

HDFS Java API编程 —— 文件读写

任务描述

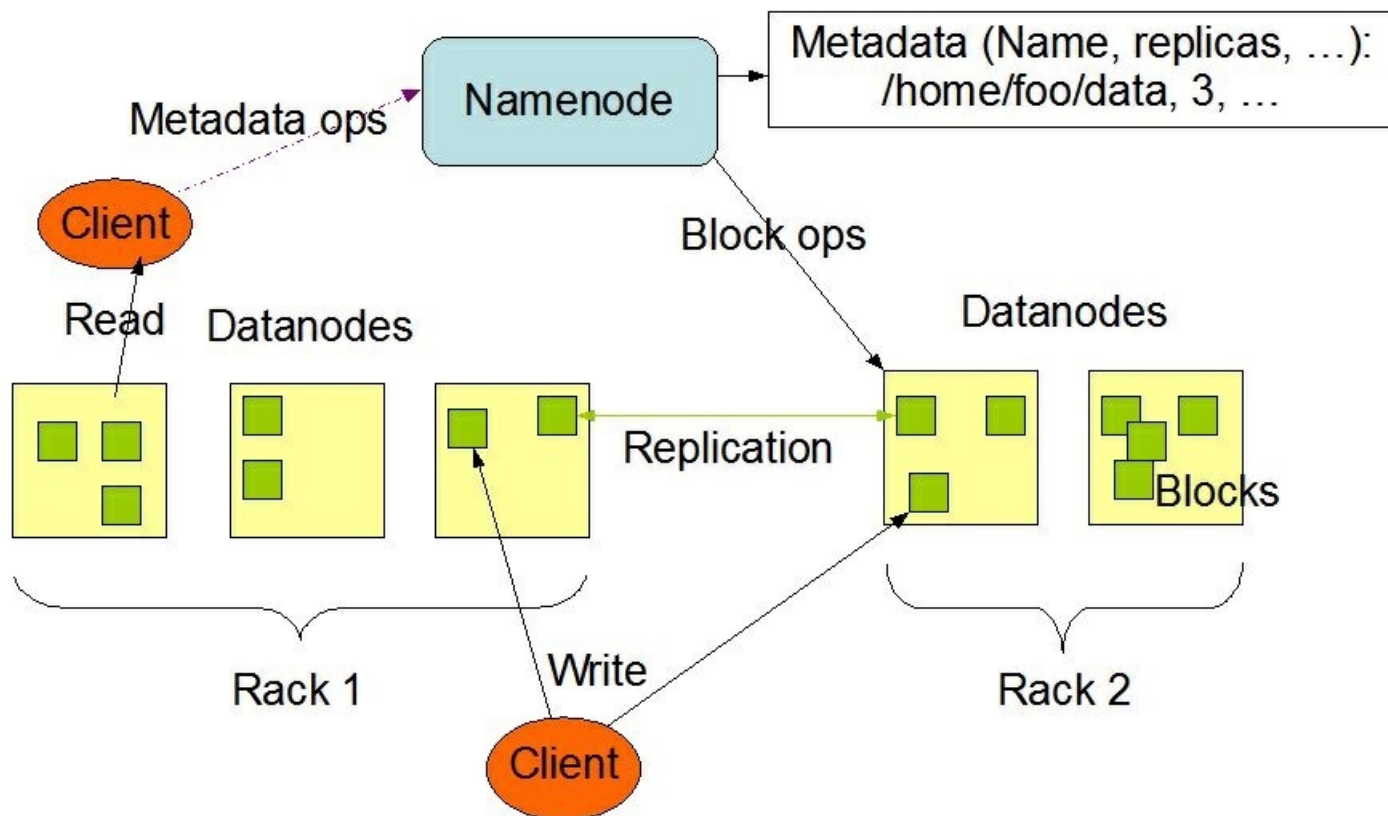
利用HDFS文件系统开放的API对HDFS系统进行文件的创建和读写

相关知识

HDFS文件系统

HDFS设计成能可靠地在集群中大量机器之间存储大量的文件，它以块序列的形式存储文件。文件中除了最后一个块，其他块都有相同的大小（一般64M）。属于文件的块为了故障容错而被复制到不同节点备份（备份数量有复制因子决定）。块的大小和读写是以文件为单位进行配置的。HDFS中的文件是一次写的，并且任何时候都只有一个写操作，但是可以允许多次读。

HDFS Architecture



HDFS文件创建和读写

HDFS文件创建和操作可分为三个步骤：

获取FileSystem对象；

```
Configuration conf = new Configuration(); //configuration类实现hadoop各模块之间值的传递
```

```
FileSystem fs = FileSystem.get(conf); //获取文件系统
```

```
Path file = new Path(xxx); //创建文件
```

通过FSDataOutputStream进行写入；

```
FSDDataOutputStream outputStream = fs.create(file); //获取输出流
```

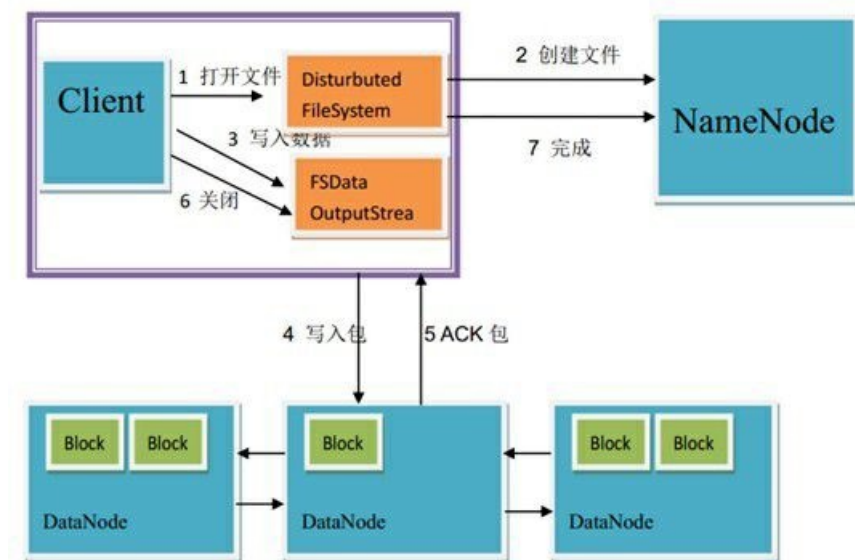
```
outputStream.writeUTF("XXX");//可以写入任意字符
```

```
outputStream.close();//记得关闭输出流
```

通过FSDDataInputStream将文件内容输出。

```
FSDDataInputStream inputStream = fs.open(file); //获取输入流
```

```
String data = inputStream.readUTF(); //读取文件
```



HDFS 写入数据流程图

编程要求

在右侧编辑器中补全代码，完成本关任务，具体要求如下：

获取hadoop的系统设置，并在其中创建HDFS文件，文件路径为/user/hadoop/myfile;

在myfile文件中添加字符串https://www.educoder.net;

读取刚刚创建myfile文件中的内容，并输出。

测试说明

本关的评测预设文件时/user/hadoop/myfile，所以创建文档的路径必须设置为/user/hadoop/myfile才能评测，否则会评测失败。

注：由于启动服务、编译等耗时，以及hdfs文件操作过程资源消耗较大且时间较长，因而评测时间较长，在40s左右。

实训使用软件环境为：JavaJDK1.8，hadoop2.7.4。

开始你的任务吧，gook luck!