

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

ЗВІТ

З ЛАБОРАТОРНОЇ РОБОТИ №2

З ДИСЦИПЛІНИ: « ПРОГРАМУВАННЯ ІНТЕЛЕКТУАЛЬНИХ
ІНФОРМАЦІЙНИХ СИСТЕМ»

Виконав:
ІП-15 Мєшков А.І.

Перевірів:
Курченко О.А.

Київ 2023

ЗАВДАННЯ

1. Створити стовпчикову і звичайну бд
2. Розрахувати сумарну затримку по містах
3. Порахувати кількість польотів по містах
4. Знайти місто з найменшою і найбільшою затримкою
5. Знайти всі польоти з затримкою більше за середній час затримки
6. Заміряти вбудованими методами об'єм БД та швидкість виконання запитів. Порівняти звичайну і стовпчикову

ХІД РОБОТИ

1. Створити стовпчикову і звичайну бд

- Встановлюємо Docker Desktop.
- Встановлюємо базу даних за допомогою команд:
 - o `docker run -d -p 3307:3306 --shm-size=512m -e PM1=mcs1 --hostname=mcs1 --name mcs1 mariadb/columnstore`
 - o `docker exec -it mcs1 provision mcs1`
- Клонуємо репозиторій з даними за допомогою команд:
 - o `cd /tmp`
 - o `git clone https://github.com/mariadb-corporation/mariadb-columnstore-sample-data.git`
 - o `cd mariadb-columnstore-sample-data`
- Додаємо дані польотів(flights.csv).
- Запускаємо проєкт:
 - o `./run_project`

2. Розрахувати сумарну затримку по містах

- Для стовпчикової бази даних виконаємо команду:
 - `USE columnstore_bts; SELECT dest, SUM(arr_delay) AS sum_delay FROM flights GROUP BY dest ORDER BY sum_delay DESC;`
- Для звичайної бази даних виконаємо команду:
 - `USE innodb_bts; SELECT dest, SUM(arr_delay) AS sum_delay FROM flights GROUP BY dest ORDER BY sum_delay DESC;`

Результат виконання запитів:

Для стовпчикової бд

ADQ	50.00
ITH	41.00
PPG	-16.00
EKO	-80.00
IMT	-110.00
SCC	-151.00
ADK	-293.00
CDV	-339.00
RKS	-351.00
HLN	-357.00
BRD	-371.00
INL	-416.00
BJI	-416.00
YAK	-419.00
ABR	-525.00
BTM	-601.00
MSO	-763.00
FAI	-1128.00
VEL	-1128.00
LWS	-1221.00
CNY	-1596.00
BLI	-1664.00
ANC	-6346.00
+-----+-----+	
315 rows in set (0.285 sec)	

Для звичайної бд

ADQ	50.00
ITH	41.00
PPG	-16.00
EKO	-80.00
IMT	-110.00
SCC	-151.00
ADK	-293.00
CDV	-339.00
RKS	-351.00
HLN	-357.00
BRD	-371.00
BJI	-416.00
INL	-416.00
YAK	-419.00
ABR	-525.00
BTM	-601.00
MSO	-763.00
VEL	-1128.00
FAI	-1128.00
LWS	-1221.00
CNY	-1596.00
BLI	-1664.00
ANC	-6346.00
+-----+-----+	
315 rows in set (3.094 sec)	

3. Порахувати кількість польотів по містах

- Для стовпчикової бази даних виконаємо команду:
 - `USE columnstore_bts; SELECT dest, COUNT(*) AS flight_count FROM flights GROUP BY dest ORDER BY flight_count DESC;`
- Для звичайної бази даних виконаємо команду:
 - `USE innodb_bts; SELECT dest, COUNT(*) AS flight_count FROM flights GROUP BY dest ORDER BY flight_count DESC;`

Результат виконання запитів:

Для стовпчикової бд

VEL	101
APN	101
ESC	101
PUB	101
HYS	101
CDC	101
INL	101
HOB	96
DVL	92
STC	82
MQT	77
MMH	77
ADQ	67
GUM	65
PBG	59
ILG	46
BGM	38
OTH	33
UST	29
BGR	22
ADK	21
PPG	18
ITH	6
+-----+-----+	
315 rows in set (0.223 sec)	

Для звичайної бд

HYS	101
APN	101
INL	101
ESC	101
VEL	101
CDC	101
PUB	101
HOB	96
DVL	92
STC	82
MQT	77
MMH	77
ADQ	67
GUM	65
PBG	59
ILG	46
BGM	38
OTH	33
UST	29
BGR	22
ADK	21
PPG	18
ITH	6
+-----+-----+	
315 rows in set (0.647 sec)	

4. Знайти місто з найменшою і найбільшою затримкою

- Для стовпчикової бази даних виконаємо команди:

```
o USE columnstore_bts;
  SELECT dest, MIN(arr_delay) AS min_delay FROM flights GROUP BY dest
  ORDER BY min_delay ASC LIMIT 1;
  SELECT dest, MAX(arr_delay) AS max_delay FROM flights GROUP BY dest
  ORDER BY max_delay ASC LIMIT 1;
```

- Для звичайної бази даних виконаємо команду:

```
o USE innodb_bts;
  SELECT dest, MIN(arr_delay) AS min_delay FROM flights GROUP BY dest
  ORDER BY min_delay ASC LIMIT 1;
  SELECT dest, MAX(arr_delay) AS max_delay FROM flights GROUP BY dest
  ORDER BY max_delay ASC LIMIT 1;
```

Результат виконання запитів:

Для стовпчикової бд

```
+-----+-----+
| dest | min_delay |
+-----+-----+
| SEA  |    -82.00 |
+-----+-----+
1 row in set (0.253 sec)
```

```
+-----+-----+
| dest | max_delay |
+-----+-----+
| VEL  |     19.00 |
+-----+-----+
1 row in set (0.226 sec)
```

Для звичайної бд

```
+-----+-----+
| dest | min_delay |
+-----+-----+
| SEA  |    -82.00 |
+-----+-----+
1 row in set (2.640 sec)
```

```
+-----+-----+
| dest | max_delay |
+-----+-----+
| VEL  |     19.00 |
+-----+-----+
1 row in set (2.095 sec)
```

5. Знайти всі польоти з затримкою більше за середній час затримки

- Для стовпчикової бази даних виконаємо команду:
 - ```
USE columnstore_bts;
SELECT COUNT(*) FROM flights WHERE dep_delay > (SELECT AVG(dep_delay)
FROM flights WHERE dep_delay IS NOT NULL);
```
- Для звичайної бази даних виконаємо команду:
  - ```
USE innodb_bts;  
SELECT COUNT(*) FROM flights WHERE dep_delay > (SELECT AVG(dep_delay)  
FROM flights WHERE dep_delay IS NOT NULL );
```

Результат виконання запитів:

Для стовпчикової бд

```
+-----+  
| COUNT(*) |  
+-----+  
|   247480 |  
+-----+  
1 row in set (0.299 sec)
```

Для звичайної бд

```
+-----+  
| COUNT(*) |  
+-----+  
|   247480 |  
+-----+  
1 row in set (3.118 sec)
```

6. Заміряти вбудованими методами об'єм БД та швидкість виконання запитів. Порівняти звичайну і стовпчикову

- Для заміру виконаємо команду:

```
SELECT table_schema 'Database Name', ROUND (SUM(data_length +  
index_length) / 1024/1024, 2) 'Size in MB' FROM  
information_schema.tables GROUP BY table_schema;
```

Результат виконання запитів:

Database Name	Size in MB
calpontsys	0.00
columnstore_bts	0.00
infinidb_querystats	0.05
information_schema	0.22
innodb_bts	0.23
mysql	11.23
performance_schema	0.00
sys	0.03

- Порівняння швидкості:

№ Завдання	Стовпчикова, сек	Звичайна, сек
2	0.284	3.094
3	0.223	0.647
4.1	0.253	2.640
4.2	0.226	2.095
5	0.299	3.118