

## “计算机组织结构”作业 08 参考答案

1. 考虑一个有 10 个磁盘，每个磁盘 100GB 的 RAID 阵列。对于 RAID0、1、2、3、4、5、6 中的各级，有效数据的存储容量为多大？

RAID0: 1000G

RAID1: 500G

RAID2: 600G

RAID3: 900G

RAID4: 900G

RAID5: 900G

RAID6: 800G

2. 假定有两个用来存储 10TB 数据的 RAID 系统，每个磁盘的大小均为 2TB。系统 A 使用 RAID 1 技术，系统 B 使用 RAID 5 技术。请问：[刘璟，121250083]

- a) 系统 A 需要比系统 B 多用多少存储容量（单位：TB）？

A 系统需 20TB 存储容量；B 系统采用 6 个磁盘，需 12TB 存储容量，多用 8TB

- b) 假定一个应用需要向磁盘写入一块数据，若磁盘读或写一块数据的时间为 30ms，则最坏情况下，在系统 A 上写入一块数据需要多少时间（单位：毫秒）？

30ms

- c) 如果问题 b) 是在系统 B 上写入一块数据，需要多少时间（单位：毫秒）？

120ms（最坏情况：两读两写）

- d) 哪个系统更加可靠？

A 更加可靠。

从直观角度来说，RAID1 通过镜像对整个数据进行了备份，只要其中一份数据所在磁盘出现了问题，通过另一份数据就可以很方便的进行恢复；而 B 系统中，如果同一对应位置的任意两块磁盘同时存在故障，相应的数据就无法进行恢复了。

具体而言，假设每一位损坏的概率为  $p$ ，则

- RAID1 的可靠性为：  $R1 = 1 - p^2$

- RAID5 的可靠性为：  $R5 = (1-p)^n C_n^0 + p(1-p)^n C_n^1 = [(n-1)p + 1](1-p)^{n-1}$ ， $n$  为磁盘数

当  $n=2$  时两者相等， $n>2$  时，对  $n$  求偏导：

-  $\partial R1 / \partial n = 0$

-  $\partial R5 / \partial n = (1-p)^{n-1} \{p + [(n-1)p + 1] \ln(1-p)\}$

∵  $1-p > 0$ ，∴ 再对  $p + [(n-1)p + 1] \ln(1-p)$  对  $n$  求导，得到二阶导  $p \ln(1-p) < 0$

∴ 曲线以  $n$  为自变量的  $R5$  曲线单调递减

此时有  $R1 - R5 = 1 - p^2 - (2p+1)(1-p)^2 = (1-p)[(1+p) - (2p+1)(1-p)] = (1-p)(2p^2) > 0$

因此 RAID1 的可靠性比 RAID5 更高，且 RAID5 的磁盘数越多，可靠性越低。

===== 分割线：以下内容不在小程序上提交 =====

3. 假定在一个使用 RAID5 的系统中，采用先更新数据块，再更新校验块的信息更新方式。如果在更新数据块和更新校验块的操作之间发生了掉电现象，那么会出现什么问题？采用什么样的信息更新方式可以避免这个问题？

对于 RAID5 来说，如果在写完数据块但未写入校验块时发生断电，则写入的数据和对应的校验信息不匹配，无法正确恢复数据。这种情况可以避免，因为 RAID5 是大数据块交叉方式，每个盘独立进行操作，所以只要同时写数据块所在盘和校验块所在盘即可。