分布式文件系统HDFS

Grissom ｜ 2025年10月

TRANSWARP
星环科技

目录›
CONTENTS

1 HDFS Shell命令

2 HDFS API

3 HDFS 运维管理

1

chapter

HDFS Shell 命令

HDFS Shell 命令

语法

- hadoop fs <args> （使用面最广，可以操作任何文件系统）
- hdfs dfs <args>（只能操作HDFS文件系统）
- 大部分用法和Linux Shell类似，可通过**help**查看帮助

HDFS URI

- 格式：scheme://authority/path
- 示例：HDFS上的一个文件/parent/child
  - URI全写：hdfs://nameservice/parent/child（用nameservice替代namenodehost）
  - URI简写：/parent/child
  - 需在配置文件中定义hdfs://namenodehost

HDFS Shell 命令

Command	Description
hadoop fs -help	Return usage output
hadoop fs -usage command	Return the help for an individual command
hadoop fs -ls [-d] [-h] [-R] <args>	Options: -d: Directories are listed as plain files. -h: Format file sizes in a human-readable fashion (eg 64.0m instead of 67108864). -R: Recursively list subdirectories encountered
hadoop fs -get [-ignorecrc] [-crc] <src> <localdst>	Copy files to the local file system. Files that fail the CRC check may be copied with the -ignorecrc option. Files and CRCs may be copied using the -crc option. Example: hadoop fs -get /user/hadoop/file localfile hadoop fs -get hdfs://nn.example.com/user/hadoop/file localfile
hadoop fs -put <localsrc> ... <dst>	Copy single src, or multiple srcs from local file system to the destination file system. Also reads input from stdin and writes to destination file system.

HDFS Shell 命令

Command	Description
hadoop fs -cp [-f] [-p | -p[topax]] URI [URI ...] <dest>	Copy files from source to destination. This command allows multiple sources as well in which case the destination must be a directory. Options: -f: Overwrite the destination if it already exists. -p: Preserve file attributes [topx] (timestamps, ownership, permission, ACL, XAttr).
hadoop fs -mv URI [URI ...] <dest>	Moves files from source to destination. This command allows multiple sources as well in which case the destination needs to be a directory. Moving files across file systems is not permitted.
hadoop fs -rm [-f] [-r |-R] [-skipTrash] URI [URI ...]	Delete files specified as args. Options: -f: the option will not display a diagnostic message or modify the exit status to reflect an error if the file does not exist. -R: the option deletes the directory and any content under it recursively. -r: the option is equivalent to -R. -skipTrash: the option will bypass trash, if enabled, and delete the specified file(s) immediately. This can be useful when it is necessary to delete files from an over-quota directory.

2

chapter

HDFS API

HDFS API

- API

java.lang.Object
    org.apache.hadoop.conf.Configured
        org.apache.hadoop.fs.**FileSystem**

Modifier and Type	Method and Description
void	**copyFromLocalFile(boolean delSrc, boolean overwrite, Path src, Path dst)** The src file is on the local disk.
void	**copyToLocalFile(boolean delSrc, Path src, Path dst, boolean useRawLocalFileSystem)** The src file is under FS, and the dst is on the local disk.
boolean	**createNewFile(Path f)** Creates the given Path as a brand-new zero-length file.
boolean	**delete(Path f)**
abstract boolean	**delete(Path f, boolean recursive)** Delete a file.
boolean	**exists(Path f)** Check if exists.
FileStatus[]	**listStatus(Path[] files)** Filter files/directories in the given list of paths using default path filter.
boolean	**mkdirs(Path f)** Call mkdirs(Path, FsPermission) with default permission.
void	**moveFromLocalFile(Path src, Path dst)** The src file is on the local disk.
void	**moveToLocalFile(Path src, Path dst)** The src file is under FS, and the dst is on the local disk.

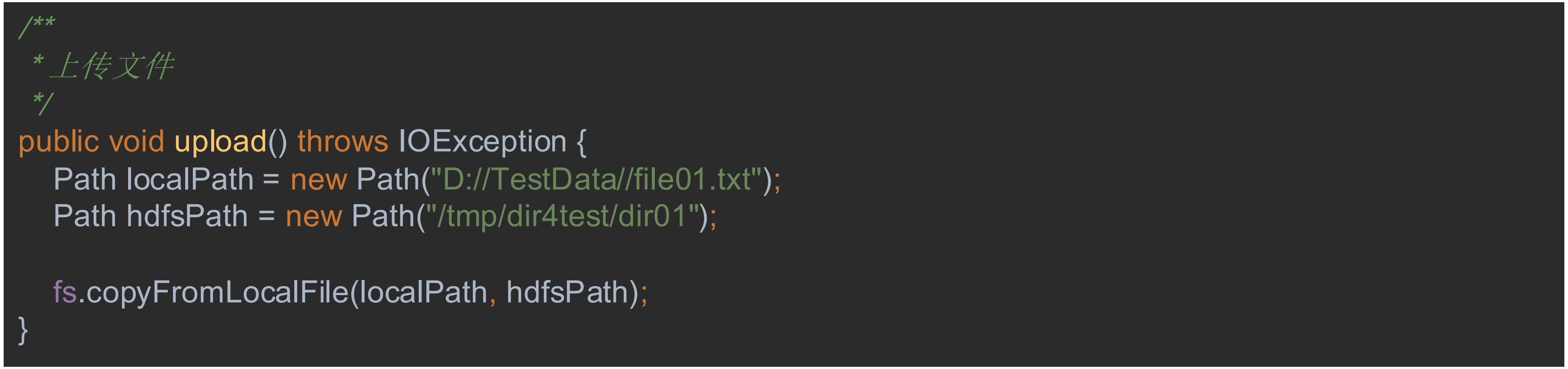
HDFS API

- API

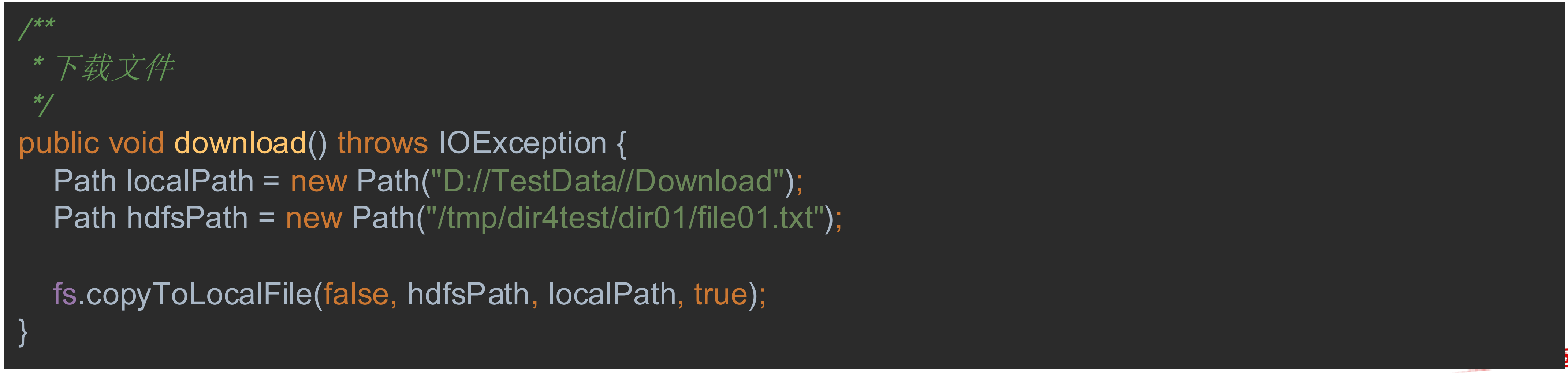
```
public void mkdir() throws IOException {
    Configuration conf = new Configuration();
    FileSystem fs = FileSystem.get(conf);

    Path path = new Path("/tmp/dir4test/dir01");

    // 创建文件目录
    fs.mkdirs(path);

    // 查看文件目录
    FileStatus[] status = fs.listStatus(path);
    for (FileStatus s : status) {
        System.out.println(s.getPath());
    }

    fs.close();
}
```

HDFS API

➢ API



```
/**
 * 上传文件
 */
public void upload() throws IOException {
    Path localPath = new Path("D://TestData//file01.txt");
    Path hdfsPath = new Path("/tmp/dir4test/dir01");

    fs.copyFromLocalFile(localPath, hdfsPath);
}
```



```
/**
 * 下载文件
 */
public void download() throws IOException {
    Path localPath = new Path("D://TestData//Download");
    Path hdfsPath = new Path("/tmp/dir4test/dir01/file01.txt");

    fs.copyToLocalFile(false, hdfsPath, localPath, true);
}
```

HDFS API

➢ API

```
/**
 * 删除文件
 */
public void delFile() throws IOException {
    Path delFilePath = new Path("/tmp/dir4test/dir01/file01.txt");

    boolean flag = fs.delete(delFilePath, false);
    if (flag) {
        System.out.println("Delete Success!");
    }
}
```

3

chapter

HDFS 运维管理

- ✓ 系统配置
- ✓ HDFS 管理命令
- ✓ UI 监控

系统配置

➢ 核心配置文件

- core-site.xml：Hadoop全局配置
- hdfs-site.xml：HDFS局部配置
- 示例：NameNode URI配置（core-site.xml）

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://nameservice:9000</value>
  </property>
</configuration>
```

➢ 环境变量文件

- Hadoop-env.sh：设置了HDFS运行所需的环境变量

系统配置

➢ hdfs-site.xml

Command	Description
dfs.namenode.name.dir	Determines where on the local filesystem the DFS name node should store the name table(fsimage). If this is a comma-delimited list of directories then the name table is replicated in all of the directories, for redundancy.
dfs.datanode.data.dir	Determines where on the local filesystem an DFS data node should store its blocks. If this is a comma-delimited list of directories, then data will be stored in all named directories, typically on different devices. Directories that do not exist are ignored.
dfs.blocksize	The default block size for new files, in bytes. You can use the following suffix (case insensitive): k(kilo), m(mega), g(giga), t(tera), p(peta), e(exa) to specify the size (such as 128k, 512m, 1g, etc.), Or provide complete size in bytes (such as 134217728 for 128 MB).
dfs.datanode.du.reserved	Reserved space in bytes per volume. Always leave this much space free for non hdfs use.
dfs.replication	Default block replication. The actual number of replications can be specified when the file is created. The default is used if replication is not specified in create time.
fs.trash.interval	Number of minutes after which the checkpoint gets deleted. If zero, the trash feature is disabled. This option may be configured both on the server and the client. If trash is disabled server side then the client side configuration is checked. If trash is enabled on the server side then the value configured on the server is used and the client configuration value is ignored.

HDFS管理命令

➢ NameNode（格式化或恢复）

```
# hdfs  namenode [-format [-clustered cid] [-force] [-nonInteractive] ] | [-recover [-force] ]
```

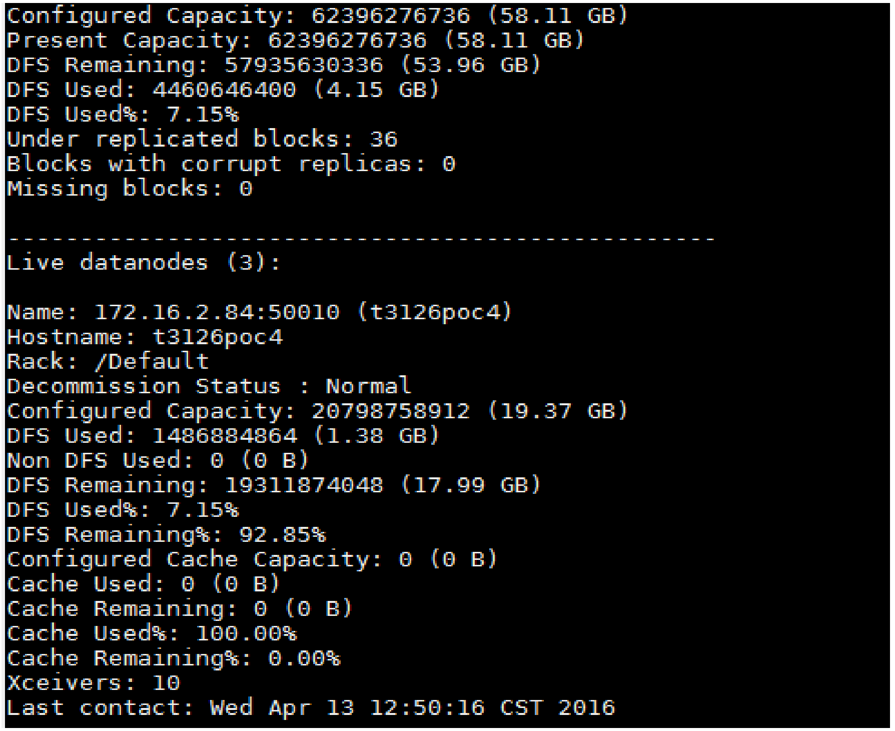
Command Options	Description
-format [-clusterid cid] [-force] [-nonInteractive]	Formats the specified NameNode. It starts the NameNode, formats it and then shut it down. -force option formats if the name directory exists. -nonInteractive option aborts if the name directory exists, unless -force option is specified.
-recover [-force]	Recover lost metadata on a corrupt filesystem.

HDFS管理命令

➢ Report（报告文件系统信息）

```
# hdfs dfsadmin [generic_options] [-report [-live] [-dead] [-decommissioning] ]
```

Command Options	Description
-report [-live] [-dead] [-decommissioning]	Reports basic filesystem information and statistics. Optional flags may be used to filter the list of displayed DataNodes.



```
Configured Capacity: 62396276736 (58.11 GB)
Present Capacity: 62396276736 (58.11 GB)
DFS Remaining: 57935630336 (53.96 GB)
DFS Used: 4460646400 (4.15 GB)
DFS Used%: 7.15%
Under replicated blocks: 36
Blocks with corrupt replicas: 0
Missing blocks: 0

-------------------------------------------------
Live datanodes (3):

Name: 172.16.2.84:50010 (t3126poc4)
Hostname: t3126poc4
Rack: /Default
Decommission Status : Normal
Configured Capacity: 20798758912 (19.37 GB)
DFS Used: 1486884864 (1.38 GB)
Non DFS Used: 0 (0 B)
DFS Remaining: 19311874048 (17.99 GB)
DFS Used%: 7.15%
DFS Remaining%: 92.85%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 10
Last contact: Wed Apr 13 12:50:16 CST 2016
```

HDFS管理命令

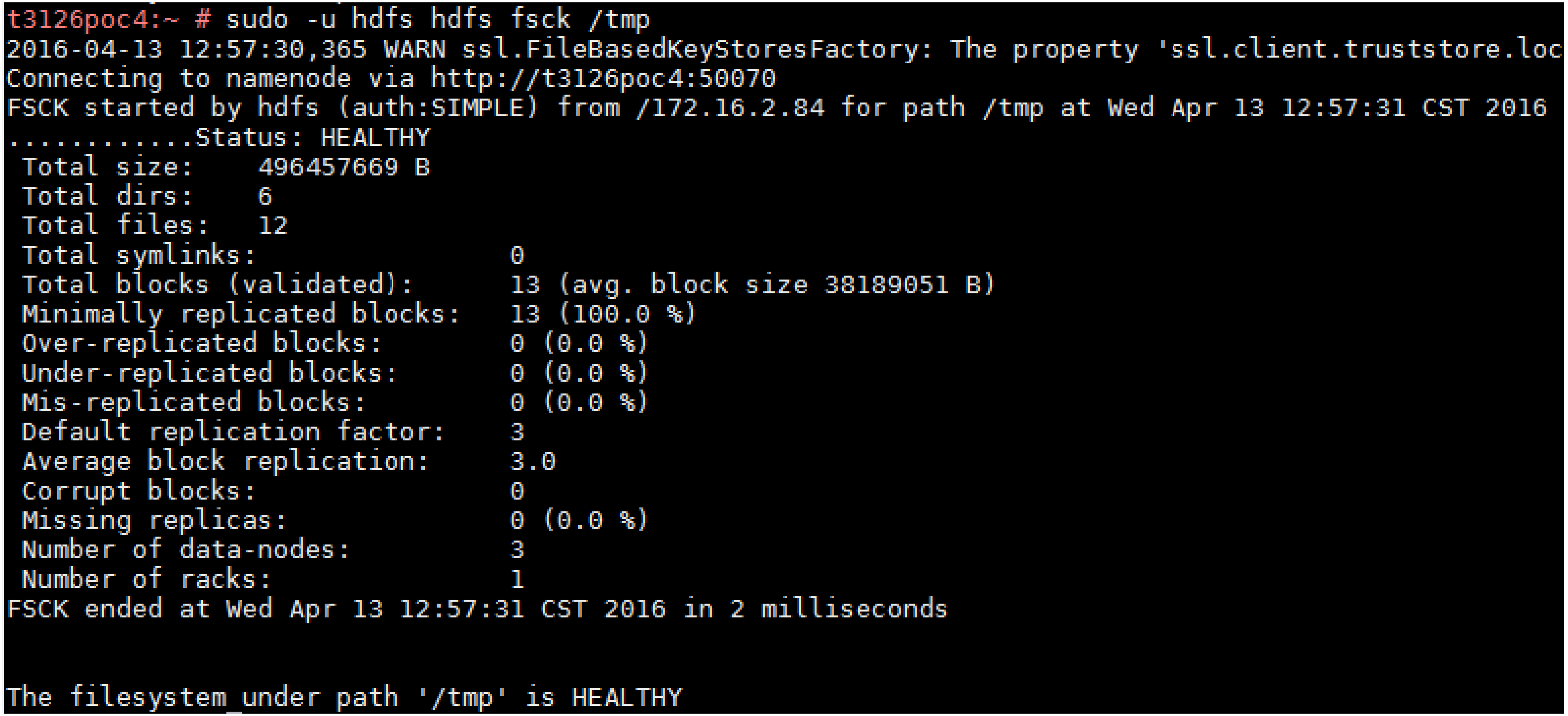
➢ Fsck（检查文件系统健康状况）

```
# hdfs fsck <path> [-move | -delete] | [-files [-blocks [-locations | -racks] ] ]
```

Command Options	Description
path	Start checking from this path.
-delete	Delete corrupted files.
-files	Print out files being checked.
-files -blocks	Print out the block report
-files -blocks -locations	Print out locations for every block.
-files -blocks -racks	Print out network topology for data-node locations.
-move	Move corrupted files to /lost+found.

HDFS管理命令

- Fsck（检查文件系统健康状况）



```
t3126poc4:~ # sudo -u hdfs hdfs fsck /tmp
2016-04-13 12:57:30,365 WARN ssl.FileBasedKeyStoresFactory: The property 'ssl.client.truststore.loc
Connecting to namenode via http://t3126poc4:50070
FSCK started by hdfs (auth:SIMPLE) from /172.16.2.84 for path /tmp at Wed Apr 13 12:57:31 CST 2016
............Status: HEALTHY
 Total size:    496457669 B
 Total dirs:    6
 Total files:   12
 Total symlinks:                0
 Total blocks (validated):      13 (avg. block size 38189051 B)
 Minimally replicated blocks:   13 (100.0 %)
 Over-replicated blocks:        0 (0.0 %)
 Under-replicated blocks:       0 (0.0 %)
 Mis-replicated blocks:         0 (0.0 %)
 Default replication factor:    3
 Average block replication:     3.0
 Corrupt blocks:                0
 Missing replicas:              0 (0.0 %)
 Number of data-nodes:          3
 Number of racks:               1
FSCK ended at Wed Apr 13 12:57:31 CST 2016 in 2 milliseconds


The filesystem under path '/tmp' is HEALTHY
```

HDFS管理命令

➢ Safemode（安全模式)

- NameNode启动会自动进入安全模式（也支持手动进入），该模式下只支持读操作
- 检测Block上报率超过阈值，才会离开安全模式
- 在TDH中，为避免用户错误退出安全模式，增加了检查变量，只有设置变量后，命令才可以正确执行
- 慎用hdfs dfsadmin leave，想了解变量设置，请联系3221723229（QQ）

```
# hdfs  dfsadmin  [generic_options] [-safemode enter | leave | get | wait]
```

Note: Safe mode maintenance command. Safe mode is a Namenode state in which it

1. does not accept changes to the name space (read-only)
2. does not replicate or delete blocks.

Safe mode is entered automatically at Namenode startup, and leaves safe mode automatically when the configured minimum percentage of blocks satisfies the minimum replication condition. Safe mode can also be entered manually, but then it can only be turned off manually as well.

HDFS管理命令

➢ NameNode HA（主备切换）

```
# hdfs haadmin -failover [--forcefence] [--forceactive] <serviceId> <serviceId>
# hdfs haadmin -getServiceState <serviceId>
# hdfs haadmin -transitionToActive <serviceId> [--forceactive]
# hdfs haadmin -transitionToStandby <serviceId>
```

Command Options	Description
-failover	initiate a failover between two NameNodes
-getServiceState	determine whether the given NameNode is Active or Standby
-transitionToActive	transition the state of the given NameNode to Active
-transitionToStandby	transition the state of the given NameNode to Standby

HDFS管理命令

- Decommission or Recommission（DataNode退役和服役)

```
# hdfs  dfsadmin  [generic_options]  -refreshNodes
```

Notes: Re-read the hosts and exclude files to update the set of Datanodes that are allowed to connect to the Namenode and those that should be decommissioned or recommissioned.

Command Options	Description
dfs.hosts	Names a file that contains a list of hosts that are permitted to connect to the namenode. The full pathname of the file must be specified. If the value is empty, all hosts are permitted.
dfs.hosts.exclude	Names a file that contains a list of hosts that are not permitted to connect to the namenode. The full pathname of the file must be specified. If the value is empty, no hosts are excluded.

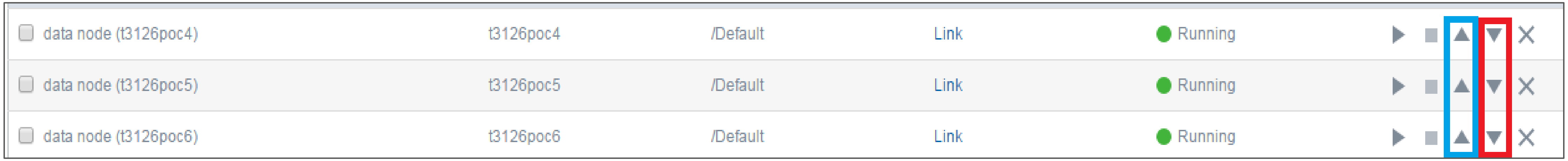
DataNode退役的基本步骤：

1. 将计划退役的DataNode列表加入dfs.hosts.exclude文件
2. hadoop dfsadmin -refreshNodes
3. 等待一段时间，这组DataNode的状态由Inservice变为Decommission
4. 将这组DataNode从dfs.hosts文件中删除
5. hadoop dfsadmin -refreshNodes

HDFS管理命令

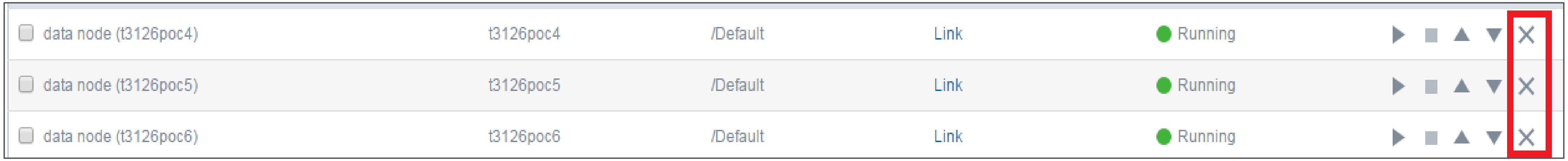
➢ Decommission or Recommission（DataNode退役和服役）

- 退役和服役（Web）



□	data node (t3126poc4)	t3126poc4	/Default	Link	● Running	▶ ■ ▲ ▼ ✕
□	data node (t3126poc5)	t3126poc5	/Default	Link	● Running	▶ ■ ▲ ▼ ✕
□	data node (t3126poc6)	t3126poc6	/Default	Link	● Running	▶ ■ ▲ ▼ ✕

- 删除DataNode（先退役再删除）



□	data node (t3126poc4)	t3126poc4	/Default	Link	● Running	▶ ■ ▲ ▼ ✕
□	data node (t3126poc5)	t3126poc5	/Default	Link	● Running	▶ ■ ▲ ▼ ✕
□	data node (t3126poc6)	t3126poc6	/Default	Link	● Running	▶ ■ ▲ ▼ ✕

HDFS管理命令

➢ Balancer（数据重分布）

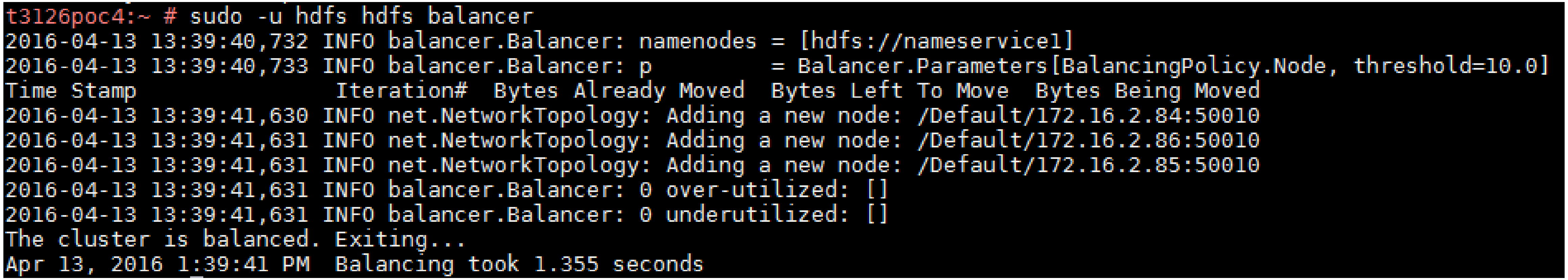
```
# hdfs balancer [-threshold <threshold>]
                [-exclude [-f <hosts-file> | <comma-separated list of hosts>] ]
                [-include [-f <hosts-file> | <comma-separated list of hosts>] ]
```

Command Options	Description
-threshold <threshold>	Percentage of disk capacity. This overwrites the default threshold.
-exclude -f <hosts-file> | <comma-separated list of hosts>	Excludes the specified datanodes from being balanced by the balancer.
-include -f <hosts-file> | <comma-separated list of hosts>	Includes only the specified datanodes to be balanced by the balancer.

HDFS管理命令

➢ Balancer（数据重分布）

- 集群平衡的标准：每个DataNode的存储使用率和集群总存储使用率的差值均小于阈值
- 默认阈值为10，设置值为0~100



```
t3126poc4:~ # sudo -u hdfs hdfs balancer
2016-04-13 13:39:40,732 INFO balancer.Balancer: namenodes = [hdfs://nameservice1]
2016-04-13 13:39:40,733 INFO balancer.Balancer: p         = Balancer.Parameters[BalancingPolicy.Node, threshold=10.0]
Time Stamp               Iteration#  Bytes Already Moved  Bytes Left To Move  Bytes Being Moved
2016-04-13 13:39:41,630 INFO net.NetworkTopology: Adding a new node: /Default/172.16.2.84:50010
2016-04-13 13:39:41,631 INFO net.NetworkTopology: Adding a new node: /Default/172.16.2.86:50010
2016-04-13 13:39:41,631 INFO net.NetworkTopology: Adding a new node: /Default/172.16.2.85:50010
2016-04-13 13:39:41,631 INFO balancer.Balancer: 0 over-utilized: []
2016-04-13 13:39:41,631 INFO balancer.Balancer: 0 underutilized: []
The cluster is balanced. Exiting...
Apr 13, 2016 1:39:41 PM  Balancing took 1.355 seconds
```

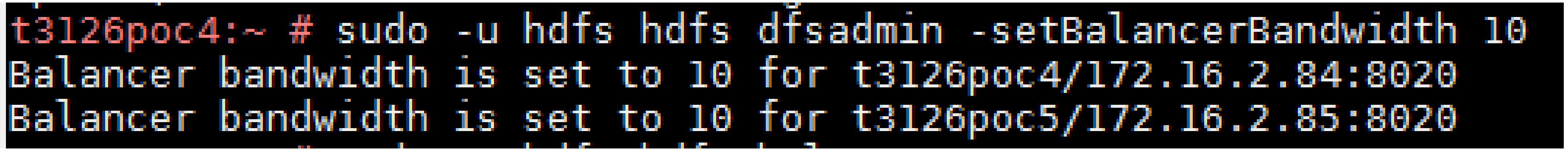
HDFS管理命令

BalancerBandwidth

- 默认带宽为1M/s，主要为了Balance的同时不影响HDFS操作
- 建议Balance的时候，带宽设为10M/s，并且停止操作HDFS

```
# hdfs  dfsadmin  [generic_options]  [-setBalancerBandwidth <bandwidth in bytes per second>]
```

Command Options	Description
-setBalancerBandwidth <bandwidth in bytes per second>	Changes the network bandwidth used by each datanode during HDFS block balancing. <bandwidth> is the maximum number of bytes per second that will be used by each datanode. This value overrides the dfs.balance.bandwidthPerSec parameter. NOTE: The new value is not persistent on the DataNode.



```
t3126poc4:~ # sudo -u hdfs hdfs dfsadmin -setBalancerBandwidth 10
Balancer bandwidth is set to 10 for t3126poc4/172.16.2.84:8020
Balancer bandwidth is set to 10 for t3126poc5/172.16.2.85:8020
```

HDFS管理命令

➢ Distcp（分布式拷贝）

- 大规模集群内部和集群之间拷贝的工具
- 使用MapReduce实现文件分发、错误处理恢复，以及报告生成

```
# hadoop distcp options [source_path...] <target_path>
```

Notes: distcp (distributed copy) is a tool used for large inter/intra-cluster copying. It uses MapReduce to effect its distribution, error handling and recovery, and reporting.

Command Options	Description
-m <num_maps>	Maximum number of simultaneous copies
-overwrite	Overwrite destination
-bandwidth	Specify bandwidth per map, in MB/second.

HDFS管理命令

➢ Quota（配额限制）

- HDFS允许管理员对用户的目录设置Quota，主要从两个维度：文件数量和文件大小
- 限制指定目录及子目录中的文件总数
- 限制指定目录中的所有文件的容量大小，需要考虑副本数

```
# hdfs dfsadmin -setSpaceQuota <N> <directory>...<directory>
```

Notes: Set the space quota to be N bytes for each directory.

```
# hdfs dfsadmin -clrSpaceQuota <directory>...<directory>
```

Notes: Remove any space quota for each directory.

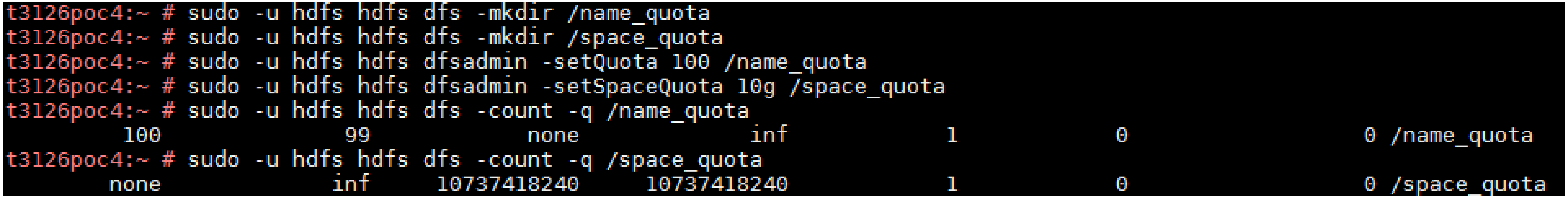
```
# hadoop fs -count -q [-h] [-v] <directory>...<directory>
```

Notes: With the -q option, also report the name quota value set for each directory, the available name quota remaining, the space quota value set, and the available space quota remaining. The -h option shows sizes in human readable format. The -v option displays a header line.

HDFS管理命令

➢ Quota（配额限制）

- 示例：hadoop fs -count -q
  - 输出：数量quota | 数量剩余 | 空间quota | 空间剩余 | 目录数量 | 文件数量 | 目录逻辑空间大小 | 路径



```
t3126poc4:~ # sudo -u hdfs hdfs dfs -mkdir /name_quota
t3126poc4:~ # sudo -u hdfs hdfs dfs -mkdir /space_quota
t3126poc4:~ # sudo -u hdfs hdfs dfsadmin -setQuota 100 /name_quota
t3126poc4:~ # sudo -u hdfs hdfs dfsadmin -setSpaceQuota 10g /space_quota
t3126poc4:~ # sudo -u hdfs hdfs dfs -count -q /name_quota
         100              99            none             inf            1            0                  0 /name_quota
t3126poc4:~ # sudo -u hdfs hdfs dfs -count -q /space_quota
        none             inf     10737418240     10737418240            1            0                  0 /space_quota
```

HDFS管理命令

➢ Snapshot（快照）

- HDFS快照是只读的，记录文件系统在某个时间点的副本
- HDFS快照可应用于根目录或其他子目录

```
# hdfs lsSnapshottableDir
```

Notes: Get all the snapshottable directories where the current user has permission to take snapshtos.

```
# hdfs snapshotDiff <path> <fromSnapshot> <toSnapshot>
```

Notes: Get the differences between two snapshots. This operation requires read access privilege for all files/directories in both snapshots.

```
# hdfs dfsadmin -allowSnapshot <path>
```

Notes: Allowing snapshots of a directory to be created.

```
# hdfs dfsadmin -disallowSnapshot <path>
```

Notes: Disallowing snapshots of a directory to be created.

HDFS管理命令

➢ Snapshot（快照）

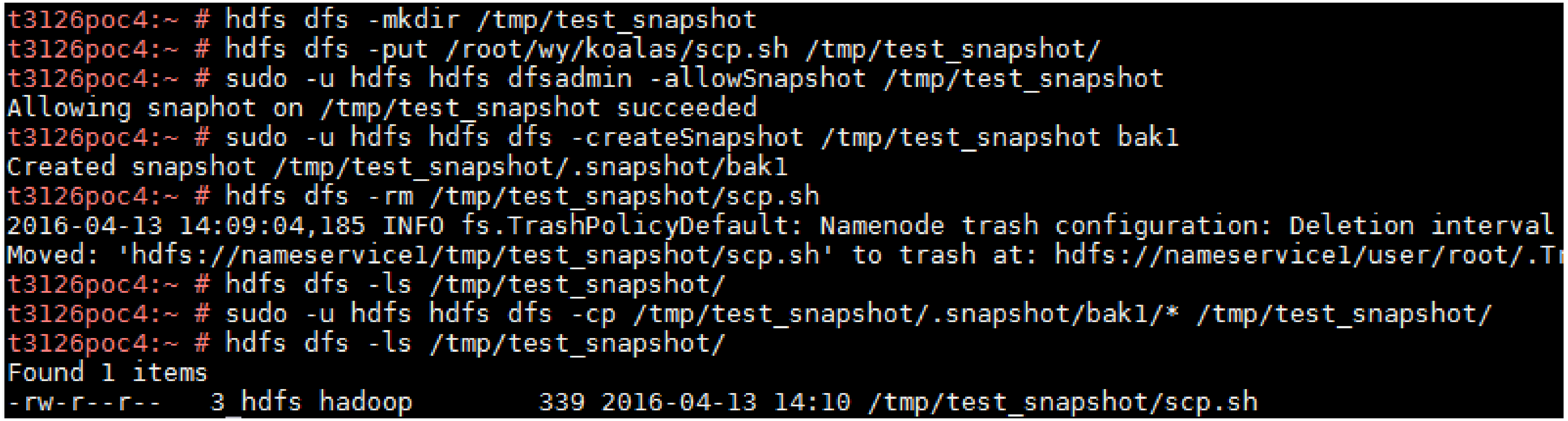
- 创建好的Snapshot文件夹在源文件夹下，命名为.snapshot/[<snapshotName>]
- 恢复的时候，直接使用cp命令即可

# hdfs dfs -createSnapshot <path> [<snapshotName>]

Notes: Create a snapshot of a snapshottable directory. This operation requires owner privilege of the snapshottable directory.

# hdfs dfs -deleteSnapshot <path> <snapshotName>

Notes: Delete a snapshot of from a snapshottable directory. This operation requires owner privilege of the snapshottable directory.



```
t3126poc4:~ # hdfs dfs -mkdir /tmp/test_snapshot
t3126poc4:~ # hdfs dfs -put /root/wy/koalas/scp.sh /tmp/test_snapshot/
t3126poc4:~ # sudo -u hdfs hdfs dfsadmin -allowSnapshot /tmp/test_snapshot
Allowing snaphot on /tmp/test_snapshot succeeded
t3126poc4:~ # sudo -u hdfs hdfs dfs -createSnapshot /tmp/test_snapshot bak1
Created snapshot /tmp/test_snapshot/.snapshot/bak1
t3126poc4:~ # hdfs dfs -rm /tmp/test_snapshot/scp.sh
2016-04-13 14:09:04,185 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval
Moved: 'hdfs://nameservice1/tmp/test_snapshot/scp.sh' to trash at: hdfs://nameservice1/user/root/.Tr
t3126poc4:~ # hdfs dfs -ls /tmp/test_snapshot/
t3126poc4:~ # sudo -u hdfs hdfs dfs -cp /tmp/test_snapshot/.snapshot/bak1/* /tmp/test_snapshot/
t3126poc4:~ # hdfs dfs -ls /tmp/test_snapshot/
Found 1 items
-rw-r--r--   3 hdfs hadoop        339 2016-04-13 14:10 /tmp/test_snapshot/scp.sh
```

Q&A

TRANSWARP
星环科技