

# 实验 4: HDFS 编程实践

(教材 4.3 节)

## 本地运行

1. 使用 Eclipse 创建 Java 项目, 并为项目添加 HDFS 需要的 JAR 包. (教材 p93)

需添加的 JAR 包 (以 Hadoop2.7.3 为例, 注意自己的版本号):

`/usr/local/hadoop/share/hadoop/common/hadoop-common-2.7.3.jar`

`/usr/local/hadoop/share/hadoop/common/hadoop-nfs-2.7.3.jar`

`/usr/local/hadoop/share/hadoop/common/lib/*`

`/usr/local/hadoop/share/hadoop/hdfs/hadoop-hdfs-2.7.3.jar`

`/usr/local/hadoop/share/hadoop/hdfs/hadoop-hdfs-nfs-2.7.3.jar`

`/usr/local/hadoop/share/hadoop/hdfs/lib/*`

2. 创建类 `HDFSFileIfExist`, 依照教材编写 `main()`, 实现查看 HDFS 中是否存在一个文件的功能.

a) 教材代码查找的是在 HDFS 中对应的用户目录下是否存在 `test`.

(用户目录为 `/user/你的用户名`, 如 `/user/hadoop`)

b) 可通过修改代码中的 `fileName` 来实现对某个路径下的文件进行查询, 如:

`String fileName = "/test";` //查看 HDFS 中的根目录下是否存在 `test`

3. 编译, 运行程序.

a) 确保 HDFS 已启动 (启动命令: `/usr/local/hadoop/sbin/start-dfs.sh`)

b) 参照教材使用 Eclipse 编译运行, 也可使用命令行编译运行:

`javac -cp /usr/local/hadoop/share/hadoop/common/*:/usr/local/hadoop/share/hadoop/common/lib/*:/usr/local/hadoop/share/hadoop/hdfs/*:/usr/local/hadoop/share/hadoop/hdfs/lib/* HDFSFileIfExist.java`

`java -cp ./usr/local/hadoop/share/hadoop/common/*:/usr/local/hadoop/share/hadoop/common/lib/*:/usr/local/hadoop/share/hadoop/hdfs/*:/usr/local/hadoop/share/hadoop/hdfs/lib/* HDFSFileIfExist`

c) 执行程序后控制台会输出 文件存在 或 文件不存在.

4. 将程序打成 JAR 包, 在本地部署的 Hadoop 上运行.

## 集群上运行

1. 修改代码.

原 `conf.set("fs.defaultFS", "hdfs://localhost:9000");`

改为 `conf.set("fs.defaultFS", "hdfs://hadoop1:9000");`

或 `conf.set("fs.defaultFS", "hdfs://10.11.6.91:9000");`

2. 参照之前的步骤在本地将程序编译, 打成 JAR 包.

3. 将 JAR 包拷贝至集群.

a) 在本地控制台 (不要连接集群) 执行命令: `scp 你的 JAR 包路径 hadoopX@10.11.6.91:~`

(如 `scp /home/hadoop/HDFSFileIfExist.jar hadoop1@10.11.6.91:~`)

- b) 若安装了虚拟机, 既可在虚拟机中按照上述方法完成该步骤, 也可将 JAR 包拷贝至 Windows, 再使用 Xftp 或 WinSCP 上传 JAR 包至集群.
- c) 完成拷贝后, 可使用 shell 命令查看是否上传成功.  
首先使用 Xshell 连接至 10.11.6.91, 再执行命令 `ls`  
若出现 `HDFSFileIfExist.jar` 则表示拷贝成功.

#### 4. 参照教材 p102 运行程序.

**注意** JAR 包的路径是 `~/HDFSFileIfExist.jar`, 与之前不同.

## 附: 集群的使用

注意:

- (1) 集群的使用采用互斥锁的方式: 先在QQ群中声明ON (加锁), 再使用集群, 使用完毕后声明OFF (释放锁)。
- (2) 请每位同学在集群的HDFS上创建一个 `/user/<你的姓名>` 的目录作为自己的工作目录, 在集群上运行程序时, 各位同学需使用该目录作为输入/输出路径。
- (3) 集群VPN账号采用哈希方法分配:

```
i = 座位号 % 10 + 1;  
if i < 10  
    VPN用户名 = "dase"+"0"+i;  
else  
    VPN用户名 = "dase"+i;  
VPN密码 = VPN用户名;
```

例如: 座位号为9、19的同学, VPN账号为dase10; 座位号为10、20的同学, VPN账号为dase01