

Cloud Data Warehouse

Getting Started

Product Documentation



Copyright Notice

©2013-2024 Tencent Cloud. All rights reserved.

Copyright in this document is exclusively owned by Tencent Cloud. You must not reproduce, modify, copy or distribute in any way, in whole or in part, the contents of this document without Tencent Cloud's the prior written consent.

Trademark Notice



All trademarks associated with Tencent Cloud and its services are owned by Tencent Cloud Computing (Beijing) Company Limited and its affiliated companies. Trademarks of third parties referred to in this document are owned by their respective proprietors.

Service Statement

This document is intended to provide users with general information about Tencent Cloud's products and services only and does not form part of Tencent Cloud's terms and conditions. Tencent Cloud's products or services are subject to change. Specific products and services and the standards applicable to them are exclusively provided for in Tencent Cloud's applicable terms and conditions.

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](#).

Cloud Data Warehouse ClickHouse

[Product Documentation](#) [Console](#)

Purchase Notes

Instructions Not sure which model to purchase? You can see [Billing Overview](#) or [contact us](#) for help.

Basic Configuration

Billing Mode: **Pay-as-you-go**

Region: **singapore** (selected), **siliconvalley** (available). Cloud products in different regions are not interconnected over private networks and the region cannot be changed after you purchase the cluster. We recommend you select the region nearest your customer to reduce access latency.

Availability Zone: **ap-singapore-1** (selected), **ap-singapore-2**, **ap-singapore-3**

Network: **vpc-r1do81wx | test-vpc** (selected), **subnet-7ibf8vhg | test-subnet** (selected). 253 subnet IPs in total, 217 available. If the existing networks do not meet your needs, you can [create a VPC](#) or [create a subnet](#) in the console. After purchasing Cloud Data Warehouse, you can change the VPC and subnet in the console.

Cluster Configuration

Cluster Name: 6 to 36 characters; supports Chinese characters, letters, digits, -, and _

Kernel Version: **21.3.9.84**

High Availability: **Enable** In a high-availability cluster, each node (shard) provides 2 replicas, and there are 3 ZooKeeper nodes by default. In a non-high availability cluster, each node (shard) provides 1 replica, which is not recommended for production environments.

Compute Node Type: **Standard** (selected), **Large-storage** (Sold out), **High-performance**

Compute Spec: **4-core 16 GB**

Storage Spec: **CLOUD_HSSD** **GB** A single node supports 200 to 320000 GB.

Compute Nodes:

If high availability is enabled, the number of nodes must be a multiple of 2.
 The number of nodes to configure must not exceed the number of available IPs (212) of the selected VPC or subnet. Switch to another subnet or VPC and try again if the number of IPs is insufficient.

ZooKeeper Node ⓘ

Compute Spec: 4-core 16 GB

Storage Spec: CLOUD_HSSD - 100 + GB

A single node supports 100 to 32000 GB.

Hot-Cold Tiered Storage Configuration

Cold Backup Storage Enable

After cold backup storage is enabled, data will be classified as hot and cold based on the capacity policy. When more than 90% of the node storage capacity is used, earlier data will be migrate to the COS disk you configured for cold backup.

Log Configuration

CLS Enable

After you enable CLS, the system will charge you on a pay-as-you-go basis. View [Purchase Guide](#) for details.

Other Configuration

Tag (Optional) ⓘ

Tag key: [input] Tag value: [input] [Delete]

+ Add

Configuration Cost

Activate

Configuration items

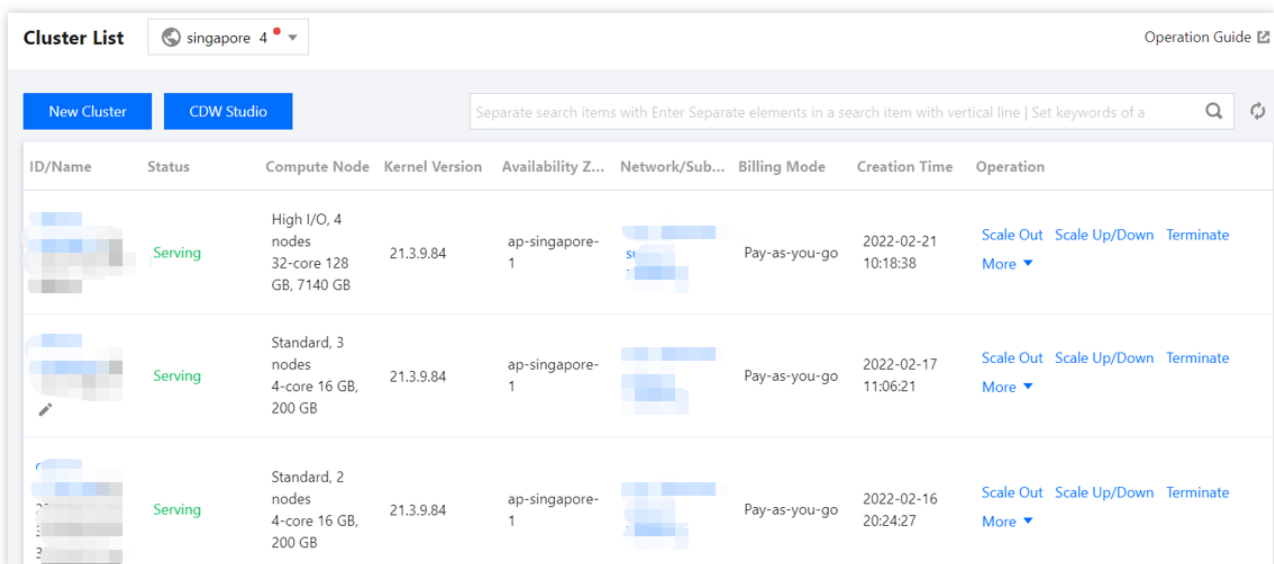
Configuration Item	Description
Billing Mode	Pay-as-you-go: postpaid , where a bill is generated hourly based on resource usage and then you pay for what you use.
Region	Currently, Cloud Data Warehouse is available in the Shanghai, Hong Kong (China), Beijing, Guangzhou, Singapore, and Silicon Valley regions. 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.

<p>Compute Node Type</p>	<p>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.</p>
<p>ZooKeeper Node Type</p>	<p>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.</p>

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

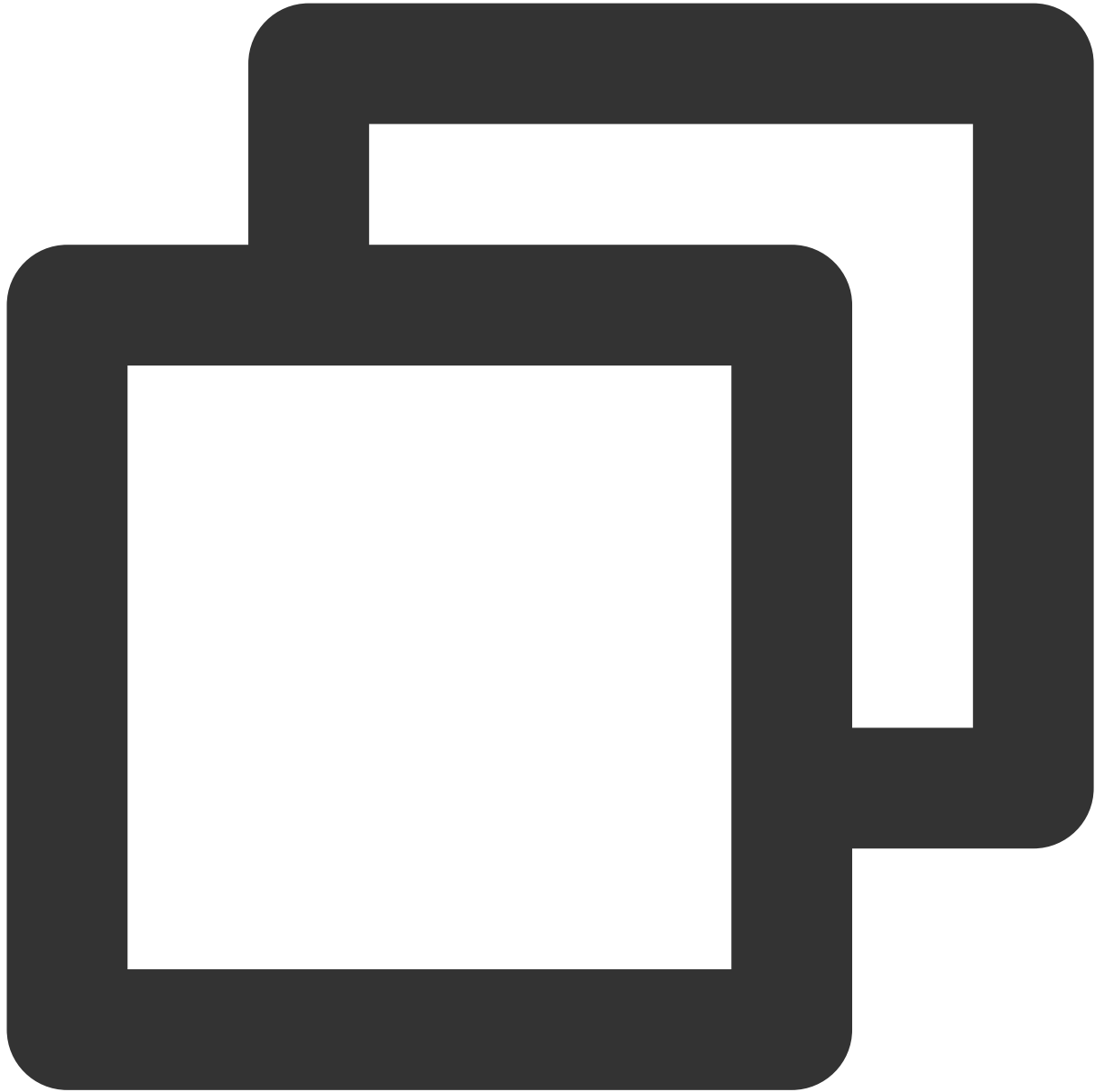
Viewing Cluster Information

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

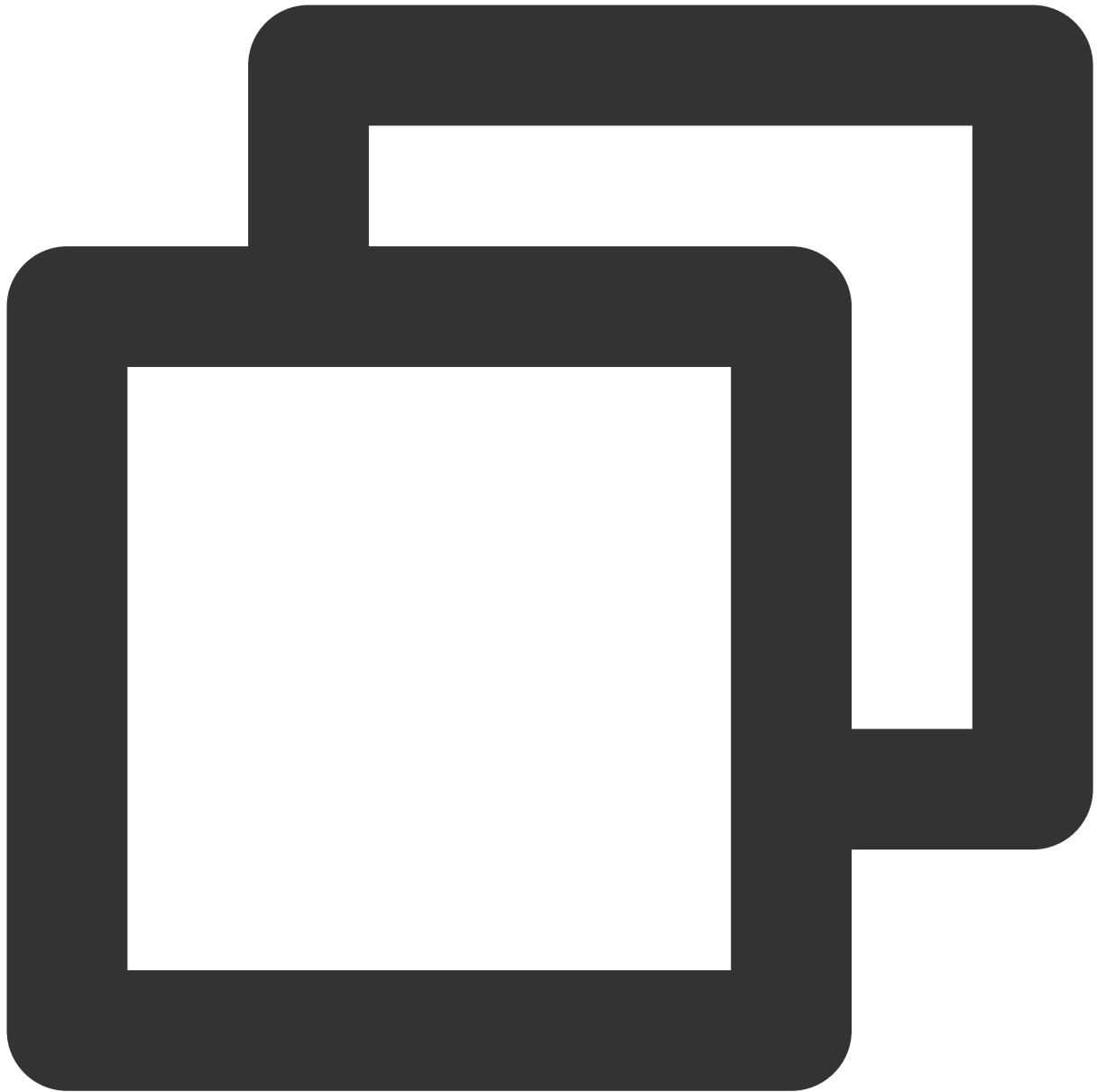
Import a data file to a ClickHouse cluster and view the imported data. Prepare the following `account.csv` file:



```
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
```

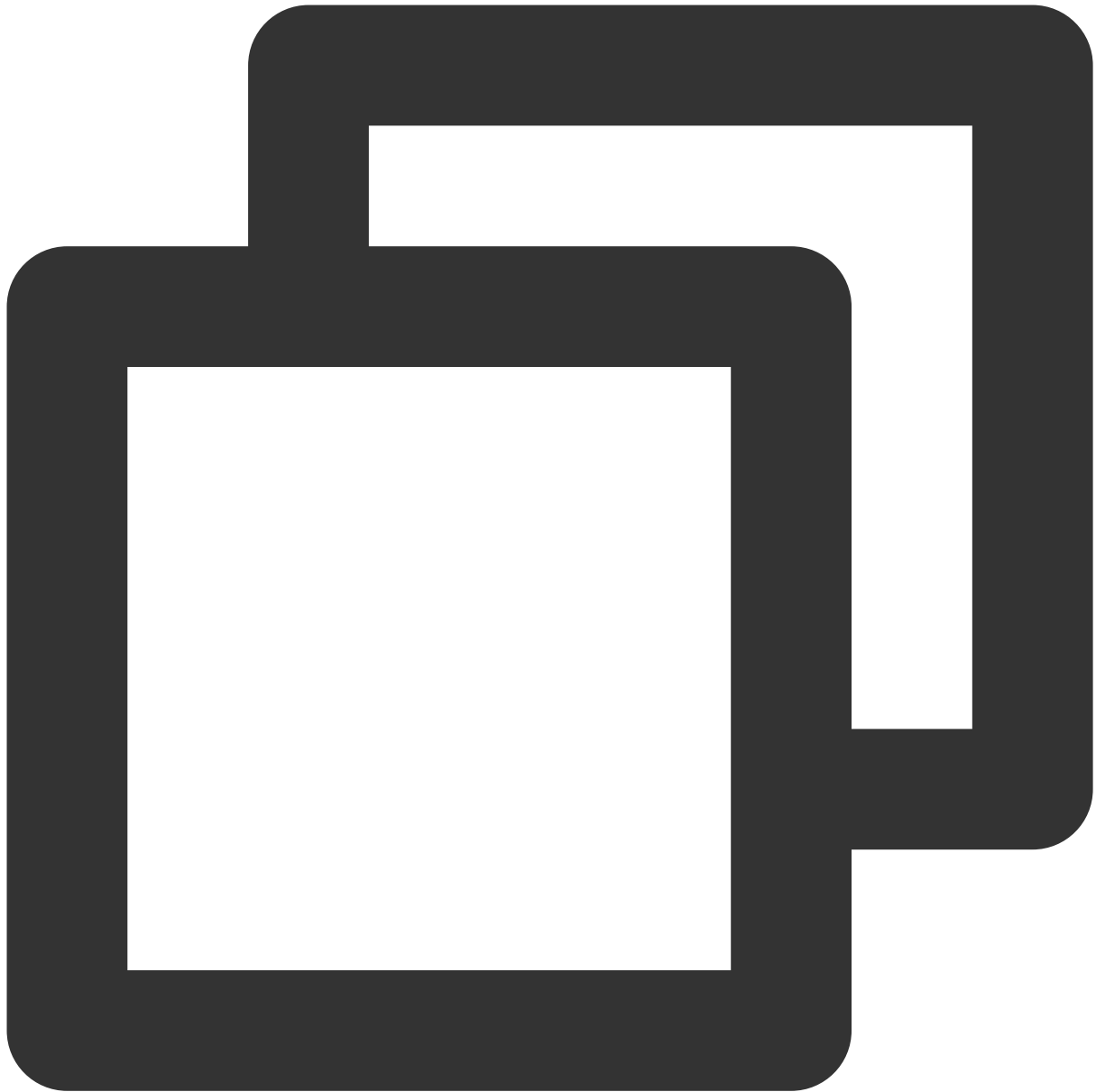
Install the client.



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

Access the cluster from the client.

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



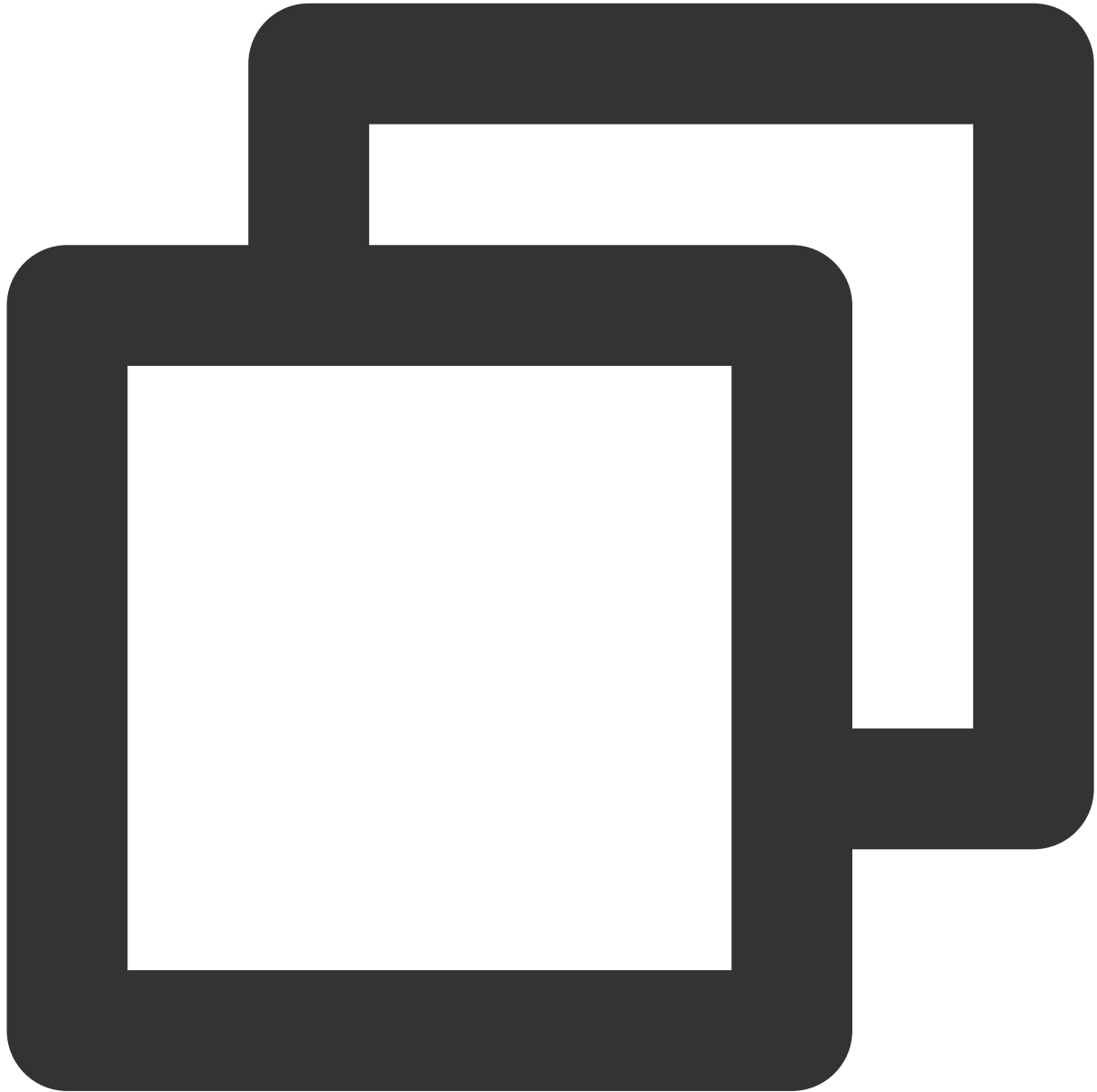
```
clickhouse-client -hxxx.xxx.xxx.xxx --port 9000
```

```
[root@VM-1-18-centos ~]# clickhouse-client -h 172.16.1.45 --port 9000
ClickHouse client version 20.7.2.30 (official build).
Connecting to 172.16.1.45:9000 as user default.
Connected to ClickHouse server version 20.7.2 revision 54437.

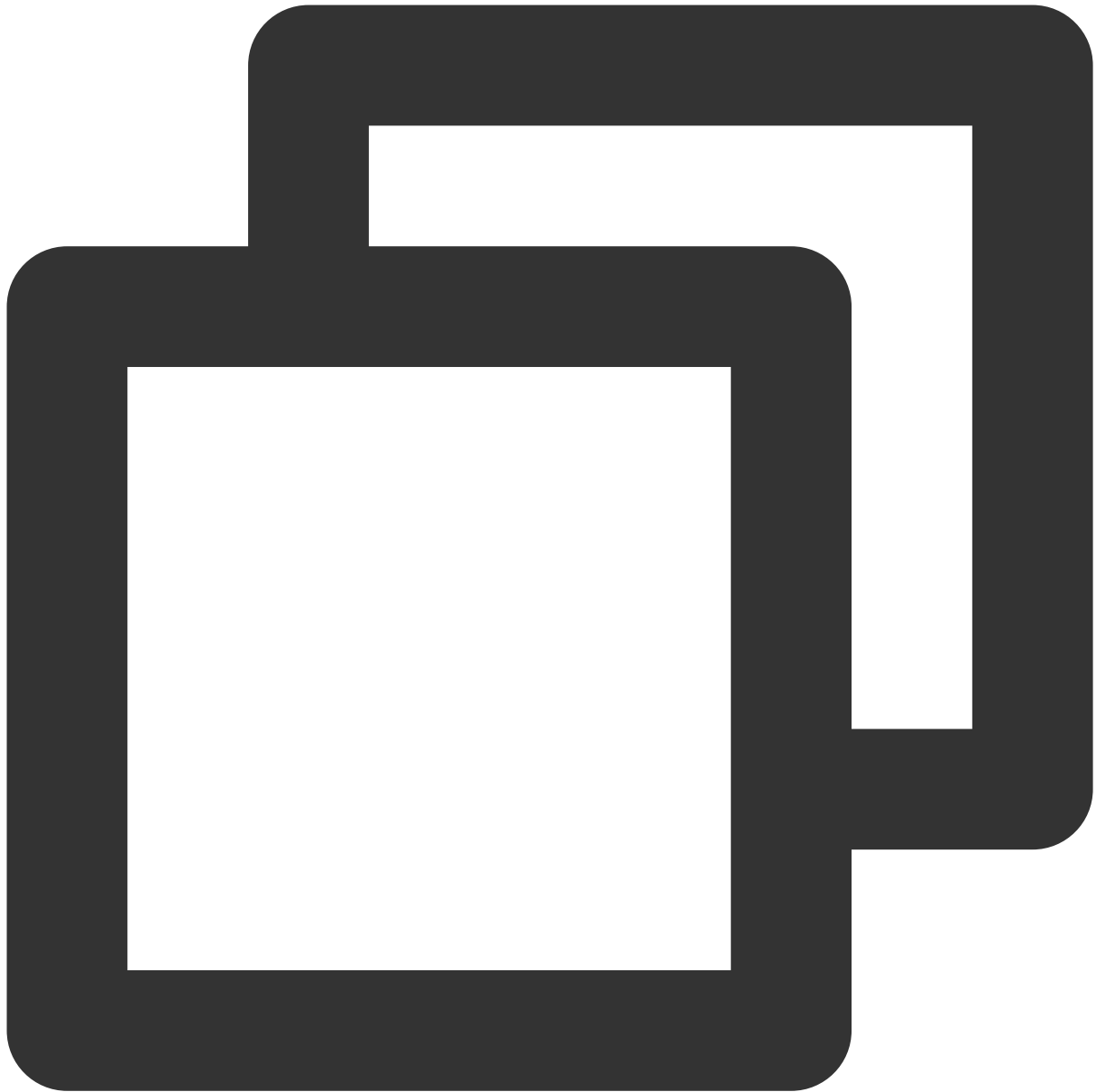
172.16.1.45 :)
```

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.



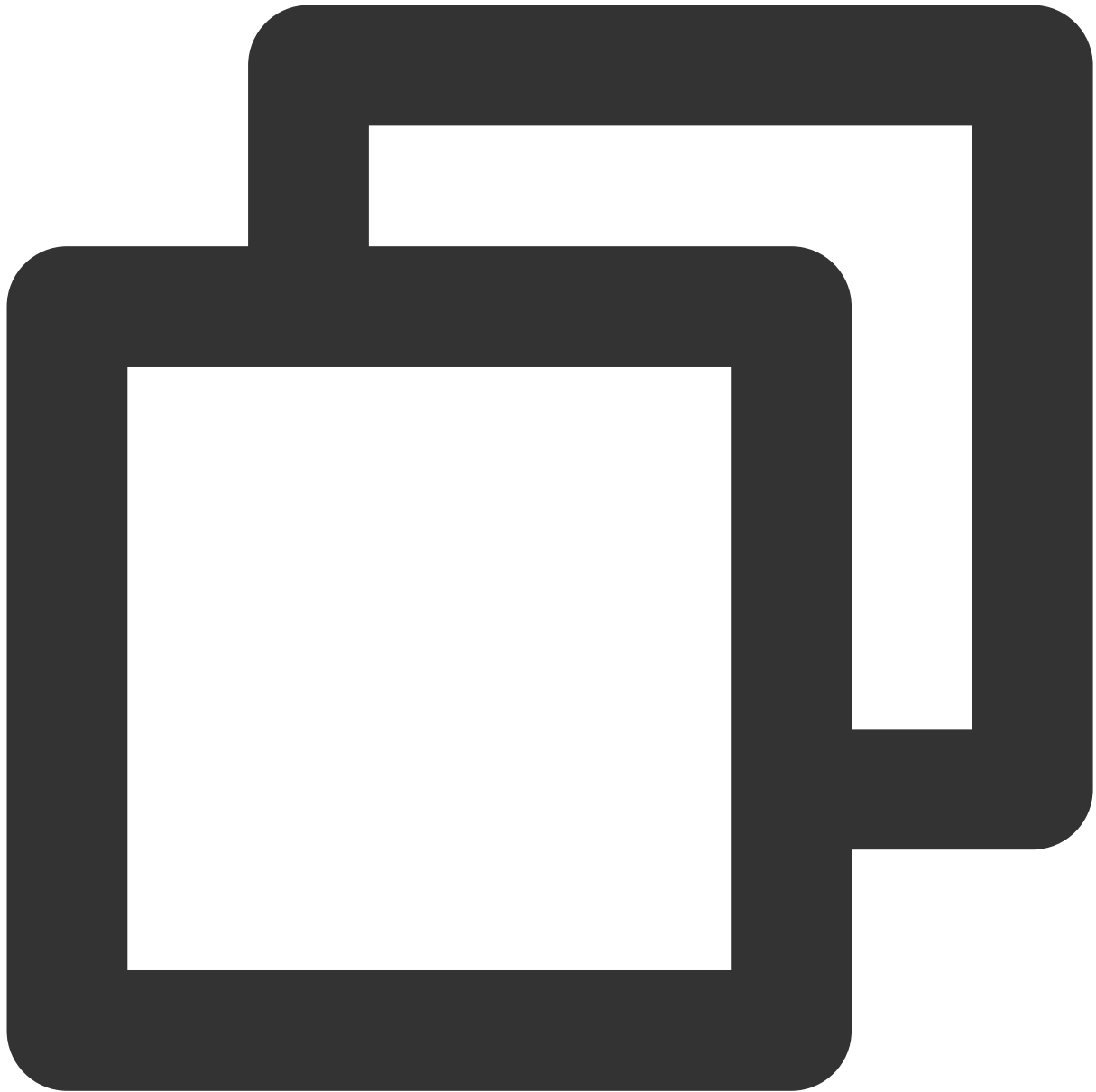
```
echo "select version()=21.3.9.83" | curl 'http://xxx.xxx.xxx.xxx:8123/' --data-b
```



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

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

Query the system cluster.



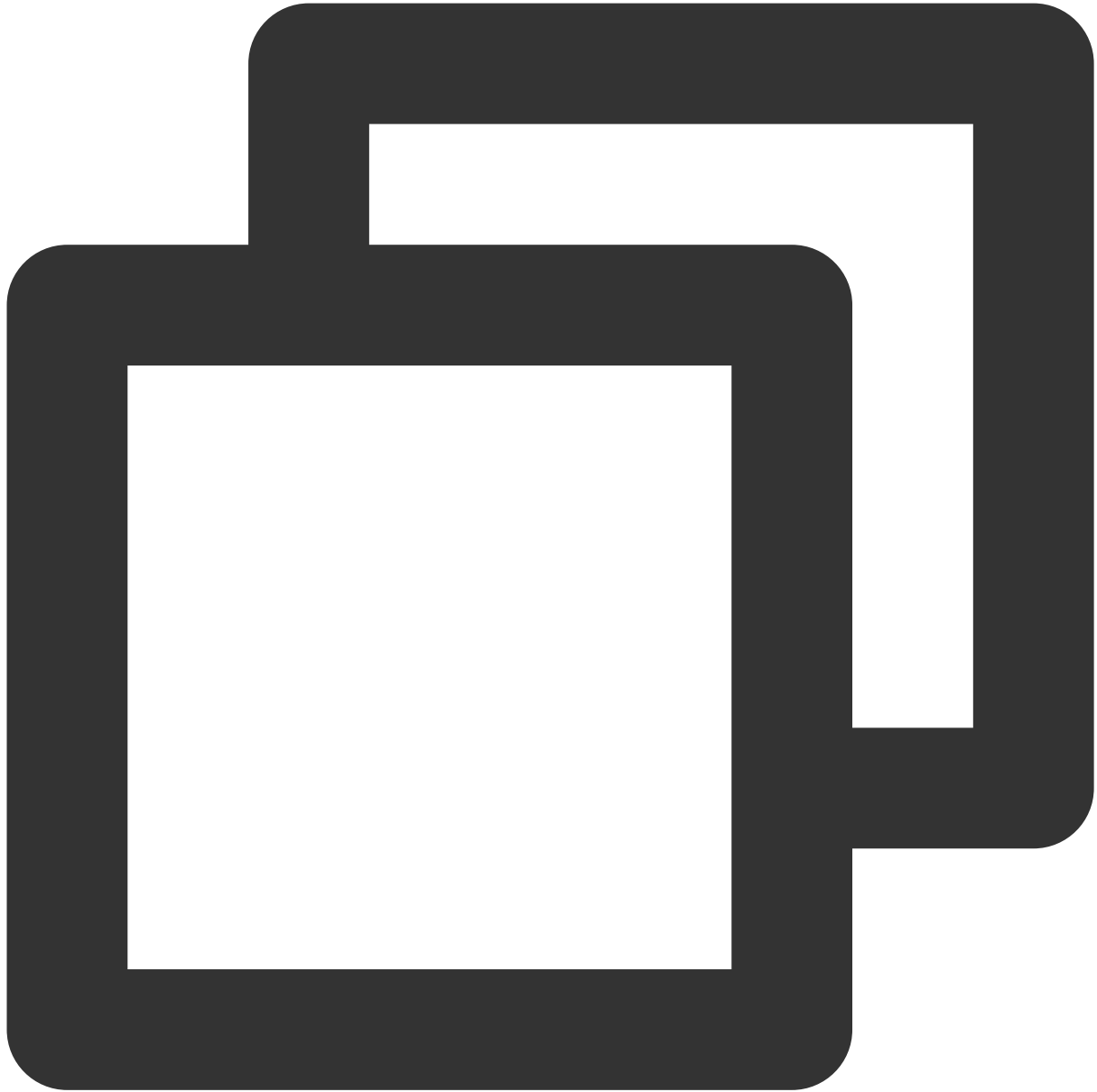
```
echo "select * from system.clusters" | curl 'http://xxx.xxx.xxx.xxx:8123/' --dat
```

```
[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

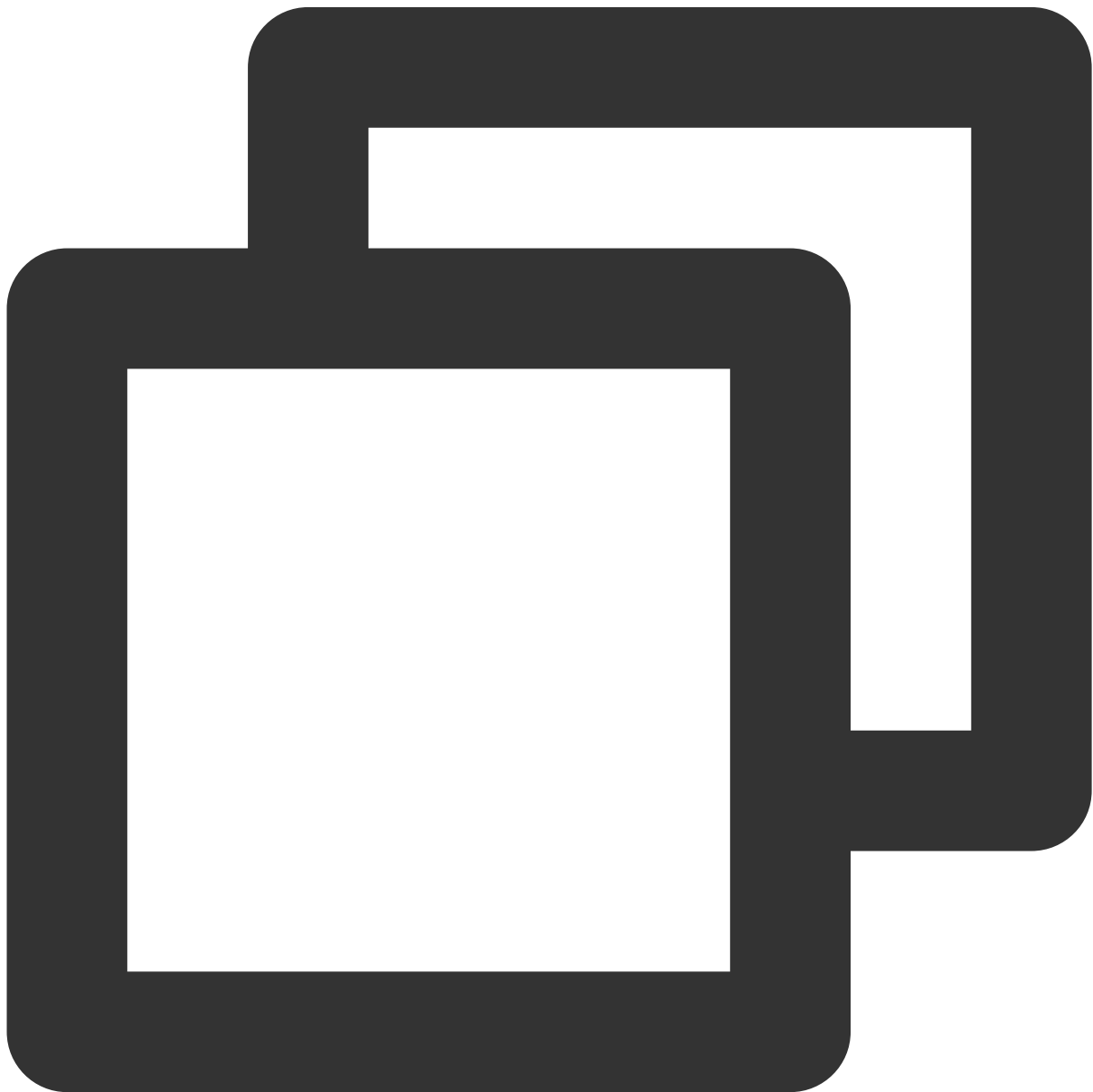


```
CREATE DATABASE IF NOT EXISTS testdb ON CLUSTER default_cluster;
```

```
) 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
	9000	0		1	0
	9000	0		0	0

Create a table in HA mode



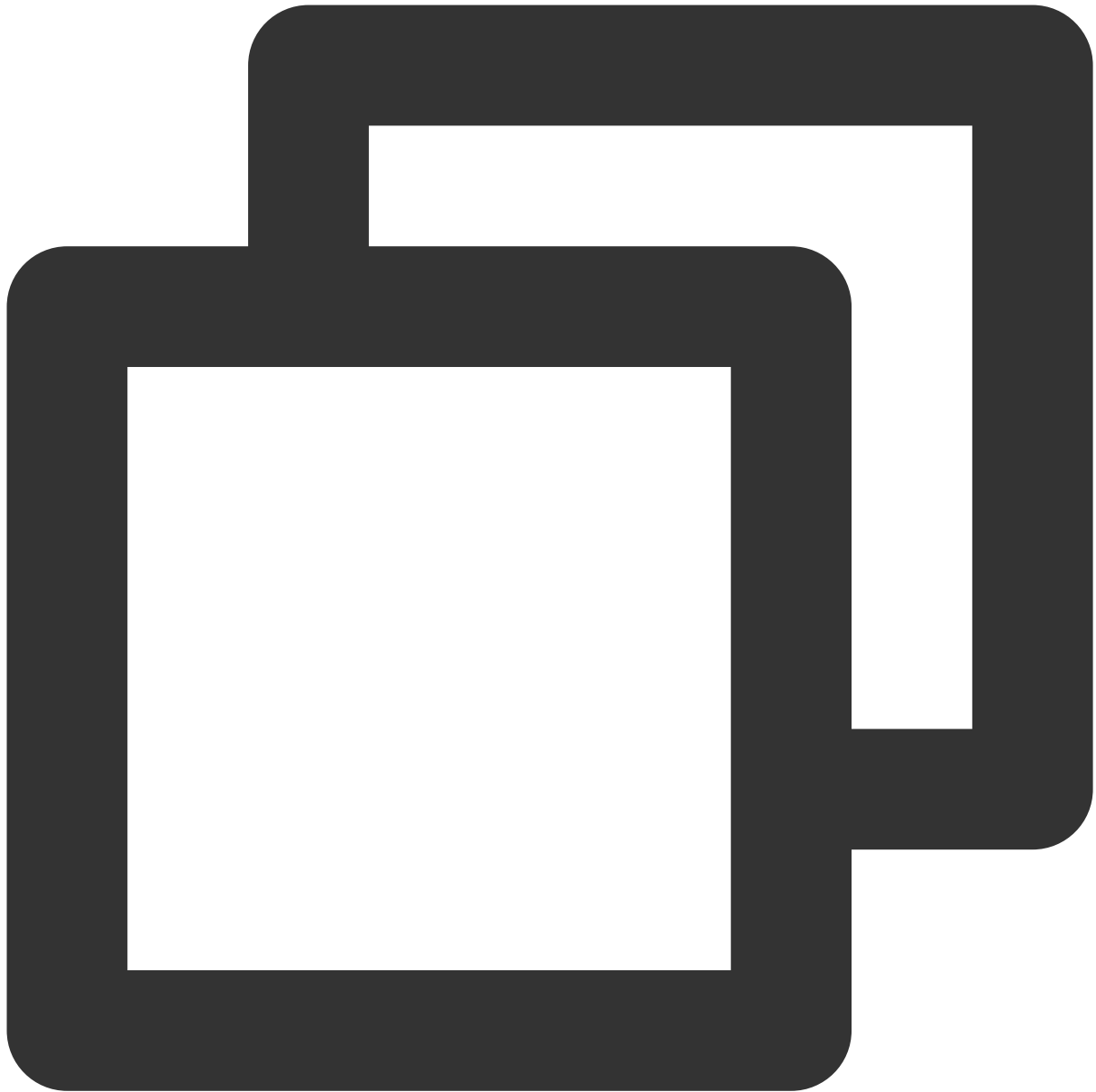
```
CREATE TABLE testdb.account ON CLUSTER default_cluster(accountid UInt16,name String
```

```
CREATE TABLE testdb.account ON CLUSTER default_cluster
(
  `accountid` UInt16,
  `name` String,
  `address` String,
  `year` UInt64
)
ENGINE = ReplicatedMergeTree('/clickhouse/tables/{layer}-{shard}/testdb/account', '{replica}')
ORDER BY accountid
```

host	port	status	error	num_hosts_remaining	num_hosts_active
	9000	0		1	0
	9000	0		0	0

2 rows in set. Elapsed: 0.112 sec.

Create a database in non-HA mode



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



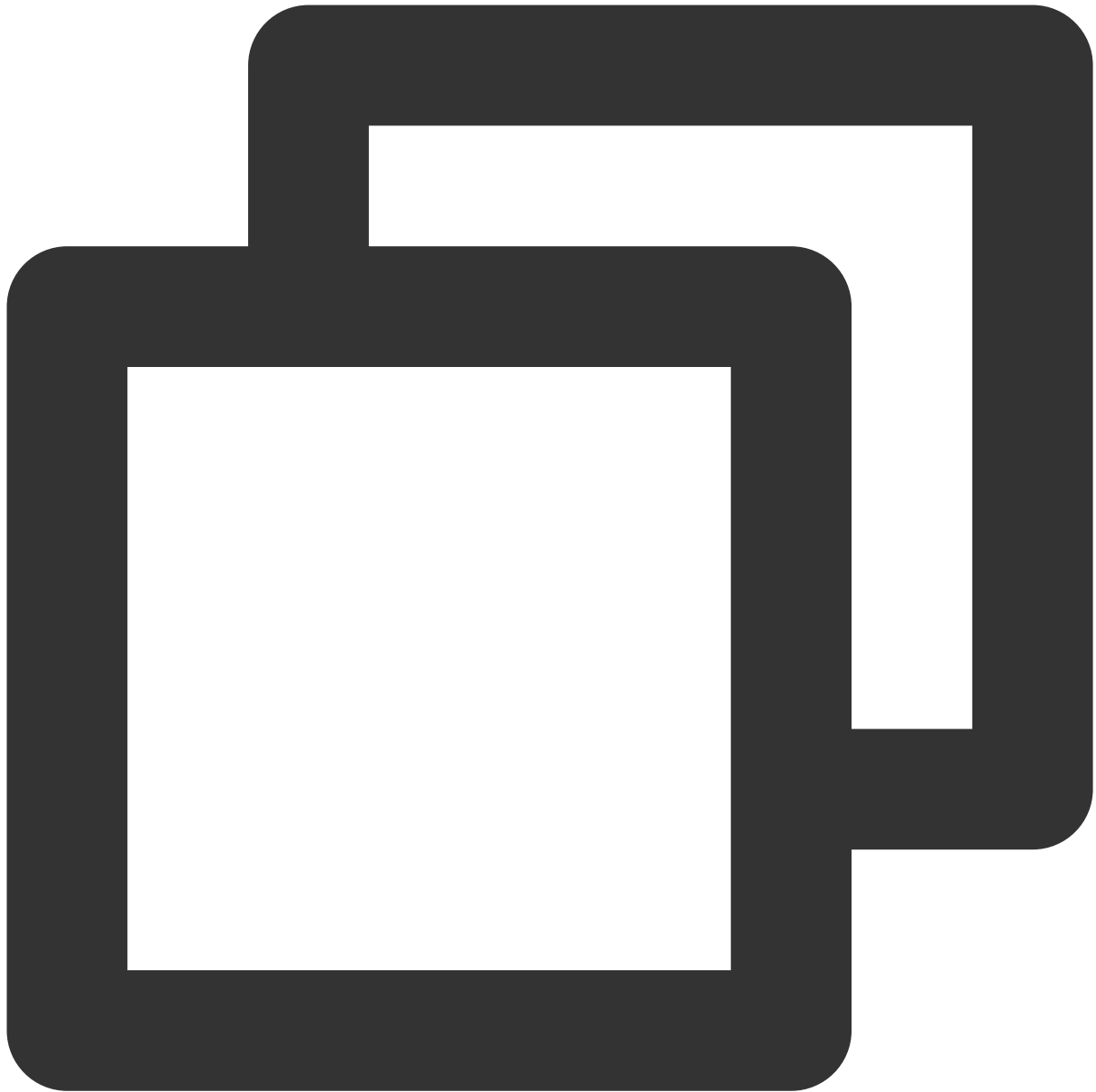
```
CREATE DATABASE IF NOT EXISTS testdb ON CLUSTER default_cluster
```

```
Query id: 14ff4b38-9979-4c90-a67d-93b57235919b
```

host	port	status	error	num_hosts_remaining	num_hosts_active
	29000	0		1	0
	19000	0		0	0

```
2 rows in set. Elapsed: 0.222 sec.
```

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
	9000	0		0	0

2 rows in set. Elapsed: 0.247 sec.

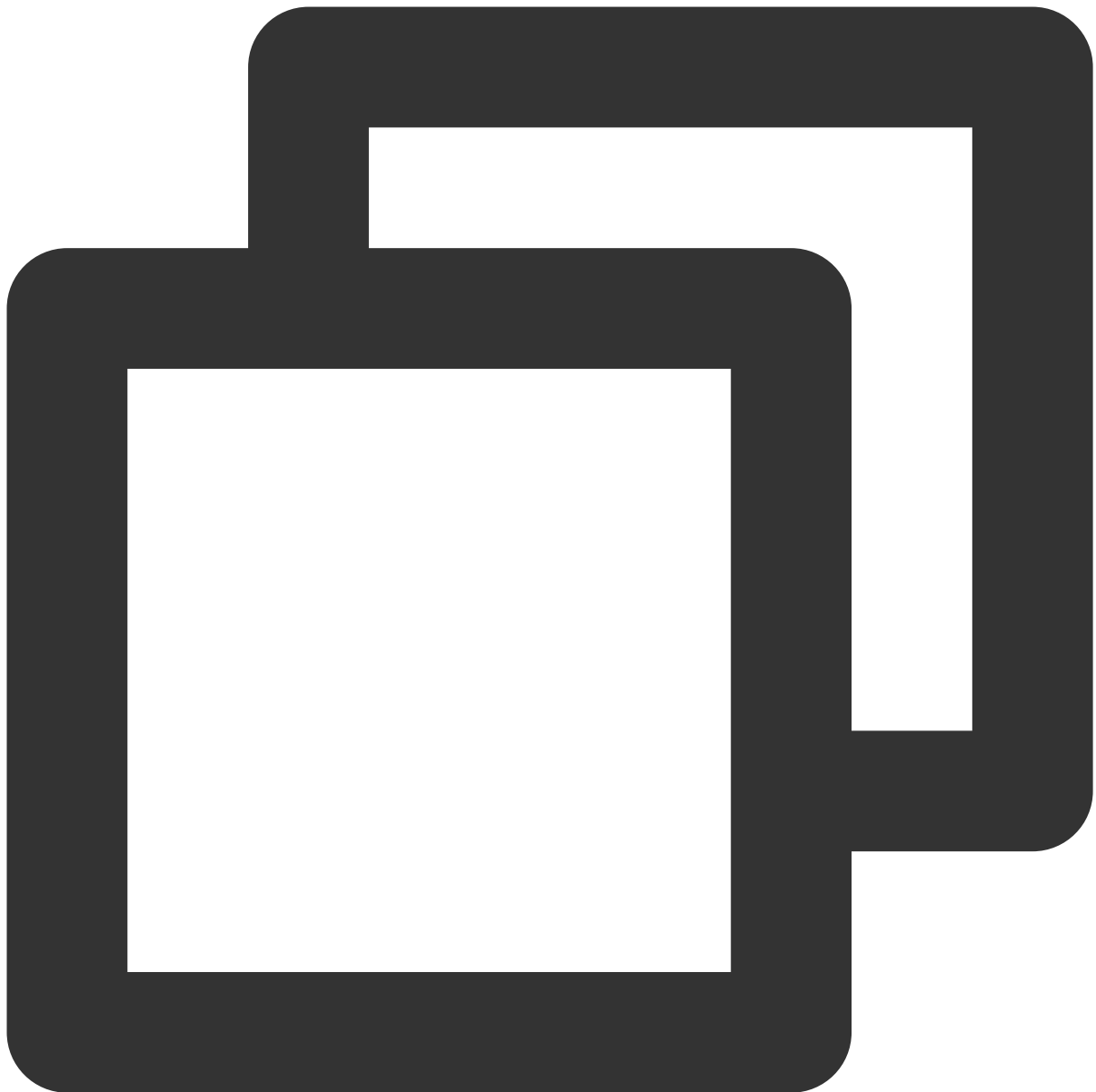
Importing data

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.
```