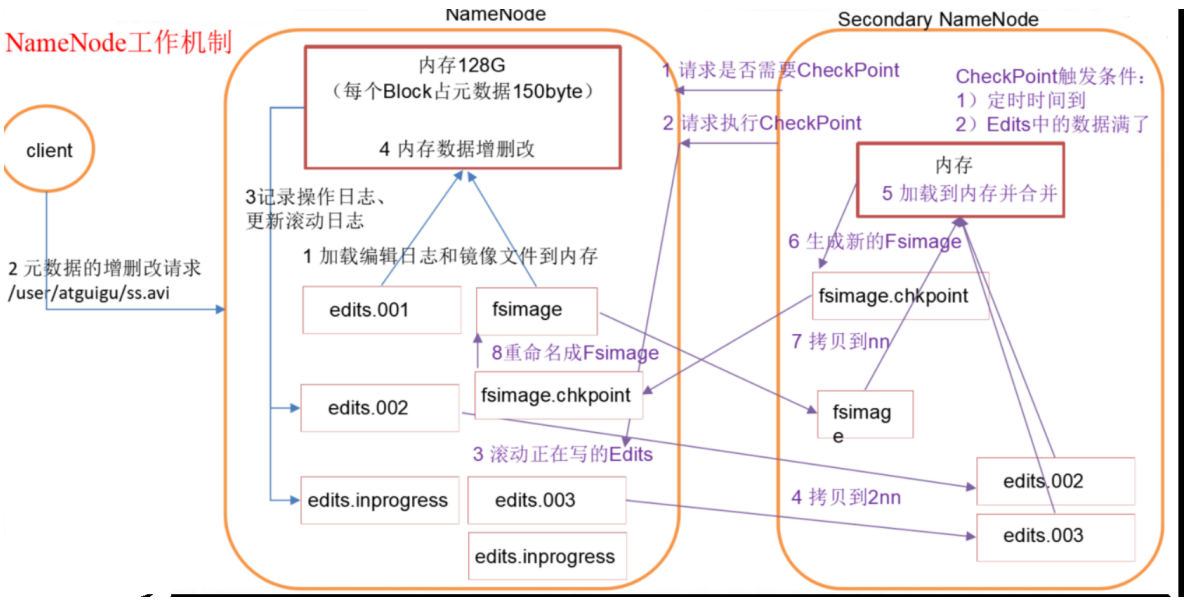
> 对未来真正的慷慨，是把一切献给现在

NameNode和SecondaryNameNode工作机制



第一阶段：NameNode启动

（1）第一次启动NameNode格式化后，创建**Fsimage和Edits文件**。如果不是第一次启动，直接加载编辑日志和镜像文件到内存。

> Fsimage文件（镜像文件）：HDFS文件系统元数据的一个永久性的检查点，其中包含HDFS文件系统的所有目录和文件idnode的序列化信息。

> Edits文件（编辑日志）：存放HDFS文件系统的所有更新操作的路径，文件系统客户端执行的所有写操作首先会被记录到edits文件中。

（2）客户端对元数据进行增删改的请求。

（3）NameNode记录操作日志，更新滚动日志。

【日志是为了到时候重启的时候恢复HDFS】

【可以手动滚动日志 hdfs dfsadmin -rollEdits】

（4）NameNode在内存中对数据进行增删改。

第二阶段：Secondary NameNode工作

> （1）Secondary NameNode询问NameNode是否需要CheckPoint。直接带回NameNode是否检查结果。
> （2）Secondary NameNode请求执行CheckPoint。
> （3）NameNode滚动正在写的Edits日志。
> （4）将滚动前的编辑日志和镜像文件拷贝到Secondary NameNode。
> （5）Secondary NameNode加载编辑日志和镜像文件到内存，并合并。
> （6）生成新的镜像文件fsimage.chkpoint。
> （7）拷贝fsimage.chkpoint到NameNode。
> （8）NameNode将fsimage.chkpoint重新命名成fsimage。

> 由于Edits中记录的操作会越来越多，**Edits文件会越来越大，导致NameNode在启动加载Edits时会很慢**，所以需要对Edits和Fsimage进行合并（所谓合并，就是将Edits和Fsimage加载到内存中，照着Edits中的操作一步步执行，最终形成新的Fsimage）。SecondaryNameNode的作用就是帮助NameNode进行Edits和Fsimage的合并工作。
>
> SecondaryNameNode首先会询问NameNode**是否需要CheckPoint**（触发CheckPoint需要满足两个条件中的任意一个，定时时间到和Edits中数据写满了）。直接带回NameNode是否检查结果。SecondaryNameNode执行CheckPoint操作，首先会让NameNode滚动Edits并生成一个空的edits.inprogress，**滚动Edits的目的是给Edits打个标记**，以后所有新的操作都写入edits.inprogress，其他未合并的Edits和Fsimage会拷贝到SecondaryNameNode的本地，然**后将拷贝的Edits和Fsimage加载到内存中进行合并**，生成fsimage.chkpoint，然后将fsimage.chkpoint拷贝给NameNode，重命名为Fsimage后替换掉原来的Fsimage。**NameNode在启动时就只需要加载之前未合并的Edits和Fsimage即可**，因为合并过的Edits中的元数据信息已经被记录在Fsimage中。

Fsimage镜像文件和Edits编辑日志解析

NameNode被格式化之后，将在/opt/module/hadoop-2.7.2/data/tmp/dfs/name/current目录中产生如下文件

```
fsimage_0000000000000000000
fsimage_0000000000000000000.md5
seen_txid
VERSION
```

（1）Fsimage文件：HDFS文件系统元数据的一个永久性的检查点，其中包含HDFS文件系统的所有目录和文件idnode的序列化信息。

（2）Edits文件：存放HDFS文件系统的所有更新操作的路径，文件系统客户端执行的所有写操作首先会被记录到Edits文件中。

（3）seen_txid文件保存的是一个数字，就是最后一个edits_的数字

（4）每次NameNode启动的时候都会将Fsimage文件读入内存，加载Edits里面的更新操作，保证内存中的元数据信息是最新的、同步的，可以看成NameNode启动的时候就将Fsimage和Edits文件进行了合并。

oiv查看Fsimage镜像文件

使用oiv命令，语法如下:

```
hdfs oiv -p 文件类型 -i镜像文件 -o 转换后文件输出路径
hdfs oiv -p XML -i fsimage_0000000000000000025 -o /opt/module/hadoop-2.7.2/fsimage.xml
```

oev查看Edits编辑日志文件

使用oev命令，语法如下:

```
hdfs oev -p 文件类型 -i编辑日志 -o 转换后文件输出路径

hdfs oev -p XML -i edits_0000000000000000012-0000000000000000013 -o /opt/module/hadoop-2.7.2/edits.xml
```

CheckPoint时间设置

触发SecondaryNameNode执行checkpoint保存数据的条件有两种：时间和次数

【1】通常SNN每隔一小时执行一次，配置hdfs-default.xml

```
<property>
  <name>dfs.namenode.checkpoint.period</name>
  <value>3600</value>
</property >
```

【2】当操作次数达到100完次时候

```
<property>
  <name>dfs.namenode.checkpoint.txns</name>
  <value>1000000</value>
<description>操作动作次数</description>
</property>

<property>
  <name>dfs.namenode.checkpoint.check.period</name>
  <value>60</value>
<description> 1分钟检查一次操作次数</description>
</property >
```