#任务1：最热门的国家签到排行（除中国）

cat weibo\_locate.2018-05/\* | cut -f5,8,9 | sort -u > temp\_user\_location.txt

rm -rf temp\_country.txt

touch temp\_country.txt

cat temp\_user\_location.txt | while read line

do echo $line

do echo $line1

lati1=$(echo $line1 | awk '{print $2}')

long1=$(echo $line1 | awk '{print $3}')

wget "http://api.map.baidu.com/geocoder?callback=renderReverse&location=$lati1,$long1&output=json&pois=1" -O temp\_location.txt -q

result1=$(grep cityCode temp\_location.txt | grep -Eo [0-9]+ )

if [ $result1 -ne 0 ]; then echo "China" >> temp\_country.txt

else wget "http://dev.virtualearth.net/REST/v1/Locations/$lati1,$long1?o=xml&key=AvwLdi5YcwldiXf5pWlES9OAEHKJvrE3sWPju56N\_-RO-1nIbehzyQh4xKHDkyGn" -O temp\_location.txt -q

result1=$(grep -Eo "<CountryRegion>[A-Za-z -]+</CountryRegion>" temp\_location.txt | grep -Eo ">[A-Za-z -]+<" | sed 's/<//g' | sed 's/>//g' | head -1)

echo $result1 >> temp\_country.txt

fi

done

paste -d"\t" temp\_user\_location.txt temp\_country.txt > country\_result.txt

cat country\_result.txt | cut -f1,4 | sort -u | cut -f2 | sort | uniq -c | sort -k1,1rn > temp\_result\_columns.txt

cat temp\_result\_columns.txt | sed -e 's/[^0-9]\* //'| cut -d' ' -f1 > temp\_result\_column1.txt

cat temp\_result\_columns.txt | awk '{$1="";print $0}' | sed 's/^ //g' > temp\_result\_column2.txt

paste -d"\t" temp\_result\_column2.txt temp\_result\_column1.txt > hw10.1\_result

sed -i " " '/China/d' hw10.1\_result

#任务2： 最热门的中国城市排行

grep China country\_result.txt > temp\_china\_users.txt

rm -rf temp\_city.txt

touch temp\_city.txt

cat temp\_china\_users.txt | while read line

do echo $line

lati=$(echo $line | awk '{print $2}')

long=$(echo $line | awk '{print $3}')

wget "http://api.map.baidu.com/geocoder?callback=renderReverse&location=$lati,$long&output=json&pois=1" -O temp\_china\_location.txt -q

result=$(grep "\"city\"" temp\_china\_location.txt | awk -F: '{print $2}' | sed 's/"//g' | sed 's/,//g')

echo $result >> temp\_city.txt

done

paste -d"\t" temp\_china\_users.txt temp\_city.txt > city\_result.txt

cat city\_result.txt | cut -f1,5 | sort -u | cut -f2 | sed '/^$/d'| sort | uniq -c | awk '{print $2"\t"$1}' | sort -k2,2rn > hw10.2\_ result

#任务3：最热门的中国省排名

grep China country\_result.txt > temp\_china\_users.txt

rm -rf temp\_province.txt

touch temp\_province.txt

cat temp\_china\_users.txt | while read line

do echo $line

lati=$(echo $line | awk '{print $2}')

long=$(echo $line | awk '{print $3}')

wget "http://api.map.baidu.com/geocoder?callback=renderReverse&location=$lati,$long&output=json&pois=1" -O temp\_china\_location.txt -q

result1=$(grep province temp\_china\_location.txt | grep 省 | awk -F: '{print $2}' | sed 's/"//g' | sed 's/,//g')

echo $result1 >> temp\_province.txt

done

paste -d"\t" temp\_china\_users.txt temp\_province.txt | cut -f1,5 | sort -u | cut -f2 | sed '/^$/d'| sort | uniq -c | awk '{print $2"\t"$1}' | sort -k2,2rn > hw10.3\_result

#任务4：微博用户签到城市数量排名和签到数量排名

cat weibo\_locate.2018-05/\* | cut -f4,5 | sort | uniq > temp\_user\_weibo.txt

cut -f2 temp\_user\_weibo.txt | sort | uniq -c | awk '{print $2"\t"$1}' > temp\_weibo\_count.txt

cut -f1,5 city\_result.txt | sort | uniq -c | awk '{print $2"\t"$1}' | sort -k2,2rn > user\_city\_sign.txt

join -t $'\t' user\_city\_sign.txt temp\_weibo\_count.txt > hw10.4\_result

#任务5：5月有跨国旅行的人在微博总用户中的比例

foreign=$(grep -v China country\_result.txt | cut -f1 | sort | uniq | wc -l)

total=$(cut -f1 temp\_user\_location.txt | sort | uniq | wc -l)

echo $foreign $total | awk '{print $1/$2\*100"%"}'

#任务6：跨国旅行的人喜欢什么手机

cut -f5,15 weibo\_freshdata.2018-05-17 | sort -u > temp\_user\_device.txt

grep -v China country\_result.txt | cut -f1 | uniq > temp\_non\_china\_users.txt

join temp\_non\_china\_users.txt temp\_user\_device.txt > temp\_foreign\_device.txt

rm -rf temp\_device\_merged.txt

touch temp\_device\_merged.txt

grep -i -e "iPhone" -e "iPad" -e "iPod" temp\_foreign\_device.txt | awk '{print $1" 苹果"}' >> temp\_device\_merged.txt

grep -i -e "HUAWEI" -e "华为" -e "荣耀" -e "麦芒" temp\_foreign\_device.txt | awk '{print $1" 华为"}' >> temp\_device\_merged.txt

grep -e "小米" -e "红米" temp\_foreign\_device.txt | awk '{print $1" 小米"}' >> temp\_device\_merged.txt

grep -i -e "OPPO" temp\_foreign\_device.txt | awk '{print $1" OPPO"}' >> temp\_device\_merged.txt

grep -i -e "VIVO" temp\_foreign\_device.txt | awk '{print $1" VIVO"}' >> temp\_device\_merged.txt

grep -i -v -e "iPhone" -e "iPad" -e "iPod" -e "HUAWEI" -e "华为" -e "荣耀" -e "麦芒" -e "小米" -e "红米" -e "OPPO" -e "VIVO" temp\_foreign\_device.txt >> temp\_device\_merged.txt

cat temp\_device\_merged.txt | sed 's/^[0-9]\*//g'| sort | uniq -c | sort -k1,1rn > temp\_phone\_columns.txt

cat temp\_phone\_columns.txt | sed -e 's/[^0-9]\* //'| cut -d' ' -f1 > temp\_phone\_column1.txt

cat temp\_phone\_columns.txt | awk '{$1="";print $0}' | sed 's/^ //g' > temp\_phone\_column2.txt

paste -d"\t" temp\_phone\_column2.txt temp\_phone\_column1.txt > hw10.6\_result