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：

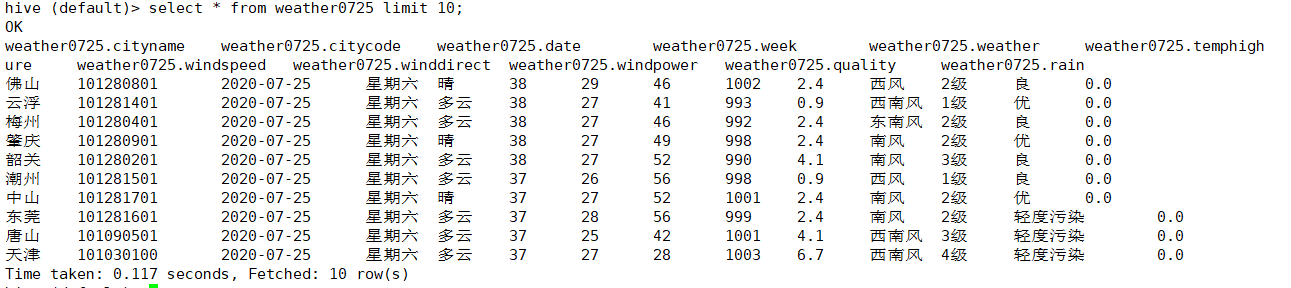
hive (default)> insert into temphigh

> 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

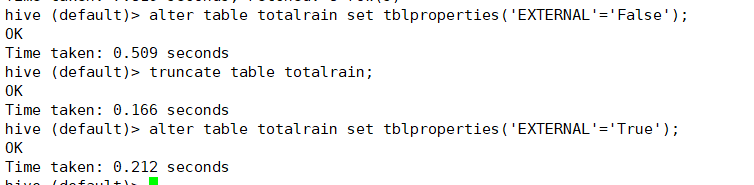
FIELDS TERMINATED BY '\t'

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：

