Graded Exercise 2

CSC230 Database Technologies for Analytics

05 November 2021

These questions refer to documents on this courses' Moodle site:

- I have placed a document on Moodle titled *qa230-02nov2021.pdf.* It is in Week 3 in "Tuesday, November 2: Notes on data warehouses."

- I have placed a SQL script named *northwoods-exercise.sql* on Moodle in Week 3.

---

1. What does this query produce?

    (a) How many columns does the resulting table contain?

    (b) What kind of information do the columns contain?

    (c) Which values come first in these columns?

```
SELECT t.name, f.name FROM
    upfall AS f INNER JOIN trip AS t ON t.stop = f.id
    ORDER BY t.name, f.name;
```

    (a) 2 columns

    (b) The first column contains the names of trips and the second column contains the names of waterfalls.

(c) The rows are ordered first by the names of the trips and then, for each trip, the names of the waterfalls tourists will see on each of those trips.

2. Write a query that will produce a table that lists tours of the Upper Peninsula together with the number of waterfalls that travelers can see on each of the tours.

The query will produce this table:

Tour	Number of waterfalls
Munising	6
M-28	6
US-2	4

```
SELECT t.name AS 'Tour',
  COUNT(t.name) AS 'Number_of_waterfalls' FROM
  upfall AS f INNER JOIN trip AS t ON t.stop = f.id
  GROUP BY t.name;
```

3. Here is a query followed by the table it produces.

```
SELECT f.name, c.name
  FROM upfall f INNER JOIN county c ON f.county_id = c.id;
```

name	name
Munising Falls	Alger
Tannery Falls	Alger
Alger Falls	Alger
Wagner Falls	Alger
Horseshoe Falls	Alger
Miners Falls	Alger
Little Miners	Alger
Scott Falls	Alger
Twin Falls #1	Alger
Twin Falls #2	Alger
Canyon Falls	Baraga
Agate Falls	Ontonagon
Bond Falls	Ontonagon
Fumee Falls	Dickinson
Kakabika Falls	Gogebic
Rapid River Fls	Delta

Write a new query that produces this table:

Name of waterfall	Name of county
Alger Falls	Alger
Horseshoe Falls	Alger
Little Miners	Alger
Miners Falls	Alger
Munising Falls	Alger
Scott Falls	Alger
Tannery Falls	Alger
Twin Falls #1	Alger
Twin Falls #2	Alger
Wagner Falls	Alger
Canyon Falls	Baraga
Rapid River Fls	Delta
Fumee Falls	Dickinson
Kakabika Falls	Gogebic
Agate Falls	Ontonagon
Bond Falls	Ontonagon

```
SELECT f.name AS 'Name_of_waterfall', c.name AS 'Name_of_county'
  FROM upfall f INNER JOIN county c ON f.county_id = c.id
  ORDER BY c.name, f.name;
```

4. Write a query that produces the whole gov_unit table, with rows sorted first by type and then by name.

   (a) There is only one record in which the parent_id is **NULL**. What is the name and type in that record?

   (b) There are six records whose type is 'County'. What is the parent_id in those records?

   (c) There are three records whose parent_id is 2. These are cities or townships in which county?

   (a) It is Michigan State.

   (b) The parent_id is 3. This is the id of the Michigan State record.

   (c) These are cities and townships in Alger County (the row with id = 2.

5. Describe the table that this query produces.

```
SELECT a.name AS 'City/Township',
       b.name AS 'County',
       c.name AS 'State'
  FROM gov_unit a INNER JOIN gov_unit b
    ON a.parent_id = b.id INNER JOIN gov_unit c
    ON b.parent_id = c.id
  WHERE a.type IN ('City', 'Township') AND b.type = 'County'
  ORDER BY c.name, b.name, a.name;
```

The table has three columns. The first contains the names of cities or townships in ascending alphabetical order. The second and third columns contain the counties and state in which those towns are found.

6. A datum is a reference that cartographers use in making maps. It includes a description of the shape of the earth and point on the earth from which distances and directions to all other points can be computed.

   Write a query that produces a table that contains the name of each datum referenced in the *northwoods* database. The table should not contain **NULL** and should not repeat the name of any datum.

   You will have to determine which table or tables contain the relevant information.

```sql
SELECT DISTINCT datum FROM upfall
    WHERE datum IS NOT NULL;
```

7. Some of the descriptions of waterfalls are long. It might sometimes be convenient to look at just the first part of a description.

   Write a query that produces a result that contains two columns. The first column will contain the names of waterfalls in alphabetical order. The second column contains that first N words (or first N characters) of the waterfalls' descriptions.

   Look online for documentation for MySQL's string functions. You might find the **LEFT**(), **SUBSTRING**(), or **SUBSTRING_INDEX**() functions helpful.

```sql
SELECT name, SUBSTRING_INDEX(description, ' ', 8) AS 'Description'
    FROM upfall ORDER BY name;
```

8. The letters in each of these groups are the first letters of which words?

   (a) EDW
   (b) ETL (also called ELT)
   (c) OLAP
   (d) OLTP

EDW Enterprise Data Warehouse

ETL/ELT Extract, Transform, Load

OLAP OnLine Analytical Processing

OLTP OnLine Transactional Processing

9. We read an article from IBM about data warehouses. That article told us that companies choose to use cloud data warehouses "to reduce their on-premises data center footprint."

   What does that mean?

   You might find help here in the section in the article from Oracle about reasons for choosing a cloud data warehouse.

   A company that chooses to use resources in the cloud frees itself from the costs of owning and operating its own computing hardware and software, and the costs of acquiring and maintaining the expertise that it would need to manage that hardware and software.

   Effectively, the company rents just as much as it needs. When the company's needs grow, it rents more.

   The company does not have to come up with as money up front. The company reduces its financial risk.

10. Search through the articles about data warehouses from IBM and Oracle for the word "data mining" until you find a sentence that defines that term.

   You can search in the PDF document by typing Control-F.

   What is data mining?

   According to the authors of the articles that we read, data mining is "finding unseen patterns and relationships in data."

11. What is an advantage of a star schema over a snowflake schema?

_____

It is a simple schema. It allows fast queries.

12. Both the IBM and Oracle articles contain sections titled "Benefits of a data warehouse."

Make a case for using a data warehouse in two or three sentences.

Do not try to list all of the benefits. Read what the experts at IBM and Oracle think. Exercise your own judgment. Be selective. Choose what you think is most important.