云存储应用技术

第二章：存储技术基础

丁烨

dingye@dgut.edu.cn

网络空间安全学院

2019-09-12

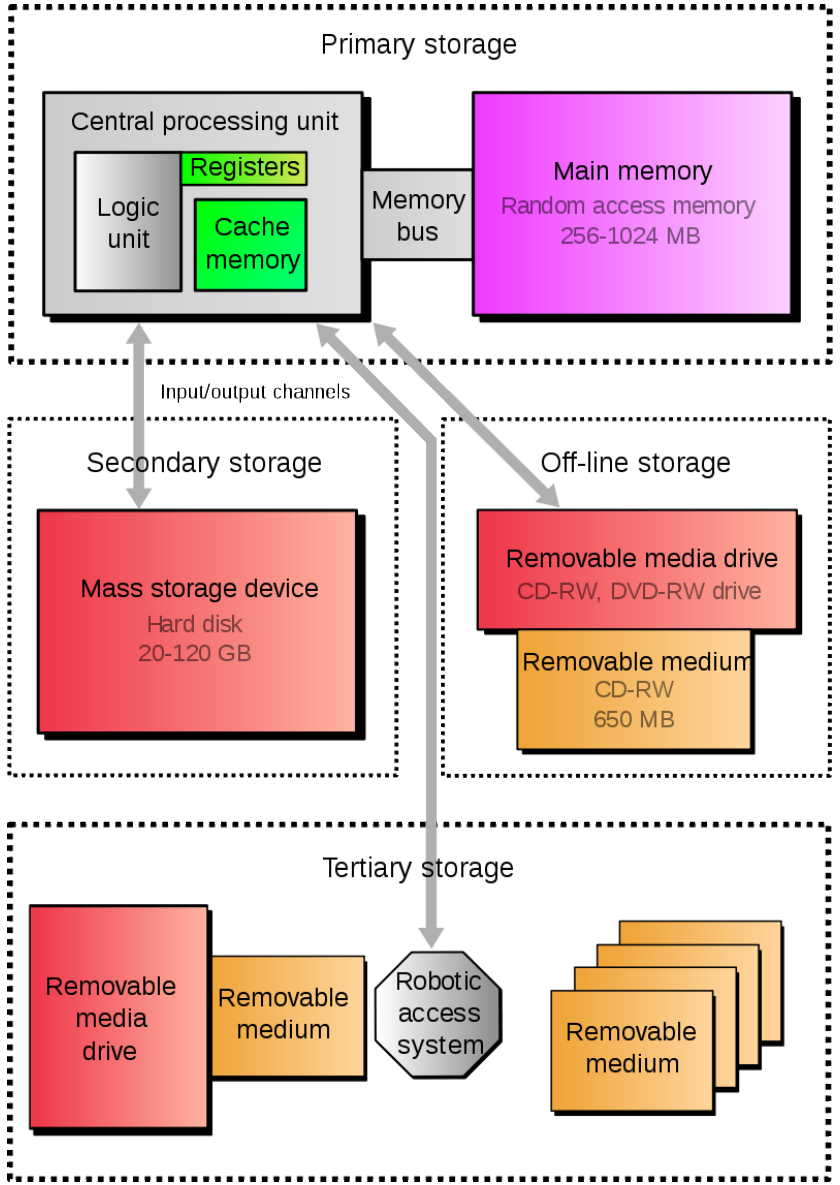
東莞理工學院
DONGGUAN UNIVERSITY OF TECHNOLOGY

目录

直连存储（DAS）

直连存储（DAS）的概念

- 计算机存储
- 计算机存储主要分为四类：
- 一级存储：与 CPU 直接连通，CPU 会不断读取存储在这里的指令集，并在需要时运行这些指令集。例如：内存
- 二级存储：和 CPU 没有直接连通，而是使用 I/O 通道来连接，并使用 Cache 将数据发送至一级存储。例如：机械硬盘、固态硬盘
- 三级存储：可直接插入或自计算机拔除的大规模存储设备，例如：磁带
- 离线存储：需要人工操作才可以访问的存储设备，例如：光盘、U 盘



直连存储（DAS）

直连存储（DAS）的概念

- 直连式存储（Direct-Attached Storage，DAS）
- 指直接和计算机相连接的数据储存方式
- 固态硬盘、机械硬盘、光盘等与计算机直接相连的设备都属于直连式存储设备
- 通常来说，二级、三级存储都属于 DAS
- 不通过网络传输的离线存储（例如 U 盘）也通常属于 DAS

直连存储（DAS）

直连存储（DAS）的概念

- 主要优点：
- 简单、直接、快速

- 主要缺点：
- 可控制的存储容量是很有限的
- 一个存储设备只能由一台计算机访问

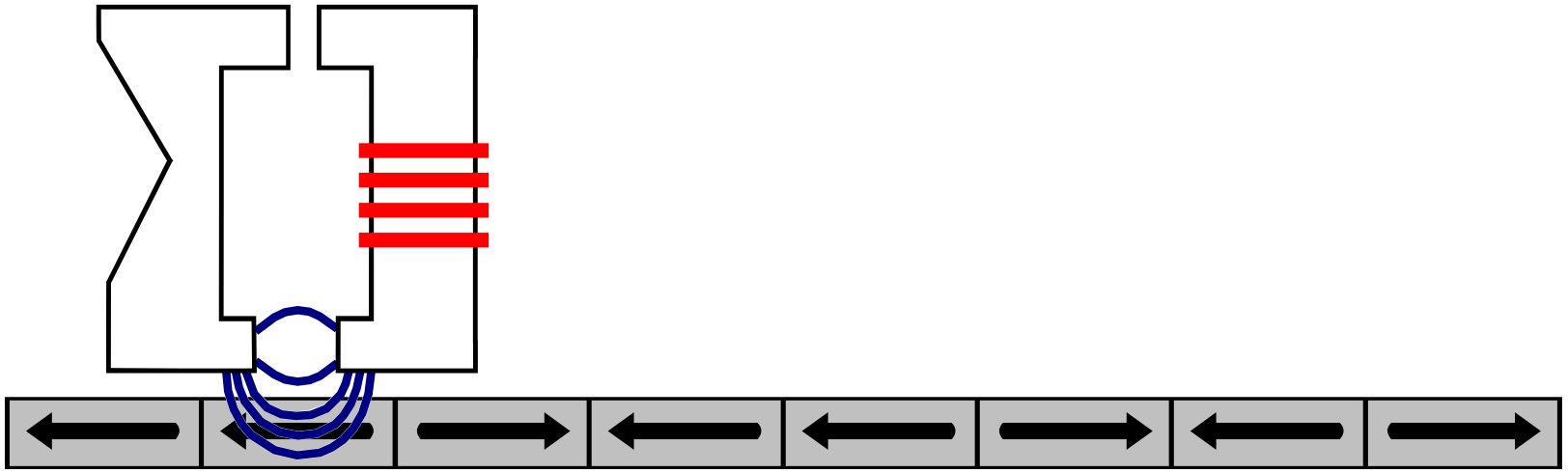
机械硬盘（HDD）

- 机械硬盘（Hard Disk Drive，HDD）
- 1956 年由 IBM 的 Rey Johnson 研发成功并逐步开始使用
- 在 1960 年代初成为通用式计算机中主要的辅助存放设备
- 随着技术的进步，硬盘也成为服务器及个人计算机的主要组件

直连存储（DAS）

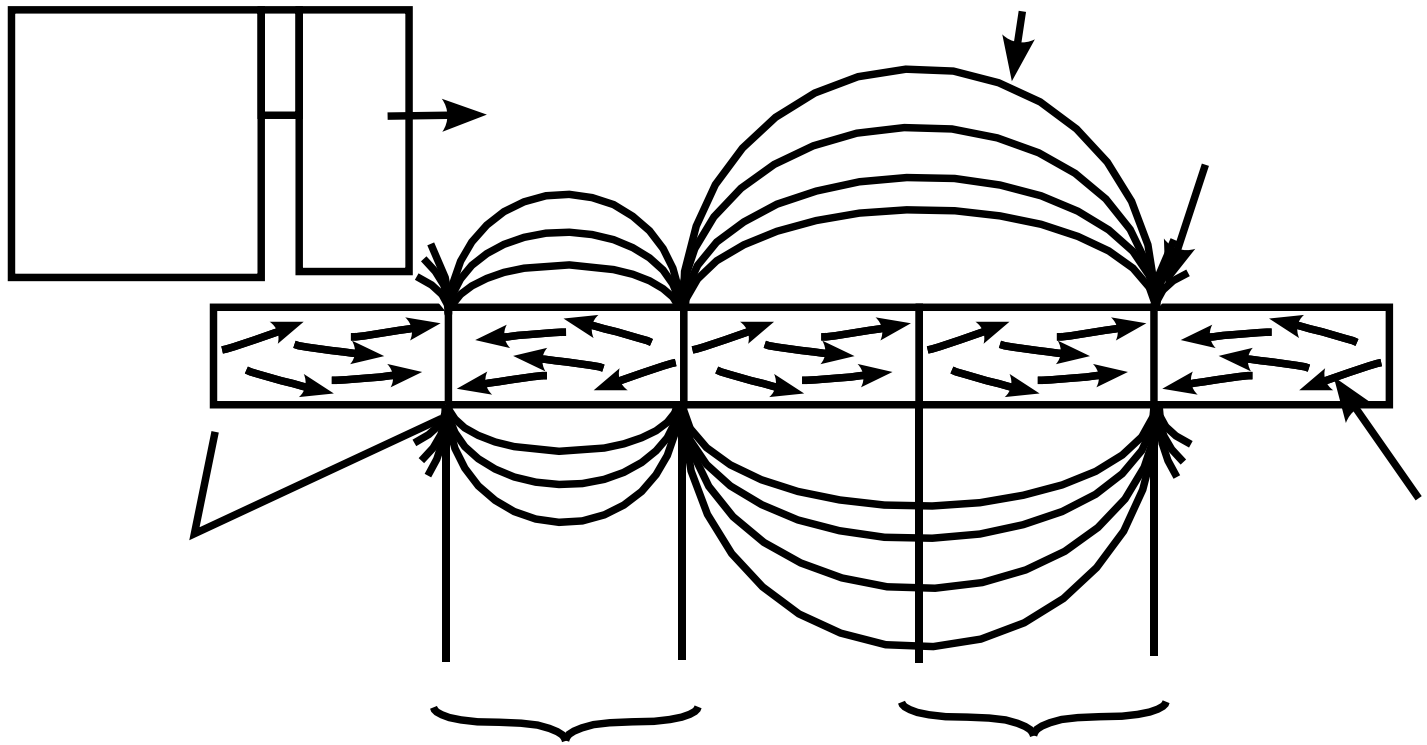
机械硬盘（HDD）

- 机械硬盘的基本原理：电磁感应



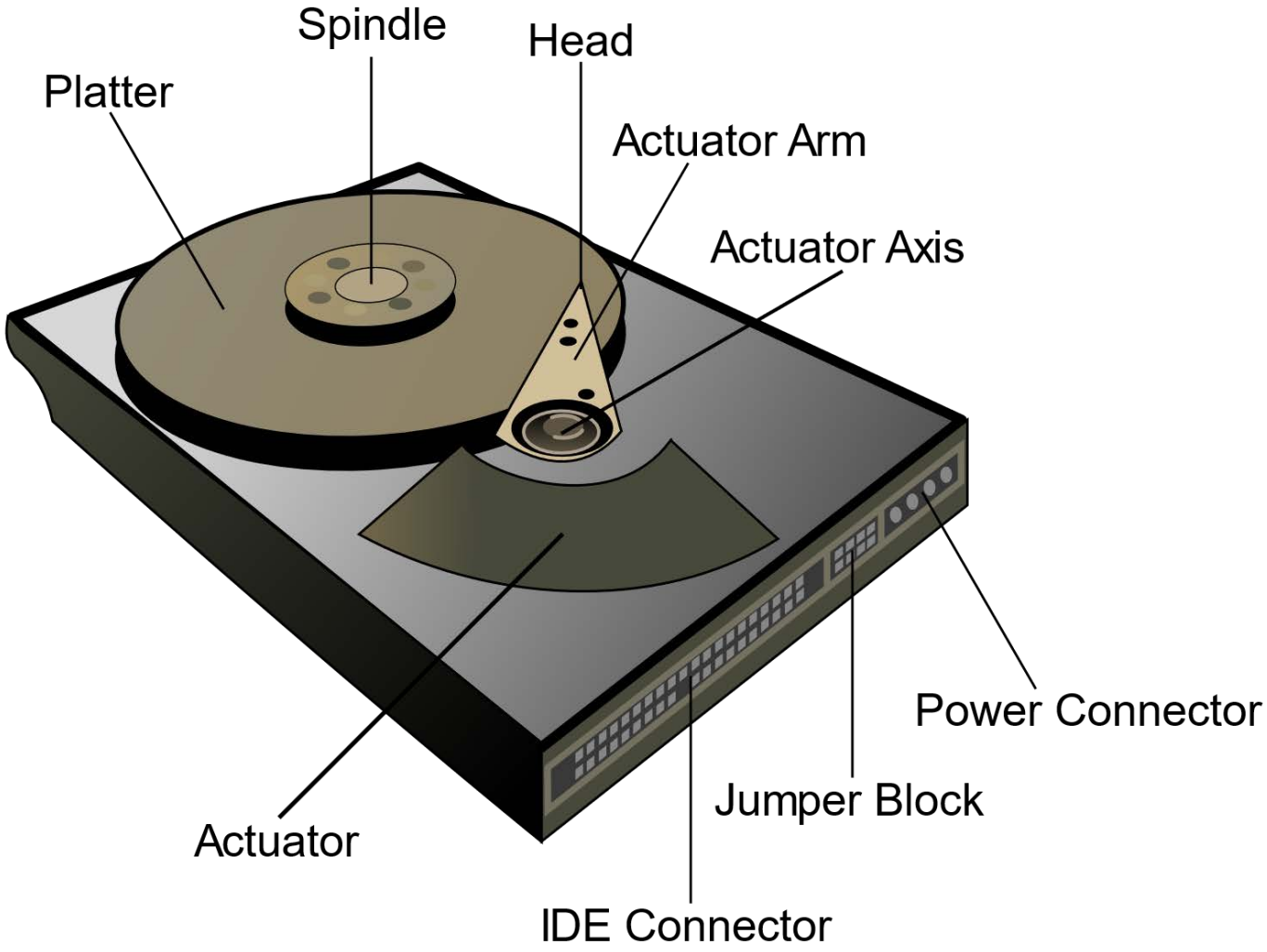
机械硬盘（HDD）

❖ 机械硬盘的基本原理：电磁感应

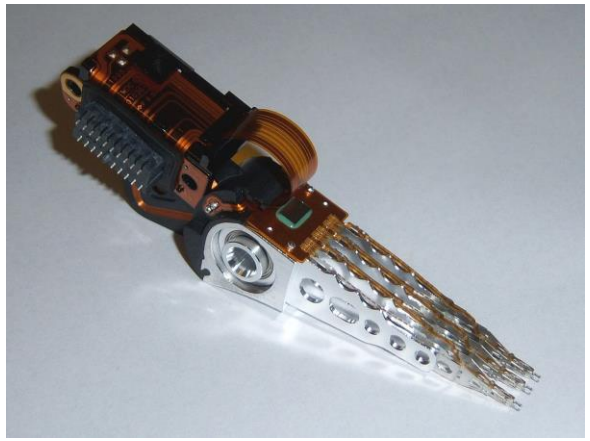


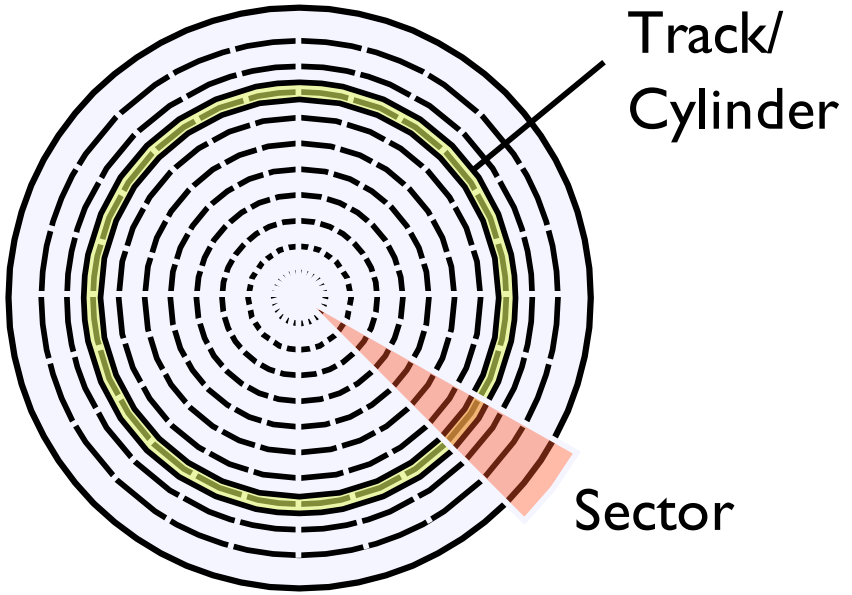
机械硬盘（HDD）

- 机械硬盘的结构
- Platter：盘片
- Spindle：主轴
- Actuator：步进电动机
- Actuator Head：磁头
- Actuator Arm：传动臂
- Actuator Axis：驱动轴
- Power Connector：电源接口
- Data Connector：数据接口
- Jumper Block：跳线

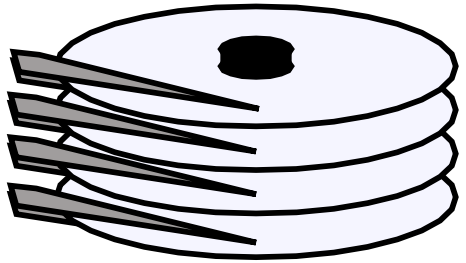


- 机械硬盘的结构
- 磁道（Track）
- 柱面（Cylinder）
- 扇区（Sector）
- 磁头（Heads）
- 盘片（Platters）



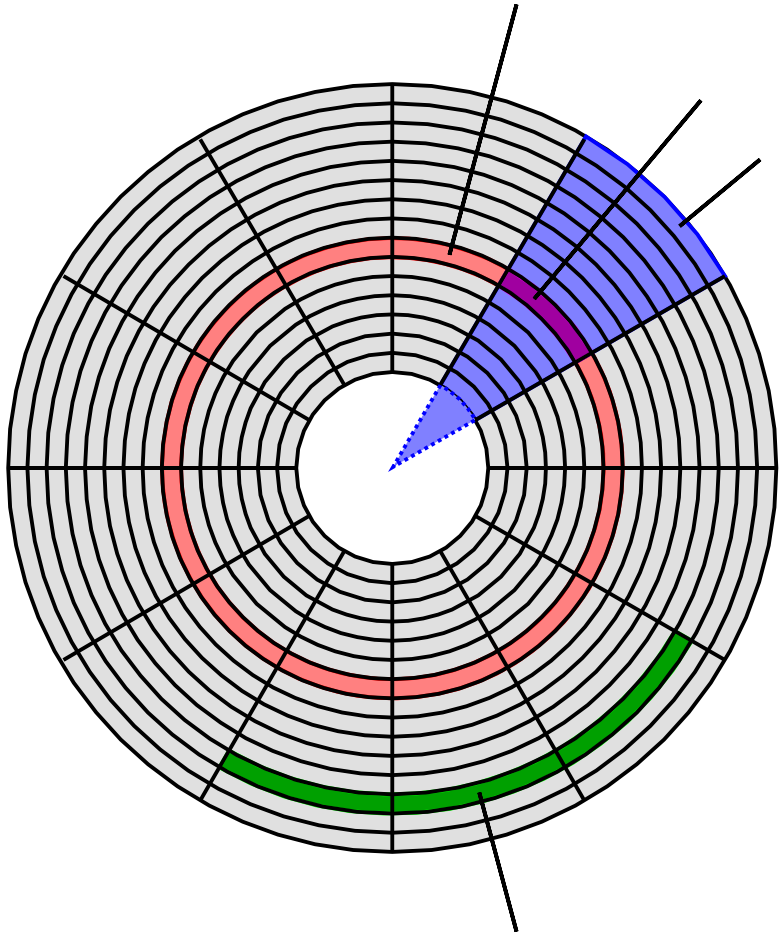


- 每个盘片都有两面
- 因此也会相对应每盘片有 2 个磁头



- 机械硬盘的结构
- A：磁道
- B：扇面
- C：扇区
- D：簇（扇区组）

- 在硬盘上定位某一数据记录位置：C 扇区
- 定位使用了三维定位

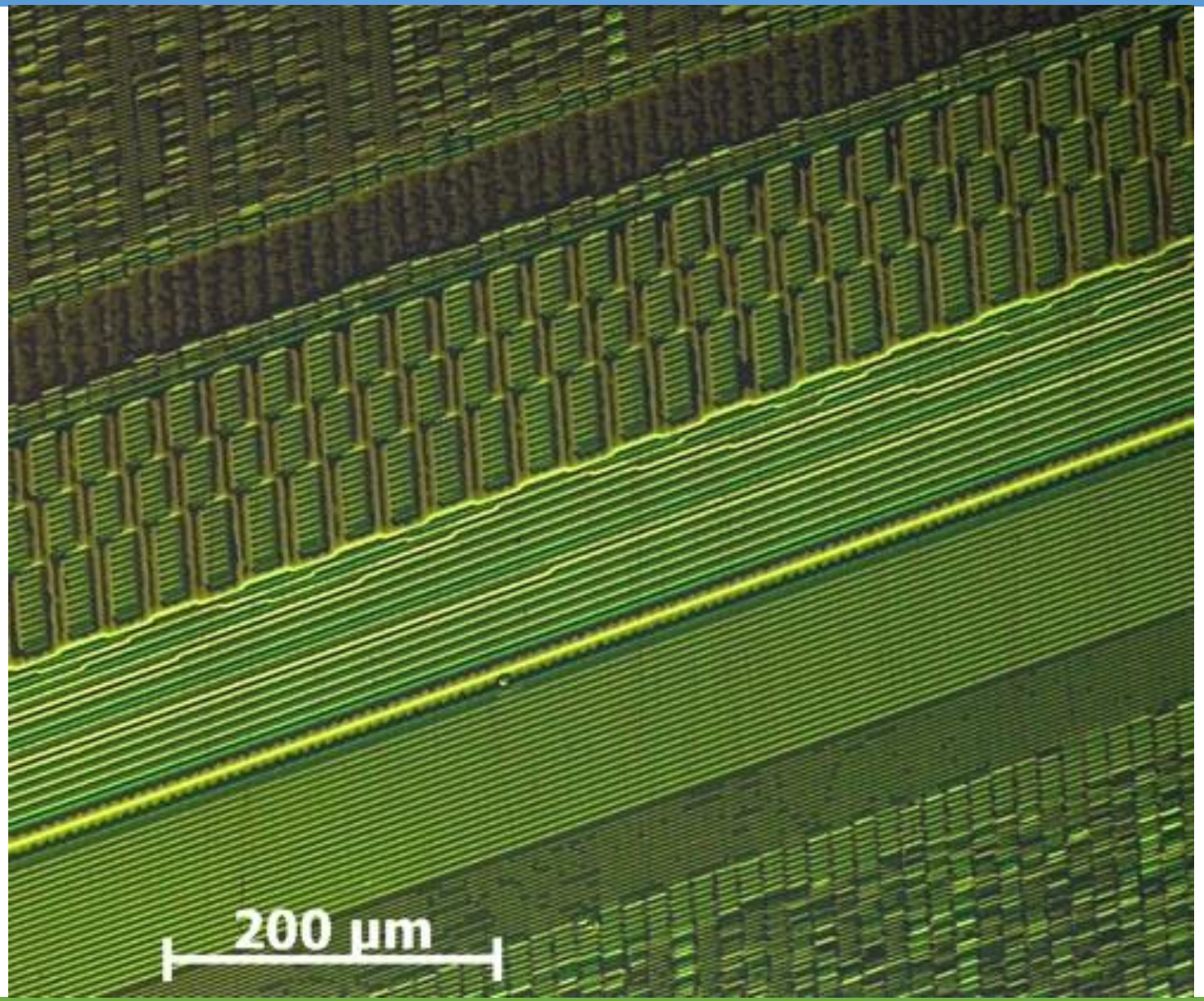


直连存储（DAS）

机械硬盘（HDD）

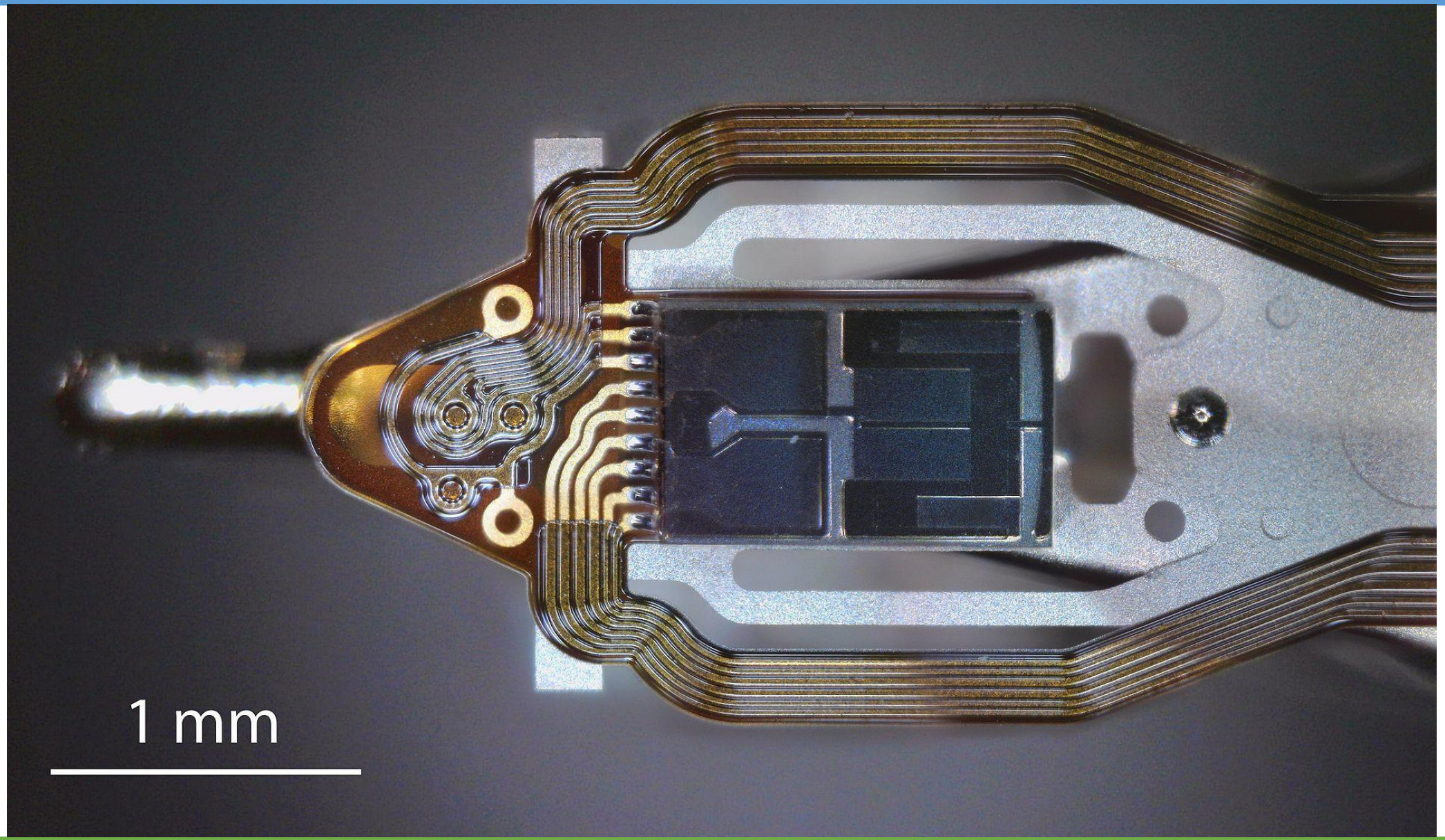
直连存储（DAS）

机械硬盘（HDD）



直连存储（DAS）

机械硬盘（HDD）



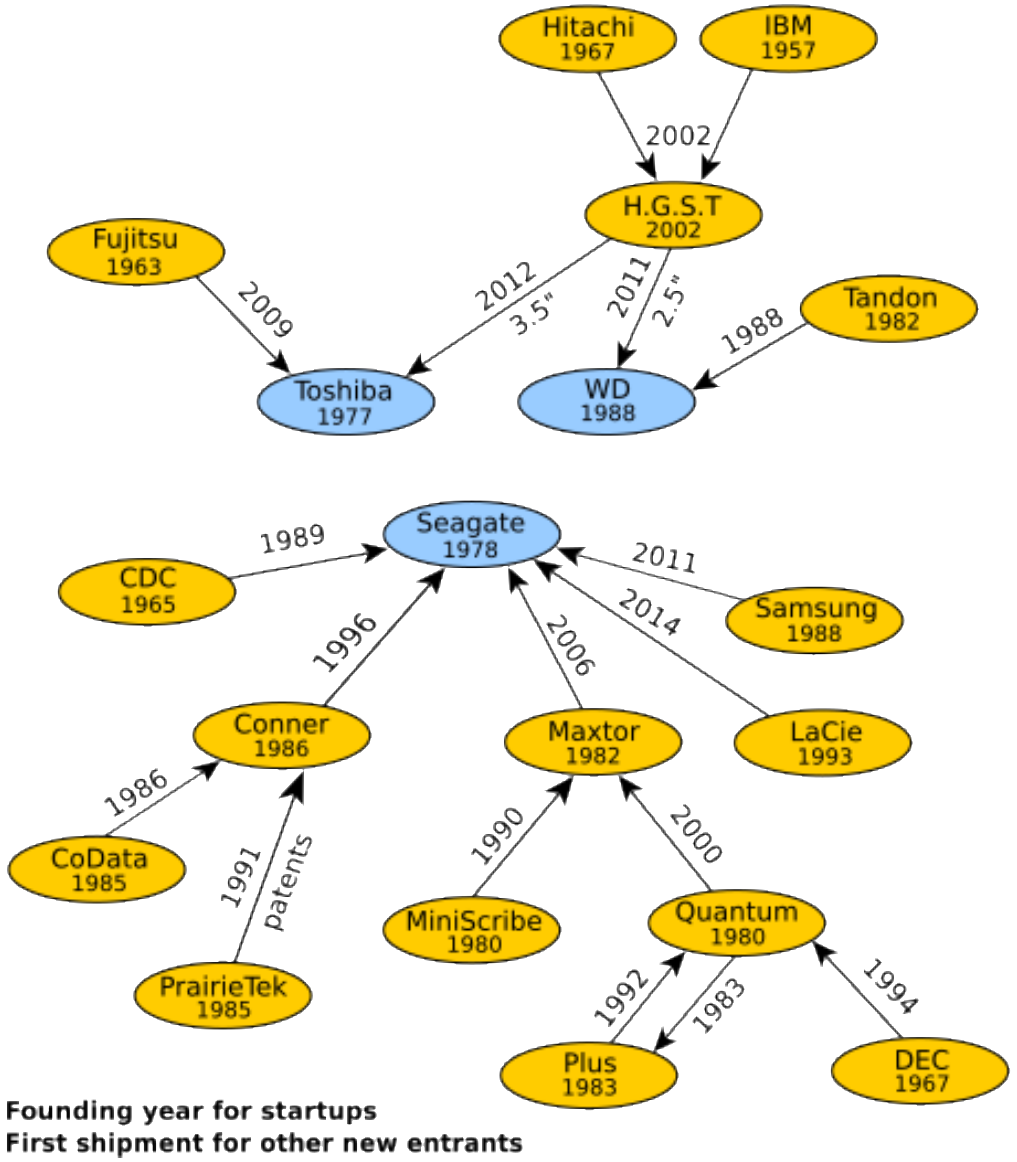
直连存储（DAS）

机械硬盘（HDD）

参数	1957 年	2019 年
最大容量	3.75 MB	16 TB
体积	68 立方呎（1.9 立方米）	2.1 立方英寸（34 立方厘米）
重量	2,000 磅（901 千克）	2.2 盎司（62 克）
访问速率	600 毫秒	2.5 - 10 毫秒
价格	9,200 USD/MB	0.032 USD/GB
数据密度	2,000 Bits/平方呎	1.3 TBits/平方呎
可靠性	2000 小时 MTBF	2,500,000 小时 MTBF

直连存储（DAS）

机械硬盘（HDD）



直连存储（DAS）

机械硬盘（HDD）

转速（RPM）	延迟（ms）
15,000	2
10,000	3
7,200	4.16
5,400	5.55
4,800	6.25

- 机械硬盘的转速曾经是一个重要指标
- 现在因为高速机械硬盘普遍被 SSD 替代，转速已经不是很重要了

京东物流 希捷（SEAGATE）SAS 300GB 15K 3.5企业级硬盘（ST3300657SS）

价格含增值税专票，顺丰包邮！支持货到付款。

京 东 价 ￥1380.00 降价通知　累计评价 200+

优 惠 券 满2000减10　满5000减30　满8000减50　更多>>

促　　销 赠品 ×1 （赠完即止）

满减 满2000元减10元，满5000元减30元，满8000元减50元

希捷(Seagate)1TB 64MB 7200RPM 台式机机械硬盘 SATA接口 希捷酷鱼BarraCuda系列 (ST1000DM010)

希捷SSD首销日，享【限量手办/3A游戏兑换券/40元E卡】；移动硬盘粉丝日五重礼不停立即购买

京 东 价 ￥289.00 降价通知　累计评价 138万+

促　　销 换购 购买1件可优惠换购热销商品 立即换购 >>

增值业务 以旧换新，卖了换钱　礼品包装

配 送 至 广东深圳市南山区 有货　支持 京准达 | 211限时达 | 自提

由 京东 发货，并提供售后服务。23:10前下单，预计明天(07月27日)送达

重　　量 0.435kg

选择颜色 希捷酷鱼-办公家用　希捷酷狼-NAS存储　希捷酷鹰-安防监控　希捷银河-企业存储

选择版本 1T　2T高速　2T　3T　4T　6T　8T　10T　12T　14T

套　　装 优惠套装1

增值保障 全保换2年 ¥19　延长保2年 ¥12　标准版安装 ¥198

京东服务 一年数据救援 ¥48

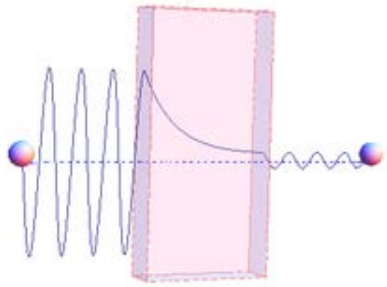
固态硬盘（SSD）

- 固态硬盘（Solid-State Drive，SSD）
- 1978 年由美商存储（STK）研发成功
- 1991 年由 SanDisk 首次投放市场
- 固态主要相对于以机械臂带动磁头转动实现读写操作的磁盘而言
- 固态存储以电位高低或者相位状态的不同记录 0 和 1

直连存储（DAS）

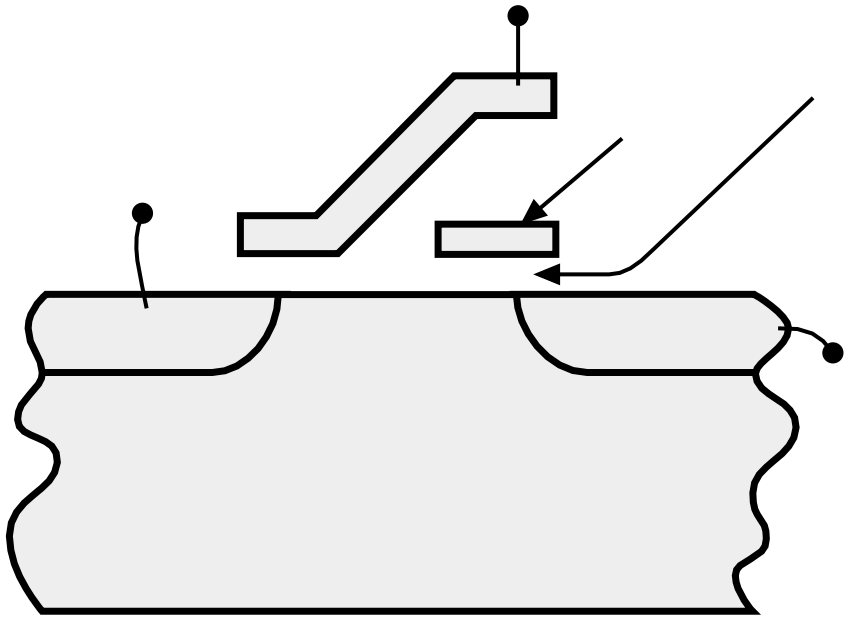
固态硬盘（SSD）

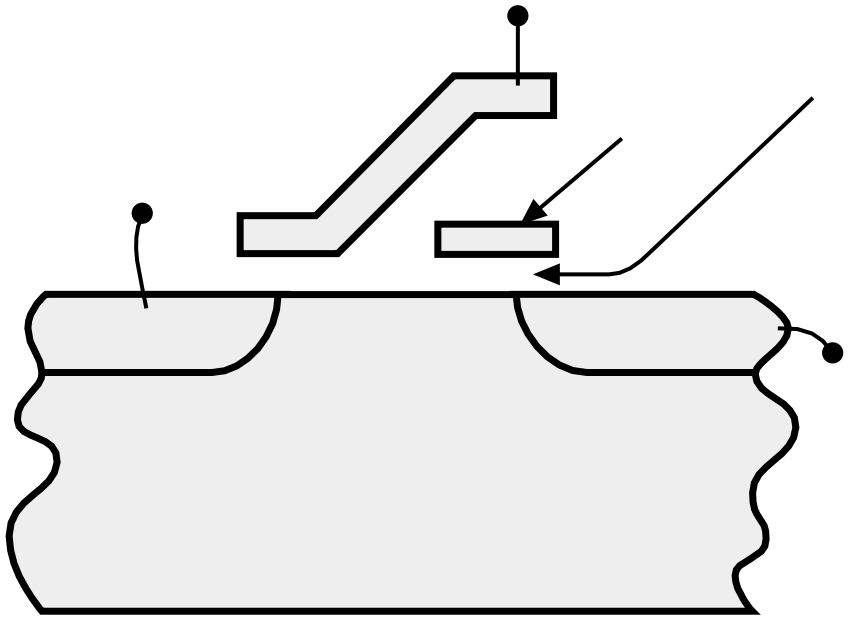
- 固态硬盘的基本原理之一：量子隧穿效应
- 在量子力学里，量子隧穿效应（Quantum Tunneling Effect）指的是，像电子等微观粒子能够穿入或穿越位势垒的量子行为，尽管位势垒的高度大于粒子的总能量
- 在经典力学里，这是不可能发生的，但使用量子力学理论却可以给出合理解释
- 量子隧穿效应是太阳核聚变所倚赖的机制。量子隧穿效应限制了太阳燃烧的速率，是太阳聚变循环的瓶颈，因此维持太阳的长久寿命
- 许多现代器件的运作都倚赖量子隧穿效应，例如，隧道二极管、场致发射、约瑟夫森结、磁隧道结等等。扫描隧道显微镜、原子钟也应用到量子隧穿效应
- 至 2017 年为止，由于对于量子隧穿效应在半导体、超导体等领域的研究或应用，已有 5 位物理学者获得诺贝尔物理学奖



固态硬盘（SSD）

❖ 固态硬盘的基本原理之一：量子隧穿效应



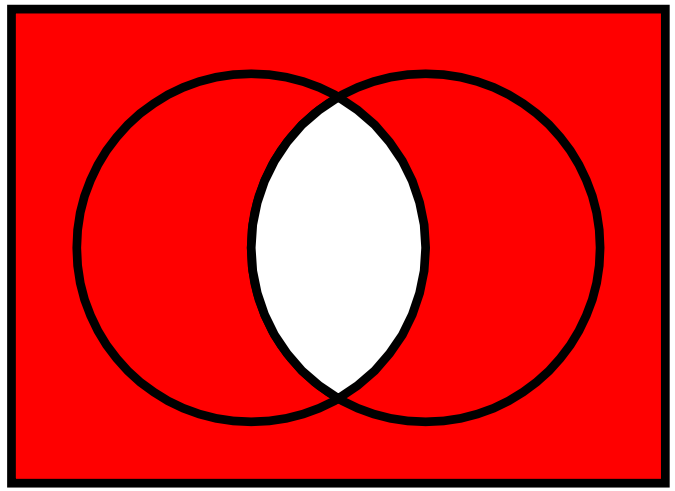


直连存储（DAS）

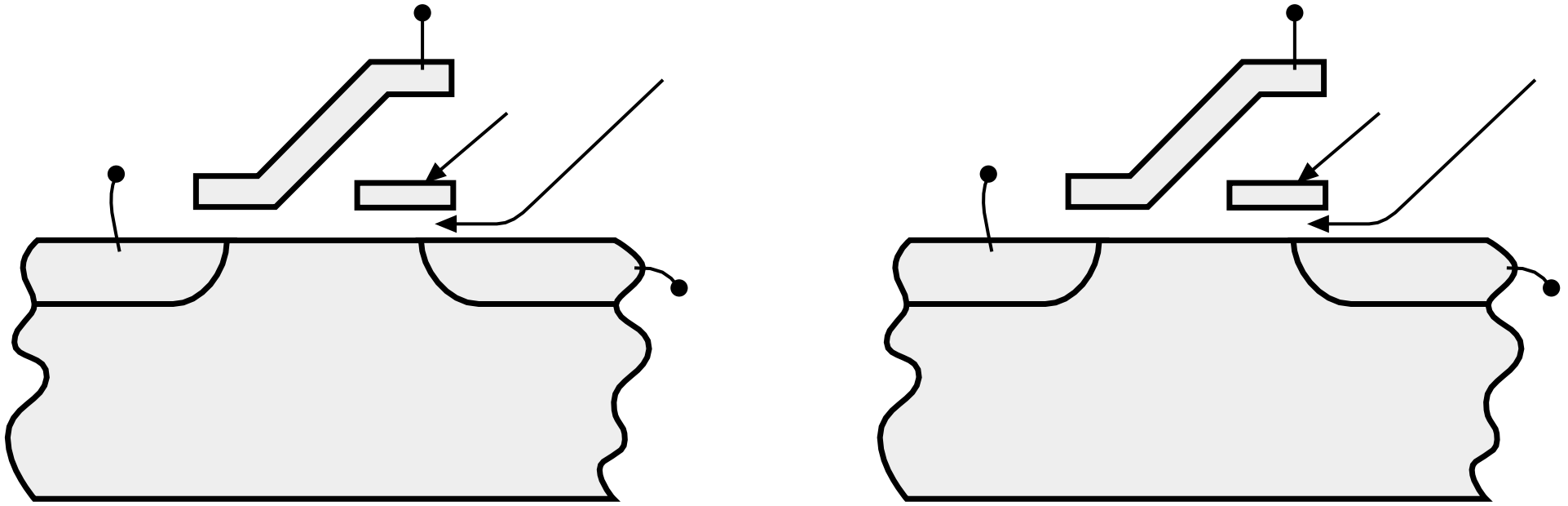
固态硬盘（SSD）

- 固态硬盘的基本原理之二：与非门
- 谢费尔竖线（Sheffer Stroke），得名于 Henry M. Sheffer，写为“ | ”
- 指示等价于合取运算的否定的逻辑运算
- 普通语言表达为“不全是即真”（Not AND，因此也常缩写为 NAND）
- 也就是说，A | B 假，当且仅当 A 与 B 都真时才成立

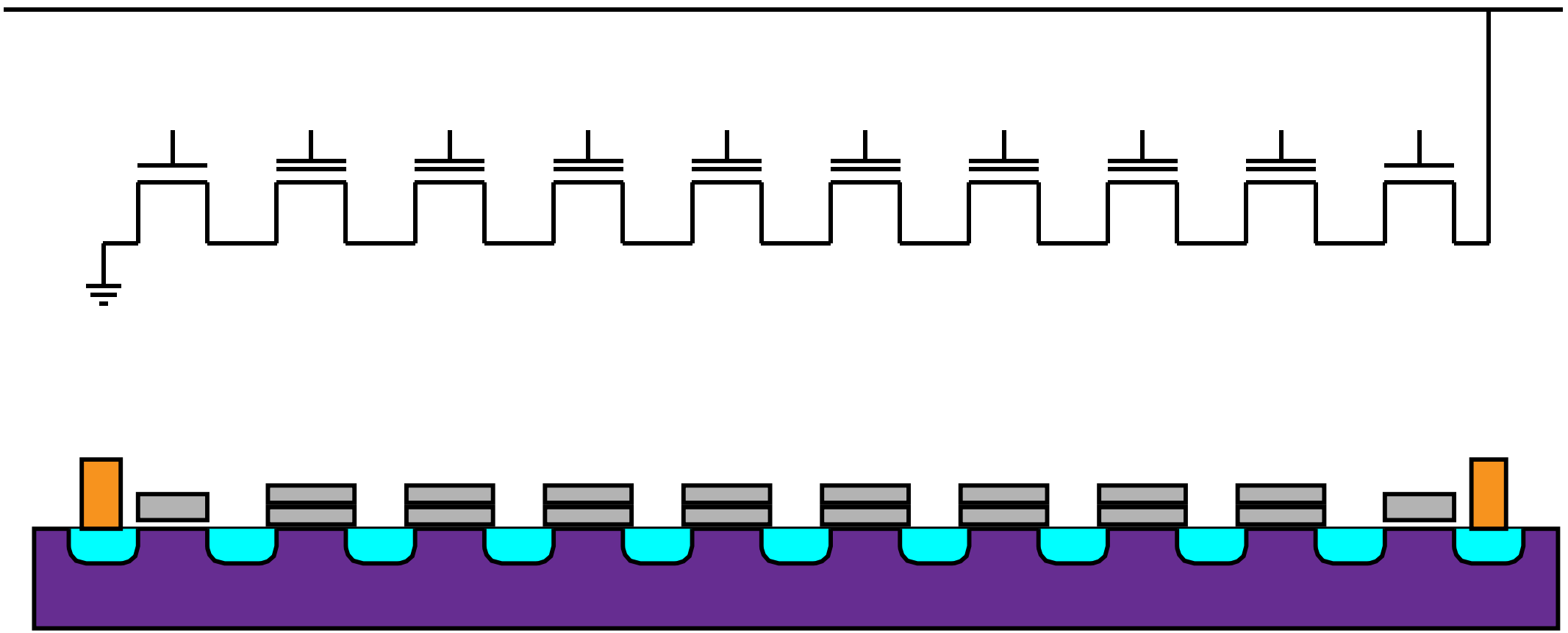
A	B	A | B
T	T	F
T	F	T
F	T	T
F	F	T



❖ 固态硬盘的基本原理之二：与非门



固态硬盘（SSD）



直连存储（DAS）

固态硬盘（SSD）

- 通过对闪存内最小的物理存储单元的电位划分不同的阶数，可以在一个存储单元内存储一至多个二进制位数
- 常见的一至四阶存储单元为 SLC、MLC、TLC 和 QLC
- 企业级 SSD 一般是 SLC 结构
- 家用 SSD 一般是 MLC 结构
- TLC 和 QLC 基本已被淘汰

SLC闪存	
电位情况	二进制值
低 电 位	0
高 电 位	1

MLC闪存	
电位情况	二进制值
最低电位	00
次低电位	01
次高电位	10
最高电位	11

TLC闪存	
电位情况	二进制值
最低电位	000
次低电位	001
第三低电位	010
第四低电位	011
第五低电位	100
第六低电位	101
次高电位	110
最高电位	111

QLC闪存	
电位情况	二进制值
最低电位	0000
次低电位	0001
第三低电位	0010
第四低电位	0011
第五低电位	0100
第六低电位	0101
第七低电位	0110
第八低电位	0111
第九低电位	1000
第十低电位	1001
第十一低电位	1010
第十二低电位	1011
第十三低电位	1100
第十四低电位	1101
次高电位	1110
最高电位	1111

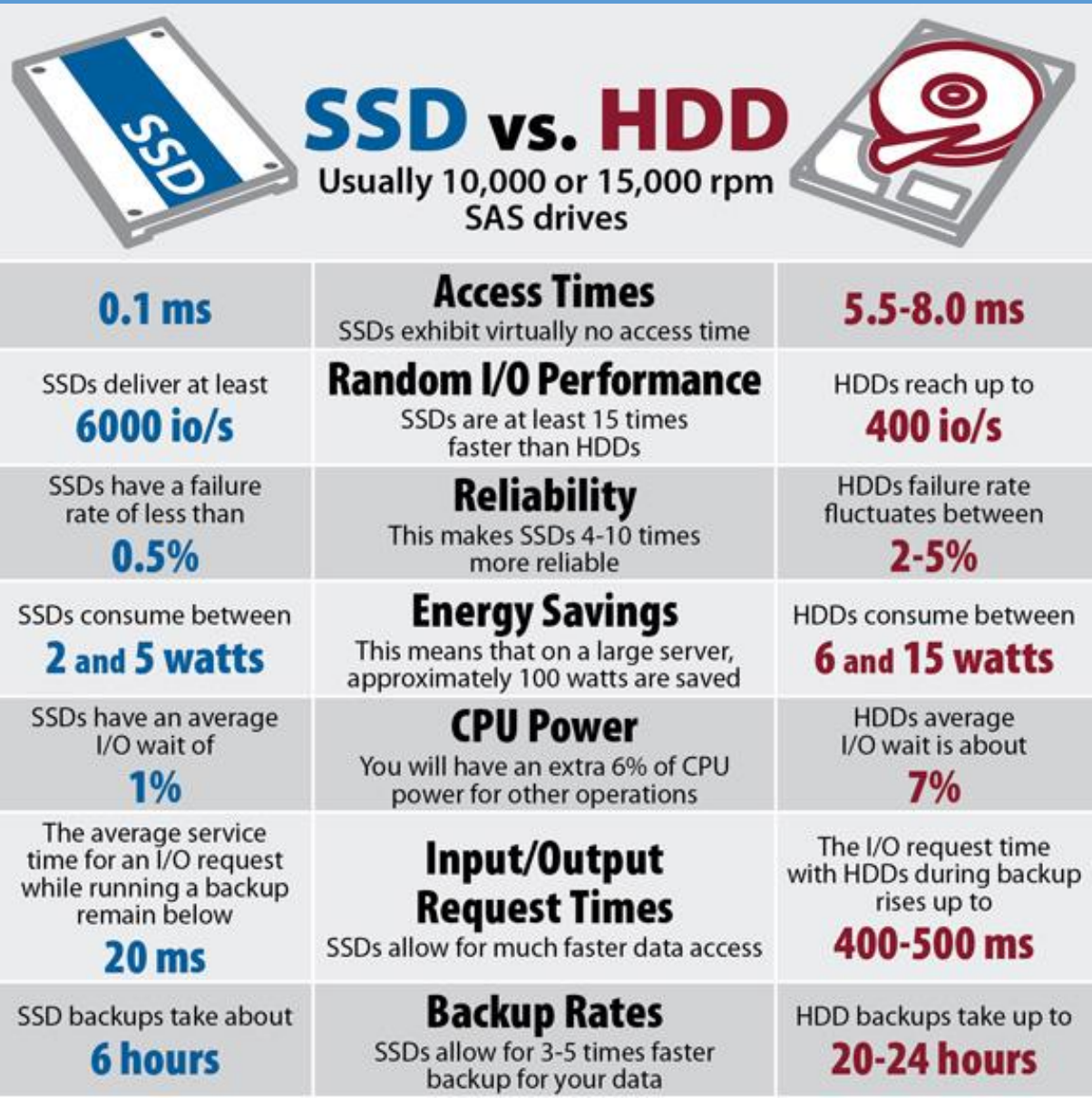
表内数据为假设低电位表示二进制的0，高电位表示二进制的1时的情况。

固态硬盘（SSD）

- 机械硬盘（HDD）vs. 固态硬盘（SSD）

- 机械硬盘主要优点：
- 便宜

- 固态硬盘主要优点：
- 除了贵，其他都是优点



❖ 机械硬盘（HDD）vs. 固态硬盘（SSD）

¥289.00

降价通知　5.6万

自营 希捷(Seagate)1TB 64MB 7200RPM 台式机机械硬盘 SATA接口 希捷酷鱼BarraCuda系列(ST1000DM010)

品牌闪购 ¥729.00 ¥899.00

距闪购结束还剩: 2天 07 : 10 : 42

自营 三星（SAMSUNG）1TB SSD 固态硬盘 SATA3.0接口 860 QVO（MZ-76Q1T0B）

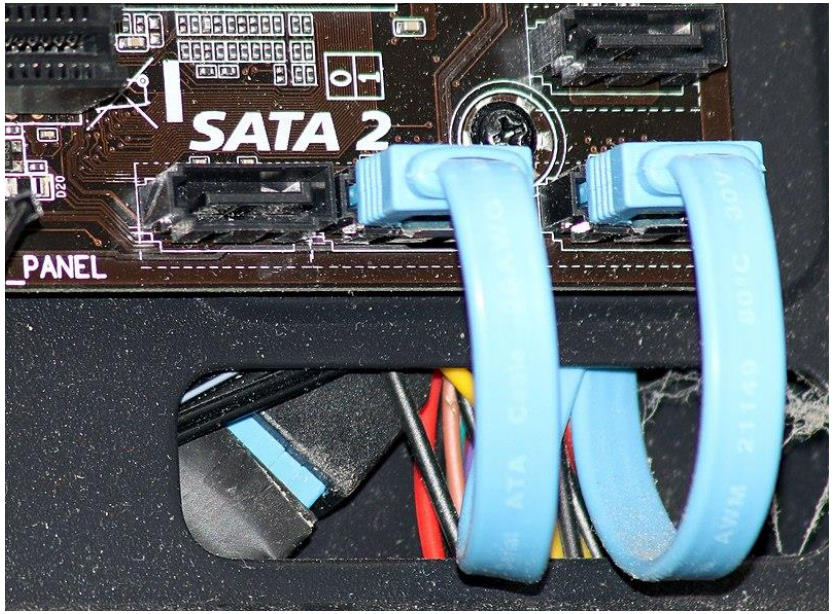
0.7万

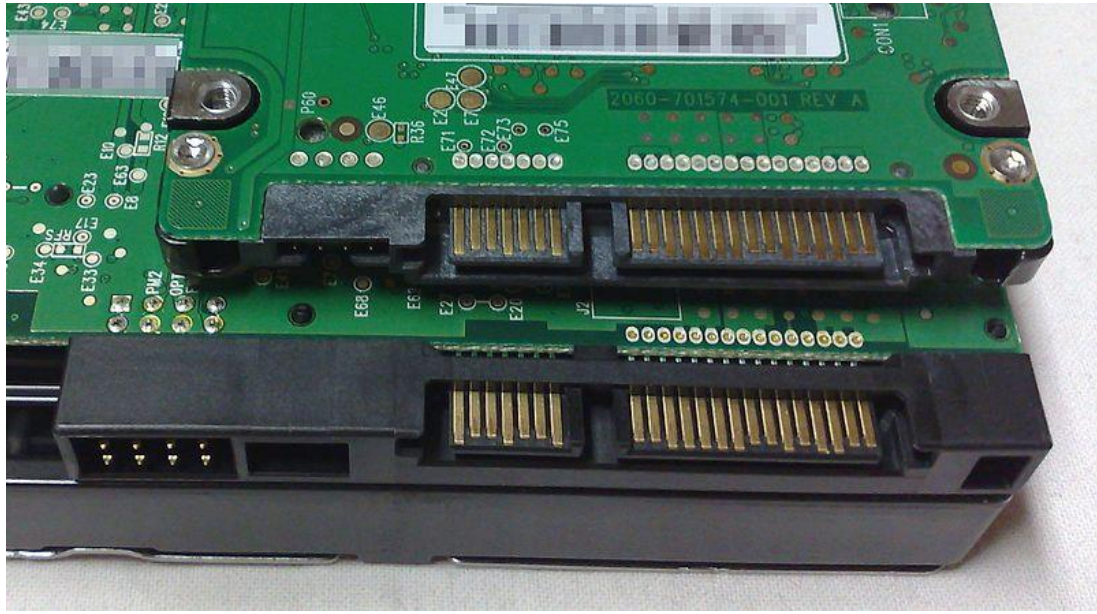
总线标准

- 总线标准
- 总线（Bus）是指计算机组件间规范化的交换数据的方式
- 即以一种通用的方式为各组件提供数据传送和控制逻辑
- 从另一个角度来看，如果说主板（Mother Board）是一座城市，那么总线就像是城市里的公共汽车（Bus），能按照固定行车路线，传输来回不停运作的比特（Bit）
- 这些线路在同一时间内都仅能负责传输一个比特。因此，必须同时采用多条线路才能发送更多数据，而总线可同时传输的数据数就称为宽度，以比特为单位，总线宽度越大，传输性能就越佳
- 总线的带宽（即单位时间内可以传输的总数据数）为：
- 总线带宽 = 频率×宽度（bytes/sec）

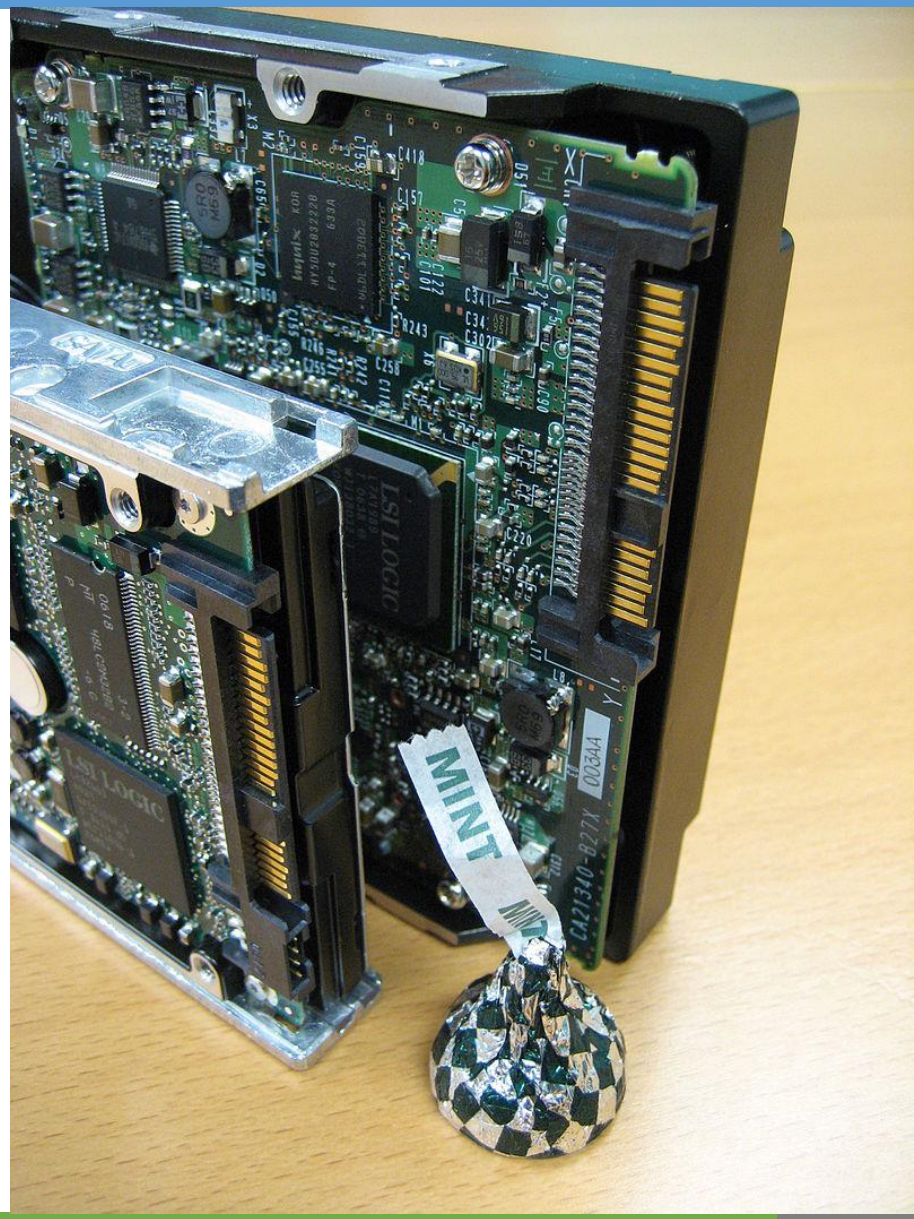
❖ Serial ATA（SATA）

- ❖ 串行 ATA（Serial Advanced Technology Attachment，SATA）是一种计算机总线，负责主板和大容量存储设备（如硬盘及光盘驱动器）之间的数据传输，主要用于个人计算机
- ❖ 串行 ATA 与串行 SCSI（Serial Attached SCSI，SAS）的两者排线兼容



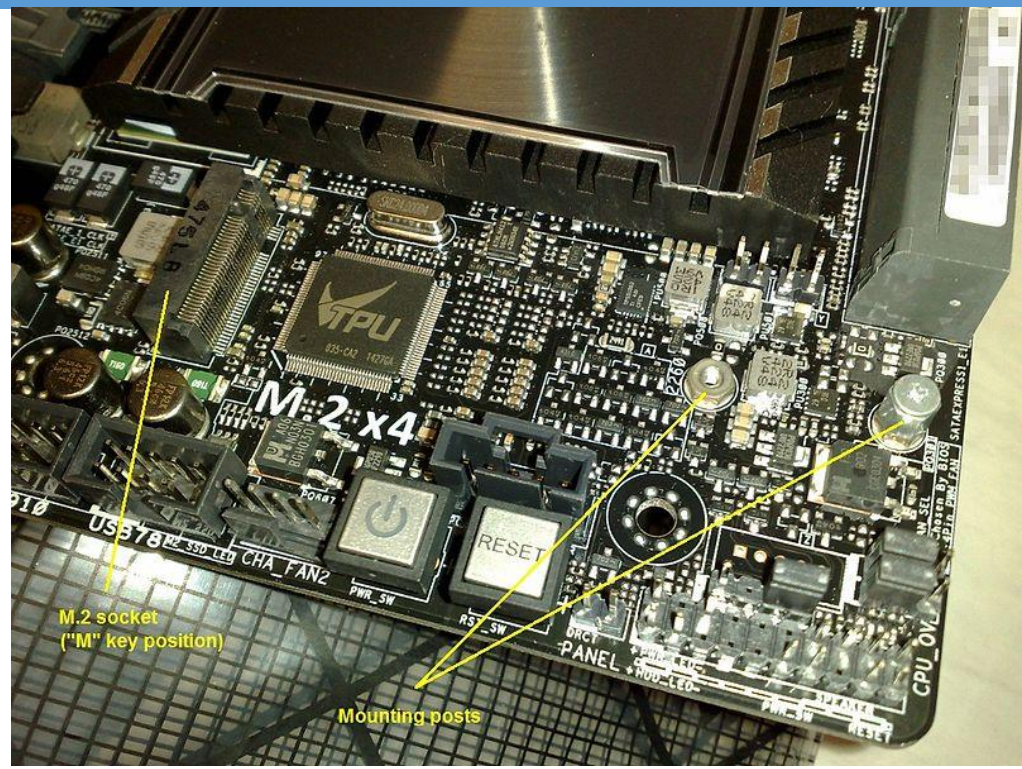


- Serial Attached SCSI（SAS）
- 串行 SCSI（Serial Attached SCSI）由并行 SCSI 物理存储接口演化而来
- 与并行方式相比，序列方式能提供更快速的通信传输速度以及更简易的配置
- SAS 支持与 SATA设备兼容，且两者可以使用相类似的电缆
- 服务器多采用 SAS 接口标准



- mSATA
- SATA 3.1 于 2011 年 7 月发布
- 其中引入 mSATA 新标准
- mSATA 为 SATA 在移动计算设备的固态硬盘接口
- 外型与 Mini PCI Express 相同（但两者并不兼容）

- M.2
- M.2，也称为 Next Generation Form Factor（NGFF），是计算机内部扩展卡及相关连接器规范
- 其采用了全新的物理布局和连接器，将取代 PCI Express Mini 及与 PCI Express Mini 兼容的 mSATA 标准
- M.2 具有灵活的物理规范，允许更多种类的模块宽度与长度，并与更高级的接口相配，使 M.2 比 mSATA 更适合日常应用，尤其是用于超级本或平板电脑等设备的固态硬盘
- 理论上 M.2 接口最多可提供 PCI Express x4 的带宽



直连存储（DAS）

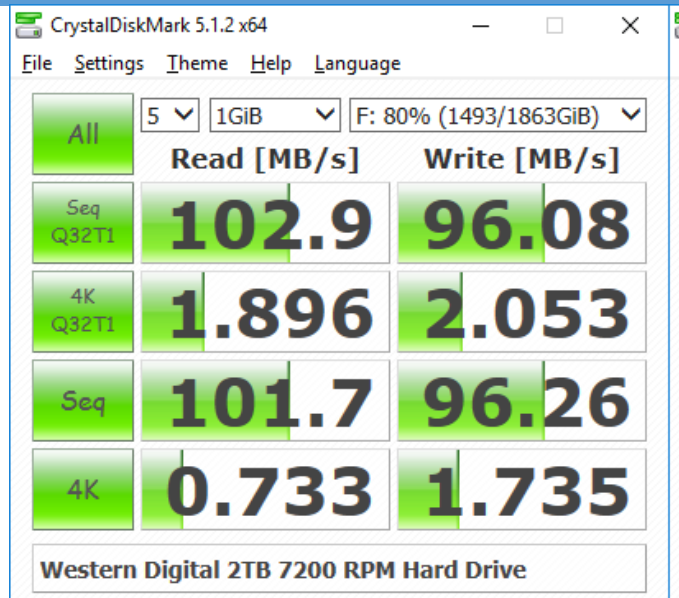
总线标准

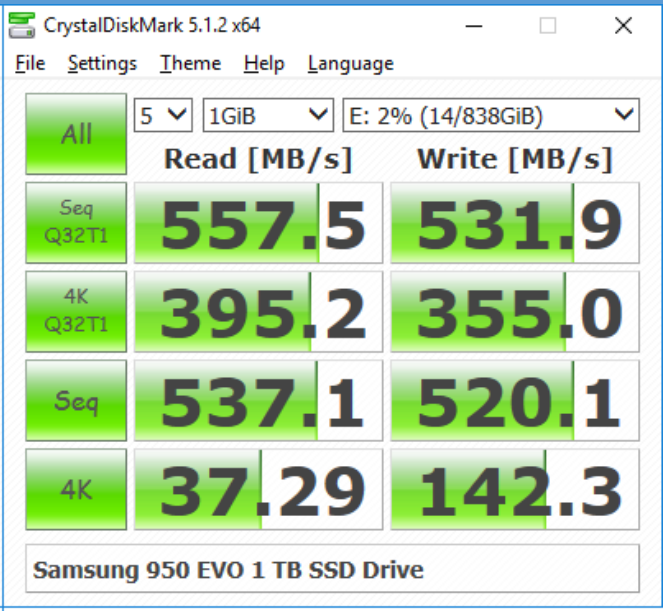
- NVM Express（NVMe）
- 非易失性内存主机控制器接口规范（Non-Volatile Memory Host Controller Interface Specification，NVMHCIS），是一个逻辑设备接口规范
- NVMe 通过 PCI Express（PCIe）总线访问附加的非易失性存储器介质（例如 SSD）

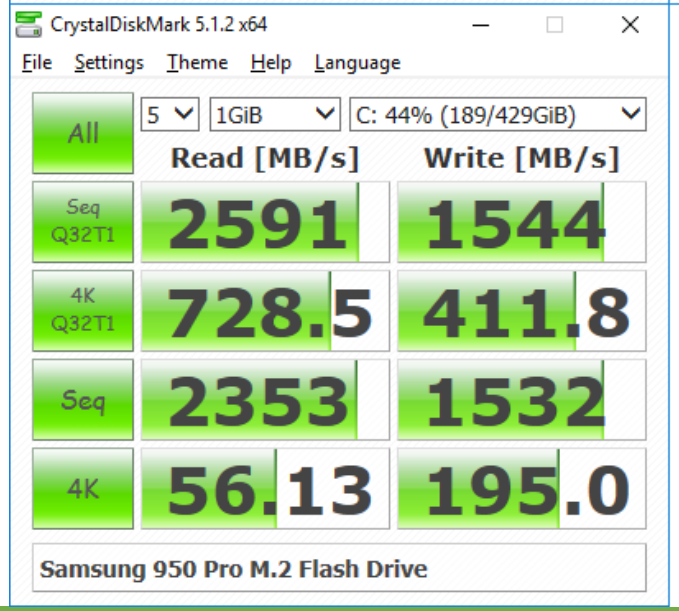
直连存储（DAS）

总线标准

- NVMe（M.2）
- vs.
- SSD（SATA）
- vs.
- HDD（SATA）







Performance Summary:

Hard Drive: 103 MB/sec read, 96 MB/sec write speed
SSD Drive: 558 MB/sec read, 532 MB/sec write speed
M.2 Drive: 2591 MB/sec read, 1544 MB/sec write speed

In Read Performance:
SSD is 5x Faster than HDD
M.2 is 5x Faster than SSD
M.2 is 25x Faster than HDD

直连存储（DAS）

总线标准

- 目前最高速的个人电脑存储标准：
- M.2（笔记本及部分台式机）或 PCI Express（台式机）接口
- NVMe 协议

英睿达（Crucial）1TB SSD固态硬盘 M.2接口(NVMe协议) P1系列-五年质保｜美光出品专注精品

京东价 ¥699.00 降价通知

累计评价 10万+

促销 换购 购买1件可优惠换购热销商品 立即换购 >>

增值业务 以旧换新，卖了换钱

配送至 广东深圳市南山区 有货 支持 自提 99元免基础运费(20kg内)

由 京东 发货，英睿达京东自营旗舰店 提供售后服务. 7天内调货完成，预计08月01日送达

重量 0.05kg

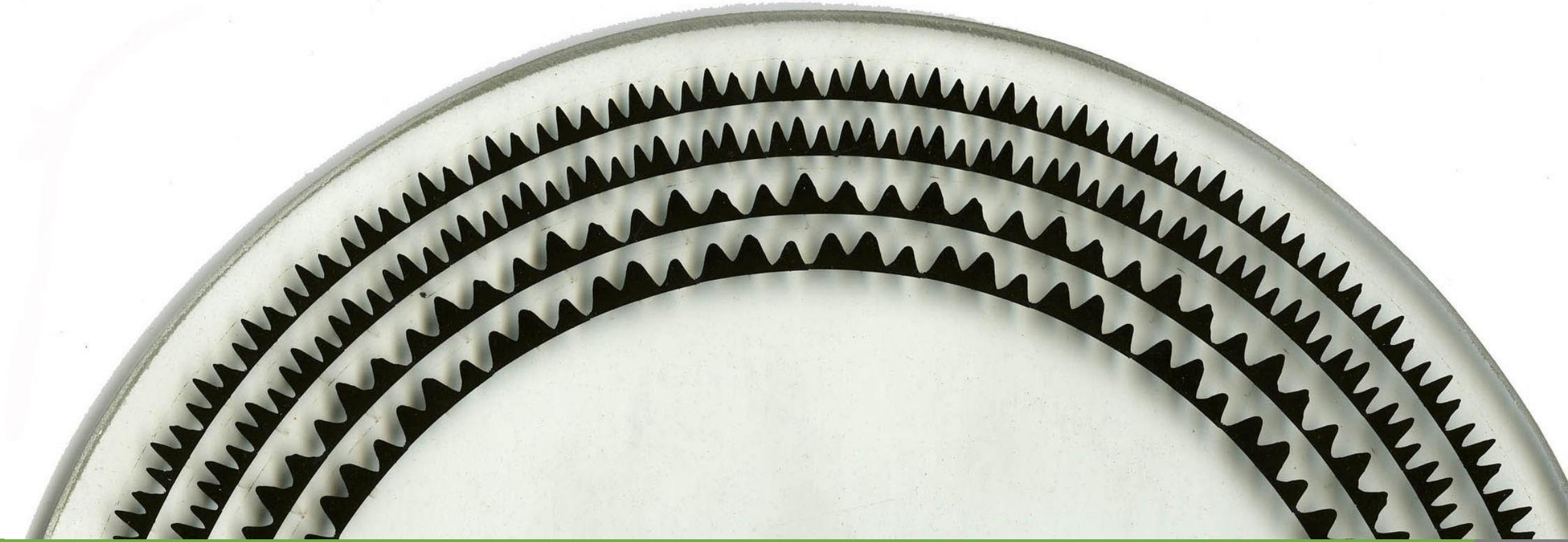
选择颜色 美光原厂精品/性能至上/五年质保 | 游戏性能出众/为美光代言/五年保 | 三年只换不修/为电脑升级助力

选择版本 120GB | 240GB/250GB | 480GB/500GB | 960GB/1TB | 2TB

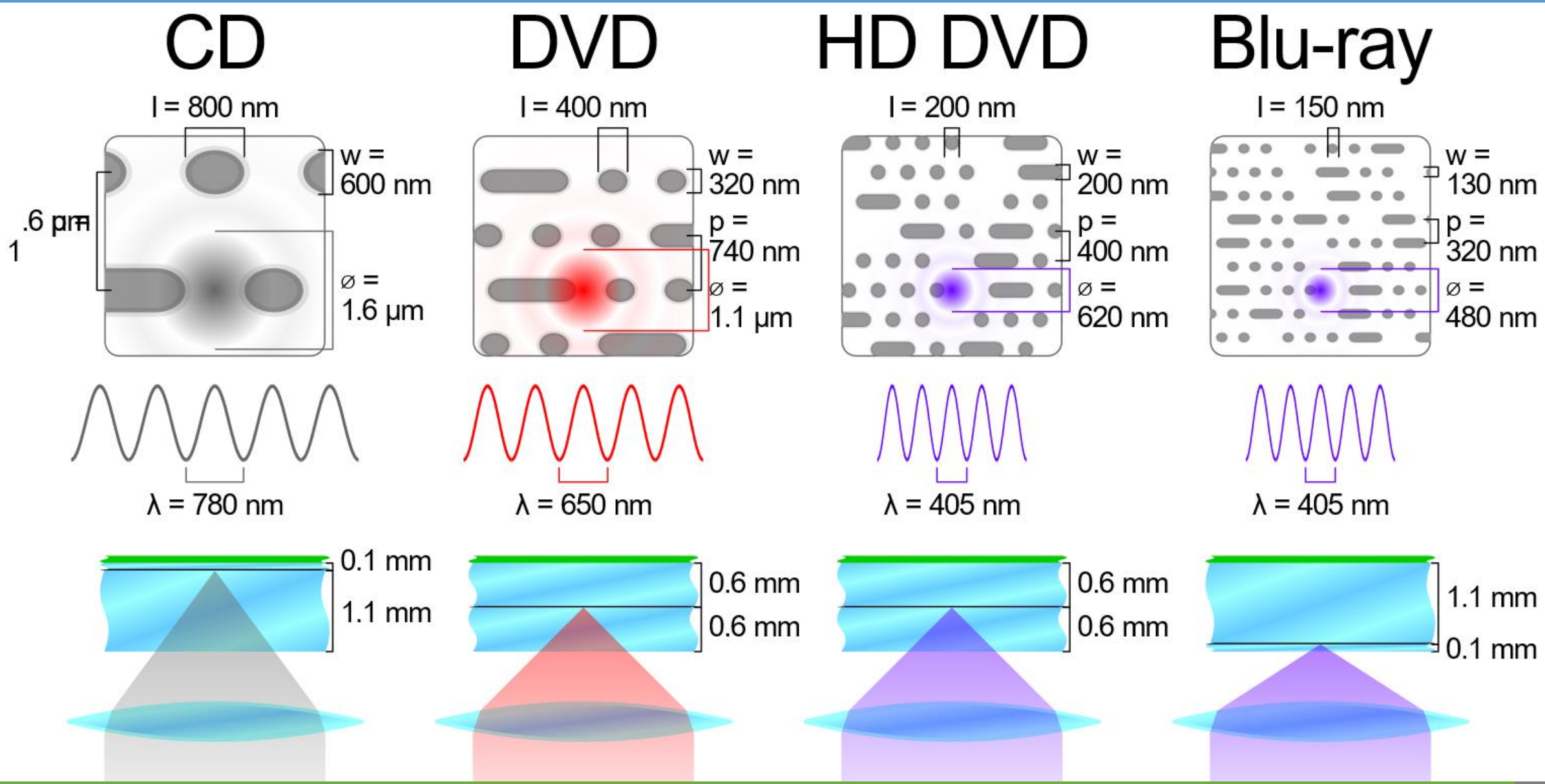
增值保障 全保换2年 ¥49 | 延长保2年 ¥19 | 标准版安装 ¥198

光盘

- 光盘
- 用激光扫描的记录和读出方式保存信息的一种介质
- 大约在 1990 年代中期时开始普及
- 通常为只读介质



光盘


CD
DVD
HD DVD
Blu-ray
l = 800 nm
w = 600 nm
1.6 µm
⌀ = 1.6 µm
l = 400 nm
w = 320 nm
p = 740 nm
⌀ = 1.1 µm
l = 200 nm
w = 200 nm
p = 400 nm
⌀ = 620 nm
l = 150 nm
w = 130 nm
p = 320 nm
⌀ = 480 nm
λ = 780 nm
λ = 650 nm
λ = 405 nm
λ = 405 nm
0.1 mm
1.1 mm
0.6 mm
0.6 mm
0.6 mm
0.6 mm
1.1 mm
0.1 mm

光盘

- 光盘容量
- CD：700 MB
- DVD 单层：4.7 GB
- DVD 双层：8.5 GB
- HD DVD 单层：15 GB
- HD DVD 双层：30 GB
- Blu-ray 单层：25 GB
- Blu-ray 双层：50 GB
- Blu-ray 四层：128 GB

直连存储（DAS）

总结

- 直连式存储（Direct-Attached Storage，DAS）
- 指直接和计算机相连接的数据储存方式
- 固态硬盘、机械硬盘、光盘等与计算机直接相连的设备都属于直连式存储设备
- 通常来说，二级、三级存储都属于 DAS
- 不通过网络传输的离线存储（例如 U 盘）也通常属于 DAS

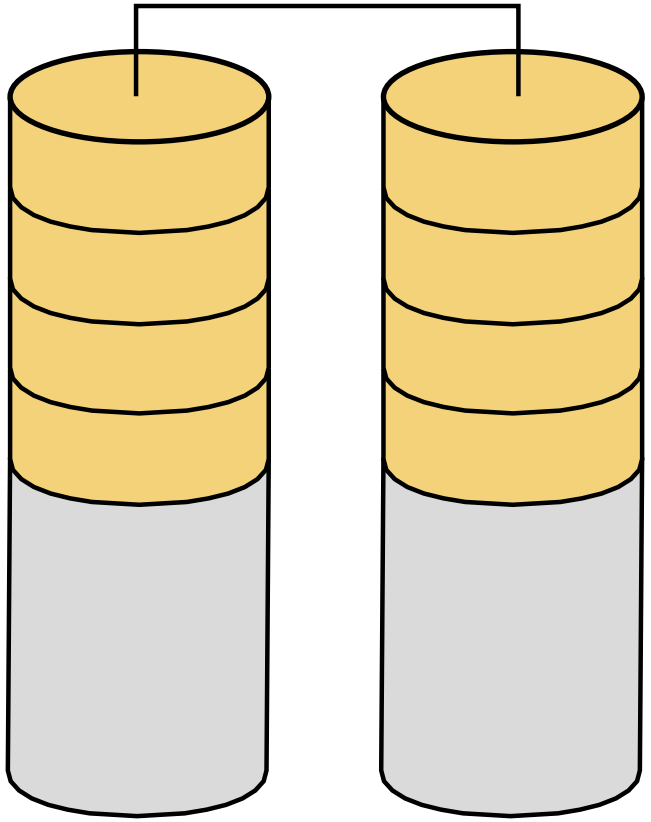
目录

磁盘阵列（RAID）简介

- 磁盘阵列（Redundant Array of Inexpensive Disks，RAID）
- 利用虚拟化存储技术把多个硬盘组合起来，成为一个或多个硬盘阵列组，目的提升性能或数据冗余或是两者同时提升
- 在运作中，取决于 RAID 层级不同，数据会以多种模式分散于各个硬盘，RAID 层级的命名会以 RAID 开头并带数字，例如：RAID 0，RAID 1，RAID 5，RAID 6，RAID 7，RAID 01，RAID 10，RAID 50，RAID 60
- 每种等级都有其理论上的优缺点，不同的等级在两个目标间获取平衡，分别是增加数据可靠性以及增加存储器（群）读写性能
- 简单来说，RAID 把多个硬盘组合成为一个逻辑硬盘，因此，操作系统只会把它当作一个硬盘。RAID 常被用在服务器上，并且常使用完全相同的硬盘作为组合
- 由于硬盘价格的不断下降与 RAID 功能更加有效地与主板集成，它也成为普通用户的一个选择，特别是需要大容量存储空间的工作，例如视频与音频制作等

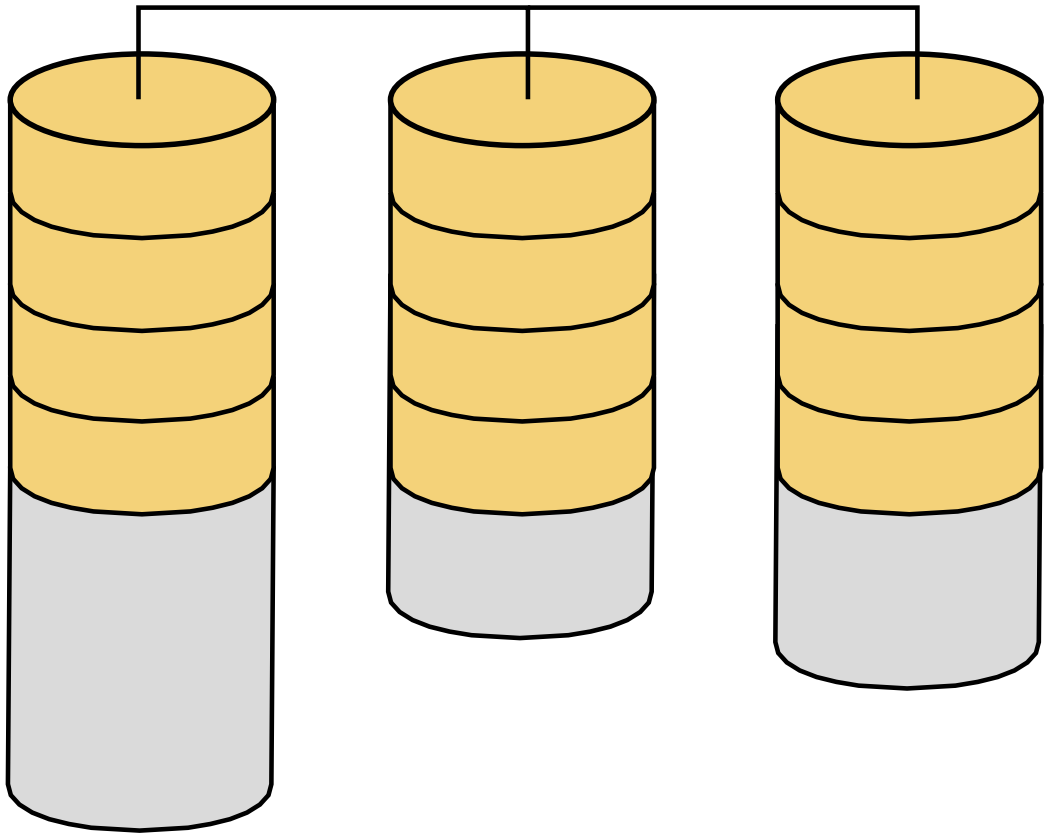
RAID 标准

- RAID 0
- 将两个以上的硬盘合并起来，成为一个大容量的硬盘
- RAID 0 在 Windows 上亦称为带区卷（Striped Volumes）
- 在访问数据时，两块硬盘可以同时读写，所以在所有的 RAID 级别中，RAID 0 的速度是最快的
- 但是 RAID 0 既没有冗余功能，也不具备容错能力，如果一个磁盘（物理）损坏，所有数据都会丢失，危险程度与 JBOD 相当



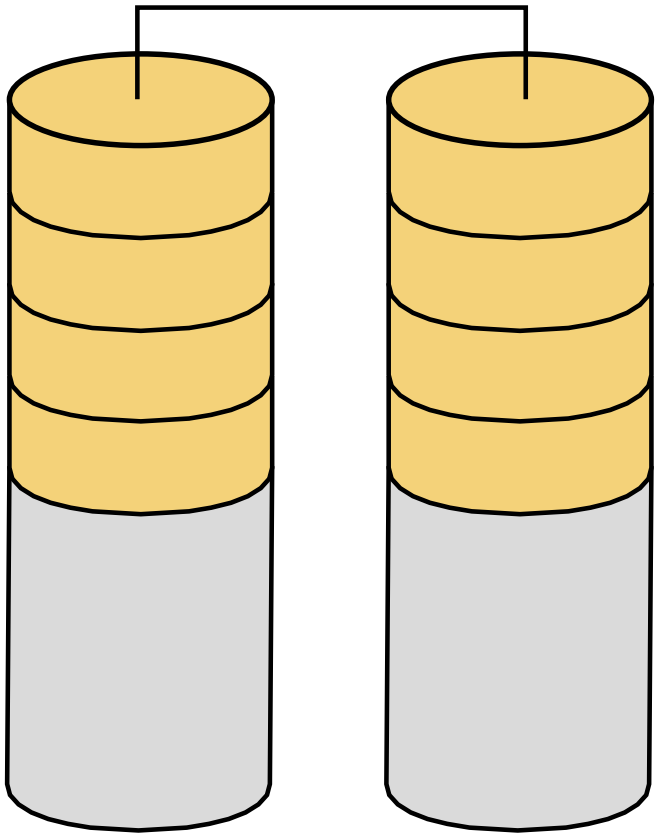
RAID 标准

- JBOD（Just a Bunch of Disks）
- 在分类上，JBOD 并不是 RAID 的等级
- 数据的存放机制是由第一块硬盘开始依序往后存放，即操作系统看到的是一个大硬盘（由许多小硬盘组成的）
- 如果一块硬盘损毁，则该硬盘上的所有数据都将无法找回
- 若第一块硬盘损坏，通常所有数据都会丢失，因为大部分文件系统将分区表（Partition Table）存在硬盘前端（即第一块硬盘），失去分区表即失去一切数据
- 它的特点是不会像 RAID 0，每次访问都要读写全部硬盘



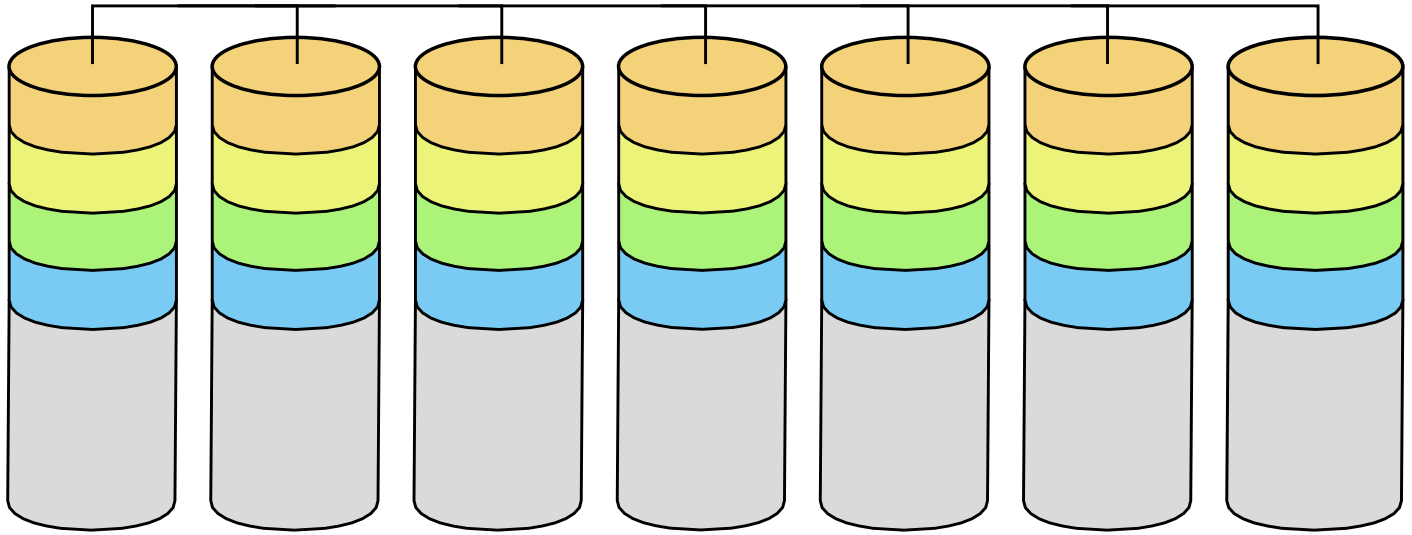
RAID 标准

- RAID 1
- 两块以上的硬盘相互作镜像
- 两块硬盘可以同时访问，理论上读写速度与 RAID 0 相同
- 当主硬盘（物理）损坏时，镜像硬盘则代替主硬盘工作
- 因为有镜像硬盘做数据备份，所以 RAID 1 的数据安全性在所有的 RAID 级别上来说是最好的
- 但无论用多少硬盘做 RAID 1，都仅算一个硬盘的容量，所以 RAID 1 是所有 RAID 级别中硬盘利用率最低的



RAID 标准

- RAID 2
- RAID 2 是 RAID 1 的改良版，以汉明码（Hamming Code）的方式将数据进行编码后，将校验信息写入另一块硬盘中
- 因为使用了错误修正码（ECC，Error Correction Code），而不是将数据完全镜像，因此硬盘利用率比 RAID 1 要高
- RAID 2 至少要三块硬盘方能运作



RAID 标准

- 奇偶校验位（Parity Bit）
- 一个表示给定位数的二进制数中 1 的个数是奇数还是偶数的二进制数
- 奇偶校验位是最简单的错误检测码

- 以偶校验位来说，如果一组给定数据位中 1 的个数是奇数，补一个 bit 为 1，使得总的 1 的个数是偶数
- 例：0000 001，补一个 bit 为 1，得到：0000 0011

- 以奇校验位来说，如果给定一组数据位中 1 的个数是奇数，补一个 bit 为 0，使得总的 1 的个数是奇数
- 例：0000 001，补一个 bit 为 0，得到：0000 0010

RAID 标准

- 循环冗余校验（Cyclic Redundancy Check，CRC）
- 由 W. Wesley Peterson 于 1961 年发表
- 是一种根据网络数据包或计算机文件等数据产生简短固定位数校验码的一种散列函数
- 主要用来检测或校验数据传输或者保存后可能出现的错误
- 生成的数字在传输或者存储之前计算出来并且附加到数据后面，然后接收方进行检验确定数据是否发生变化
- 一般来说，循环冗余校验的值都是 32 位的整数
- 由于本函数尤其适用于检测传输通道干扰引起的错误，因此获得广泛应用

Thank you for downloading Ubuntu Server

Your download should start automatically. If it doesn't, download now.

▼ Verify your download

Run this command in your terminal in the directory the iso was downloaded to verify the SHA256 checksum:

```
$ echo "ea6ccb5b57813908c006f42f7ac8eaa4fc603883a2d07876cf9ed74610ba2f53 *ubuntu
  18.04.2-live-server-amd64.iso" | shasum -a 256 --check
```

You should get the following output:

```
ubuntu-18.04.2-live-server-amd64.iso: OK
```

Or follow this tutorial to learn how to verify downloads

RAID 标准

- 汉明码（Hamming Code）
- 汉明码是一种错误修正码（ECC，Error Correction Code）
- 由理查德·卫斯里·汉明（Richard Wesley Hamming）于 1950 年发明
- 相比而言，简单的奇偶检验码除了不能纠正错误之外，也只能侦测出奇数个的错误
- 汉明码有很多实现方案

RAID 标准

❖ 可以纠错一位（Single Error Correcting）的汉明码

1. 从 1 开始给数字的数据位（从左向右）标上序号：1、2、3、4、5…
2. 位置序号中所有为二的幂的位（序号 1、2、4、8 等，即二进制表示中只有一个 1 的序号）是校验位
3. 所有其它位置是数据位
4. 校验位通过二进制异或来得到校验信息

RAID 标准

- 可以纠错一位（Single Error Correcting）的汉明码
- 对 1100 0010 进行汉明编码
- 列出表格，从左往右填入数字，但 2 的幂的位置不填

位	01	02	03	04	05	06	07	08	09	10	11	12
值			1		1	0	0		0	0	1	0

RAID 标准

- 可以纠错一位（Single Error Correcting）的汉明码
- 对 1100 0010 进行汉明编码
- 把数据位有 1 的位的序号写为二进制

位	01	02	03	04	05	06	07	08	09	10	11	12
值			1		1	0	0		0	0	1	0
位			0011		0101						1011	

RAID 标准

- 可以纠错一位（Single Error Correcting）的汉明码
- 对 1100 0010 进行汉明编码
- 收集所有二进制数字，求异或

位	01	02	03	04	05	06	07	08	09	10	11	12
值			1		1	0	0		0	0	1	0
位			0011		0101						1011	

$$0011 \oplus 0101 \oplus 1011 = 1101$$

RAID 标准

- 可以纠错一位（Single Error Correcting）的汉明码
- 对 1100 0010 进行汉明编码
- 把 1101 倒序填入表格中的校验位

位	01	02	03	04	05	06	07	08	09	10	11	12
值	1	0	1	1	1	0	0	1	0	0	1	0
位			0011		0101						1011	

$$0011 \oplus 0101 \oplus 1011 = 1101$$

RAID 标准

- 可以纠错一位（Single Error Correcting）的汉明码
- 对 1100 0010 进行汉明编码：1011 1001 0010
- 数据错误：1011 1001 0110

位	01	02	03	04	05	06	07	08	09	10	11	12
值	1	0	1	1	1	0	0	1	0	1	1	0
位			0011		0101					1010	1011	

$$0011 \oplus 0101 \oplus 1010 \oplus 1011 = 0111$$

RAID 标准

- 可以纠错一位（Single Error Correcting）的汉明码
- 对 1100 0010 进行汉明编码：1011 1001 0010
- 数据错误：1011 1001 0110

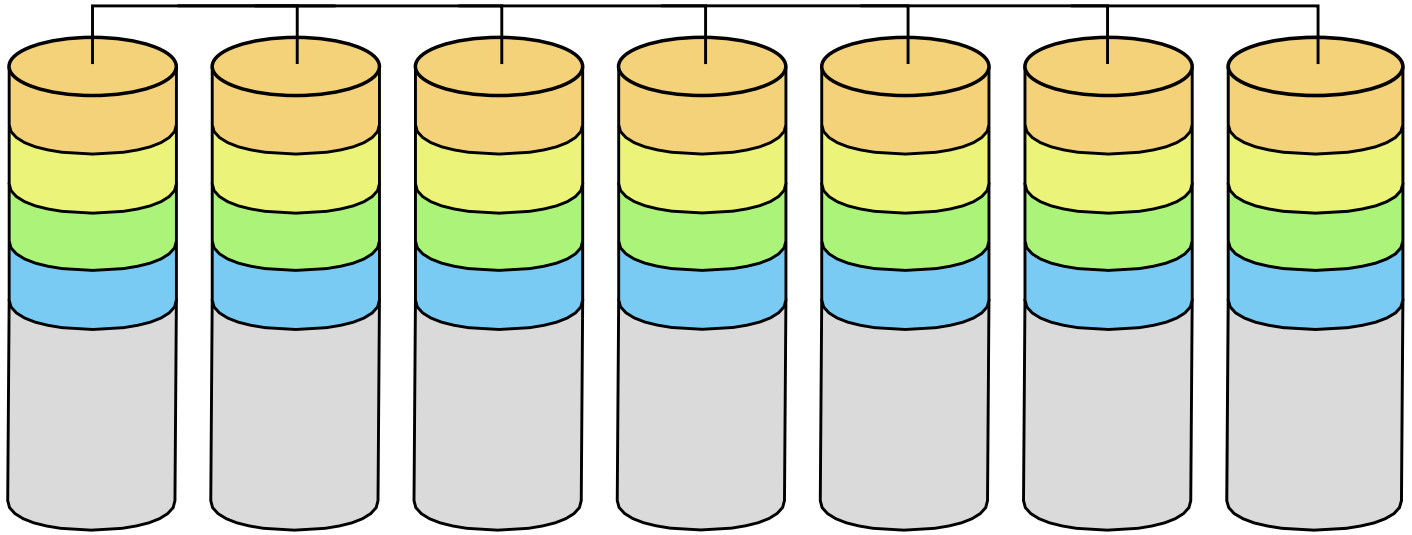
位	01	02	03	04	05	06	07	08	09	10	11	12
值	1	0	1	1	1	0	0	1	0	1	1	0
位			0011		0101					1010	1011	

$$0011 \oplus 0101 \oplus 1010 \oplus 1011 = 0111$$

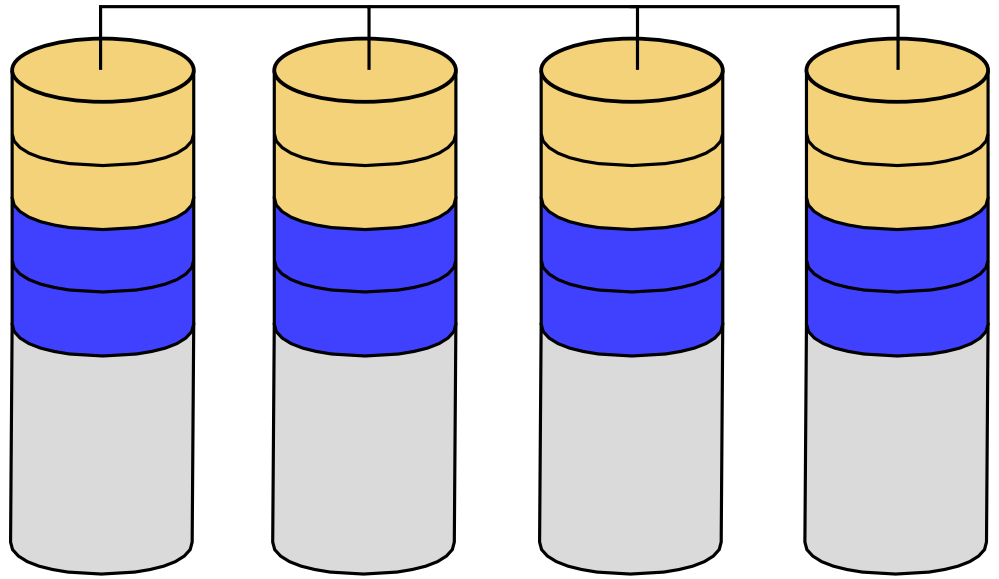
$$2 + 8 \rightarrow 10$$

RAID 标准

- RAID 2
- RAID 2 是 RAID 1 的改良版，以汉明码（Hamming Code）的方式将数据进行编码后，将校验信息写入另一块硬盘中
- 因为使用了错误修正码（ECC，Error Correction Code），而不是将数据完全镜像，因此硬盘利用率比 RAID 1 要高
- RAID 2 至少要三块硬盘方能运作

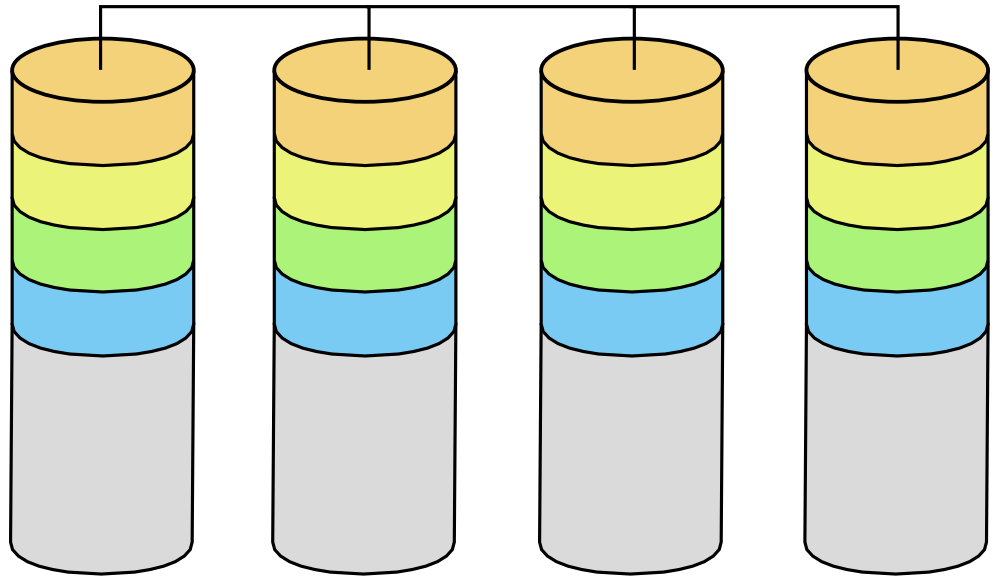


- RAID 3
- RAID 3 是 RAID 2 的改良版，采用 Bit-interleaving（比特交错存储）技术，通过编码将数据分割后分别存在不同硬盘中
- 相对 RAID 2 的汉明码，RAID 3 的硬盘利用率进一步提高
- 安全性相对 RAID 2 有所下降



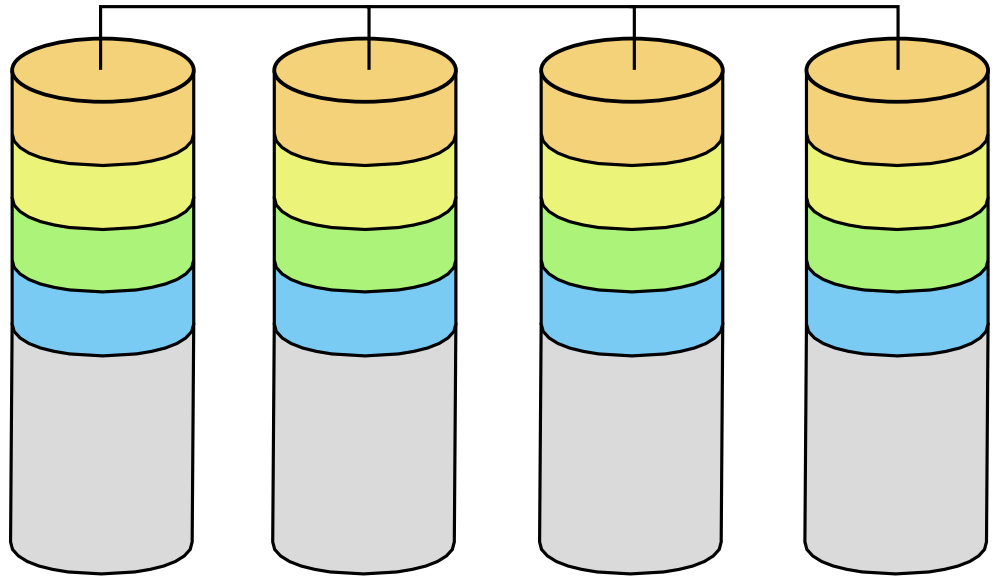
RAID 标准

- RAID 4
- RAID 4 是 RAID 3 的改良版，采用 Block-interleaving（块交错存储）技术，通过编码将数据分割后分别存在不同硬盘中
- 相对 RAID 3，RAID 4 对硬盘的访问速度更快
- 硬盘利用率并没有显著提高
- 安全性相对 RAID 3 有所下降



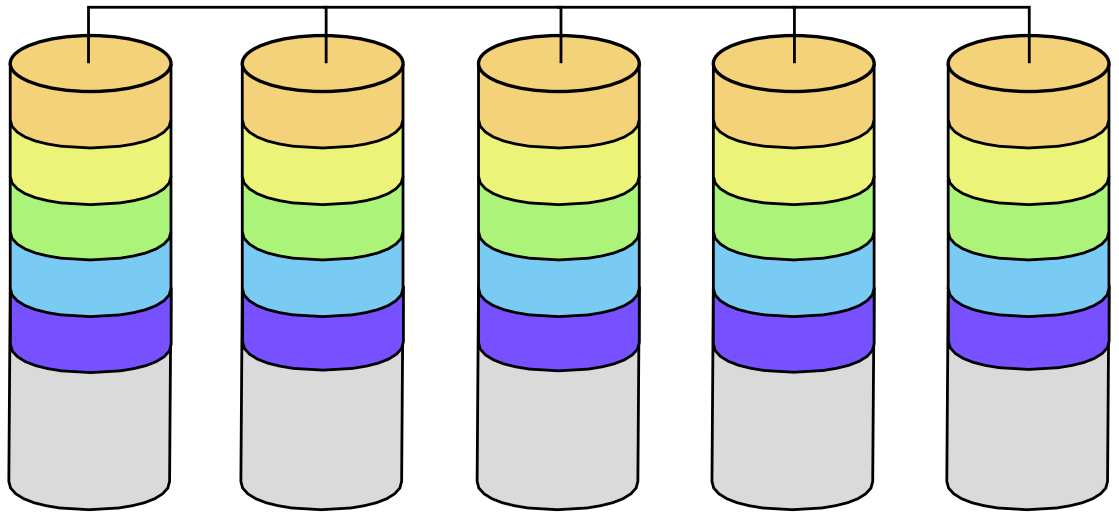
RAID 标准

- RAID 5
- RAID 5 是 RAID 3/4 的改良版，校验信息不再存放于一块独立的硬盘中，而是交错存放于每一块硬盘中
- 相对 RAID 3/4，RAID 5 的安全性有所提高
- 硬盘利用率并没有显著提高



RAID 标准

- RAID 6
- RAID 6 是 RAID 5 的改良版，提供了额外一份校验信息
- 相对 RAID 5，RAID 6 的安全性进一步提高
- RAID 6 的安全性相对 RAID 2 来说没有明显下降
- 硬盘利用率显著高于 RAID 2
- RAID 0、1、6 是目前最常用的 RAID 标准



RAID 实现方案

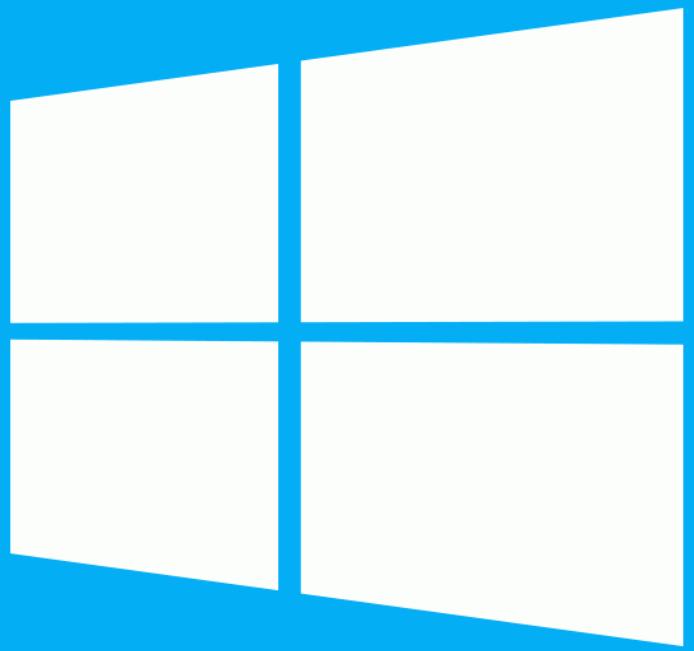
- 硬件磁盘阵列（Hardware RAID）
- RAID 卡上内置处理器，不需要服务器的 CPU 运算
- 优点是读写性能最快，不占用服务器资源，可用于任何操作系统，也能在系统断电后，透过备份电池模块（BBU，Backup Battery Unit）以及非易失性存储器（NVRAM）将硬盘读写日志档（Journal）包含的剩余读写作业先纪录在存储器中，等待电力供应撤销后，再由NVRAM取回日志档数据，接着再完成读写作业，将剩余读写作业安全完成以确保读写完整性
- 缺点是其售价很高，通常只用于 RAID 5 和 RAID 6

磁盘阵列（RAID）

RAID 实现方案

- 软件磁盘阵列（Software RAID）
- 主要由 CPU 处理数组存储作业，缺点为耗损较多 CPU 资源运算 RAID，优点则是价格偏低。分类有 3 种：
- 基于主板的磁盘阵列：只需要主板支持即可（通常是芯片组内置的 RAID 功能，如 Intel Matrix RAID，Intel Rapid Storage Technology），不需要任何磁盘阵列卡。若主板损坏，可能难以购买同款主板重建 RAID
- 硬件辅助磁盘阵列（Hardware-assisted RAID）：需要一张基于 Fake RAID 的 RAID 卡，以及厂商所提供的驱动程序，但此类 RAID 卡仍然通过 CPU 进行运算。这款 RAID 较易迁移到其他计算机。RAID 功能靠运行于操作系统的厂商驱动程序和 CPU 运算提供
- 操作系统的 RAID 功能：如 Linux、FreeBSD、Windows Server 等操作系统内置 RAID 功能

在 Windows 上使用 RAID



Windows 10

磁盘阵列（RAID）

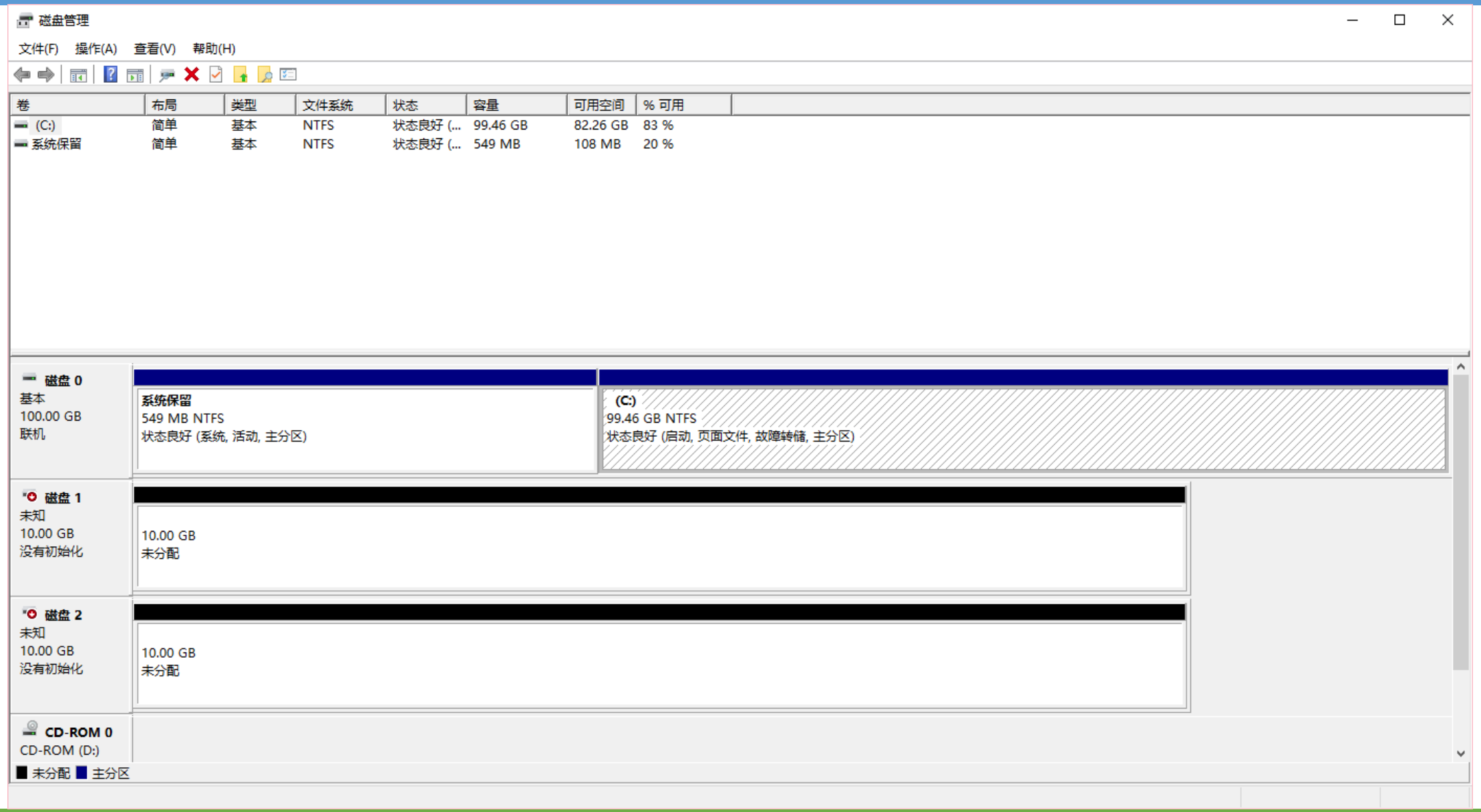
在 Windows 上使用 RAID

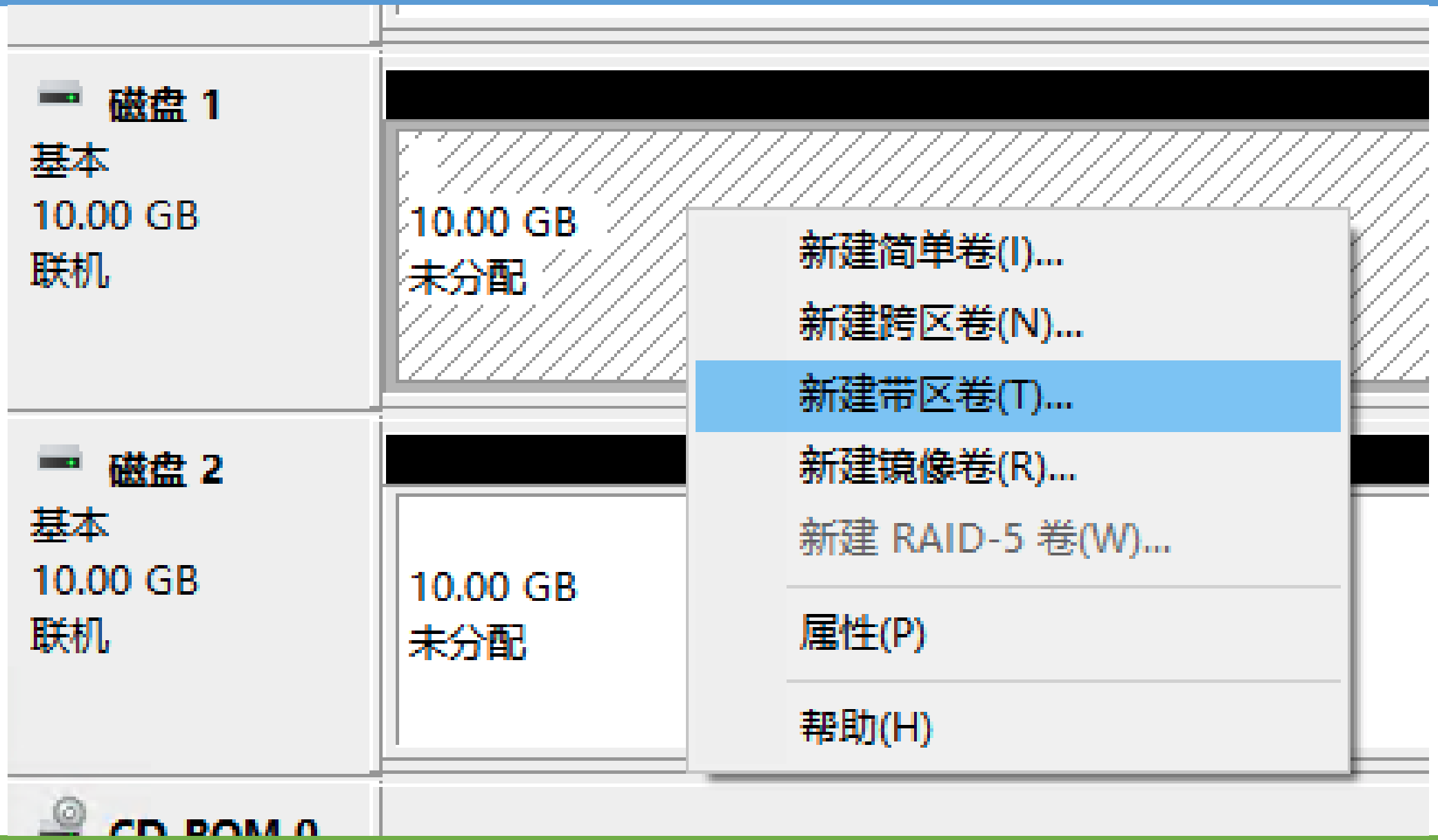
- 磁盘分区（Disk Partitioning）
- 将磁盘划分为若干逻辑部分，方便常用的数据存放在磁盘外道
- 磁盘分区可做看作是逻辑卷管理、软件 RAID 前身的一项简单技术
- 如果使用固态硬盘，磁盘分区会降低访问速度

Partition	Name	File System	Mount Point	Label	Size	Used	Unused	Flags
/dev/sda1	EFI system partition	fat32	/boot/efi	SYSTEM_DRV	260.00 MiB	49.77 MiB	210.23 MiB	boot, esp
/dev/sda2	Microsoft reserved partition	unknown			16.00 MiB	---	---	msftres
/dev/sda3	Basic data partition	ntfs		Windows	293.83 GiB	187.58 GiB	106.26 GiB	msftdata
/dev/sda11		btrfs	/home		183.17 GiB	114.90 GiB	68.27 GiB	
/dev/sda10		linux-swap			11.72 GiB	6.26 MiB	11.71 GiB	
unallocated		unallocated			2.93 GiB	---	---	
/dev/sda4	Basic data partition	ntfs		データやもの	264.30 GiB	220.86 GiB	43.44 GiB	msftdata
/dev/sda9		btrfs	/		127.64 GiB	77.26 GiB	50.38 GiB	
/dev/sda5	Basic data partition	ntfs		LENOVO	25.00 GiB	3.44 GiB	21.56 GiB	msftdata
/dev/sda6	Basic data partition	ntfs		WINRE_DRV	1000.00 MiB	520.09 MiB	479.91 MiB	hidden, diag
/dev/sda7	Basic data partition	ntfs		LENOVO_PART	20.71 GiB	12.40 GiB	8.31 GiB	diag
/dev/sda8	Basic data partition	fat32		LRS_ESP	1000.00 MiB	526.93 MiB	473.07 MiB	hidden

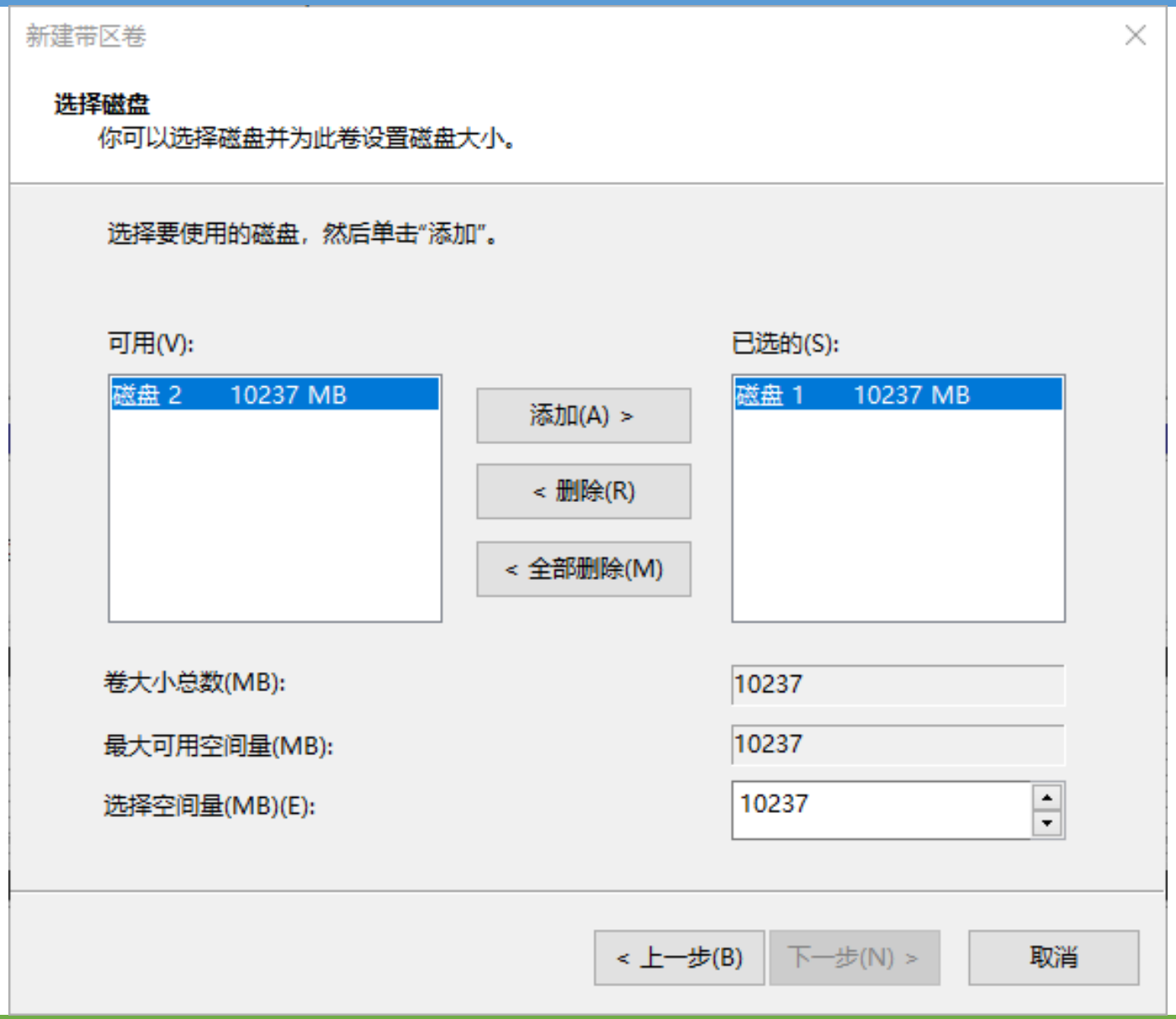
磁盘阵列（RAID）

在 Windows 上使用 RAID

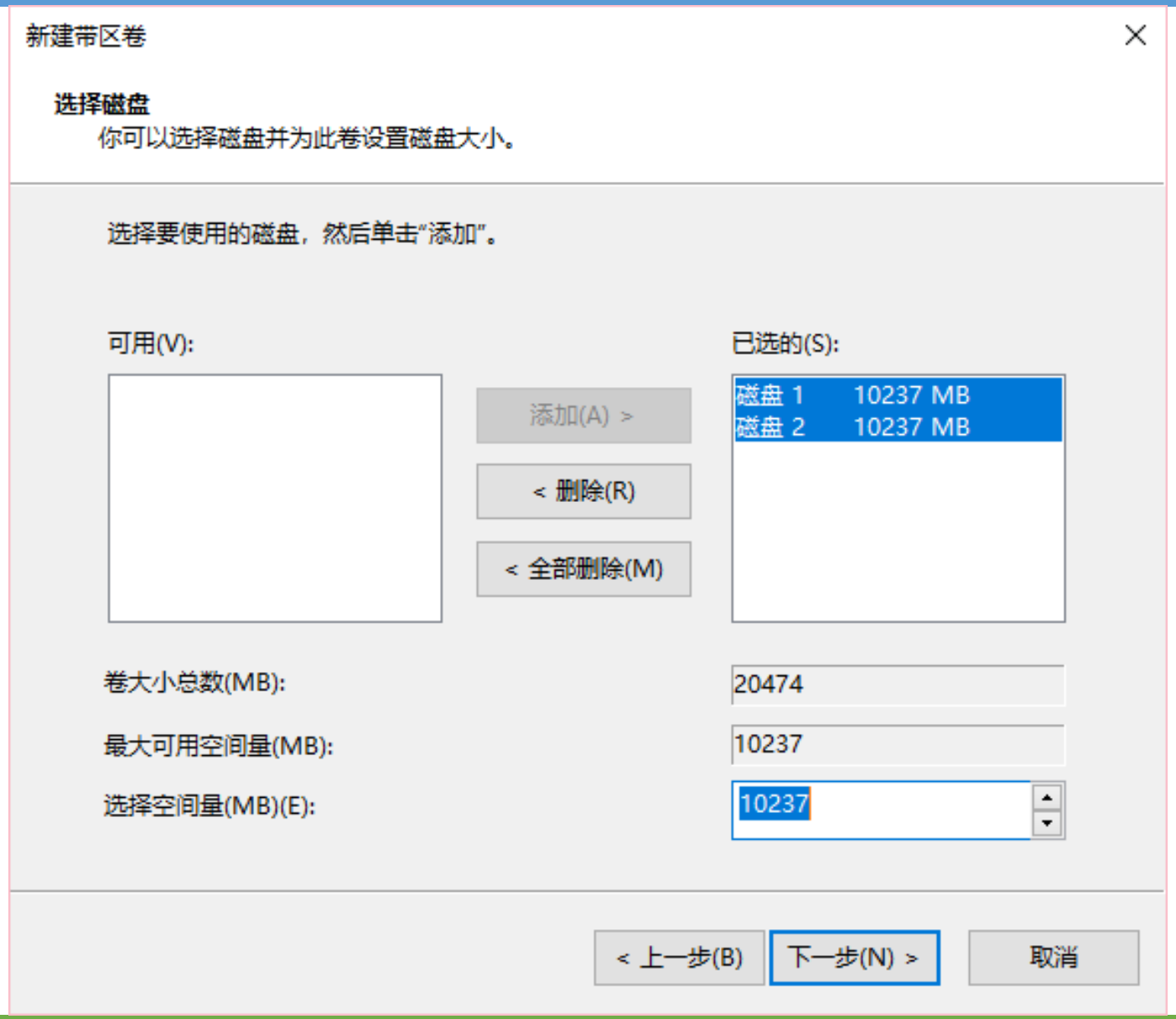


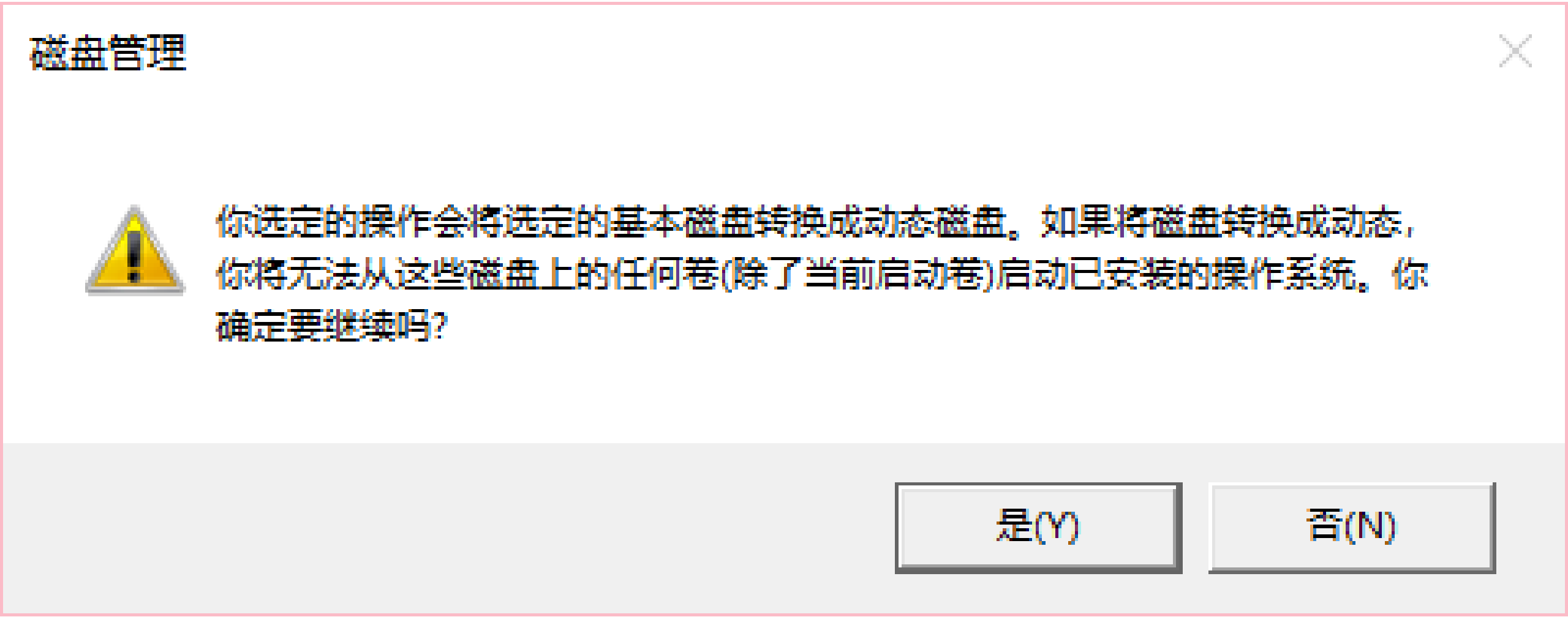

磁盘 1
基本
10.00 GB
联机
10.00 GB
未分配
新建简单卷(I)...
新建跨区卷(N)...
新建带区卷(T)...
新建镜像卷(R)...
新建 RAID-5 卷(W)...
属性(P)
帮助(H)
磁盘 2
基本
10.00 GB
联机
10.00 GB
未分配

在 Windows 上使用 RAID



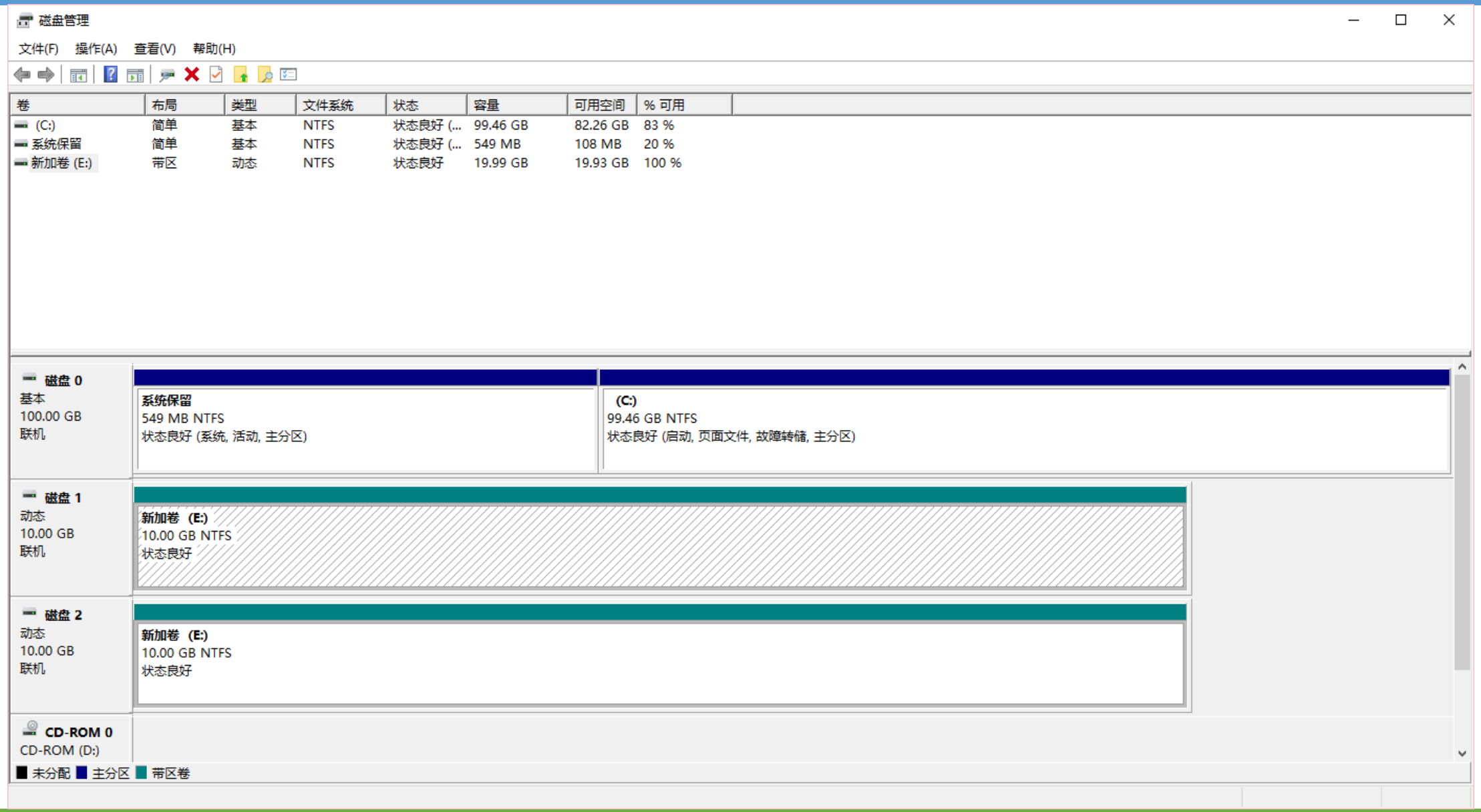
在 Windows 上使用 RAID




磁盘管理
你选定的操作会将选定的基本磁盘转换成动态磁盘。如果将磁盘转换成动态，
你将无法从这些磁盘上的任何卷(除了当前启动卷)启动已安装的操作系统。你
确定要继续吗?
是(Y)
否(N)

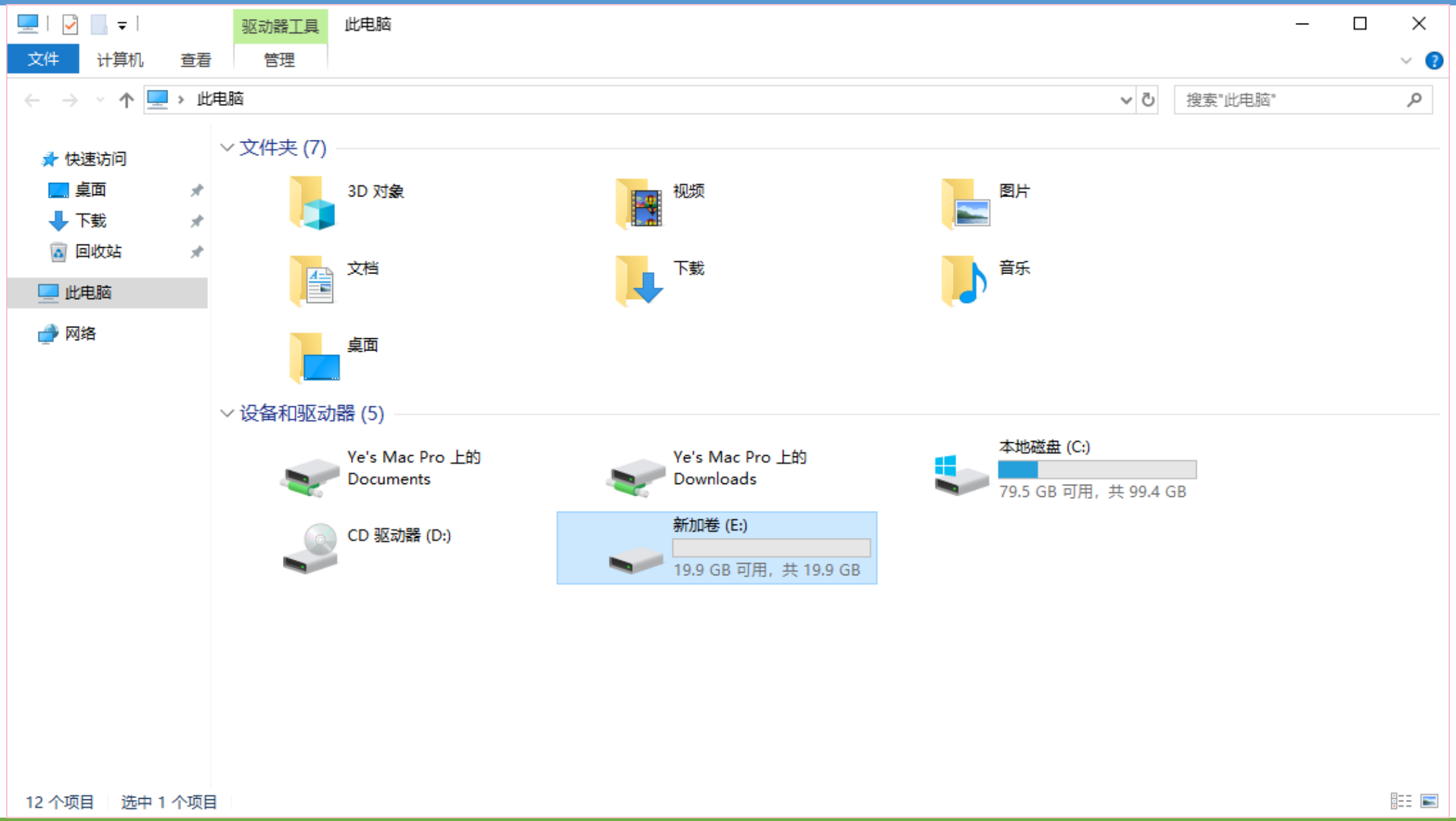
磁盘阵列（RAID）

在 Windows 上使用 RAID



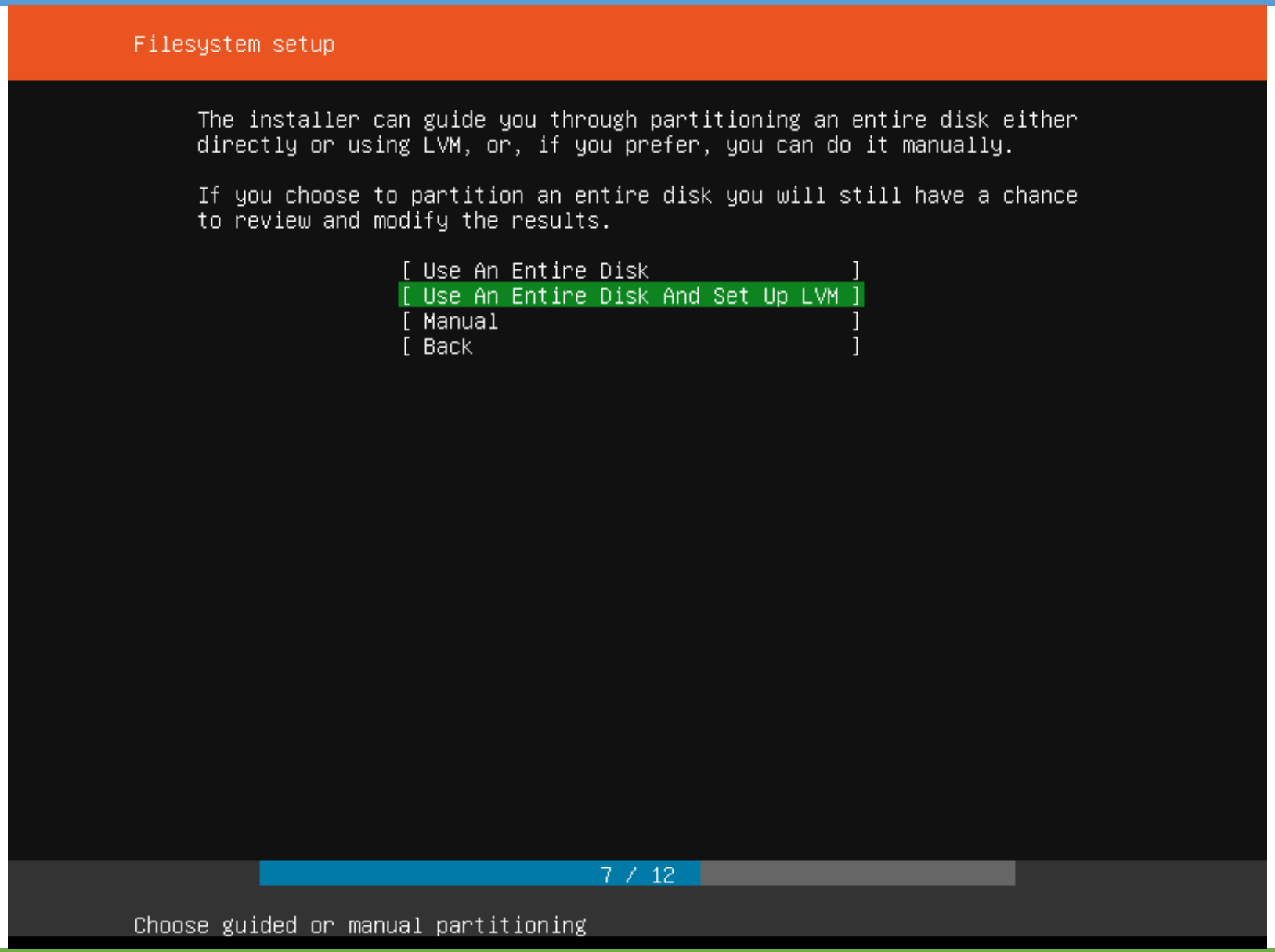
磁盘阵列（RAID）

在 Windows 上使用 RAID

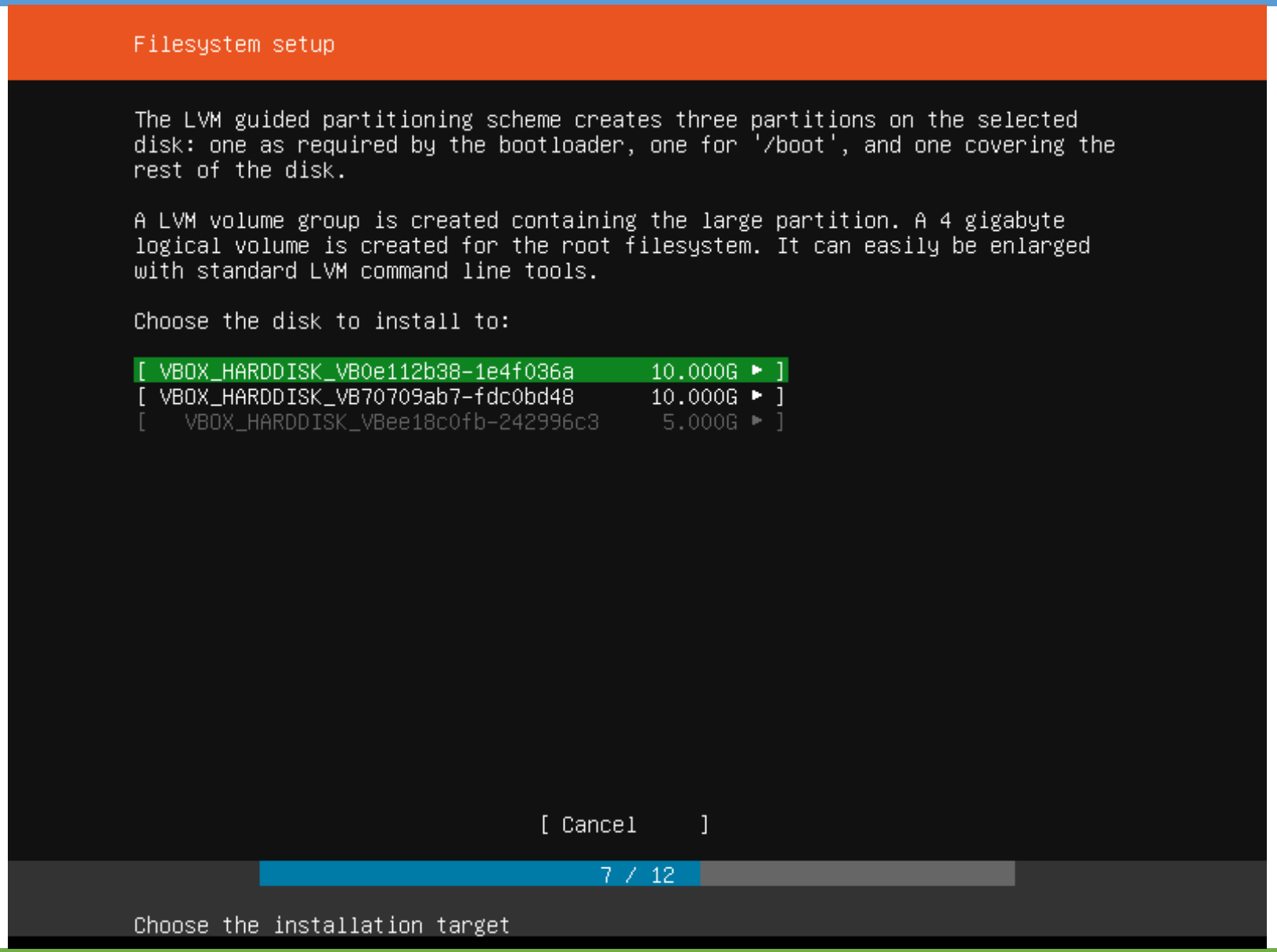


在 Linux 上使用 RAID

在 Linux 上使用 RAID

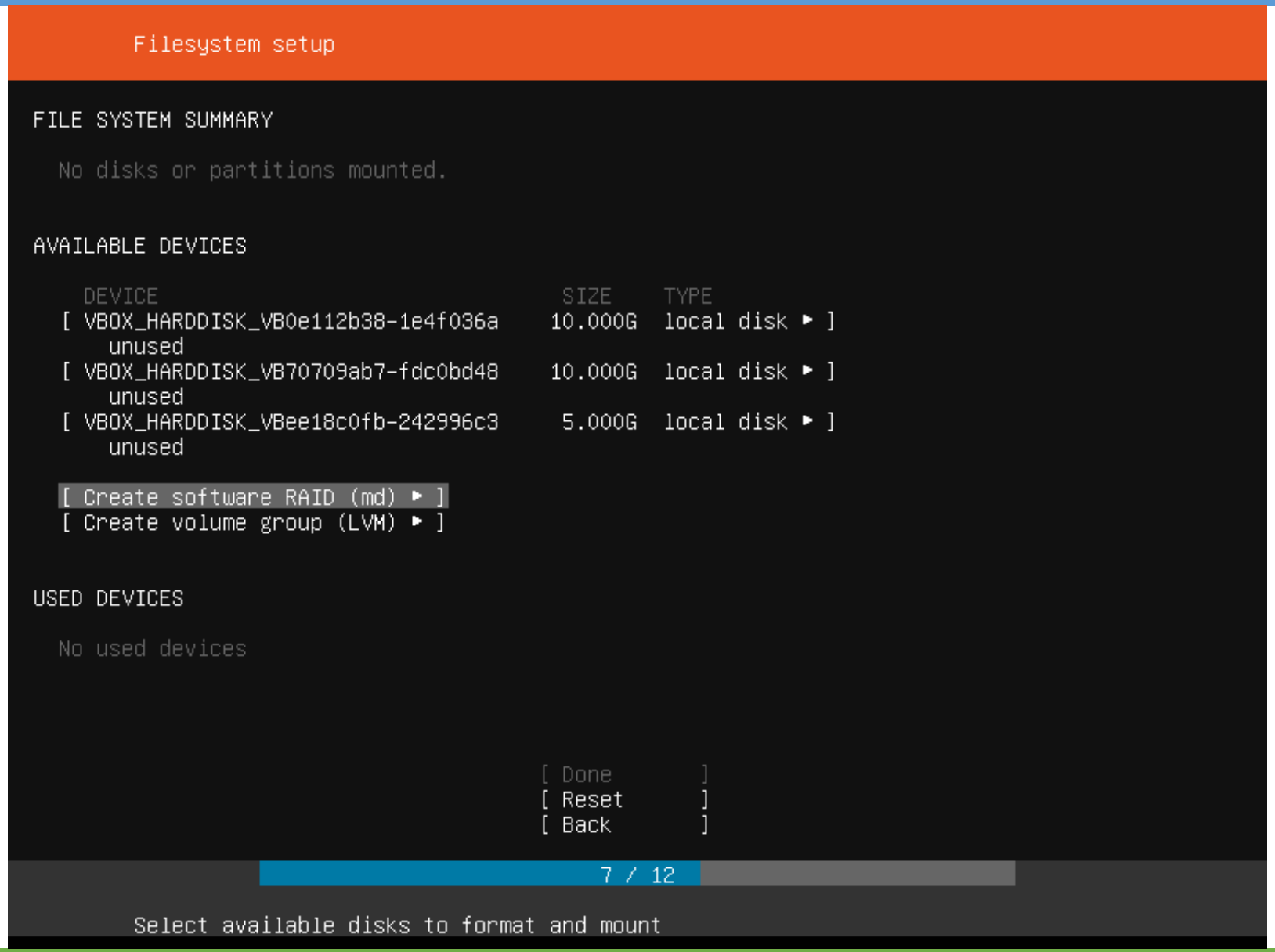


在 Linux 上使用 RAID



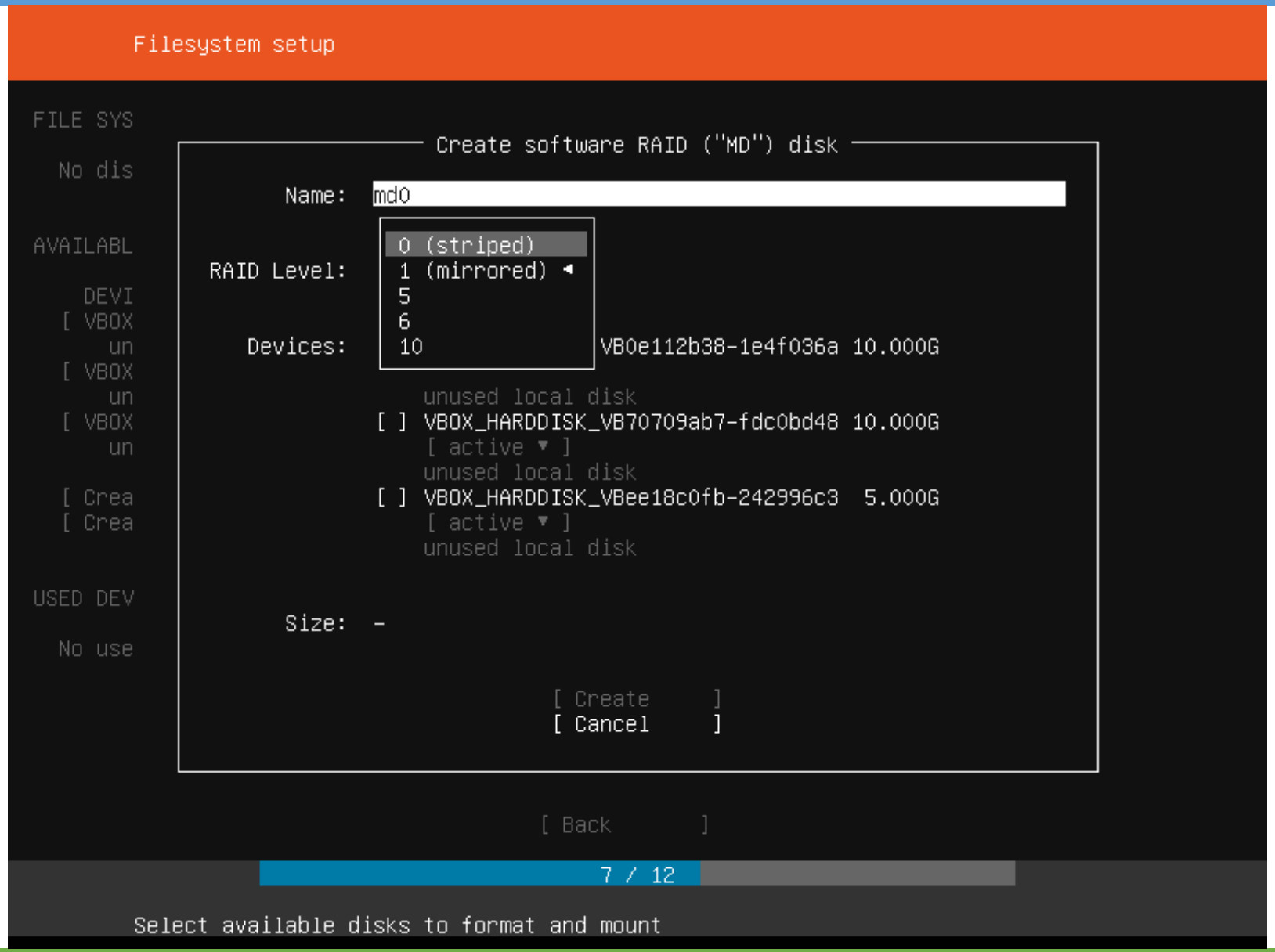
磁盘阵列（RAID）

在 Linux 上使用 RAID



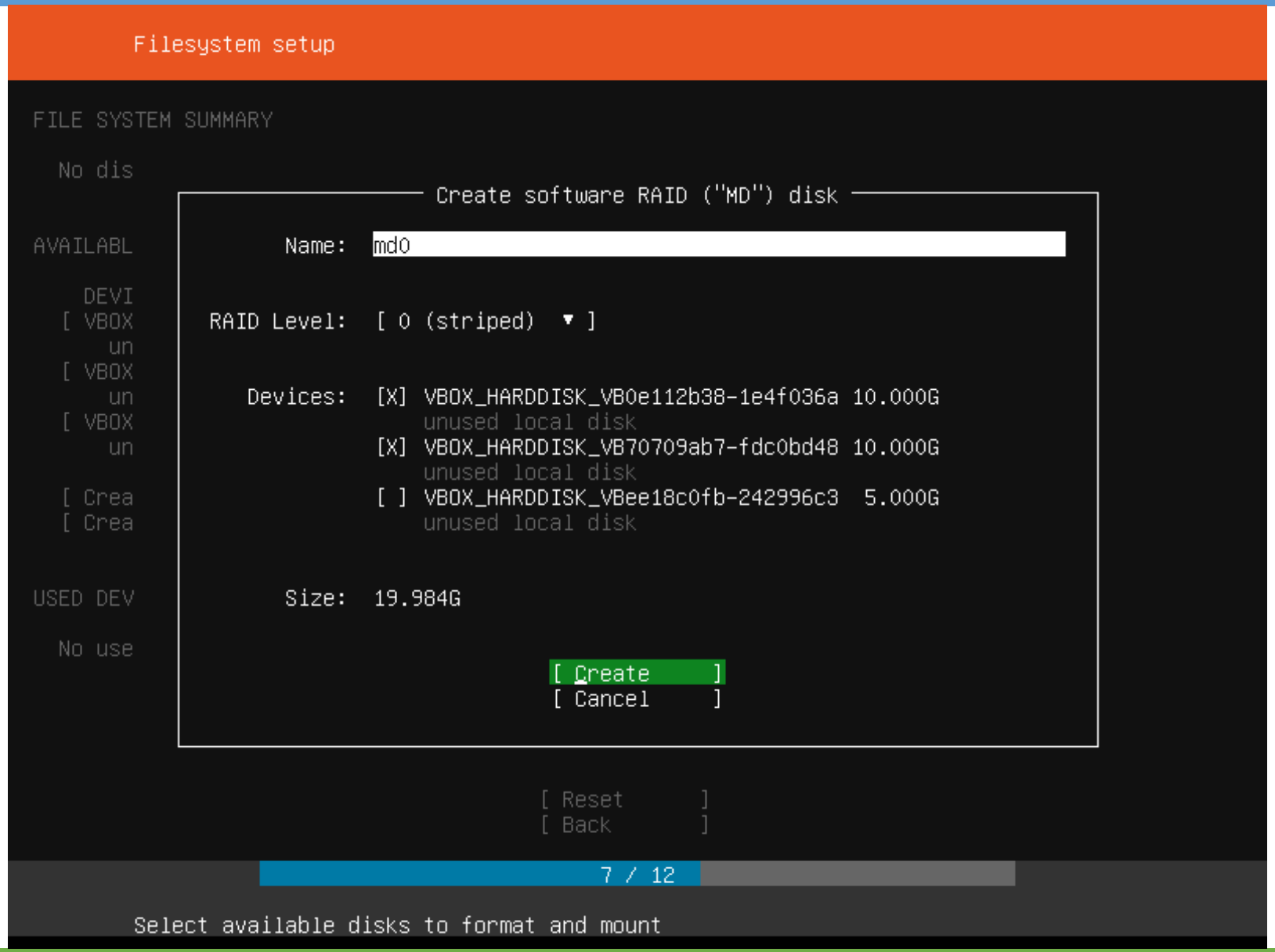
磁盘阵列（RAID）

在 Linux 上使用 RAID



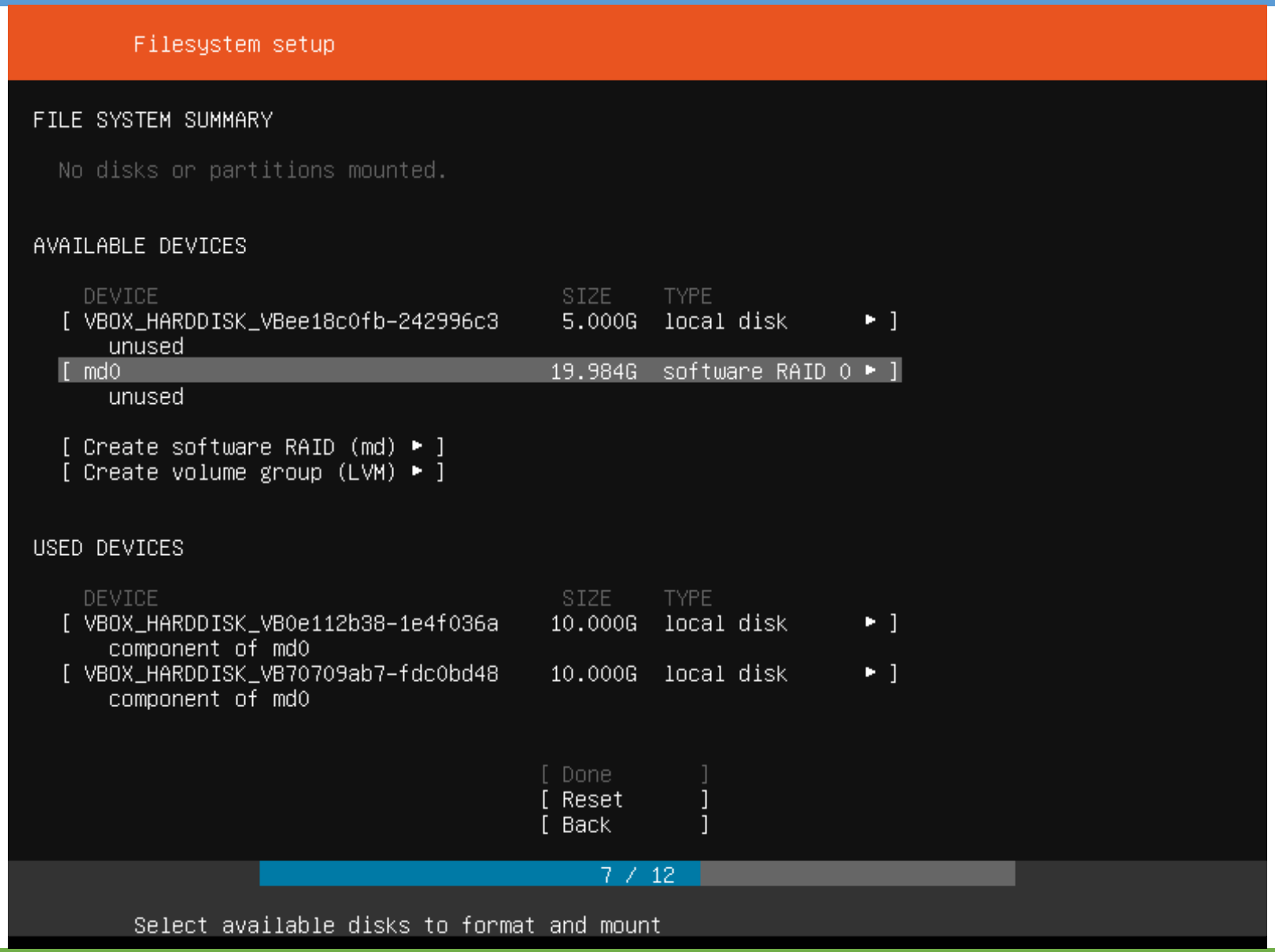
磁盘阵列（RAID）

在 Linux 上使用 RAID

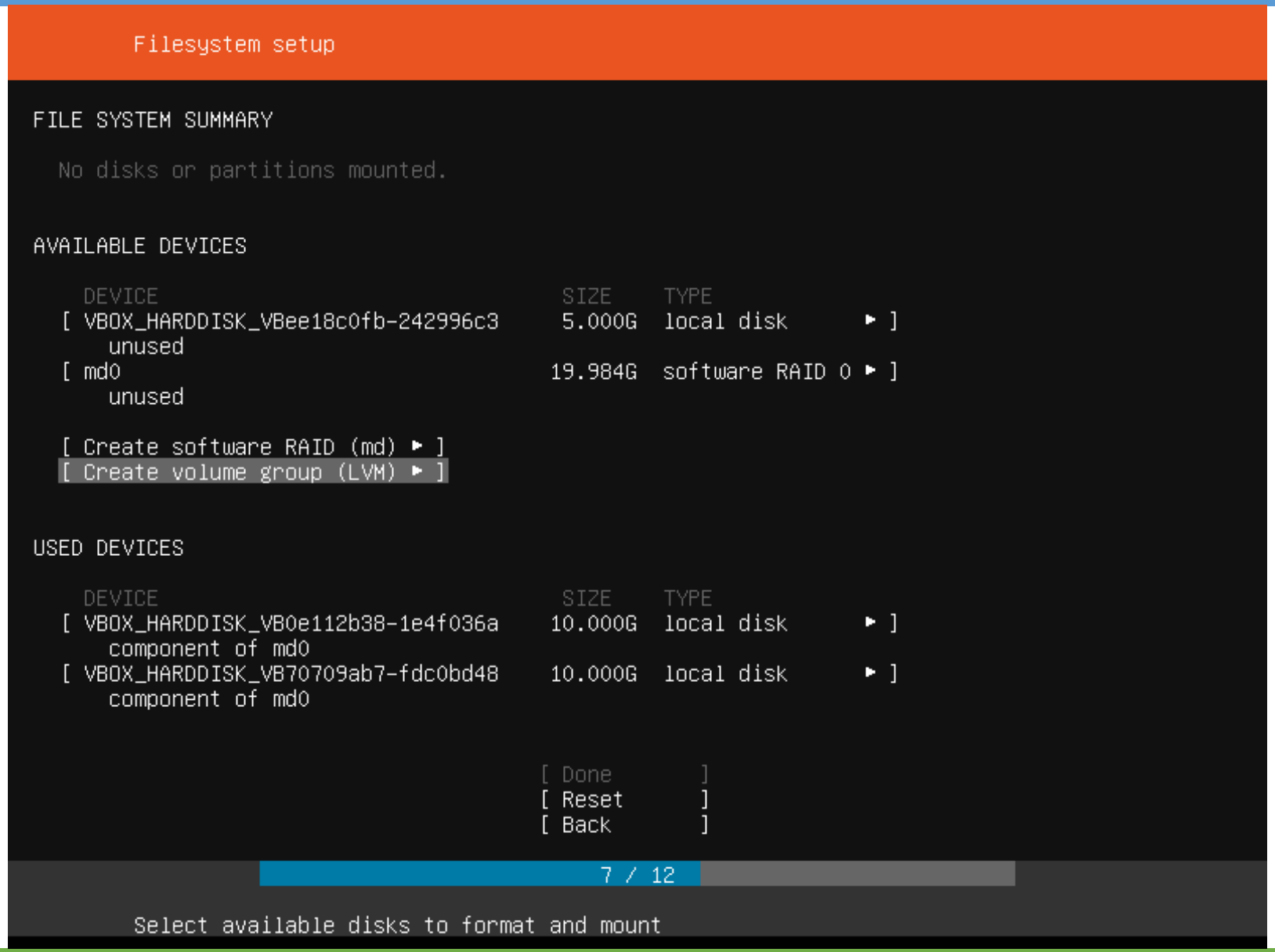


磁盘阵列（RAID）

在 Linux 上使用 RAID



逻辑卷管理器（LVM）


Filesystem setup
FILE SYSTEM SUMMARY
No disks or partitions mounted.
AVAILABLE DEVICES
DEVICE SIZE TYPE
[VBOX_HARDDISK_VBee18c0fb-242996c3 5.000G local disk ▸]
unused
[md0 19.984G software RAID 0 ▸]
unused
[Create software RAID (md) ▸]
[Create volume group (LVM) ▸]
USED DEVICES
DEVICE SIZE TYPE
[VBOX_HARDDISK_VB0e112b38-1e4f036a 10.000G local disk ▸]
component of md0
[VBOX_HARDDISK_VB70709ab7-fdc0bd48 10.000G local disk ▸]
component of md0
[Done]
[Reset]
[Back]
7 / 12
Select available disks to format and mount

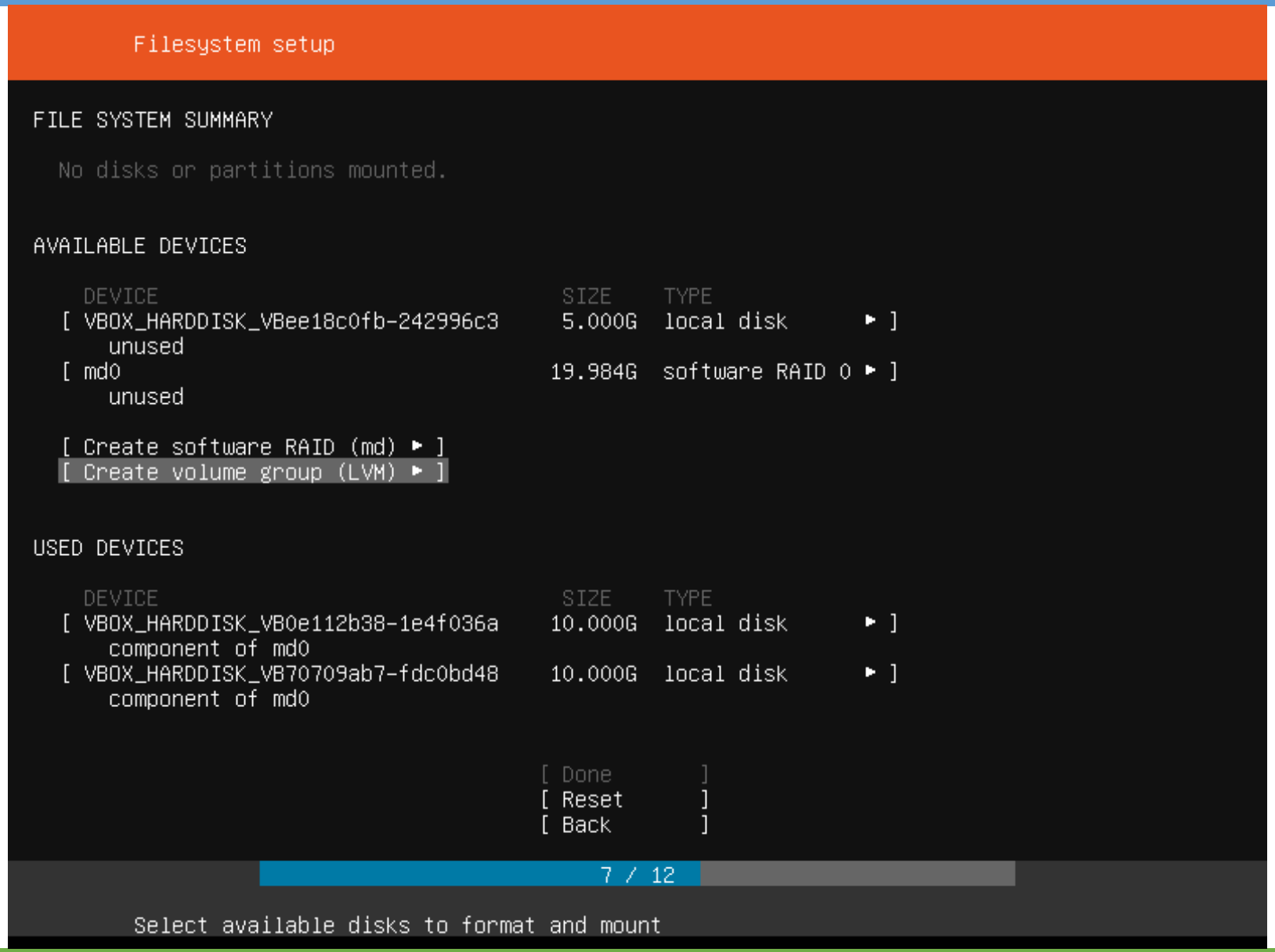
逻辑卷管理器（LVM）

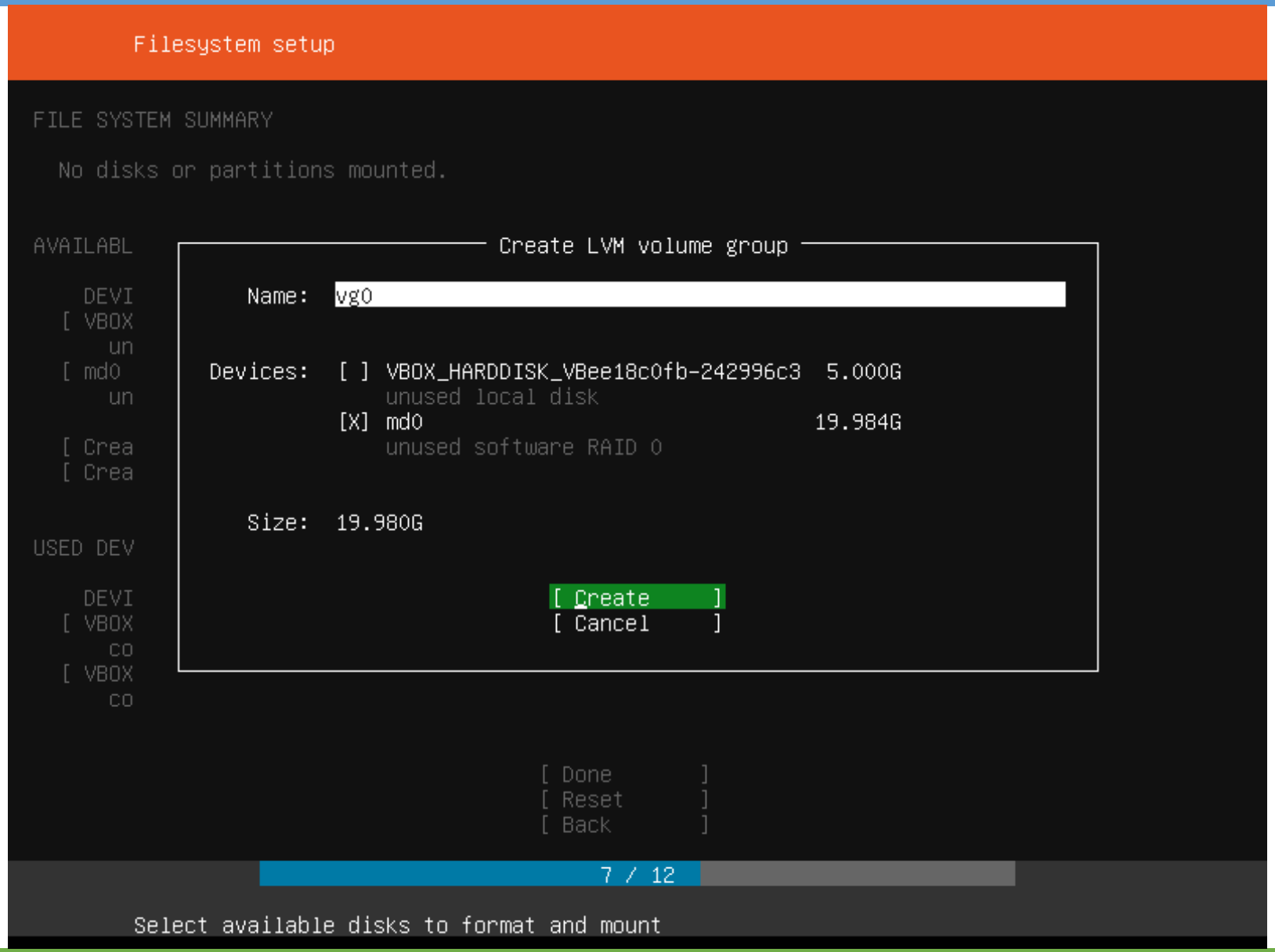
- 逻辑卷管理器（Logical Volume Manager，LVM）
- Linux 核心所提供的功能
- 它在硬盘的硬盘分区之上，又创建一个逻辑层，以方便系统管理硬盘分割系统
- 最先由 IBM 开发，在 AIX 系统上实现，OS/2 操作系统与 HP-UX 也支持这个功能
- 1998 年，Heinz Mauelshagen 根据在 HP-UX 上的逻辑卷管理器，开发了第一个 Linux 版本的逻辑卷管理器

逻辑卷管理器（LVM）

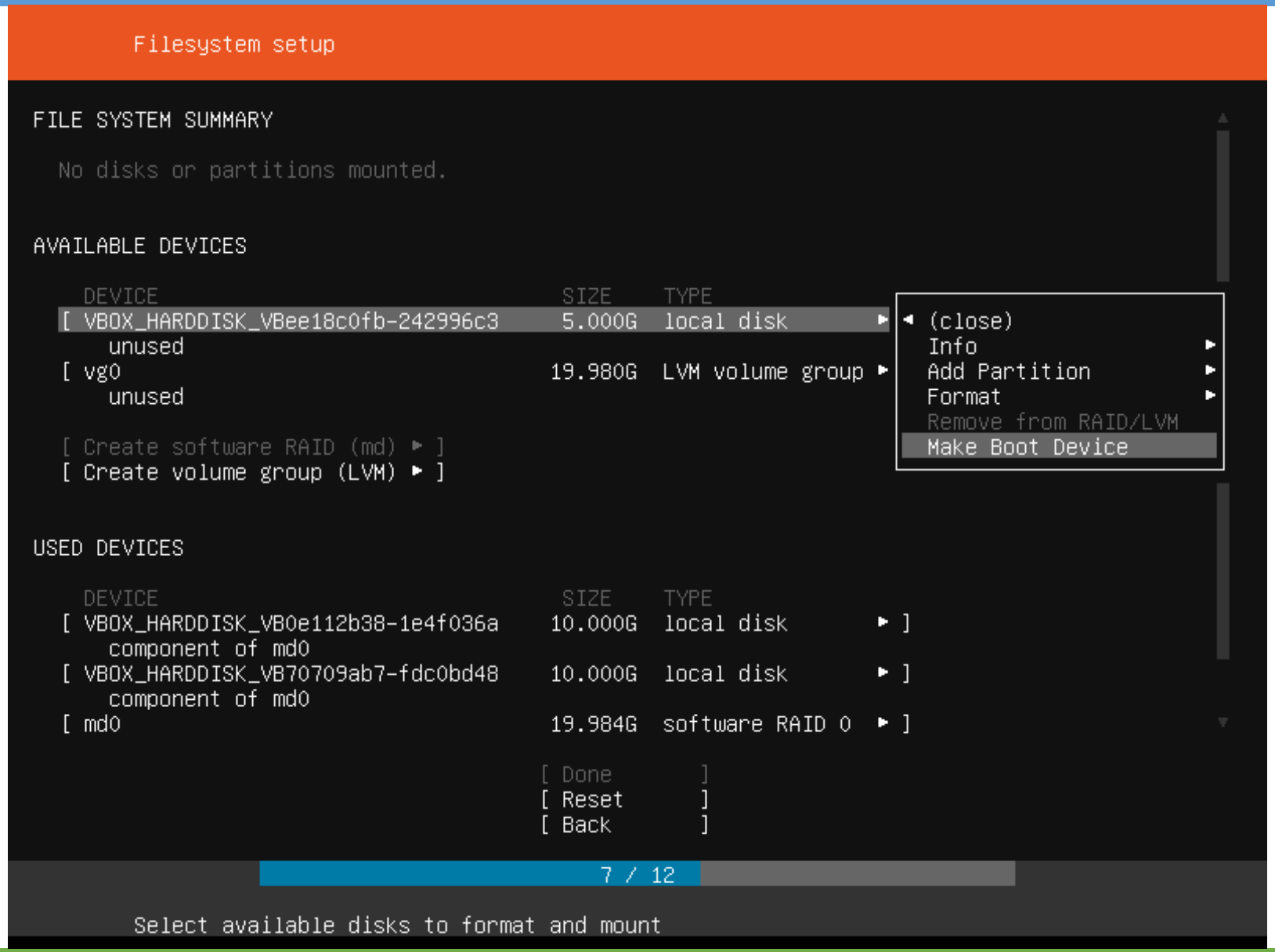
- LVM 基本术语：
- PV：物理卷，PV 处于 LVM 系统最低层，它可以是整个硬盘，或者与磁盘分区具有相同功能的设备（如 RAID），但和基本的物理存储介质相比较，多了与 LVM 相关管理参数
- VG：卷组，创建在 PV 之上，由一个或多个 PV 组成，可以在 VG 上创建一个或多个“LVM 分区”（逻辑卷），功能类似非 LVM 系统的物理硬盘
- LV：逻辑卷，从 VG 中分割出的一块空间，创建之后其大小可以伸缩，在 LV 上可以创建文件系统
- PE：物理区域，每一个 PV 被划分为基本单元（也被称为 PE），具有唯一编号的 PE 是可以被 LVM 寻址的最小存储单元，默认为 4 MB

逻辑卷管理器（LVM）


Filesystem setup
FILE SYSTEM SUMMARY
No disks or partitions mounted.
AVAILABLE DEVICES
DEVICE SIZE TYPE
[VBOX_HARDDISK_VBee18c0fb-242996c3 5.000G local disk ▸]
unused
[md0 19.984G software RAID 0 ▸]
unused
[Create software RAID (md) ▸]
[Create volume group (LVM) ▸]
USED DEVICES
DEVICE SIZE TYPE
[VBOX_HARDDISK_VB0e112b38-1e4f036a 10.000G local disk ▸]
component of md0
[VBOX_HARDDISK_VB70709ab7-fdc0bd48 10.000G local disk ▸]
component of md0
[Done]
[Reset]
[Back]
7 / 12
Select available disks to format and mount

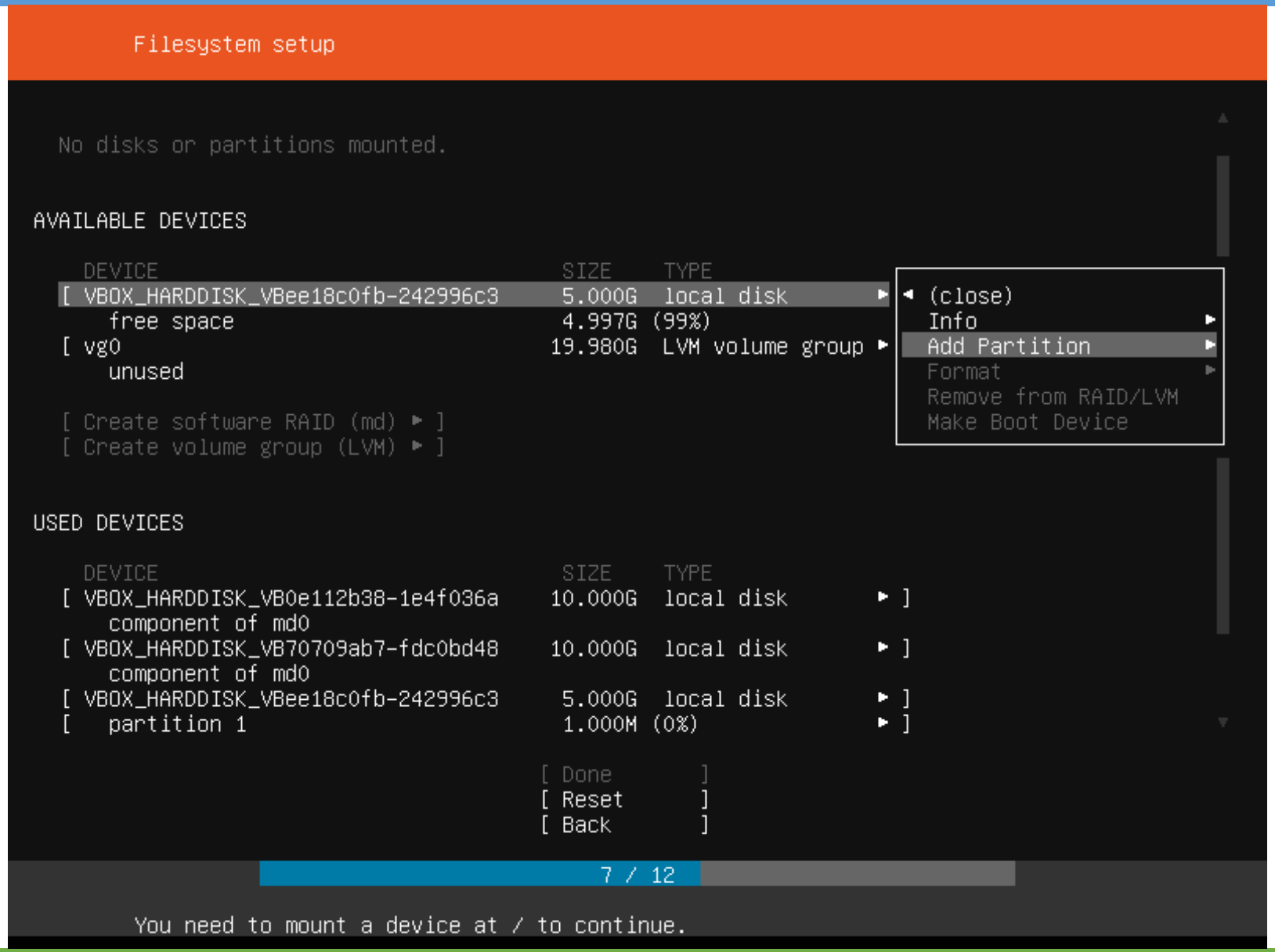

Filesystem setup
FILE SYSTEM SUMMARY
No disks or partitions mounted.
AVAILABL
DEVI
[VBOX
un
[md0
un
[Crea
[Crea
USED DEV
DEVI
[VBOX
co
[VBOX
co
Create LVM volume group
Name: vg0
Devices: [ ] VBOX_HARDDISK_VBee18c0fb-242996c3 5.000G
unused local disk
[X] md0 19.984G
unused software RAID 0
Size: 19.980G
[Create]
[Cancel]
[Done]
[Reset]
[Back]
7 / 12
Select available disks to format and mount

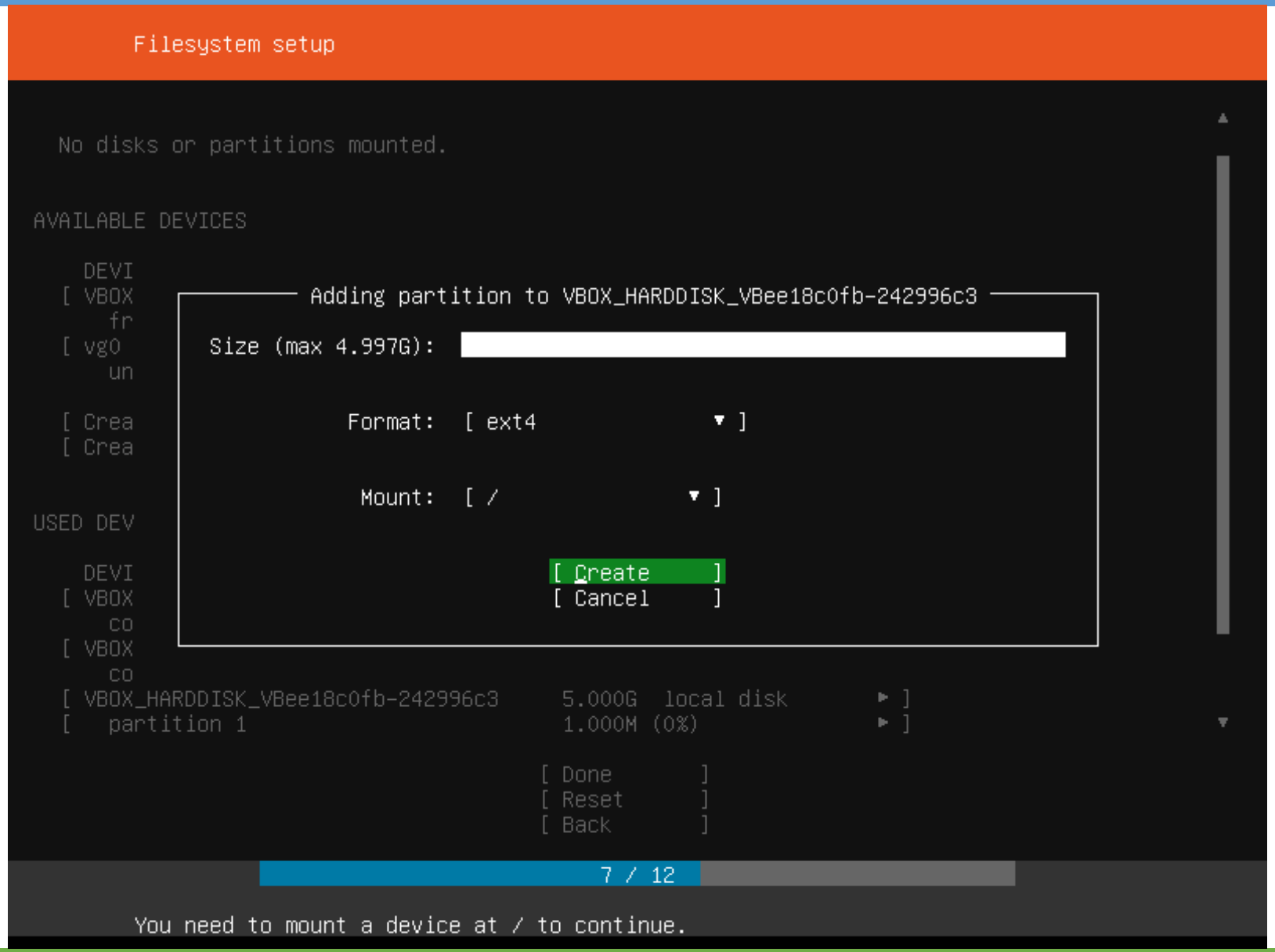
逻辑卷管理器（LVM）


Filesystem setup
FILE SYSTEM SUMMARY
No disks or partitions mounted.
AVAILABLE DEVICES
DEVICE SIZE TYPE
[VBOX_HARDDISK_VBee18c0fb-242996c3 5.000G local disk ▸
unused
[vg0 19.980G LVM volume group ▸
unused
[Create software RAID (md) ▸]
[Create volume group (LVM) ▸]
◂ (close)
Info ▸
Add Partition ▸
Format ▸
Remove from RAID/LVM
Make Boot Device
USED DEVICES
DEVICE SIZE TYPE
[VBOX_HARDDISK_VB0e112b38-1e4f036a 10.000G local disk ▸]
component of md0
[VBOX_HARDDISK_VB70709ab7-fdc0bd48 10.000G local disk ▸]
component of md0
[md0 19.984G software RAID 0 ▸]
[Done]
[Reset]
[Back]
7 / 12
Select available disks to format and mount

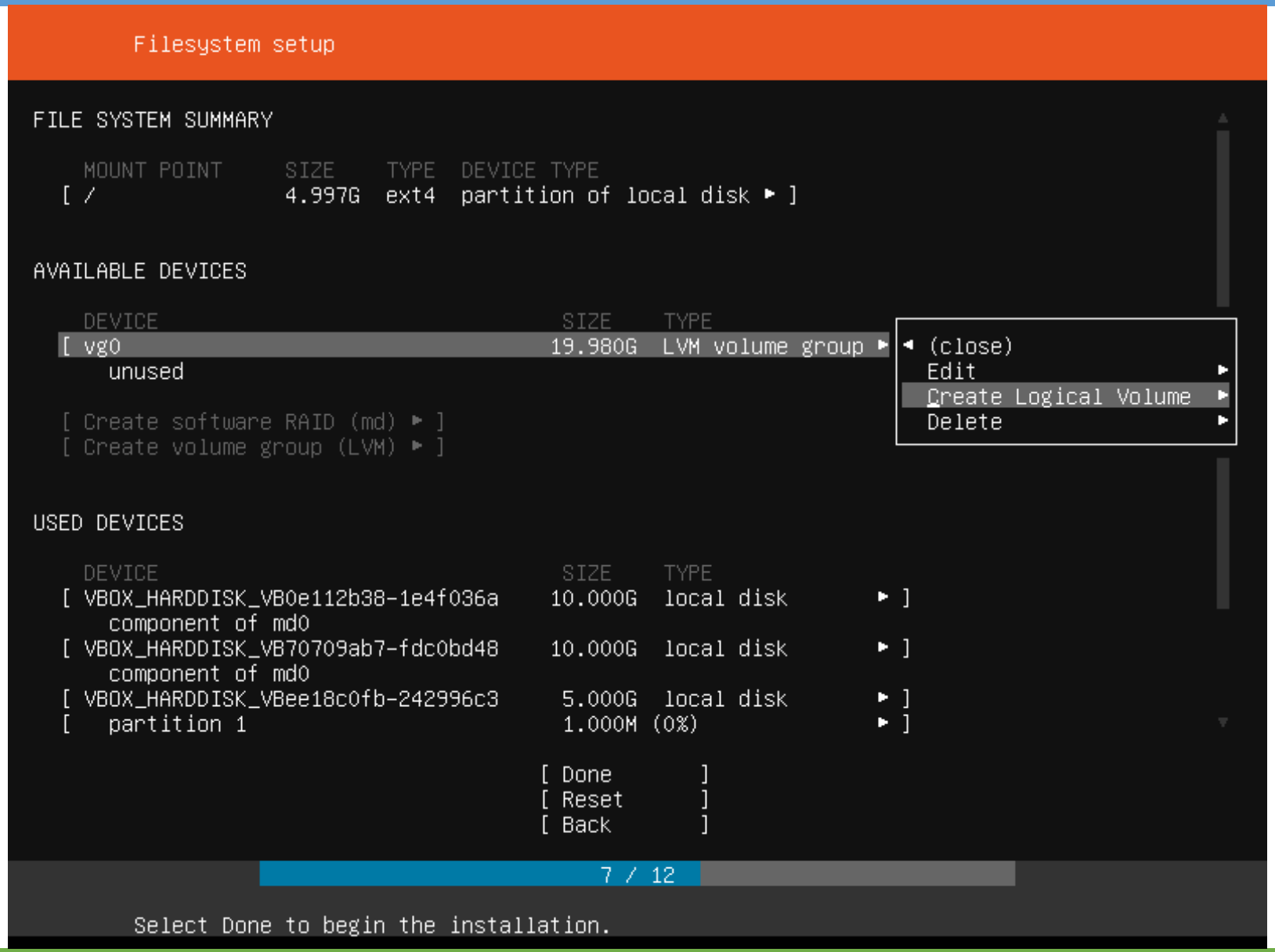
磁盘阵列（RAID）

逻辑卷管理器（LVM）



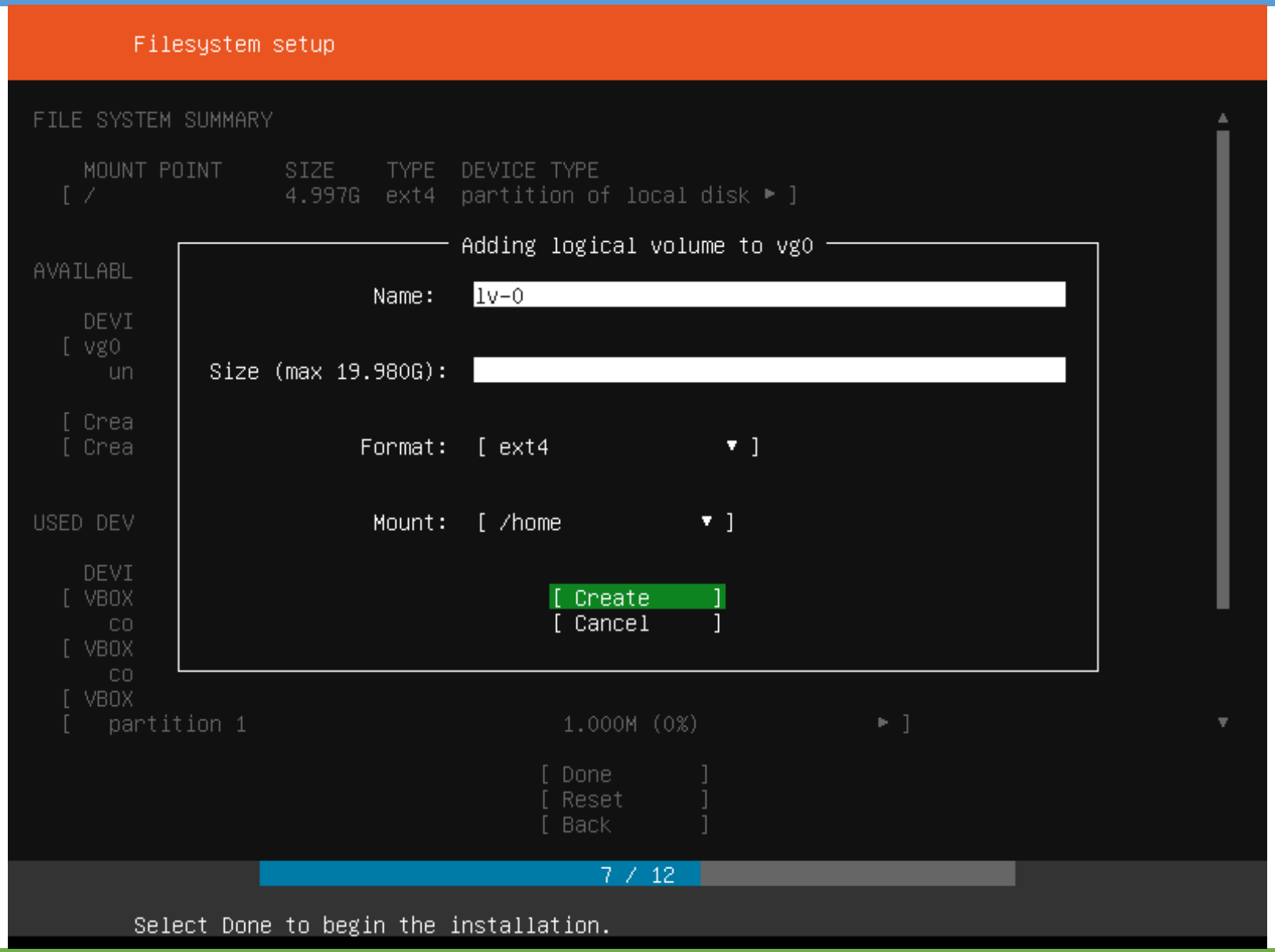

Filesystem setup
No disks or partitions mounted.
AVAILABLE DEVICES
DEVI
[VBOX
fr
[vg0
un
[Crea
[Crea
USED DEV
DEVI
[VBOX
co
[VBOX
co
[VBOX_HARDDISK_VBee18c0fb-242996c3 5.000G local disk ▸]
[partition 1 1.000M (0%) ▸]
[Done]
[Reset]
[Back]
Adding partition to VBOX_HARDDISK_VBee18c0fb-242996c3
Size (max 4.997G):
Format: [ext4 ▾]
Mount: [/ ▾]
[Create]
[Cancel]
7 / 12
You need to mount a device at / to continue.

逻辑卷管理器（LVM）


Filesystem setup
FILE SYSTEM SUMMARY
MOUNT POINT SIZE TYPE DEVICE TYPE
[/ 4.997G ext4 partition of local disk ▸]
AVAILABLE DEVICES
DEVICE SIZE TYPE
[vg0 19.980G LVM volume group ▸
unused
[Create software RAID (md) ▸]
[Create volume group (LVM) ▸]
◂ (close)
Edit ▸
Create Logical Volume ▸
Delete ▸
USED DEVICES
DEVICE SIZE TYPE
[VBOX_HARDDISK_VB0e112b38-1e4f036a 10.000G local disk ▸]
component of md0
[VBOX_HARDDISK_VB70709ab7-fdc0bd48 10.000G local disk ▸]
component of md0
[VBOX_HARDDISK_VBee18c0fb-242996c3 5.000G local disk ▸]
[partition 1 1.000M (0%) ▸]
[Done]
[Reset]
[Back]
7 / 12
Select Done to begin the installation.

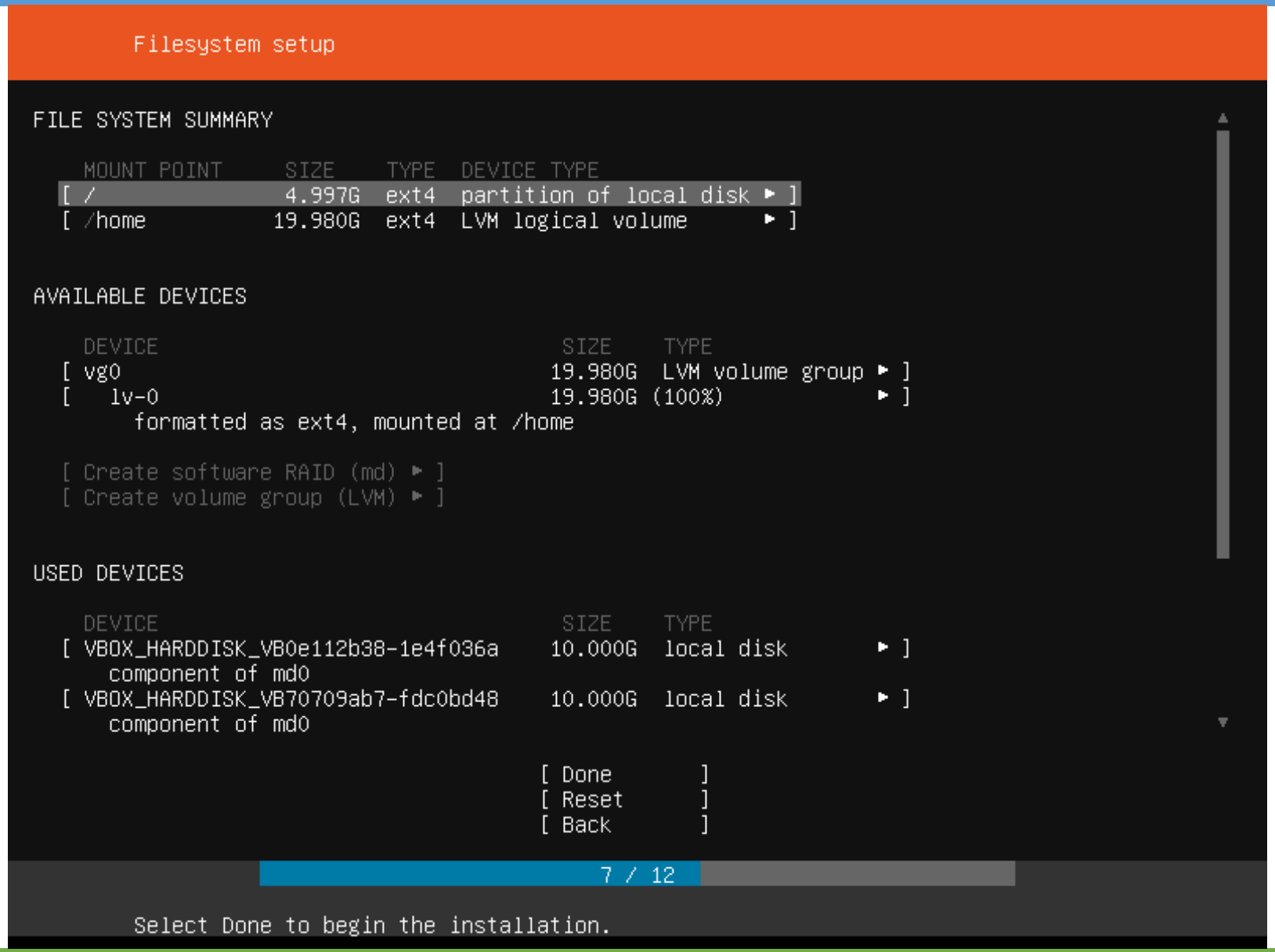
磁盘阵列（RAID）

逻辑卷管理器（LVM）



磁盘阵列（RAID）

逻辑卷管理器（LVM）



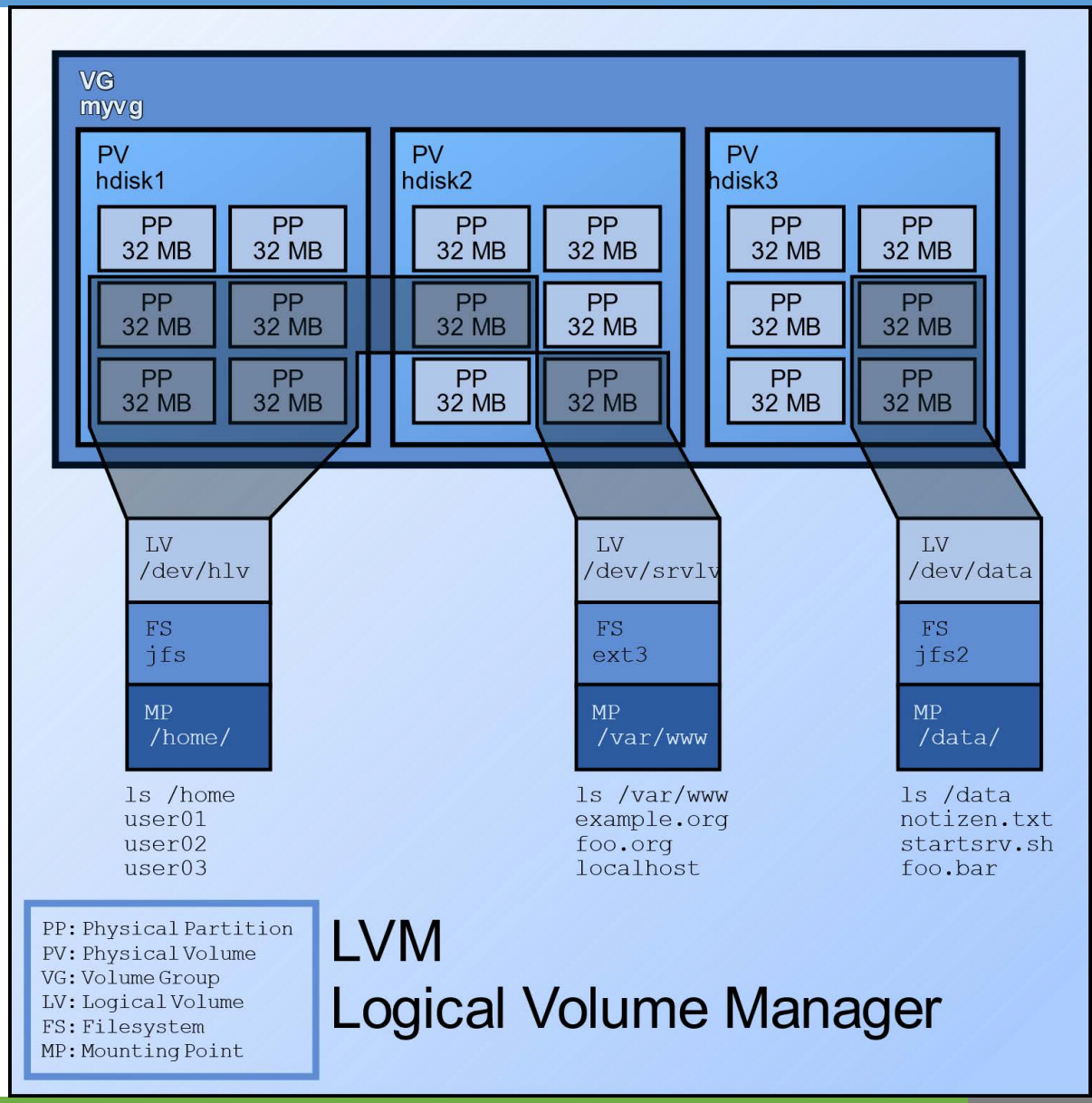
逻辑卷管理器（LVM）

```
  Memory usage:   3%                IP address for enp0s3: 10.0.2.15
  Swap usage:     0%

154 packages can be updated.
69 updates are security updates.



The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

valency@course:~$ df -h
Filesystem             Size  Used Avail Use% Mounted on
udev                   1.9G     0  1.9G   0% /dev
tmpfs                  395M  996K  394M   1% /run
/dev/sda2              4.9G  1.9G  2.8G  41% /
tmpfs                  2.0G     0  2.0G   0% /dev/shm
tmpfs                  5.0M     0  5.0M   0% /run/lock
tmpfs                  2.0G     0  2.0G   0% /sys/fs/cgroup
/dev/mapper/vg0-lv--0   20G   45M   19G   1% /home
/dev/loop0              91M   91M     0 100% /snap/core/6350
tmpfs                  395M     0  395M   0% /run/user/1000
valency@course:~$ sudo cat /proc/mdstat
[sudo] password for valency:
Personalities : [raid0] [linear] [multipath] [raid1] [raid6] [raid5] [raid4] [raid10]
md0 : active raid0 sdc[1] sdb[0]
      20953088 blocks super 1.2 512k chunks

unused devices: <none>
valency@course:~$ _
```

逻辑卷管理器（LVM）

- LVM 最大的优点是：
- 将多个硬盘整合成了一套虚拟的存储系统
- 存储系统可以随时动态修改大小

- 如何动态调整 LVM 大小：
- https://www.rootusers.com/how-to-increase-the-size-of-a-linux-lvm-by-expanding-the-virtual-machine-disk/



总结

- 三大云服务平台
- Google Cloud：https://cloud.google.com
- Amazon Web Services：https://aws.amazon.com
- 阿里云：https://www.aliyun.com

总结

- 课外阅读
- 《云存储技术——分析与实践》，刘洋著，经济管理出版社
- http://product.dangdang.com/24247525.html

- 《Ahead in the Cloud》，Stephen Orban（GM of AWS）
- https://www.amazon.com/Ahead-Cloud-Practices-Navigating-Enterprise/dp/1981924310/

- 《Cloud Computing: Concepts, Technology & Architecture》，Thomas Erl
- https://www.amazon.com/Cloud-Computing-Concepts-Technology-Architecture/dp/0133387526/

Thanks!