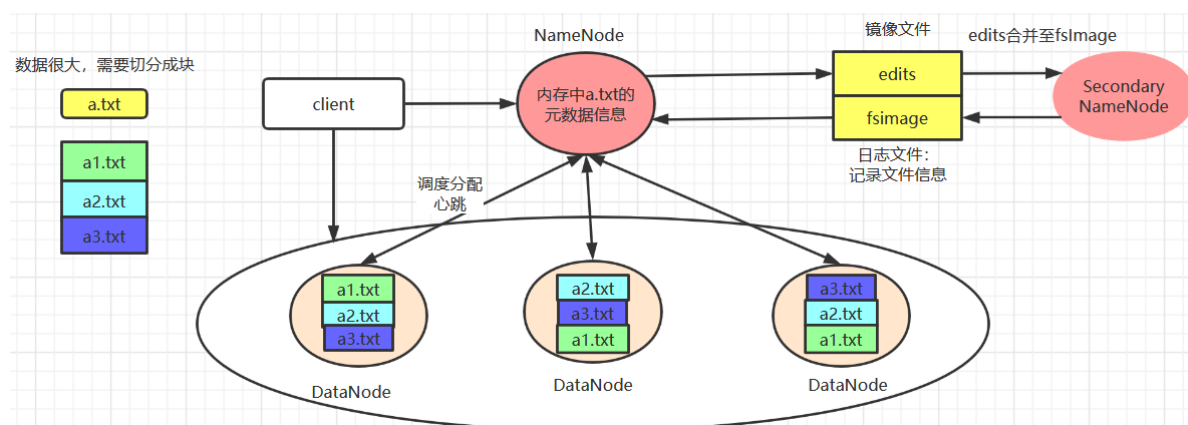


HDFS

具体组件介绍



Client客户端

- 文件切分——Blocks
- 与NameNode交互，获取文件位置信息
- 与DataNode交互，读入或写入数据
- 提供一些开关HDFS命令

NameNode

- 在内存中保存整个文件系统的**名称空间**和文件数据块的地址映射
- 整个HDFS可存储的文件数受限于**NameNode的内存大小**
- 元数据管理

1.元数据信息：文件名、文件目录结构、文件属性、每个文件块列表以及列表中块与块所在的DataNode之间的映射关系，数据会定期持久化到本地磁盘的fsImage文件和edits文件中

2.文件操作：DataNode负责文件内容的读写请求，数据流不经过NameNode。

3.副本机制：文件存在哪些DataNode上是由NameNode上决定的。

4.心跳机制：DataNode周期性向NameNode发送心跳信号和块状态报告

Seondary NameNode

- 辅助NameNode，分担部分工作
- 紧急情况下,辅助恢复NameNode
- 并非NameNode的热备份

FsImage和Edits

日志备份，持久化存储

- Edits：存储元数据关联命令的执行目录
- FsImage：将Edits文件中记录的内容合并后再进行记录

core-site.xml

设置CheckPoint，用于Secondary NameNode自动获取Edits文件并于FsImage，然后再将FsImage传送给DataNode

```
<property>
  <name>fs.checkpoint.period</name>
  <value>60</value>
  <description>The number of seconds between two periodic
checkpoints.
  </description>
</property>

<property>
  <name> fs.checkpoint.size</name>
  <value>1048576</value>
  <description>The size of the current edit log (in bytes) that
triggers
  a period-ic checkpoint even if the fs.checkpoint.period hasn't
expired.</description>
</property>
```

HDFS副本机制和机架感知

配置文件：hdfs-site.xml

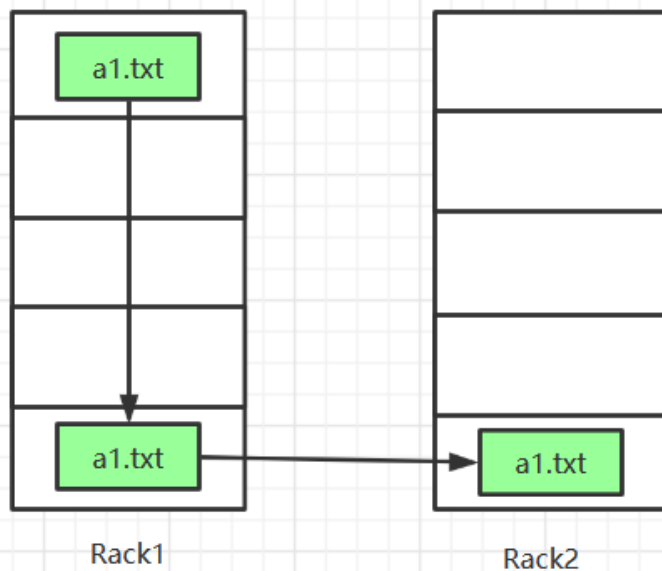
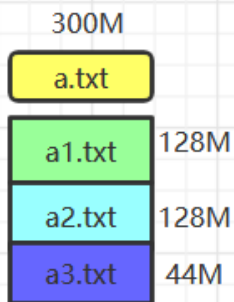
- 文件切片大小设置：默认128M
- 文件切片副本数：一个Block默认副本值为3

```
<property>
  <name>dfs.replication</name>
  <value>3</value>
</property>
```

机架感知Rack—数据备份

- 本机架备份，另外一个机架也备份

数据很大，需要切分成块



DataNode作用

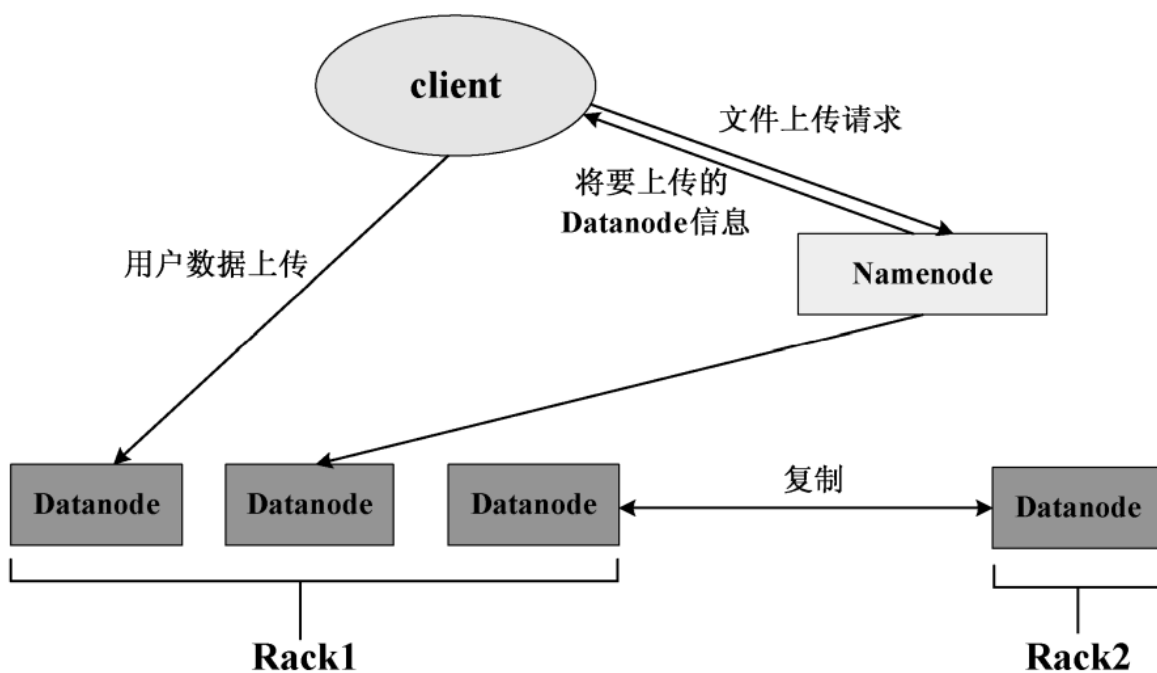
- 存储实际的数据块
- 执行数据端的读写操作

1.DataNode以数据块形式存储HDFS文件

2.DataNode响应HDFS客户端**读写请求**

3.DataNode周期性向NameNode汇报**心跳信息**、**数据块信息**、**缓存数据块信息**

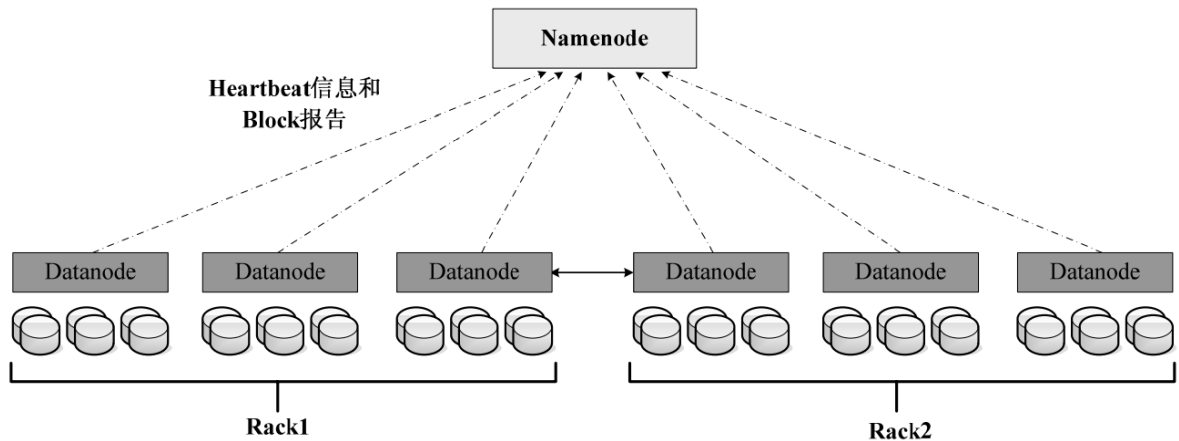
用户信息上传过程



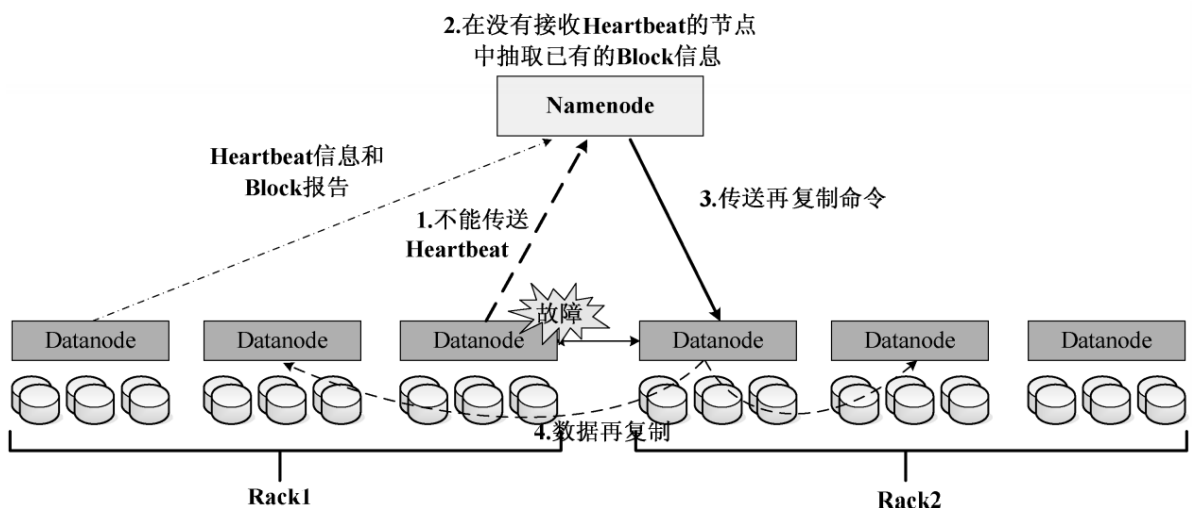
DataNode心跳汇报和再复制

DataNode需要每隔一定周期向NameNode汇报自己的状态，未收到心跳的DataNode将会被切断与NameNode的联系，判为无法使用。

NameNode会根据DataNode中原有的block信息找到该block被复制的DataNode的具体位置，以 block 类别新添复制的DataNode，并向相应的 DataNode传达再复制命令。收到再复制命令的 DataNode，会在其他 DataNode上按照缺少的复制数量进行 block 复制。

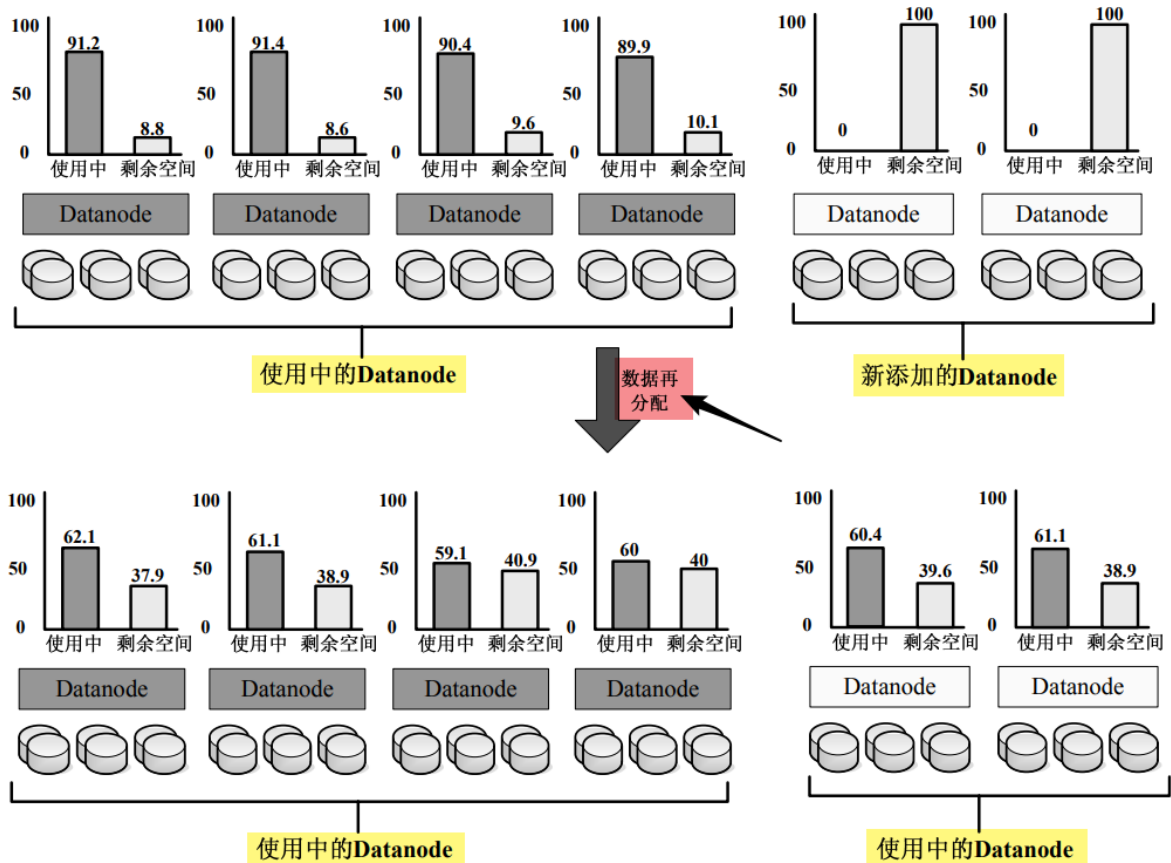


DataNode再分配—高扩展性



新添加的DataNode却不能立马被使用，因为数据保存在其它DataNode中，如果直接将新数据添加到新DataNode中，会使得DataNode的利用率不平衡。

数据再分配可以优化使用效率，达到平衡，但也使得大量数据从使用中的节点移动到新增节点的过程中所导致的网络负载变大的问题，一般再分配设置为根据管理员手动执行，不是自动执行。



DataNode的block管理

hdfs-site.xml—设置block大小

```
<property>
  <name>dfs.block.size</name>
  <value>134217728</value>
  <description>The 128MB block size.</description>
</property>
```

core-site.xml—修改数据存储位置

```
<property>
  <name>hadoop.tmp.dir</name>
  <value>/opt/hadoopData</value>
</property>
```

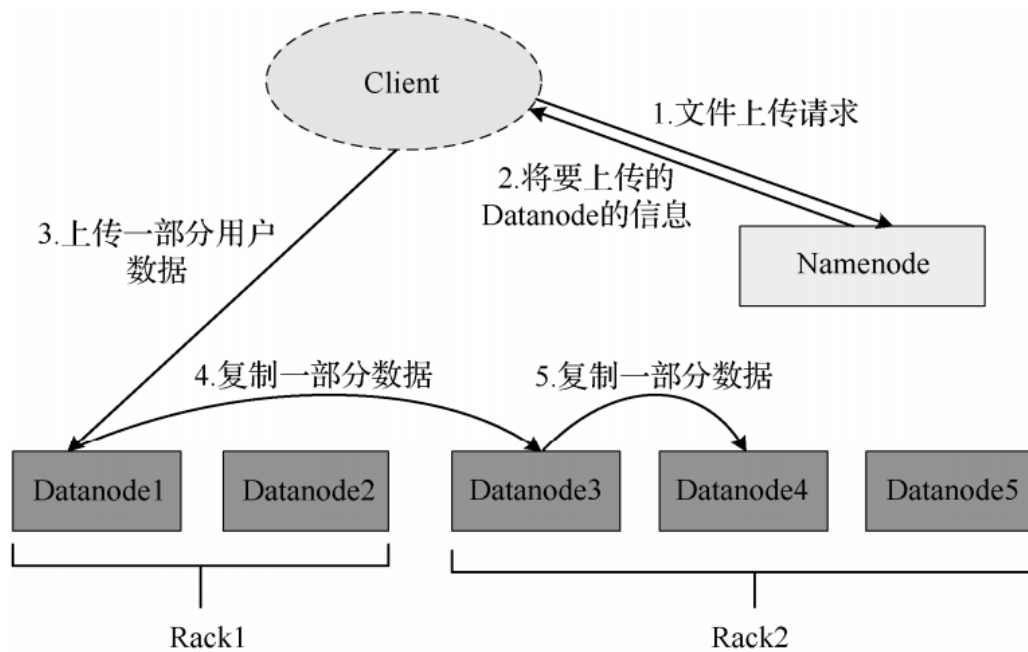
DataNode数据复制

防止服务器和磁盘故障带来数据丢失

复制的流水线过程

- 上传数据需要1,2步从NameNode获取要上传的DataNode的信息
- Hadoop上传数据的同时将数据复制到多个DataNode上

- client 和 Datanode 会将 3 步、4 步、5 步过程执行到完成所有的数据传输，传输的同时对数据进行复制



参考文献

《基于Hadoop的大数据分析和处理%20魏祖宽.pdf》