

HDFS10 快照Snapshot管理

快照顾名思义，就是相当于对我们的hdfs文件系统做一个备份，我们可以通过快照对我们指定的文件夹设置备份，但是添加快照之后，并不会立即复制所有文件，而是指向同一个文件。当写入发生时，才会产生新文件。

快照Snapshot

Hdfs的快照（snapshot）是在某一时间点对指定文件系统拷贝，快照采用只读模式，可以对重要数据进行恢复、防止用户错误性的操作。

快照分为两种：

- 建立文件系统的索引，每次更新文件不会真正的改变文件，而是新开辟一个空间用来保存更改的文件
- 拷贝所有的文件系统。

Hdfs属于前者。Hdfs的快照的特征如下：

- 快照的创建是瞬间的，代价为 $O(1)$ ，取决于子节点扫描文件目录的时间。
- 当且仅当做快照的文件目录下有文件更新时才会占用小部分内存，占用内存的大小为 $O(M)$ ，其中 M 为更改文件或者目录的数量；
- 新建快照的时候，Datanode中的block不会被复制，快照中只是记录了文件块的列表和大小信息。
- 快照不会影响正常的hdfs的操作。对做快照之后的数据进行的更改将会按照时间顺序逆序的记录下来，用户访问的还是当前最新的数据，快照里的内容为快照创建的时间点时文件的内容减去当前文件的内容。

快照的基本语法

- 开启指定目录的快照功能

```
hdfs dfsadmin -allowSnapshot 路径
```

- 禁用指定目录的快照功能（默认就是禁用状态）

```
hdfs dfsadmin -disallowSnapshot
```

- 给某个路径创建快照snapshot

```
hdfs dfs -createSnapshot 路径
```

- 指定快照名称进行创建快照snapshot

```
hdfs dfs -createSanpshot 路径 名称
```

- 给快照重新命名

```
hdfs dfs -renameSnapshot 路径 旧名称 新名称
```

- 列出当前用户所有可快照目录

```
hdfs lsSnapshottableDir
```

- 比较两个快照的目录不同之处

```
hdfs snapshotDiff 路径1 快照名称1 快照名称2
```

- 删除快照snapshot

```
hdfs dfs -deleteSnapshot 路径1 快照名称
```

- 通过浏览器查看

<http://xxxx:50070/dfshealth.html#tab-snapshot>

Snapshot Summary

Snapshottable directories: 1

Path	Snapshot Number	Snapshot Quota	Modification Time	Permission	Owner	Group
/root	1	65536	2020/10/23 下午9:35:11	rwxt-xr-x	root	supergroup

Snapshotted directories: 1

Snapshot ID	Snapshot Directory	Modification Time
s20201023-213511.360	/root/.snapshot/s20201023-213511.360	2020/10/23 下午9:35:11

```
[hadoop@master /]$ hdfs dfsadmin -allowSnapshot /root
Allowing snapshot on /root succeeded
[hadoop@master /]$ hdfs dfs -createSnapshot /root
Created snapshot /root/.snapshot/s20201023-213511.360
```