

密级：机密(版权所有，翻版必究)



分册名称：

第 册/共 册

大数据高端人才培养计划

HDFS 设置配额



沈阳昊宸科技有限公司

2017 年 3 月 30 日

变更履历

修改编号	版本	修改内容	修改人	修改日期
			dulm	



目 录

未找到目录项。



Hadoop 分布式文件系统(HDFS)允许管理员为每个目录设置配额。新建立的目录没有配额。最大的配额是 **Long.Max_Value**。配额为 1 可以强制目录保持为空。

目录配额是对目录树上该目录下的名字数量做硬性限制。如果创建文件或目录时超过了配额，该操作会失败。重命名不会改变该目录的配额；如果重命名操作会导致违反配额限制，该操作将会失败。如果尝试设置一个配额而现有文件数量已经超出了这个新配额，则设置失败。

配额和 **fsimage** 保持一致。当启动时，如果 **fsimage** 违反了某个配额限制（也许 **fsimage** 被偷偷改变了），则启动失败并生成错误报告。设置或删除一个配额会创建相应的日志记录。

下面的新命令或新选项是用于支持配额的。前两个是管理员命令。

```
#把每个目录配额设为N。这个命令会在每个目录上尝试，如果N不是一个正的长整型数，目录不存在或是文件名，或者目录超过配额，则会产生错误报告。
dfsadmin -setquota <N> <directory>...<directory>
#为每个目录删除配额。这个命令会在每个目录上尝试，如果目录不存在或者是文件，则会产生错误报告。如果目录原来没有设置配额不会报错。
dfsadmin -clrquota <directory>...<director>
#使用-q选项，会报告每个目录设置的配额，以及剩余配额。如果目录没有设置配额，会报告none和inf。
fs -count -q <directory>...<directory>
```

在多人共用 **HDFS** 的环境下，配额设置非常重要。特别是在 **Hadoop** 处理大量资料的环境，如果没有配额管理，很容易把所有的空间用完造成别人无法存取。**Hdfs** 的配额设定是针对目标而不是针对账号，所有在管理上最好让每个账号仅操作某一个目录，然后对目录设置配置。

- Name Quotas: 设置某一个目录下文件总数
- Space Quotas: 设置某一个目录下可使用空间大小

Name Quotas

计算公式: $QUOTA - (DIR_COUNT + FILE_COUNT) = REMAINING_QUOTA$

这里的 10000 是指 $DIR_COUNT + FILE_COUNT = 10000$ ，最大值为 **Long.Max_Value**

启用设定: **hadoop dfsadmin -setQuota 10000 /user/seamon**

清除设定: **hadoop dfsadmin -clrQuota /user/seamon**

Space Quotas

计算公式: $SPACE_QUOTA - CONTENT_SIZE = REMAINING_SPACE_QUOTA$

可以使用 m, g, t 代表 MB, GB, TB

启用设定: **hadoop dfsadmin -setSpaceQuota 1g /user/seamon/**

清除设定: **hadoop dfsadmin -clrSpaceQuota /user/seamon**

- 这里需要特别注意的是“Space Quota”的设置所看的不是 **Hdfs** 的文件大小，而是写入 **Hdfs** 所有 block 块的大小，假设一个文件被切分为 2 个 blocks，在 **core-site.xml** 里面设置 **dfs.block.size=64MB**，**dfs.replication=3**，那么该文件所需要的存储空间为: $2 * 64M * 3 = 384MB$

- 如果一个小文件（例如，1k 大小的文件）被上传到 **hdfs**，该文件并不能占满一整个 **blok**，但是按照 **hdfs** 配置规则也需要按照一个 **blok** 计算，即存储空间为： $1 \times 64\text{MB} \times 3 = 192\text{MB}$

其他注意事项：

hdfs 的配额管理是跟着目录走，如果目录被重命名，配额依然有效。

麻烦的是，在设置完配额以后，如果超过限制，虽然文件不会写入到 **hdfs**，但是文件名依然存在，只是文件 **size** 为 0。当加大配额设置后，还需要将之前的空文件删除才能进一步写入。

如果新设置的 **quota** 值，小于该目录现有的 **Name Quotas** 及 **Space Quotas**，系统并不会给出错误提示，但是该目录的配置会变成最新设置的 **quota**。

