

# Hadoop-v2Distcp介绍

创建：未知用户 (twiki\_maotianming)，最新修改：马彬彬 于 2015-04-27

- 背景及介绍
- 使用说明
- 常用参数解释
  - ugi相关
  - 限速功能
  - 写慢节点规避
- 实际例子
  - 拷贝jx集群数据到yx集群
  - 拷贝多个集群的输入到yx集群
  - 采用更新方式（-update）拷贝jx集群数据到yx集群
  - 采用覆盖方式（-overwrite）拷贝jx集群数据到yx集群
  - 保留源文件属性
  - 忽略传输失败
  - 限制最大的map数
  - 采用文件列表作为输入
  - 限制每次传输的大小及文件数量
  - 删除目的集群中有这样的文件，但源集群中没有这样的文件

## 背景及介绍

---

为了适应大规模集群之间数据传输的需要：需要有一种工具能满足在短时间传输文件数据量大，文件数量多，且必须有一种机制保证两个集群数据的一致，即使出现某一个文件传输错误，并且还会有重试机制来保证数据正确。

**DistCp**（分布式拷贝）是用于大规模集群内部和集群之间拷贝的工具。它使用方便，传输高效，并且能错误处理和断点续传。它使用**Map/Reduce**实现文件分发，错误处理和恢复，以及报告生成。它把文件和目录的列表作为**map**任务的输入，每个任务会完成源列表中部分文件的拷贝。

## 使用说明

---

Distcp通过hadoop shell接口启动，命令参数如下，srcurl可以为多个。

**distcp [OPTIONS] <srcurl>\* <desturl>**

OPTIONS:

-p[rbugp]	Preserve status
	r: replication number
	b: block size
	u: user
	g: group
	p: permission
	-p alone is equivalent to -prbugp
-i	Ignore failures
-log <logdir>	Write logs to <logdir>
-m <num_maps>	Maximum number of simultaneous copies
-su <user,pwd>	UGI of source FS

-du <user,pwd>	UGI of destination FS
-overwrite	Overwrite destination
-update	Overwrite if src size different from dst size
-f <urilist_uri>	Use list at <urilist_uri> as src list
-filelimit <n>	Limit the total number of files to be <= n
-sizelimit <n>	Limit the total size to be <= n bytes
-delete	Delete the files existing in the dst but not in src

## 常用参数解释

- -p参数是保留dst文件的权限位是否与src集群的文件权限位一致，若不加此选项则在dst集群上生成的文件的权限信息将与distcp的ugi信息一致。
- -i 若在拷贝的过程中出现错误，则忽略此错误，否则就会进行4次重试直到整个任务失败。
- -m 设置最大的map数，非capacity
- -overwrite 将强行覆盖dst目录中和src重名的文件
- -update 将比较同名文件大小，若不一致则覆盖
- -f 可以将带传输的文件或目录写到平台上存放的一个文件中，多目录可以采用这种方式
- -filelimit 和 -sizelimit 是限制传输的文件数和大小，感觉用途不大

## ugi相关

2.11.6后的hadoop-v2版本支持src和dst的ugi不同。此时存在三种ugi:

- src集群的ugi: 在命令行用-su <user,pwd>指定;
- dst集群的ugi: 在命令行用-du <user,pwd>指定;
- 运行作业的ugi: 即原来的hadoop.job.ugi可在hadoop-site.xml或命令行用-D进行配置。

举例如下:

```
./hadoop distcp -su usera,passwd -du userb,passwd hdfs://A:8020/a/src hdfs://B:8020/b/src
```

若作业运行在c集群，则需要在该client的hadoop-site.xml，或命令行用-D进行配置。

通常情况下运行作业的ugi是src或dst中的一个。

**src的ugi**需对**src**的目录及其下面得文件全部有读权限，可以不是同一个用户；

对源集群的src文件必须可以保证读操作，否则**dist-cp**就会失败

**dst的ugi**需对**dst**的目录下有可执行权限，可以不是同一个用户。

对目的集群的文件夹也必须是可以保证写权限，否则**dist-cp**就会失败

若带上-p选项，则dst的ugi必须能将文件chown成src文件对应的属性的能力。

-p选项是保留源文件的属性，默认创建文件的属性是和dist-cp的ugi信息一致的。若带上此选项则在文件传输后将调用chown和chmod命令将其源文件的属性。但是，若distcp的ugi用户没有权限在dst集群上操作的话，则此选项无效。

## 限速功能

---

支持在每个并发线程（map）中控制速度，启动distcp时加入 `-D distcp.map.speed.kb=3000` ,即可将各个map传输速度控制在3000KB/s以内。

注意：

1. 此功能仅仅对各个map限速，不是对整个拷贝计算进行限速。
2. 通过设置map capacity和distcp.map.speed.kb来达到控制总的传输速度

## 写慢节点规避

---

支持在每个并发线程（map）中限制写文件速度以规避慢节点，通过在hadoop-site中添加参数 `dfs.client.slow.write.limit`单位Mb用以规避写过程中的慢节点。

注意：

1. 此功能仅仅对各个map中写文件的速度进行限制，不是对整个拷贝计算进行限速。

## 实际例子

---

### 拷贝jx集群数据到yx集群

---

```
./hadoop distcp "hdfs://jx-spi-test9.jx.baidu.com:64310/user/test/input" "hdfs://yx-mapred-a001-v1.yx01.baidu.com:64310/user/test/input"
```

将jx集群的/user/test/input目录下的文件及目录拷贝到yx集群的/user/test/input目录下。由于jx和yx集群的namenode的服务端口设为64310，所以此处也必须保持一致。

注意：

1. 若jx的/user/test/input是文件而不是目录，并且yx集群不存在/user/test/input不存在，则将拷贝到yx的是/user/test/input文件。
2. 反之不满足情况1，并且yx的/user/test/input目录不存在，则将拷贝的是/user/test/input这个目录。
3. 反之不满足情况1，并且yx的/user/test/input目录存在，则将拷贝的是/user/test/input/input这个目录。
4. 若源目录和目的目录下存在相同的文件，则计算会失败，这种情况就需要采用-update或-overwrite方式

另外，本地集群（启动distcp的集群）可以不用hdfs开头 直接写平台路径即可。

即如果在jx一台客户端上运行distcp可以写成这样：

```
./hadoop distcp -update "/user/test/input" "hdfs://yx-mapred-a001-v1.yx01.baidu.com:64310/user/test/input"
```

---

## 拷贝多个集群的输入到yx集群

```
./hadoop distcp "hdfs://jx-spi-test9.jx.baidu.com:64310/user/test/input" "hdfs://ai-logpf-master3-v.ai01:54310/user/test/input2""hdfs://yx-mapred-a001-v1.yx01.baidu.com:64310/user/test/input"
```

将多个集群的指定目录下的文件拷贝到yx集群，若多个源集群指定目录下的文件存在一样的则计算是失败的。

---

## 采用更新方式（-update）拷贝jx集群数据到yx集群

```
./hadoop distcp -update "hdfs://jx-spi-test9.jx.baidu.com:64310/user/test/input" "hdfs://yx-mapred-a001-v1.yx01.baidu.com:64310/user/test/input"
```

将jx集群的/user/test/input目录下的文件及目录拷贝到yx集群的/user/test/input目录下，若yx集群目的目录下存在同名文件则比较大小，若不一样则覆盖掉yx集群上的文件。

---

## 采用覆盖方式（-overwrite）拷贝jx集群数据到yx集群

```
./hadoop distcp -overwrite "hdfs://jx-spi-test9.jx.baidu.com:64310/user/test/input" "hdfs://yx-mapred-a001-v1.yx01.baidu.com:64310/user/test/input"
```

将jx集群的/user/test/input目录下的文件及目录拷贝到yx集群的/user/test/input目录下，若yx集群目的目录下存在同名文件则直接覆盖掉yx集群上的文件。

---

## 保留源文件属性

```
./hadoop distcp -prug "hdfs://jx-spi-test9.jx.baidu.com:64310/user/test/input" "hdfs://yx-mapred-a001-v1.yx01.baidu.com:64310/user/test/input"
```

将jx集群的/user/test/input目录下的文件及目录拷贝到yx集群的/user/test/input目录下，拷贝完毕后将yx上的/user/test/input文件副本（r），用户（u），组（g）设置为和jx集群对应文件一致。

注意：

运行distcp的ugi账户必须对目的集群即yx集群有root权限，可以将其chown和chmod为jx集群对应的属性。

如果未设置-p选项，则生成的文件将是启动distcp的ugi属性。副本数将是/user/test/input的默认副本数。

若只设置-p参数，则是所有属性都保持一致，即等于-prbugp,其中r:副本数，b: 块大小；u: 用户；g: 组；p: 权限。

## 忽略传输失败

---

```
./hadoop distcp -i "hdfs://jx-spi-test9.jx.baidu.com:64310/user/test/input" "hdfs://yx-mapred-a001-v1.yx01.baidu.com:64310/user/test/input"
```

如果没有加入-i选项，则若有文件传输失败，则后续的文件接着处理，但返回tasktracker失败，因而会在其它tasktracker重试，重试4次则整个计算失败。

若加入-i，则即使遇到传输失败也会返回成功，因而不在于其它tasktracker重试。整个计算返回成功。

## 限制最大的map数

---

```
./hadoop distcp -m 30 "hdfs://jx-spi-test9.jx.baidu.com:64310/user/test/input" "hdfs://yx-mapred-a001-v1.yx01.baidu.com:64310/user/test/input"
```

-m 设置最大的map数，非capacity，默认情况下distcp是按照256M一个map处理传输文件，一个文件若大于256M则将其整个放入一个map里，但若小于256M则将多个文件放入一个map凑足256M（最后一个文件超过256M也算进去）。

若设置了-m则总待传输大小/m数得到一个参考值，distcp将按照这个参考值和256M的大者来决定一个map的参考大小。此例中限制map不会超过30，目录总大小如果有30G，则平均每个map要传输的大小为1GB。

## 采用文件列表作为输入

---

```
./hadoop distcp -f "hdfs://jx-spi-test9.jx.baidu.com:64310/user/test/input_list.txt" "hdfs://yx-mapred-a001-v1.yx01.baidu.com:64310/user/test/input"
```

在jx集群上/user/test/input\_list.txt里面保存的是hdfs路径的文件，这里的路径可以是目录。

## 限制每次传输的大小及文件数量

---

通过-filelimit限制文件数量，distcp按递归遍历顺序依次取目录下的文件，当超过限制时后面的文件不在加入这次拷贝；

```
./hadoop distcp -filelimit 100 "hdfs://jx-spi-test9.jx.baidu.com:64310/user/test/input_list.txt" "hdfs://yx-mapred-a001-v1.yx01.baidu.com:64310/user/test/input"
```

限制一次拷贝的文件个数不超过100。

通过-sizelimit限制文件大小累加量（单位字节），distcp按递归遍历顺序依次取目录下的文件，当超过限制时后面的文件不在加入这次拷贝；

```
./hadoop distcp -sizelimit 10000000 "hdfs://jx-spi-test9.jx.baidu.com:64310/user/test/input_list.txt" "hdfs://yx-mapred-a001-v1.yx01.baidu.com:64310/user/test/input"
```

限制一次拷贝的文件大小累加量不超过10000000字节。

一般这种方式配合-update参数，就可以实现distcp分批次拷贝。

删除目的集群中有这样的文件，但源集群中没有这样的文件

```
./hadoop distcp -delete -overwrite "hdfs://jx-spi-test9.jx.baidu.com:64310/user/test/input" "hdfs://yx-mapred-a001-v1.yx01.baidu.com:64310/user/test/input"
```

如果jx集群/user/test/input下没有data1.txt文件，而yx（目的集群）/user/test/input有这样的文件，则-delete参数将删除yx集群上的/user/test/input/data1.txt文件。

注意：按目前的语义，此参数需和-overwrite一起使用。

附件列表


名称	大小	创建人	日期	
sync.tar.gz	1 kB	系统管理员	2014-10-25 13:05:24	更多
分布式拷贝工具distCp使用指南.doc	53 kB	系统管理员	2014-10-25 13:05:23	更多查看

赞

成为第一个赞同者

无标签

评论



系统管理员 发表：

本页面由老系统（<http://wiki.babel.baidu.com/>）迁移而来，原始页面请访问：[Hadoop-v2Distcp介绍](#)，迁移时间：2014-10-25 13:05:23