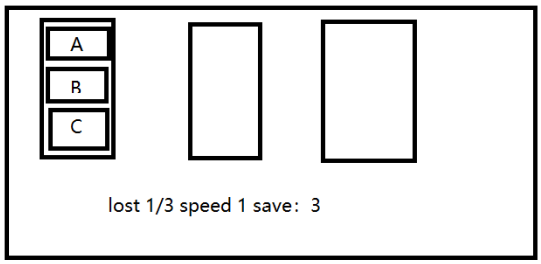
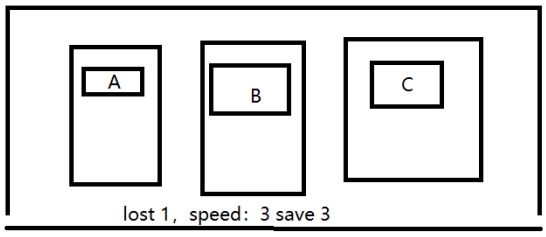
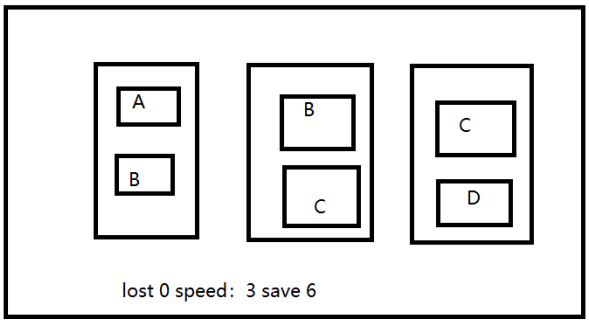
存储策略







可以看出，没有一种策略是完美的，每一种都有弊端和优势，各个参数之间相互联系，相互作用，因此，我们没有最优方案，只能安装我们的侧重点选择，例如数据非常重要但是不需要太快，我们就要选择多备份，数据不重要但是追求速度，我们就要多分割。

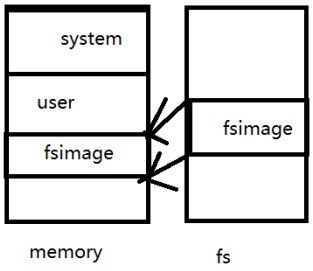
如何检查数据错误

使用校验码，例如
10101010前面有4个1，那么有偶数个1，我们在后面添加0，变为101010100
如果数据变成 101010110那么我们发现除了校验码有奇数个1而校验位为0（应该有偶数个11），所以校验错误。
这仅仅是简单的一种校验方式。
我们的数据块采用CRC循环冗余校验码校验码来进行检验。

NameNode存储什么

FsImage

存储文件的元数据如文件名，文件目录结构，文件属性（生成时间,副本数,文件权限），以及每个文件的块列表和块所在DataNode等。
这个文件的RW方式是 `FMAP->mmap`



使用这种方式，主要是因为fsimage文件变化的特点：

1. 保证速度
2. 因为文件系统的客观条件，和日志不一样，不能顺序追加，因为修改居多，同时即便是追加，也不是在结尾追加，所以不能用文件的顺序读写来加速。
3. 为了让第二NN来复制fsimage文件来做备份。

opLog

操作的日志，通过顺序的读写的方式追加到磁盘中，采用顺序读写的方式来进行磁盘加速。

为什么块大小为128MB

寻道时间为RW时间的0.01
寻道时间 10ms ， rw时间1s，磁盘速度 100MBps，那么块大小100MB取整数128MB