实验步骤：

Part-1 I/O实验部分

1. 部署实验用的hadoop-2.0.0-nkfs集群，详见文档2；(注意在core-site.xml中设置n、k值)
2. 写入实验数据 (写入nkfs与hdfs的时间对比)

time ./bin/hdfs dfs -put /home/fengqingqing/exp/data/OANC-GrAF/data/large\_txt nkfs:///

记录数据写入时间

完成后可用 ./bin/hdfs dfs -ls hdfs:///nkfs\_base 查看到输入数据被放在hdfs:///nkfs\_base/origin/目录下

3、读出数据（记录数据读时间）（从nkfs与hdfs读出的时间对比）

1. 运行time ./bin/hadoop cn.ict.magicube.fs.shell.NKFSShell -doraid 执行数据分片

记录数据分片时间

1. 运行 ./bin/hdfs dfs -ls hdfs:///nkfs\_base/parities 可查看各个文件分片，分片之后的原始数据被放在hdfs:///nkfs\_base/parities/xxx/part-xxx-xxx/origin目录下，各个分片分别为：

hdfs:///nkfs\_base/parities/xxx/part-xxx-xxx/parity\_x

运行./bin/hadoop cn.ict.magicube.fs.shell.NKFSShell -showlocations nkfs:/// 可查看各分片所在位置

5、（人为破坏掉部分分片，造成数据丢失）：

./bin/hdfs dfs -rm hdfs:///nkfs\_base/parities/xxx/part-xxx-xxx/parity\_x

1. 运行time ./bin/hadoop cn.ict.magicube.fs.shell.NKFSShell –dofix 可还原受损文件

记录数据恢复时间

7、用time ./bin/hdfs dfs -get hdfs:///nkfs\_base/parities/xxx/part-xxx-xxx/origin /tmp/xxx获取恢复之后的数据到本地

8、可用md5sum效验恢复的数据，与之前的原始数据对比，可得到一致的效验结果。

9、执行完以上操作之后，执行./bin/hadoop cn.ict.magicube.fs.shell.NKFSShell –reset，运行的结果，删除hdfs:///nkfs\_base下所有内容

Part-2 job执行实验部分

Wordcount

1. nkfs上运行wordcount

time ./bin/hadoop jar ./share/hadoop/mapreduce/hadoop-mapreduce-examples-2.0.0-alpha.jar wordcount nkfs:///2g out2x

real 0m43.965s

user 0m3.864s

sys 0m0.192s

或者

./bin/hadoop jar ./share/hadoop/mapreduce/hadoop-mapreduce-examples-2.0.0-alpha.jar wordcount hdfs:///nkfs\_base/origin/xxx output

1. 在hdfs上运行wordcount

time ./bin/hadoop jar ./share/hadoop/mapreduce/hadoop-mapreduce-examples-2.0.0-alpha.jar wordcount hdfs:///input out2gh

real 0m43.957s

user 0m3.858s

sys 0m0.193s

其它jobs:

注：HDFS文件操作：http://yu06206.iteye.com/blog/1396376

**在HDFS上创建目录**命令：user@namenode:hadoop$ bin/hadoop dfs -mkdir /文件名

**上传一个文件到HDFS**命令：user@namenode:hadoop$ bin/hadoop dfs -put 文件名 /user/yourUserName/

目录 myfiles/能够这样被拷贝进HDFS中：   
Shell代码   
   1. someone@anynode:hadoop$ bin/hadoop -put myfiles /user/myUsername

-put  的另外一种写法是 -copyFromLocal. 它们的功能和用法是一样的。

**从 HDFS 中导出数据**

命令：user@namenode:hadoop$ bin/hadoop dfs -cat foo

Step 2: 将HDFS中的文件拷贝到本地系统中。   
"get"命令有跟"put"命令相反的功能，它能够将HDFS中文件或目录拷贝到本地系统中。“get”命令的别名叫做copyToLocal.   
  
  
Shell代码   
   1. someone@anynode:hadoop$ bin/hadoop dfs -get foo localFoo

跟 "put"命令一样，"get"操作既可以操作文件，也可以操作目录。

**HDFS全局状态信息**

命令：bin/hadoop dfsadmin -report

获取帮助 - 跟 dfs 模块是一样的, 你可以使用 bin/hadoop dfsadmin -help命令来获取特定的命令的一些用法。

附录：

gb 28物理集群上的实验结果记录

time ./bin/hdfs dfs -put /home/fengqingqing/exp/data/OANC-GrAF/data/2g nkfs:///

real 0m43.084s

user 0m14.463s

sys 0m3.820s

time ./bin/hdfs dfs -get hdfs:///nkfs\_base/origin/2g /tmp/

real 0m21.602s

user 0m10.530s

sys 0m6.239s

在hdfs上执行同样操作：

time ./bin/hdfs dfs -put /home/fengqingqing/exp/data/OANC-GrAF/data/2g hdfs:///input

real 0m23.780s

user 0m13.819s

sys 0m2.593s