Cloud Data Warehouse

Getting Started

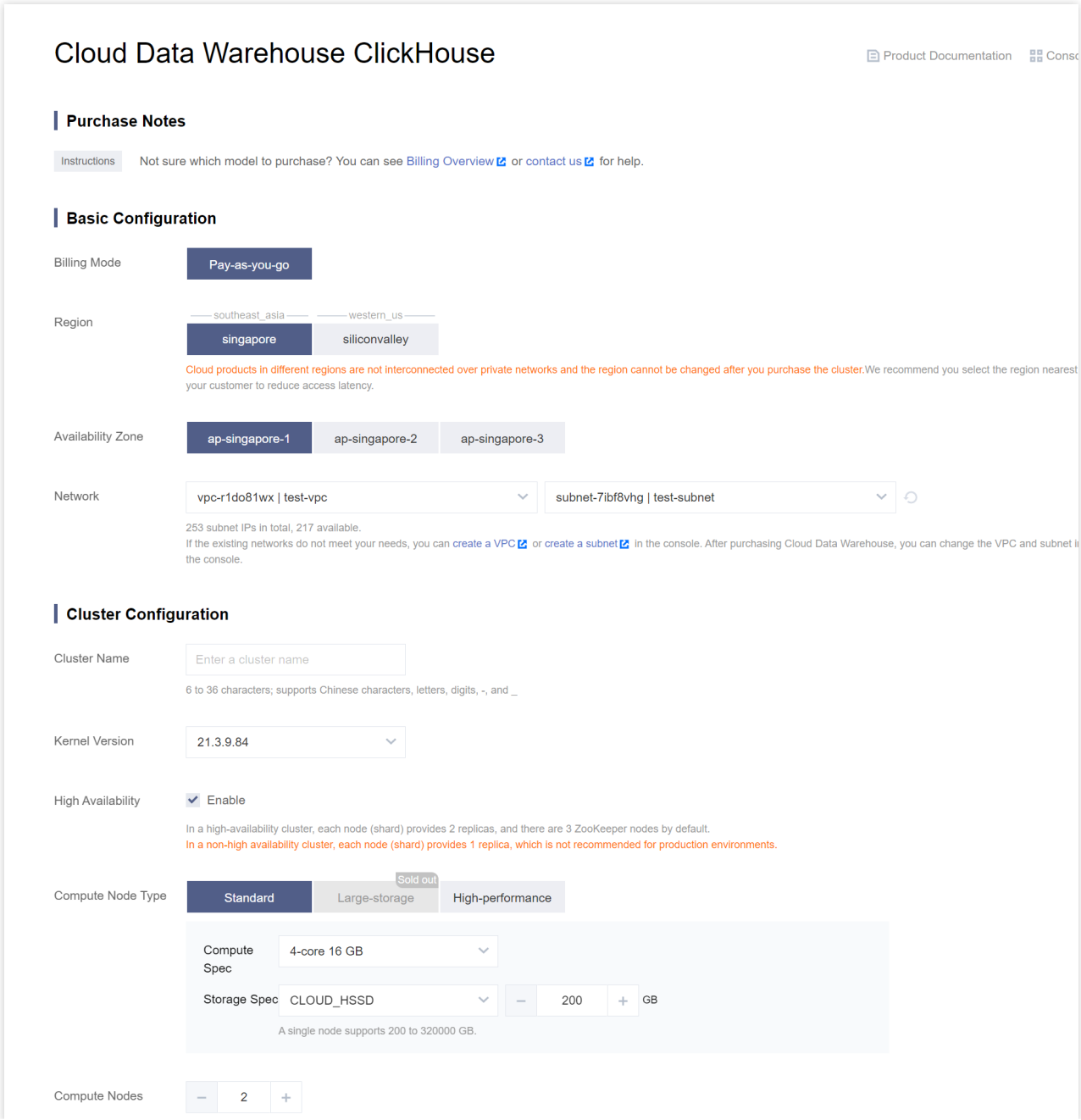
Product Documentation

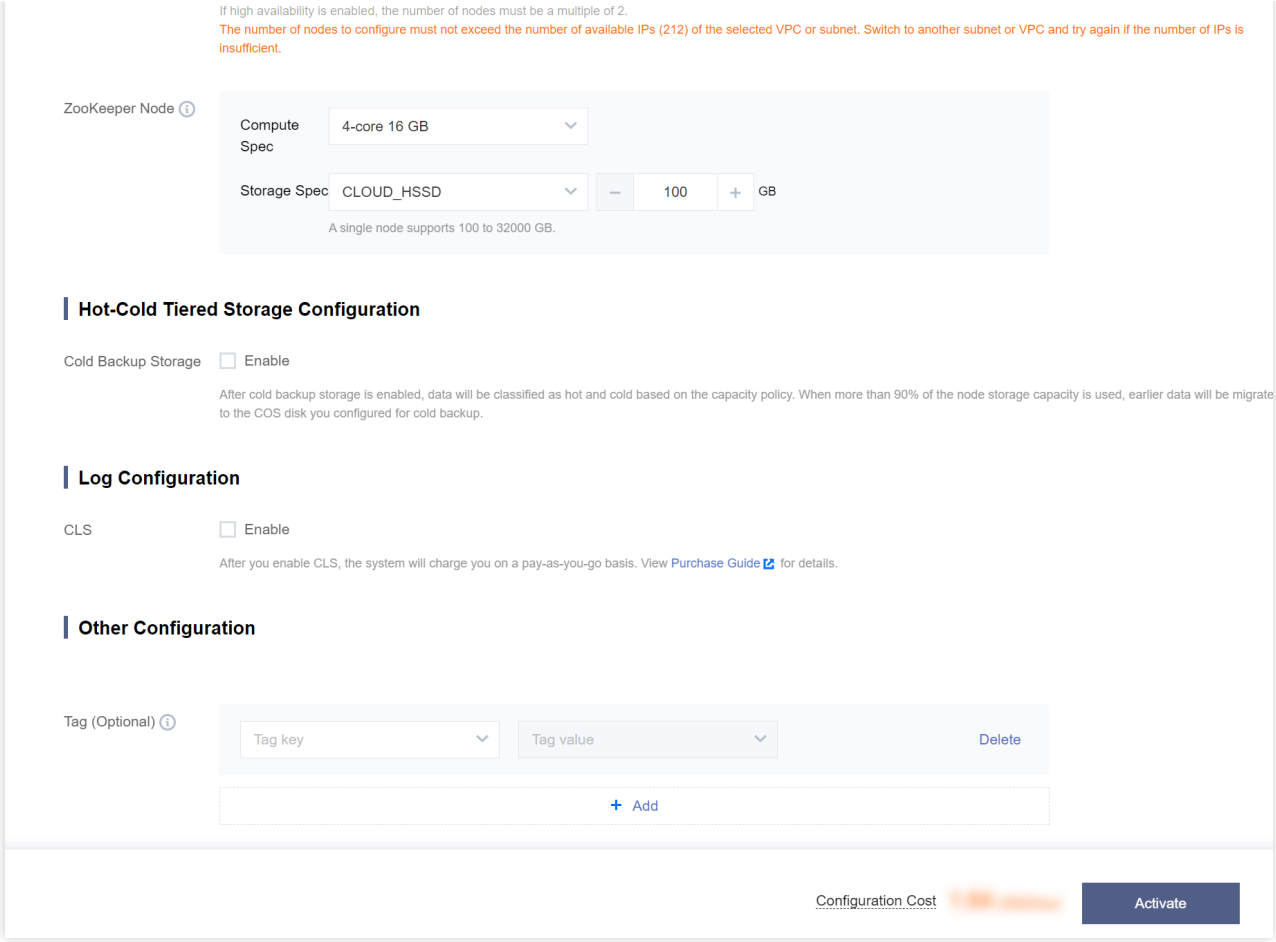
Getting Started

Last updated：2024-01-19 16:46:19

Creating Cluster

1. Log in to the Cloud Data Warehouse overview page and click **Buy Now** or log in to the Cloud Data Warehouse console and click **New Cluster** with your Tencent Cloud account.
2. On the purchase page, configure and purchase a cluster as prompted. For more information on configuration items, see Configuration Items.





Configuration items

Configuration Item	Description
Billing Mode	Pay-as-you-go: postpaid, where a bill is generated hourly based on resource usage and then you pay for what you use.
Region	Currently, Cloud Data Warehouse is available in the Shanghai, Hong Kong (China), Beijing, Guangzhou, Singapore, and Silicon Valley regions. We recommend you select a region closest to your users, and you cannot change the region after the purchase.
Availability Zone	Select availability zones in different regions as needed on the purchase page.
Network	A VPC is an isolated, highly secure, and dedicated network environment. You can create a VPC and subnet or select an existing one.
High Availability	In HA mode, each shard has two replicas; in non-HA mode, each shard has only one replica, where the entire cluster will fail if the replica fails. Therefore, we recommend you use the HA mode for production environments.

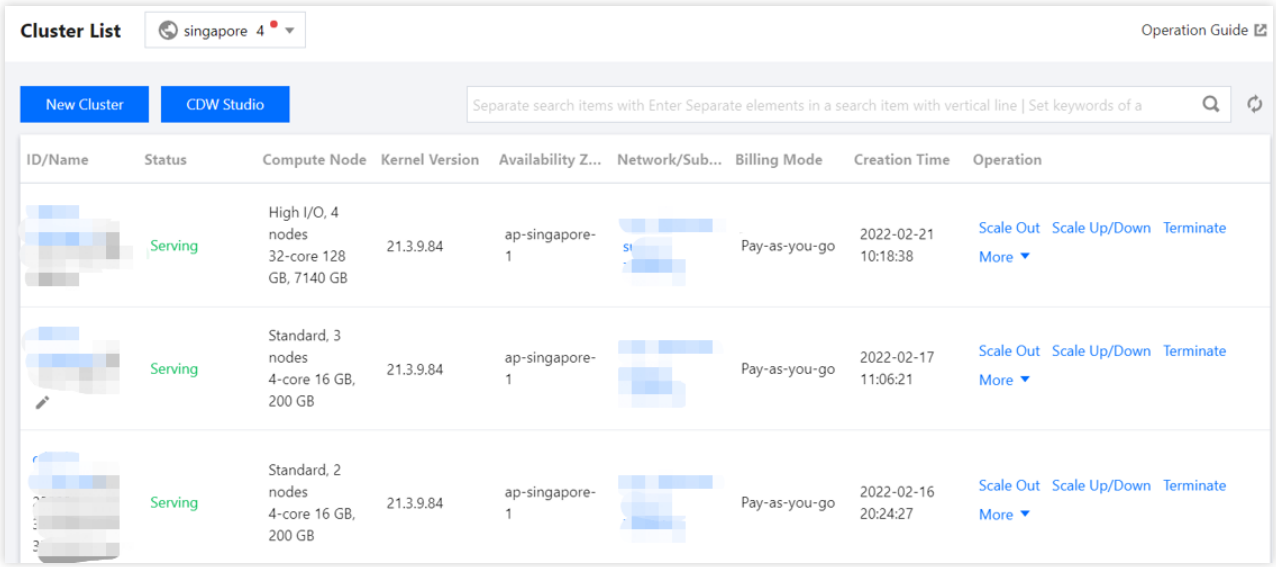
Compute Node Type	There are three types of compute nodes: Standard: 4-core 16 GB, 8-core 32 GB, 16-core 64 GB, 24-core 96 GB, 32-core 128 GB, 64-core 256 GB, 90-core 224 GB, and 128-core 256 GB. Storage-Optimized: 32-core 128 GB (with twelve 3720 GB SATA HDDs) and 64-core 256 GB (with twenty-four 3720 GB SATA HDDs), 84-core 320 GB (with twenty-four 3720 GB SATA HDDs). High-Performance: 32-core 128 GB (with two 3570 GB NVMe SSDs), 64-core 256 GB (with four 3570 GB NVMe SSDs), and 84-core 320 GB (with four 3570 GB NVMe SSDs). The higher the specification, the better the performance. You can select an appropriate specification as needed.
ZooKeeper Node Type	There are 4-core 16 GB, 8-core 32 GB, 16-core 64 GB, 24-core 96 GB, 32-core 128 GB, 64-core 256 GB, 90-core 224 GB, and 128-core 256 GB ZooKeeper nodes. The heavier the load, the higher the specification needed. You can select an appropriate specification as needed.

Note:

You can enable dedicated Grafana monitoring, cluster logging, tiered storage of hot/cold data, and auto-renewal features as needed.

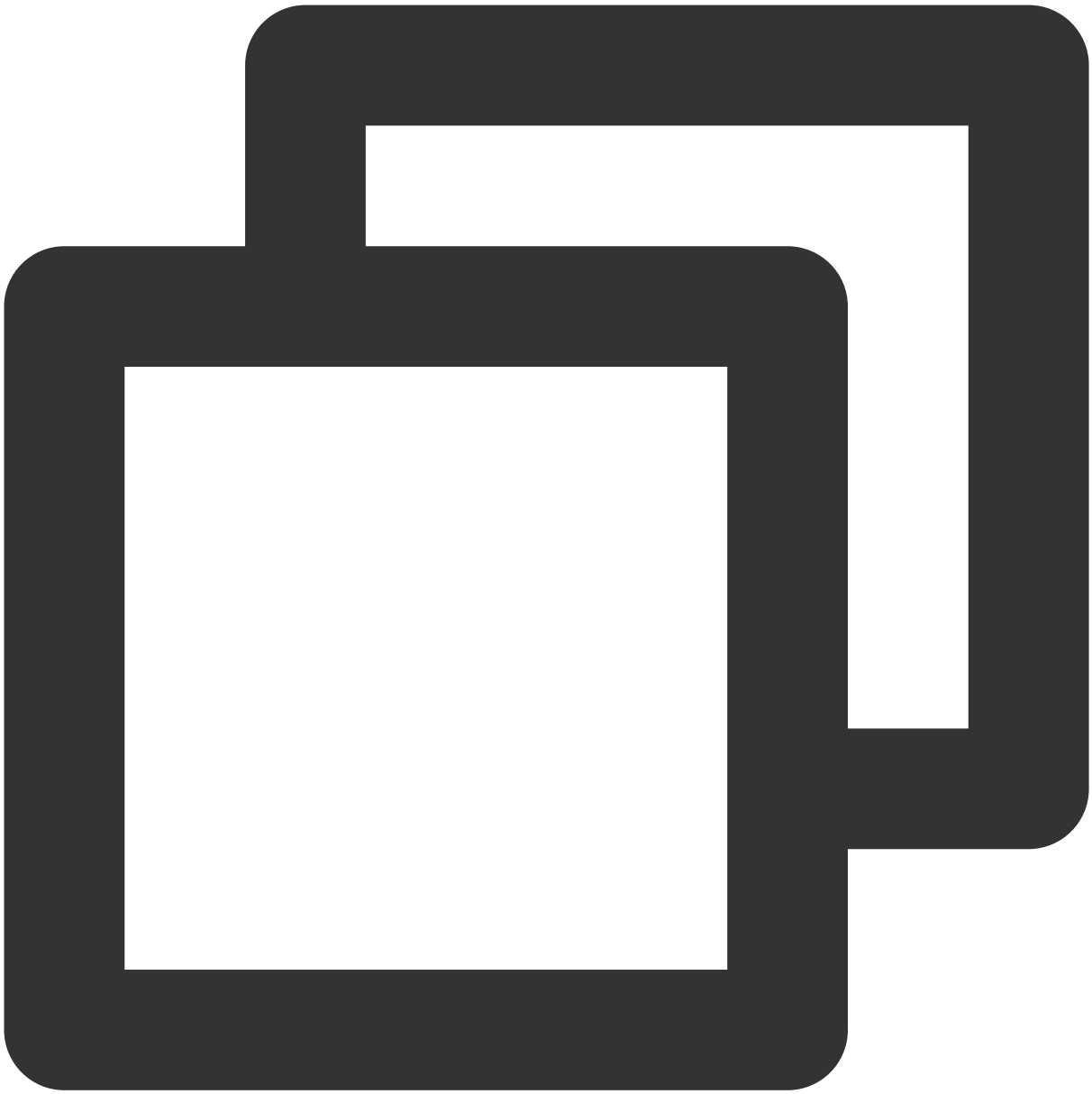
Viewing Cluster Information

After the cluster is created, go to the Cloud Data Warehouse console, select the region where the cluster resides, and view the cluster status and information as shown below:



Using ClickHouse

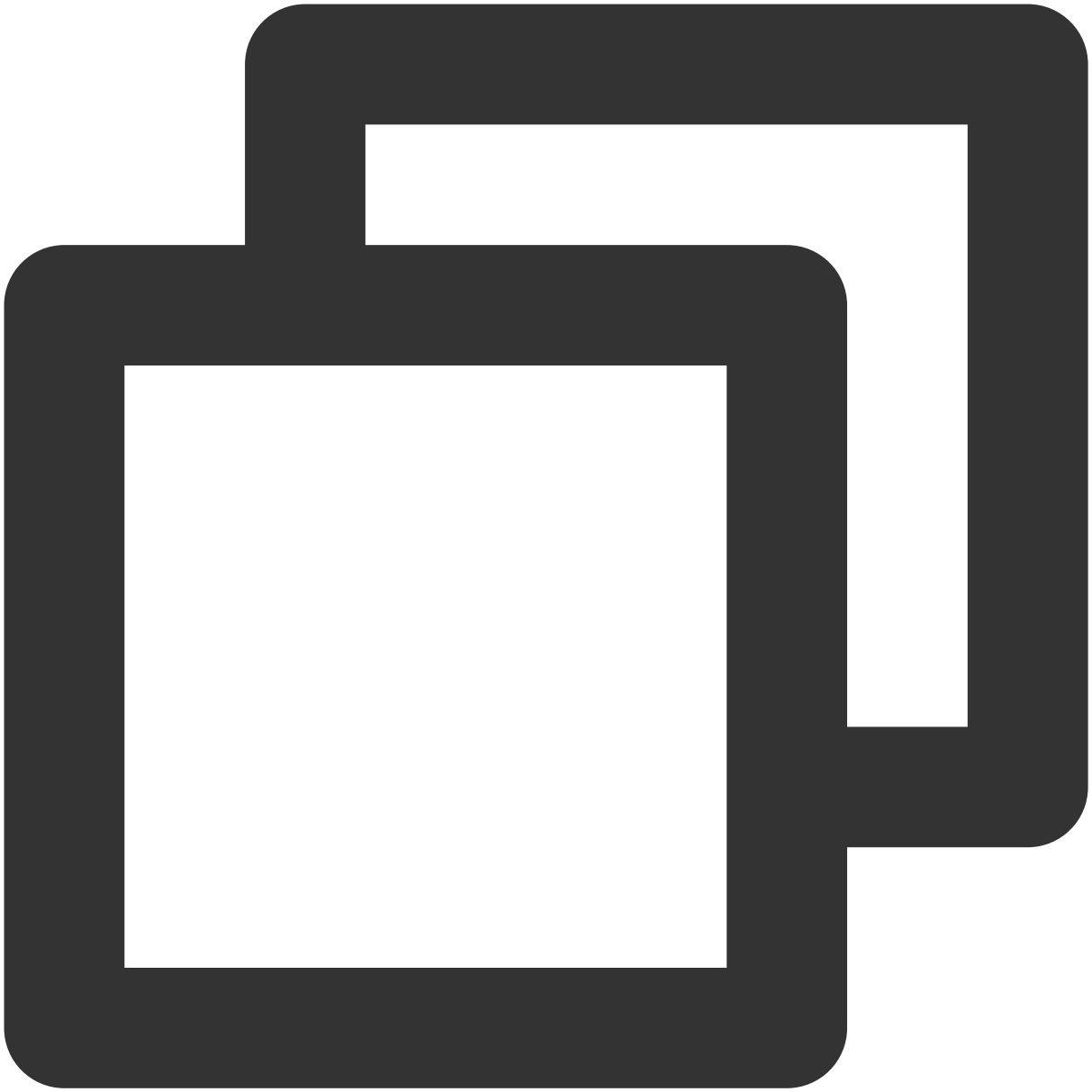
Import a data file to a ClickHouse cluster and view the imported data. Prepare the following `account.csv` file:



```
AccountId, Name, Address, Year
1, 'GHua', 'WuHan Hubei', 1990
2, 'SLiu', 'ShenZhen Guangzhou', 1991
3, 'JPong', 'Chengdu Sichuan', 1992
```

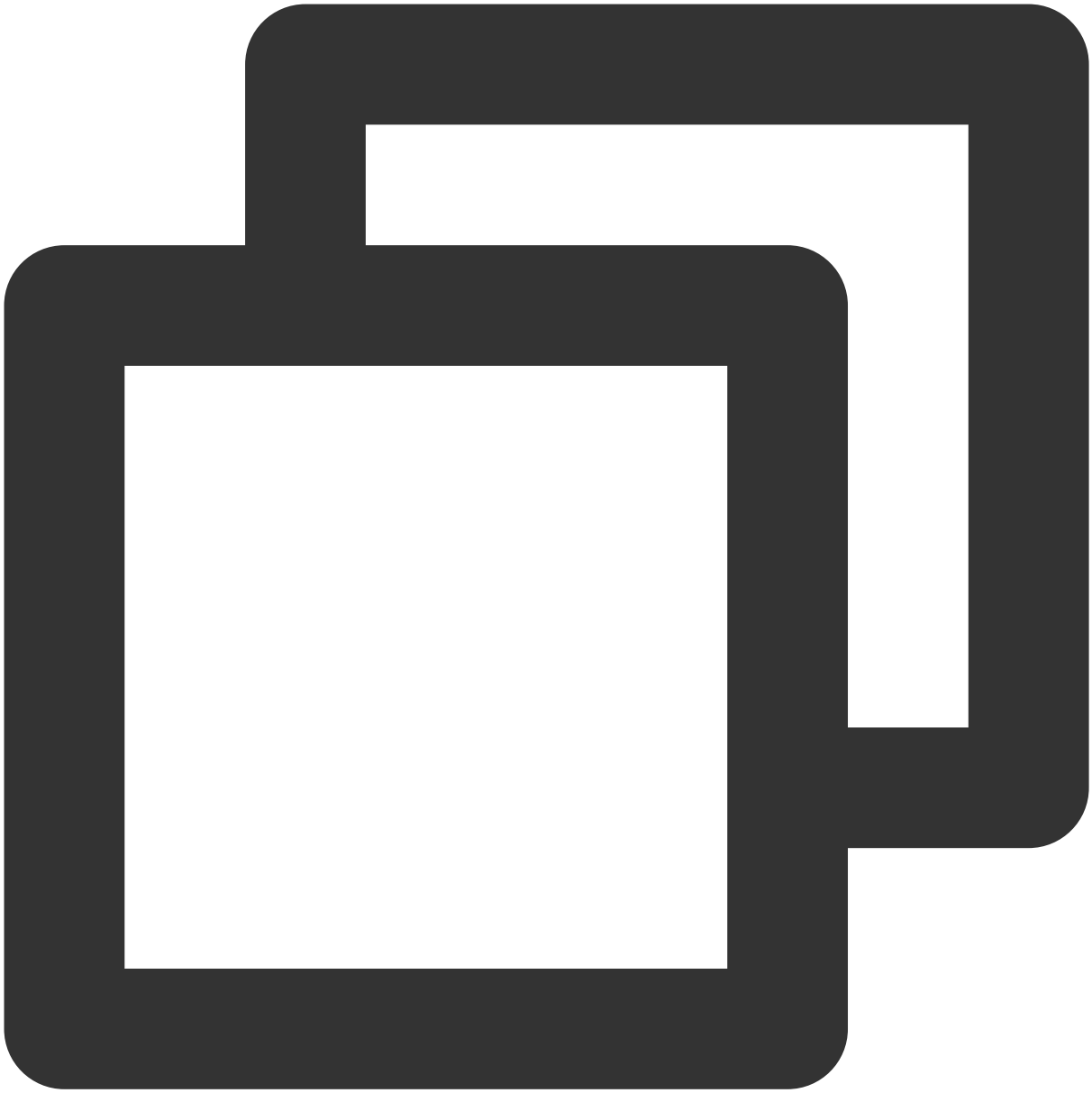
Connecting to cluster

1. Download a ClickHouse client.



```
wget https://repo.yandex.ru/clickhouse/rpm/stable/x86_64/clickhouse-client-20.7.2.3
wget https://repo.yandex.ru/clickhouse/rpm/stable/x86_64/clickhouse-common-static-2
```

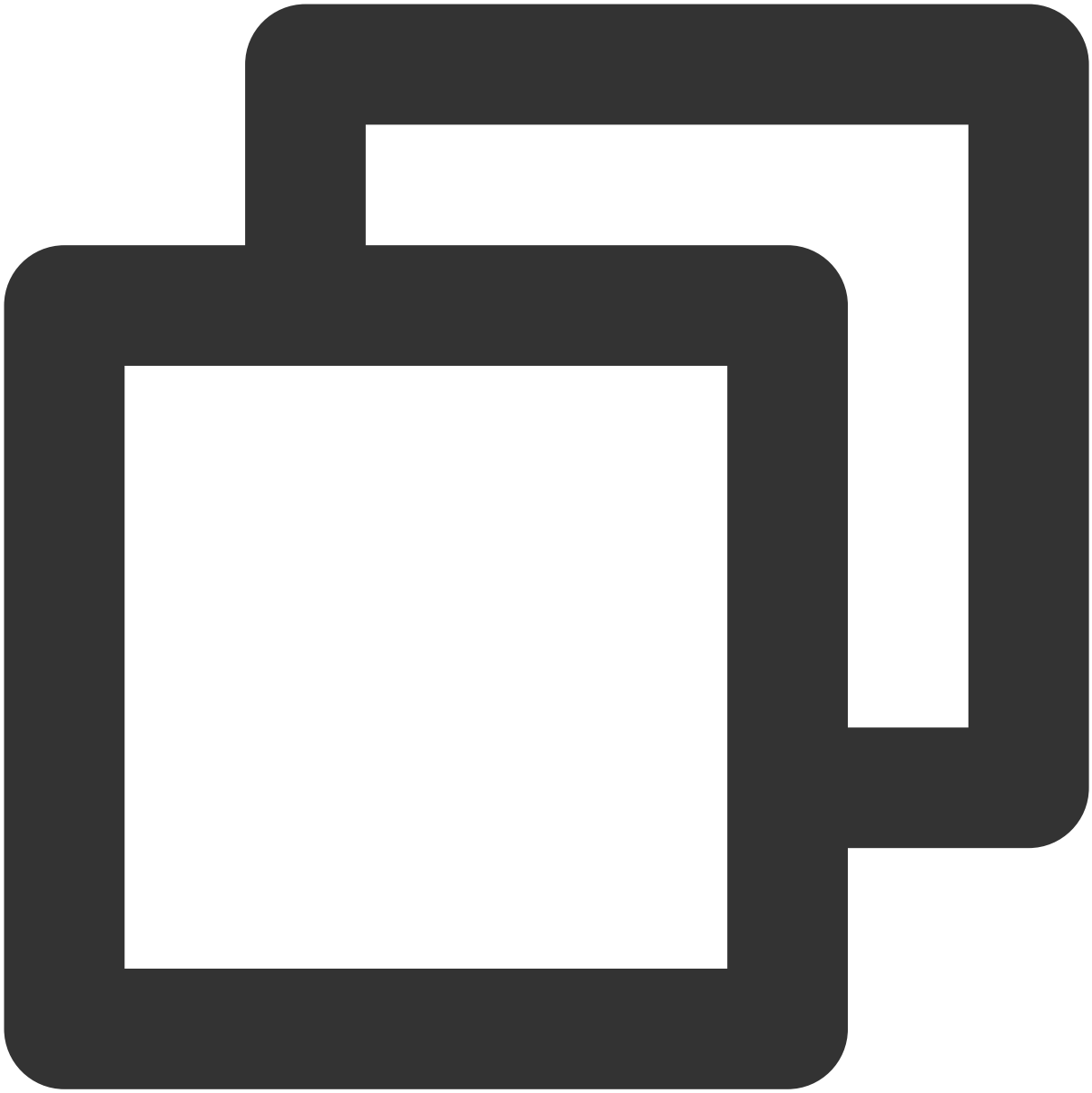
Install the client.



```
rpm -ivh *.rpm
```

Access the cluster from the client.

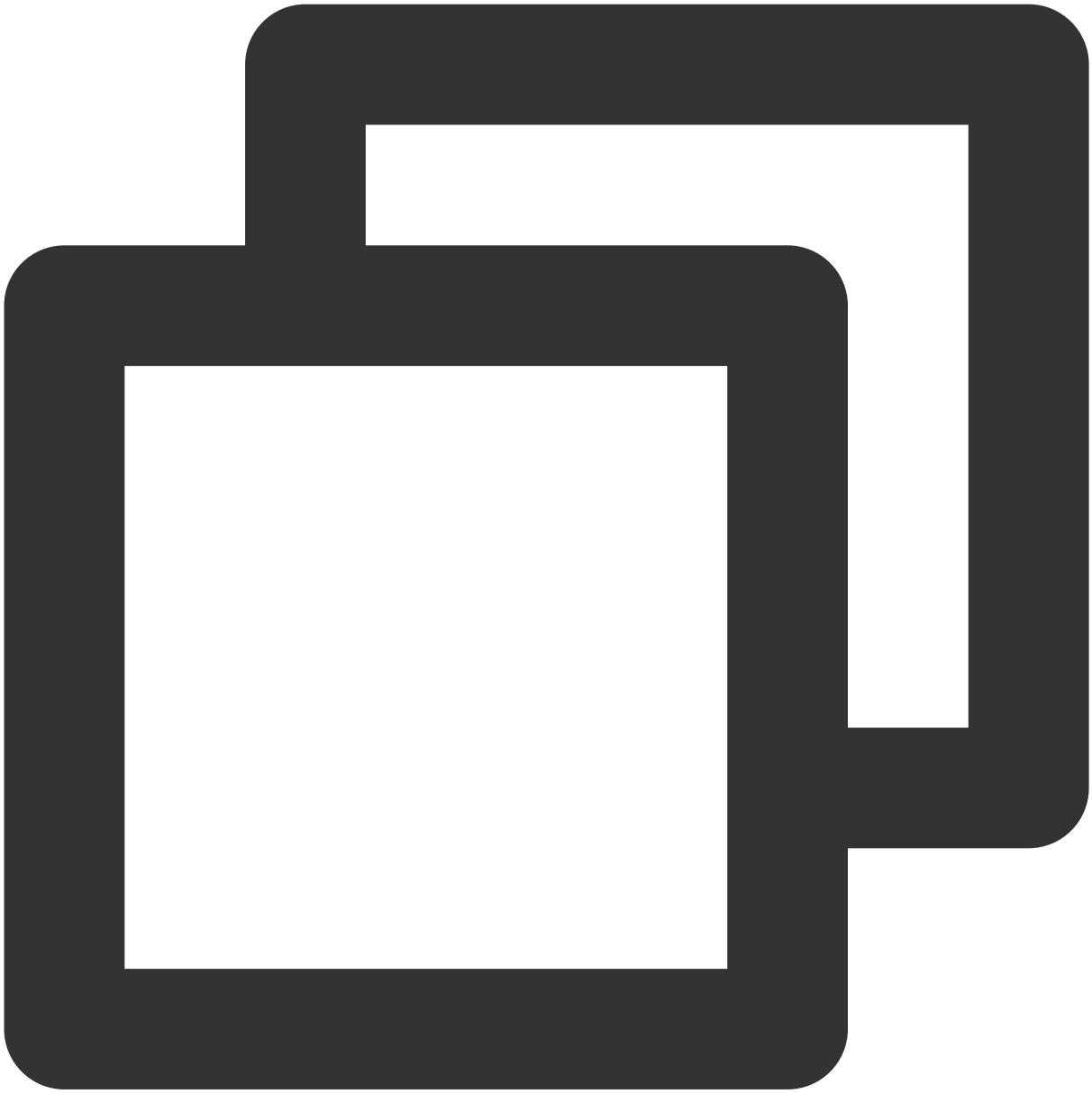
View the node IP address in the console and select the TCP port `9000` .



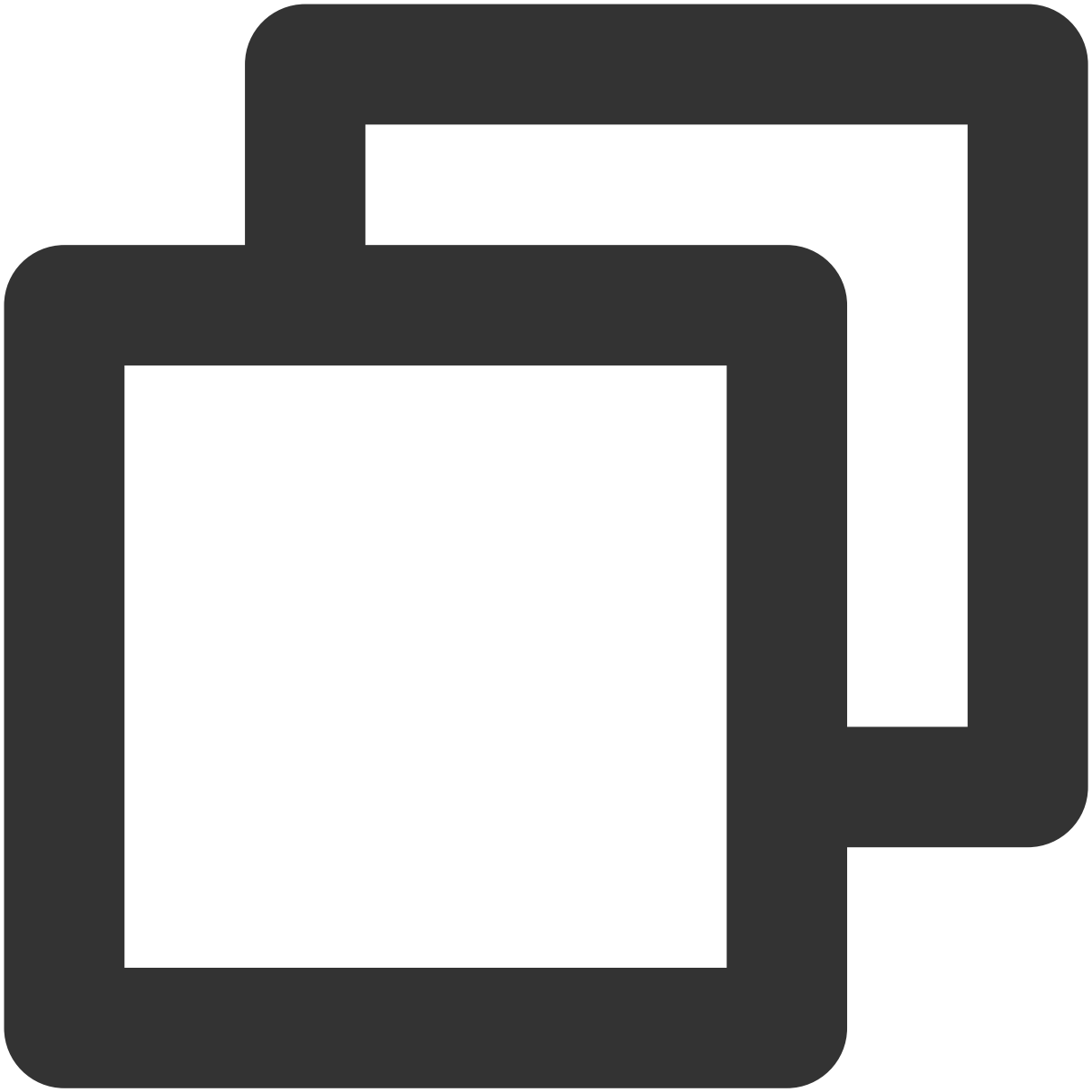
```
clickhouse-client -hxxx.xxx.xxx.xxx --port 9000
```

```
[root@VM-1-18-centos ~]# clickhouse-client -h [redacted] --port 9000
ClickHouse client version 20.7.2.30 (official build).
Connecting to [redacted]:9000 as user default.
Connected to ClickHouse server version 20.7.2 revision 54437.

172.16.1.45 :)
```

Select the HTTP port `8123` and get the specific access IP address in **Cluster Access Address (HTTP)** on the cluster details page.
Query and confirm the engine version of the cluster.



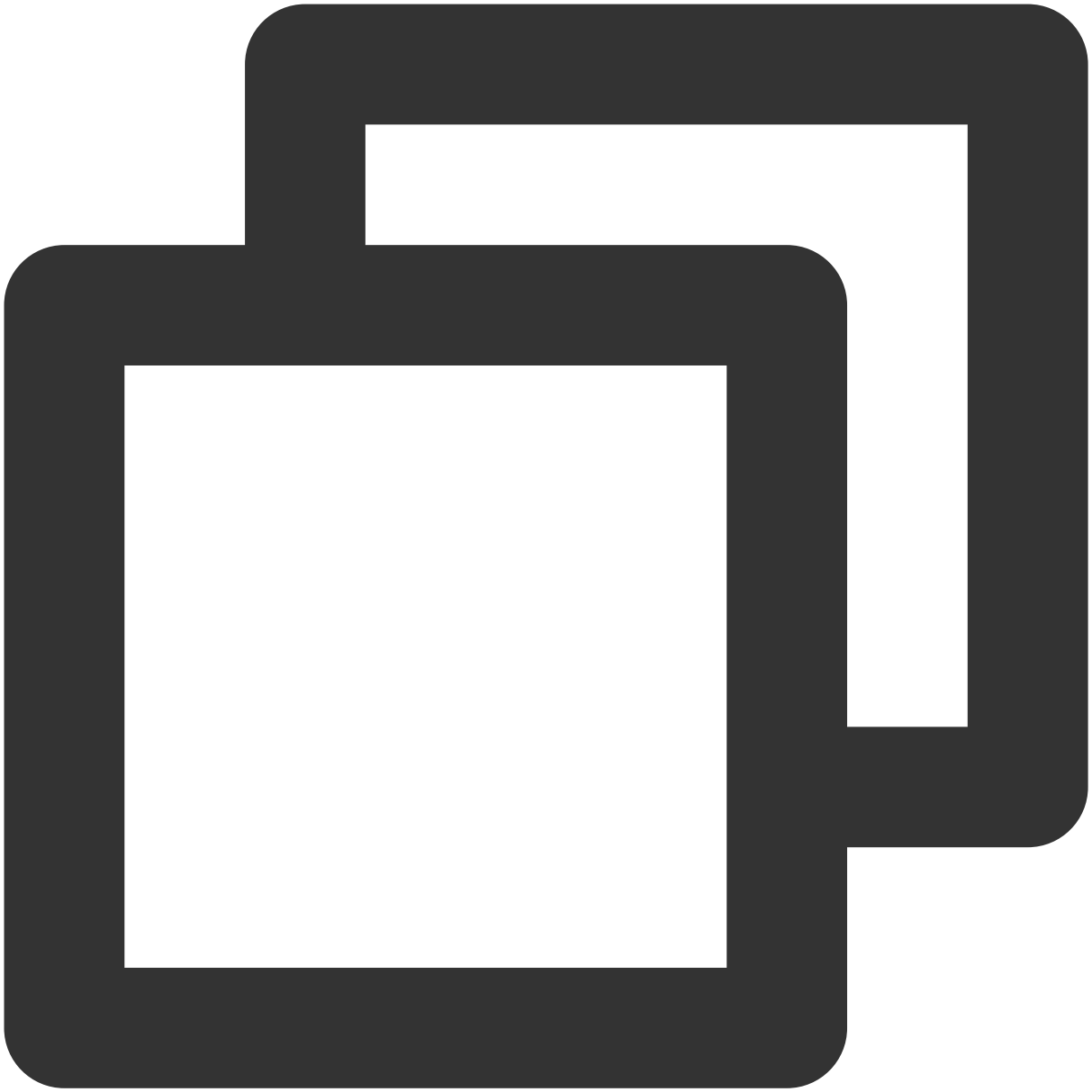
```
echo "select version()=21.3.9.83"  |  curl  'http://xxx.xxx.xxx.xxx:8123/' --data-b
```



```
echo "select version()"  |  curl  'http://xxx.xxx.xxx.xxx:8123/' --data-binary @-
```

```
[root@VM-0-126-centos ~]# echo "select version()"  |  curl  'http:/        :8123/' --data-binary @-
21.3.9.83
[root@VM-0-126-centos ~]# echo "select version()='21.3.9.83'"  |  curl  'http:/        :8123/' --data-binary @-
1
[root@VM-0-126-centos ~]# echo "select version()='20.3.9.83'"  |  curl  'http:/        :8123/' --data-binary @-
0
```

Query the system cluster.



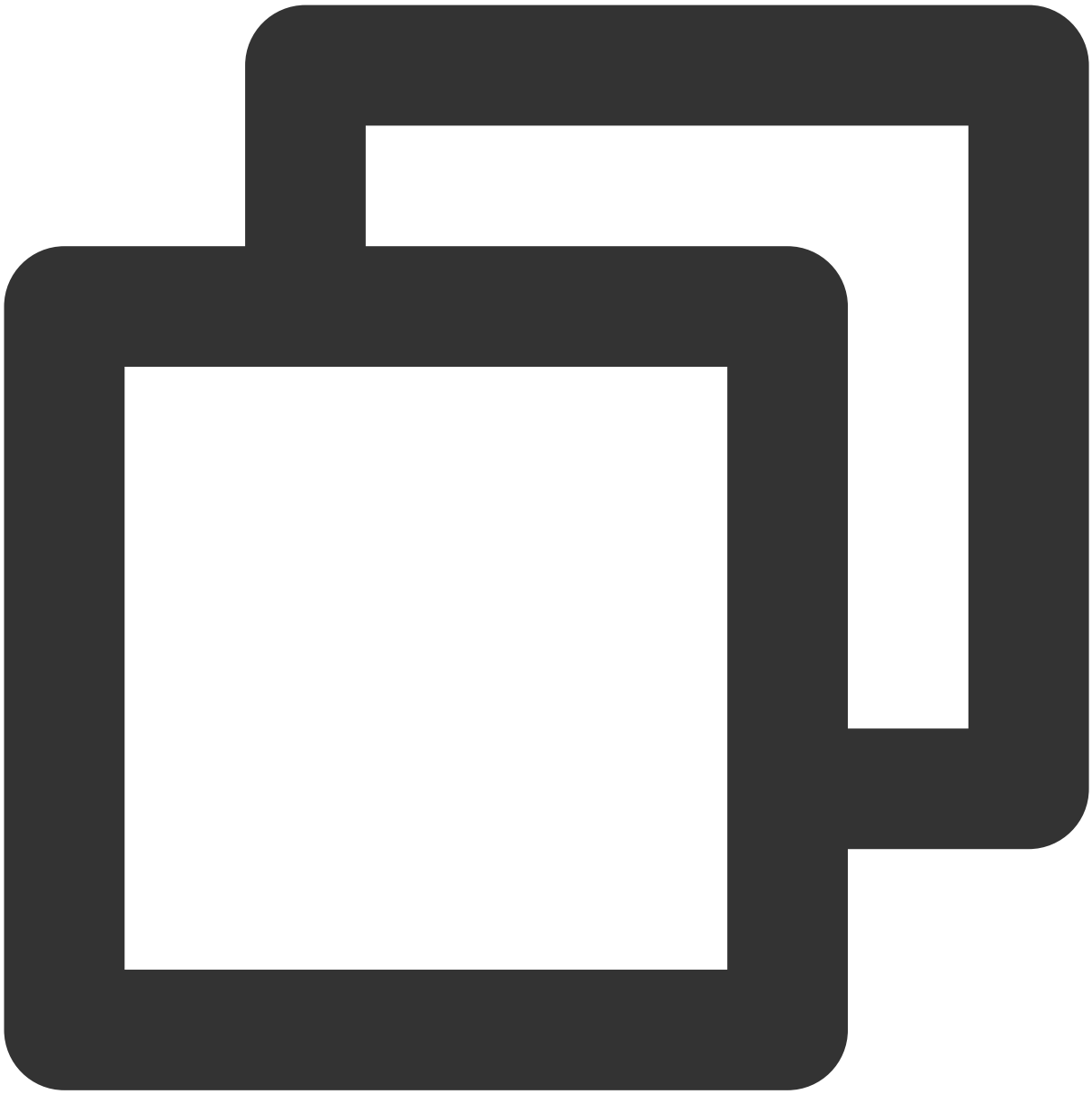
```
echo "select * from system.clusters"  |  curl  'http://xxx.xxx.xxx.xxx:8123/' --dat
```

```
[root@VM-0-126-centos ~]# echo "select * from system.clusters"  |  curl  'http://          :8123/' --data-binary @-
default_cluster 1       1       1                                       9000    1       default         0       0
default_cluster 1       1       2                                       9000    0       default         0       0
```

Creating data table

Use the ClickHouse client to connect to the cluster and create databases and tables.

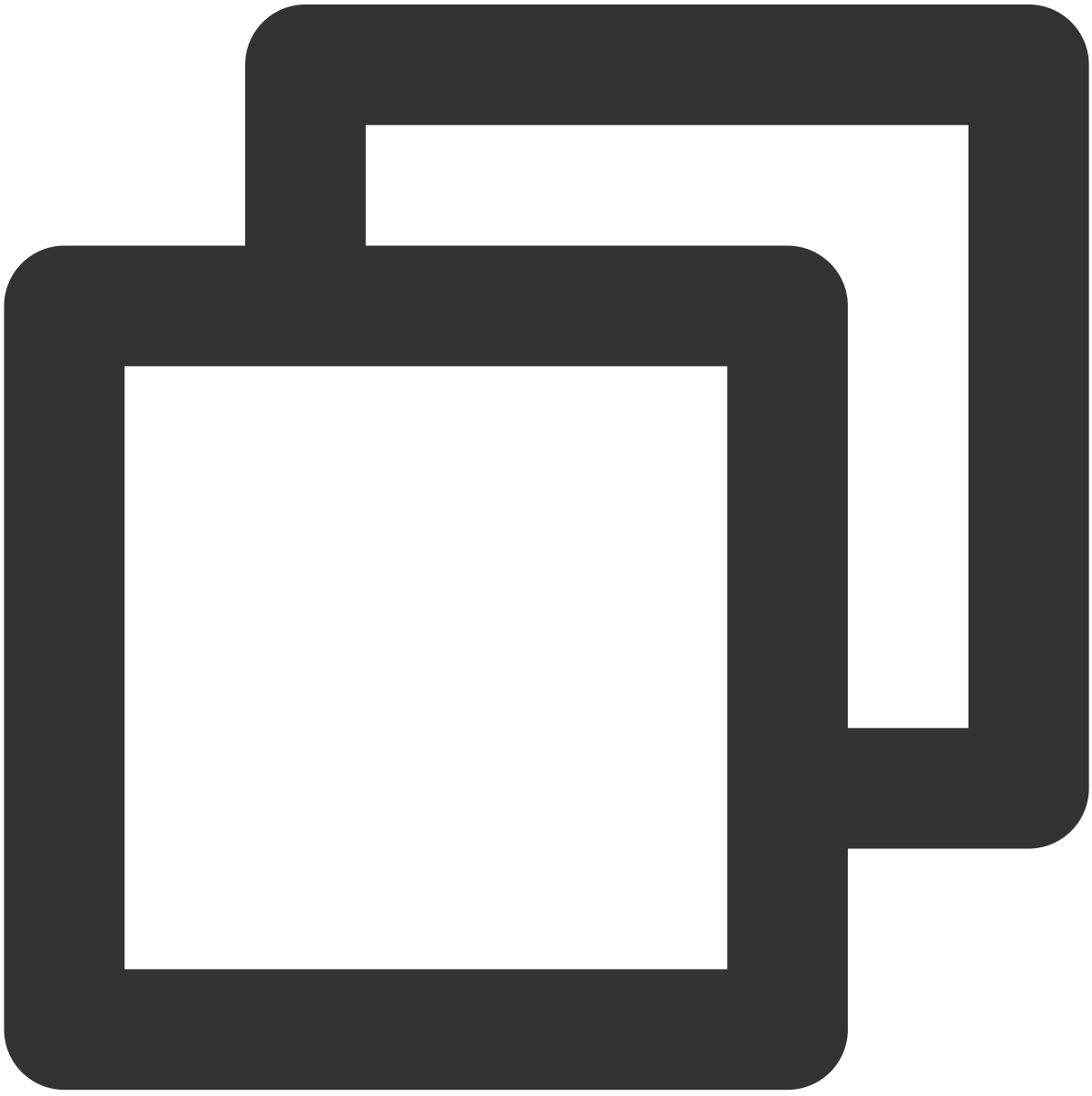
Create a database in HA mode



```
CREATE DATABASE IF NOT EXISTS testdb ON CLUSTER default_cluster;
```

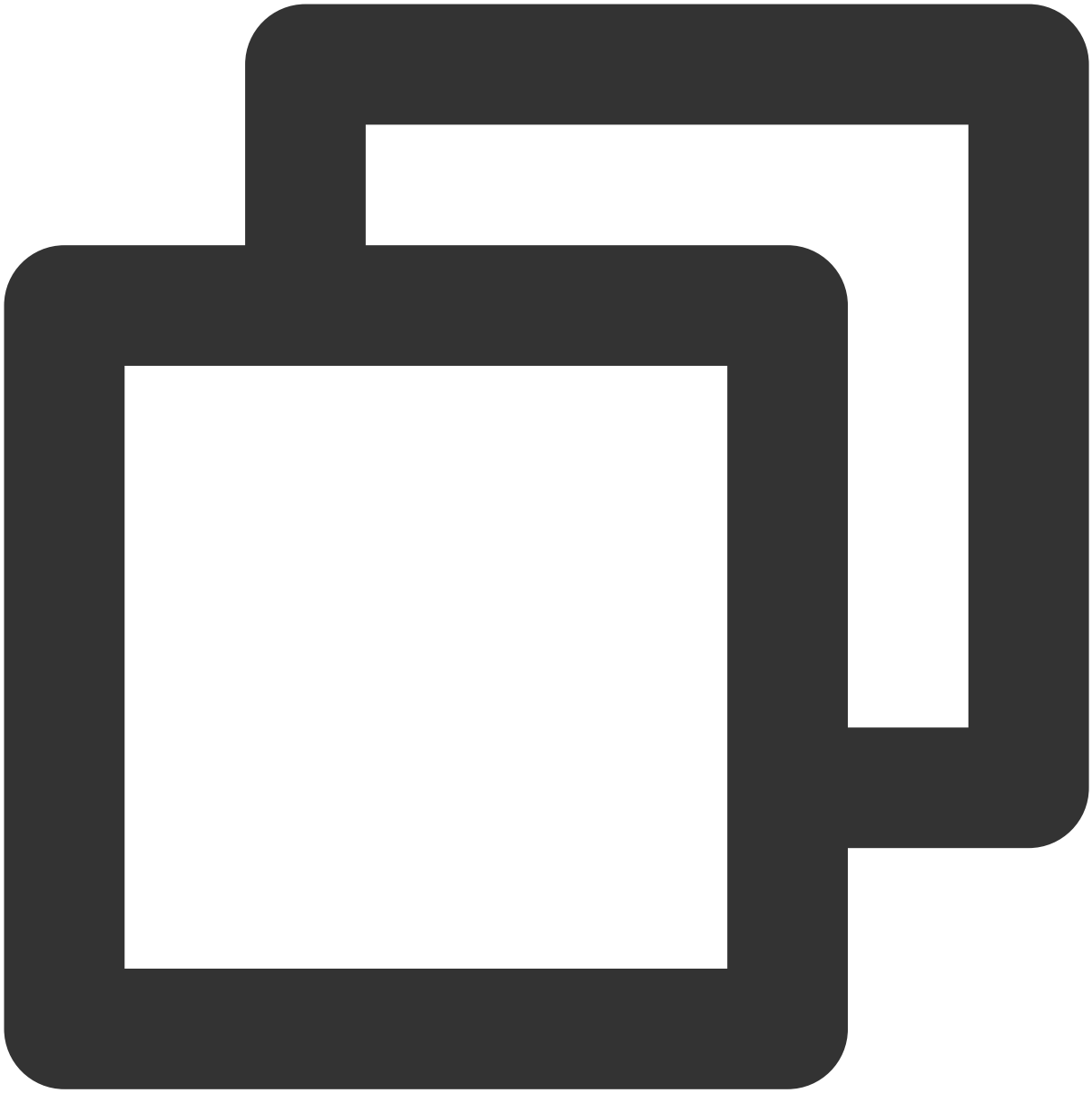
```
) CREATE DATABASE IF NOT EXISTS testdb ON CLUSTER default_cluster;

CREATE DATABASE IF NOT EXISTS testdb ON CLUSTER default_cluster

┌─host──────┬─port─┬─status─┬─error─┬─num_hosts_remaining─┬─num_hosts_active─┐
│ [redacted] │ 9000 │      0 │       │                   1 │                0 │
│ [redacted] │ 9000 │      0 │       │                   0 │                0 │
└───────────┴──────┴────────┴───────┴─────────────────────┴──────────────────┘
```

Create a table in HA mode

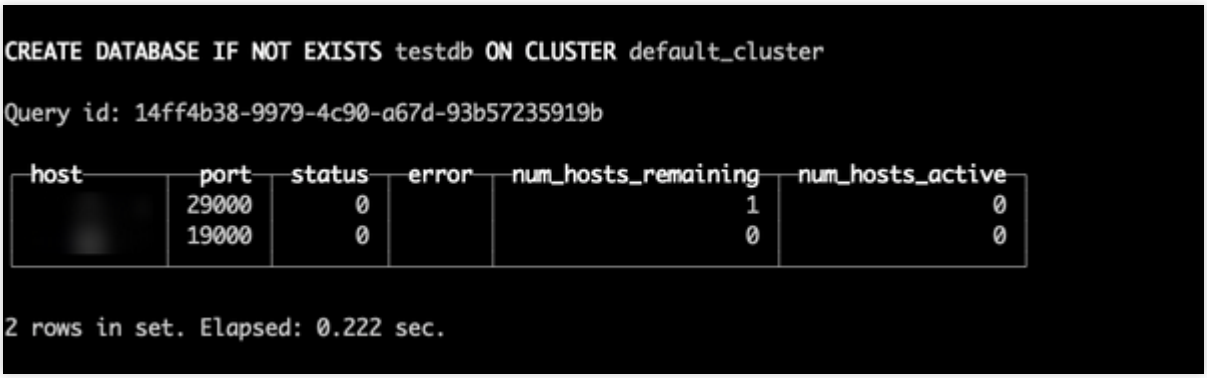


```
CREATE TABLE testdb.account ON CLUSTER default_cluster(accountid UInt16,name String
```

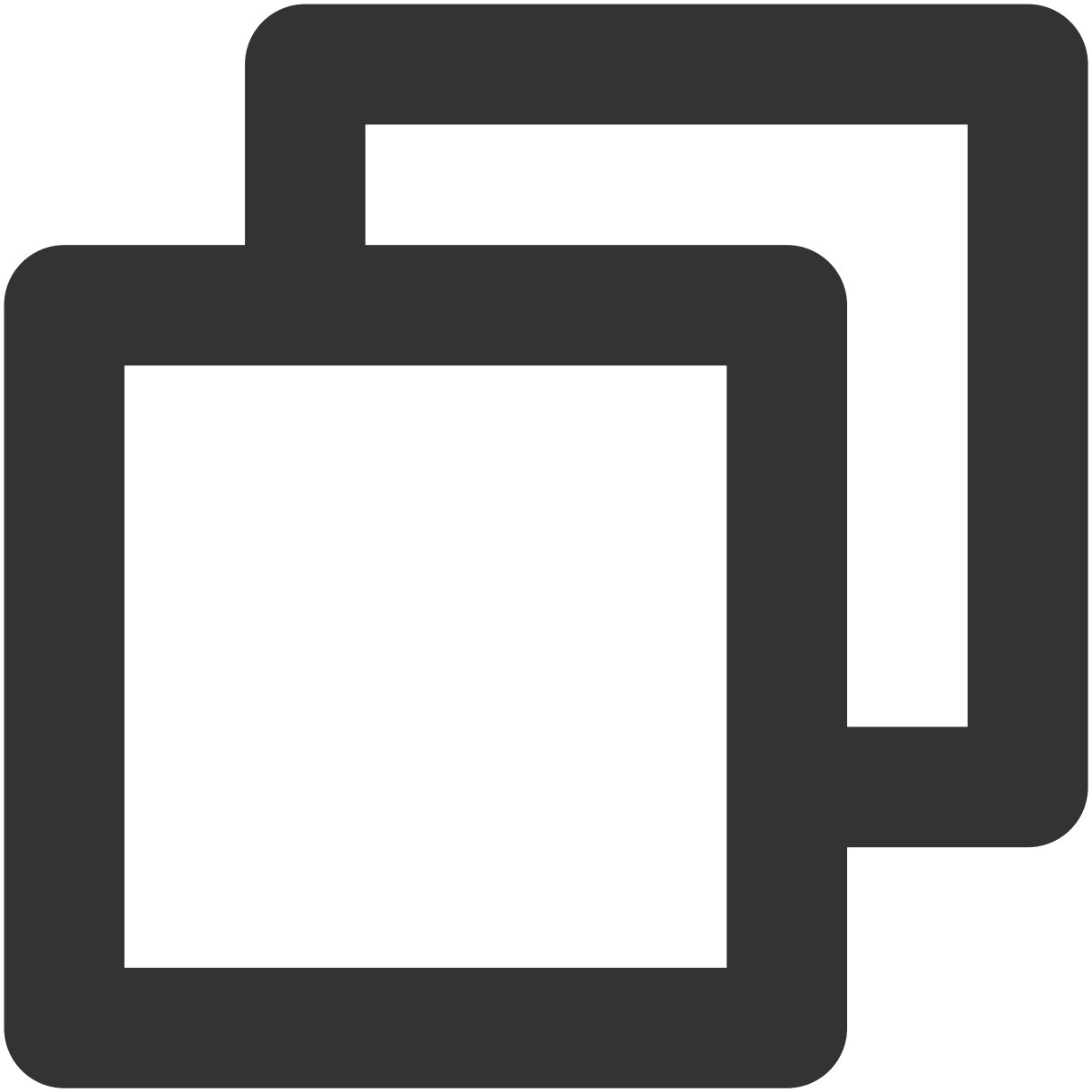
Create a database in non-HA mode



```
CREATE DATABASE IF NOT EXISTS testdb ON CLUSTER default_cluster;
```



Create a table in non-HA mode

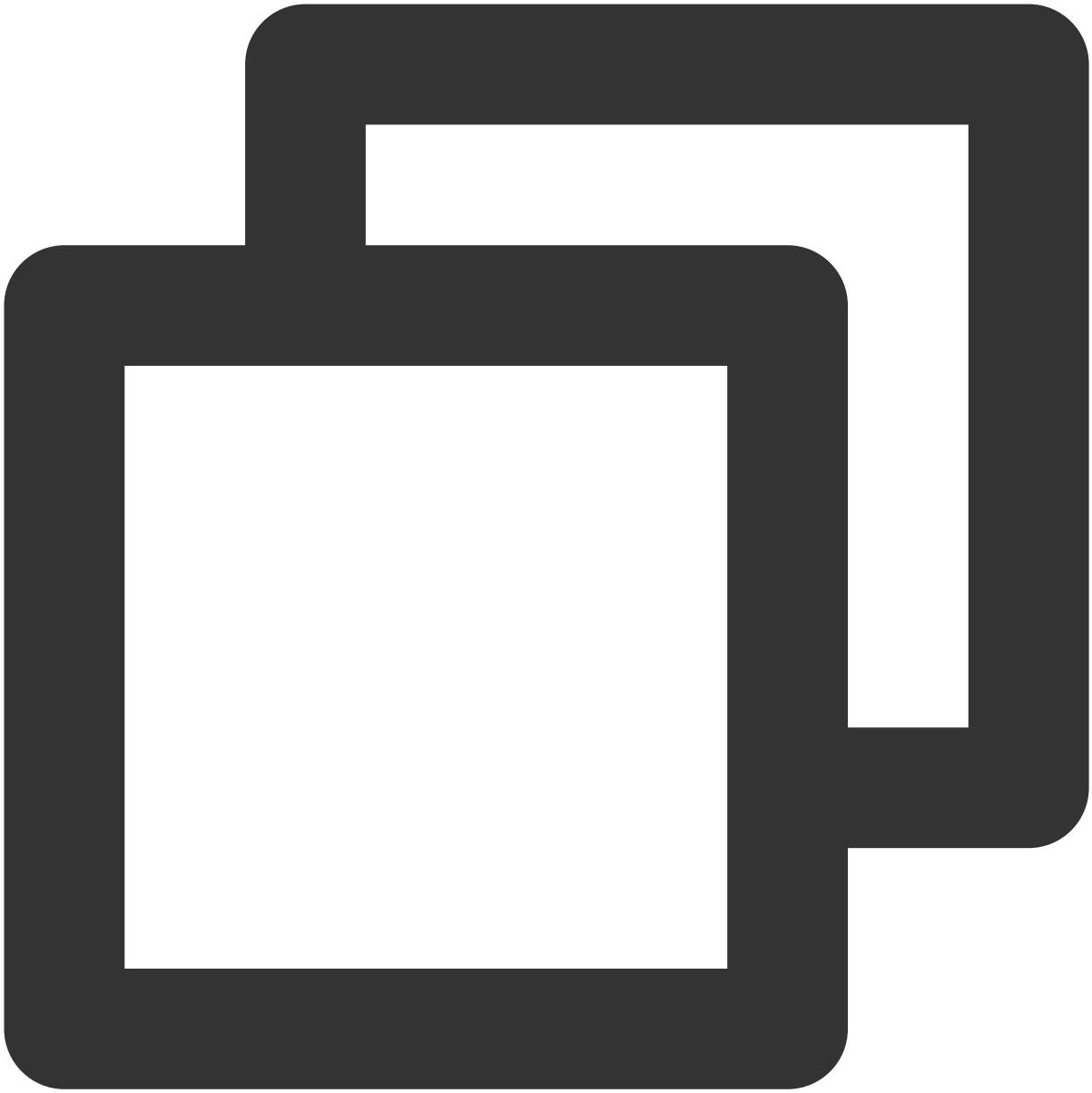


```
CREATE TABLE testdb.account ON CLUSTER default_cluster(accountid UInt16, name Strin
```

```
CREATE TABLE testdb.account ON CLUSTER default_cluster
(
    `accountid` UInt16,
    `name` String,
    `address` String,
    `year` UInt64
)
ENGINE = MergeTree
ORDER BY accountid

Query id: 6c79e383-2e1b-4922-a0e0-45cf2b8a0a46

┌─host──────┬─port─┬─status─┬─error─┬─num_hosts_remaining─┬─num_hosts_active─┐
│           │ 9000 │      0 │       │                   1 │                1 │
└───────────┴──────┴────────┴───────┴─────────────────────┴──────────────────┘
┌─host──────┬─port─┬─status─┬─error─┬─num_hosts_remaining─┬─num_hosts_active─┐
│ [redacted]│ 9000 │      0 │       │                   0 │                0 │
└───────────┴──────┴────────┴───────┴─────────────────────┴──────────────────┘

2 rows in set. Elapsed: 0.247 sec.
```

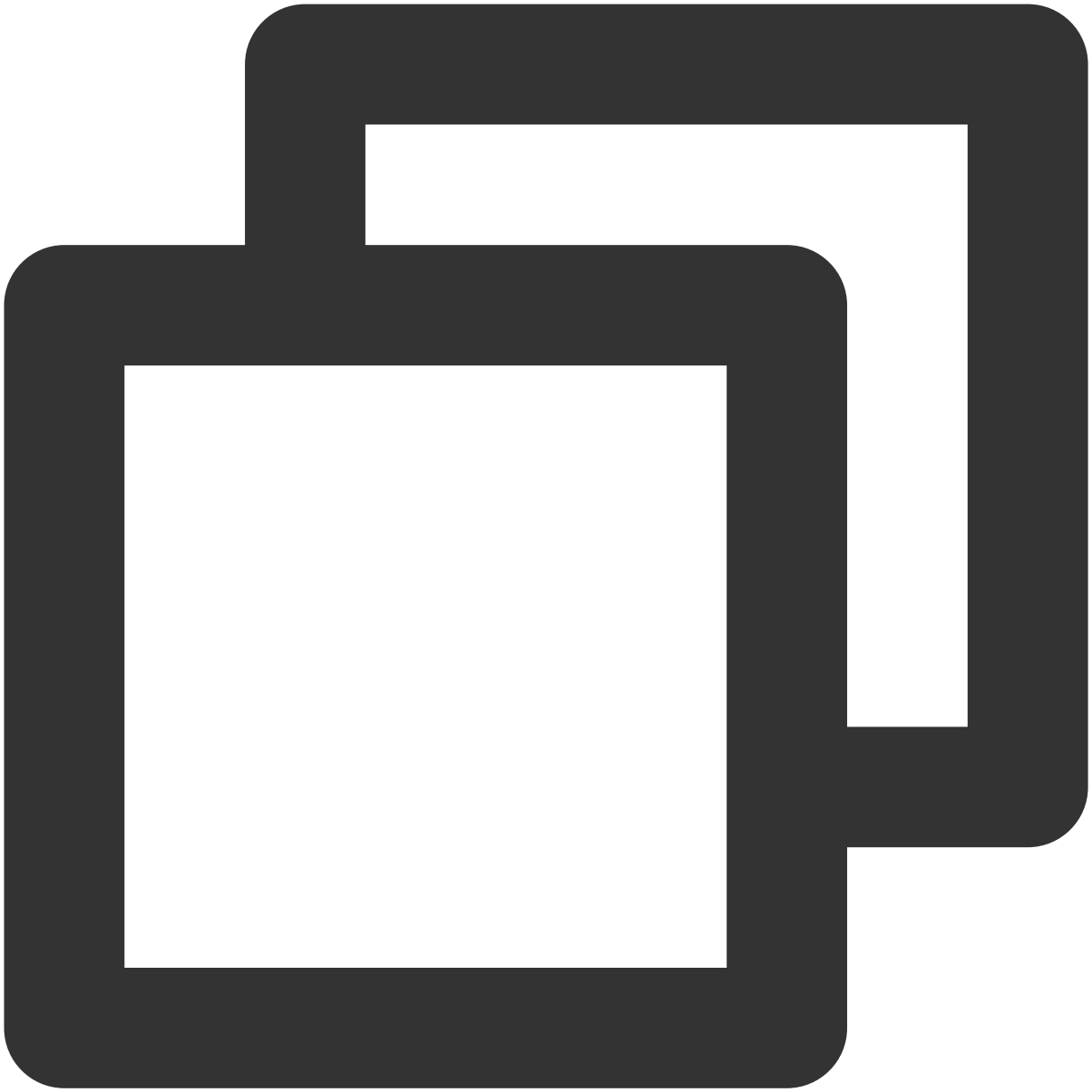
Importing data

Place the prepared data in the `/data` directory of the CVM instance connected to the ClickHouse cluster and run the following command to import the data.

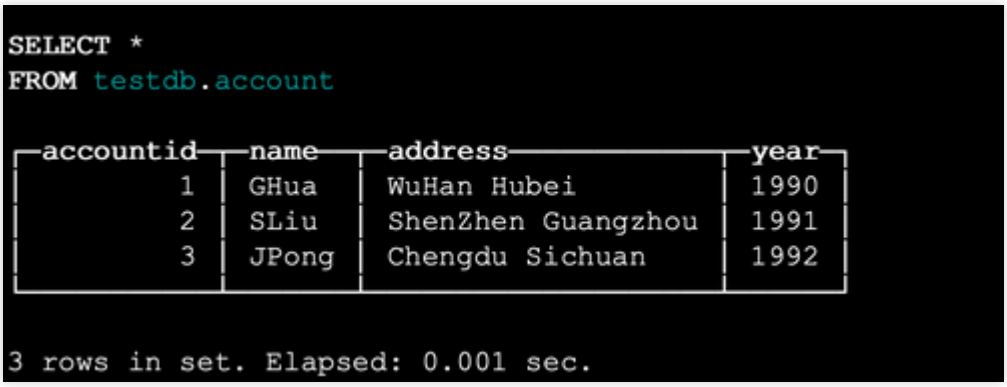


```
cat /data/account.csv | clickhouse-client - hxxx.xxx.xxx.xxx --database=testdb --qu
```

Querying data



```
select * from testdb.account;
```



```
SELECT *
FROM testdb.account

┌─accountid─┬─name──┬─address──────────┬─year─┐
│         1 │ GHua  │ WuHan Hubei      │ 1990 │
│         2 │ SLiu  │ ShenZhen Guangzhou │ 1991 │
│         3 │ JPong │ Chengdu Sichuan  │ 1992 │
└───────────┴───────┴──────────────────┴──────┘

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