

## Java API操作---上传文件

本关任务：使用HDFSAPI上传文件至集群。

### 相关知识

为了完成本关任务，你需要掌握：FSDaataInputStream对象如何使用。

### FSDaataOutputStream对象

我们知道在Java中要将数据输出到终端，需要文件输出流，HDFS的JavaAPI中也有类似的对象。

FileSystem类有一系列新建文件的方法，最简单的方法是给准备新建的文件制定一个path对象，然后返回一个用于写入数据的输出流：

```
public FSDaataOutputStream create(Path p)throws IOException
```

该方法有很多重载方法，允许我们指定是否需要强制覆盖现有文件，文件备份数量，写入文件时所用缓冲区大小，文件块大小以及文件权限。

注意：create()方法能够为需要写入且当前不存在的目录创建父目录，即就算传入的路径是不存在的，该方法也会为你创建一个目录，而不会报错。如果有时候我们并不希望它这么做，可以先用exists()方法先判断目录是否存在。

我们在写入数据的时候经常想要知道当前的进度，API也提供了一个Progressable用于传递回调接口，这样我们就可以很方便的将写入datanode的进度通知给应用了。

```
package org.apache.hadoop.util;
```

```
public interface Progressable{
```

```
    public void progress();
```

```
}
```

接下来我们通过一个例子来体验FSDaataOutputStream的用法：

1) 查看服务状态。在命令终端，执行命令jps

2) 启动Hadoop服务。启动Hadoop可以一次启动所有Hadoop服务，执行命令start-all.sh

也可以单独启动HDFS，执行命令start-dfs.sh

shell指令：

```
mkdir /develop
```

```
mkdir /develop/input
```

```
cd /develop/input
```

touch hello.txt

vim hello.txt 插入数据 wq 保存退出

start-dfs.sh

Java 代码如下

```
public class FileSystemUpload {  
    public static void main(String[] args) throws IOException {  
        /******* Begin *****/  
        File localPath = new File("/develop/hello.txt");  
        String hdfsPath = "hdfs://localhost:9000/user/hadoop/hello.txt";  
        InputStream in = new BufferedInputStream(new FileInputStream(localPath));  
        Configuration config = new Configuration();  
        FileSystem fs = FileSystem.get(URI.create(hdfsPath), config);  
        long fileSize = localPath.length() > 65536 ? localPath.length() / 65536 : 1;  
        FSDataOutputStream out = fs.create(new Path(hdfsPath), new Progressable() {  
            long fileCount = 0;  
            public void progress() {  
                System.out.println("总进度" + (fileCount / fileSize) * 100 + "%");  
                fileCount++;  
            }  
        });  
        IOUtils.copyBytes(in, out, 2048, true);  
    }  
}
```

运行结果:

```
WARN [main] - Unable to load native-hadoop library f  
总进度0%  
总进度100%
```

## 测试说明

平台会运行你的 `java` 程序，并查看集群的文件将文件信息输出到控制台，第一行属于警告信息可以忽略。

预期输出：

```
WARN [main] - Unable to load native-hadoop library for y  
总进度0%  
总进度100%  
迢迢牵牛星，皎皎河汉女。  
纤纤擢素手，札札弄机杼。  
终日不成章，泣涕零如雨。  
河汉清且浅，相去复几许？  
盈盈一水间，脉脉不得语。  
《迢迢牵牛星》
```

开始你的任务吧，祝你成功！