

11 ДЗ - Мутация данных и манипуляции с партициями.

Собираем Dockerfile.

clickhouse-25.2.1.Dockerfile:

```
FROM clickhouse/clickhouse-server:25.2.1
MAINTAINER Maksim Kulikov <max.uoles@rambler.ru>

RUN apt-get update -y --fix-missing
RUN DEBIAN_FRONTEND=noninteractive apt-get -yq upgrade
RUN apt-get install nano mc python3 pip kafkacat -y
RUN pip install clickhouse_driver

EXPOSE 8123 9000

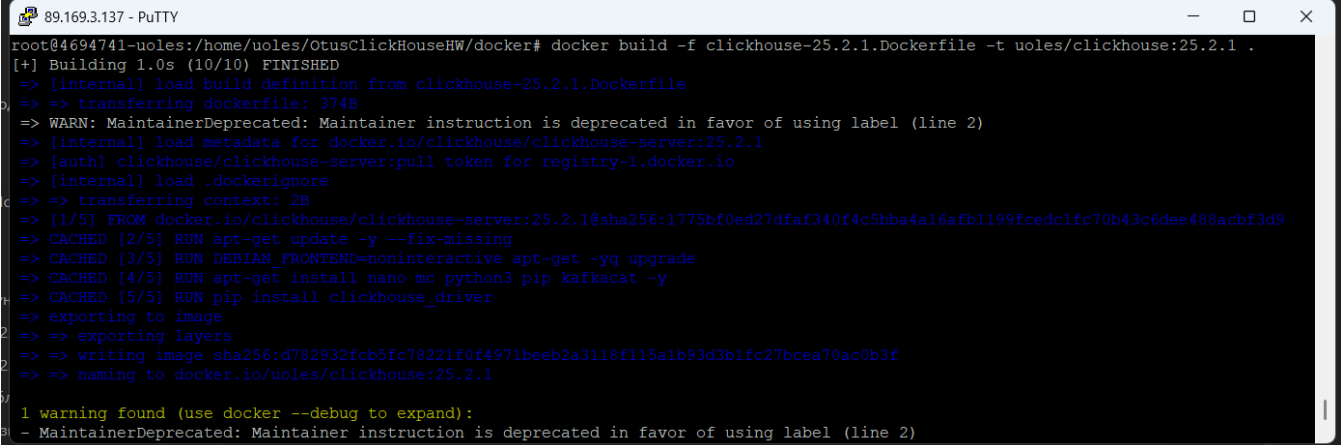
ENTRYPOINT ["/entrypoint.sh"]
```

Собираем образ командой:

```
docker build -f clickhouse-25.2.1.Dockerfile -t uoles/clickhouse:25.2.1 .
```

Запускаем:

```
docker run -d \
-p 18123:8123 \
-p 19000:9000 \
-e CLICKHOUSE_DB=my_database \
-e CLICKHOUSE_USER=username \
-e CLICKHOUSE_DEFAULT_ACCESS_MANAGEMENT=1 \
-e CLICKHOUSE_PASSWORD=password \
--name clickhouse-server \
--ulimit nofile=262144:262144 \
uoles/clickhouse:25.2.1
```



```
89.169.3.137 - PuTTY
root@4694741-uoles:/home/uoles/OtusClickHouseHW/docker# docker build -f clickhouse-25.2.1.Dockerfile -t uoles/clickhouse:25.2.1 .
[+] Building 1.0s (10/10) FINISHED
=> [internal] load build definition from clickhouse-25.2.1.Dockerfile
=> => transferring dockerfile: 374B
=> WARN: MaintainerDeprecated: Maintainer instruction is deprecated in favor of using label (line 2)
=> [internal] load metadata for docker.io/clickhouse/clickhouse-server:25.2.1
=> [auth] clickhouse/clickhouse-server:pull token for registry-1.docker.io
=> [internal] load .dockerignore
=> => transferring context: 2B
=> [1/5] FROM docker.io/clickhouse/clickhouse-server:25.2.1@sha256:1775bf0ed27dfaf340f4c5bba4a16afb1199fcedc1fc70b43c6dee488acbf3d9
=> CACHED [2/5] RUN apt-get update -y --fix-missing
=> CACHED [3/5] RUN DEBIAN_FRONTEND=noninteractive apt-get -yq upgrade
=> CACHED [4/5] RUN apt-get install nano mc python3 pip kafkacat -y
=> CACHED [5/5] RUN pip install clickhouse_driver
=> exporting to image
=> => exporting layers
=> => writing image sha256:d782932fcb5fc78221f0f4971beeb2a3118f115alb93d3b1fc27bcea70ac0b3f
=> => naming to docker.io/uoles/clickhouse:25.2.1

1 warning found (use docker --debug to expand):
- MaintainerDeprecated: Maintainer instruction is deprecated in favor of using label (line 2)
```

Создание пользователя и роли.

Создаем таблицу user_activity:

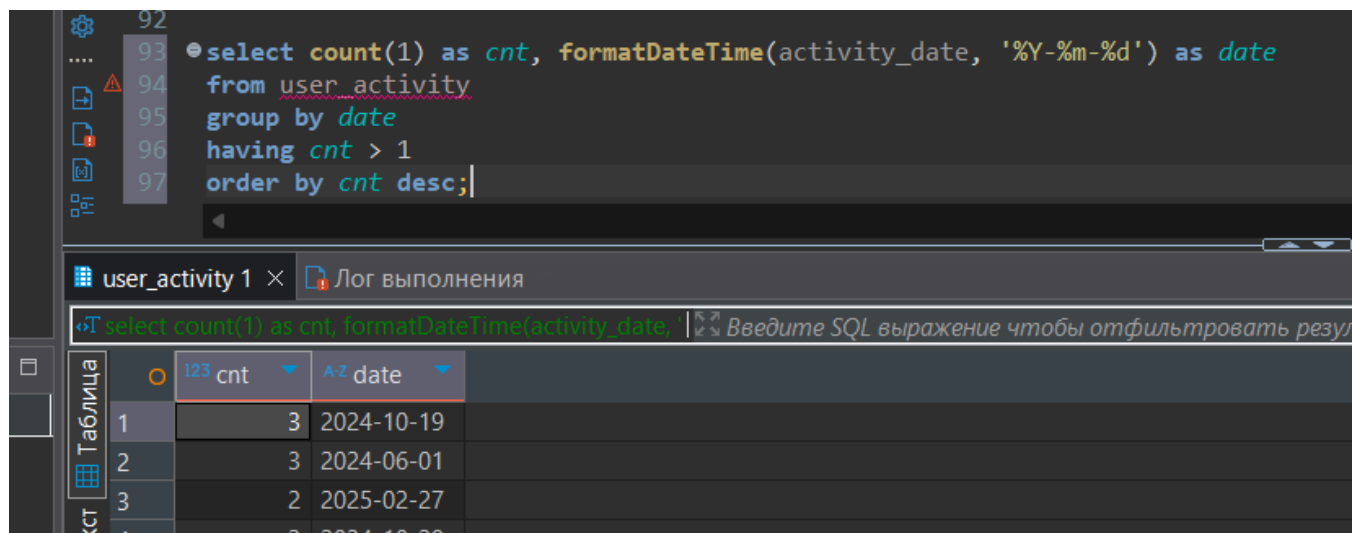
```
CREATE TABLE user_activity
(
    activity_date DateTime,
    user_id UInt32,
    activity_type LowCardinality(String)
)
ENGINE = MergeTree
PARTITION BY toYYYYMMDD(activity_date)
ORDER BY (activity_date);
```

Заполняем таблицу данными:

```
INSERT INTO user_activity
SELECT
    now() - toIntervalSecond(rand() % (365 * 24 * 60 * 60)) AS activity_date,
    randUniform(1,1000)::Int AS user_id,
    arrayElement(['login', 'logout', 'edit', 'delete'], randUniform(1,5)::Int) AS
activity_type
FROM system.numbers
LIMIT 100;
```

Ищем данные для изменения:

```
select count(1) as cnt, formatDateTime(activity_date, '%Y-%m-%d') as date
from user_activity
group by date
having cnt > 1
order by cnt desc;
```



The screenshot shows a database IDE interface. The top pane displays a SQL query: `select count(1) as cnt, formatDateTime(activity_date, '%Y-%m-%d') as date from user_activity group by date having cnt > 1 order by cnt desc;`. The bottom pane shows the results of the query in a table format. The table has two columns: 'cnt' and 'date'. The results are as follows:

	cnt	date
1	3	2024-10-19
2	3	2024-06-01
3	2	2025-02-27
4	2	2024-10-20

Выбираем дату и вносим изменения:

99

100

101

102

103

```
select *
from user_activity
where formatDateTime(activity_date, '%Y-%m-%d') = '2024-06-01'
```

user_activity 1 ×

Лог выполнения

select * from user_activity where formatDateTime(act

Введите SQL выражение чтобы отфи

Таблица

Текст

	activity_date	123 user_id	A-Z activity_type
1	2024-06-01 10:38:39	585	edit
2	2024-06-01 14:32:39	769	edit
3	2024-06-01 23:37:28	161	login

Обновляем поле activity_type:

```
alter table user_activity
update activity_type = 'edit'
where formatDateTime(activity_date, '%Y-%m-%d') = '2024-06-01'
and activity_type = 'login'
```

98

99

100

101

102

103

```
select *
from user_activity
where formatDateTime(activity_date, '%Y-%m-%d') = '2024-06-01'
```

user_activity 1 ×

Лог выполнения

select * from user_activity where formatDateTime(act

Введите SQL выражение чтобы отфильт

Таблица

Текст

	activity_date	123 user_id	A-Z activity_type
1	2024-06-01 10:38:39	585	edit
2	2024-06-01 14:32:39	769	edit
3	2024-06-01 23:37:28	161	edit

Проверяем наличие мутаций:

55

56

57

58

59

```
select command, is_done
from system.mutations
where table = 'user_activity'
```

mutations 1 ×

Лог выполнения

select command, is_done from system.mutations whe

Введите SQL выражение чтобы отфильтровать результаты

Таблица

	A-Z command	123 is_done
1	(UPDATE activity_type = 'edit' WHERE (formatDateTime(activity_date, '%Y-%m-%d') = '2024-06-01') AND (activity_type = 'login'))	1

Удаляем партиции за определенный месяц.

Выбираем данные для удаления:

114

115

116

117

118

119

```
select *
from system.parts
where table = 'user_activity' and partition like '202405%'
```

parts 1

Лог выполнения

select * from system.parts where table = 'user_activity'

Введите SQL выражение чтобы отфильтровать результаты

	partition	name	uuid	part_type
1	20240511	20240511_45_45_0_82	00000000-0000-0000-0000-000000000000	Compact
2	20240514	20240514_40_40_0_82	00000000-0000-0000-0000-000000000000	Compact
3	20240515	20240515_57_57_0_82	00000000-0000-0000-0000-000000000000	Compact
4	20240516	20240516_58_58_0_82	00000000-0000-0000-0000-000000000000	Compact
5	20240517	20240517_20_20_0_82	00000000-0000-0000-0000-000000000000	Compact
6	20240525	20240525_56_56_0_82	00000000-0000-0000-0000-000000000000	Compact
7	20240527	20240527_15_15_0_82	00000000-0000-0000-0000-000000000000	Compact
8	20240529	20240529_28_28_0_82	00000000-0000-0000-0000-000000000000	Compact

Удаляем партиции и проверяем данные:

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

```
select *
from system.parts
where table = 'user_activity' and partition like '202405%'

ALTER TABLE user_activity DROP PARTITION '20240511';
ALTER TABLE user_activity DROP PARTITION '20240514';
ALTER TABLE user_activity DROP PARTITION '20240515';
ALTER TABLE user_activity DROP PARTITION '20240516';

ALTER TABLE user_activity DROP PARTITION '20240517';
ALTER TABLE user_activity DROP PARTITION '20240525';
ALTER TABLE user_activity DROP PARTITION '20240527';
ALTER TABLE user_activity DROP PARTITION '20240529';
```

parts 1

Лог выполнения

select * from system.parts where table = 'user_activity'

Введите SQL выражение чтобы отфильтровать результаты

	partition	name	uuid	part_type	active	
1	20240511	20240511_45_45_1_82	00000000-0000-0000-0000-000000000000	Compact	0	
2	20240514	20240514_40_40_1_82	00000000-0000-0000-0000-000000000000	Compact	0	
3	20240515	20240515_57_57_1_82	00000000-0000-0000-0000-000000000000	Compact	0	
4	20240516	20240516_58_58_1_82	00000000-0000-0000-0000-000000000000	Compact	0	
5	20240517	20240517_20_20_1_82	00000000-0000-0000-0000-000000000000	Compact	0	
6	20240525	20240525_56_56_1_82	00000000-0000-0000-0000-000000000000	Compact	0	
7	20240527	20240527_15_15_1_82	00000000-0000-0000-0000-000000000000	Compact	0	
8	20240529	20240529_28_28_1_82	00000000-0000-0000-0000-000000000000	Compact	0	

За май данных в выборке нет:

129
130
131
132
133

```
select *  
from user_activity  
where formatDateTime(activity_date, '%Y-%m') = '2024-05'
```

user_activity 1 ×

Лог выполнения

select * from user_activity where formatDateTime(activity_date, '%Y-%m') = '2024-05'

Введите SQL выражение чтобы отфильтровать

Таблица

activity_date	123 user_id	A-Z activity_type