

无线大数据 (H3C WA1.0) 历史数据迁移方案

谢超凡 2017-06-27 发表

无线大数据 (H3C WA1.0版本) 系统现阶段没有平滑升级方案, 而且存在历史版本过多的情况。所以, 针对使用WA 1.0版本之前版本的局点, 涉及需要无线大数据系统迁移时, 需要在新集群部署WA 1.0版本的无线大数据软件, 同时, 如果客户要求保留历史数据, 需要参考本案例进行历史数据迁移。

无线大数据采集的数据主要是AP的设备在线信息和上网认证数据, 分别保存在HDFS和HBase中, 所以, 历史数据迁移主要是迁移HDFS和HBase中的数据。

在恢复这两项数据之前请先查看新环境中的HBase是否正常初始化:

I 查看HBase中的存在的数据表

hbase shell->list即可查看hbase中的表, 需要有如下5张表:

```
[root@node2 ~]# hbase shell
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/hdp/2.3.4.0-3485/hadoop/lib/slf4j-log4j12-1.7.10.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/hdp/2.3.4.0-3485/zookeeper/lib/slf4j-log4j12-1.6.1.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type 'exit<RETURN>' to leave the HBase Shell
Version 1.1.2.3.4.0-3485, r2b7f4d38f4dcca6eb81f99b5d37b2ea3beef09e, Wed Dec 16 03:49:28 UTC 2015

hbase(main):001:0> list
TABLE
ambarismoketest
wirelessbigdata:ap_sta_info
wirelessbigdata:endpoint_sta_info
wirelessbigdata:locus_sta_info
wirelessbigdata:zone_sta_info
5 row(s) in 0.2310 seconds

-> ["ambarismoketest", "wirelessbigdata:ap_sta_info", "wirelessbigdata:endpoint_sta_info", "wirelessbigdata:locus_sta_info", "wirelessbigdata:zone_sta_info"]
hbase(main):002:0> []
```

I 查看wirelessbigdata表空间下表的结构

describe 'wirelessbigdata:表名'查看表结构。除了endpoint_sta_info (有两个列族), 其他4张表都只有一个列族。

```
hbase(main):003:0> describe 'wirelessbigdata:zone_sta_info'
Table wirelessbigdata:zone_sta_info is ENABLED
wirelessbigdata:zone_sta_info
COLUMN FAMILIES DESCRIPTION
{NAME => 'info', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_FACTOR => '6', IN_MEMORY => 'false', BLOCKCACHE => 'true'}
1 row(s) in 0.1070 seconds

hbase(main):004:0> []
```

如果按照上述方法检查后发现HBase没有正常初始化, 请使用如下步骤初始化Hbase:

I 建立新数据库:

create_namespace 'wirelessbigdata' # 39;, 建立名为wirelessbigdata的表空间。

I 创建数据表

create # 39;tablename# 39;,{NAME=># 39;n1# 39;},{NAME=># 39;n2# 39;},.....,{NAME=># 39;nn# 39;} 在wirelessbigdata下建立表, n1,n2为列族, 视具体情况而定。(如果Hbase中数据表存在, 但是列族有问题, 需要先删除错误的数据库)

1、数据备份、恢复

之后备份恢复均使用表zone_sta_info为例子。

1.1 数据备份:

使用su hdfs切换到hdfs用户进行数据备份。进入hdfs用户后输入hbase

org.apache.hadoop.hbase.mapreduce.Driver export wirelessbigdata:zone_sta_info

/backup/zone_sta_info。

(/backup/zone_sta_info为备份到hdfs的地址, 可以自定, 但是不要使用wirelessdata)。

```
[root@node2 ~]# su hdfs
[hdfs@node2 root]$ hbase org.apache.hadoop.hbase.mapreduce.Driver export wirelessbigdata:zone_sta_info /backup/zone_sta_info
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/hdp/2.3.4.0-3485/hadoop/lib/slf4j-log4j12-1.7.10.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/hdp/2.3.4.0-3485/zookeeper/lib/slf4j-log4j12-1.6.1.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2017-05-05 08:59:16.007 INFO [main] mapreduce.Export: versions=1, starttime=0, endtime=9223372036854775807, keepDeletedCells=false
2017-05-05 08:59:17.791 INFO [main] impl.TimelineClientImpl: Timeline service address: http://node2.hde.h3c.com:8188/ws/v1/timeline/
2017-05-05 08:59:17.971 INFO [main] client.RPCProxy: Connecting to ResourceManager at node2.hde.h3c.com/100.4.84.82:8050
2017-05-05 08:59:19.460 INFO [main] zookeeper.RecoverableZooKeeper: Process identifier-hconnection-0x26b0a240 connecting to ZooKeeper ensemble
2017-05-05 08:59:19.467 INFO [main] zookeeper.ZooKeeper: Client environment:zookeeper.version=3.4.6-3485--1, built on 12/16/2015 02:35 GMT
2017-05-05 08:59:19.467 INFO [main] zookeeper.ZooKeeper: Client environment:host.name=node2.hde.h3c.com
2017-05-05 08:59:19.467 INFO [main] zookeeper.ZooKeeper: Client environment:java.version=1.7.0_79
```

```

2017-05-05 08:59:35,795 INFO [main] mapreduce.Job: map 100% reduce 0%
2017-05-05 08:59:35,807 INFO [main] mapreduce.Job: Job job_1493523081704_0014 completed successfully
2017-05-05 08:59:35,937 INFO [main] mapreduce.Job: Counters: 30
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=160678
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=98
  HDFS: Number of bytes written=115520
  HDFS: Number of read operations=4
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
Job Counters
  Launched map tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=5005
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=5005
  Total vcore-seconds taken by all map tasks=5005
  Total megabyte-seconds taken by all map tasks=10250240
Map-Reduce Framework
  Map input records=331
  Map output records=331
  Input split bytes=98
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=37
  CPU time spent (ms)=2220
  Physical memory (bytes) snapshot=266747904
  Virtual memory (bytes) snapshot=2480250880
  Total committed heap usage (bytes)=650641408
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=115520

```

导出成功后使用hadoop dfs -ls /backup/zone_sta_info可以查看备份文件。

```

[hdfs@node2 root]$ hadoop dfs -ls /backup/zone_sta_info
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

Found 2 items
-rw-r--r--  3 hdfs hdfs          0 2017-05-05 08:59 /backup/zone_sta_info/_SUCCESS
-rw-r--r--  3 hdfs hdfs    115520 2017-05-05 08:59 /backup/zone_sta_info/part-m-00000
[hdfs@node2 root]$

```

zone_sta_info已成功备份到hdfs，之后需再从hdfs到本地，请参考下一节hdfs备份。

1.2 数据备份的备用方法：

如HBase某张表过大无或者其他原因导致的数据导出的job不执行，可以使用如下备用方法

su hdfs->hdfs dfs -ls /apps/hbase/data/data/wirelessbigdata/表名/，可以看到数据分片。

```

[hdfs@node2 root]$ hdfs dfs -ls /apps/hbase/data/data/wirelessbigdata/locus_sta_info
Found 3 items
drwxr-xr-x  - hbase hdfs          0 2017-05-12 14:14 /apps/hbase/data/data/wirelessbigdata/locus_sta_info/.tabledesc
drwxr-xr-x  - hbase hdfs          0 2017-05-12 14:14 /apps/hbase/data/data/wirelessbigdata/locus_sta_info/.tmp
drwxr-xr-x  - hbase hdfs          0 2017-05-16 19:49 /apps/hbase/data/data/wirelessbigdata/locus_sta_info/e917038a537b37acacee
303a63f8bf15

```

hdfs dfs -get /apps/hbase/data/data/wirelessbigdata/表名/分片名 /tmp/xx，(tmp/xx为本地文件夹，需要事先建立)，用get命令将所有hdfs中的数据分片复制到本地目录下。之后将所有分片拷贝到新节点上，在新节点上执行hdfs dfs -copyFromLocal /tmp/xx /apps/hbase/data/data/wirelessbigdata/表名/(注意上文，先建立数据库和表)。将所有分片导入到hdfs之后，exit。

su hbase->hbase hbck -repair，由于导入分片之后会产生不一致，所以请使用hbase用户进行repair。之后，查看表状态为okay即可，进入hbase shell使用scan可以看到表中数据。

```

Deployed on: node2.hde.h3c.com,16020,1494937210850
Table hbase:meta is okay.
  Number of regions: 1
  Deployed on: node3.hde.h3c.com,16020,1494936465164
Table wirelessbigdata:ap_sta_info is okay.
  Number of regions: 1
  Deployed on: node1.hde.h3c.com,16020,1494935639352
Table hbase:namespace is okay.
  Number of regions: 1
  Deployed on: node1.hde.h3c.com,16020,1494935639352
Table wirelessbigdata:zone_sta_info is okay.
  Number of regions: 1
  Deployed on: node3.hde.h3c.com,16020,1494936465164
Table wirelessbigdata:locus_sta_info is okay.
  Number of regions: 1
  Deployed on: node2.hde.h3c.com,16020,1494937210850
Table ambarismoketest is okay.
  Number of regions: 1
  Deployed on: node3.hde.h3c.com,16020,1494936465164
0 inconsistencies detected.

```

1.3 数据恢复：

注意两点：第一，在导入前，要确保HBase中有同名表，否则导入时会提示“找不到表”；第二，同名表的列族也需要和历史数据的列族相同。

```

hbase(main):007:0> create_namespace 'wirelessbigdata'
0 row(s) in 0.0190 seconds

hbase(main):008:0> create 'wirelessbigdata:zone_sta_info',{NAME=>'info'}
0 row(s) in 1.3130 seconds

=> Hbase::Table - wirelessbigdata:zone_sta_info

```

我们首先将之前备份的文件放入新节点的tmp目录下，su HDFS 切换到HDFS用户，并使用hadoop fs -

copyFromLocal /tmp/backup /导入hdfs中（详见下一节HDFS文件恢复）。

```
hadoop fs -copyFromLocal /tmp/backup /  
hbase org.apache.hadoop.hbase.mapreduce.Driver import zone_test /backup/zone_sta_info
```

然后，输入hbase org.apache.hadoop.hbase.mapreduce.Driver import wirelessbigdata:zone_sta_info /backup/zone_sta_info从HDFS中将恢复HBase的表数据。

```
[root@node2 ~]# su hdfs  
[hdfs@node2 root]# hbase org.apache.hadoop.hbase.mapreduce.Driver import zone_test /backup/zone_sta_info  
SLF4J: Class path contains multiple SLF4J bindings.  
SLF4J: Found binding in [jar:file:/usr/hdp/2.3.4.0-3485/hadoop/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jar:file:/usr/hdp/2.3.4.0-3485/zookeeper/lib/slf4j-log4j12-1.6.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.  
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]  
2017-05-05 09:18:20,114 INFO [main] impl.TimelineClientImpl: Timeline service address: http://node2.hde.h3c.com:8188/ws/v1/timeline/  
2017-05-05 09:18:20,294 INFO [main] client.RMWProxy: Connecting to ResourceManager at node2.hde.h3c.com/100.4.84.82:8050  
2017-05-05 09:18:21,751 INFO [main] input.FileInputFormat: Total input paths to process : 1  
2017-05-05 09:18:21,822 INFO [main] mapreduce.JobSubmitter: number of splits:1  
2017-05-05 09:18:22,051 INFO [main] mapreduce.JobSubmitter: Submitting tokens for job: job_1493523081704_0015  
2017-05-05 09:18:22,444 INFO [main] impl.YarnClientImpl: Submitted application application_1493523081704_0015  
2017-05-05 09:18:22,505 INFO [main] mapreduce.Job: The url to track the job: http://node2.hde.h3c.com:8088/proxy/application_1493523081704_0015/  
2017-05-05 09:18:22,506 INFO [main] mapreduce.Job: Running job: job_1493523081704_0015  
2017-05-05 09:18:30,607 INFO [main] mapreduce.Job: Job job_1493523081704_0015 running in uber mode : false  
2017-05-05 09:18:30,608 INFO [main] mapreduce.Job: map 0% reduce 0%  
2017-05-05 09:18:37,687 INFO [main] mapreduce.Job: map 100% reduce 0%
```

完成后进入hbase shell->scan 'table'查看表 zone_sta_info。

2、数据备份、恢复

2.1 文件备份：

使用Xshell连接至HDFS数据所在节点，使用root用户登录，键入hadoop dfs -ls /wirelessdata即可查看HDFS中的数据。

```
[root@node2 ~]# su hdfs  
[hdfs@node2 root]# hbase org.apache.hadoop.hbase.mapreduce.Driver import zone_test /backup/zone_sta_info  
[root@node2 wirelessdata]# hadoop dfs -ls /wirelessdata  
DEPRECATED: Use of this script to execute hdfs command is deprecated.  
Instead use the hdfs command for it.  
  
Found 10 items  
-rw-r--r-- 3 hdfs hdfs 111058 2017-05-02 14:46 /wirelessdata/aruba_online_2017-04-01-11-54-55.csv  
-rw-r--r-- 3 hdfs hdfs 104750 2017-05-02 14:46 /wirelessdata/aruba_online_2017-04-01-11-59-55.csv  
-rw-r--r-- 3 hdfs hdfs 94059 2017-05-02 14:47 /wirelessdata/aruba_online_2017-04-01-12-04-55.csv  
-rw-r--r-- 3 hdfs hdfs 85007 2017-05-02 14:47 /wirelessdata/aruba_online_2017-04-01-12-09-55.csv  
-rw-r--r-- 3 hdfs hdfs 82065 2017-05-02 14:48 /wirelessdata/aruba_online_2017-04-01-12-14-55.csv  
-rw-r--r-- 3 hdfs hdfs 63915 2017-05-02 14:48 /wirelessdata/aruba_online_2017-04-01-12-19-55.csv  
-rw-r--r-- 3 hdfs hdfs 52468 2017-05-02 14:48 /wirelessdata/aruba_online_2017-04-01-12-24-55.csv  
-rw-r--r-- 3 hdfs hdfs 45050 2017-05-02 14:49 /wirelessdata/aruba_online_2017-04-01-12-29-55.csv  
-rw-r--r-- 3 hdfs hdfs 43237 2017-05-02 14:49 /wirelessdata/aruba_online_2017-04-01-12-34-55.csv  
-rw-r--r-- 3 hdfs hdfs 38870 2017-05-02 14:50 /wirelessdata/aruba_online_2017-04-01-12-39-55.csv
```

如

上图我们可以看到HDFS中的数据，我们可以键入hadoop fs -get /wirelessdata /tmp拷贝数据到本地的tmp文件夹。（下图是将数据拷贝到root目录，不推荐这么做，因为root存在权限问题可能会导致失败）

```
[root@node2 ~]# hadoop fs -get /wirelessdata /root  
[root@node2 ~]# cd /root  
[root@node2 ~]# ll  
total 126180  
-rw-r--r-- 1 root root 920 Apr 1 11:56 anaconda-ks.cfg  
-rw-r--r-- 1 hdfs hadoop 38870 May 4 16:04 aruba_online_2017-04-01-12-39-55.csv  
-rw-r--r-- 1 root root 8815 Apr 1 11:56 install.log  
-rw-r--r-- 1 root root 3384 Apr 1 11:55 install.log.syslog  
-rw-r--r-- 1 root root 29 Apr 16 19:09 nohup.out  
-rwxrwxrwx 1 root root 60865379 Apr 28 16:49 WA_Spark-1.0.0.bin  
-rwxrwxrwx 1 root root 68268932 Apr 28 16:49 WA_Web-1.0.0.bin  
drwxr-xr-x 2 root root 4096 May 4 16:38 wirelessdata  
[root@node2 ~]# ll wirelessdata  
total 720  
-rw-r--r-- 1 root root 111058 May 4 16:38 aruba_online_2017-04-01-11-54-55.csv  
-rw-r--r-- 1 root root 104750 May 4 16:38 aruba_online_2017-04-01-11-59-55.csv  
-rw-r--r-- 1 root root 94059 May 4 16:38 aruba_online_2017-04-01-12-04-55.csv  
-rw-r--r-- 1 root root 85007 May 4 16:38 aruba_online_2017-04-01-12-09-55.csv  
-rw-r--r-- 1 root root 82065 May 4 16:38 aruba_online_2017-04-01-12-14-55.csv  
-rw-r--r-- 1 root root 63915 May 4 16:38 aruba_online_2017-04-01-12-19-55.csv  
-rw-r--r-- 1 root root 52468 May 4 16:38 aruba_online_2017-04-01-12-24-55.csv  
-rw-r--r-- 1 root root 45050 May 4 16:38 aruba_online_2017-04-01-12-29-55.csv  
-rw-r--r-- 1 root root 43237 May 4 16:38 aruba_online_2017-04-01-12-34-55.csv  
-rw-r--r-- 1 root root 38870 May 4 16:38 aruba_online_2017-04-01-12-39-55.csv
```

之后我们查看tmp目录，可以看到目录下有之前我们从HDFS拷贝的wirelessdata文件夹，文件夹中有需要备份的.csv文件。

2.2 文件恢复

我们将前文放入tmp文件夹下的wirelessdata拷贝到物理机上（例子中改名为wirelessdata2，应使用wirelessdata），用Xftp从物理机拷贝到新节点tmp文件夹下，使用命令hadoop fs -copyFromLocal /tmp/wirelessdata2 /，即可将备份的本地数据导入到HDFS。

```
[root@node2 ~]# hadoop fs -copyFromLocal /root/wirelessdata2 /
[root@node2 ~]# hadoop fs -ls /
Found 10 items
drwxrwxrwx - yarn      hadoop      0 2017-04-17 08:36 /app-logs
drwxr-xr-x - hdfs      hdfs      0 2017-04-17 08:34 /apps
drwxr-xr-x - hdfs      hdfs      0 2017-04-17 08:34 /hdp
drwxr-xr-x - mapred    hdfs      0 2017-04-17 08:34 /mapred
drwxrwxrwx - mapred    hadoop     0 2017-04-17 08:34 /mr-history
drwxrwxrwx - spark     hadoop     0 2017-05-04 17:08 /spark-history
drwxrwxrwx - hdfs      hdfs      0 2017-04-17 08:34 /tmp
drwxr-xr-x - hdfs      hdfs      0 2017-05-04 16:17 /user
drwxr-xr-x - hdfs      hdfs      0 2017-05-04 16:59 /wirelessdata
drwxr-xr-x - root      hdfs      0 2017-05-04 17:08 /wirelessdata2
```