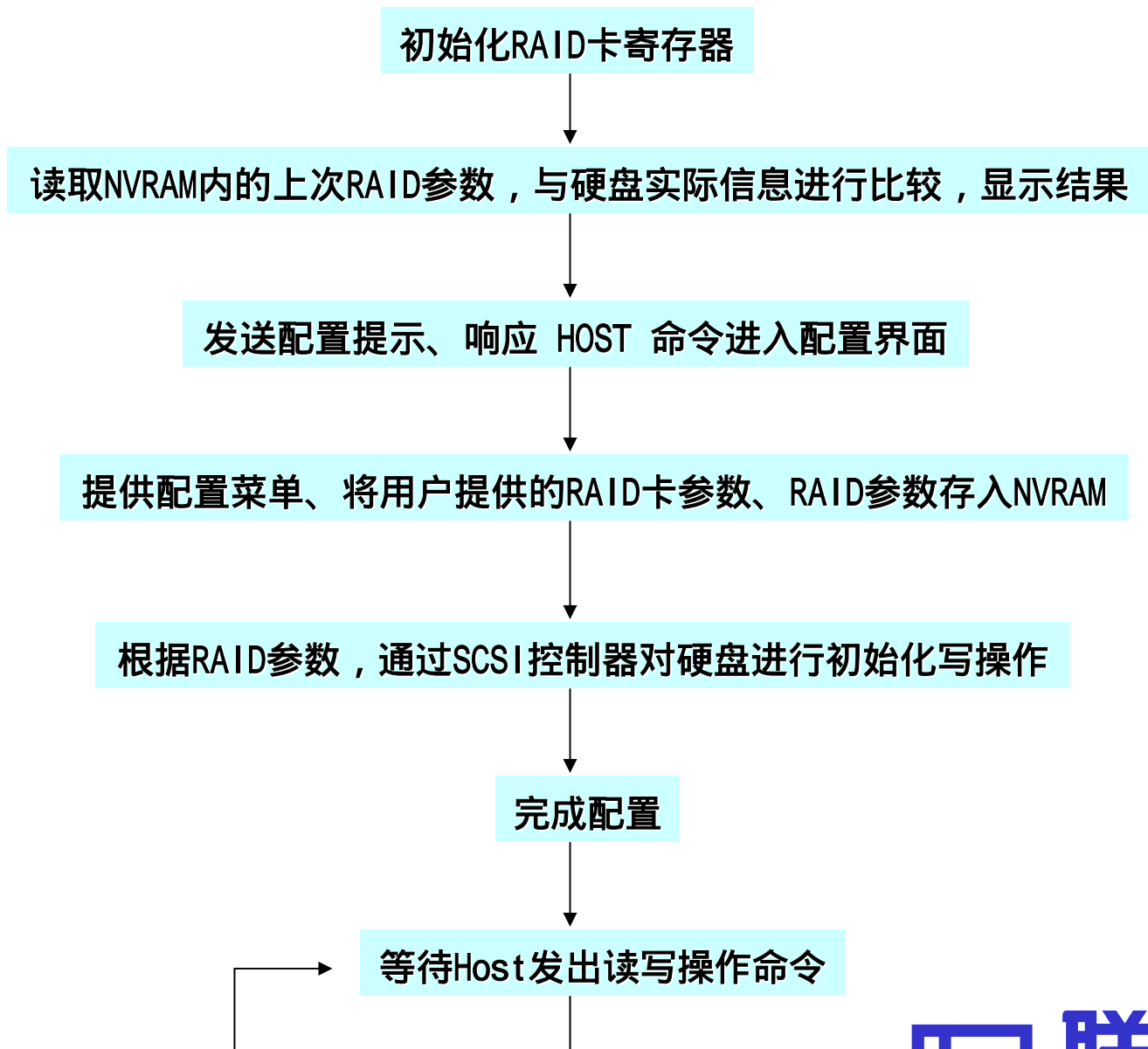
Mylex的RAID卡需要ECC的；Adaptec的RAID卡没限定

带RAID卡SCSI系统硬件框图



RAID 处理器

- 是一个PCI从设备，接受并执行来自系统的命令。同时占用PCI中断，代表SCSI磁盘子系统向系统。提出中断请求，请求占用PCI总线，返回对系统。命令的响应，如输送SCSI硬盘上的数据。
- 作为RAID卡的CPU，通过执行闪存中的Firmware，控制SCSI控制器、Cache Memory以及指示报警电路，来实现RAID卡的功能，运作流程如流程图所示。

RAID**处理器
运作
流程
简图**

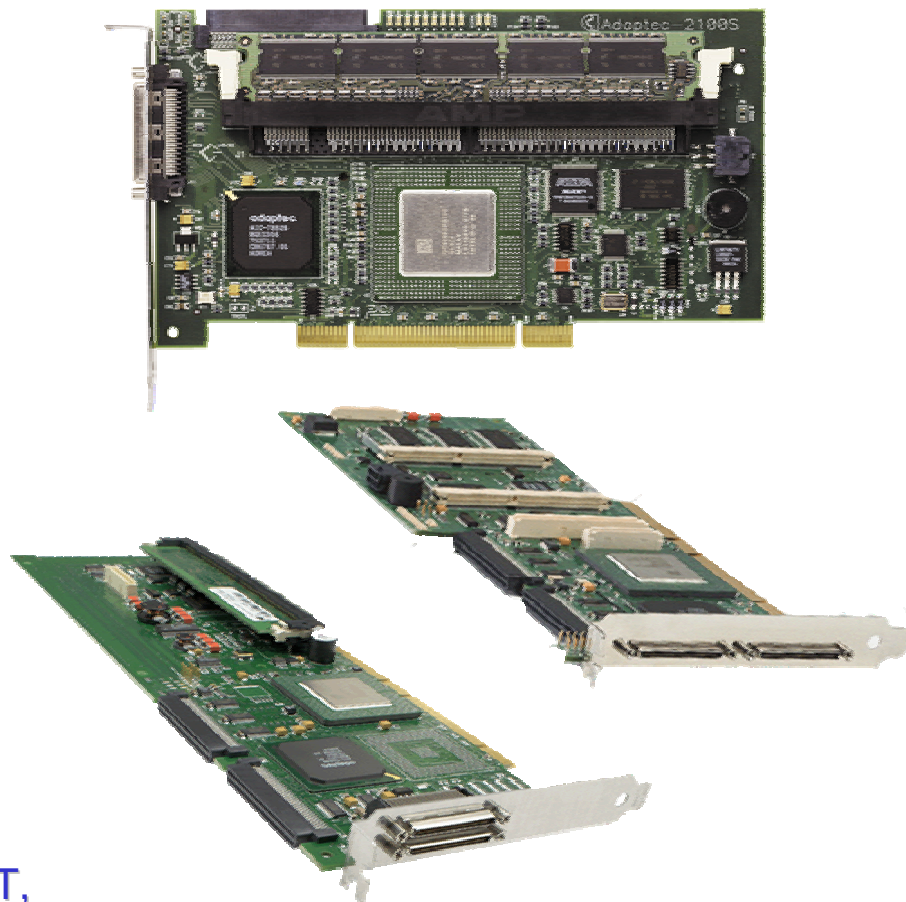
培训内容

- RAID术语
- RAID技术的实现
- IDE RAID与SCSI RAID技术及应用
- *常用RAID卡产品介绍*
- 问题与讨论

SCSI RAID 卡产品介绍

➤ Adaptec

- ❖ 2100S
 - ❖ 单通道Ultra160 SCSI RAID
 - ❖ i960RS/32-Bit/33MHz
- ❖ 3200S
 - ❖ 双通道Ultra160 SCSI RAID
 - ❖ i960RN/64-Bit/33MHz
- ❖ 3210S
 - ❖ 双通道Ultra160 SCSI RAID
 - ❖ IOP303/64-Bit/66MHz
- ❖ 3410S
 - ❖ 四通道SCSI RAID
 - ❖ IOP303/64-Bit/66MHz
 - ❖ RAID 0, 1, 0/1, 5, 0/5
 - ❖ 支持 Windows 2000 , Windows NT, NetWare , SCO UNIX以及 Linux等多种操作系统



SCSI RAID 卡产品介绍

➤ Mylex(IBM)

❖ AcceleRAID 160

- ❖ 单通道Ultra160 SCSI RAID
- ❖ i960RS/32-Bit/33MHz

❖ AcceleRAID 170

- ❖ 单通道Ultra160 SCSI RAID
- ❖ i960RM/32-Bit/33MHz

❖ AcceleRAID 352

- ❖ 双通道Ultra160 SCSI RAID
- ❖ i960RN/64-Bit/33MHz
- ❖ RAID 0, 1, 0+1, 3, 5, 10, 30, 50, JBOD
- ❖ 支持 Windows 2000 , Windows NT, NetWare , Solaris , SCO UNIX等多种操作系统

SCSI RAID 卡产品介绍

➤ AMI(LSI)

- ❖ Express 300(Series 490)
 - ❖ 单通道U2W SCSI RAID
 - ❖ i960RM/32-Bit/33MHz
- ❖ Express 500(Series 475)
 - ❖ 单通道Ultra160 SCSI RAID
 - ❖ i960RM /32-Bit/33MHz
- ❖ Elite 1600(Series 493)
 - ❖ 双通道Ultra160 SCSI RAID
 - ❖ i960RN/64-Bit/66MHz
 - ❖ RAID 0, 1, 3, 5, 10, 30, 50
 - ❖ 支持 Windows 2000 , Windows NT, NetWare , Solaris , SCO UNIX以及 Linux等多种操作系统

IDE RAID 卡产品介绍

➤ Promise

❖ FastTrak66

- ❖ 双通道ATA/66 RAID
- ❖ 支持四颗硬盘
- ❖ RAID 0, 1, 0+1, JBOD

❖ FastTrak100

- ❖ 双通道ATA/100 RAID
- ❖ 支持四颗硬盘
- ❖ RAID 0, 1, 0+1, JBOD
- ❖ 支持 Windows 2000, Windows NT, NetWare以及 Linux等多种操作系统



万全服务器机型与RAID卡对应表

RAID卡 机型	Adaptec 2100s	Adaptec 3210s	Mylex 170	Mylex 352	Adaptec 2400(IDE RAID	FastTrack 66(IDE RAID)	FastTrack 100(IDE RAID)
2300/2600/2460R							
2300T							
2600T							
3200C/3250R							
4200 /4600R		X					
2400c(370DL3主 板)							
1600B							
1060							

培训内容

- RAID术语
- RAID技术的实现
- IDE RAID与SCSI RAID技术及应用
- 常用RAID卡产品介绍
- 问题与讨论

联想万全与您共同进步！

问题与讨论



我们的目标
您清楚了
吗？

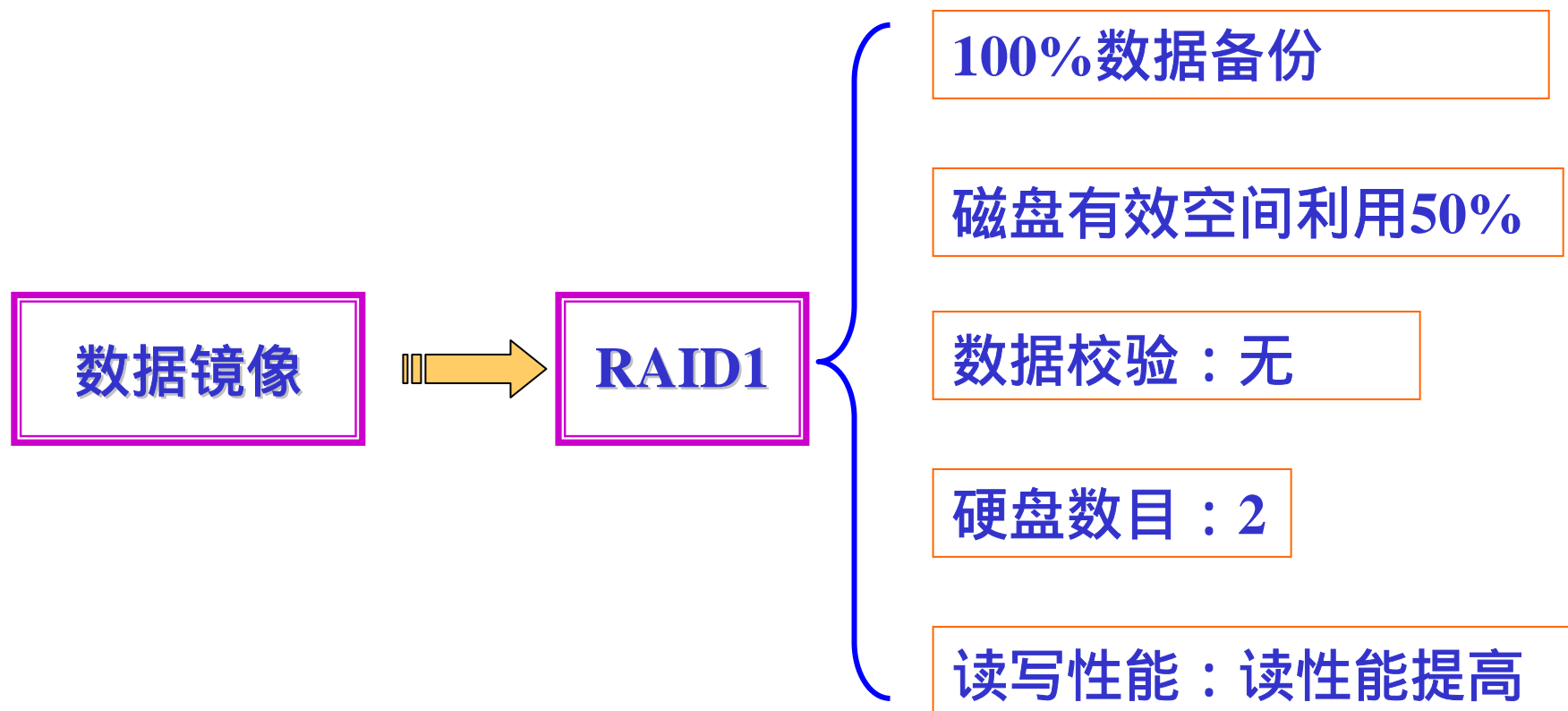


什么是RAID

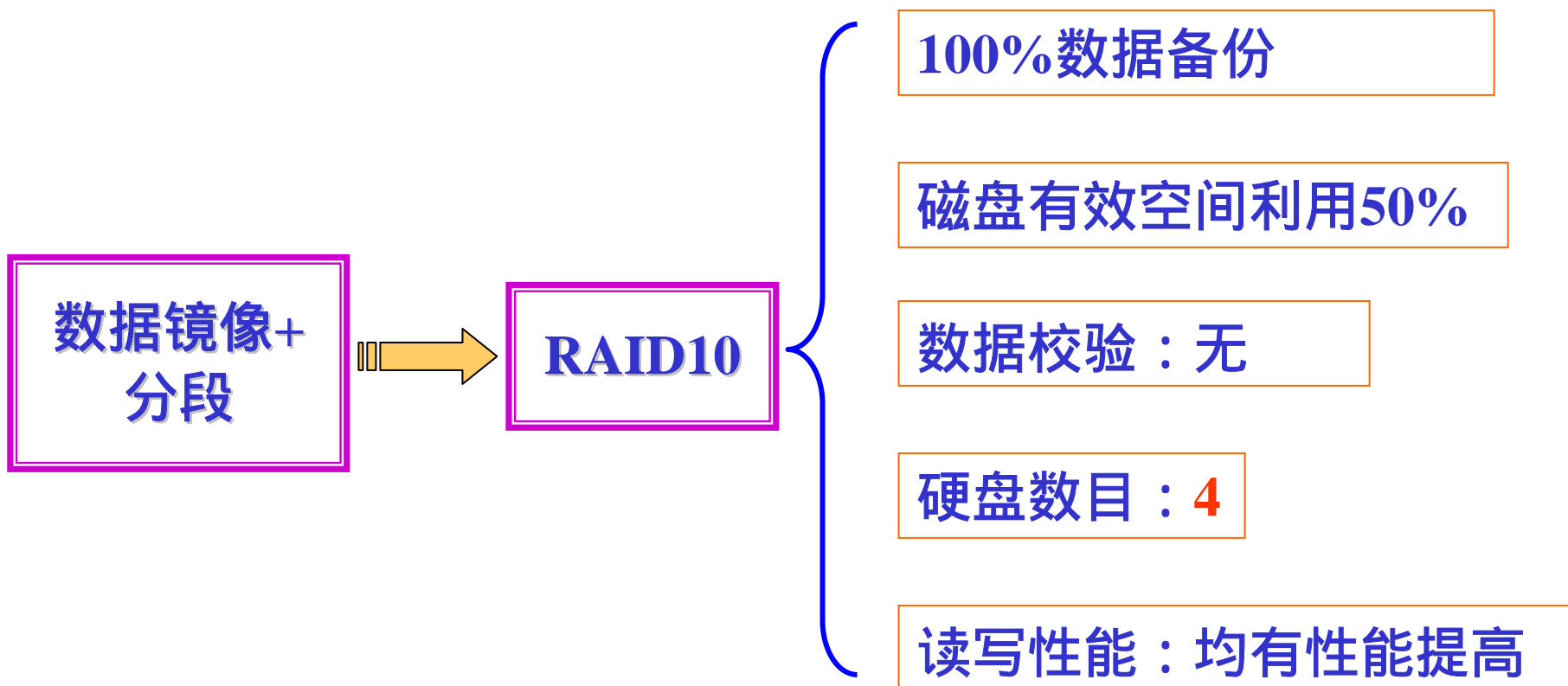
➤ RAID就是廉价冗余磁盘阵列。

- ❖ 一种在配件、存储子系统和系统级别上通过硬件和软件冗余来保护重要数据和改善系统性能的方法；
- ❖ 利用磁盘分段、高速缓存等技术提高磁盘存取速度，同时通过磁盘镜像、数据冗余提供数据保护和备份。

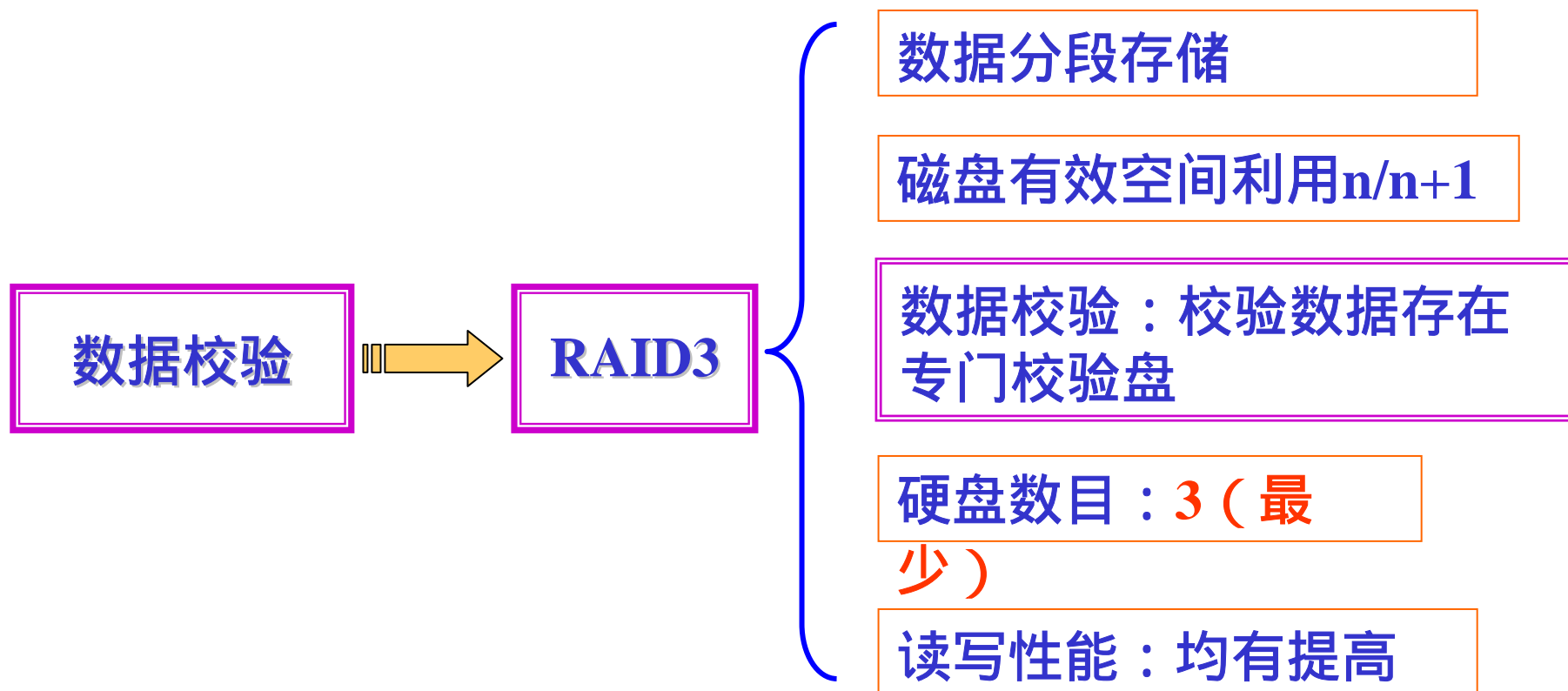
RAID技术如何实现数据保护



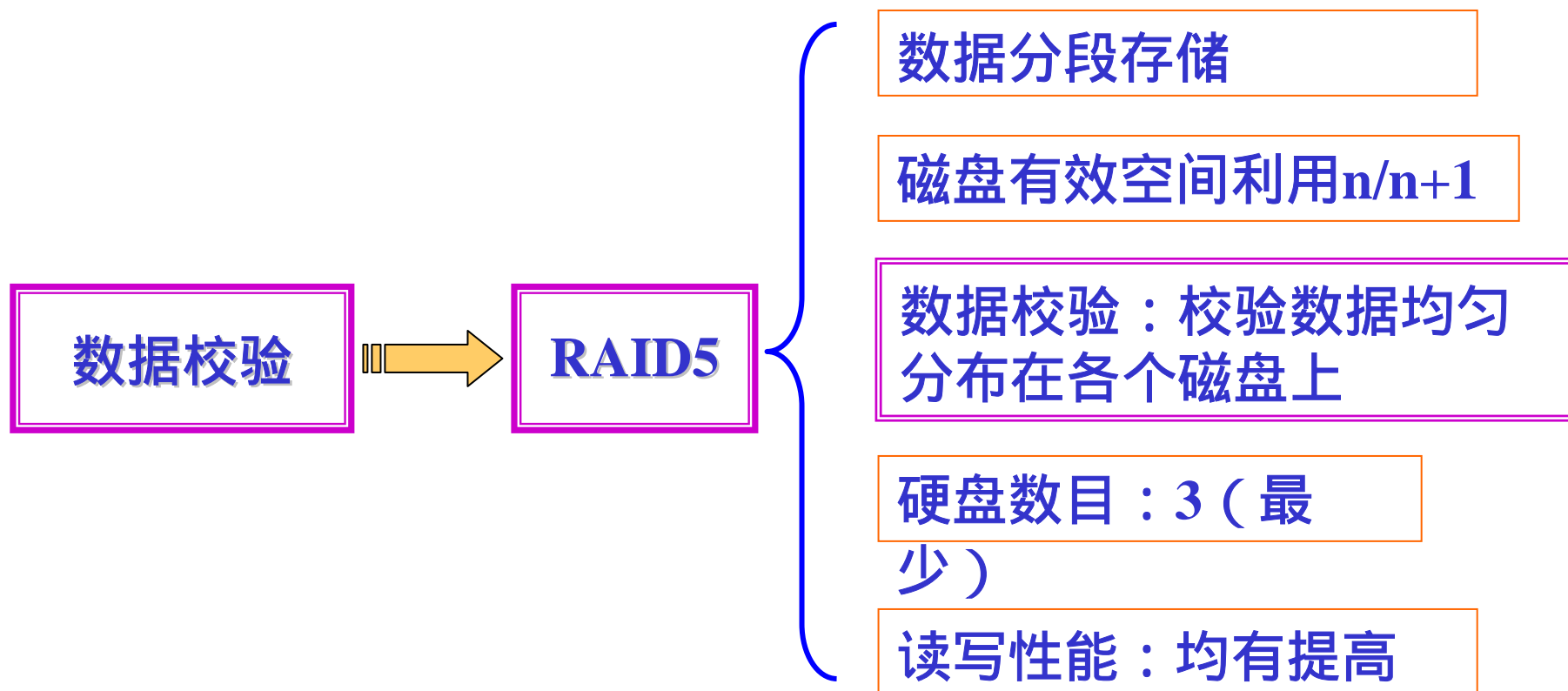
RAID技术如何实现数据保护



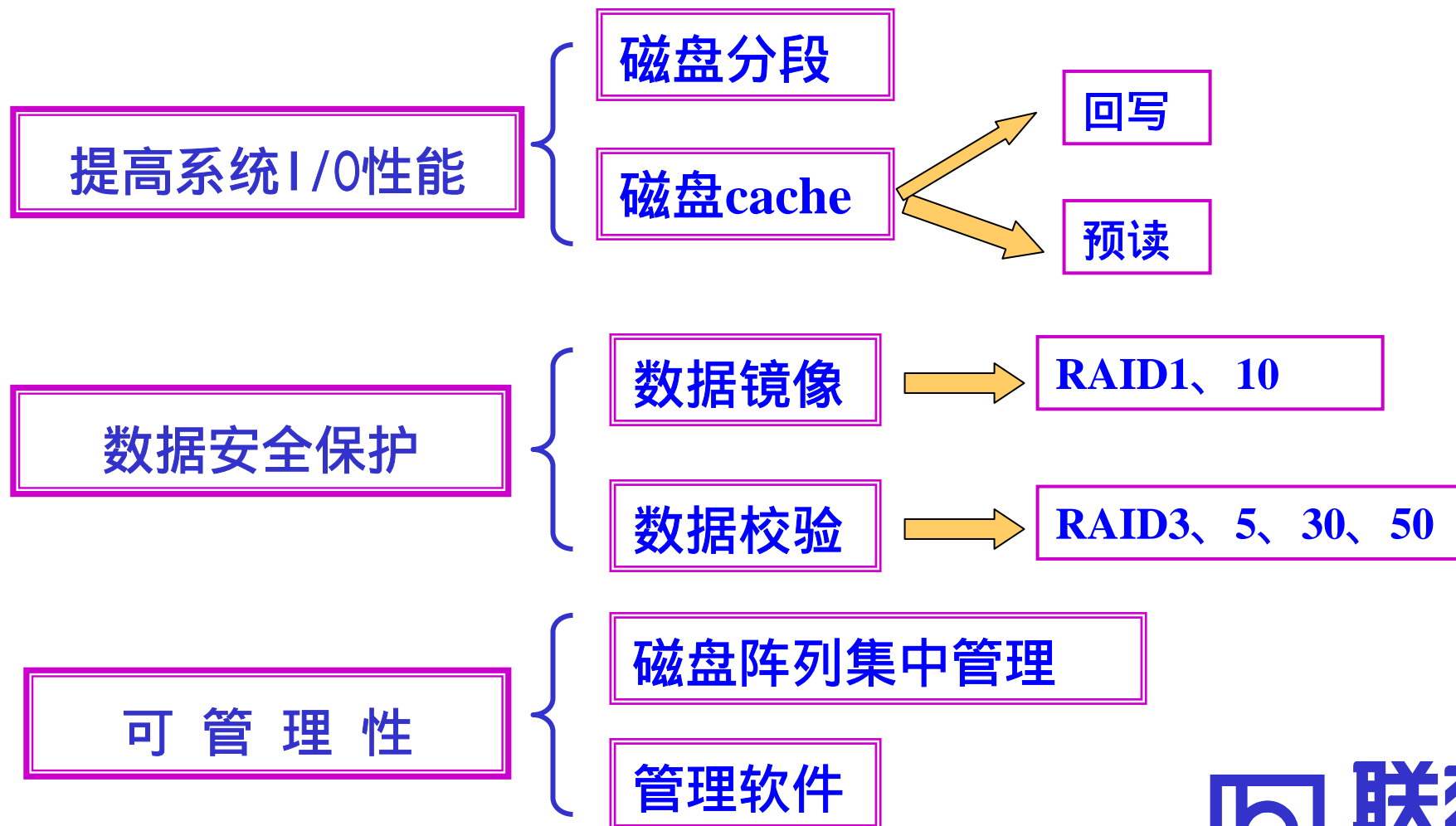
RAID技术如何实现数据保护



RAID技术如何实现数据保护



RAID 技术的价值



万全服务器实力在！信心在！

谢谢！

