Report on Lab Work 2.
Gleb Cyganov, J4132C

1. Table schemas

transactions_d

```
DESCRIBE TABLE  gtsyganov_371942.transactions_d

Query id: 6d78084f-8b0c-4f52-bde7-15e02842e7dc

┌─name─────────┬─type─────┬─default_type─┬─default_expression─┐
│ user_id_out  │ Int64    │              │                    │
│ user_id_in   │ Int64    │              │                    │
│ important    │ Int8     │              │                    │
│ amount       │ Float64  │              │                    │
│ datetime     │ DateTime │              │                    │
└──────────────┴──────────┴──────────────┴────────────────────┘

5 rows in set. Elapsed: 0.002 sec.
```

v_avg_by_day

```
DESCRIBE TABLE  gtsyganov_371942.v_avg_by_day

Query id: f5138b60-8f3a-4506-9221-dff917ab5531

┌─name────────┬─type──────────────────────────────┐
│ user_id     │ Int64                             │
│ avg_income  │ AggregateFunction(avg, Float64)   │
│ avg_outcome │ AggregateFunction(avg, Float64)   │
│ date        │ Date                              │
└─────────────┴───────────────────────────────────┘

4 rows in set. Elapsed: 0.003 sec.
```

v_sum_by_month

```
DESCRIBE TABLE  gtsyganov_371942.v_sum_by_month

Query id: 4e3880b3-43f6-473b-89f7-a4028d386b0e

┌─name────┬─type────┬─default_type─┬─default_expre─┐
│ user_id │ Int64   │              │               │
│ month   │ UInt8   │              │               │
│ income  │ Float64 │              │               │
│ outcome │ Float64 │              │               │
└─────────┴─────────┴──────────────┴───────────────┘

4 rows in set. Elapsed: 0.003 sec.
```

v_count_saldo

```
DESCRIBE TABLE  gtsyganov_371942.v_count_saldo

Query id: 024488b5-2f47-4385-b109-4a6c4bd91b55

┌─name────┬─type────┬─default_type─┬─default_expression─┐
│ user_id │ Int64   │              │                    │
│ saldo   │ Float64 │              │                    │
└─────────┴─────────┴──────────────┴────────────────────┘

2 rows in set. Elapsed: 0.003 sec.
```

v_avg_by_month

```
DESCRIBE TABLE  gtsyganov_371942.v_avg_by_month

Query id: 3d97262b-0a68-4899-b01b-5f226ed32f84

┌─name────────┬─type──────────────────────────────┐
│ user_id     │ Int64                             │
│ avg_income  │ AggregateFunction(avg, Float64)   │
│ avg_outcome │ AggregateFunction(avg, Float64)   │
│ month       │ UInt8                             │
└─────────────┴───────────────────────────────────┘

4 rows in set. Elapsed: 0.003 sec.
```

2. Chosen sharding expression justification

SQL for table creation is presented below. I used the month of the transaction as the sharding key. First of all, in almost all tasks, I use month to evaluate data. Secondly, a month allows to evenly distribute transactions.

```sql
-- Create table for transactions dataset
CREATE TABLE gtsyganov_371942.transactions ON CLUSTER kube_clickhouse_cluster
(
    user_id_out Int64,
    user_id_in Int64,
    important Bool,
    amount Float64,
    datetime DateTime
)
ENGINE = MergeTree()
PARTITION BY toYYYYMM(datetime)
ORDER BY (user_id_out, user_id_in);

-- Create table distibuted by month
CREATE TABLE gtsyganov_371942.transactions_d ON CLUSTER kube_clickhouse_cluster AS
gtsyganov_371942.transactions
ENGINE = Distributed(
    kube_clickhouse_cluster,
    gtsyganov_371942,
    transactions,
    intHash64(toYYYYMM(datetime))
);
```

3. List the chosen MVs and provide their creation queries.

Chosen MV: 1, 3, 4.

MV 1. Average amount for incoming and outcoming transactions by months and days for each user.

```
SHOW CREATE TABLE gtsyganov_371942.v_avg_by_day

Query id: b9d9e02b-994d-4359-a90b-10563453aefa

┌─statement──────────────────────────────────────────────────────────────
──────────────────────────────────────────────────────────────────────────┐
│ CREATE MATERIALIZED VIEW gtsyganov_371942.v_avg_by_day
(
    `user_id` Int64,
    `avg_income` AggregateFunction(avg, Float64),
    `avg_outcome` AggregateFunction(avg, Float64),
    `date` Date
)
ENGINE = AggregatingMergeTree
ORDER BY (user_id, date)
SETTINGS index_granularity = 8192 AS
SELECT
    A.user_id,
    A.avg_income,
    B.avg_outcome,
    A.date
FROM
(
    SELECT
        user_id_in AS user_id,
        avgState(amount) AS avg_income,
        toDate(datetime) AS date
    FROM gtsyganov_371942.transactions_d
    GROUP BY
        user_id_in,
        toDate(datetime)
) AS A
INNER JOIN
(
    SELECT
        user_id_out AS user_id,
        avgState(amount) AS avg_outcome,
        toDate(datetime) AS date
    FROM gtsyganov_371942.transactions_d
    GROUP BY
        user_id_out,
        toDate(datetime)
) AS B ON (A.user_id = B.user_id) AND (A.date = B.date) │
└────────────────────────────────────────────────────────────────────────
──────────────────────────────────────────────────────────────────────────┘

1 rows in set. Elapsed: 0.003 sec.
```

```
SHOW CREATE TABLE gtsyganov_371942.v_avg_by_month

Query id: 2439093b-2dcc-46f3-bd35-051f5790baed

┌─statement─────────────────────────────────────────────────────────────
─────────────────────────────────────────────────────────────────────────┐
│ CREATE MATERIALIZED VIEW gtsyganov_371942.v_avg_by_month
(
    `user_id` Int64,
    `avg_income` AggregateFunction(avg, Float64),
    `avg_outcome` AggregateFunction(avg, Float64),
    `month` UInt8
)
ENGINE = AggregatingMergeTree
ORDER BY (user_id, month)
SETTINGS index_granularity = 8192 AS
SELECT
    A.user_id,
    A.avg_income,
    B.avg_outcome,
    A.month
FROM
(
    SELECT
        user_id_in AS user_id,
        avgState(amount) AS avg_income,
        toMonth(datetime) AS month
    FROM gtsyganov_371942.transactions_d
    GROUP BY
        user_id_in,
        toMonth(datetime)
) AS A
INNER JOIN
(
    SELECT
        user_id_out AS user_id,
        avgState(amount) AS avg_outcome,
        toMonth(datetime) AS month
    FROM gtsyganov_371942.transactions_d
    GROUP BY
        user_id_out,
        toMonth(datetime)
) AS B ON (A.user_id = B.user_id) AND (A.month = B.month) │
└───────────────────────────────────────────────────────────────────────
─────────────────────────────────────────────────────────────────────────┘

1 rows in set. Elapsed: 0.003 sec.
```

MV 3. The sums for incoming and outcoming transactions by months for each user.

```
SHOW CREATE TABLE gtsyganov_371942.v_sum_by_month

Query id: c67699fd-58d7-4503-8c9c-6ec1513d6fe1

┌─statement──────────────────────────────────────────────────────────────
────────────────────────────────────────────────────────────────────────┐
│ CREATE MATERIALIZED VIEW gtsyganov_371942.v_sum_by_month
(
    `user_id` Int64,
    `month` UInt8,
    `income` Float64,
    `outcome` Float64
)
ENGINE = AggregatingMergeTree
ORDER BY (user_id, month)
SETTINGS index_granularity = 8192 AS
SELECT
    A.user_id,
    A.month,
    A.amount AS income,
    B.amount AS outcome
FROM
(
    SELECT
        user_id_out AS user_id,
        toMonth(datetime) AS month,
        sum(amount) AS amount
    FROM gtsyganov_371942.transactions_d
    GROUP BY
        user_id,
        toMonth(datetime)
) AS A
INNER JOIN
(
    SELECT
        user_id_in AS user_id,
        toMonth(datetime) AS month,
        sum(amount) AS amount
    FROM gtsyganov_371942.transactions_d
    GROUP BY
        user_id,
        toMonth(datetime)
) AS B ON (A.user_id = B.user_id) AND (A.month = B.month) │
└────────────────────────────────────────────────────────────────────────
────────────────────────────────────────────────────────────────────────┘

1 rows in set. Elapsed: 0.005 sec.
```

MV 4. Users' saldo for the current moment.

```
SHOW CREATE TABLE gtsyganov_371942.v_count_saldo

Query id: 18422e5b-347d-42f6-8bf7-27870349fb11

┌─statement──────────────────────────────────────────────────────────────
─────────────────────────────────────────────────────────────────────────┐
│ CREATE MATERIALIZED VIEW gtsyganov_371942.v_count_saldo
(
    `user_id` Int64,
    `saldo` Float64
)
ENGINE = AggregatingMergeTree
ORDER BY (user_id, saldo)
SETTINGS index_granularity = 8192 AS
SELECT
    A.user_id,
    B.income - A.outcome AS saldo
FROM
(
    SELECT
        user_id_out AS user_id,
        sum(amount) AS outcome
    FROM gtsyganov_371942.transactions_d
    GROUP BY user_id_out
) AS A
INNER JOIN
(
    SELECT
        user_id_in AS user_id,
        sum(amount) AS income
    FROM gtsyganov_371942.transactions_d
    GROUP BY user_id_in
) AS B ON A.user_id = B.user_id │
└────────────────────────────────────────────────────────────────────────
─────────────────────────────────────────────────────────────────────────┘

1 rows in set. Elapsed: 0.013 sec.
```

Checking outputs

```
SELECT
    user_id,
    avgMerge(avg_income) AS avg_income,
    avgMerge(avg_outcome) AS avg_outcome,
    date
FROM gtsyganov_371942.v_avg_by_day
WHERE user_id = 123
GROUP BY
    user_id,
    date
ORDER BY
    user_id ASC,
    date ASC
LIMIT 10

Query id: 1439e038-bc95-4f97-824f-8e165d4a624a
```

user_id	avg_income	avg_outcome	date
123	608.6175000000001	713.675	2018-01-01
123	474.99000000000007	490.375	2018-01-02
123	634.9333333333333	630.1575	2018-01-03
123	783.03	308.9775	2018-01-04
123	748	508.6875	2018-01-05
123	388.02	528.025	2018-01-06
123	347.34749999999997	350.0833333333333	2018-01-08
123	86.055	671.75	2018-01-09
123	440.4375	338.07666666666665	2018-01-10
123	206.01	520.9350000000001	2018-01-12

```
SELECT
    user_id,
    avgMerge(avg_income) AS avg_income,
    avgMerge(avg_outcome) AS avg_outcome,
    month
FROM gtsyganov_371942.v_avg_by_month
WHERE user_id = 123
GROUP BY
    user_id,
    month
ORDER BY
    user_id ASC,
    month ASC

Query id: 000e2f3c-d8e8-483f-bcac-b43af70ca078
```

user_id	avg_income	avg_outcome	month
123	526.6100909090909	499.4343137254902	1
123	508.9664444444445	501.82118279569886	2
123	536.8673737373737	498.4857281553394	3
123	473.42282608695655	497.6565979381442	4
123	504.83386792452836	526.4030693069302	5
123	477.97121621621613	466.22937499999966	6
123	478.2986734693878	528.2217204301076	7
123	546.943	516.4867741935485	8
123	484.94424242424236	534.2464705882352	9
123	508.6058064516129	487.6297142857144	10
123	479.80273584905655	539.9660465116277	11
123	480.5748148148148	499.34930232558133	12

```
SELECT
    user_id,
    max(income),
    max(outcome)
FROM gtsyganov_371942.v_sum_by_month
WHERE user_id IN (123, 456, 789)
GROUP BY user_id

Query id: 6718e6de-f8d7-46ad-a117-3158610ac83f

┌─user_id─┬──────max(income)─┬─────max(outcome)─┐
│     456 │         56548.96 │ 56062.18999999999 │
│     789 │ 61235.41000000001 │ 58830.340000000004 │
│     123 │ 64416.05999999999 │ 60163.729999999996 │
└─────────┴──────────────────┴──────────────────┘
```

```
SELECT
    user_id,
    saldo
FROM gtsyganov_371942.v_count_saldo
WHERE saldo < 0
LIMIT 20

Query id: ea4b33a8-4615-4acd-a108-72b6190caf11

┌─user_id─┬─────────────saldo─┐
│       3 │  -29191.36999999988 │
│       4 │ -27961.040000000037 │
│       8 │  -32012.45000000007 │
│      11 │ -26776.730000000098 │
│      13 │  -35514.23999999999 │
│      16 │ -21623.420000000042 │
│      17 │  -8049.420000000042 │
│      19 │  -33154.34999999998 │
│      20 │ -13543.839999999967 │
│      22 │ -10816.089999999967 │
│      23 │  -64272.48999999999 │
│      25 │  -9908.689999999944 │
│      27 │ -25708.169999999925 │
│      29 │   -8753.09999999986 │
│      31 │  -4068.819999999949 │
│      32 │ -11041.580000000075 │
│      35 │ -22544.060000000056 │
│      37 │ -12389.600000000093 │
│      38 │  -24298.43000000005 │
│      39 │  -17492.32000000018 │
└─────────┴──────────────────┘
```