

RAID 独立冗余磁盘阵列

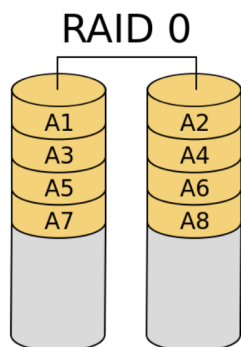
RAID 将一组硬盘连结起来，组成一个阵列，以避免单个硬盘损坏而带来的数据损失，同时亦提供了比单个硬盘高的可用性及容错性。

条带化技术，分散存储在多个盘上。(做切割数据的，存在盘上的对应位置，在外观看来就是条带状的)

raid级别：从读写速度，可靠性来考虑，代表raid的组成方式是不一样的，没有上下级之分一：RAID级别

1.RAID0

RAID0 俗称“条带”，它将两个或多个硬盘组成一个逻辑硬盘，容量是所有硬盘之和，因为是多个硬盘组合成一个，故可并行写操作，写入速度提高，但此方式硬盘数据没有冗余，没有容错，一旦一个物理硬盘损坏，则所有数据均丢失。因而，RAID0 适合于对数据量大，但安全性要求不高的场景，比如音像、视频文件的存储等。



性能提升：读写

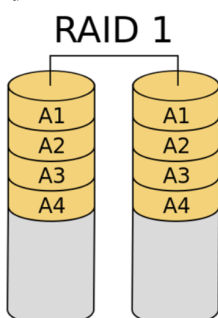
冗余能力：不具备

空间利用率：n

至少两块盘

2.RAID1

RAID1 俗称“镜像”，它最少由两个硬盘组成，且两个硬盘上存储的数据均相同，以实现数据冗余。RAID1 读操作速度有所提高，写操作理论上与单硬盘速度一样，但由于数据需要同时写入所有硬盘，实际上稍为下降。容错性是所有组合方式里最好的，只要有一块硬盘正常，则能保持正常工作。但它对硬盘容量的利用率则是最低，只有 50%，因而成本也是最高。RAID1 适合对数据安全性要求非常高的场景，比如存储数据库数据文件之类。



性能提升：写性能下降，读性能提高

冗余能力：具备

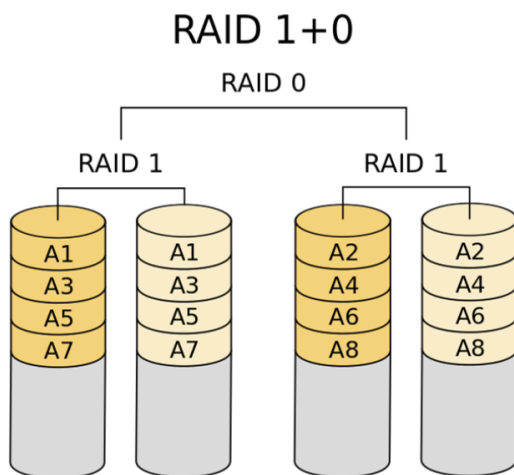
空间利用率：1/2

正好两个

3.RAID 1+0

RAID1+0 是先将数据进行镜像操作，然后再对数据进行分组， RAID 1 在这里就是一个冗余的备份阵列，而 RAID 0 则负责数据的读写阵列。至少要四块盘，两块做 raid1,另两块做 raid0， RAID 10 对存储容量的利用率和 RAID 1 一样低，只有 50%.Raid1+0 方案造成了 50%的磁盘浪费，但是它提供了 200%的速度和单磁盘损坏的数据安全性，并且当同时损坏的磁盘不在同一 Raid1 中，就能保证数据安全性、 RAID 10 能提供比RAID 5 更好的性能。这种新结构的可扩充性不好，使用此方案比较昂贵，

RAID1+0（同组不能都坏掉）用的多



性能表现：读写提升

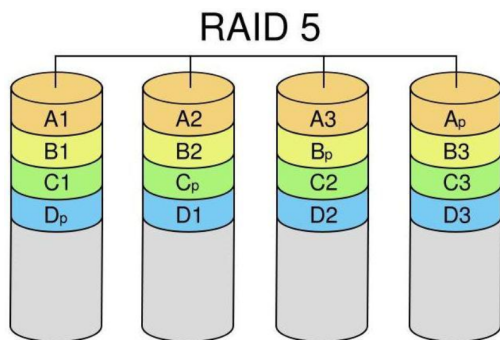
冗余能力：具备

空间利用率：1/2

至少4块

4.RAID 5

RAID5 最少由三个硬盘组成，它将数据分散存储于阵列中的每个硬盘，并且还伴有一个数据校验位，数据位与校验位通过算法能相互验证，当丢失其中的一位时， RAID 控制器能通过算法，利用其它两位数据将丢失的数据进行计算还原。因而 RAID5 最多能允许一个硬盘损坏，有容错性。 RAID5 相对于其它的组合方式，在容错与成本方面有一个平衡，因而受到大多数使用者的欢迎。一般的磁盘阵列，最常使用的就是 RAID5 这种方式。



性能表现：读写提升

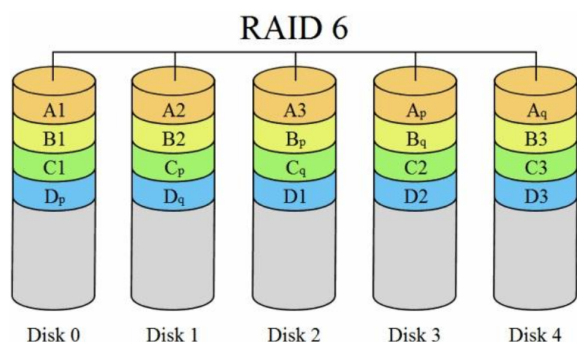
冗余能力：具备

空间利用率： $(n-1)/n$

至少3块

5. RAID 6

RAID6 是在 RAID5 的基础上改良而成的， RAID6 再将数据校验位增加一位，所以允许损坏的硬盘数量也由 RAID5 的一个增加到二个。由于同一阵列中两个硬盘同时损坏的概率非常少，所以， RAID6 用增加一块硬盘的代价，换来了比 RAID5 更高的数据安全性。



Disk 0 Disk 1 Disk 2 Disk 3 Disk 4

以上是当前主流的 RAID 组合方式，至于如何选择，则是结合到使用者对数据安全性、数据冗余、数据可用性、投入成本等各方面的综合衡量。建议：

- 1、系统盘配置为RAID 1（镜像卷）
- 2、存储盘配置为RAID 10

硬件raid:

- 1、阵列卡就是用来实现RAID功能的板卡，通常是由I/O处理器、硬盘控制器、硬盘连接器和缓存等一系列零组件构成的
- 2、不同的RAID卡支持的RAID功能不同，例如支持RAID0、RAID1、RAID5、RAID10等
- 3、RAID卡的接口类型
 - IDE接口、SCSI接口、SATA接口和SAS接口
- 4、缓存（Cache）是RAID卡与外部总线交换数据的场所，RAID卡先将数据传送到缓存，再由缓存和外边数据总线交换数据，缓存的大小与速度是直接关系到RAID卡的实际传输速

度的重要因素，不同的RAID卡出厂时配备的内存容量不同，一般为几兆到数百兆容量不等。

通过PCIe3.0接口技术，阵列卡带宽高达12Gbp/S。大容量高速缓存，支持1、5、6、10等多种Raid阵列选项



阵列卡介绍

阵列卡的全名叫磁盘阵列卡 是用来做 RAID 的。磁盘阵列是一种把若干硬磁盘驱动器按照一定要求组成一个整体，整个磁盘阵列由阵列控制器管理的系统。冗余磁盘阵列 RAID(Redundant Array of Independent Disks)技术 1987 年由加州大学伯克利分校提出，最初的研制目的是为了组合小的廉价磁盘来代替大的昂贵磁盘，以降低大批量数据存储的费用（当时 RAID 称为 dundant Array of Inexpensive Disks 廉价的磁盘阵列），同时也希望采用冗余信息的方式，使得磁盘失效时不会使对数据的访问受损失，从而开发出一定水平的数据保护技术。

阵列卡接口类型

接口类型目前主要有几种： IDE 接口、 SCSI 接口、 SATA 接口和 SAS 接口。

1.IDE 接口

IDE 的英文全称为 “Integrated Drive Electronics” ，即 “电子集成驱动器” ，它的本意是指把 “硬盘控制器” 与 “盘体” 集成在一起的硬盘驱动器。把盘体与控制器集成在一起的做法减少了硬盘接口的电缆数目与长度，数据传输的可靠性得到了增强，硬盘制造起来变得更容易，因为硬盘生产厂商不需要再担心自己的硬盘是否与其它厂商生产的控制器兼容，对用户而言，硬盘安装起来也更为方便。 IDE 这一接口技术从诞生至今就一直在不断发展，性能也不断的提高，其拥有的价格低廉、兼容性强的特点，为其造就了其它类型硬盘无法替代的地位。 IDE 代表着硬盘的一种类型，但在实际的应用中，人们也习惯用 IDE 来称呼最早出现 IDE类型硬盘 ATA-1，这种类型的接口随着接口技术的发展已经被淘汰了，而其后发展分支出更多类型的硬盘接口，比如 ATA、 Ultra ATA、 DMA、 Ultra DMA 等

接口都属于 IDE 硬盘。此外，由于 IDE 口属于并行接口，因此为了和 SATA 口硬盘相区别，IDE 口硬盘也叫 PATA 口硬盘。

2.SCSI 接口

SCSI 的英文全称为 “Small Computer System Interface”（小型计算机系统接口），是同 IDE 完全不同的接口，IDE 接口是普通 PC 的标准接口，而 SCSI 并不是专门为硬盘设计的接口，是一种广泛应用于小型机上的高速数据传输技术。SCSI 接口具有应用范围广、多任务、带宽大、CPU 占用率低，以及支持热插拔等优点，但较高的价格使得它很难如 IDE 硬盘般普及，因此 SCSI 硬盘主要应用于中、高端和高档工作站中。SCSI 硬盘和普通 IDE 硬盘相比有很多优点：接口速度快，并且由于主要用于服务器，因此硬盘本身的性能也比较高，硬盘转速快，缓存容量大，CPU 占用率低，扩展性远优于 IDE 硬盘，并且支持热插拔。

3.SATA 接口

使用 SATA (Serial ATA) 口的硬盘又叫串口硬盘，是目前 PC 硬盘的主流。2001 年，由 Intel、APT、Dell、IBM、希捷、迈拓这几大厂商组成的 Serial ATA 委员会正式确立了 SerialATA 1.0 规范，2002 年，虽然串行 ATA 的相关设备还未正式上市，但 Serial ATA 委员会已抢先确立了 Serial ATA 2.0 规范。Serial ATA 采用串行连接方式，串行 ATA 总线使用嵌入式时钟信号，具备了更强的纠错能力，与以往相比其最大的区别在于能对传输指令（不仅仅是数据）进行检查，如果发现错误会自动矫正，这在很大程度上提高了数据传输的可靠性。串行接口还具有结构简单、支持热插拔的优点。

串口硬盘是一种完全不同于并行 ATA 的硬盘接口类型，由于采用串行方式传输数据而知名。相对于并行 ATA 来说，就具有非常多的优势。首先，Serial ATA 以连续串行的方式传送数据，一次只会传送 1 位数据。这样能减少 SATA 接口的针脚数目，使连接电缆数目变少，效率也会更高。实际上，Serial ATA 仅用四支针脚就能完成所有的工作，分别用于连接电缆、连接地线、发送数据和接收数据，同时这样的架构还能降低系统能耗和减小系统复杂性。其次，Serial ATA 的起点更高、发展潜力更大，Serial ATA 1.0 定义的数据传输率为 150MB/s，这比并行 ATA（即 ATA/133）所能达到 133MB/s 的最高数据传输率还高，而在 Serial ATA 2.0 的数据传输率达到 300MB/s，SATA Revision 3.0 可达到 750 MB/s 的最高数据传输率。

4.SAS 接口

SAS 是新一代的 SCSI 技术，和现在流行的 Serial ATA(SATA)硬盘相同，都是采用串行技术以获得更高的传输速度，并通过缩短连结线改善内部空间等。SAS 是并行 SCSI 接口之后开发出的全新接口。此接口的设计是为了改善存储系统的效能、可用性和扩充性，提供与串行 ATA (Serial ATA，缩写为 SATA)硬盘的兼容性。SAS 的接口技术可以向下兼容 SATA。SAS 系统的背板(Backpanel)既可以连接具有双端口、高性能的 SAS 驱动器，也可以连接高容量、低成本的 SATA 驱动器。因为 SAS 驱动器的端口与 SATA 驱动器的端口形

状看上去类似，所以 SAS 驱动器和 SATA 驱动器可以同时存在于一个存储系统之中。但需要注意的是，SATA 系统并不兼容 SAS，所以 SAS 驱动器不能连接到 SATA 背板上。由于 SAS 系统的兼容性，IT 人员能够运用不同接口的硬盘来满足各类应用在容量上或效能上的需求，因此在扩充存储系统时拥有更多的弹性，让存储设备发挥最大的投资效益。串行 SCSI 是点到点的结构，可以建立磁盘到控制器的直接连接。具有以下特点：

□ 更好的性能

点到点的技术减少了地址冲突以及菊花链连结的减速；

为每个设备提供了专用的信号通路来保证最大的带宽；

全双工方式下的数据操作保证最有效的数据吞吐量；

□ 简便的线缆连结

更细的电缆搭配更小的连接器；

□ 更好的扩展性

可以同时连结更多的磁盘设备。

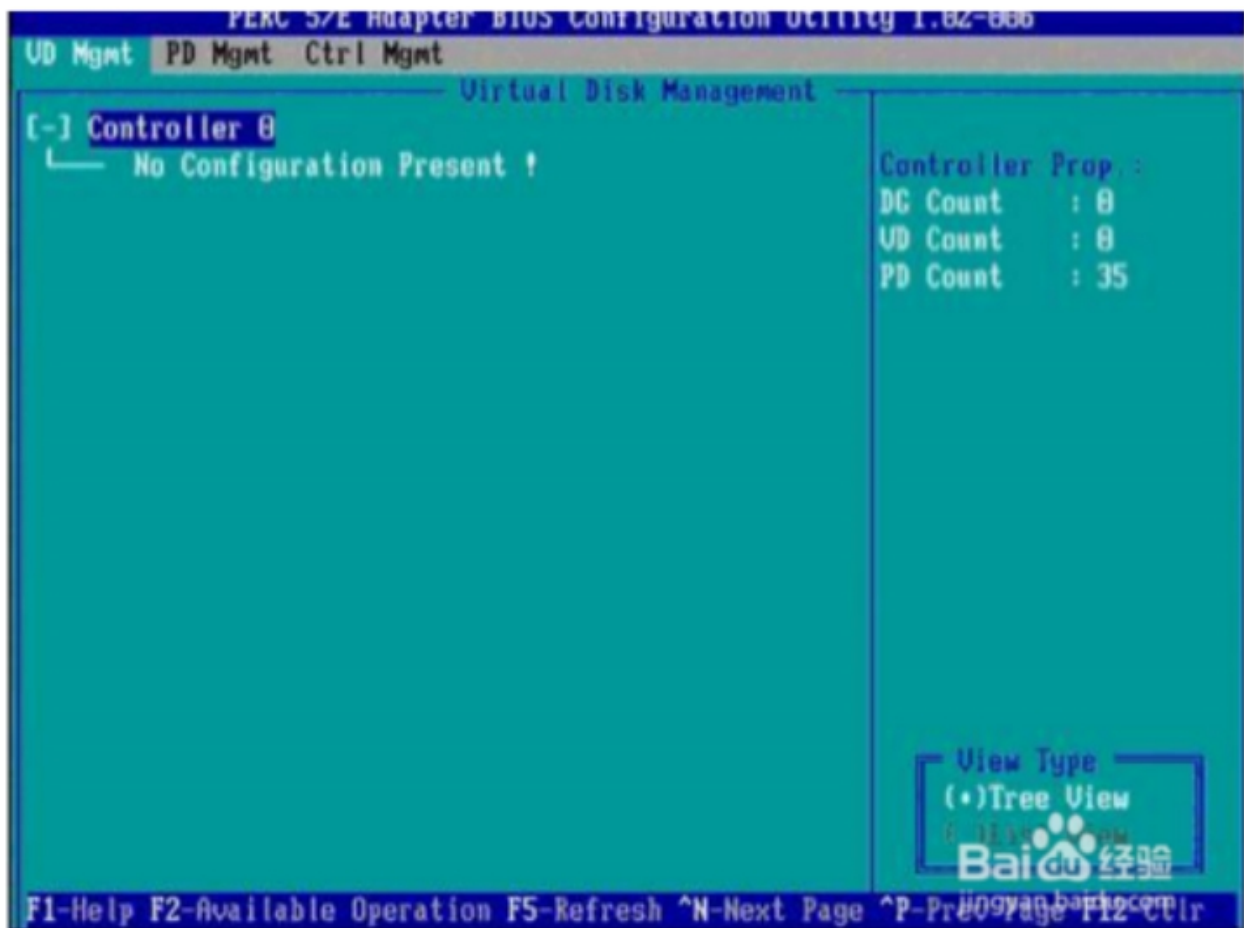
阵列卡的缓存

缓存 (Cache) 是 RAID 卡与外部总线交换数据的场所，RAID 卡先将数据传送到缓存，再由缓存和外边数据总线交换数据。它是 RAID 卡电路板上的一块存储芯片，与硬盘盘片相比，具有极快的存取速度，实际上就是相对低速的硬盘盘片与相对高速的外部设备(例如内存)之间的缓冲器。缓存的大小与速度是直接关系到 RAID 卡的实际传输速度的重要因素，大缓存能够大幅度地提高数据命中率从而提高 RAID 卡整体性能。多数 RAID 卡都配备了一定数量的内存作为高速缓存使用。不同的 RAID 卡出厂时配备的内存容量不同，一般为几兆到数百兆容量不等，这取决于磁盘阵列产品的应用范围。

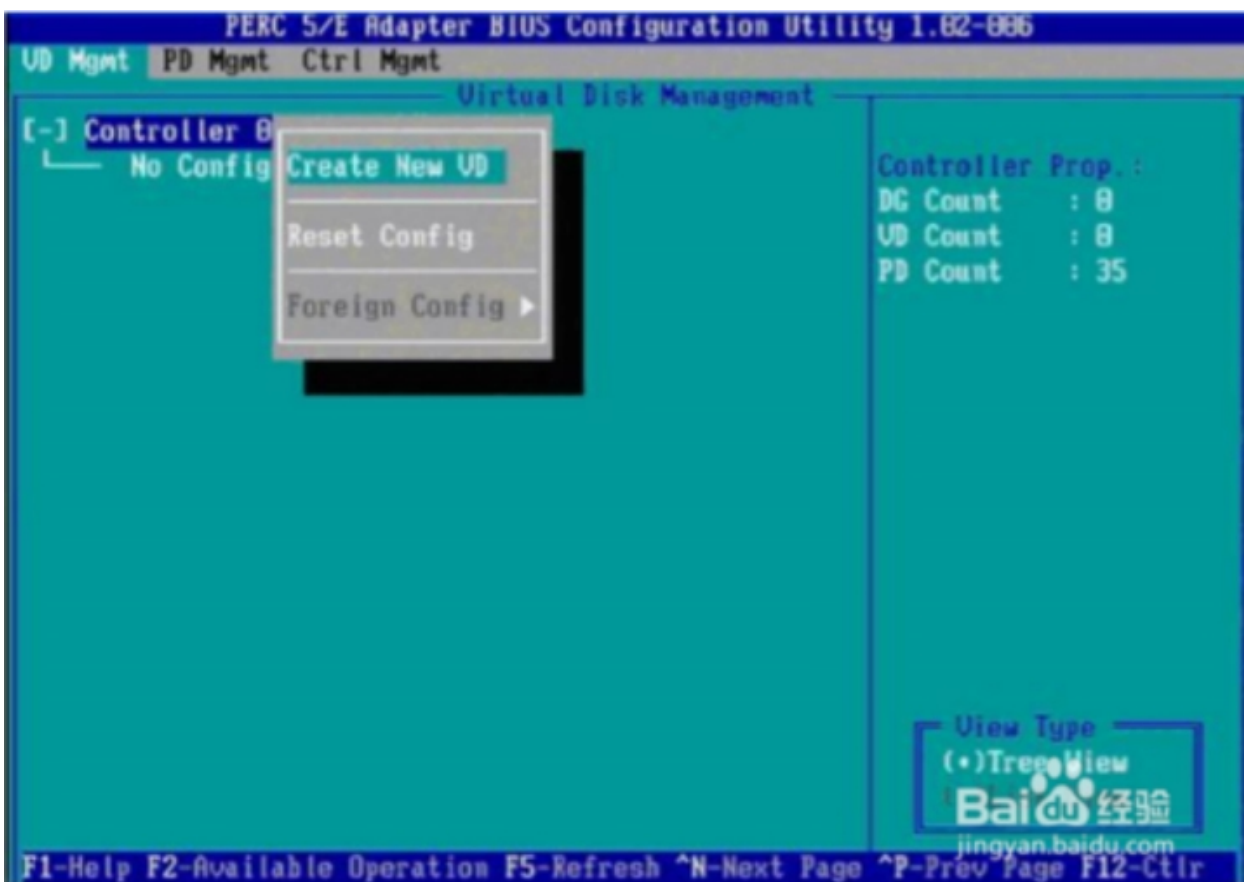
Write Through 和 Write Back 是阵列卡缓存的两种使用方式，也称为透写和回写。Write Through 也是 RAID 阵列卡的默认模式。当选用 write through 方式时，系统的写磁盘操作并不利用阵列卡的 Cache，而是直接与磁盘进行数据的交互。而 write Back 方式则利用阵列 Cache 作为系统与磁盘间的二传手，系统先将数据交给 Cache，然后再由 Cache 将数据传给磁盘。

制作硬件RAID案例

按照屏幕下方的虚拟磁盘管理器提示，按 F2 展开虚拟磁盘创建菜单



在虚拟磁盘创建窗口，按回车键选择“Create New VD” 创建新虚拟磁盘



在 RAID Level 选项按回车，可以出现能够支持的 RAID 级别，RAID 卡能够支持的级别有 RAID0/1/5/10/50，根据具体配置的硬盘数量不同，这个位置可能出现的选项也会有所

区别。选择不同的级别，选项会有所差别。选择好需要配置的 RAID 级别（我们这里以 RAID5 为例），按回车确认

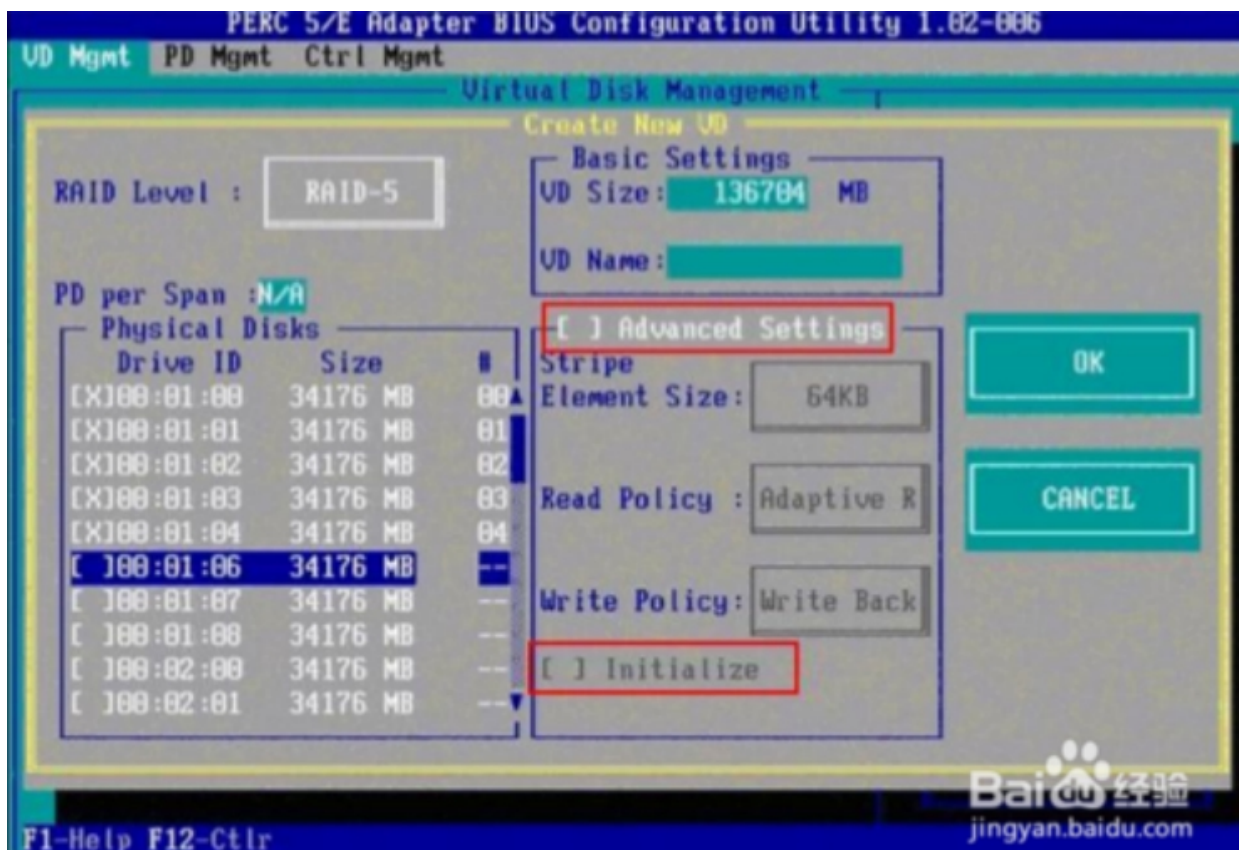


确认 RAID 级别以后，按向下方向键，将光标移至 Physical Disks 列表中，上下移动至需要选择的硬盘位置，按空格键来选择（移除）列表中的硬盘，当选择的硬盘数量达到这个 RAID 级别所需的要求时，Basic Settings 的 VD Size 中可以显示这个 RAID 的默认容量信息。有 X 标志为选中的硬盘。

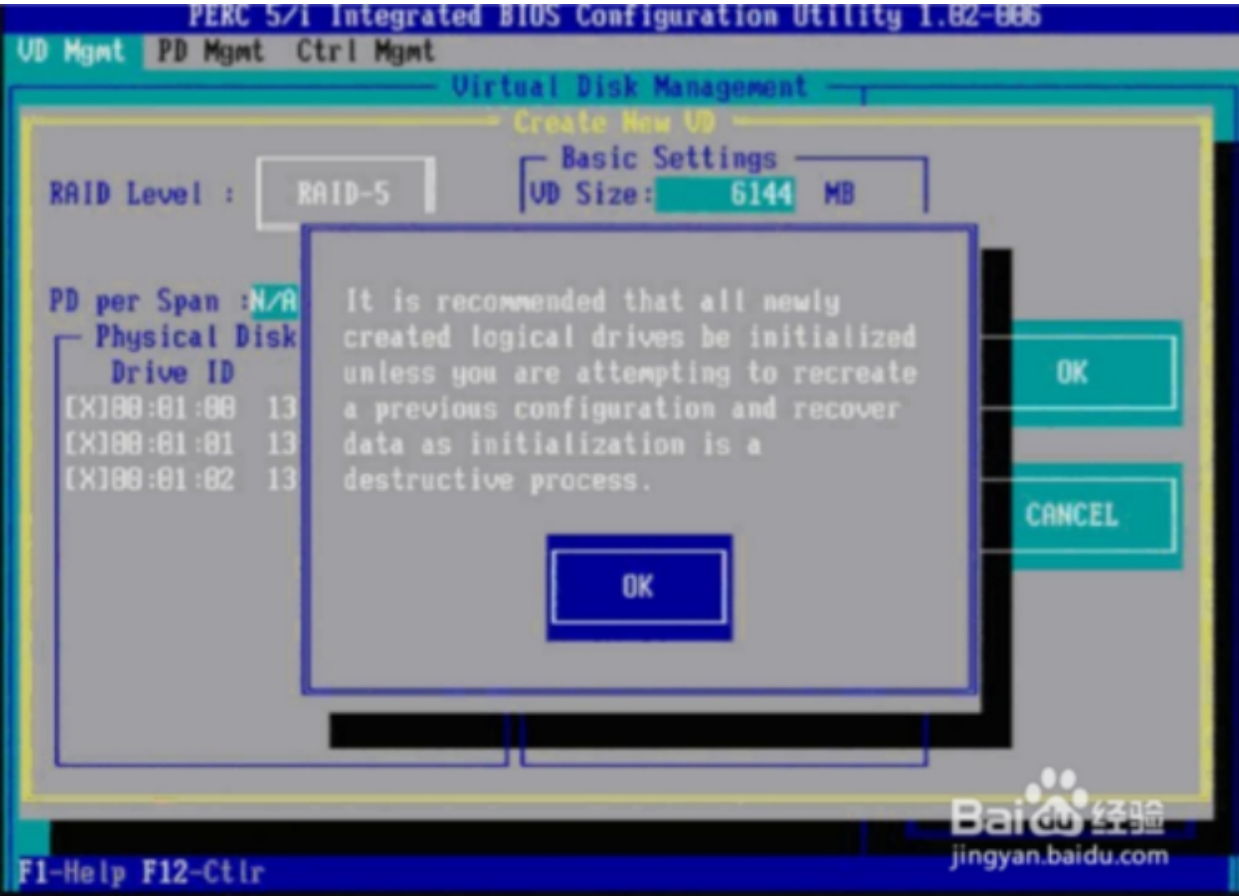
选择完硬盘后按 Tab 键，可以将光标移至 VD Size 栏，VD Size 可以手动设定大小，也就是说可以不用将所有的容量配置在一个虚拟磁盘中。如果这个虚拟磁盘没有使用我们所配置的 RAID5 阵列所有的容量，剩余的空间可以配置为另外的一个虚拟磁盘，但是配置下一个虚拟磁盘时必须返回 VD Mgmt 创建 VD Name 根据需要设置，也可以为空



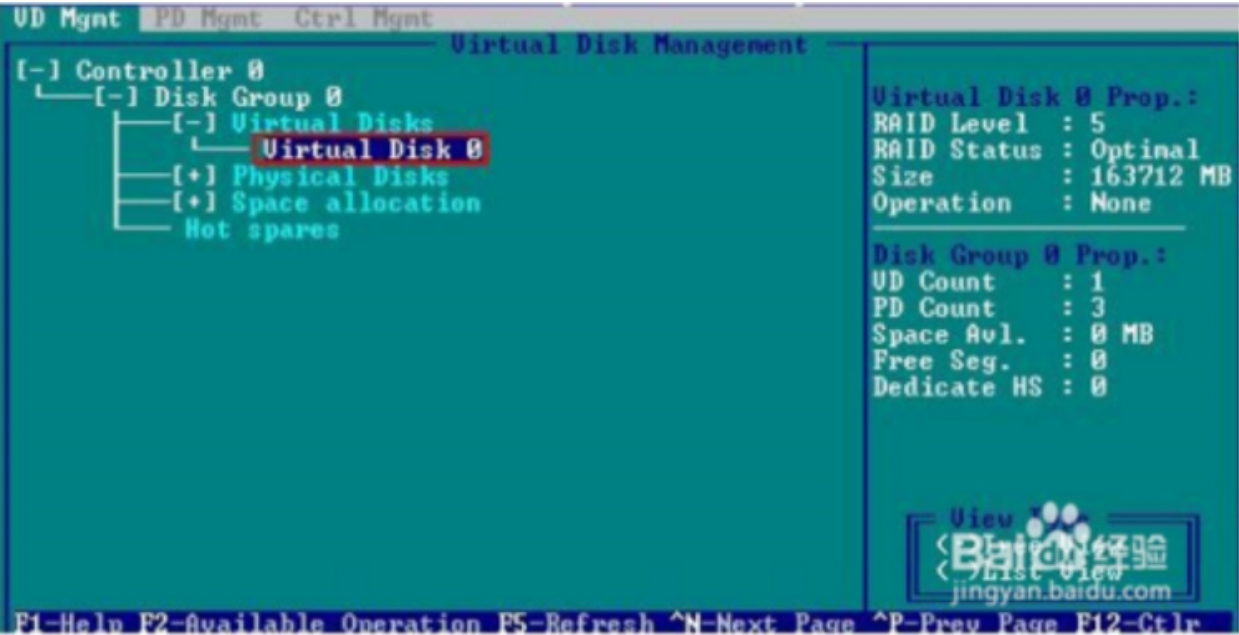
修改高级设置，选择完 VD Size 后，可以按向下方向键，或者 Tab 键，将光标移至 Advanced Settings 处，按空格键开启（禁用）高级设置。如果开启后（红框处有 X 标志为开启），可以修改 Stripe Element Size 大小，以及阵列的 Read Policy 与 Write Policy，Initialize 处可以选择是否在阵列配置的同时进行初始



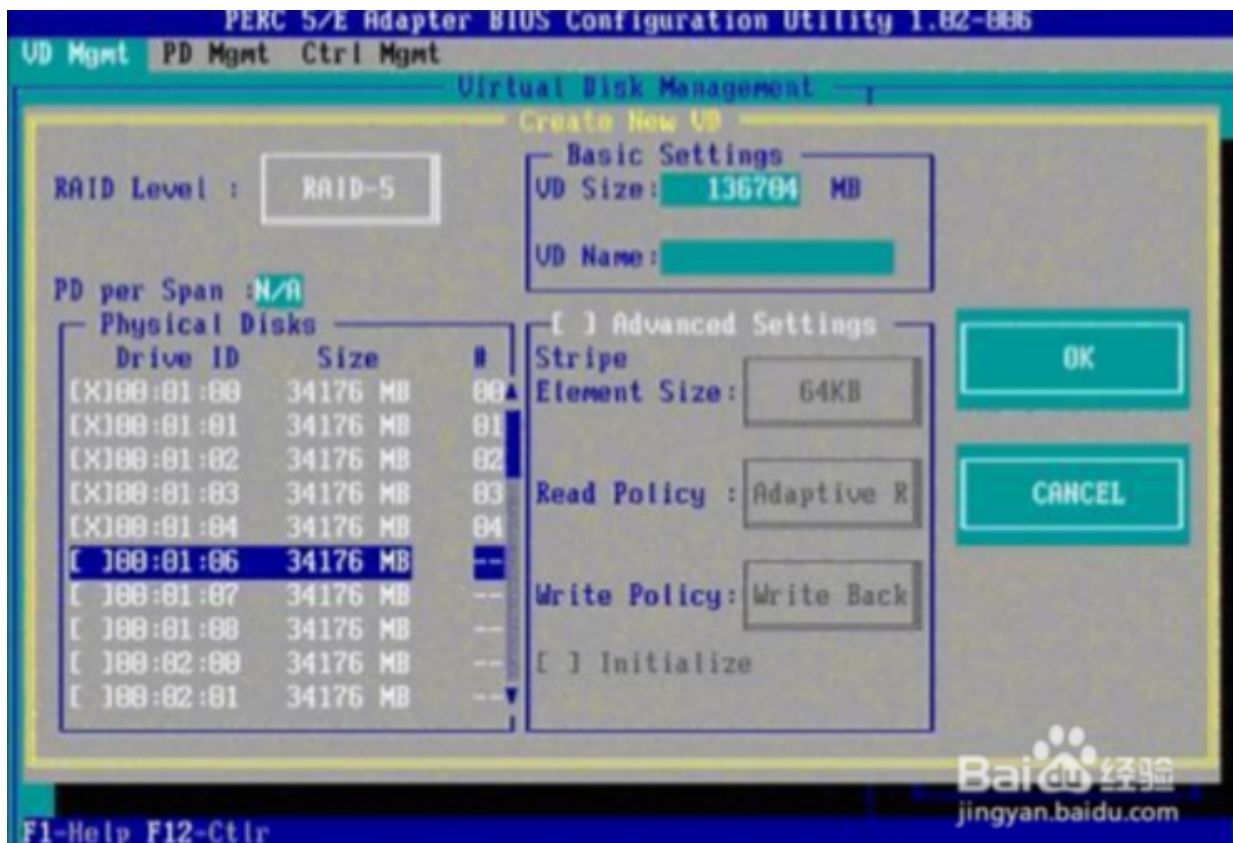
配置确认完成后，按 Tab 键，将光标移至 OK 处，按回车，会出现如下的提示，如果是一个全新的阵列，建议进行初始化操作，如果配置阵列的目的是为了恢复之前的数据，则不要进行初始化。按回车确认即可继续



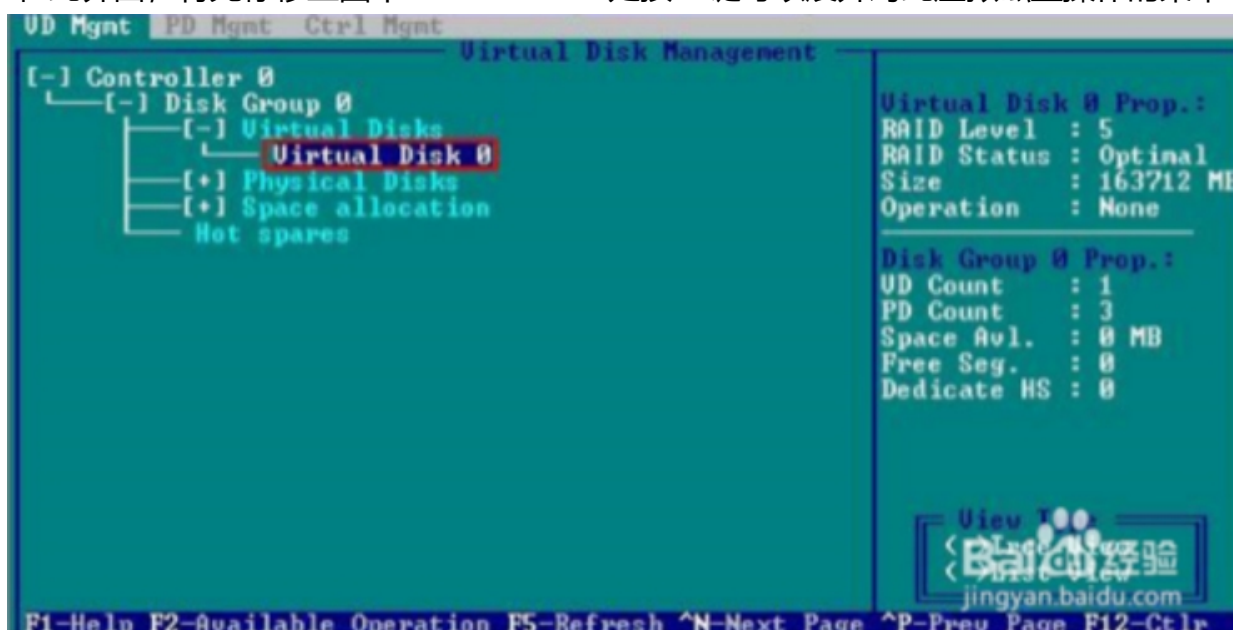
配置完成后，会返回至 VD Mgmt 主界面，将光标移至图中 Virtual Disk 0 处，按回车



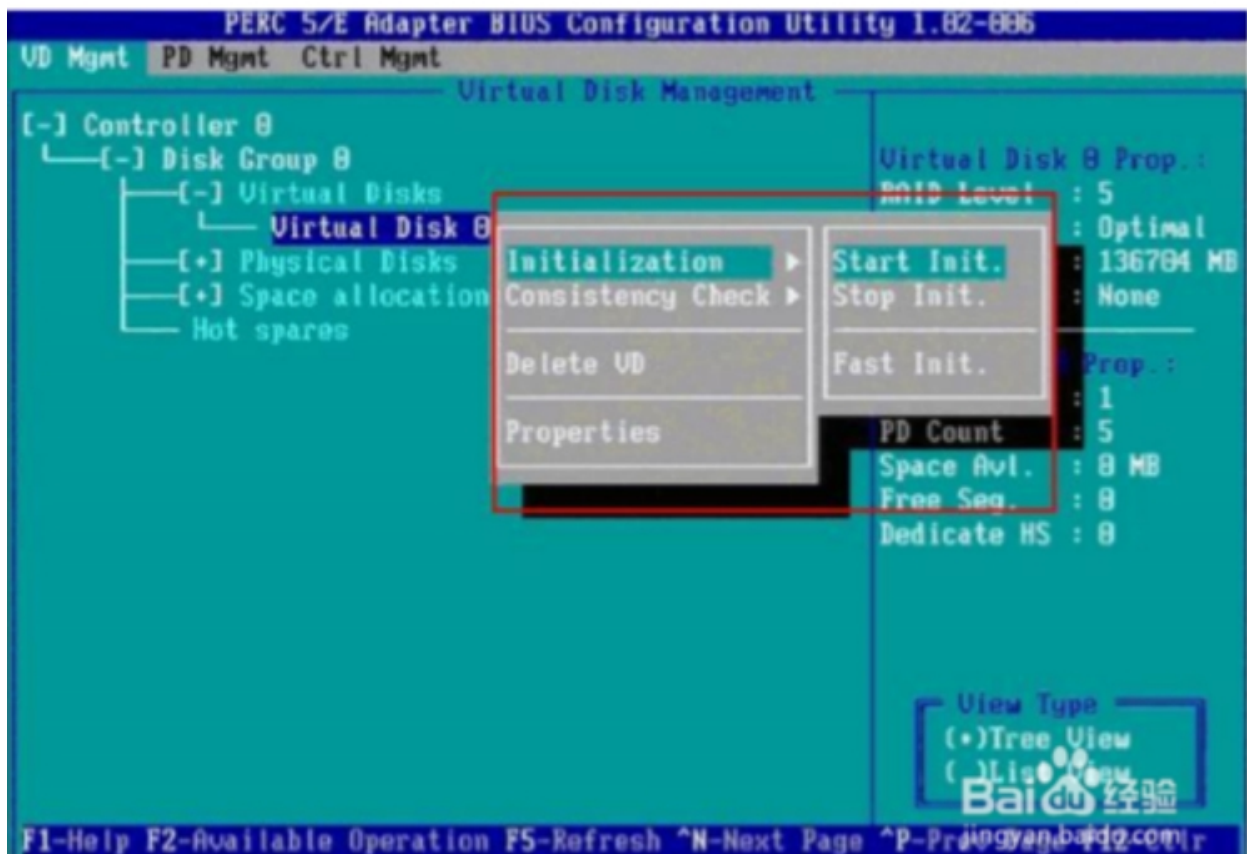
可以看到刚才配置成功的虚拟磁盘信息，查看完成后按 esc 键可以返回主界面



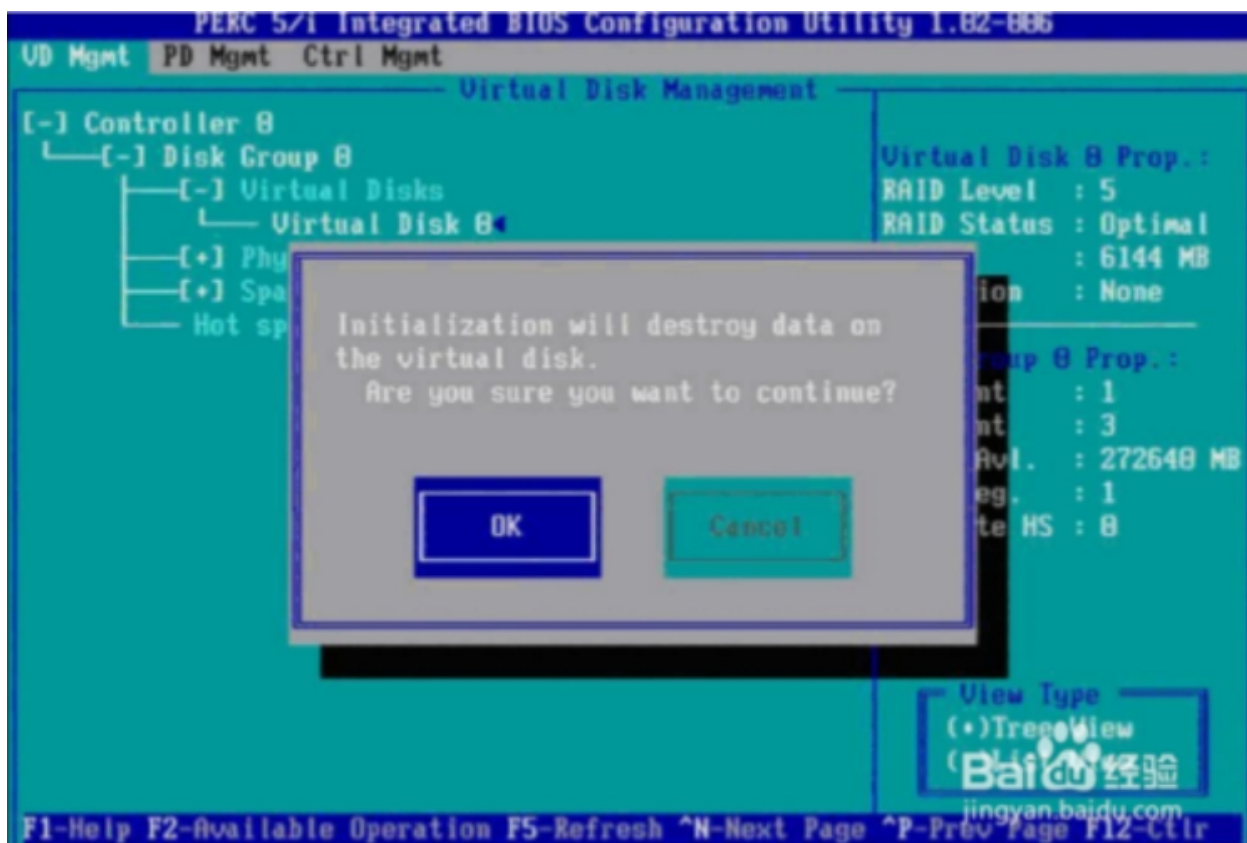
在此界面，将光标移至图中Virtual Disk 0处按F2键可以展开对此虚拟磁盘操作的菜单



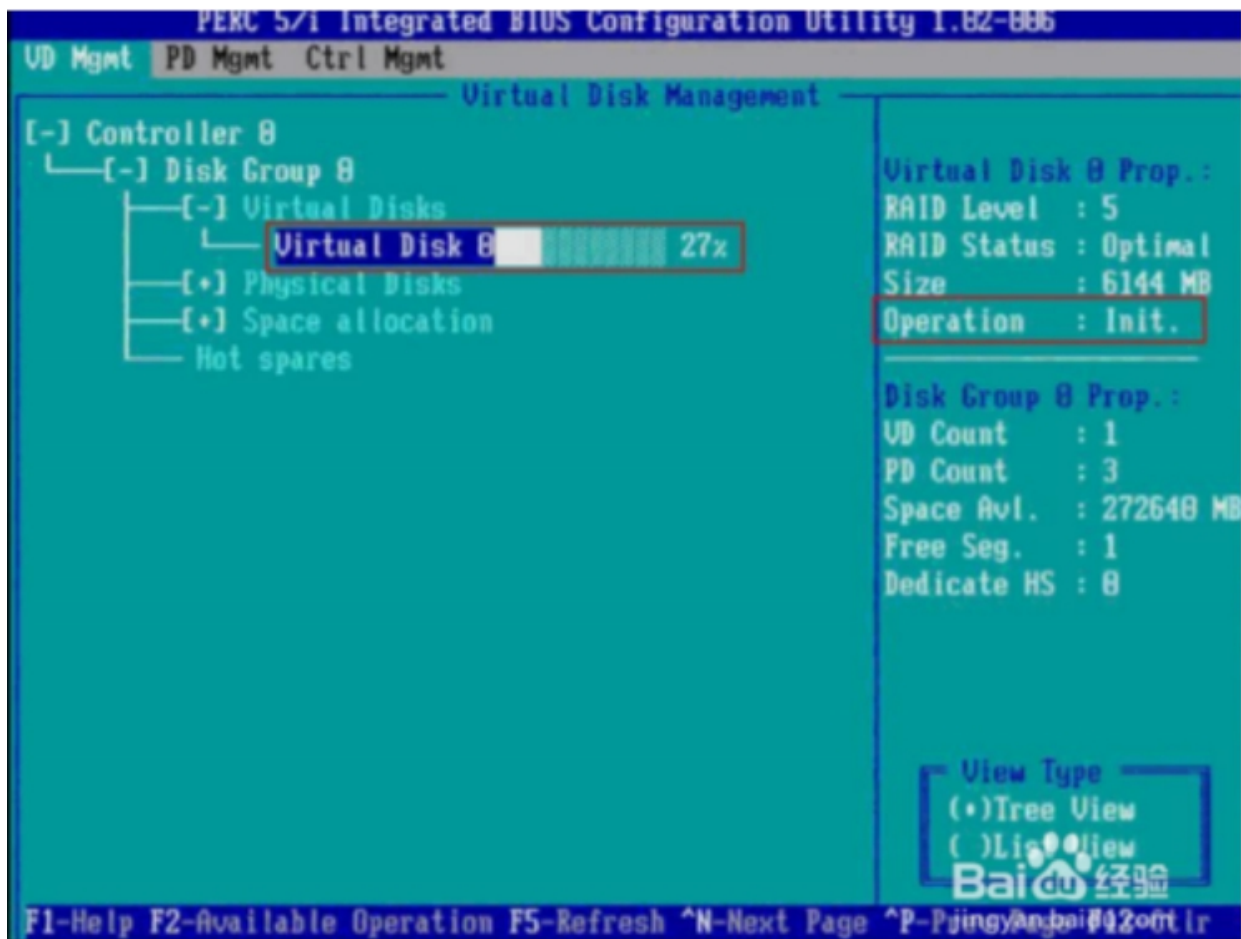
如下图红框所示，可以对刚才配置成功的虚拟磁盘（Virtual Disk 0）进行初始化（Initialization），一致性校验（Consistency Check），删除，查看属性等操作



如果我们要对此虚拟磁盘进行初始化，可以将光标移至 Initialization 处，回车后选择 Start Init。此时会弹出提示窗口，初始化将会清除所有数据，如果确认要进行初始化操作，在 OK 处按回车即可继续。



确认后可以看到初始化的进度，左边红框处为百分比表示，右边红框处表示目前所作的操作。等待初始化进行为 100%，虚拟磁盘的配置完成



软件raid:

MD:multi disks 多磁盘, 内核中的模块, 可以读取配置文件, 来实现
软件RAID, MD要模拟一个假的RAID设备 (逻辑
raid) /dev/md# (metadevice)
mdadm

创建模式 -C

专业选项:

- l: 指定级别
- n:设备个数
- a:{yes|no} 自动为其创建设备文件
- c:chunk大小, 默认为64k, (数据块)
- x:指定空闲盘的个数

创建raid0: mdadm -C /dev/md0 -a yes -l 0 -n 2 /dev/sdb{1,2}

创建raid1: mdadm -C /dev/md1 -a yes -n 2 -l 1 /dev/sdb{5,6}

创建raid5: mdadm -C /dev/md2 -a yes -l 5 -n 3 /dev/sdb{5,6,7}

增加热备磁盘: mdadm /dev/md2 -a /dev/sdb8

查看指定信息： mdadm -D /dev/md# 查看指定RAID设备的详细信息

查看raid状态： cat /proc/mdstat

管理模式： 选项： -a(--add), -d(--del), -r(--remove), -f(--fail)

模拟损坏： mdadm /dev/md1 -f /dev/sdb5

移除损坏的磁盘： mdadm /dev/md1 -r /dev/sdb5

添加新的硬盘到已有阵列： mdadm /dev/md1 -a /dev/sdb7

注意：

1、新增加的硬盘需要与原硬盘大小一致

2、如果原有阵列缺少工作磁盘（如raid1只有一块在工作，raid5只有2块在工作），

这时新增加的磁盘直接变为工作磁盘，如果原有阵列工作正常，则新增加的磁盘为热备磁盘。

停止阵列： mdadm -S /dev/md1

监控模式 -F 不常用

增长模式 -G 用于增加磁盘，为阵列扩容

将raid5的热备磁盘增加到阵列工作磁盘中： mdadm -G /dev/md2 -n 4

装配模式 -A 软RAID是基于系统的，当原系统损坏了，需要重新装配RAID

将停止的阵列重新装配； mdadm -A /dev/md1 /dev/sdb5 /dev/sdb6

自动装配： mdadm运行时会自动检查/etc/mdadm.conf 文件并尝试自动装配，

因此第一次配置raid后可以将信息导入到/etc/mdadm.conf 中，

命令如下：

echo ' DEVICE /dev/hd[a-z] /dev/sd*[a-z]' > /etc/mdadm.conf

mdadm --examine --scan --config=mdadm.conf >>

/etc/mdadm.conf

实验： fdisk /dev/sd# 进行分区 修改类型为fd

raid0实验： 2G 2:1G

mdadm -C /dev/md0 -a yes -l 0 -n 2 /dev/sda{5,6} 创建

cat /proc/mdstat 查看内核中启动的raid信息

mkfs -t ext4 /dev/md0 格式化

mount 挂载使用

raid1 2G 2:2G


```
mdadm -C /dev/md1 -a yes -n 2 -l 0 /dev/sda# /dev/sda#  
cat /proc/mdstat  
mkfs -t ext4 /dev/md1  
mount
```

raid5实验: 1.mdadm -C /dev/md0 -a yes -n4 -l5 /dev/{sda1, 2, 3, 4}

-C (create)v -a yes 创建过程中如果有设备没有, 会帮着创建

-n4(用了几块硬盘) -l (级别)

2.cat /proc/mdstat 查看raid详细信息(启动中的raid)

3.mkfs -t ext4 /dev/md0

4.mount /dev/md0 /tmp

5.模拟损坏和修复

1) mdadm -vDs (mdadm --detail --scan查看, 数据, 扫描) /dev/md0

2)配置文件的设置/etc/mdadm.conf (mdadm -vDs +DEVICE /dev/sdb1 /dev/sdc1...)

停止: mdadm -S /dev/md0 启动: mdadm -A /dev/md0

3) 模拟磁盘损坏: mdadm /dev/md0 -f /dev/sda1(指定第一块盘坏掉)

4) 移除磁盘: cat /proc/mdstat >> mdadm /dev/md0 -r /dev/sda1

5) 添加磁盘: mdadm /dev/md0 -a /dev/sda2 cat /proc/mdstat

lsmod 列出内核模块

mdadm --detail /dev/md#

watch 'cat /proc/mdstat' -n 指定刷新时间

DM:device mapper 逻辑设备

linear 线性的 mirror snapshot (快照)访问的一种路径

多路径 (使用负载均衡或者是备用) multipath

实验步骤:

1、添加硬盘

2、fdisk 分区 分区类型id要改成fd(linux raid)

3、mdadm 命令创建raid阵列

4、手动生成/etc/mdadm.conf文件

5、格式化raid设备文件

6、挂载raid设备文件

7、编写/etc/fstab文件，实现开机自动挂载设备