

```
root@testPSQ501:/home/administrator# apt-get install -y apt-transport-https ca-certificates curl gnupg
Чтение списков пакетов... Готово
Построение дерева зависимостей... Готово
Чтение информации о состоянии... Готово
Уже установлен пакет ca-certificates самой новой версии (20240203~22.04.1).
Уже установлен пакет curl самой новой версии (7.81.0-1ubuntu1.21).
Уже установлен пакет gnupg самой новой версии (2.2.27-3ubuntu2.5).
gnupg помечен как установленный вручную.
Следующий пакет устанавливался автоматически и больше не требуется:
  clickhouse-common-static
Для его удаления используйте «sudo apt autoremove».
Следующие пакеты будут обновлены:
  apt-transport-https
Обновлено 1 пакетов, установлено 0 новых пакетов, для удаления отмечено 0 пакетов, и 100 пакетов не обновлено.
Необходимо скачать 1 510 В архивов.
После данной операции объем занятого дискового пространства возрастет на 0 В.
Пол:1 http://ru.archive.ubuntu.com/ubuntu jammy-updates/universe amd64 apt-transport-https all 2.4.14 [1 510 В]
Получено 1 510 В за 0с (83,4 kB/s)
(Чтение базы данных ... на данный момент установлено 160088 файлов и каталогов.)
Подготовка к распаковке .../apt-transport-https_2.4.14_all.deb ...
Распаковывается apt-transport-https (2.4.14) на замену (2.4.12) ...
Настраивается пакет apt-transport-https (2.4.14) ...
Scanning processes ...
Scanning candidates ...
Scanning linux images ...
```

```
No VM guests are running outdated hypervisor (qemu) binaries on this host.
root@testPSQS01:/home/administrator# curl -fsSL 'https://packages.clickhouse.com/rpm/lts/repodata/repomd.xml.key' | sudo gpg --dearmor -o /usr/share/keyrings/clickhouse-keyring.gpg
root@testPSQS01:/home/administrator#
# Get the system architecture
root@testPSQS01:/home/administrator# ARCH=$(dpkg --print-architecture)
root@testPSQS01:/home/administrator# echo "deb [signed-by=/usr/share/keyrings/clickhouse-keyring.gpg arch=${ARCH}] https://packages.clickhouse.com/deb stable main" | sudo tee /etc/apt/sources.list.d/clickhouse.list
deb [signed-by=/usr/share/keyrings/clickhouse-keyring.gpg arch=amd64] https://packages.clickhouse.com/deb stable main
root@testPSQS01:/home/administrator# apt-get update
Cyw:1 http://ru.archive.ubuntu.com/ubuntu jammy InRelease
Non:2 http://ru.archive.ubuntu.com/ubuntu jammy-updates InRelease [128 kB]
Non:3 http://ru.archive.ubuntu.com/ubuntu jammy-backports InRelease [127 kB]
Non:4 http://ru.archive.ubuntu.com/ubuntu jammy-security InRelease [129 kB]
Cyw:5 https://repo.puls.ru/repository/puls-apt-jammy jammy InRelease
Non:6 http://ru.archive.ubuntu.com/ubuntu jammy-updates/main amd64 Packages [3 165 kB]
Non:7 http://ru.archive.ubuntu.com/ubuntu jammy-updates/main amd64 c-n-f Metadata [19.1 kB]
Non:8 http://ru.archive.ubuntu.com/ubuntu jammy-updates/universe amd64 Packages [1 248 kB]
Non:9 http://ru.archive.ubuntu.com/ubuntu jammy-updates/universe amd64 c-n-f Metadata [39.1 kB]
Non:10 http://ru.archive.ubuntu.com/ubuntu jammy-updates/multiverse amd64 c-n-f Metadata [768 B]
Non:11 https://apt.postgresql.org/pub/repos/apt jammy-pgdg InRelease [167 kB]
```

```
root@testPSQS01:/home/administrator# apt-get install -y clickhouse-server clickhouse-client
Чтение списков пакетов... Готово
Построение дерева зависимостей... Готово
Чтение информации о состоянии... Готово
Будут установлены следующие дополнительные пакеты:
  clickhouse-common-static
Предлагаемые пакеты:
  clickhouse-common-static-dbg
Следующие НОВЫЕ пакеты будут установлены:
  clickhouse-client clickhouse-server
Следующие пакеты будут обновлены:
  clickhouse-common-static
Обновлено 1 пакетов, установлено 2 новых пакетов, для удаления отмечено 0 пакетов, и 100 пакетов не обновлено.
Необходимо скачать 216 MB архивов.
После данной операции объем занятого дискового пространства возрастет на 243 MB.
Пол:1 https://packages.clickhouse.com/deb/stable/main/amd64/clickhouse-common-static\_amd64\_25.12.4.35 [216 MB]
Пол:2 https://packages.clickhouse.com/deb/stable/main/amd64/clickhouse-client\_amd64\_25.12.4.35 [185 kB]
Пол:3 https://packages.clickhouse.com/deb/stable/main/amd64/clickhouse-server\_amd64\_25.12.4.35 [214 kB]
Получено 216 MB за 8с (27,2 MB/s)
(Чтение базы данных ... на данный момент установлено 160088 файлов и каталогов.)
Подготовка к распаковке .../clickhouse-common-static_25.12.4.35_amd64.deb ...
Распаковывается clickhouse-common-static (25.12.4.35) на замену (24.8.1.2684) ...
Выбор ранее не выбранного пакета clickhouse-client.
Подготовка к распаковке .../clickhouse-client_25.12.4.35_amd64.deb ...
Распаковывается clickhouse-client (25.12.4.35) ...
Выбор ранее не выбранного пакета clickhouse-server.
Подготовка к распаковке .../clickhouse-server_25.12.4.35_amd64.deb ...
Распаковывается clickhouse-server (25.12.4.35)
```

```

testPSQS01 :) CREATE TABLE trips
(
  CREATE TABLE trips
(
  `trip_id` UInt32,
  `vendor_id` Enum8('1' = 1, '2' = 2, '3' = 3, '4' = 4, 'CMT' = 5, 'VTS' = 6, 'DDS' = 7, 'B02512' =
),
  `pickup_date` Date,
  `pickup_datetime` DateTime,
  `dropoff_date` Date,
  `dropoff_datetime` DateTime,
  `store_and_fwd_flag` UInt8,
  `rate_code_id` UInt8,
  `pickup_longitude` Float64,
  `pickup_latitude` Float64,
  `dropoff_longitude` Float64,
  `dropoff_latitude` Float64,
  `passenger_count` UInt8,
  `trip_distance` Float64,
  `fare_amount` Float32,
  `extra` Float32,
  `mta_tax` Float32,
  `tip_amount` Float32,
  `tolls_amount` Float32,
  `ehail_fee` Float32,
  `improvement_surcharge` Float32,
  `total_amount` Float32,
  `payment_type` Enum8('UNK' = 0, 'CSH' = 1, 'CRE' = 2, 'NOC' = 3, 'DIS' = 4),
  `trip_type` UInt8,
  `pickup` FixedString(25),
  `dropoff` FixedString(25),

```

```
`dropoff_ctcode` UInt8,  
`dropoff_borocode` UInt8,  
`dropoff_ct2010` String,  
`dropoff_boroc2010` String,  
`dropoff_cdeligibil` String,  
`dropoff_ntacode` FixedString(4),  
`dropoff_ntaname` String,  
`dropoff_puma` UInt16
```

```
)  
ENGINE = MergeTree  
PARTITION BY toYYYYMM(pickup_date)  
ORDER BY pickup_datetime
```

Query id: b5b21c3b-36cb-44c3-ac11-4ab17ba74761

Ok.

0 rows in set. Elapsed: 0.009 sec.

testPSQS01 :) █


```

INSERT INTO trips
SETTINGS input_format_try_infer_datetimes = 0
SELECT *
FROM s3('https://datasets-documentation.s3.eu-west-3.amazonaws.com/nyc-taxi/trips_{1..2}.gz', 'TabSeparatedWithNames', '
' = 1, '2' = 2, '3' = 3, '4' = 4, 'CMT' = 5, 'VTS' = 6, 'DDS' = 7, 'B02512' = 10, 'B02598' = 11, 'B02617' = 12, 'B02682'
Date,\n    \pickup_datetime\` DateTime,\n    \dropoff_date\` Date,\n    \dropoff_datetime\` DateTime,\n    \store_and
\pickup_longitude\` Float64,\n    \pickup_latitude\` Float64,\n    \dropoff_longitude\` Float64,\n    \dropoff_lati
\trip_distance\` Float64,\n    \fare_amount\` Float32,\n    \extra\` Float32,\n    \mta_tax\` Float32,\n    \tip_am
ehail_fee\` Float32,\n    \improvement_surcharge\` Float32,\n    \total_amount\` Float32,\n    \payment_type\` Enum8
4),\n    \trip_type\` UInt8,\n    \pickup\` FixedString(25),\n    \dropoff\` FixedString(25),\n    \cab_type\` Enum8
p_nyct2010_gid\` Int8,\n    \pickup_ctlabel\` Float32,\n    \pickup_borocode\` Int8,\n    \pickup_ct2010\` String,\n
ibil\` String,\n    \pickup_ntacode\` FixedString(4),\n    \pickup_ntaname\` String,\n    \pickup_puma\` UInt16,\n
el\` Float32,\n    \dropoff_borocode\` UInt8,\n    \dropoff_ct2010\` String,\n    \dropoff_boroc2010\` String,\n
\` FixedString(4),\n    \dropoff_ntaname\` String,\n    \dropoff_puma\` UInt16\n`)
SETTINGS input_format_try_infer_datetimes = 0

Query id: 9536a873-7d09-4254-8a77-a220e19c937c

ok.

1999657 rows in set. Elapsed: 16.683 sec. Processed 2.00 million rows, 163.07 MB (119.86 thousand rows/s., 9.77 MB/s.)
Peak memory usage: 1001.33 MiB.

testPS0S01 :) SELECT count() FROM trips

```

```
testPSQS01 :) select count() from trips where payment_type = 1
```

```
SELECT count()  
FROM trips  
WHERE payment_type = 1
```

```
Query id: 49b01862-7d30-4e87-8a56-9d99a358b5be
```

```
1. 

| count() |
|---------|
| 1232637 |

 -- 1.23 million
```

```
1 row in set. Elapsed: 0.016 sec. Processed 2.00 million rows, 2.00 MB (125.15 million rows/s., 125.15 MB/s.)  
Peak memory usage: 323.87 KiB.
```

```
testPSQS01 :) █
```

0 rows in set. Elapsed: 0.002 sec.

testPSQS01 :) SET max_insert_threads=2;

SET max_insert_threads = 2

Query id: 20e69c4a-1f3c-4304-9348-bba03ea4d8a9

Ok.

0 rows in set. Elapsed: 0.001 sec.


```

INSERT INTO trips
SETTINGS input_format_try_infer_datetimes = 0
SELECT *
FROM s3('https://datasets-documentation.s3.eu-west-3.amazonaws.com/nyc-taxi/trips_{1..2}.gz', 'TabSeparatedWithNames', '\n
' = 1, '2' = 2, '3' = 3, '4' = 4, 'CMT' = 5, 'VTS' = 6, 'DDS' = 7, 'B02512' = 10, 'B02598' = 11, 'B02617' = 12, 'B02682' =
Date,\n      \pickup_datetime\ DateTime,\n      \dropoff_date\ Date,\n      \dropoff_datetime\ DateTime,\n      \store_and_f
\pickup_longitude\ Float64,\n      \pickup_latitude\ Float64,\n      \dropoff_longitude\ Float64,\n      \dropoff_latitud
\trip_distance\ Float64,\n      \fare_amount\ Float32,\n      \extra\ Float32,\n      \mta_tax\ Float32,\n      \tip_amount
ehail_fee\ Float32,\n      \improvement_surcharge\ Float32,\n      \total_amount\ Float32,\n      \payment_type\ Enum8('U
4),\n      \trip_type\ UInt8,\n      \pickup\ FixedString(25),\n      \dropoff\ FixedString(25),\n      \cab_type\ Enum8('y
p nyct2010_gid\ Int8,\n      \pickup_ctlabel\ Float32,\n      \pickup_borocode\ Int8,\n      \pickup_ct2010\ String,\n
ibil\ String,\n      \pickup_ntacode\ FixedString(4),\n      \pickup_ntaname\ String,\n      \pickup_puma\ UInt16,\n
el\ Float32,\n      \dropoff_borocode\ UInt8,\n      \dropoff_ct2010\ String,\n      \dropoff_borocod2010\ String,\n      \d
\ FixedString(4),\n      \dropoff_ntaname\ String,\n      \dropoff_puma\ UInt16\n')
SETTINGS input_format_try_infer_datetimes = 0

```

Query id: 7d71edf1-808e-4c27-97a4-1a97dd78c971

Ok.

1999657 rows in set. Elapsed: 11.123 sec. Processed 2.00 million rows, 163.07 MB (179.78 thousand rows/s., 14.66 MB/s.)
Peak memory usage: 1.10 GiB.

testPSQS01 :) █

```

SELECT
    vendor_id,
    payment_type,
    count()
FROM trips AS t1
INNER JOIN trips AS t2 ON t1.pickup_date = t2.dropoff_date
WHERE cab_type = 'yellow'
GROUP BY
    vendor_id,
    payment_type

```

Query id: 31416ed7-5a67-4cc4-9323-7f51728e2f9d

	vendor_id	payment_type	count()
1.	1	CSH	15400564572
2.	2	CSH	17497640809
3.	1	CRE	9468272745
4.	2	CRE	10814862971
5.	1	NOC	176169378
6.	2	NOC	9121775
7.	1	DIS	60530280
8.	2	DIS	7563087

8 rows in set. Elapsed: 69.553 sec. Processed 4.00 million rows, 14.00 MB (57.50 thousand rows/s., 201.25 KB/s.)
Peak memory usage: 74.34 MiB.

```
testPSQS01 :) SELECT * FROM system.settings WHERE name = 'max_threads'
```

```

SELECT *
FROM system.settings
WHERE name = 'max_threads'

```

Query id: cc9f28eb-8577-4d34-befe-25bbf717b861

Row 1:

```

name:          max_threads
value:         2
changed:       1
description:    The maximum number of query processing threads, excluding threads for retrieving data from rep

```

```
testPSQS01 :) SET max_threads=2;
```

```
SET max_threads = 2
```

```
Query id: 1033de3d-be24-4f29-bf55-b692158ca91e
```

```
Ok.
```

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

```
testPSQS01 :) select
^Ivendor_id,
^Ipayment_type,
^Icount()
from trips as t1
^Iinner join trips as t2 on t1.pickup_date = t2.dropoff_date
where cab_type = 'yellow'
group by vendor_id, payment_type;
```

```
SELECT
    vendor_id,
    payment_type,
    count()
FROM trips AS t1
INNER JOIN trips AS t2 ON t1.pickup_date = t2.dropoff_date
WHERE cab_type = 'yellow'
GROUP BY
    vendor_id,
    payment_type
```

```
Query id: 11ad0d8c-4cca-428e-a083-70761ebf0783
```

	vendor_id	payment_type	count()
1.	1	CSH	15400564572
2.	2	CSH	17497640809
3.	1	CRE	9468272745
4.	2	CRE	10814862971
5.	1	NOC	176169378
6.	2	NOC	9121775
7.	1	DIS	60530280
8.	2	DIS	7563087

```
8 rows in set. Elapsed: 69.623 sec. Processed 4.00 million rows, 14.00 MB (57.44 thousand rows/s., 201.05 KB/s.)
Peak memory usage: 74.34 MiB.
```

```
group by t1.vendor_id, t1.payment_type,
```

```
SELECT
    t1.vendor_id,
    t1.payment_type,
    count()
FROM trips AS t1
INNER JOIN
(
    SELECT dropoff_date
    FROM trips
    WHERE payment_type = 'CSH'
) AS t2 ON t1.pickup_date = t2.dropoff_date
WHERE (t1.cab_type = 'yellow') AND (t1.payment_type = 'CRE')
GROUP BY
    t1.vendor_id,
    t1.payment_type
```

Query id: 069ee379-270f-4b9a-b12d-4cffb14177e4

	vendor_id	payment_type	count()
1.	1	CRE	5818232787
2.	2	CRE	6641775495

2 rows in set. Elapsed: 29.531 sec. Processed 4.00 million rows, 16.00 MB (135.43 thousand rows/s., 541.71 KB/s.)
Peak memory usage: 37.83 MiB.

```
testPSQS01 :) SET max_threads=2;

SET max_threads = 2

Query id: 1d2f10a4-a297-49ff-9841-0c2ef1689114

Ok.

0 rows in set. Elapsed: 0.001 sec.

testPSQS01 :) select t1.vendor_id, t1.payment_type, count()
from trips as t1
^Inner join ( SELECT dropoff_date FROM trips WHERE payment_type = 'CSH' ) as t2 on t1.pickup_date = t2.dropoff_date
where t1.cab_type = 'yellow' and t1.payment_type = 'CRE'
group by t1.vendor_id, t1.payment_type;

SELECT
    t1.vendor_id,
    t1.payment_type,
    count()
FROM trips AS t1
INNER JOIN
(
    SELECT dropoff_date
    FROM trips
    WHERE payment_type = 'CSH'
) AS t2 ON t1.pickup_date = t2.dropoff_date
WHERE (t1.cab_type = 'yellow') AND (t1.payment_type = 'CRE')
GROUP BY
    t1.vendor_id,
    t1.payment_type

Query id: 7f386193-6f10-49dd-9a9b-ff410402b07a

1. | vendor_id | payment_type | count() |
2. | 1         | CRE         | 5818232787 |
   | 2         | CRE         | 6641775495 |

2 rows in set. Elapsed: 16.964 sec. Processed 4.00 million rows, 16.00 MB (235.75 thousand rows/s., 943.00 KB/s.)
Peak memory usage: 41.74 MiB.
```

```
1 Для ускорения вставки использовал SET max_insert_threads=2 - заметно влияние.
2
3 Для SELECT использовал сложный запрос с join, т.к. линейный запрос выполнялся слишком быстро.
4 Не заметно было влияние даже параметра max_threads.
5
6 select
7     t1.vendor_id,
8     t1.payment_type,
9     count()
10 from trips as t1
11     inner join (
12         SELECT
13             dropoff_date
14         FROM trips
15         WHERE payment_type = 'CSH'
16     ) as t2 on t1.pickup_date = t2.dropoff_date
17 where t1.cab_type = 'yellow' and t1.payment_type = 'CRE'
18 group by t1.vendor_id, t1.payment_type;
19
20 1) max_threads - ускорение
21 2) partial_merge_join_rows_in_right_blocks=65536->131072 - нет влияния
22 3) max_parallel_replicas - изменять нет смысла, т.к. нет реплик
23 4) allow_asynchronous_read_from_io_pool_for_merge_tree=1 - вышло даже чуть дольше. Но и для SSD не рекомендуется менять дефолтное 0.
```