在Linux中添加硬盘时，需要考虑的是容量

1当磁盘的容量小于2T时，可以直接使用

对硬盘进行分区操作，前提是您把fdisk -l 弄明白了；通过fdisk -l ，我们能找出机器中所有硬盘个数及设备名称；比如上面的例子，我们会看到两个设备一个是/dev/hda ，另一个是/dev/sda ；

简单介绍一下Linux的磁盘的分区问题：

通常会使用接下来的命令提示进行分区操作，

磁盘容量与主分区、扩展分区、逻辑分区的关系：

硬盘的容量＝主分区的容量＋扩展分区的容量

扩展分区的容量＝各个逻辑分区的容量之和

一块物理硬盘只能有: 一到四个主分区(但其中只能有一个是活动的主分区),或一到三个主分区,和一个扩展分区。分别对应hda1,hda2,hda3,hda4。

基本上使用fdisk命令即可分区。

2.当容量大于2T时，上面个的操作就变得力不从心了

需要使用GPT分区：

　　使用 parted 建立大小超过2T的分区  
  
　　1，parted /dev/sdb  
　　可以输入p打印磁盘信息，查看分区的情况，找到起始和结束位置。  
  
　　2，mklabel gpt  
　　设置分区类型为gpt  
  
　　3，mkpart primary 0% 100%  
　　primary指分区类型为主分区，0是分区开始位置，100%是分区结束位置。相同的命令为：mkpart primary 0-1 或者是：mkpart  primary 0  XXXXXX结束的空间  
  
　　4，print  
　　打印当前分区,查看分区设置是否正确  
　　5，quit  
　　完成后用quit命令退出。  
  
 　　6，mkfs.ext3  /dev/sdb1  
  
　　格式化磁盘。如果支持EXT4可选择格式化为EXT4，格式化成ext3的时候又出现问题。Ext3默认的block size的大小为4k,最大只能支持8T的空间。格式化时指定block size的大小为8K这样最大可以支持16T的空间。相应命令：mkfs.ext3 -b 8192 /dev/sdb2  
  
 　　7，mount  /dev/sdb1  /home/  
  
挂载到/home目录下  
  
 　　8，自动挂载分区  
  
当在系统里创建了一个新的分区后，因为mount挂载在重启服务后会失效，所以需要将分区信息写到/etc/fstab文件中让其永久挂载  
，编辑/etc/fstab里加入：输入红色内容  
/dev/sdb1(磁盘分区)                /home(mount目录)                   ext3(文件格式)                 defaults                    0             0  
保存退出，重启后/dev/sdb1 就会自动挂载到/mnt/abc目录下

以上就完成了基本的容量扩充的功能。

下面引出在HDFS中的扩充的问题，一般的在hadoop的系统搭建完成时，我们都需要先进行一个HDFS的namenode 的格式化

这主要是在物理上还未存在，或者用网络磁盘来描述更加合适；还有格式化，并不是传统意义上的磁盘清理，而是一些清除与准备工作。

namemode是hdfs系统中的管理者，它负责管理文件系统的命名空间，维护文件系统的文件树以及所有的文件和目录的元数据，元数据

同时为了保证操作的可靠性，还引入了操作日志，所以，namenode会持久化这些数据到本地。对于第一次使用HDFS时，需要执行-format命令才能正常使用namenode节点。

声明一下

1、hadoop namenode -format命令只对namenode有效，对其它节点无效。  
2、集群增加节点，只要在namenode节点的配置文件中，将[IP](https://www.baidu.com/s?wd=IP&tn=44039180_cpr&fenlei=mv6quAkxTZn0IZRqIHckPjm4nH00T1YdujR4uWbzrHwhnWRdP16d0ZwV5Hcvrjm3rH6sPfKWUMw85HfYnjn4nH6sgvPsT6KdThsqpZwYTjCEQLGCpyw9Uz4Bmy-bIi4WUvYETgN-TLwGUv3EPHm3nHTzPjD3" \t "_blank)和端口以及数据目录配置一下就可以了。可以用更新或是重启服务的命令，那么新增的datanode就会自动追加到集中。  
3、称这样的系统为高度自由可伸缩系统。增或减节点都能被智能控制节点监测到，并将相应的节点的变化，通过冗余机制给予解决。

主要是在hdfs-site.xml中进行配置

dfs.data.dir

这是DataNode结点被指定要存储数据的本地文件系统路径。DataNode结点上 的这个路径没有必要完全相同，因为每台机器的环境很可能是不一样的。但如果每台机器上的这 个路径都是统一配置的话，会使工作变得简单一些。默认的情况下，它的值hadoop.tmp.dir, 这个路径只能用于测试的目的，因为，它很可能会丢失掉一些数据。所以，这个值最好还是被覆盖。

<property>  
    <name>dfs.name.dir</name>  
    <value>/pvdata/hadoopdata/name/,/opt/hadoopdata/name/</value>  
</property>

**dfs.name.dir**

 这是NameNode结点存储hadoop文件系统信息的本地系统路径。这个值只对NameNode有效，DataNode并不需要使用到它。上面对于/temp类型的警告，同样也适用于这里。在实际应用中，它最好被覆盖掉。

<property>  
    <name>dfs.data.dir</name>  
    <value>/dev/sda3/hadoopdata/,/dev/sda1/hadoopdata/</value>  
</property>

可将扩展的容量添加至dfs.name.dir的value中。