第二天作业

谭雨璇 2015141462202

上午：

- 主节点的文件目录如下：

```
[root@hdp-node-01 hadoop-2.6.5]# tree hdpdata/
hdpdata/
├── dfs
│   ├── data
│   │   ├── current
│   │   │   ├── BP-347816925-192.168.190.101-1530076630118
│   │   │   │   ├── current
│   │   │   │   │   ├── finalized
│   │   │   │   │   │   └── subdir0
│   │   │   │   │   │       └── subdir0
│   │   │   │   │   │           ├── blk_1073741825
│   │   │   │   │   │           └── blk_1073741825_1001.meta
│   │   │   │   │   ├── rbw
│   │   │   │   │   └── VERSION
│   │   │   │   ├── dncp_block_verification.log.curr
│   │   │   │   ├── dncp_block_verification.log.prev
│   │   │   │   └── tmp
│   │   │   └── VERSION
│   │   └── in_use.lock
│   ├── name
│   │   ├── current
│   │   │   ├── edits_0000000000000000001-0000000000000000010
│   │   │   ├── edits_0000000000000000011-0000000000000000012
│   │   │   ├── edits_0000000000000000013-0000000000000000014
│   │   │   ├── edits_0000000000000000015-0000000000000000016
│   │   │   ├── edits_0000000000000000017-0000000000000000018
│   │   │   ├── edits_0000000000000000019-0000000000000000020
│   │   │   ├── edits_0000000000000000021-0000000000000000022
│   │   │   ├── edits_0000000000000000023-0000000000000000024
│   │   │   ├── edits_0000000000000000025-0000000000000000026
│   │   │   ├── edits_0000000000000000027-0000000000000000028
│   │   │   ├── edits_0000000000000000029-0000000000000000030
│   │   │   ├── edits_0000000000000000031-0000000000000000032
│   │   │   ├── edits_0000000000000000033-0000000000000000034
│   │   │   ├── edits_0000000000000000035-0000000000000000036
│   │   │   ├── edits_0000000000000000037-0000000000000000038
```

```
│   │   │   ├── edits_0000000000000000017-0000000000000000018
│   │   │   ├── edits_0000000000000000019-0000000000000000020
│   │   │   ├── edits_0000000000000000021-0000000000000000022
│   │   │   ├── edits_0000000000000000023-0000000000000000024
│   │   │   ├── edits_0000000000000000025-0000000000000000026
│   │   │   ├── edits_0000000000000000027-0000000000000000028
│   │   │   ├── edits_0000000000000000029-0000000000000000030
│   │   │   ├── edits_0000000000000000031-0000000000000000032
│   │   │   ├── edits_0000000000000000033-0000000000000000034
│   │   │   ├── edits_0000000000000000035-0000000000000000036
│   │   │   ├── edits_0000000000000000037-0000000000000000038
│   │   │   ├── edits_0000000000000000039-0000000000000000040
│   │   │   ├── edits_0000000000000000041-0000000000000000042
│   │   │   ├── edits_0000000000000000043-0000000000000000044
│   │   │   ├── edits_0000000000000000045-0000000000000000046
│   │   │   ├── edits_0000000000000000047-0000000000000000048
│   │   │   ├── edits_inprogress_0000000000000000049
│   │   │   ├── fsimage_0000000000000000000
│   │   │   ├── fsimage_0000000000000000000.md5
│   │   │   ├── seen_txid
│   │   │   └── VERSION
│   │   └── in_use.lock
│   └── namesecondary
│       ├── current
│       │   ├── edits_0000000000000000001-0000000000000000002
│       │   ├── edits_0000000000000000003-0000000000000000004
│       │   ├── fsimage_0000000000000000002
│       │   ├── fsimage_0000000000000000002.md5
│       │   ├── fsimage_0000000000000000004
│       │   ├── fsimage_0000000000000000004.md5
│       │   └── VERSION
│       └── in_use.lock
└── nm-local-dir
    ├── filecache
    ├── nmPrivate
    └── usercache

18 directories, 41 files
[root@hdp-node-01 hadoop-2.6.5]#
```

HDFS metadata 主要存储两种类型的文件。fsimage：记录某一永久性检查点（Checkpoint）时整个 HDFS 的元信息；edits：记录所有对 HDFS 的写操作。

Checkpoint：
HDFS 会定期的对最近的 fsimage 和一批新 edits 文件进行 Checkpoint，Checkpoint 发生后会将前一次 Checkpoint 后的所有 edits 文件合并到新的 fsimage 中，HDFS 会保存最近两次 checkpoint 的 fsimage。Namenode 启动时会把最新的 fsimage 加载到内存中。

分析 dfs-name 目录：
1、VERSION
layoutVersion：HDFS metadata 版本号，通常只有 HDFS 增加新特性时才会更新这个版本号
namespaceID/clusterID/blockpoolID ：这三个 ID 在整个 HDFS 集群全局唯一，作用是引导 Datanode 加入同一个集群。在 HDFS Federation 机制下，会有多个 Namenode，所以不同 Namenode 直接 namespaceID 是不同的，分别管理一组 blockpoolID，但是整个集群中，clusterID 是唯一的，每次 format namenode 会生成一个新的，也可以使用-clusterid 手工指定 ID
storageType ：有两种取值 NAME_NODE /JOURNAL_NODE，对于 JournalNode 的参数 dfs.journalnode.edits.dir，其下的 VERSION 文件显示的是 JOURNAL_NODE
cTime ： HDFS 创建时间，在升级后会更新该值
2、edits_start transaction ID-end transaction ID
一个文件内的修改记录
3、edits_inprogress__start transaction ID
当前正在被追加的修改记录，HDFS 默认会为该文件提前申请 1MB 空间以提升性能
4、fsimage_end transaction ID
每次 checkpoint（合并所有 edits 到一个 fsimage 的过程）产生的最终的 fsimage，同时会生成一个.md5 的文件用来对文件做完整性校验
5、seen_txid
保存最近一次 fsimage 或者 edits_inprogress 的 transaction ID。需要注意的是，这并不是 Namenode 当前最新的 transaction ID，该文件只有在 checkpoing(merge of edits into a fsimage) 或者 edit log roll(finalization of current edits_inprogress and creation of a new one)时才会被更新。
这个文件的目的在于判断在 Namenode 启动过程中是否有丢失的 edits，由于 edits 和 fsimage 可以配置在不同目录，如果 edits 目录被意外删除了，最近一次 checkpoint 后的所有 edits 也就丢失了，导致 Namenode 状态并不是最新的，为了防止这种情况发生，Namenode 启动时会检查 seen_txid，如果无法加载到最新的 transactions，Namenode 进程将不会完成启动以保护数据一致性。
6、in_use.lock
防止一台机器同时启动多个 Namenode 进程导致目录数据不一致

- 从节点的文件目录如下：

```
[root@hdp-node-02 hadoop-2.6.5]# tree hdpdata/
hdpdata/
├── dfs
│   └── data
│       └── current
│           ├── BP-1371741619-192.168.190.101-1530069449224
│           │   ├── current
│           │   │   ├── dfsUsed
│           │   │   ├── finalized
│           │   │   ├── rbw
│           │   │   └── VERSION
│           │   ├── dncp_block_verification.log.curr
│           │   └── tmp
│           └── VERSION
└── nm-local-dir
    ├── filecache
    ├── nmPrivate
    └── usercache

12 directories, 4 files
```

分析 dfs-data 目录：

1、BP-random integer-NameNode-IP address-creation time

BP 代表 BlockPool 的意思，就是上面 Namenode 的 VERSION 中的集群唯一 blockpoolID，如果是 Federation HDFS，则该目录下有两个 BP 开头的目录，IP 部分和时间戳代表创建该 BP 的 NameNode 的 IP 地址和创建时间戳

2、VERSION

与 Namenode 类似，其中 storageType 是 DATA_NODE

3、finalized/rbw 目录

这两个目录都是用于实际存储 HDFS BLOCK 的数据，里面包含许多 block_xx 文件以及相应的.meta 文件，.meta 文件包含了 checksum 信息。

rbw 是“replica being written”的意思，该目录用于存储用户当前正在写入的数据。

4、dncp_block_verification.log

该文件用于追踪每个 block 最后修改后的 checksum 值，该文件会定期滚动，滚动后会移到.prev 文件

5、in_use.lock

防止一台机器同时启动多个 Datanode 进程导致目录数据不一致

下午：

- -ls

-ls [-d] [-h] [-R] [<path> ...] :

显示目录下包含的内容

```
[root@hdp-node-01 ~]# hdfs dfs -ls /
18/06/27 01:44:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 3 items
drwxr-xr-x   - root supergroup          0 2018-06-27 01:18 /mytest
drwxr-xr-x   - root supergroup          0 2018-06-27 01:18 /user
drwxr-xr-x   - root supergroup          0 2018-06-27 01:20 /wordcount
```

- -mkdir

-mkdir [-p] <path> ... :

在给定的路径下创建目录

```
[root@hdp-node-01 ~]# hdfs dfs -mkdir /abc
18/06/27 01:45:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[root@hdp-node-01 ~]# hdfs dfs -ls /
18/06/27 01:46:04 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 4 items
drwxr-xr-x   - root supergroup          0 2018-06-27 01:45 /abc
drwxr-xr-x   - root supergroup          0 2018-06-27 01:18 /mytest
drwxr-xr-x   - root supergroup          0 2018-06-27 01:18 /user
drwxr-xr-x   - root supergroup          0 2018-06-27 01:20 /wordcount
```

- -moveFromLocal

-moveFromLocal <localsrc> ... <dst> :

从本地剪切粘贴到 hdfs

```
[root@hdp-node-01 ~]# hdfs dfs -moveFromLocal /home/hadoop/test.txt /mytest/test
18/06/27 01:49:03 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[root@hdp-node-01 ~]# hdfs dfs -ls /mytest/test
18/06/27 01:49:18 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r--   3 root supergroup         19 2018-06-27 01:49 /mytest/test/test.txt
```

- -moveToLocal

-moveToLocal <src> <localdst> :

该命令尚未实现

- -appenToFile

-appendToFile <localsrc> ... <dst> :

把多个 src 文件的内容追加到 dst 文件当中

- -cat

-cat [-ignoreCrc] <src> ... :

输出匹配的 src 文件内容

```
[root@hdp-node-01 ~]# hdfs dfs -cat /test0.txt
18/06/27 04:23:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hello world!
```

```
[root@hdp-node-01 ~]# hdfs dfs -cp /test0.txt /test1.txt
18/06/27 04:27:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[root@hdp-node-01 ~]# hdfs dfs -cat /*.txt
18/06/27 04:27:05 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hello world!

hello world!
```

- -tail

-tail [-f] <file> :

输出 file 文件最后 1kb 内容

```
[root@hdp-node-01 ~]# hdfs dfs -tail /test0.txt
18/06/27 04:29:12 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
hello world!
```

- -text

-text [-ignoreCrc] <src> ... :

把某些特定格式的文件转化成文本格式输出

```
[root@hdp-node-01 ~]# hdfs dfs -text /test0.txt
18/06/27 04:33:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hello world!
```

- -chgrp

-chgrp [-R] GROUP PATH... :

修改文件所属的组

- -chmod

-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH... :

修改文件的权限

```
[root@hdp-node-01 ~]# hdfs dfs -chmod 111 /test.txt
18/06/27 04:41:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[root@hdp-node-01 ~]# hdfs dfs -ls /test.txt
18/06/27 04:41:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
---x--x--x   3 root supergroup          0 2018-06-27 01:58 /test.txt
```

- -chown

-chown [-R] [OWNER][:[GROUP]] PATH... :

修建文件所属的用户（和组）

- -copyFromLocal

-copyFromLocal [-f] [-p] [-l] <localsrc> ... <dst> :

同-put，把本地的 localsrc 的文件拷贝到 dst 位置

```
[root@hdp-node-01 /]# hdfs dfs -copyFromLocal ~/test.txt /mytest
18/06/27 04:53:29 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[root@hdp-node-01 /]# hdfs dfs -ls /mytest
18/06/27 04:53:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x   - root supergroup          0 2018-06-27 01:49 /mytest/test
-rw-r--r--   3 root supergroup         14 2018-06-27 04:53 /mytest/test.txt
```

- -copyToLocal

-copyToLocal [-p] [-ignoreCrc] [-crc] <src> ... <localdst> :

同-get，把 src 文件拷贝到本地的 localdst 位置

```
[root@hdp-node-01 ~]# hdfs dfs -copyToLocal /test.txt /home/hadoop
18/06/27 05:08:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[root@hdp-node-01 ~]# ls /home/hadoop
Desktop  Documents  Downloads  Music  Pictures  Public  Templates  test.txt  test.txt~  Videos
```

- -cp

-cp [-f] [-p | -p[topax]] <src> ... <dst> :

在 hdfs 内部进行 src 到 dst 的拷贝

```
[root@hdp-node-01 ~]# hdfs dfs -cp /test0.txt /test1.txt
18/06/27 04:27:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[root@hdp-node-01 ~]# hdfs dfs -cat /*.txt
18/06/27 04:27:05 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hello world!

hello world!
```

- -mv

-mv <src> ... <dst> :

在 hdfs 内部把 src 文件移到 dst 位置，当 src 文件为多个时，dst 必须为文件夹

```
mv: `/test1.txt': File exists
[root@hdp-node-01 ~]# hdfs dfs -mv /*.txt /user
18/06/27 05:16:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[root@hdp-node-01 ~]# hdfs dfs -ls user
18/06/27 05:16:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
ls: `user': No such file or directory
[root@hdp-node-01 ~]# hdfs dfs -ls /
18/06/27 05:16:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 4 items
drwxr-xr-x   - root supergroup          0 2018-06-27 01:45 /abc
drwxr-xr-x   - root supergroup          0 2018-06-27 04:53 /mytest
drwxr-xr-x   - root supergroup          0 2018-06-27 05:16 /user
drwxr-xr-x   - root supergroup          0 2018-06-27 01:20 /wordcount
[root@hdp-node-01 ~]# hdfs dfs -ls /user
18/06/27 05:17:01 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 4 items
drwxr-xr-x   - root supergroup          0 2018-06-27 01:18 /user/root
---x--x--x   3 root supergroup          0 2018-06-27 01:58 /user/test.txt
-rw-r--r--   3 root supergroup         14 2018-06-27 04:22 /user/test0.txt
-rw-r--r--   3 root supergroup         14 2018-06-27 04:27 /user/test1.txt
```

- -get

[-get [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]:

把 src 文件拷贝到本地的 localdst 位置

- -getmerge

- -put

[-put [-f] [-p] [-l] <localsrc> ... <dst>]:

把本地的 localsrc 的文件拷贝到 dst 位置

- -rm

-rm [-f] [-r|-R] [-skipTrash] <src> ... :

删除匹配的 src 文件

```
[root@hdp-node-01 ~]# hdfs dfs -rm /user/*.txt
18/06/27 05:25:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/06/27 05:25:49 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /user/test.txt
18/06/27 05:25:49 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /user/test0.txt
18/06/27 05:25:49 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /user/test1.txt
[root@hdp-node-01 ~]# hdfs dfs -ls /user
18/06/27 05:26:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
drwxr-xr-x   - root supergroup          0 2018-06-27 01:18 /user/root
```

- -rmdir

-rmdir [--ignore-fail-on-non-empty] <dir> ... :

删除空目录 dir

```
[root@hdp-node-01 ~]# hdfs dfs -mkdir /user/empty
18/06/27 05:28:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[root@hdp-node-01 ~]# hdfs dfs -ls /user
18/06/27 05:28:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x   - root supergroup          0 2018-06-27 05:28 /user/empty
drwxr-xr-x   - root supergroup          0 2018-06-27 01:18 /user/root
[root@hdp-node-01 ~]# hdfs dfs -rmdir /user/empty
18/06/27 05:29:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[root@hdp-node-01 ~]# hdfs dfs -ls /user/
18/06/27 05:29:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
drwxr-xr-x   - root supergroup          0 2018-06-27 01:18 /user/root
```

- -df

-df [-h] [<path> ...] :

统计文件系统的可用空间信息，-h 转换成方便阅读的形式

```
[root@hdp-node-01 ~]# hdfs dfs -df -h /
18/06/27 05:34:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Filesystem                Size  Used  Available  Use%
hdfs://hdp-node-01:9000  17.4 G  64 K     12.7 G    0%
```

- -du

-du [-s] [-h] <path> ... :

统计文件夹的大小信息，-s 转换成所有文件总的大小

```
[root@hdp-node-01 ~]# hdfs dfs -du -s -h /
18/06/27 05:35:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
52  /
[root@hdp-node-01 ~]# hdfs dfs -du -s -h /user/*
18/06/27 05:35:39 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
0  /user/root
```

- -count

-count [-q] [-h] <path> ... :

统计 path 目录下的文件夹个数、文件个数、文件总大小

```
[root@hdp-node-01 ~]# hdfs dfs -count /
18/06/27 05:39:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
           11            3                 52 /
```

- -setrep

-setrep [-R] [-w] <rep> <path> ... :

```
[root@hdp-node-01 ~]# hdfs dfs -setrep 3 /mytest/test.txt
18/06/27 05:43:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Replication 3 set: /mytest/test.txt
```