

Отчет по ДЗ Мутации данных и манипуляции с партициями

Задачи:

1. Создание таблицы:
Создайте таблицу `user_activity` с полями:
`user_id` (UInt32) — идентификатор пользователя
`activity_type` (String) — тип активности (например, 'login', 'logout', 'purchase')
`activity_date` (DateTime) — дата и время активности
Используйте MergeTree как движок таблицы и настройте партиционирование по дате активности (`activity_date`).
2. Заполнение таблицы:
Вставьте несколько записей в таблицу `user_activity`. Используйте различные `user_id`, `activity_type` и `activity_date`.
3. Выполнение мутаций:
Выполните мутацию для изменения типа активности у пользователя(-ей)
4. Проверка результатов:
Напишите запрос для проверки изменений в таблице `user_activity`.
Убедитесь, что тип активности у пользователей изменился. Приложите логи отслеживания мутаций в системной таблице.
5. Манипуляции с партициями:
Удалите партицию за определённый месяц. Проверьте текущее состояние таблицы после удаления партиции. Убедитесь, что данные за указанный месяц были удалены.
6. Исследуйте, как работают другие типы мутаций. Попробуйте создать новую партицию и вставить в неё данные.
7. Изучите возможность использования TTL (Time to Live) для автоматического удаления старых партиций

1. Создаем таблицу согласно ТЗ:

```
5c26017c72da :) create table user_activity ON CLUSTER '{cluster}'
(
    user_id UInt32,
    activity_type String,
    activity_date DateTime
)
ENGINE = ReplicatedMergeTree
order by (user_id, activity_date)
PARTITION BY toYYYYMM(activity_date);
```

```
CREATE TABLE user_activity ON CLUSTER '{cluster}'
(
    `user_id` UInt32,
    `activity_type` String,
    `activity_date` DateTime
)
ENGINE = ReplicatedMergeTree
PARTITION BY toYYYYMM(activity_date)
ORDER BY (user_id, activity_date)
```

Query id: 816ab15f-bd3e-4389-8907-ff929fc47d53

	host	port	status	error	num_hosts_remaining	num_hosts_active
1.	shard2_replica1	9000	0		3	0
2.	shard1_replica2	9000	0		2	0
3.	shard1_replica1	9000	0		1	0
4.	shard2_replica2	9000	0		0	0

4 rows in set. Elapsed: 0.173 sec.

5c26017c72da :) █

2. Заполняем таблицу случайными данными:

```
5c26017c72da :) INSERT INTO user_activity (user_id, activity_type, activity_date) VALUES
(1, 'login', '2023-10-01 10:00:00'),
(2, 'logout', '2023-10-01 11:00:00'),
(1, 'purchase', '2023-10-02 12:00:00'),
(3, 'login', '2023-10-03 13:00:00'),
(2, 'purchase', '2023-10-04 14:00:00'),
(1, 'logout', '2023-10-05 15:00:00');
```

```
INSERT INTO user_activity (user_id, activity_type, activity_date) FORMAT Values
```

Query id: 5b70e71c-3d70-4248-ab0c-24c0618223db

Ok.

6 rows in set. Elapsed: 0.011 sec.

5c26017c72da :) █

3. Выполняем апдейт данных. Заменяем 'logout' на 'exit':

```
5c26017c72da :) ALTER TABLE user_activity
UPDATE activity_type = 'exit'
WHERE activity_type = 'logout';

ALTER TABLE user_activity
  (UPDATE activity_type = 'exit' WHERE activity_type = 'logout')

Query id: c9b78479-6c7b-48f1-aa7c-bf7a49ea62ef

Ok.

0 rows in set. Elapsed: 0.019 sec.

5c26017c72da :) █
```

4. Проверяем что получилось:

```
SELECT *
FROM user_activity

Query id: e5c41f40-6220-4c4c-a5f0-3f5822480e06
```

	user_id	activity_type	activity_date
1.	1	login	2023-10-01 10:00:00
2.	1	purchase	2023-10-02 12:00:00
3.	1	exit	2023-10-05 15:00:00
4.	2	exit	2023-10-01 11:00:00
5.	2	purchase	2023-10-04 14:00:00
6.	3	login	2023-10-03 13:00:00

6 rows in set. Elapsed: 0.002 sec.

```
5c26017c72da :) █
```

4.1. Логи мутации из system.mutations. Видим что она завершилась.

```
5c26017c72da :) select * from system.mutations limit 1 format Vertical;

SELECT *
FROM system.mutations
LIMIT 1
FORMAT Vertical

Query id: 14726b5c-a9df-4390-9bd5-e67d2ee8e684

Row 1:
-----
database:                default
table:                   user_activity
mutation_id:              0000000000
command:                 (UPDATE activity_type = 'exit' WHERE activity_type = 'logout')
create_time:              2025-03-18 18:09:13
block_numbers.partition_id: ['202310']
block_numbers.number:     [1]
parts_to_do_names:        []
parts_to_do:              0
is_done:                  1
is_killed:                0
latest_failed_part:
latest_fail_time:         1970-01-01 00:00:00
latest_fail_reason:
latest_fail_error_code_name:

1 row in set. Elapsed: 0.003 sec.

5c26017c72da :)
```

5. Удаляем партицию за определённый месяц и проверяем результат. Так как у нас была всего одна партиция, таблица стала пустой

```
5c26017c72da :) ALTER TABLE user_activity DROP PARTITION 202310;
```

```
ALTER TABLE user_activity  
(DROP PARTITION 202310)
```

```
Query id: 565ab8ad-5fc9-43a3-9d0a-0c91518d868e
```

```
Ok.
```

```
0 rows in set. Elapsed: 0.060 sec.
```

```
5c26017c72da :) SELECT * FROM user_activity;
```

```
SELECT *  
FROM user_activity
```

```
Query id: eb22e02b-7b38-41f4-b685-b59d8028e47a
```

```
Ok.
```

```
0 rows in set. Elapsed: 0.001 sec.
```

```
5c26017c72da :) SELECT count() FROM user_activity;
```

```
SELECT count()  
FROM user_activity
```

```
Query id: e6b7c46c-083d-412a-a1c0-8cf893536edf
```

```
1. count()  
    0
```

```
1 row in set. Elapsed: 0.002 sec.
```

```
5c26017c72da :) █
```

6. Исследуем другие мутации. Заполняем таблицу данными за разные месяцы и еще раз проворачиваем операцию с удалением партии. Видим что удалилась

```
5c26017c72da :) SELECT * FROM user_activity;

SELECT *
FROM user_activity

Query id: 0eeb1edf-a085-40bc-a397-13b3f2ef81b4

+----+-----+-----+
| user_id | activity_type | activity_date |
+----+-----+-----+
1.      | 1            | login         | 2023-10-01 10:00:00 |
2.      | 1            | purchase      | 2023-10-02 12:00:00 |
3.      | 1            | logout        | 2023-10-05 15:00:00 |
4.      | 2            | logout        | 2023-10-01 11:00:00 |
5.      | 2            | purchase      | 2023-10-04 14:00:00 |
6.      | 3            | login         | 2023-10-03 13:00:00 |
7.      | 1            | login         | 2023-11-01 10:00:00 |
8.      | 1            | purchase      | 2023-11-02 12:00:00 |
9.      | 1            | logout        | 2023-11-05 15:00:00 |
10.     | 2            | logout        | 2023-11-01 11:00:00 |
11.     | 2            | purchase      | 2023-11-04 14:00:00 |
12.     | 3            | login         | 2023-11-03 13:00:00 |
+----+-----+-----+

12 rows in set. Elapsed: 0.003 sec.

5c26017c72da :) ALTER TABLE user_activity DROP PARTITION 202310;

ALTER TABLE user_activity
(DROP PARTITION 202310)

Query id: 9eac8745-7998-42dc-ac4e-6f0a1cfb0f96

Ok.

0 rows in set. Elapsed: 0.025 sec.

5c26017c72da :) SELECT * FROM user_activity;

SELECT *
FROM user_activity

Query id: bc8a65c7-9721-492c-b0b0-2f4498699b61

+----+-----+-----+
| user_id | activity_type | activity_date |
+----+-----+-----+
1.      | 1            | login         | 2023-11-01 10:00:00 |
2.      | 1            | purchase      | 2023-11-02 12:00:00 |
3.      | 1            | logout        | 2023-11-05 15:00:00 |
4.      | 2            | logout        | 2023-11-01 11:00:00 |
5.      | 2            | purchase      | 2023-11-04 14:00:00 |
6.      | 3            | login         | 2023-11-03 13:00:00 |
+----+-----+-----+

6 rows in set. Elapsed: 0.002 sec.

5c26017c72da :) █
```

одна партиция:

7. Посмотрим как работает TTL. Добавляю TTL таблице. Хочу, что бы она хранила данные за три года.

```
5c26017c72da :) ALTER TABLE user_activity on cluster '{cluster}' MODIFY TTL toDate(activity_date) + toIntervalYear(3);

ALTER TABLE user_activity ON CLUSTER '{cluster}'
(MODIFY TTL toDate(activity_date) + toIntervalYear(3))

Query id: 8c1953cb-79ff-432e-bbd4-0349d1e50d97
```

	host	port	status	error	num_hosts_remaining	num_hosts_active
1.	shard2_replica1	9000	0		3	2
2.	shard2_replica2	9000	0		2	2

	host	port	status	error	num_hosts_remaining	num_hosts_active
3.	shard1_replica2	9000	0		1	0
4.	shard1_replica1	9000	0		0	0

4 rows in set. Elapsed: 1.396 sec.

5c26017c72da :) █

- 7.1. Делаем оптимайз и смотрим на результат. Данные в таблице остались:

```
5c26017c72da :) optimize table user_activity final;

OPTIMIZE TABLE user_activity FINAL

Query id: 1400ba40-fb40-41b1-a2b6-1a023bdd8286

Ok.

0 rows in set. Elapsed: 0.018 sec.

5c26017c72da :) SELECT * FROM user_activity;

SELECT *
FROM user_activity

Query id: d9f2f70d-5e4b-4873-bca2-70600b56241c
```

	user_id	activity_type	activity_date
1.	1	login	2023-11-01 10:00:00
2.	1	purchase	2023-11-02 12:00:00
3.	1	logout	2023-11-05 15:00:00
4.	2	logout	2023-11-01 11:00:00
5.	2	purchase	2023-11-04 14:00:00
6.	3	login	2023-11-03 13:00:00

6 rows in set. Elapsed: 0.002 sec.

5c26017c72da :) █

7.2. Меняю TTL на 1 год:

```
5c26017c72da :) ALTER TABLE user_activity ON CLUSTER '{cluster}' MODIFY TTL toDate(activity_date) + toIntervalYear(1);

ALTER TABLE user_activity ON CLUSTER '{cluster}'
(MODIFY TTL toDate(activity_date) + toIntervalYear(1))

Query id: 0df18ec6-71d8-4bd7-9d6a-67b0db7388b8
```

	host	port	status	error	num_hosts_remaining	num_hosts_active
1.	shard2_replica1	9000	0		3	2
2.	shard2_replica2	9000	0		2	2

	host	port	status	error	num_hosts_remaining	num_hosts_active
3.	shard1_replica2	9000	0		1	0
4.	shard1_replica1	9000	0		0	0

4 rows in set. Elapsed: 1.237 sec.

```
5c26017c72da :)
```

7.3. Делаем оптимайз и видим, что TTL сработал и из таблицы удалились все данные, которые старше 1 года (т.е. 2024 года):

```
5c26017c72da :) optimize table user_activity final;

OPTIMIZE TABLE user_activity FINAL

Query id: aa21ae97-30cc-4c98-88c2-0beb82b09ca2

Ok.

0 rows in set. Elapsed: 0.037 sec.

5c26017c72da :) SELECT * FROM user_activity;

SELECT *
FROM user_activity

Query id: 41196aca-7fb3-421a-a394-4ef1a0d1f32f

Ok.

0 rows in set. Elapsed: 0.002 sec.

5c26017c72da :) SELECT count(*) FROM user_activity;

SELECT count(*)
FROM user_activity

Query id: 5ad1ec25-d10c-4970-a72f-420bb550e7ee
```

	count()
1.	0

1 row in set. Elapsed: 0.004 sec.

```
5c26017c72da :) █
```