

## 1. Топология с 3 шардами и фактором репликации 2

В этой конфигурации у нас будет 3 шарда, каждый из которых реплицируется на два сервера. Всего будет использоваться 6 серверов.

```
<clickhouse>
  <remote_servers>
    <sharded_cluster>
      <shard>
        <internal_replication>true</internal_replication>
        <replica>
          <host>clickhouse4</host>
          <port>9000</port>
        </replica>
        <replica>
          <host>clickhouse5</host>
          <port>9000</port>
        </replica>
      </shard>
      <shard>
        <internal_replication>true</internal_replication>
        <replica>
          <host>clickhouse6</host>
          <port>9000</port>
        </replica>
        <replica>
          <host>clickhouse7</host>
          <port>9000</port>
        </replica>
      </shard>
      <shard>
        <internal_replication>true</internal_replication>
        <replica>
          <host>clickhouse8</host>
          <port>9000</port>
        </replica>
        <replica>
          <host>clickhouse9</host>
          <port>9000</port>
        </replica>
      </shard>
    </sharded_cluster>
  </remote_servers>
  <zookeeper>
    <node index="1">
      <host>zookeeper</host>
      <port>2181</port>
    </node>
  </zookeeper>
  <macros>
    <cluster>sharded_cluster</cluster>
    <shard>01</shard>
    <replica>01</replica>
  </macros>
</clickhouse>
```

Содержимое macros для каждого узла:

1. Для clickhouse4:

```
<macros>
  <cluster>sharded_cluster</cluster>
  <shard>01</shard>
  <replica>01</replica>
</macros>
```

2. Для clickhouse5:

```
<macros>
  <cluster>sharded_cluster</cluster>
  <shard>01</shard>
  <replica>02</replica>
</macros>
```

3. Для clickhouse6:
 

```

      <macros>
        <cluster>sharded_cluster</cluster>
        <shard>02</shard>
        <replica>01</replica>
      </macros>
      
```
4. Для clickhouse7:
 

```

      <macros>
        <cluster>sharded_cluster</cluster>
        <shard>02</shard>
        <replica>02</replica>
      </macros>
      
```
5. Для clickhouse8:
 

```

      <macros>
        <cluster>sharded_cluster</cluster>
        <shard>03</shard>
        <replica>01</replica>
      </macros>
      
```
6. Для clickhouse9:
 

```

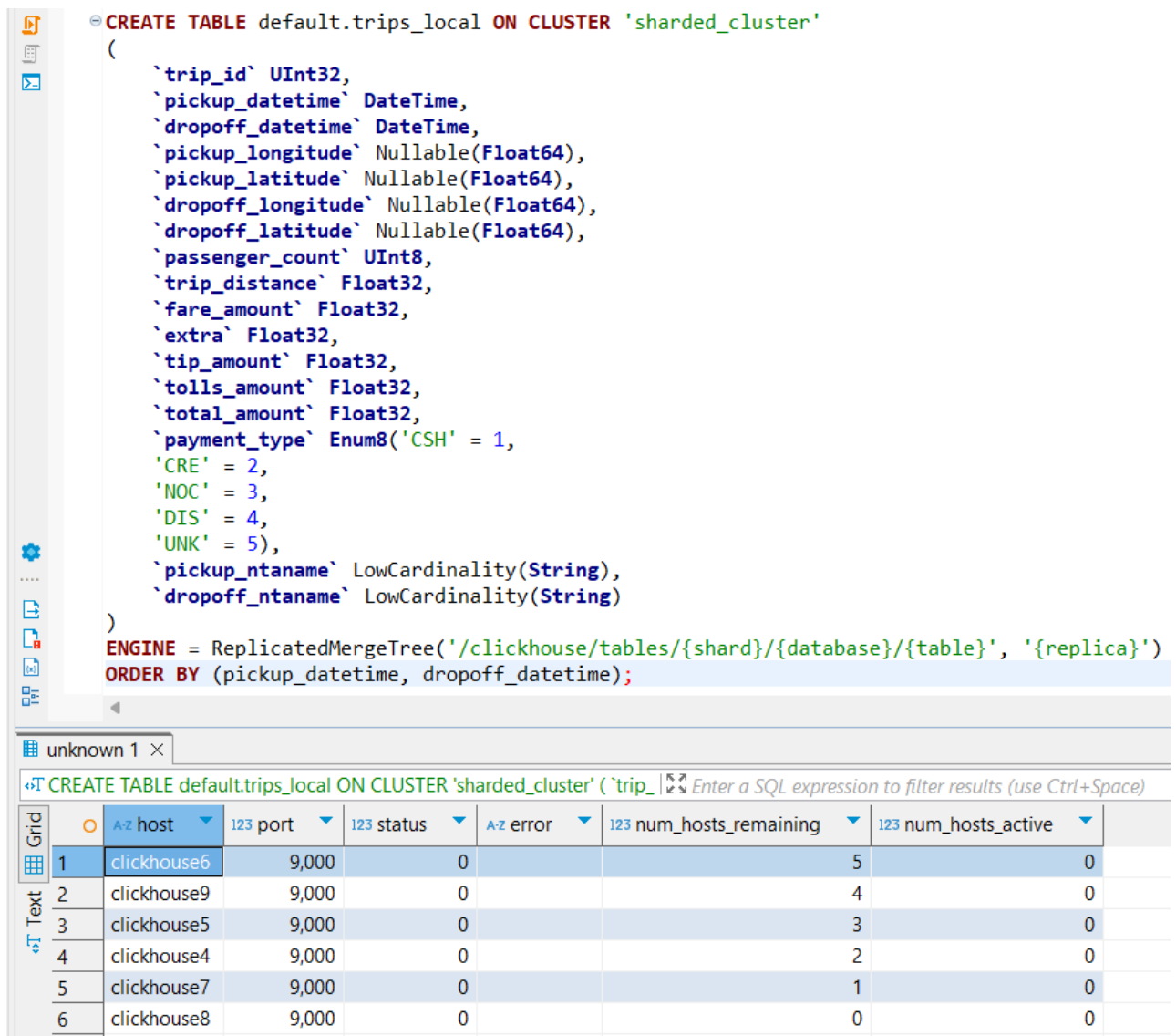
      <macros>
        <cluster>sharded_cluster</cluster>
        <shard>03</shard>
        <replica>02</replica>
      </macros>
      
```

### Создание таблицы «default.trips\_local» с табличным движком «ReplicatedMergeTree»

```

CREATE TABLE default.trips_local ON CLUSTER 'sharded_cluster'
(
  `trip_id` UInt32,
  `pickup_datetime` DateTime,
  `dropoff_datetime` DateTime,
  `pickup_longitude` Nullable(Float64),
  `pickup_latitude` Nullable(Float64),
  `dropoff_longitude` Nullable(Float64),
  `dropoff_latitude` Nullable(Float64),
  `passenger_count` UInt8,
  `trip_distance` Float32,
  `fare_amount` Float32,
  `extra` Float32,
  `tip_amount` Float32,
  `tolls_amount` Float32,
  `total_amount` Float32,
  `payment_type` Enum8('CSH' = 1,
    'CRE' = 2,
    'NOC' = 3,
    'DIS' = 4,
    'UNK' = 5),
  `pickup_ntaname` LowCardinality(String),
  `dropoff_ntaname` LowCardinality(String)
)
ENGINE = ReplicatedMergeTree('/clickhouse/tables/{shard}/{database}/{table}', '{replica}')
ORDER BY (pickup_datetime, dropoff_datetime);

```



```

CREATE TABLE default.trips_local ON CLUSTER 'sharded_cluster'
(
    `trip_id` UInt32,
    `pickup_datetime` DateTime,
    `dropoff_datetime` DateTime,
    `pickup_longitude` Nullable(Float64),
    `pickup_latitude` Nullable(Float64),
    `dropoff_longitude` Nullable(Float64),
    `dropoff_latitude` Nullable(Float64),
    `passenger_count` UInt8,
    `trip_distance` Float32,
    `fare_amount` Float32,
    `extra` Float32,
    `tip_amount` Float32,
    `tolls_amount` Float32,
    `total_amount` Float32,
    `payment_type` Enum8('CSH' = 1,
        'CRE' = 2,
        'NOC' = 3,
        'DIS' = 4,
        'UNK' = 5),
    `pickup_ntaname` LowCardinality(String),
    `dropoff_ntaname` LowCardinality(String)
)
ENGINE = ReplicatedMergeTree('/clickhouse/tables/{shard}/{database}/{table}', '{replica}')
ORDER BY (pickup_datetime, dropoff_datetime);

```

unknown 1 ×

CREATE TABLE default.trips\_local ON CLUSTER 'sharded\_cluster' ( `trip\_id` *Enter a SQL expression to filter results (use Ctrl+Space)*

Grid	A-Z host	123 port	123 status	A-Z error	123 num_hosts_remaining	123 num_hosts_active
1	clickhouse6	9,000	0		5	0
2	clickhouse9	9,000	0		4	0
3	clickhouse5	9,000	0		3	0
4	clickhouse4	9,000	0		2	0
5	clickhouse7	9,000	0		1	0
6	clickhouse8	9,000	0		0	0

### Создание таблицы «default.trips» с табличным движком «Distributed»

```

CREATE TABLE IF NOT EXISTS default.trips ON CLUSTER 'sharded_cluster'
(
    `trip_id` UInt32,
    `pickup_datetime` DateTime,
    `dropoff_datetime` DateTime,
    `pickup_longitude` Nullable(Float64),
    `pickup_latitude` Nullable(Float64),
    `dropoff_longitude` Nullable(Float64),
    `dropoff_latitude` Nullable(Float64),
    `passenger_count` UInt8,
    `trip_distance` Float32,
    `fare_amount` Float32,
    `extra` Float32,
    `tip_amount` Float32,
    `tolls_amount` Float32,
    `total_amount` Float32,
    `payment_type` Enum8('CSH' = 1, 'CRE' = 2, 'NOC' = 3, 'DIS' = 4, 'UNK' = 5),
    `pickup_ntaname` LowCardinality(String),
    `dropoff_ntaname` LowCardinality(String)
)
ENGINE = Distributed('sharded_cluster', 'default', 'trips_local', rand());

```

```
CREATE TABLE IF NOT EXISTS default.trips ON CLUSTER 'sharded_cluster'
(
    `trip_id` UInt32,
    `pickup_datetime` DateTime,
    `dropoff_datetime` DateTime,
    `pickup_longitude` Nullable(Float64),
    `pickup_latitude` Nullable(Float64),
    `dropoff_longitude` Nullable(Float64),
    `dropoff_latitude` Nullable(Float64),
    `passenger_count` UInt8,
    `trip_distance` Float32,
    `fare_amount` Float32,
    `extra` Float32,
    `tip_amount` Float32,
    `tolls_amount` Float32,
    `total_amount` Float32,
    `payment_type` Enum8('CSH' = 1, 'CRE' = 2, 'NOC' = 3, 'DIS' = 4, 'UNK' = 5),
    `pickup_ntaname` LowCardinality(String),
    `dropoff_ntaname` LowCardinality(String)
)
ENGINE = Distributed('sharded_cluster', 'default', 'trips_local', rand());
```

unknown 1 x

CREATE TABLE IF NOT EXISTS default.trips ON CLUSTER 'sharded\_cluster' Enter a SQL expression to filter results (use Ctrl+Sp

	A-Z host	123 port	123 status	A-Z error	123 num_hosts_remaining	123 num_hosts_active
1	clickhouse5	9,000	0		5	0
2	clickhouse6	9,000	0		4	0
3	clickhouse9	9,000	0		3	0
4	clickhouse4	9,000	0		2	0
5	clickhouse7	9,000	0		1	0
6	clickhouse8	9,000	0		0	0

Вставка данных в таблицу «default.trips»

```
INSERT INTO default.trips
SELECT
```

```
    trip_id,
    pickup_datetime,
    dropoff_datetime,
    pickup_longitude,
    pickup_latitude,
    dropoff_longitude,
    dropoff_latitude,
    passenger_count,
    trip_distance,
    fare_amount,
    extra,
    tip_amount,
    tolls_amount,
    total_amount,
    payment_type,
    pickup_ntaname,
    dropoff_ntaname
```

```
FROM
```

```
gcs('https://storage.googleapis.com/clickhouse-public-datasets/nyc-taxi/trips_{0..2}.gz',
    'TabSeparatedWithNames'
);
```



## SHOW CREATE TABLE default.trips

The screenshot shows a database client interface with a top toolbar containing icons for file operations and a search bar. Below the toolbar is a tab labeled "CREATE 1" with a close button. The main area is divided into two panes. The left pane, titled "statement", contains a table with one row and two columns: "id" and "statement". The row contains the value "1" and the SQL statement for creating the "trips" table. The right pane, titled "Value", displays the full SQL statement for creating the "trips" table, including column definitions, data types, and engine settings.

id	statement
1	CREATE TABLE default.trips( `trip_id` UInt32, `pickup_datetime` DateTime, `dropoff_datetime` DateTime, `pickup_longitude` Nullable(Float64), `pickup_latitude` Nullable(Float64), `dropoff_longitude` Nullable(Float64), `dropoff_latitude` Nullable(Float64), `passenger_count` UInt8, `trip_distance` Float32, `fare_amount` Float32, `extra` Float32, `tip_amount` Float32, `tolls_amount` Float32, `total_amount` Float32, `payment_type` Enum8('CSH' = 1, 'CRE' = 2, 'NOC' = 3, 'DIS' = 4, 'UNK' = 5), `pickup_ntaname` LowCardinality(String), `dropoff_ntaname` LowCardinality(String) ) ENGINE = Distributed('sharded_cluster', 'default', 'trips_local', rand())

## 2. Топология с 4 шардами и фактором репликации 3

В этой конфигурации у нас будет 4 шарда, каждый из которых реплицируется на три сервера. Всего будет использоваться 12 серверов.

```
<clickhouse>
  <remote_servers>
    <sharded_cluster_sh4r3>
      <shard>
        <internal_replication>true</internal_replication>
        <replica>
          <host>clickhouse10</host>
          <port>9000</port>
        </replica>
        <replica>
          <host>clickhouse11</host>
          <port>9000</port>
        </replica>
        <replica>
          <host>clickhouse12</host>
          <port>9000</port>
        </replica>
      </shard>
      <shard>
        <internal_replication>true</internal_replication>
        <replica>
          <host>clickhouse13</host>
          <port>9000</port>
        </replica>
        <replica>
          <host>clickhouse14</host>
          <port>9000</port>
        </replica>
        <replica>
          <host>clickhouse15</host>
          <port>9000</port>
        </replica>
      </shard>
      <shard>
        <internal_replication>true</internal_replication>
        <replica>
          <host>clickhouse16</host>
          <port>9000</port>
        </replica>
        <replica>
          <host>clickhouse17</host>
          <port>9000</port>
        </replica>
        <replica>
          <host>clickhouse18</host>
          <port>9000</port>
        </replica>
      </shard>
      <shard>
        <internal_replication>true</internal_replication>
        <replica>
          <host>clickhouse19</host>
          <port>9000</port>
        </replica>
        <replica>
          <host>clickhouse20</host>
          <port>9000</port>
        </replica>
        <replica>
          <host>clickhouse21</host>
          <port>9000</port>
        </replica>
      </shard>
    </sharded_cluster_sh4r3>
  </remote_servers>
  <zookeeper>
    <node index="1">
      <host>zookeeper</host>
```

```

        <port>2181</port>
    </node>
</zookeeper>
<macros>
    <cluster>sharded_cluster_sh4r3</cluster>
    <shard>01</shard>
    <replica>01</replica>
</macros>
</clickhouse>

```

Содержимое macros для каждого узла:

1. Для clickhouse10:
 

```

      <macros>
          <cluster>sharded_cluster</cluster>
          <shard>01</shard>
          <replica>01</replica>
      </macros>
      
```
2. Для clickhouse11:
 

```

      <macros>
          <cluster>sharded_cluster</cluster>
          <shard>01</shard>
          <replica>02</replica>
      </macros>
      
```
3. Для clickhouse12:
 

```

      <macros>
          <cluster>sharded_cluster</cluster>
          <shard>01</shard>
          <replica>03</replica>
      </macros>
      
```
4. Для clickhouse13:
 

```

      <macros>
          <cluster>sharded_cluster</cluster>
          <shard>02</shard>
          <replica>01</replica>
      </macros>
      
```
5. Для clickhouse14:
 

```

      <macros>
          <cluster>sharded_cluster</cluster>
          <shard>02</shard>
          <replica>02</replica>
      </macros>
      
```
6. Для clickhouse15:
 

```

      <macros>
          <cluster>sharded_cluster</cluster>
          <shard>02</shard>
          <replica>03</replica>
      </macros>
      
```
7. Для clickhouse16:
 

```

      <macros>
          <cluster>sharded_cluster</cluster>
          <shard>03</shard>
          <replica>01</replica>
      </macros>
      
```
8. Для clickhouse17:
 

```

      <macros>
          <cluster>sharded_cluster</cluster>
          <shard>03</shard>
          <replica>02</replica>
      </macros>
      
```
9. Для clickhouse18:
 

```

      <macros>
          <cluster>sharded_cluster</cluster>
          <shard>03</shard>
          <replica>03</replica>
      </macros>
      
```



## Создание таблицы «default.trips\_local» с табличным движком «ReplicatedMergeTree»

```
CREATE TABLE default.trips_local ON CLUSTER 'sharded_cluster'
(
    `trip_id` UInt32,
    `pickup_datetime` DateTime,
    `dropoff_datetime` DateTime,
    `pickup_longitude` Nullable(Float64),
    `pickup_latitude` Nullable(Float64),
    `dropoff_longitude` Nullable(Float64),
    `dropoff_latitude` Nullable(Float64),
    `passenger_count` UInt8,
    `trip_distance` Float32,
    `fare_amount` Float32,
    `extra` Float32,
    `tip_amount` Float32,
    `tolls_amount` Float32,
    `total_amount` Float32,
    `payment_type` Enum8('CSH' = 1,
        'CRE' = 2,
        'NOC' = 3,
        'DIS' = 4,
        'UNK' = 5),
    `pickup_ntaname` LowCardinality(String),
    `dropoff_ntaname` LowCardinality(String)
)
ENGINE = ReplicatedMergeTree('/clickhouse/tables/{shard}/{database}/{table}', '{replica}')
ORDER BY (pickup_datetime, dropoff_datetime);
```



unknown 1 ×						
DROP TABLE IF EXISTS default.trips_local ON CLUSTER 'sharded_cluster' Enter a SQL expression to filter results (use Ctrl+Space)						
	Az host	123 port	123 status	Az error	123 num_hosts_remaining	123 num_hosts_active
1	clickhouse17	9,000	0		11	0
2	clickhouse14	9,000	0		10	0
3	clickhouse12	9,000	0		9	0
4	clickhouse21	9,000	0		8	0
5	clickhouse13	9,000	0		7	0
6	clickhouse16	9,000	0		6	0
7	clickhouse18	9,000	0		5	0
8	clickhouse10	9,000	0		4	0
9	clickhouse19	9,000	0		3	0
10	clickhouse20	9,000	0		2	0
11	clickhouse15	9,000	0		1	0
12	clickhouse11	9,000	0		0	0

## Создание таблицы «default.trips» с табличным движком «Distributed»

```
DROP TABLE IF EXISTS default.trips ON CLUSTER 'sharded_cluster_sh4r3';
CREATE TABLE IF NOT EXISTS default.trips ON CLUSTER 'sharded_cluster_sh4r3'
(
    `trip_id` UInt32,
    `pickup_datetime` DateTime,
    `dropoff_datetime` DateTime,
    `pickup_longitude` Nullable(Float64),
    `pickup_latitude` Nullable(Float64),
    `dropoff_longitude` Nullable(Float64),
    `dropoff_latitude` Nullable(Float64),
    `passenger_count` UInt8,
    `trip_distance` Float32,
    `fare_amount` Float32,
    `extra` Float32,
    `tip_amount` Float32,
    `tolls_amount` Float32,
    `total_amount` Float32,
    `payment_type` Enum8('CSH' = 1, 'CRE' = 2, 'NOC' = 3, 'DIS' = 4, 'UNK' = 5),
    `pickup_ntaname` LowCardinality(String),
    `dropoff_ntaname` LowCardinality(String)
)
ENGINE = Distributed('sharded_cluster_sh4r3', 'default', 'trips_local', rand());
```

The screenshot shows a database client interface with a SQL editor and a results grid. The SQL editor contains the following code:

```
DROP TABLE IF EXISTS default.trips ON CLUSTER 'sharded_cluster_sh4r3';
CREATE TABLE IF NOT EXISTS default.trips ON CLUSTER 'sharded_cluster_sh4r3'
(
    `trip_id` UInt32,
    `pickup_datetime` DateTime,
    `dropoff_datetime` DateTime,
    `pickup_longitude` Nullable(Float64),
    `pickup_latitude` Nullable(Float64),
    `dropoff_longitude` Nullable(Float64),
    `dropoff_latitude` Nullable(Float64),
    `passenger_count` UInt8,
    `trip_distance` Float32,
    `fare_amount` Float32,
    `extra` Float32,
    `tip_amount` Float32,
    `tolls_amount` Float32,
    `total_amount` Float32,
    `payment_type` Enum8('CSH' = 1, 'CRE' = 2, 'NOC' = 3, 'DIS' = 4, 'UNK' = 5),
    `pickup_ntaname` LowCardinality(String),
    `dropoff_ntaname` LowCardinality(String)
)
ENGINE = Distributed('sharded_cluster_sh4r3', 'default', 'trips_local', rand());
```

Below the SQL editor, the results are displayed in a table grid. The grid has 7 columns: A-Z host, port, status, error, num\_hosts\_remaining, and num\_hosts\_active. The data is as follows:

	A-Z host	port	status	error	num_hosts_remaining	num_hosts_active
1	clickhouse17	9,000	0		11	0
2	clickhouse14	9,000	0		10	0
3	clickhouse12	9,000	0		9	0
4	clickhouse21	9,000	0		8	0
5	clickhouse13	9,000	0		7	0
6	clickhouse16	9,000	0		6	0
7	clickhouse18	9,000	0		5	0
8	clickhouse10	9,000	0		4	0
9	clickhouse15	9,000	0		3	0
10	clickhouse20	9,000	0		2	0
11	clickhouse19	9,000	0		1	0
12	clickhouse11	9,000	0		0	0

Вставка данных в таблицу «default.trips»

```
INSERT INTO default.trips
SELECT
    trip_id,
    pickup_datetime,
    dropoff_datetime,
    pickup_longitude,
    pickup_latitude,
    dropoff_longitude,
    dropoff_latitude,
    passenger_count,
    trip_distance,
    fare_amount,
    extra,
    tip_amount,
    tolls_amount,
    total_amount,
    payment_type,
    pickup_ntaname,
    dropoff_ntaname
FROM
    gcs('https://storage.googleapis.com/clickhouse-public-datasets/nyc-taxi/trips_{0..2}.gz',
    'TabSeparatedWithNames'
    );
```

▶

▶

▶

▶

⚙

⋮

📄

📄

📄

📄

```
INSERT INTO default.trips
SELECT
    trip_id,
    pickup_datetime,
    dropoff_datetime,
    pickup_longitude,
    pickup_latitude,
    dropoff_longitude,
    dropoff_latitude,
    passenger_count,
    trip_distance,
    fare_amount,
    extra,
    tip_amount,
    tolls_amount,
    total_amount,
    payment_type,
    pickup_ntaname,
    dropoff_ntaname
FROM
    gcs('https://storage.googleapis.com/clickhouse-public-datasets/nyc-taxi/trips_{0..2}.gz',
    'TabSeparatedWithNames'
    );
```

Statistics 1 ×

Name	Value	
Updated Rows	3751098	
Query	INSERT INTO default.trips	
	SELECT	
	trip_id,	
	pickup_datetime,	
	dropoff_datetime,	
	pickup_longitude,	
	pickup_latitude,	
	dropoff_longitude,	
	dropoff_latitude,	
	passenger_count,	
	trip_distance,	
	fare_amount,	
	extra,	
	tip_amount,	
	tolls_amount,	
	total_amount,	
	payment_type,	
	pickup_ntaname,	
	dropoff_ntaname	
	FROM	
	gcs('https://storage.googleapis.com/clickhouse-public-datasets/nyc-taxi/trips_{0..2}.gz',	
	'TabSeparatedWithNames'	
	)	
Start time	Mon Nov 25 21:59:17 MSK 2024	
Finish time	Mon Nov 25 21:59:27 MSK 2024	

## Проверка выполнения

```
SELECT * FROM system.clusters;
```

**SELECT \* FROM system.clusters;**

clusters 1 ×

Enter a SQL expression to filter results (use Ctrl+Space)

	cluster	shard_num	shard_weight	internal_replication	replica_num	host_name	host_address	port
1	default	1	1	0	1	localhost	127.0.0.1	9,000
2	sharded_cluster_sh4r3	1	1	1	1	clickhouse10	172.21.0.9	9,000
3	sharded_cluster_sh4r3	1	1	1	2	clickhouse11	172.21.0.8	9,000
4	sharded_cluster_sh4r3	1	1	1	3	clickhouse12	172.21.0.11	9,000
5	sharded_cluster_sh4r3	2	1	1	1	clickhouse13	172.21.0.7	9,000
6	sharded_cluster_sh4r3	2	1	1	2	clickhouse14	172.21.0.10	9,000
7	sharded_cluster_sh4r3	2	1	1	3	clickhouse15	172.21.0.6	9,000
8	sharded_cluster_sh4r3	3	1	1	1	clickhouse16	172.21.0.13	9,000
9	sharded_cluster_sh4r3	3	1	1	2	clickhouse17	172.21.0.3	9,000
10	sharded_cluster_sh4r3	3	1	1	3	clickhouse18	172.21.0.4	9,000
11	sharded_cluster_sh4r3	4	1	1	1	clickhouse19	172.21.0.12	9,000
12	sharded_cluster_sh4r3	4	1	1	2	clickhouse20	172.21.0.14	9,000
13	sharded_cluster_sh4r3	4	1	1	3	clickhouse21	172.21.0.5	9,000

```
SHOW CREATE TABLE default.trips
```

**SHOW CREATE TABLE default.trips**

CREATE 1 ×

Enter a SQL expression to filter results (use Ctrl+Space)

	statement	Value
1	CREATE TABLE default.trips( `trip_id` UInt32, `pickup_datetime` DateTime, `dropoff_datetime` DateTime, `pickup_longitude` Nullable(Float64), `pickup_latitude` Nullable(Float64), `dropoff_longitude` Nullable(Float64), `dropoff_latitude` Nullable(Float64), `passenger_count` UInt8, `trip_distance` Float32, `fare_amount` Float32, `extra` Float32, `tip_amount` Float32, `tolls_amount` Float32, `total_amount` Float32, `payment_type` Enum8('CSH' = 1, 'CRE' = 2, 'NOC' = 3, 'DIS' = 4, 'UNK' = 5), `pickup_ntaname` LowCardinality(String), `dropoff_ntaname` LowCardinality(String) ) ENGINE = Distributed('sharded_cluster_sh4r3', 'default', 'trips_local', rand())	CREATE TABLE default.trips ( `trip_id` UInt32, `pickup_datetime` DateTime, `dropoff_datetime` DateTime, `pickup_longitude` Nullable(Float64), `pickup_latitude` Nullable(Float64), `dropoff_longitude` Nullable(Float64), `dropoff_latitude` Nullable(Float64), `passenger_count` UInt8, `trip_distance` Float32, `fare_amount` Float32, `extra` Float32, `tip_amount` Float32, `tolls_amount` Float32, `total_amount` Float32, `payment_type` Enum8('CSH' = 1, 'CRE' = 2, 'NOC' = 3, 'DIS' = 4, 'UNK' = 5), `pickup_ntaname` LowCardinality(String), `dropoff_ntaname` LowCardinality(String) ) ENGINE = Distributed('sharded_cluster_sh4r3', 'default', 'trips_local', rand())