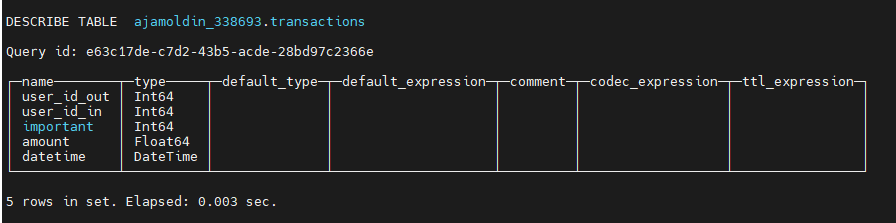
Short Report by Lab3.

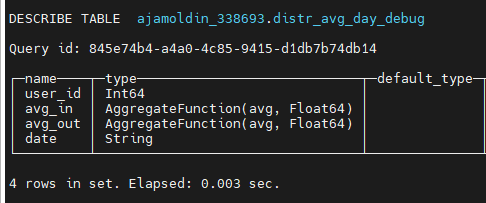
Yamoldin Alexander, J41321c

Table chemas

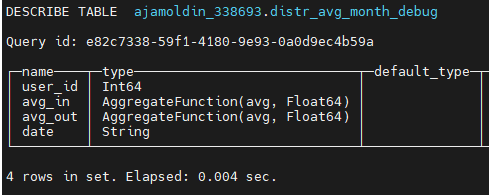
Transactions table



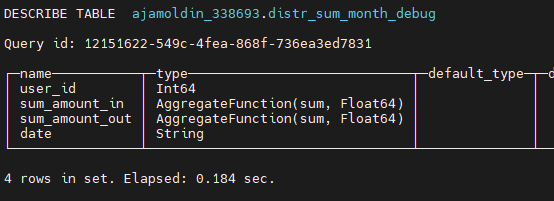
Average transactions per day



Average transactions per month



Sum transactions per month



Create table transactions

/\*создаём таблицу транзакций\*/

CREATE TABLE ajamoldin\_338693.transactions ON CLUSTER kube\_clickhouse\_cluster

(

    user\_id\_out Int64,

    user\_id\_in Int64,

    important Int64,

    amount Float64,

    datetime DateTime

)

ENGINE = MergeTree()

PARTITION BY toYYYYMM(datetime)

ORDER BY (user\_id\_out, user\_id\_in, amount)

I choose xxHash63 by datetime because in all my tasks I evaluate values per month/day so I need to keep the same data in same host.

CREATE TABLE ajamoldin\_338693.distributed\_transactions ON CLUSTER kube\_clickhouse\_cluster AS ajamoldin\_338693.transactions

ENGINE = Distributed(

    kube\_clickhouse\_cluster,

    ajamoldin\_338693,

    transactions,

    xxHash64(datetime)

)

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

Average transactions by day

CREATE MATERIALIZED VIEW ajamoldin\_338693.avg\_day\_debug ON CLUSTER kube\_clickhouse\_cluster

ENGINE = AggregatingMergeTree

ORDER BY (user\_id, date) POPULATE AS

SELECT

    user\_id\_in AS user\_id,

    avg\_amount\_in AS avg\_in,

    avg\_amount\_out AS avg\_out,

    date\_in AS date

FROM

(

    SELECT

        user\_id\_in,

        avgState(amount) AS avg\_amount\_in,

        formatDateTime(datetime, '%d-%m-%G') AS date\_in

    FROM ajamoldin\_338693.transactions

    GROUP BY

        user\_id\_in,

        date\_in

) AS t1

INNER JOIN

(

    SELECT

        user\_id\_out,

        avgState(amount) AS avg\_amount\_out,

        formatDateTime(datetime, '%d-%m-%G') AS date\_out

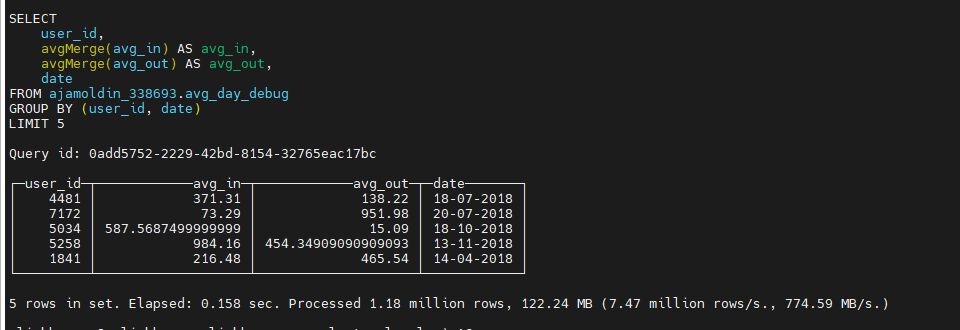
    FROM ajamoldin\_338693.transactions

    GROUP BY

        user\_id\_out,

        date\_out

) AS t2 ON (t1.user\_id\_in = t2.user\_id\_out) AND (t1.date\_in = t2.date\_out)



CREATE TABLE ajamoldin\_338693.distr\_avg\_day\_debug ON CLUSTER kube\_clickhouse\_cluster AS ajamoldin\_338693.avg\_day\_debug

ENGINE = Distributed(kube\_clickhouse\_cluster, ajamoldin\_338693, avg\_day\_debug, xxHash64(user\_id))

Average transactions by month

/\*Среднее ин/фут за месяц\*/

CREATE MATERIALIZED VIEW ajamoldin\_338693.avg\_month\_debug ON CLUSTER kube\_clickhouse\_cluster

ENGINE = AggregatingMergeTree

ORDER BY (user\_id, date) POPULATE AS

SELECT

    user\_id\_in AS user\_id,

    avg\_amount\_in AS avg\_in,

    avg\_amount\_out AS avg\_out,

    date\_in AS date

FROM

(

    SELECT

        user\_id\_in,

        avgState(amount) AS avg\_amount\_in,

        formatDateTime(datetime, '%m-%G') AS date\_in

    FROM ajamoldin\_338693.transactions

    GROUP BY

        user\_id\_in,

        date\_in

) AS t1

INNER JOIN

(

    SELECT

        user\_id\_out,

        avgState(amount) AS avg\_amount\_out,

        formatDateTime(datetime, '%m-%G') AS date\_out

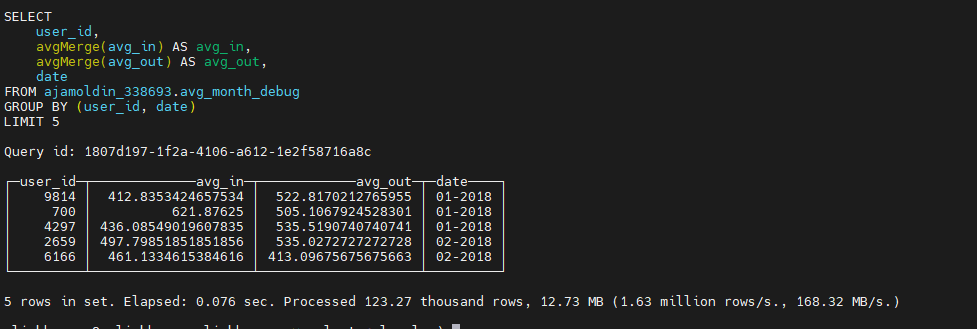
    FROM ajamoldin\_338693.transactions

    GROUP BY

        user\_id\_out,

        date\_out

) AS t2 ON (t1.user\_id\_in = t2.user\_id\_out) AND (t1.date\_in = t2.date\_out)



CREATE TABLE ajamoldin\_338693.distr\_avg\_month\_debug ON CLUSTER kube\_clickhouse\_cluster AS ajamoldin\_338693.avg\_month\_debug

ENGINE = Distributed(kube\_clickhouse\_cluster, ajamoldin\_338693, avg\_month\_debug)

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

/\*сумма за месяц\*/

CREATE MATERIALIZED VIEW ajamoldin\_338693.sum\_month\_debug ON CLUSTER kube\_clickhouse\_cluster

ENGINE = AggregatingMergeTree()

ORDER BY (user\_id, date)

POPULATE

AS SELECT

    user\_id\_in AS user\_id,

    sum\_amount\_in,

    sum\_amount\_out,

    date\_in AS date

FROM

(

    SELECT

        user\_id\_in,

        sumState(amount) AS sum\_amount\_in,

        formatDateTime(datetime, '%m-%G') AS date\_in

    FROM ajamoldin\_338693.transactions

    GROUP BY

        user\_id\_in,

        date\_in

) AS t1

INNER JOIN

(

    SELECT

        user\_id\_out,

        sumState(amount) AS sum\_amount\_out,

        formatDateTime(datetime, '%m-%G') AS date\_out

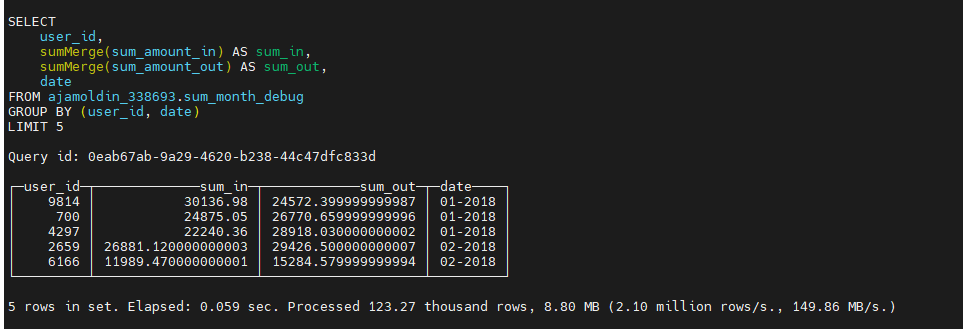
    FROM ajamoldin\_338693.transactions

    GROUP BY

        user\_id\_out,

        date\_out

) AS t2 ON (t1.user\_id\_in = t2.user\_id\_out) AND (t1.date\_in = t2.date\_out)



CREATE TABLE ajamoldin\_338693.distr\_sum\_month\_debug ON CLUSTER kube\_clickhouse\_cluster AS ajamoldin\_338693.sum\_month\_debug

ENGINE = Distributed(kube\_clickhouse\_cluster, ajamoldin\_338693, sum\_month\_debug)

Source code you can find in

<https://github.com/AAYamoldin/TrainingPrograms/blob/master/Python/ITMO_Technologies_and_Infrastructure_for_Big_Data/Lab3/чистовое.sql>