

# Домашнее задание 4

## Запуск ClickHouse сервера

```
[root@my-vpn:~# now=$(date)
[root@my-vpn:~# echo "$now"
Wed Dec 17 07:37:24 PM UTC 2025
[root@my-vpn:~# clickhouse-client -m
ClickHouse client version 18.16.1.
Connecting to localhost:9000.
Connected to ClickHouse server version 18.16.1 revision 54412.

my-vpn.aeza.network :)
```

## Создание базы

```
my-vpn.aeza.network :) Перезагрузить
INSERT INTO default.metrics_test SELECT (rand() % 999) + 1 AS int_val, generateUUIDv4() AS uuid_val, now() - toIntervalSecond(rand() / 1500) AS dt_val, multiIf(int_val <= 300, 'A', int_val <= 600, 'B', 'C') AS str_val FROM numbers(1000000);
INSERT INTO default.metrics_test SELECT (rand() % 999) + 1 AS int_val, generateUUIDv4() AS uuid_val, now() - toIntervalSecond(rand() / 1500) AS dt_val, multiIf(int_val <= 300, 'A', int_val <= 600, 'B', 'C') AS str_val FROM numbers(1000000);
Ok.
0 rows in set. Elapsed: 0.283 sec. Processed 1.05 million rows, 8.39 MB (3.71 million rows/s., 29.66 MB/s.)
my-vpn.aeza.network :)
```

## Аналитический запрос

```
[my-vpn.aeza.network :) SELECT str_val, count(), avg(int_val) FROM default.metrics_test GROUP BY str_val;

SELECT
    str_val,
    count(),
    avg(int_val)
FROM default.metrics_test
GROUP BY str_val

┌-str_val-┐┌-count()-┐┌-avg(int_val)-┐
├─┴─┴─┴─┤├─┴─┴─┴─┤├─┴─┴─┴─┤
│ B      │ │ 299951 │ │ 450.60179495984346 │
│ C      │ │ 400093 │ │ 800.0017470938006 │
│ A      │ │ 299956 │ │ 150.47985037805546 │
├─┴─┴─┴─┤├─┴─┴─┴─┤├─┴─┴─┴─┤

3 rows in set. Elapsed: 0.013 sec. Processed 1.00 million rows, 3.00 MB (74.82 million rows/s., 224.46 MB/s.)
my-vpn.aeza.network :)
```

## Системные таблицы

```
[my-vpn.aeza.network :~] SELECT cluster, shard_num, replica_num, host_name, is_local FROM system.clusters;
```

```
SELECT
  cluster,
  shard_num,
  replica_num,
  host_name,
  is_local
FROM system.clusters
```

cluster	shard_num	replica_num	host_name	is_local
test_cluster_two_shards_localhost	1	1	localhost	1
test_cluster_two_shards_localhost	2	1	localhost	1
test_shard_localhost	1	1	localhost	1
test_shard_localhost_secure	1	1	localhost	0

4 rows in set. Elapsed: 0.003 sec.

```
[my-vpn.aeza.network :~] SELECT * FROM system.macros;
```

```
SELECT *
FROM system.macros
```

macro	substitution
replica	109.237.98.160
shard	1

2 rows in set. Elapsed: 0.001 sec.

Проверка сжатия данных

```

my-vpn.aeza.network :) SELECT name,
    formatReadableSize(sum(data_compressed_bytes) AS size) AS compressed,
    formatReadableSize(sum(data_uncompressed_bytes) AS usize) AS uncompressed,
    round(usize / size, 2) AS compr_rate
FROM system.columns
WHERE database = 'default' AND table = 'metrics_test'
GROUP BY name
[ ORDER BY size DESC;

SELECT
    name,
    formatReadableSize(sum(data_compressed_bytes) AS size) AS compressed,
    formatReadableSize(sum(data_uncompressed_bytes) AS usize) AS uncompressed,
    round(usize / size, 2) AS compr_rate
FROM system.columns
WHERE (database = 'default') AND (table = 'metrics_test')
GROUP BY name
ORDER BY size DESC

┌─name─┬─compressed─┬─uncompressed─┬─compr_rate─┐
│ uuid_val │ 15.32 MiB │ 15.26 MiB │ 1 │
│ dt_val │ 3.83 MiB │ 3.81 MiB │ 1 │
│ int_val │ 17.14 KiB │ 1.91 MiB │ 113.97 │
│ str_val │ 4.40 KiB │ 976.56 KiB │ 222.07 │
└─┬─┬─┬─┬─┘

4 rows in set. Elapsed: 0.003 sec.

```

## Параллельная обработка

```

my-vpn.aeza.network :) SELECT count(*) AS count_total,
    countIf(1, str_val = 'A') AS count_A,
    countIf(1, str_val = 'B') AS count_B,
    countIf(1, str_val = 'C') AS count_C
FROM default.metrics_test
[SETTINGS max_threads = 10;

SELECT
    count(*) AS count_total,
    countIf(1, str_val = 'A') AS count_A,
    countIf(1, str_val = 'B') AS count_B,
    countIf(1, str_val = 'C') AS count_C
FROM default.metrics_test
SETTINGS max_threads = 10

┌─count_total─┬─count_A─┬─count_B─┬─count_C─┐
│ 1000000 │ 299956 │ 299951 │ 400093 │
└─┬─┬─┬─┬─┘

1 rows in set. Elapsed: 0.024 sec. Processed 1.00 million rows, 1.00 MB (41.45 million rows/s., 41.45 MB/s.)

```

## Оптимизация Order By

### Строгая сортировка

<pre> my-vpn.aeza.network :) INSERT INTO default.metrics_test SELECT modulo(rand(), 999) + 1 AS int_val, generateUUIDv4() AS uuid_val, now() - interval rand() / 1000 second AS dt_val, multiIf(int_val &lt;= 300, 'A', int_val &lt;= 600, 'B', 'C') AS str_val FROM numbers(100000000);  INSERT INTO default.metrics_test SELECT (rand() % 999) + 1 AS int_val, generateUUIDv4() AS uuid_val, now() - toIntervalSecond(rand() / 1000) AS dt_val, multiIf(int_val &lt;= 300, 'A', int_val &lt;= 600, 'B', 'C') AS str_val FROM numbers(100000000)  Ok.  0 rows in set. Elapsed: 44.166 sec. Processed 100.01 million rows, 800.06 MB (2.26 million rows/s., 18.11 MB/s.) </pre>	<p>22. Подо чтобы ве • общий 25 % по с • средне увеличил улучшил</p> <p>23. Прод таблицы</p>
---	--

## Нестрогая сортировка

<pre> my-vpn.aeza.network :) SELECT str_val, min(int_val), max(int_val), count(*) FROM default.metrics_test_rnd t WHERE t.dt_val BETWEEN toStartOfDay(now()) - interval 1 day AND toStartOfDay(now()) GROUP BY str_val FORMAT Null;  SELECT str_val, min(int_val), max(int_val), count(*) FROM default.metrics_test_rnd AS t WHERE (t.dt_val &gt;= (toStartOfDay(now()) - toIntervalDay(1))) AND (t.dt_val &lt;= toStartOfDay(now())) GROUP BY str_val FORMAT Null  3 rows in set. Elapsed: 0.056 sec. Processed 2.02 million rows, 14.15 MB (36.02 million rows/s., 251.88 MB/s.) </pre>	<p>20. На созда</p> <p>22. По чтобы • общи 25 % п • сред увели улучш</p> <p>23. Пр табли</p>
---	--

## Сравнение

```
my-vpn.aeza.network :) SELECT table,
    formatReadableSize(sum(data_compressed_bytes) AS size) AS compressed,
    formatReadableSize(sum(data_uncompressed_bytes) AS usize) AS uncompressed,
    formatReadableSize(sum(primary_key_bytes_in_memory)) AS primary_key_bytes_in_memory,
    round(usize / size, 2) AS compr_rate,
    sum(rows) AS rows
FROM system.parts
WHERE (active = 1) AND (database = 'default') AND table IN ('metrics_test', 'metrics_test_rnd')
GROUP BY table;
```

```
SELECT
    table,
    formatReadableSize(sum(data_compressed_bytes) AS size) AS compressed,
    formatReadableSize(sum(data_uncompressed_bytes) AS usize) AS uncompressed,
    formatReadableSize(sum(primary_key_bytes_in_memory)) AS primary_key_bytes_in_memory,
    round(usize / size, 2) AS compr_rate,
    sum(rows) AS rows
FROM system.parts
WHERE (active = 1) AND (database = 'default') AND (table IN ('metrics_test', 'metrics_test_rnd'))
GROUP BY table
```

table	compressed	uncompressed	primary_key_bytes_in_memory	compr_rate	rows
metrics_test_rnd	2.02 GiB	2.14 GiB	47.69 KiB	1.06	100000000
metrics_test	1.77 GiB	2.14 GiB	47.69 KiB	1.21	100000000

2 rows in set. Elapsed: 0.005 sec.