

大数据开发技术

东北林业大学

卢洋

第七章

Hadoop 2.X 新特性

1. 集群间数据拷贝;

2. 小文件存档;

3. 回收站;

4. 快照管理.

7.1

集群间数据拷贝

服务器间数据的拷贝

1. 推:

```
scp -r hello.txt root@hadoop2:/tmp
```

2. 拉:

```
scp -r root@hadoop2:/tmp/hello.txt ./hello.txt
```

3. 通过本地主机中转实现两个远程主机的文件复制; 如果两个远程主机之间ssh没有配置的情况下也可以使用该方法:

```
scp -r root@hadoop2:/tmp/hello.txt root@hadoop3:/tmp
```


集群间数据拷贝

1. 采用distcp命令实现两个Hadoop集群之间的递归数据复制
2. `bin/hadoop distcp hdfs://hadoop1:9000/user/root/hello.txt hdfs://hadoop2:9000/user/root/hello.txt`