



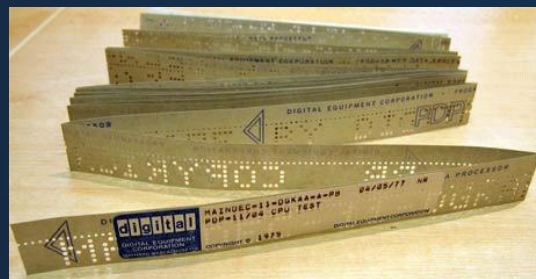
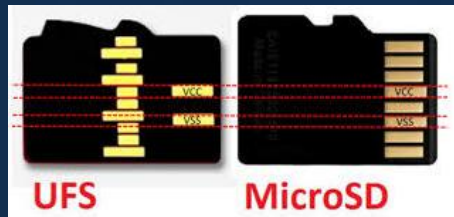
电子科技大学校企合作课

大道至简 ⚡ 极速部署

可信赖的、产品化的云平台

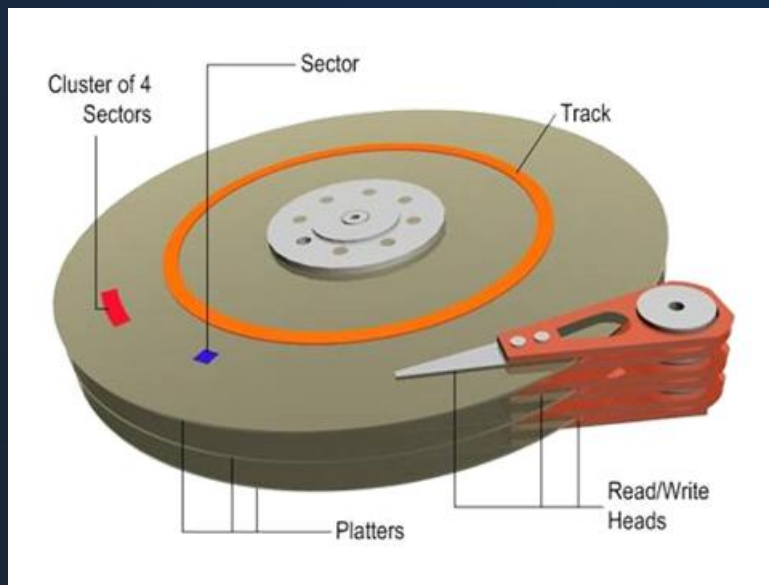
存储技术

- 1 存储介质
- 2 存储主要接口
- 3 存储设备形态
- 4 RAID
- 5 软件定义存储

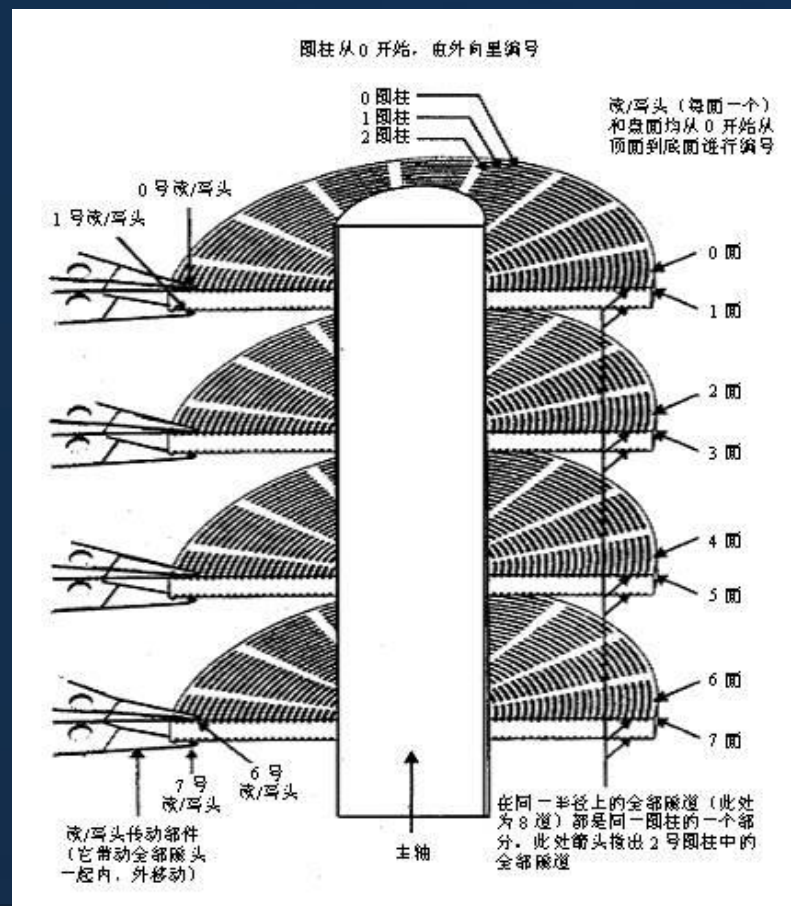


- 硬盘结构

- 磁臂/磁头、主轴、扇区 (512B)、磁道、盘片、块扇区组成块，块组织磁道，磁道铺盘片



- 主轴驱动盘面旋转，磁臂摆动尽收数据



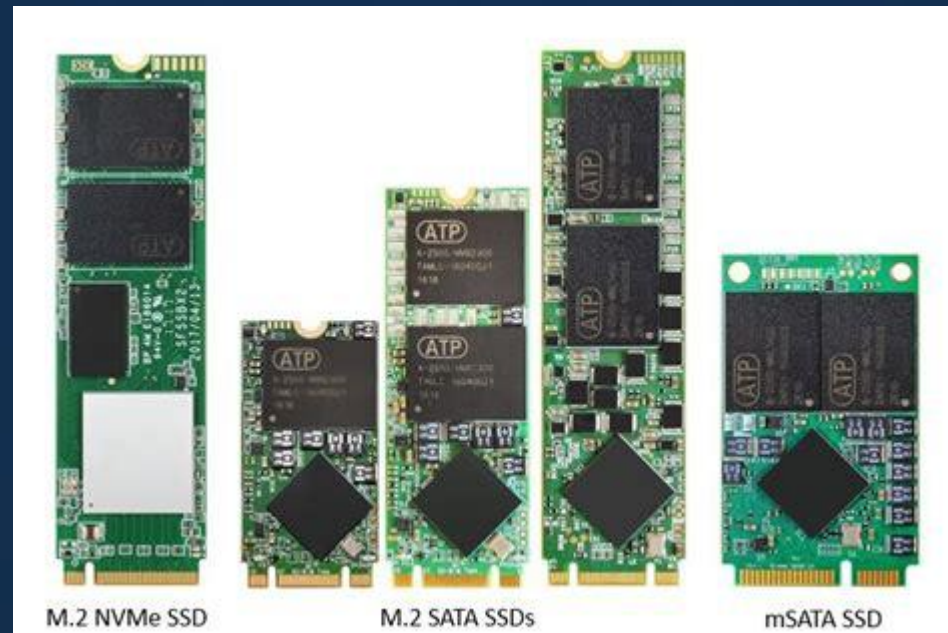
- 性能指标:
 - IOPS: 每秒的IO读写次数
 - 小文件存储(图片)、OLTP数据库、邮件服务器
 - Bandwidth: 每秒转速, 转速越快IO越快
 - 大量数据传输, 影音视频监控原始数据等
 - Latency: IO的反馈时间差
 - 主要由平均硬盘寻道时间和平均硬盘转动延迟组成



- 固态硬盘 (Solid State Drive, SSD)
 - 利用Flash芯片作为数据永久存储的硬盘
- 重要部件: NAND闪存、控制器及固件
 - NAND闪存负责重要的存储任务, 控制器和固件需要协作来完成管理数据存储、维护SSD性能和使用寿命等。
- 为什么固态硬盘厂商比机械硬盘厂商多?
 - 固态硬盘的零部件标准化, 封装过程要求低, 但需要验证
 - 机械硬盘的磁盘封装是原厂完成, 控制电路板占比不大



- 其他SSD接口形态
 - M.2 NVMe SSD
 - M.2 SATA SSD
 - mSATA SSD
- 内部环境与电气特性不支持热插拔
- 在数据中心，一般安装操作系统



- 内部 – 半高PCIe板卡

接口支持X2、X4速率

安装在服务器内部，更换不方便



- 外部 – 2.5英寸

外观兼容，背板实现兼容SAS/SATA

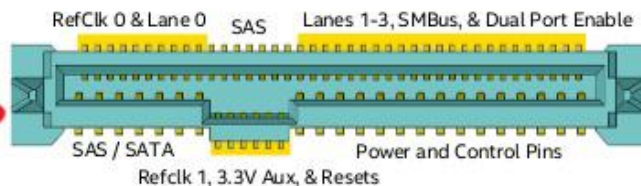
安装在服务器外部，在线热插拔

目前市场还未大量铺货

Drive Connectors

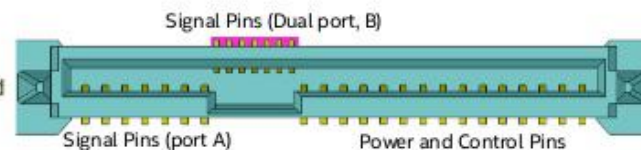
SFF-8639

- Supports SATA, SAS, and PCI Express* (PCIe*) x4 or two x2
- PCIe data, reference clock, and side band



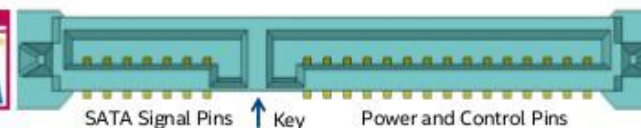
SAS

- Backwards compatible with SATA
- Dual port



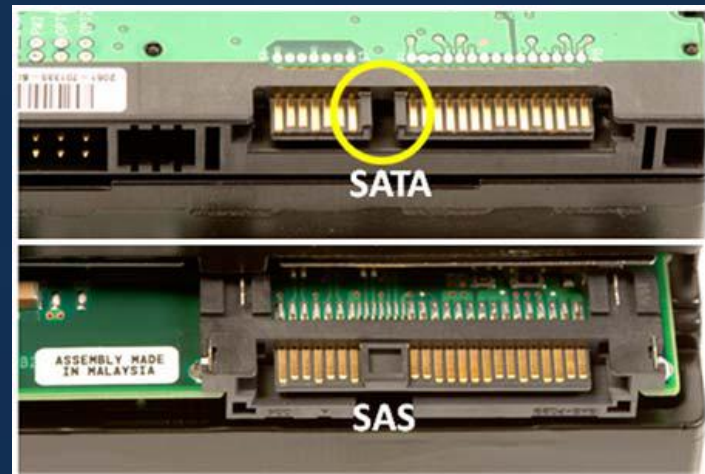
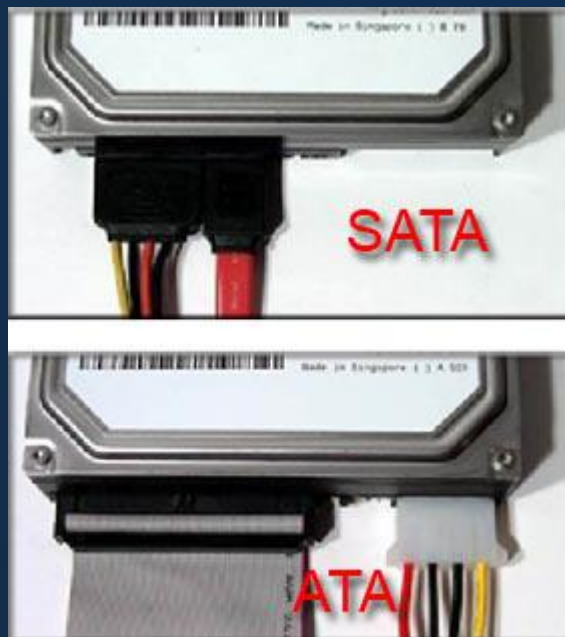
SATA

- Keyed only for SATA drives
- Separate power and data



IDF14

- ATA指令系统
 - IDE和SATA
- SCSI指令系统
 - 并行SCSI、串行SCSI (SAS)、NL-SAS
- 物理兼容
 - 物理上SAS兼容SATA
 - 接口上SAS主控芯片支持SATA



- SAS企业级（高可靠性） V.S. SATA消费级（廉价和大容量）

	300 GB 15K SAS	3000 GB 7.2K SATA	Difference
Seek Time	2 ms	4.17 ms	209% Higher
Latency - Reads	3.4 ms	9.5 ms	279% Higher
Latency - Writes	3.9 ms	10.5 ms	269% Higher
Sustained Bandwidth	204 MB/s	180 MB/s	22% Lower Bandwidth
IOPS per GB	0.64	0.05	1280% Fewer IOPS per GB
MTBF	1.6M	.8M	50% Less Reliable
Bit Error Rate	1.00E+17	1.00E+15	10,000% Worse

➤ **ATA:** 高级技术附件（Advanced Technology Attachment）是上世纪90年代桌面机标准。它不同于SCSI(一种接口技术)的一点是，通过它传输的数据，全部都要先进入RAM之后再进入CPU进行处理，之后再转存到存储里面。

➤ **SATA:**（Serial Advanced Technology Attachment）是ATA技术的升级版本，曾是桌面电脑ATA接口硬盘的主要替代技术。

因为容量大，价格便宜，在企业级服务器和存储系统中曾广泛得被使用，现在多被更加智能的NL-SAS接口的硬盘替代。

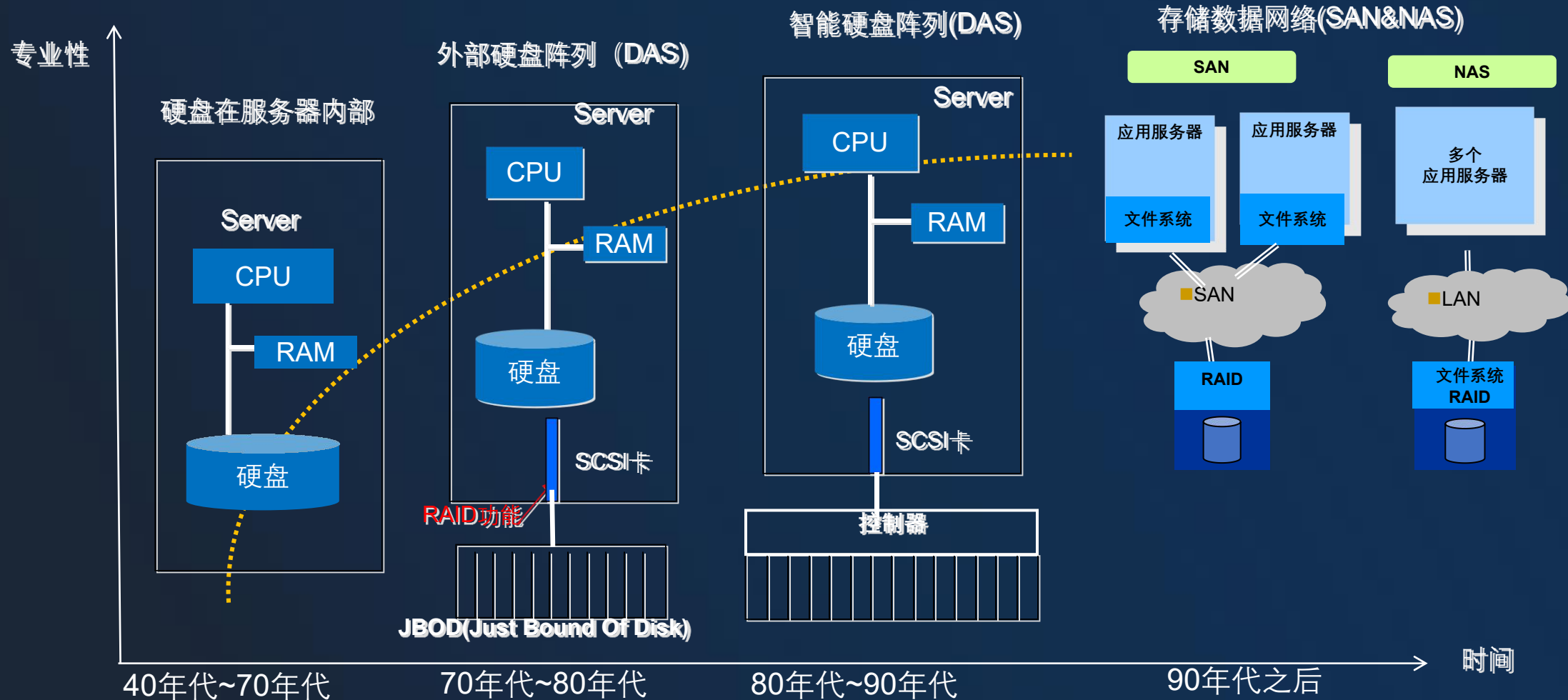
➤ **SCSI:**（Small Computer System Interface，小型计算机系统接口）是一种为小型机研制的接口技术，用于主机与外部设备之间的连接。**SCSI-3是所有存储协议的基础，其它存储协议都用到SCSI的指令集。**
优点：与主机无关、多设备并行、高带宽。缺点：允许连接设备数量少、连接距离非常有限。

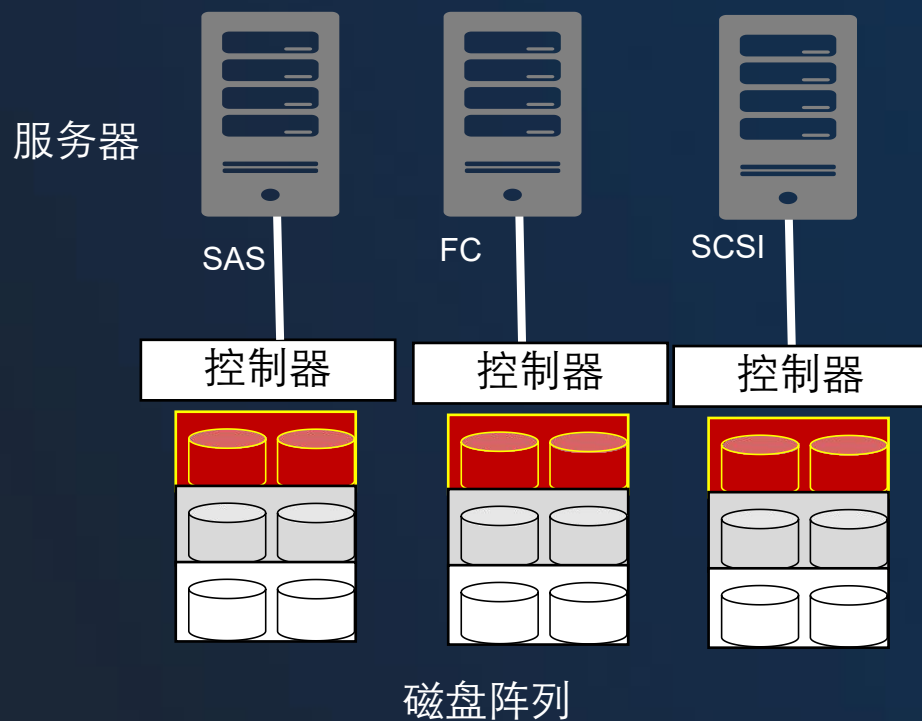
➤ **SAS:**（Serial Attached SCSI）即SCSI总线协议的串行标准，即串行连接SCSI；SAS采用串行技术以获得更高的扩充性，并兼容SATA盘。目前SAS的最高传输速率高达3Gpbs、6Gbps，支持全双工模式。

➤ **NL-SAS:** 使用的是SAS的接口，盘体使用的是SATA盘体。从而提高了传输速度，也增大了容量。

➤ **FC:** FC是光纤通道（Fiber Channel）的简称，用于服务器与共享存储设备的连接，存储控制器和驱动器之间的内部连接，是一种高性能的串行连接标准。其接口传输速率目前有4Gbps、8Gbps、16Gbps几种标准。传输介质可以选择铜缆或光纤，传输距离远，支持多种互联拓扑结构。光纤通道是构建FC SAN的基础，是FC SAN系统的硬件接口和通信接口。

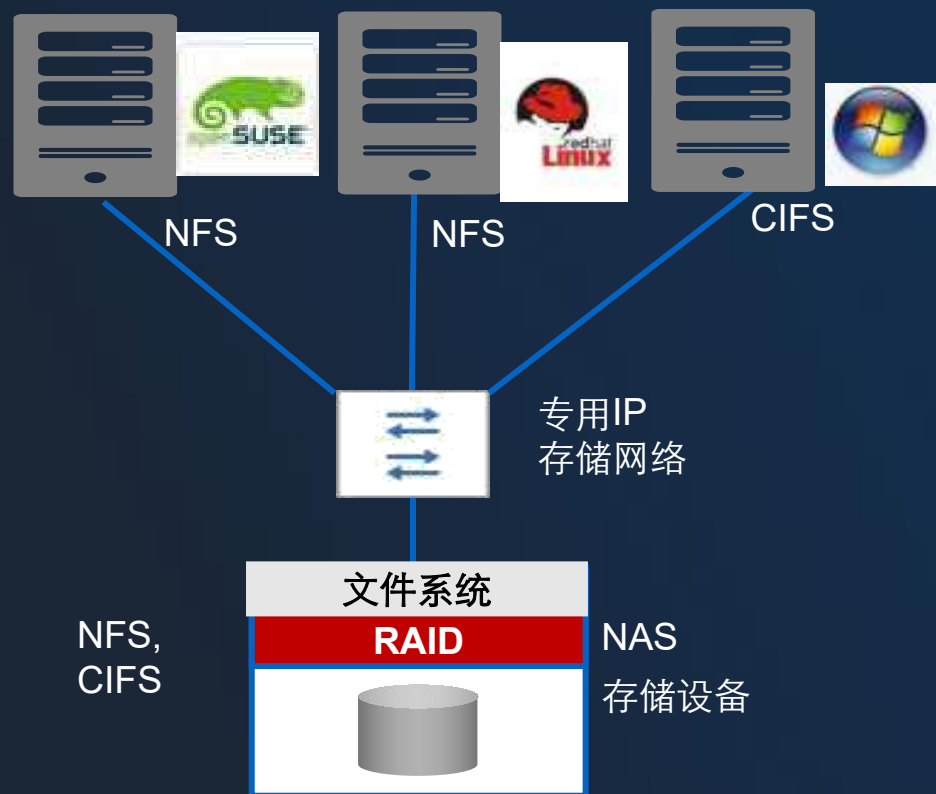
➤ **iSCSI:** （Internet Small Computer System Interface）互联网小型计算机系统接口，是一种在TCP/IP上进行数据块传输的标准，可以理解为SCSI over IP。iSCSI可构成基于IP的SAN，为用户提供高速、低价、长距离的存储解决方案。iSCSI将SCSI命令封装到TCP/IP数据包中，使I/O数据块可通过IP网络传输，是未来的发展之路。





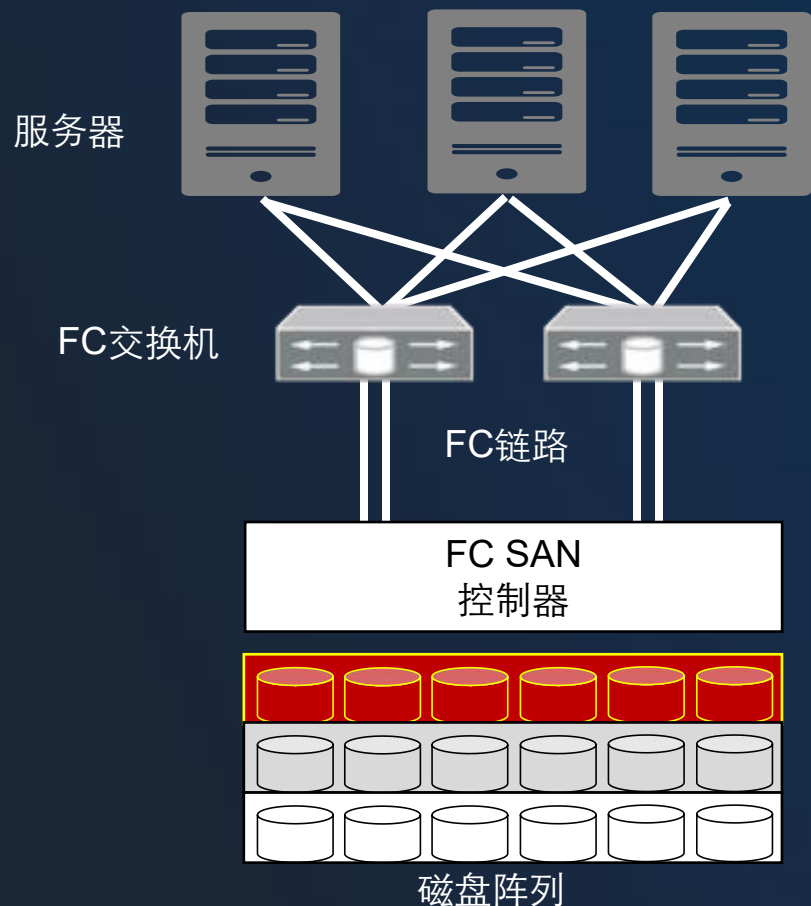
- **DAS (Direct Attached Storage)**
- 时间：70年代
- 背景：用户最早因为数据量的增多而产生存储的需求，从而产生最早最简单的存储架构直连附加存储DAS
- 连接方式：**FC,SCSI,SAS**
- 访问方式：直连式存储与服务器主机之间的连接通道通常采用SCSI连接
- 链路速率：20MB/s、40MB/s、80MB/s
- 提供快照、备份等功能

NAS应用系统架构

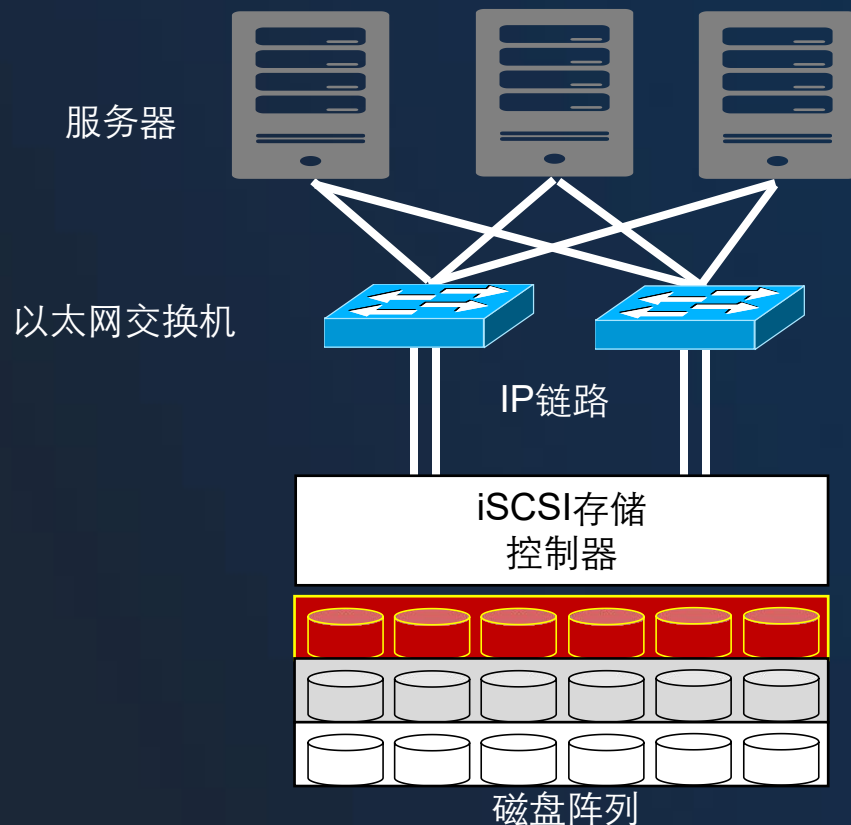


NAS(Network Attached Storage)

- 时间：90年代初
- 背景：网络飞速发展，大量数据需要共享和交换，出现专用的NAS存储设备，成为数据共享与交换的核心
- 访问方式：多台前端服务器共享后端存储设备，后端NAS设备上的存储空间通过CIFS（window系统）、NFS（Linux系统）协议共享给前端主机，可同时对同一目录或文件进行并发读写
- 文件系统位于后端存储设备
- 链路速率：1Gbps、10Gbps



- FC SAN (Fiber Channel Storage Area Network)
- 时间：90年代中后期
- 背景：为解决DAS扩展性差的问题，将存储设备网络化，可以同时连接上百台服务器
- 连接方式：FC光纤，使用专用的FC交换机
- 访问方式：后端一台存储设备的存储空间可以划分为多个LUN，每一个LUN只能属于一台前端服务器
- 链路速率：2Gbps、4Gbps、8Gbps、16Gbps
- 提供快照、容灾等高级数据保护功能



- IP SAN (IP Storage Area Network)

- 时间：2001年
- 背景：为解决FC-SAN在价格及管理上的诸多门坎而产生
- 连接方式：采用以太网作为连接链路，以太网交换机
- 访问方式：后端一台存储设备的存储空间可以划分为多个LUN，每一个LUN只能属于一台前端服务器
- 链路速率：1Gbps、10Gbps
- 提供快照、容灾等高级数据保护功能
- iSCSI 被看好的原因
 - 可以采用非常成熟的IP网络管理工具和基础建设
 - IP网络使用普遍，可为企业节省大笔建设、管理及人事成本

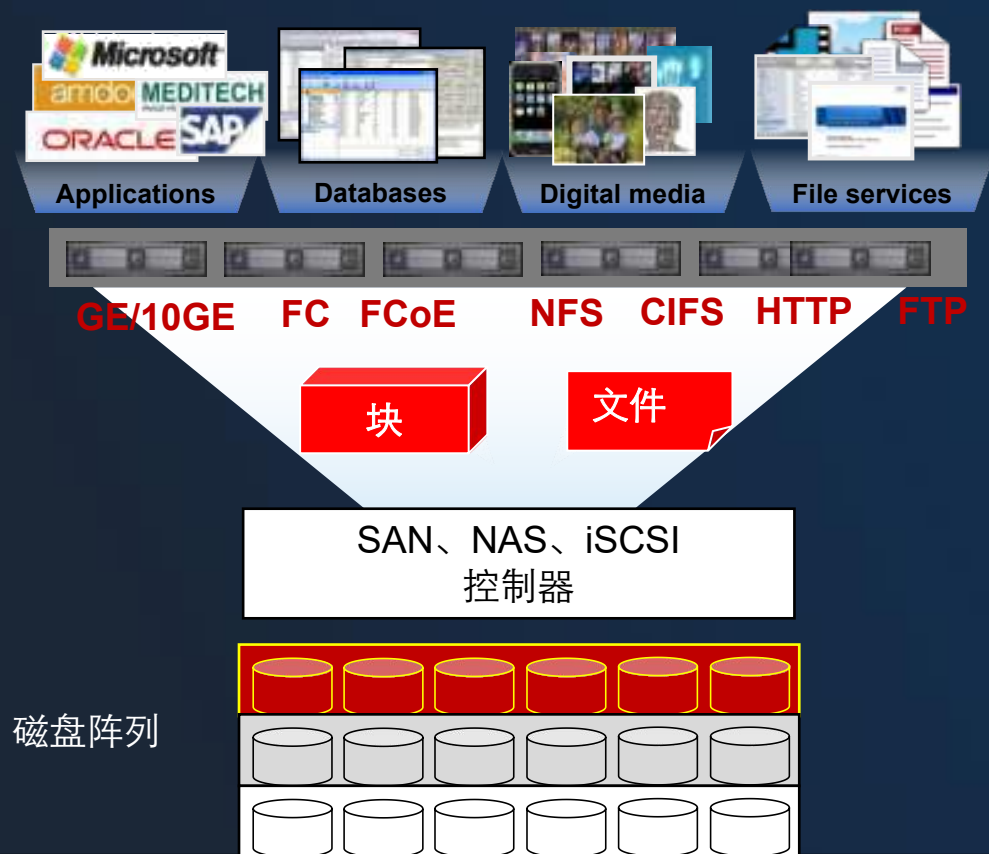
三种存储形态，各有优劣

	DAS	NAS	FC-SAN	IP-SAN
传输类型	SCSI、FC、SAS	IP	FC	IP
数据类型	块级	文件级	块级	块级
典型应用	任何	文件服务器	数据库应用	视频监控
优点	易于理解 兼容性好	易于安装 成本低	高扩展性、高性能 高可用性	高扩展性 成本低
缺点	难管理，扩展性有限 存储空间利用率不高	性能较低 对某些应用不适合	较昂贵，配置复杂 互操作性问题	性能较低

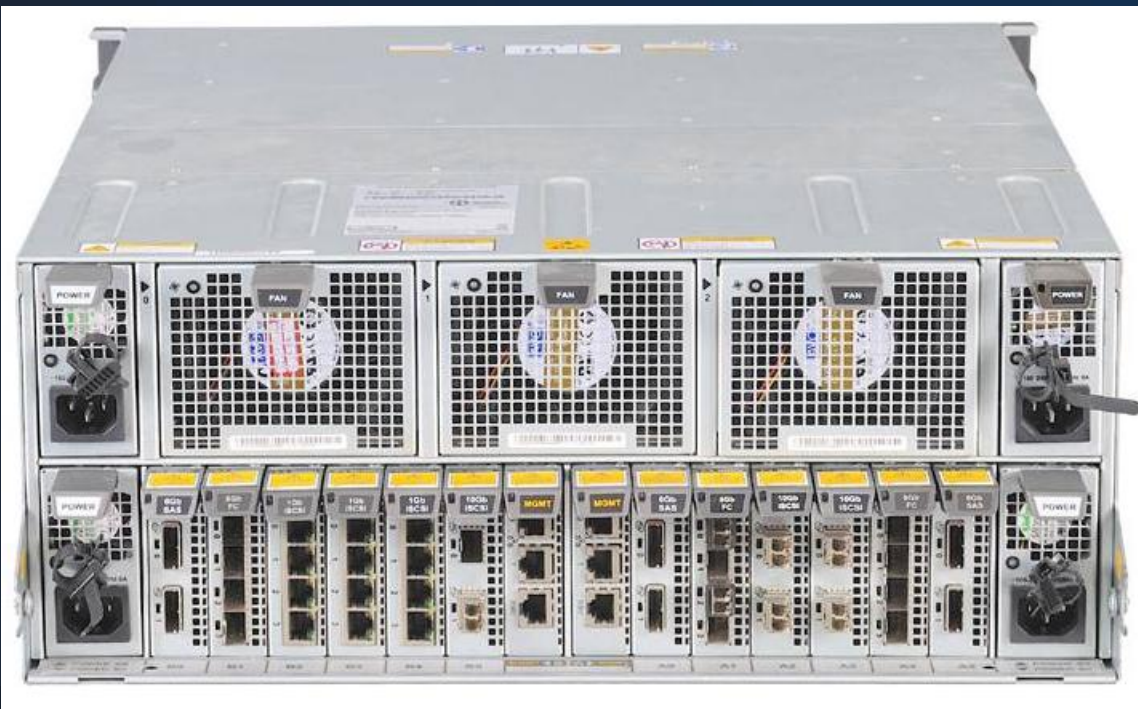
SAN与NAS并不是两种互相竞争的技术，二者通常相互补充以提供对不同类型数据的访问。SAN针对**海量的面向数据块的数据传输**，而NAS则提供**文件级的数据访问和共享服务**。越来越多的数据中心采用SAN+NAS的方式实现数据整合、高性能访问以及文件共享服务。

NAS和SAN的融合——统一存储

统一存储：一种网络存储架构，它既支持基于文件的NAS存储，又支持基于块的SAN存储。



- 时间：21世纪初
- 背景：为解决多种存储架构带来的管理、高TCO问题，支持多种传输协议，同时满足不同需求的统一存储设备应运而生
- 同时支持不同传输协议，以单一系统满足企业各式各样的存储需求
- 统一的管理界面，简化系统管理
- 前后端均为系统内部链接，整体运行效率高
- 模块化的架构，灵活配置，成本可控；轻松扩容，保护用户投资，降低整体TCO



RAID（Redundant Array of Independent Disks）即独立磁盘冗余阵列，RAID技术将多个单独的物理硬盘以不同的方式组合成一个逻辑硬盘，从而提高了硬盘的读写性能和数据安全性。

■根据不同的组合方式可以分为不同的RAID级别：

RAID 0	数据条带化，无校验
RAID 1	数据镜像，无校验
RAID 3	数据条带化读写，校验信息存放于专用硬盘
RAID 5	数据条带化，校验信息分布式存放
RAID 6	数据条带化，分布式校验并提供两级冗余

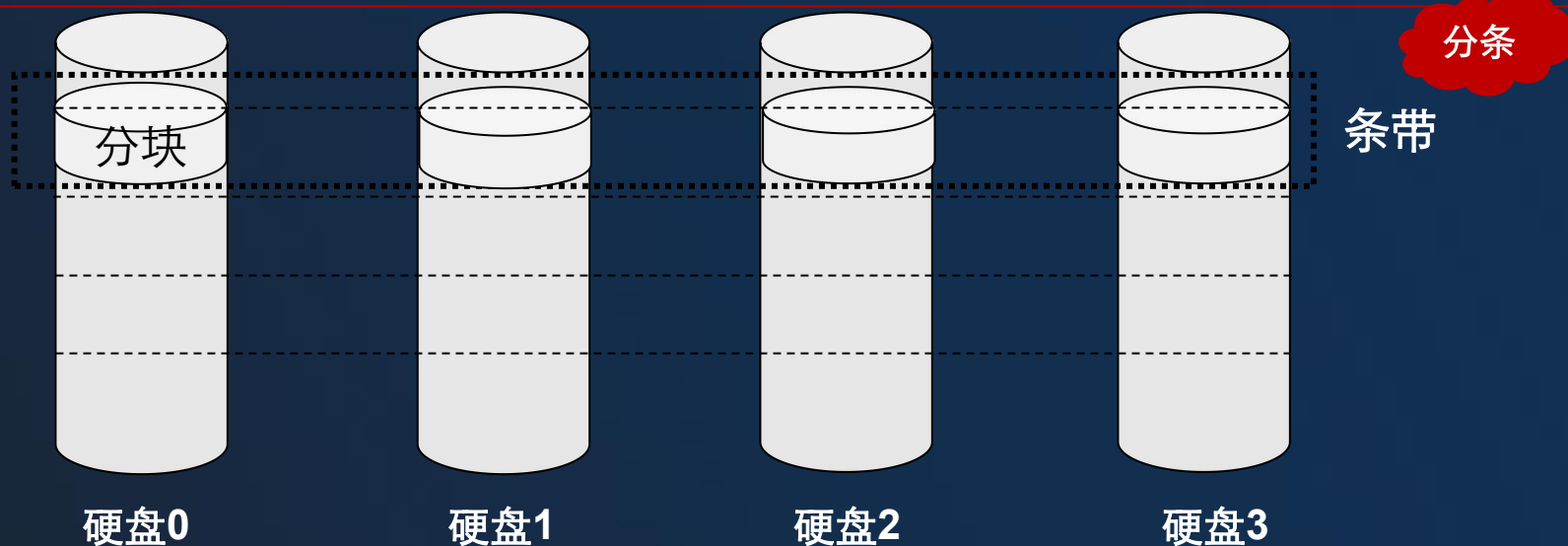
■同时采用两种不同的RAID方式还能组合成新的RAID级别：

RAID 0+1	先做RAID 0，后做RAID 1，同时提供数据条带化和镜像
RAID 10	类似于RAID 0+1，区别在于先做RAID 1，后做RAID 0
RAID 50	先做RAID 5，后做RAID 0，能有效提高RAID 5的性能

数据组织形式

分块：将一个分区分成多个大小相等的、地址相邻的块，这些块称为分块。它是组成条带的元素。

条带：同一磁盘阵列中的多个磁盘驱动器上的相同“位置”（或者说是相同编号）的分块。

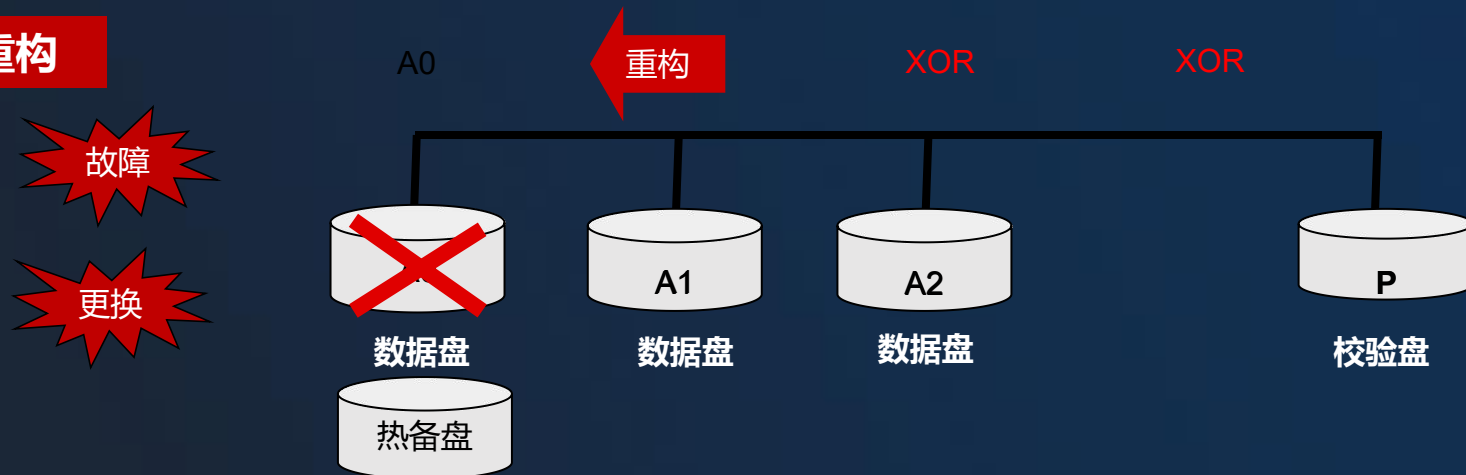


4 RAID基本概念——热备、重构

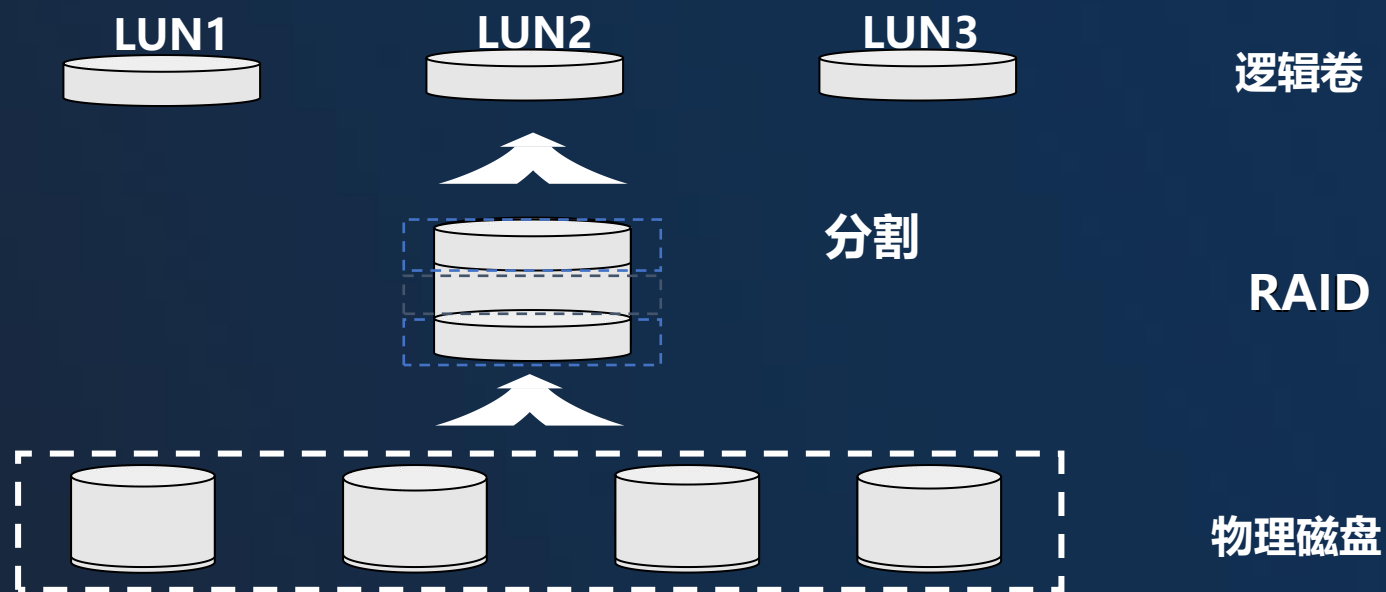
热备

- **热备 (HotSpare) 的定义:**
当冗余的RAID组中某个硬盘失效时, 在不干扰当前RAID系统的正常使用的前提下, 用RAID系统中另外一个正常的备用硬盘自动顶替失效硬盘, 及时保证RAID系统的冗余性。
- **热备一般分为两种:**
 - 全局式: 备用硬盘为系统中所有的冗余RAID组共享
 - 专用式: 备用硬盘为系统中某一组冗余RAID组专用

重构



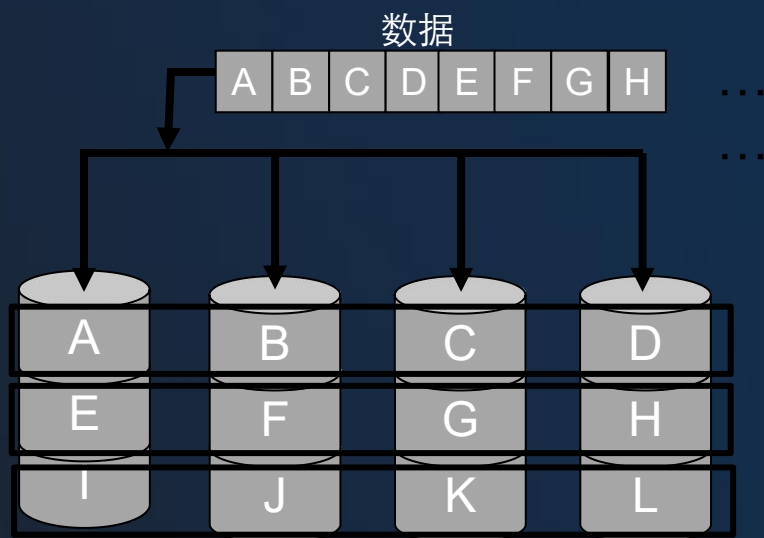
在RAID的基础上可以按照指定容量创建一个或多个逻辑卷，通过LUN（Logic Unit Number）来标识



4 RAID级别——RAID0, RAID1

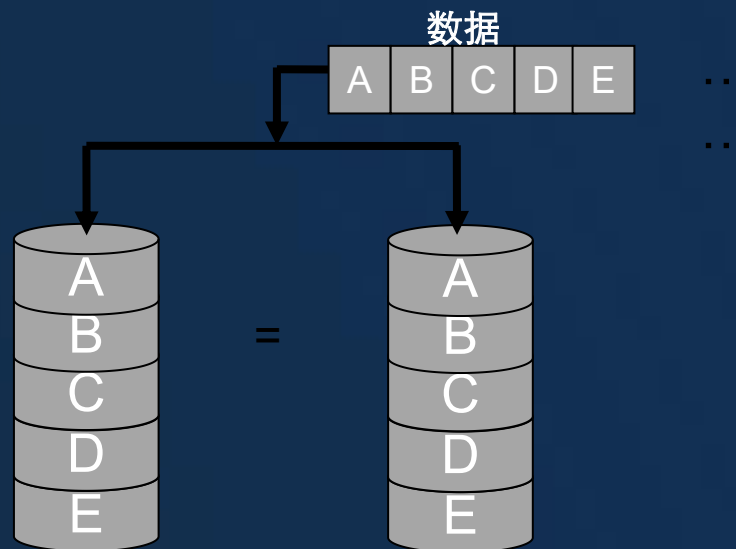
RAID0

没有容错设计的条带硬盘阵列，以条带形式将RAID组的数据均匀分布在各个硬盘中



RAID1

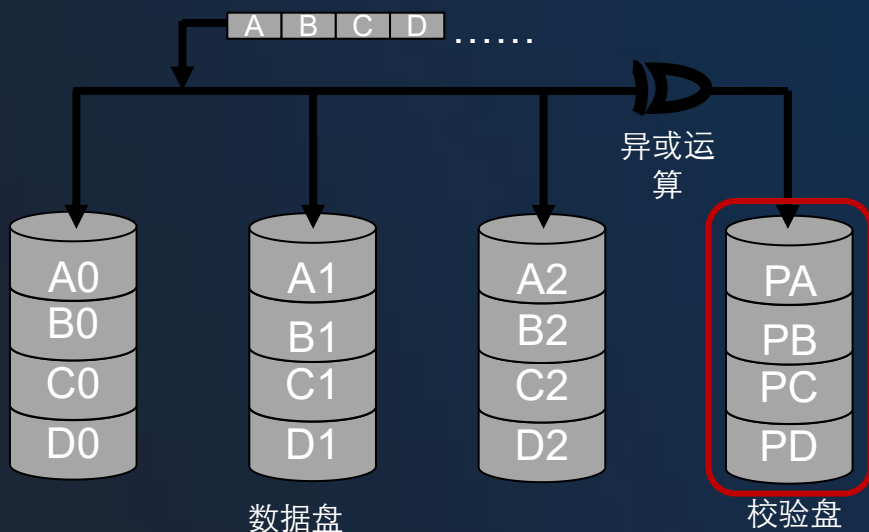
又称镜像 (Mirror)，数据同时一致写到主硬盘和镜像硬盘



4 RAID级别——RAID3, RAID5

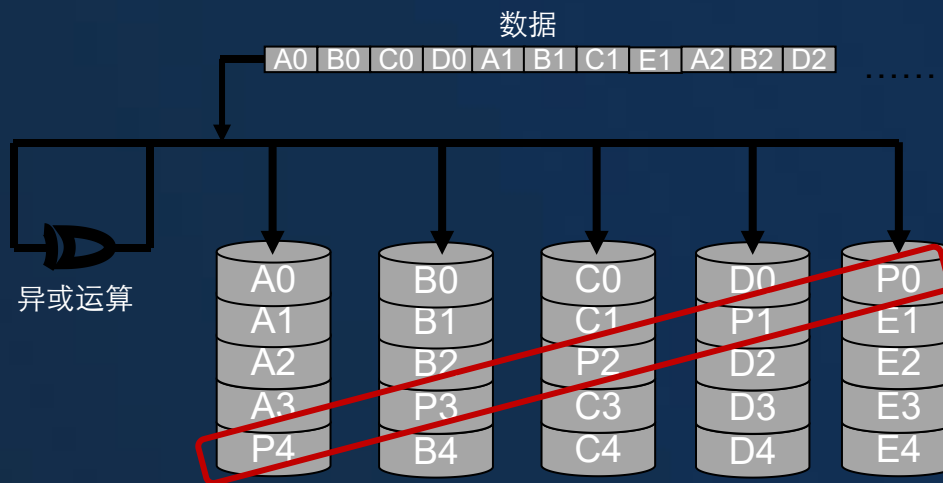
RAID3

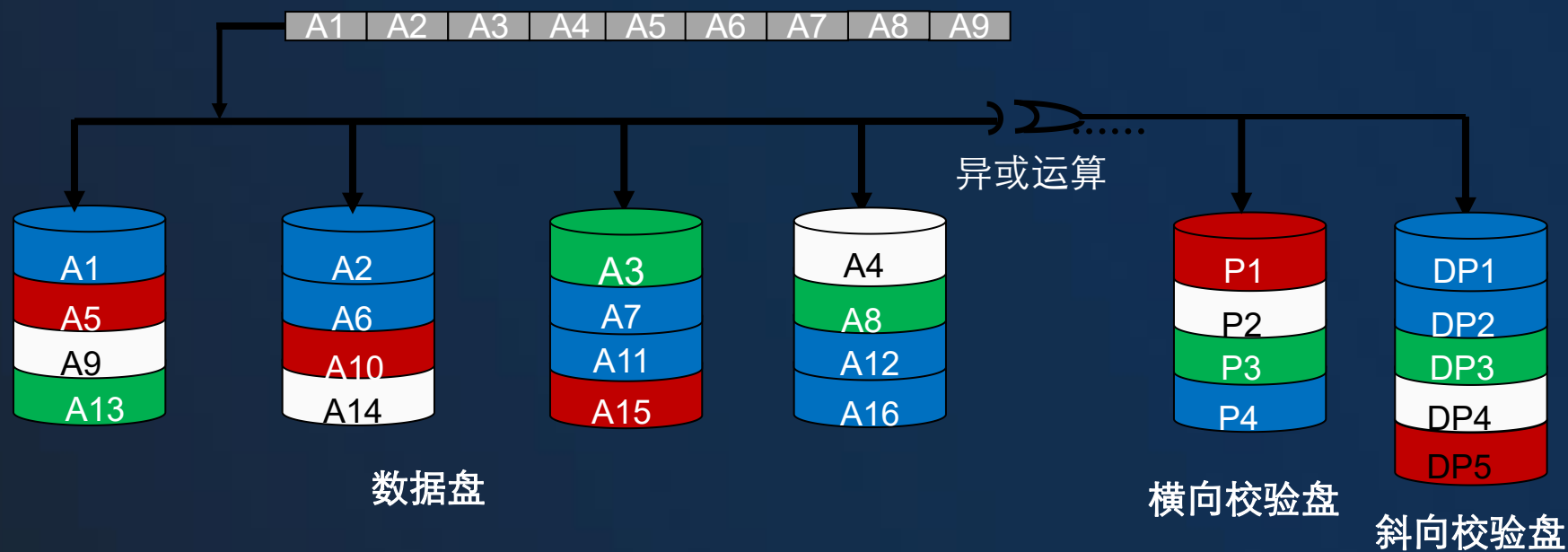
带有校验的并行数据传输阵列，数据条带化分布在数据盘中，同时使用**专用校验硬盘**存放校验数据



RAID5

与RAID 3机制类似，但校验数据均匀分布在各数据硬盘上，RAID成员硬盘上同时保存数据和校验信息，数据块和对应的校验信息保存在**不同硬盘**上。RAID 5是最常用的RAID方式之一





横向校验盘：P1—P4为各个数据盘中横向数据的校验信息

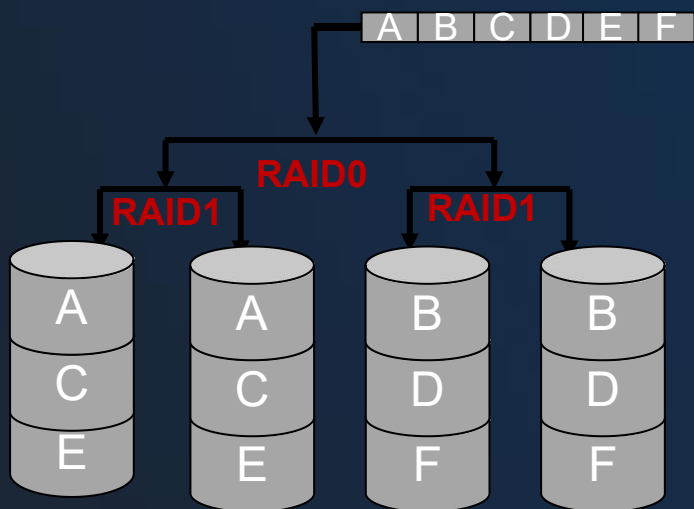
例： $P1 = A1 \text{ XOR } A2 \text{ XOR } A3 \text{ XOR } A4$

斜向校验盘：DP1—DP4为各个数据盘及横向校验盘的斜向数据的校验信息

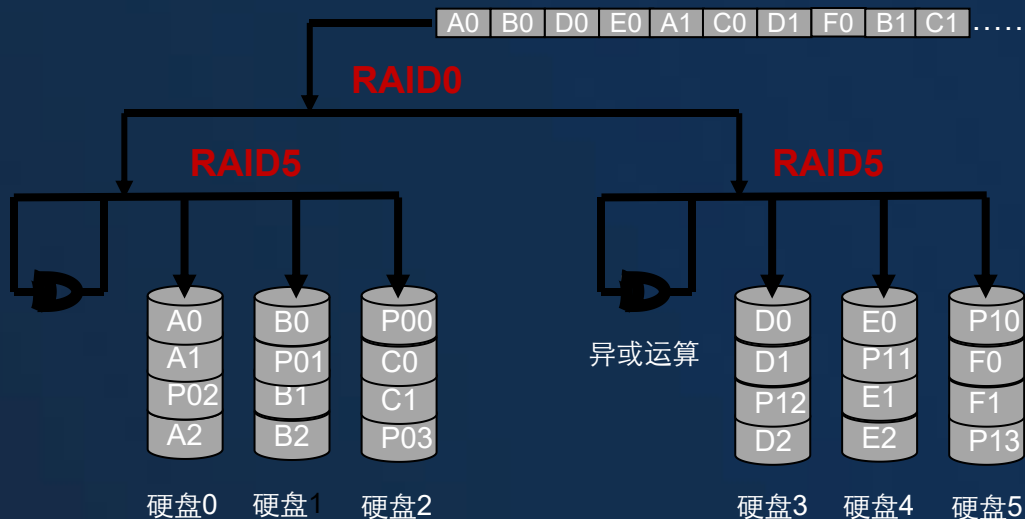
例： $DP1 = A1 \text{ XOR } A6 \text{ XOR } A11 \text{ XOR } A16$

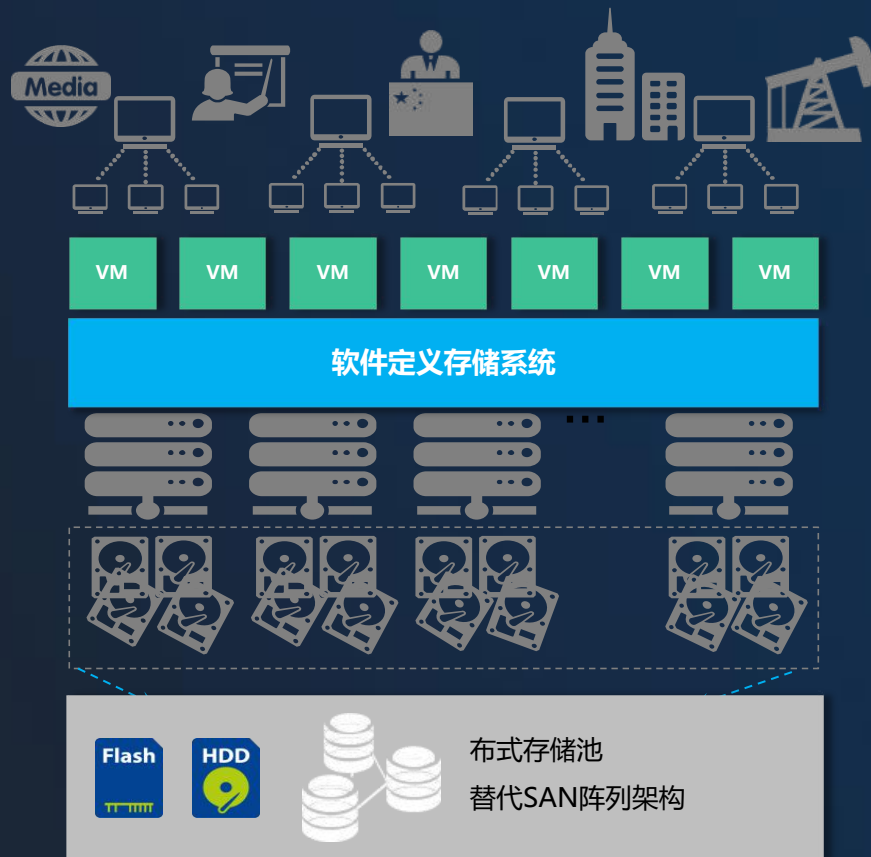
RAID10

将镜像和条带进行两级组合的RAID级别，第一级是**RAID1**镜像对，第二级为**RAID0**。RAID10也是一种应用比较广泛的RAID级别

**RAID50**

将RAID5和RAID0进行两级组合的RAID级别，最低一级是**RAID5**，第二级为**RAID0**





基于标准X86的软件定义存储解决方案

聚合每台主机本地硬盘形成集群存储池

经过闪存加速的高性能存储系统

分布式RAID结构高保障系统高可靠性

可与虚拟化集成形成超融合部署

提供统一存储服务

极简易

- 自动化部署，开机引导
1分钟快速上线
- 配置管理简单，低门槛
- 资源监控信息丰富直观，
方便运维
- 超融合产品核心模块

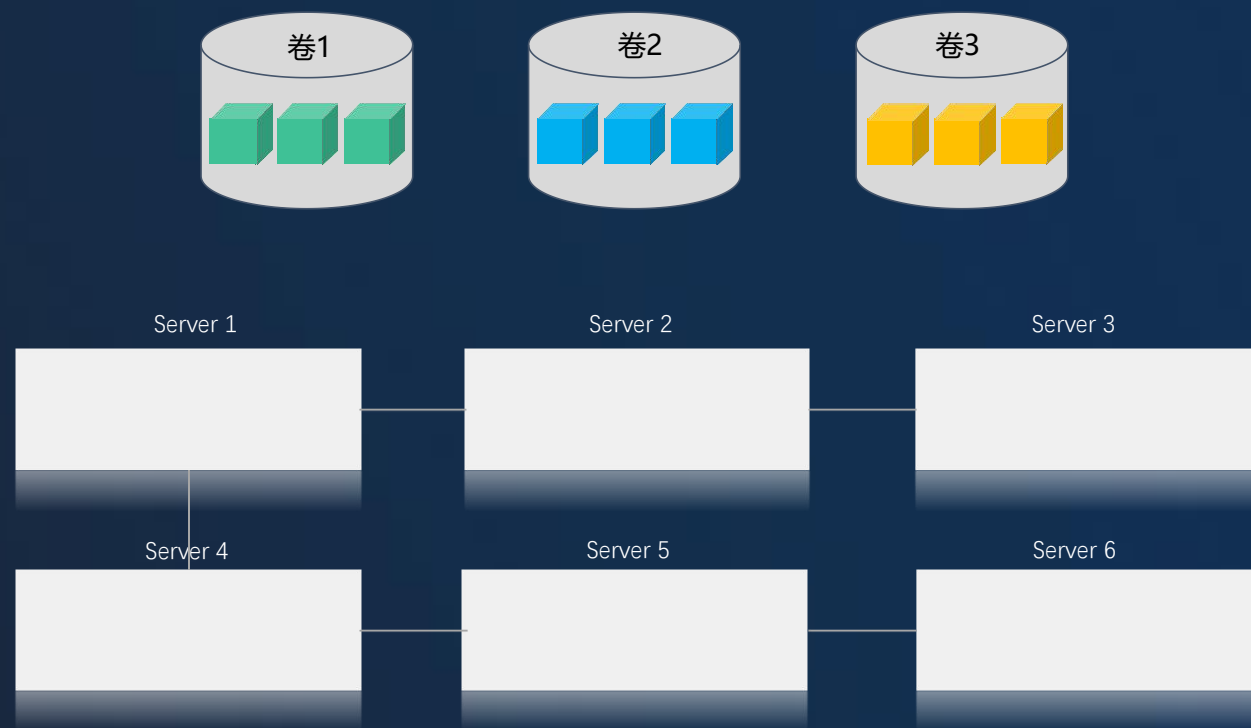
高可靠

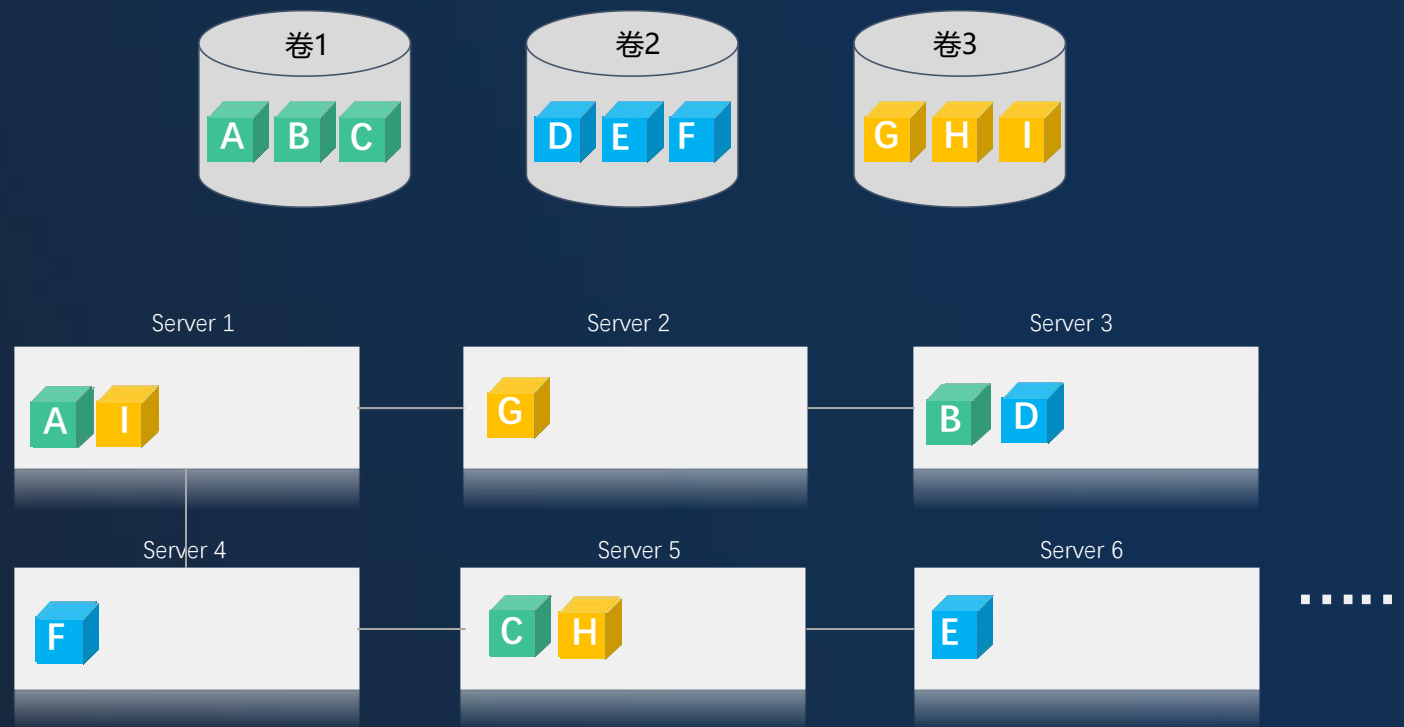
- 自动化的智能修复技术，
硬件故障自动恢复健康
- 副本机制保障节点级故障
不影响数据
- 故障域隔离保障机柜级宕
机不影响数据

低TCO

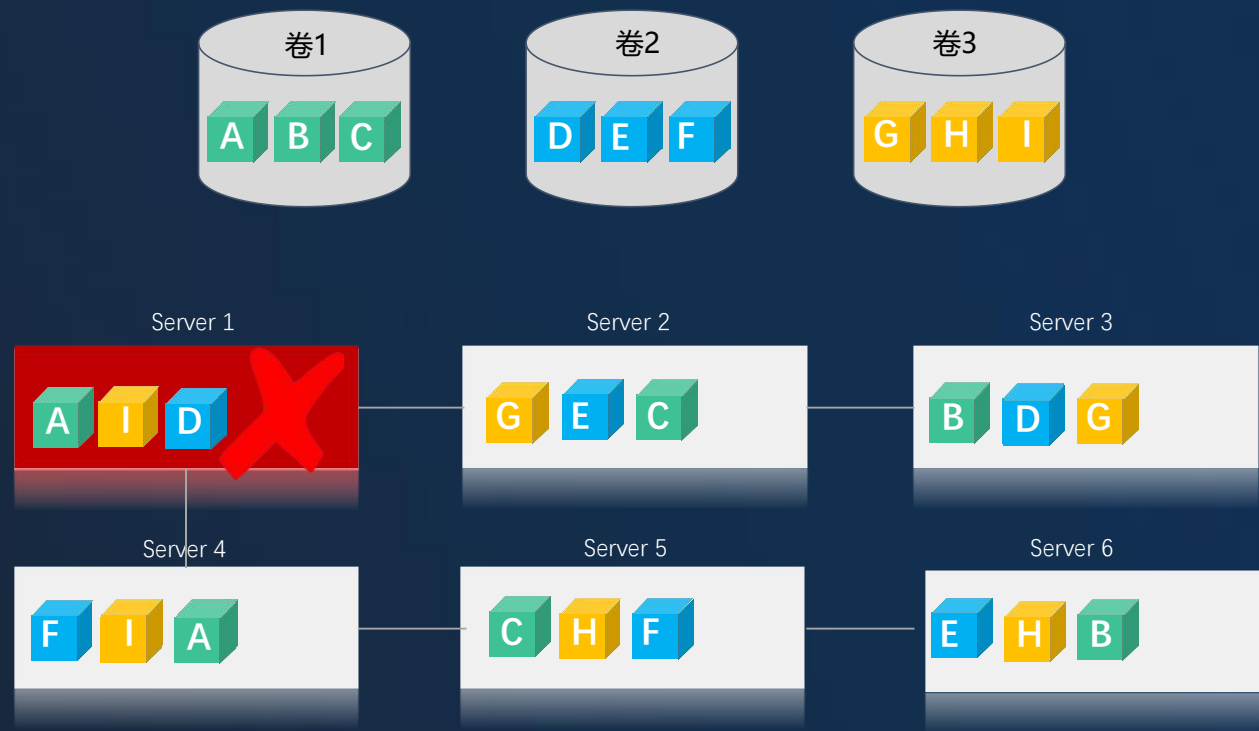
- 基于标准X86服务器，无
需专门的存储硬件
- 可随需求灵活扩展，性
能线性增加
- 大大降低前期投入，按
需采购即可

基于Ceph开发，业界使用最广泛的软件定义存储架构

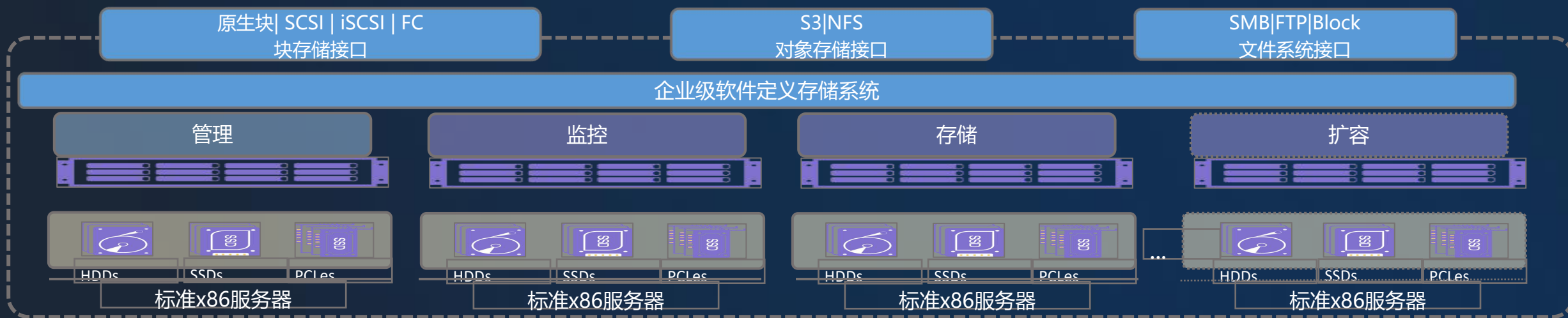




多副本存储技术类似RAID，数据存多份，保障节点级宕机时的数据可靠性



当有节点发生故障，数据可自动进行重构，保障存储集群的数据完整性



- 有几种存储组网形态？
- 每种存储形态的特点是什么？
- 常用的RAID方式有哪些？
- RAID1, RAID5, RAID10各自最少需要几块盘？
- 什么是SDS？
- SDS怎么保障数据可靠性？

01

云管理

- 1-5 云管理概念
- 1-6 云管理类型

02

云计算安全

- 2-1 云计算安全挑战
- 2-2 云计算安全方案

03 ZStack云平台

- 3-1 技术架构
- 3-2 计算管理
- 3-3 存储管理
- 3-4 网络管理
- 3-5 其他功能模块

ZStack

让每一家企业
都拥有自己的云



ZStack 微信公众号
zstack_io



ZStack 中国社区QQ群
443027683



ZStack 官网
www.zstack.io