Analytical Queries Report

Yisong Cheng; Huajing Lu

2023-04-19

Introduction

This report presents the results of two analytically queries performed on the MySQL data warehouse from Part 2. The queries were designed to provide insights on the journals with the most articles published and the number of articles per journal per year, broken down by quarter.

Set up connection

```
# 1. Library (must be installed prior to loading
library(RMySQL)
library(DBI)### MySQL

# 2. Initial settings
db_user <- 'root'
db_password <- 'cerf0312*'
db_name <- 'Journal'

db_host <- 'localhost'
db_port <- 3306 # always this port unless you change it during installation

# 3. Connect to the database
mydb <-  dbConnect(MySQL(), user = db_user, password = db_password,
                   dbname = db_name, host = db_host,
                   port = db_port, local_infile = TRUE)
```

Query I: Top Five Journals with the Most Articles Published

The top five journals with the most articles published in them for the time period are as follows:

Code part

```
SELECT
    Title, year, num_articles
FROM
    journal_facts
WHERE year <= '1999' and year >= '1977'
ORDER BY num_articles
DESC LIMIT 5
```

Table 1: 5 records

Title	year	num_articles
Acta haematologica	1977	9
Nihon Ketsueki Gakkai zasshi : journal of Japan Haematological Society	1977	7
AORN journal	1977	6
Izvestiia Akademii nauk SSSR. Seriia biologicheskaia	1977	4
Acta medica Polona	1977	3

Query II: Number of Articles per Journal per Year, Broken Down by Quarter

The following table shows the number of articles per journal per year, broken down by quarter:

Code part

```
SELECT
    year, quarter, Title, num_articles
FROM journal_facts
ORDER BY year, quarter
```

Table 2: Displaying records 1 - 10

year	quarter	Title	num_articles
1975	NA	Journal of the American Chemical Society	47
1975	NA	American journal of obstetrics and gynecology	51
1975	NA	Doklady Akademii nauk SSSR	20
1975	NA	Acta tropica	103
1975	NA	Acta poloniae pharmaceutica	15
1975	NA	Acta physiologica latino americana	12
1975	NA	Acta anatomica	4
1975	NA	Acta biochimica Polonica	16
1975	NA	Acta biologica Academiae Scientiarum Hungaricae	5
1975	NA	Acta biologica et medica Germanica	61

Table > Visiualizaiton

In this case, I suppose that a table is better than a visualization.

1) Detail: A table can present a lot more detail than a visualization can. In this case, we need to display multiple metrics or dimensions, a table can show all of them in a compact form. With a visualization, it may need multiple charts or graphs to show the same amount of detail.

2) Precision: A table can provide exact values, while a visualization can only provide estimates. This can be important in certain situations where exact numbers are critical.

3) Comparison: A table can make it easier to compare values between different rows or columns. This is especially true when the table has been sorted or filtered in a specific way. With a visualization, it may be more difficult to make precise comparisons.

```
dbDisconnect(mydb)
```

```
## [1] TRUE
```