

昨日无法的访问 50070 端口，原来是 3.x 版本以后端口改为了 9870

MobaXterm 可以创建 SSH session，完成后就可以连接上虚拟机，自动通过 FTP 连接到虚拟机，实现文件复制

FTP 和 SFTP 是两种传输文件的协议

尝试运行 Hadoop 自带示例 WordCount, 又花费了很多时间, 主要原因是文件夹建错位置:

首先在 hdfs 根目录创建 input 文件夹

```
hadoop fs -mkdir /input
```

查看新建的文件夹，在 9870 端口的 web 管理页面也能看到，每个虚拟机看到的结果相同

```
hadoop fs -ls /
```

HDFS 是 hadoop 原生态的文件系统，上述命令是在 hdfs 文件系统下创建/input 目录，所以要用 hadoop 的命令方式才能查看到文件。无论是在 centos7 的 linux 可视化界面，还是在根目录 ls，还是 MobaXTerm 下都是看不见这个 input 目录的。同时输入路径不能包括文件，只能是包括文件的上一级目录，输出路径对应的目录不能存在。

复制文件

```
hadoop dfs -put /虚拟机的路径/文件名 /input(hdfs 文件系统的 input/目录)
```

在 hadoop-3.2.1/share/hadoop/mapreduce 路径下

```
hadoop jar hadoop-mapreduce-examples-3.2.1.jar wordcount /input /output
```

执行 hadoop fs -ls / 可看到多生成了 output 与 tmp 文件夹

执行 hadoop fs -ls /output 查看 output 文件夹

执行 hadoop fs -cat /output/part-r-00000 查看报告

```
[root@node01 mapreduce]# hadoop fs -cat /output/part-r-00000
2020-06-24 14:22:14,875 INFO sasl.SaslDataTransferClient: SASL
ed = false, remoteHostTrusted = false
hate      1
i          2
love      1
you        2
[root@node01 mapreduce]#
```

hadoop fs -rm -r /root/output 是删除对应路径的目录

要学会看错误信息，不能一味地百度