

区域气候模型评价系统

1、数据准备

将气象数据上传到在本机/usr/local/src/data/weather/gsod 目录下

```
[root@neudatal gsod]# pwd
/usr/local/src/data/weather/gsod
[root@neudatal gsod]# ll
total 20
drwxr-xr-x 2 root root 4096 Dec 29 11:08 2012
drwxr-xr-x 2 root root 4096 Dec 29 11:09 2013
drwxr-xr-x 2 root root 4096 Dec 29 11:10 2014
drwxr-xr-x 2 root root 4096 Dec 29 11:12 2015
drwxr-xr-x 2 root root 4096 Dec 29 11:14 2016
[root@neudatal gsod]# ls 2012
gsod_2012.tar
```

2、数据文件列表

1) 新建脚本 generate_input_list.sh:

```
[root@neudatal weather]# pwd
/usr/local/src/data/weather
[root@neudatal weather]# vi generate_input_list.sh
```

输入以下内容到脚本 generate_input_list.sh:

```
#!/bin/bash
```

```
a=$1
```

```
rm ncdc_files.txt
```

```
hdfs dfs -rm /ncdc_files.txt
```

```
while [ $a -le $2 ]
```

```
do
```

```
    filename="/GSOD/${a}/gsod_${a}.tar"
```

```
    echo -e "$filename" >>ncdc_files.txt
```

```
    a=`expr $a + 1`
```

done

#将文件列表上传至 hdfs

hdfs dfs -put ncdc_files.txt /

```
#!/bin/bash
a=$1
rm ncdc_files.txt
hdfs dfs -rm /ncdc_files.txt

while [ $a -le $2 ]
do
    filename="/GSOD/${a}/gsod_${a}.tar"
    echo -e "$filename" >>ncdc_files.txt
    a=`expr $a + 1`
done

hdfs dfs -put ncdc_files.txt /
~
~
~
~
"generate_input_list.sh" 13L, 263C
```

2) chmod u+x generate_input_list.sh

```
[root@neudatal weather]# chmod u+x generate_input_list.sh
```

3) ./generate_input_list.sh

```
[root@neudatal weather]# ./generate_input_list.sh
```

3、数据格式转换

Hadoop 读取 gz 文件，所以需要将 tar 文件进行 gz 格式转换

1) 新建脚本 load_ncdc_map.sh，并输入以下内容：

```
#!/bin/bash
```

```
while read hdfs_file
```

```
do
```

```
echo -e "$hdfs_file"
```

```
# Retrieve file from HDFS to local disk
```

```
echo "reporter:status:Retrieving $hdfs_file" >&2
```

```
hdfs dfs -get $hdfs_file .
# Create local directory
target=`basename $hdfs_file .tar`
mkdir $target

echo "reporter:status:Un-tarring $hdfs_file to $target" >&2
tar xf `basename $hdfs_file` -C $target
# Unzip each station file and concat into one file
echo "reporter:status:Un-gzipping $target" >&2
for file in $target/*
do
    gunzip -c $file >> $target.all
    echo "reporter:status:Processed $file" >&2
done
# Put gzipped version into HDFS
echo "reporter:status:Gzipping $target and putting in HDFS" >&2
gzip -c $target.all | hdfs dfs -put - /GSOD_ALL/$target.gz
rm `basename $hdfs_file`
rm -r $target
rm $target.all

done
```

```

#!/bin/bash
while read hdfs_file
do

echo -e "$hdfs_file"
# Retrieve file from HDFS to local disk
echo "reporter:status:Retrieving $hdfs_file" >&2
hdfs dfs -get $hdfs_file .
# Create local directory
target=`basename $hdfs_file .tar`
mkdir $target

echo "reporter:status:Un-tarring $hdfs_file to $target" >&2
tar xf `basename $hdfs_file` -C $target
# Unzip each station file and concat into one file
echo "reporter:status:Un-gzipping $target" >&2
for file in $target/*
do
    gunzip -c $file >> $target.all
    echo "reporter:status:Processed $file" >&2
done
# Put gzipped version into HDFS
echo "reporter:status:Gzipping $target and putting in HDFS" >&2
gzip -c $target.all | hdfs dfs -put - /GSOD_ALL/$target.gz
rm `basename $hdfs_file`
rm -r $target
rm $target.all

done
~
~
"load_ncdc_map.sh" 29L, 830C

```

2) `chmod u+x load_ncdc_map.sh`

3) `cat ncdc_files.txt | ./load_ncdc_map.sh`

```
[root@neudatal weather]# cat ncdc_files.txt | ./load_ncdc_map.sh
```

4) 查看结果 `hdfs dfs -ls /GSOD_ALL`

```

Found 5 items
-rw-r--r-- 1 root supergroup 92449016 2016-12-30 10:27 /data/weather/gsod_2012.gz
-rw-r--r-- 1 root supergroup 93219863 2016-12-30 10:29 /data/weather/gsod_2013.gz
-rw-r--r-- 1 root supergroup 95978976 2016-12-30 10:31 /data/weather/gsod_2014.gz
-rw-r--r-- 1 root supergroup 98446127 2016-12-30 10:33 /data/weather/gsod_2015.gz
-rw-r--r-- 1 root supergroup 100308984 2016-12-30 10:36 /data/weather/gsod_2016.gz

```

4、数据处理

//将整行数据作为一个字段，因为格式比较复杂，处理后续进行
create external table weather (STN string) stored as textfile location
"/data/weather";

//创建 **view**

create view weather2 as select substr(a.c1,1,6) as
stn,substr(a.c1,15,8) as ymd, substr(a.c1,27,4) as temp ,
substr(a.c1,38,4) as dewp , regexp_replace(substr(a.c1,47,6),"#","")
as slp, regexp_replace(substr(a.c1,58,6),"#","") as stp ,
regexp_replace(substr(a.c1,69,5),"#","") as visib, substr(a.c1,81,3)
as wdsp, regexp_replace(substr(a.c1,90,4),"#","") as mxspd,
regexp_replace(substr(a.c1,96,5),"#","") as gust,
regexp_replace(substr(a.c1,104,5),"#","") as maxtemp,
substr(a.c1,113,4) as mintemp , substr(a.c1,120,4) as prcp from
(select regexp_replace(stn," ","#") as c1 from weather where
substr(stn,1,1)<>"S") a;

//查看数据

Select * from weather2 limit 100;

007052	20121130	35.2	32.1	9999.9	9999.9	999.9	1.2	4.1	999.9	44.6	30.2	0.00
007059	20120314	45.1	35.6	9999.9	9999.9	2.7	1.3	2.9	999.9	59.0	33.8	0.00
007059	20120315	46.8	38.3	9999.9	9999.9	3.0	1.9	7.0	12.0	57.2	35.6	0.00
007059	20120316	50.1	40.7	9999.9	9999.9	3.3	1.6	5.1	999.9	62.6	41.0	0.00
007059	20120317	55.7	42.7	9999.9	9999.9	3.3	1.5	4.1	999.9	77.0	41.0	0.00
007059	20120318	61.2	44.6	9999.9	9999.9	3.4	1.5	4.1	999.9	78.8	46.4	0.00
007059	20120319	52.4	27.7	9999.9	9999.9	1.6	7.2	15.0	26.0	71.6	33.8	0.00
007059	20120320	45.4	10.6	9999.9	9999.9	2.7	5.2	15.0	28.0	59.0	33.8	0.00
007059	20120321	50.6	22.0	9999.9	9999.9	2.4	1.4	5.1	999.9	75.2	32.0	0.00
007059	20120322	55.7	27.4	9999.9	9999.9	3.0	1.4	4.1	999.9	75.2	33.8	0.00
007059	20120323	58.0	35.9	9999.9	9999.9	3.0	2.0	8.9	15.9	78.8	42.8	0.00
007059	20120324	60.6	39.7	9999.9	9999.9	2.9	2.0	6.0	999.9	82.4	42.8	0.00

5、指标分析