ECE4721J - Homework 4

Methods and Tools for Big Data

Kexuan Huang 518370910126

July 3, 2022

Ex.1 Reminders on database

1. Explain what is a Join operation, and describe its most common types.[1]

`JOIN` is an SQL clause used to query and access data from multiple tables, based on logical relationships between those tables Basically, we have 5 types of `JOIN`:

- `INNER JOIN`
- `LEFT OUTER JOIN`
- `RIGHT OUTER JOIN`
- `SELF JOIN`
- `CROSS JOIN`

2. What is an aggregate operation?[2]

An aggregation operation computes a single value from a collection of values. An example of an aggregation operation is calculating the average daily temperature from a month's worth of daily temperature values.

3. Write at least three advanced nested queries on the weather database.

For schema setup, please refer to `README.md`

[1] devart
[2] Microsoft Docs

3.1 Top 5 stations with highest daily average temperature

SQL:

```sql
SELECT station.s_name AS station, weather.w_value AS value
FROM weather
    INNER JOIN station ON station.s_id = weather.w_station
WHERE weather.w_type = 'TAVG'
    AND LENGTH(weather.w_value) > 0
ORDER BY CAST(weather.w_value AS INTEGER) DESC
LIMIT 5;
```

Output:

```
+--------------------------+-------+
|         station          | value |
+--------------------------+-------+
| ELK CREEK OREGON         | 572   |
| BEVERLY HILLS CALIFORNIA | 567   |
| BEVERLY HILLS CALIFORNIA | 544   |
| COLORADO CITY COLORADO   | 492   |
| ELK CREEK OREGON         | 466   |
+--------------------------+-------+
5 rows selected (3.581 seconds)
```

3.2 Top 5 station with lowest daily minimum temperature on Augest 25, 2017

SQL:

```sql
SELECT station.s_name AS station,
    weather.w_value AS value
FROM weather
    INNER JOIN station ON station.s_id = weather.w_station
WHERE weather.w_type = 'TMIN'
    AND LENGTH(weather.w_value) > 0
    AND weather.w_value <> -999
    AND weather.w_date = '20170825'
ORDER BY CAST(weather.w_value AS INTEGER)
LIMIT 5;
```

Output:

```
+---------------------------+-------+
|          station          | value |
+---------------------------+-------+
| VOSTOK                    | -750  |
| SAN ANTONIO INCARNATE WORD | -728  |
| PROGRESS                  | -362  |
| SYOWA                     | -329  |
| MIRNYJ                    | -324  |
+---------------------------+-------+
5 rows selected (3.691 seconds)
```

3.3 Top 5 date with highest average temperature in Shanghai

SQL:

```sql
SELECT country.c_name AS country,
    station.s_name AS station,
    weather.w_date AS day,
    weather.w_value AS value
FROM station
    INNER JOIN country ON SUBSTR(station.s_id, 1, 2) = country.
       c_fips
    INNER JOIN weather ON station.s_id = weather.w_station
WHERE station.s_name LIKE 'SHANGHAI%'
    AND weather.w_type = 'TAVG'
ORDER BY CAST(weather.w_value AS INTEGER) DESC
LIMIT 5;
```

Output:

```
+---------+-------------------+----------+-------+
| country |      station      |   day    | value |
+---------+-------------------+----------+-------+
| China   | SHANGHAI/HONGQIAO | 20170721 | 356   |
| China   | SHANGHAI/HONGQIAO | 20170724 | 354   |
| China   | SHANGHAI          | 20170724 | 353   |
| China   | SHANGHAI/HONGQIAO | 20170725 | 353   |
| China   | SHANGHAI/HONGQIAO | 20170720 | 351   |
+---------+-------------------+----------+-------+
5 rows selected (2.094 seconds)
```

Ex.2 Holidays!

1. Define what is “perfect weather” according to you. Express it in terms of precipitations, average temperature, and daily temperature amplitude.

Perfect weather for me:

1. Average temperature: 15°C ~ 25°C
2. Maximum temperature: 30°C
3. Minimum temperature: 10°C
4. Precipitation: 10% ~ 20%
5. Date: July and Auguest

2. Using Drill, with or without R, determine the perfect location of your next holidays.

SQL:

```
SELECT DISTINCT(country.c_name) AS country c,
    country.c_continent AS continent
FROM station
    INNER JOIN country ON SUBSTR(station.s_id, 1, 2) = country.
       c_fips
    INNER JOIN weather ON station.s_id = weather.w_station
WHERE (
        weather.w_date > 20170701 AND weather.w_date < 20170831
        AND (
            weather.w_type = 'TAVG'
            AND CAST(weather.w_value AS FLOAT) > 150
            AND CAST(weather.w_value AS FLOAT) < 300
        )
        OR (
            weather.w_type = 'TMAX'
            AND CAST(weather.w_value AS FLOAT) < 30
        )
        OR (
            weather.w_type = 'TIN'
            AND CAST(weather.w_value AS FLOAT) > 10
        )
        OR (
            weather.w_type = 'PRCP'
            AND CAST(weather.w_value AS FLOAT) > 10
            AND CAST(weather.w_value AS FLOAT) < 20
        )
    )
LIMIT 5;
```

Output:

```
+-----------+-----------+
|  country  | continent |
+-----------+-----------+
| Belize    | NA        |
| Fiji      | OC        |
| Greece    | EU        |
| India     | AS        |
| Indonesia | AS        |
+-----------+-----------+
5 rows selected (4.472 seconds)
```

Fiji looks good to me!

Ex.3 Data visualisation