数据清洗：

1.root@node01 data]# hdfs dfs -mkdir -p /flume/record/2020-07-23/0925

2.[root@node01 data]# hdfs dfs -put ./record.list /flume/record/2020-07-23/0925

3.把etl20200618.jar.jar文件上传到到 /bigdata/install下面

4. [root@node01 soft]# hadoop jar mrxx.jar com.mr.etl.RunJob

mrxx.jar具体为对应处理数据的编号。

单天天气数据处理:

建表：以0725为例，后续每日数据都可以按同一格式分别建表

create external table if not exists weather0725(

cityname STRING,

citycode STRING,

date STRING,

week STRING,

weather STRING,

temphigh INT,

templow INT,

humidity INT,

pressure INT,

windspeed FLOAT,

winddirect STRING,

windpower STRING,

quality STRING,

rain FLOAT

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

location 'hdfs://node01:8020/opt/logs/record\_dimension/'

;

再建一个只含城市名字的表，后面统计降水量时多表协同，用cityname作为统计依据

create external table if not exists city (

city STRING

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

location 'hdfs://node01:8020/opt/logs/cityname\_dimension/'

;

再建一个空表temphigh（用hdfs上的一个空文件创建），格式和weather0725一样

在hive中：

按温度从高到低排序：

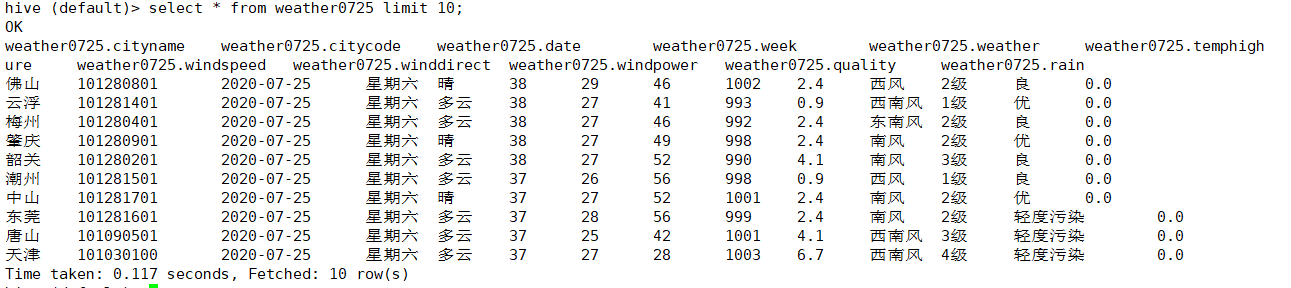
hive (default)> insert into temphigh

> select \* from weather0725

> order by temphigh desc（desc）表示降序

> ;

效果大致如图：



建一个用于统计累计降水量的表：

create external table if not exists totalrain(

cityname STRING,

rain FLOAT

)

ROW FORMAT DELIMITED

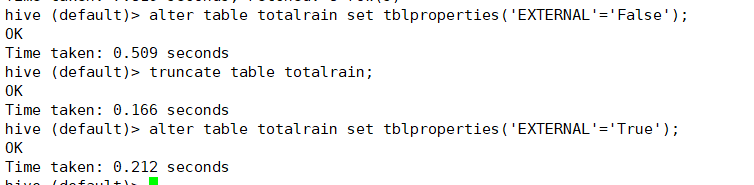
FIELDS TERMINATED BY '\t'

location 'hdfs://node01:8020/flume/record/2020-07-23/0926'

;

先把这个表的数据全部清除，但要保留表结构

Truncate只能清除内部表的数据，因此先要把外部表变为内部表



alter table totalrain set tblproperties(‘EXTERNAL’=’False’);

truncate table totalrain;

alter table totalrain set tblproperties(‘EXTERNAL’=’True’);

然后把降雨量信息统计到这个空表里，然后按城市累计降雨量从大到小排序。

insert into totalrain

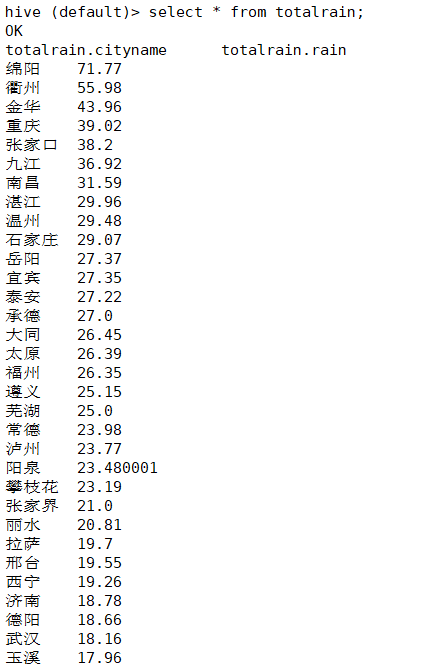
select city,sum(rain) as sumrain from weather0725

join city on city.city=weather0725.cityname

group by city.city order by sumrain desc

;

部分效果图：



导入mysql：

查看表的位置：

show create table t\_positions;

解决mysql的中文乱码问题：

进入配置文件修改配置内容，执行命令：vi /etc/my.cnf

修改配置文件的内容，在[mysqld]结束位置添加：character\_set\_server=utf8

重新启动mysql服务。

停止命令：systemctl stop mysqld.service

启动命令：systemctl start mysqld.service

建表

CREATE TABLE `temphigh` (

`cityname` varchar(255) DEFAULT NULL,

`citycode` varchar(255) DEFAULT NULL,

`date` varchar(255) DEFAULT NULL,

`week` varchar(255) DEFAULT NULL,

`weather` varchar(255) DEFAULT NULL,

`temphigh` int DEFAULT NULL,

`templow` int DEFAULT NULL,

`humidity` int DEFAULT NULL,

`pressure` int DEFAULT NULL,

`windspeed` float(255,2) DEFAULT NULL,

`winddirect` varchar(255) DEFAULT NULL,

`windpower` varchar(255) DEFAULT NULL,

`quality` varchar(255) DEFAULT NULL,

` rain` float(255,2) DEFAULT NULL

) ENGINE=InnoDB DEFAULT CHARSET=utf8;

CREATE TABLE `humidity` (

`cityname` varchar(255) DEFAULT NULL,

`citycode` varchar(255) DEFAULT NULL,

`date` varchar(255) DEFAULT NULL,

`week` varchar(255) DEFAULT NULL,

`weather` varchar(255) DEFAULT NULL,

`temphigh` int DEFAULT NULL,

`templow` int DEFAULT NULL,

`humidity` int DEFAULT NULL,

`pressure` int DEFAULT NULL,

`windspeed` float(255,2) DEFAULT NULL,

`winddirect` varchar(255) DEFAULT NULL,

`windpower` varchar(255) DEFAULT NULL,

`quality` varchar(255) DEFAULT NULL,

` rain` float(255,2) DEFAULT NULL

) ENGINE=InnoDB DEFAULT CHARSET=utf8;

排序用的临时表

CREATE TABLE `temp` (

`cityname` varchar(255) DEFAULT NULL,

`citycode` varchar(255) DEFAULT NULL,

`date` varchar(255) DEFAULT NULL,

`week` varchar(255) DEFAULT NULL,

`weather` varchar(255) DEFAULT NULL,

`temphigh` int DEFAULT NULL,

`templow` int DEFAULT NULL,

`humidity` int DEFAULT NULL,

`pressure` int DEFAULT NULL,

`windspeed` float(255,2) DEFAULT NULL,

`winddirect` varchar(255) DEFAULT NULL,

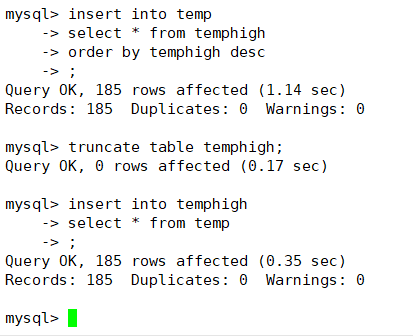
`windpower` varchar(255) DEFAULT NULL,

`quality` varchar(255) DEFAULT NULL,

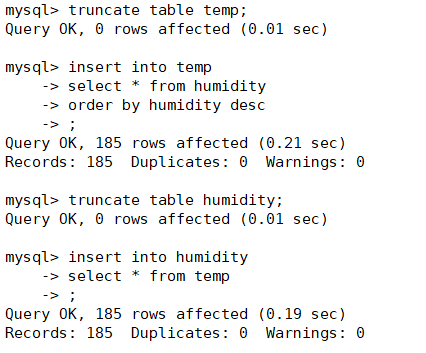
` rain` float(255,2) DEFAULT NULL

) ENGINE=InnoDB DEFAULT CHARSET=utf8;

把temphigh表中的乱序行再次排序：借用到一个临时表temp

.

对humidity表按humidity关键字重新排序，思路类似，同样借用到一个临时表temp



CREATE TABLE `totalrain` (

`cityname` varchar(255) DEFAULT NULL,

`rain` float(255,2) DEFAULT NULL

) ENGINE=InnoDB DEFAULT CHARSET=utf8;

CREATE TABLE `temp` (

`cityname` varchar(255) DEFAULT NULL,

`rain` float(255,2) DEFAULT NULL

) ENGINE=InnoDB DEFAULT CHARSET=utf8;

把temp中的降雨量无序数据存入totalrain表中：

insert into totalrain

select \* from temp

order by rain desc

;

统计空气质量：

创建一张表（创建后手动清空）记录单日天气情况

create table if not exists quality\_results (

qual STRING,

number INT

)ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

location 'hdfs://192.168.244.100:8020/sqoop/result/province/';

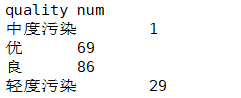
此表真正用来统计结果，需要提前清空

insert into quality\_results

select quality,sum(1) as num from weather0725

group by quality;

效果如下：



建mysql表：

CREATE TABLE `quality` (

`quality` varchar(255) DEFAULT NULL,

`number` int DEFAULT NULL

) ENGINE=InnoDB DEFAULT CHARSET=utf8;

将quality\_results表导入mysql。