Департамент образования и науки города Москвы Государственное автономное образовательное учреждение высшего образования города Москвы «Московский городской педагогический университет» Институт цифрового образования Департамент информатики управления и технологий

# Кузьмина Дарья Юрьевна БД-241м

Инструменты хранения и анализа больших данных

# <u>Лабораторная работа 2.1. Часть 2.</u> <u>Cassandra</u> Вариант 11

Направление подготовки/специальность 38.04.05 - Бизнес-информатика Бизнес-аналитика и большие данные (очная форма обучения)

Руководитель дисциплины: Босенко Т.М., доцент департамента информатики, управления и технологий, доктор экономических наук

#### Содержание

Введение	. 2
Основная часть	. 2
	• –
Заключение	21

#### Введение

#### Цель:

Получить практические навыки работы с базой данных Cassandra, изучив основные операции по управлению данными, включая создание и использование ключенейсов, таблиц, выполнение запросов CQL, а также работу с различными инструментами подключения и администрирования.

#### Задачи:

- 1. Подключиться к Cassandra через cqlsh или браузерные интерфейсы (Cassandra Web, Apache Zeppelin).
- 2. Создать ключспейс с заданной стратегией репликации (SimpleStrategy).
- 3. Создать таблицы для хранения данных (например, movies, actors) с использованием примитивных типов и коллекций.
- 4. Выполнить операции CRUD (добавление, выборка, обновление, удаление данных) в созданных таблицах.
- 5. Изучить метаданные ключспейсов и таблиц с помощью команд DESCRIBE и запросов к системным таблицам.

#### Основная часть

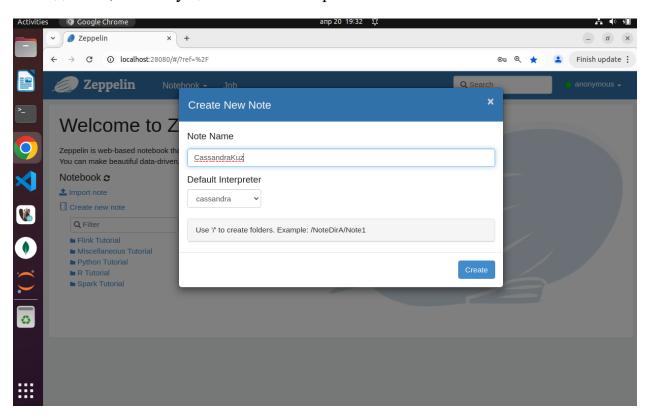
Задача 1.

В виртуальной машине проверяем, что сервер запущен:

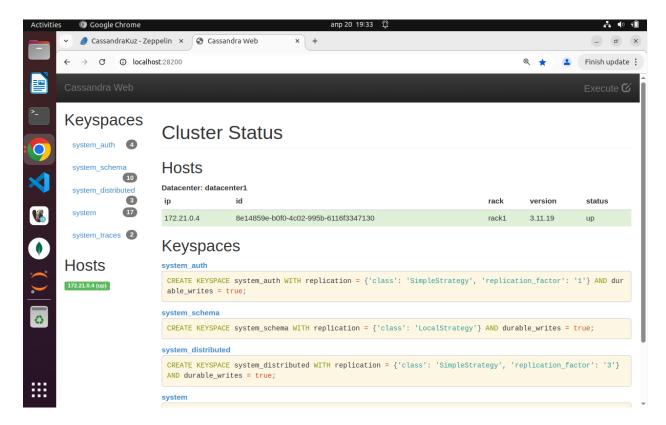
- dev@dev-vm:~/Downloads/dba/nonrel/mongo\$ cd ...
- dev@dev-vm:~/Downloads/dba/nonrel\$ cd cassandra

```
• dev@dev-vm:~/Downloads/dba/nonrel/cassandra$ docker ps
 CONTAINER ID IMAGE
                                           COMMAND
                                                                     CREATED
                                                                                       STATUS
 ORTS
                                                    NAMES
 b7804ebeb56d mongo:7.0.17-rc1-jammy "docker-entrypoint.s..."
                                                                     56 minutes ago Up 56 minutes
                                                                                                           0
 .0.0.0:27017->27017/tcp, [::]:27017->27017/tcp mongo-1
a5bb7206afbd postgres:16 "docker-entrypoir
                                           "docker-entrypoint.s..." 6 weeks ago
                                                                                       Up About an hour
  .0.0.0:5432->5432/tcp, [::]:5432->5432/tcp
                                                    postgres16
 ec4919b750da dpage/pgadmin4:latest "/entrypoint.sh"
                                                                     6 weeks ago
                                                                                       Up About an hour
  .0.0.0:80->80/tcp, [::]:80->80/tcp, 443/tcp
o dev@dev-vm:~/Downloads/dba/nonrel/cassandra$
```

Убеждаемся, что запущены 3 контейнера



Открываем ноутбук



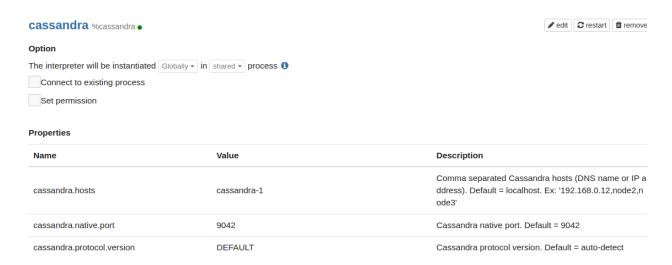
Смотрим, работает ли кассандра, проверяем доступность кластера (1 нода).



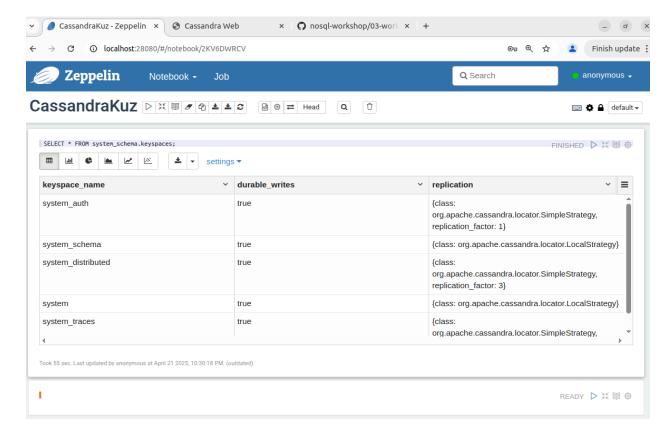
Проверим запуск

```
[cqlsh 5.0.1 | Cassandra 3.11.19 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
cassandra@cqlsh> SELECT * FROM system schema.keyspaces;
                  | durable_writes | replication
keyspace_name
system_auth |
_factor': '1'}
                               True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication
                                                                  {'class': 'org.apache.cassandra.locator.L
     system_schema
                               True
ocalStrategy'}
                               True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication
 system_distributed |
_factor': '3'}
                                                                  {'class': 'org.apache.cassandra.locator.L
            system |
                               True
ocalStrategy'}
     system_traces |
                               True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication
_factor': '2'}
(5 rows)
cassandra@cqlsh>
```

#### Посмотрим ключи по схемам



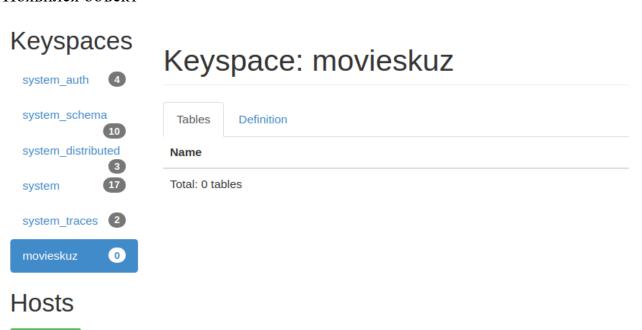
#### Меняем имя хоста



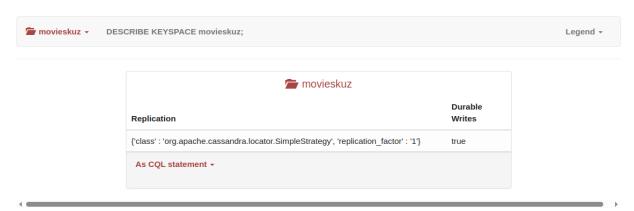
# Указываем настройки репликации



#### Появился объект



| DESCRIBE KEYSPACE movieskuz;



Took 24 sec. Last updated by anonymous at April 21 2025, 10:41:23 PM.

#### Создаем таблицы movie и actor

#### No Result Last query execution info

Took 3 sec. Last updated by anonymous at April 21 2025, 10:43:22 PM.

#### No Result Last query execution info

Took 2 sec. Last updated by anonymous at April 21 2025, 10:43:53 PM.

| DESCRIBE TABLE movie; FINISHED ▷ 💥 🞚 🐵

DESCRIBE TABLE movie; Legend ▼

<b>■</b> movie		
Column Type	Column Name	Data Type
×	movie_id	int
	cover_url	text
	genres	set <text></text>
	languages	set <text></text>
	plot_outline	text
	release_year	int
	running_time	int
	title	text
	top250_rank	int
As CQL statement ▼		

### Получим описание таблицы

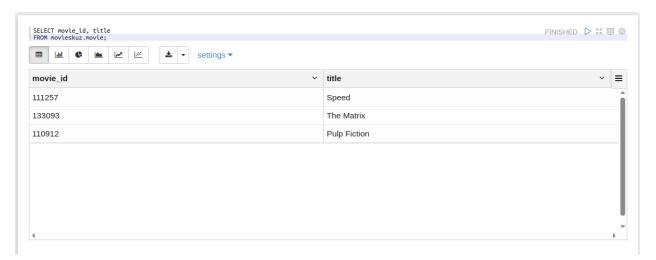
```
(Action', 'Sci-fi'),

$$Thomas A. Anderson is a man living two lives. By day he is an average computer programmer and by night a hacker known as Neo. Neo has always questioned his reality, but the truth is fra beyond his imagination. Neo finds himself targeted by the police when he is contacted by Morpheus, a legendary computer hacker branded a terrorist by the government. Morpheus awakens Neo to the real world, a ravaged wasteland where most of humanity have been captured by a race of machines that live off off the humans' body heat and electrochemical energy and who target and which the rinding stituin an artificial reality known as the Matrix. As a human rebellions's,
interposity in the property of the pro
```

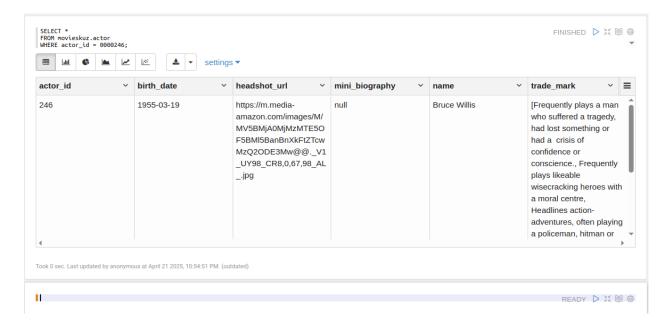
Took 0 sec. Last updated by anonymous at April 21 2025, 10:45:47 PM.

# Загоняем данные о фильмах

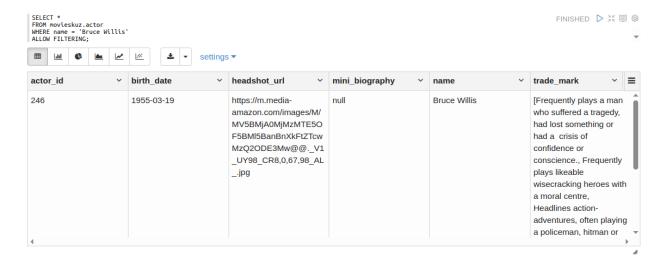
# Проверим результаты загрузки



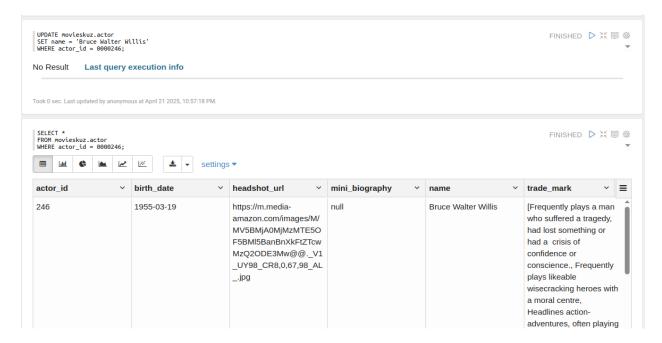
Поскольку кассандра не позволяет ограничивать поиск по столбцу, кроме первичного ключа, пропусти поиск по имени актера и выполним запрос на поле actor id



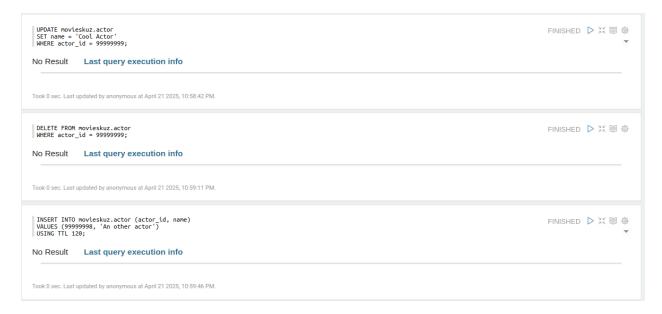
# Включим фильтрацию



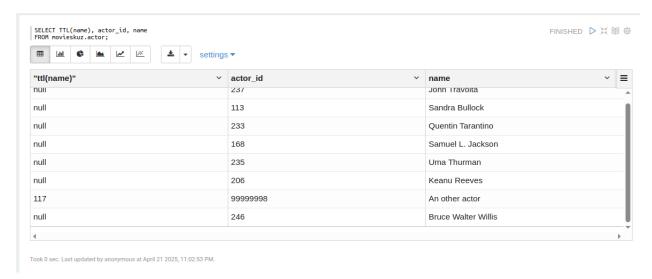
Однако это неэффективно



## Обновили данные о Брюсе Уиллисе



# Добавили и удалили пользователя, создали тестового пользователя на 120 секунд



#### Создадим динамические таблицы Фальмы на акетрам и актеры по фильмам

```
DROP TABLE IF EXISTS movieskuz.novies_by_actor;
CREATE TABLE movieskuz.novies_by_actor (actor_id int, movie_id int, title text, PRIMARY KEY (actor_id, movie_id)

No Result Last query execution info

DROP TABLE IF EXISTS movieskuz.actors_by_movie;
CREATE TABLE movieskuz.actors_by_movie;
CREATE TABLE movieskuz.actors_by_movie;
CREATE TABLE if EXISTS movieskuz.actors_by_movie;
create the content of the content
```

#### Вставляем данные в две таблицы

```
// Фильмы для актера "Bruce Willis"
INSERT INTO movieskuz.movies_by_actor (actor_id, movie_id, title)
VALUES (0000246, 0110912, 'Pulp Fiction');
INSERT INTO movieskuz.movies_by_actor (actor_id, movie_id, title)
VALUES (0000246, 1606378, 'A Good Day to Die Hard');
INSERT INTO movieskuz.movies_by_actor (actor_id, movie_id, title)
VALUES (0000246, 0217869, 'Unbreakable');
INSERT INTO movieskuz.movies_by_actor (actor_id, movie_id, title)
VALUES (0000246, 0377917, 'The Fifth Element');
INSERT INTO movieskuz.movies_by_actor (actor_id, movie_id, title)
VALUES (0000246, 0112864, 'Die Hard: With a Vengeance');
// Фильмы для актера "Keanu Reeves"
INSERT INTO movieskuz.movies_by_actor (actor_id, movie_id, title)
VALUES (0000206, 0133093, 'The Matrix');
INSERT INTO movieskuz.movies_by_actor (actor_id, movie_id, title)
VALUES (0000206, 0234215, 'The Matrix Reloaded');
INSERT INTO movieskuz.movies_by_actor (actor_id, movie_id, title)
VALUES (0000206, 0111257, 'Speed');
// Фильмы для актера "Sandra Bullock"
INSERT INTO movieskuz.movies_by_actor (actor_id, movie_id, title)
VALUES (0000113, 0111257, 'Speed');
```

Took 0 sec. Last updated by anonymous at April 21 2025, 11:08:05 PM.

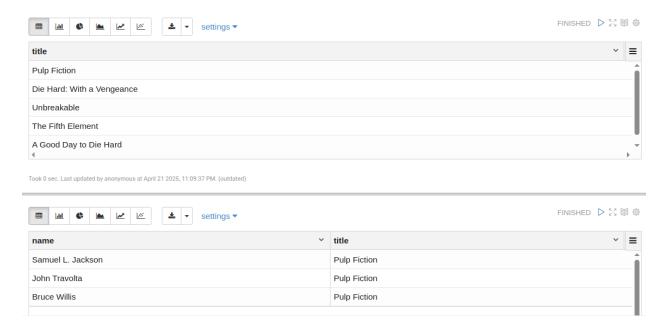
Last query execution info

No Result

```
// Актеры для фильма "The Matrix"
 INSERT INTO movieskuz.actors_by_movie (movie_id, title, actor_id, name)
 VALUES (0133093, 'The Matrix', 0000206, 'Keanu Reeves');
 INSERT INTO movieskuz.actors_by_movie (movie_id, title, actor_id, name)
 VALUES (0133093, 'The Matrix', 0000401, 'Laurence Fishburne');
 INSERT INTO movieskuz.actors_by_movie (movie_id, title, actor_id, name)
 VALUES (0133093, 'The Matrix', 0005251, 'Carrie-Anne Moss');
 INSERT INTO movieskuz.actors_by_movie (movie_id, title, actor_id, name)
VALUES (0133093, 'The Matrix', 0915989, 'Hugo Weaving');
 // Актеры для фильма "Pulp Fiction"
 INSERT INTO movieskuz.actors_by_movie (movie_id, title, actor_id, name)
 VALUES (0110912, 'Pulp Fiction', 0000237, 'John Travolta');
 INSERT INTO movieskuz.actors_by_movie (movie_id, title, actor_id, name)
 VALUES (0110912, 'Pulp Fiction', 0000168, 'Samuel L. Jackson');
 INSERT INTO movieskuz.actors_by_movie (movie_id, title, actor_id, name)
 VALUES (0110912, 'Pulp Fiction', 0000246, 'Bruce Willis');
No Result
               Last guery execution info
```

Took 0 sec. Last updated by anonymous at April 21 2025, 11:09:10 PM.

# Выведем все фильмы, в которых снимался Брюс Уиллис и всех актеров Криминального чтива



#### Использование столбцов-счетчиков

Создаем статическую таблицу рейтинга со столбцами типа counter, введем данные рейтинга



# Результат



Создадим таблицу просмотров с сортировкой по половому признаку году и месяцу просмотра, дополним ее данными.

#### No Result Last query execution info

Took 2 sec. Last updated by anonymous at April 21 2025, 11:17:38 PM.

```
// Pulp Fiction Views 2019/03
UPDATE movieskuz.movie_viewed_by_time
SET male = male + 1
WHERE movie_id = 0110912 AND year = 2019 and month = 03;
UPDATE movieskuz.movie_viewed_by_time
SET male = male + 1
WHERE movie_id = 0110912 AND year = 2019 and month = 03;
UPDATE movieskuz.movie_viewed_by_time
SET female = female + 1
WHERE movie_id = 0110912 AND year = 2019 and month = 03;
// Pulp Fiction Views 2019/04
UPDATE movieskuz.movie_viewed_by_time
SET female = female + 1
WHERE movie_id = 0110912 AND year = 2019 and month = 04;
```

#### Выведем данные



#### Криминальное чтиво за все время



#### За месяц



С января по май

# Индивидуальное задание Вариант 11

- 1. Создайте ключспейс shops с репликацией SimpleStrategy и коэффициентом репликации 1.
- 2. Создайте таблицу products в ключспейсе shops с полями product\_id (int), name (text), price (float),

category (text), stock (int) и первичным ключом product id.

- 3. Вставьте четыре продукта в таблицу products.
- 4. Выберите все продукты из таблицы products.
- 5. Обновите поле stock продукта с product\_id = 2.

Добавляем свои данные с гитхаба.

# Судя по лекции, концептуально, это задание выглядело бы так:

1. Создание ключспейса shops

Для начала работы необходимо создать ключспейс (пространство ключей) с именем "shops", используя стратегию репликации SimpleStrategy и коэффициент репликации 1 (так как мы работаем с одним узлом Cassandra):

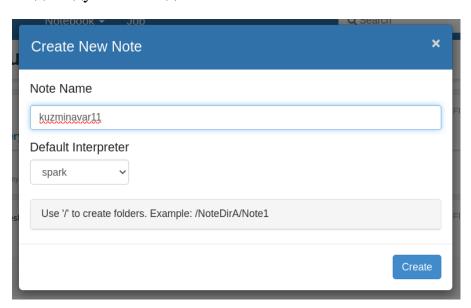
# CREATE KEYSPACE shops WITH replication = {

```
'class': 'SimpleStrategy',
  'replication_factor': 1
};
```

Эта команда создает новое пространство ключей, где:

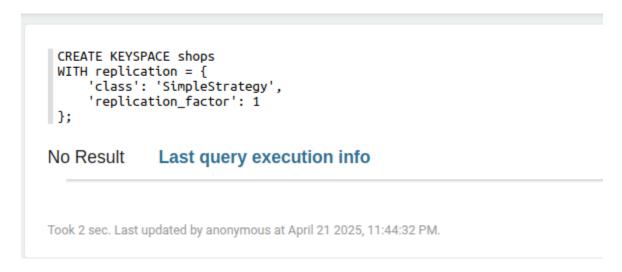
- `SimpleStrategy` простая стратегия репликации, подходящая для однодатацентровых кластеров
- `replication\_factor: 1` указывает, что каждая запись будет храниться только на одном узле

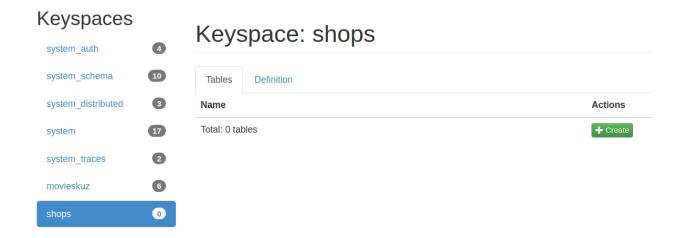
Индивидуальное задание



#### Создаем новый блокнот

Создаем новый кейспейс по заданию





# 2. Создание таблицы shops

После создания ключспейса переходим в него и создаем таблицу products:

USE shops;

```
CREATE TABLE products (

product_id int,

name text,

price float,

category text,

stock int,

PRIMARY KEY (product_id)
);
```

Таблица содержит следующие поля:

- `product\_id` уникальный идентификатор продукта (первичный ключ)
- `name` название продукта
- `price` цена продукта

- `category` категория продукта
- `stock` количество товара на складе

```
USE shops;

CREATE TABLE products (
    product_id int PRIMARY KEY,
    name text,
    price float,
    category text,
    stock int
);

No Result Last query execution info

Took 2 sec. Last updated by anonymous at April 22 2025, 12:02:02 AM.
```

# Создали таблцу шопс

DESCRIBE TABLE products;

Legend →

Column
Type
Column Name
Data Type

∴

product\_id int

category text

name text

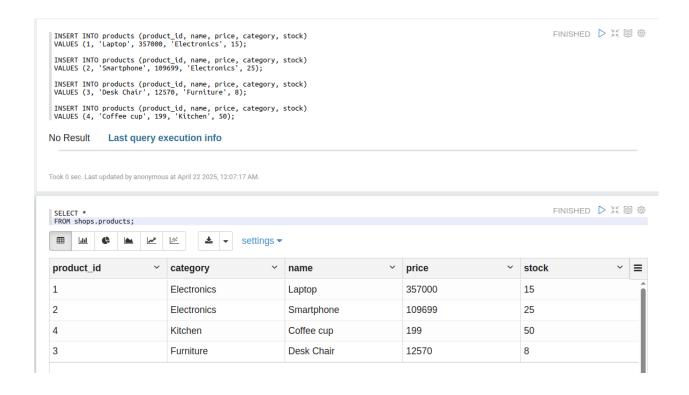
price float

stock int

As CQL statement →

# 3. Вставка данных в таблицу products

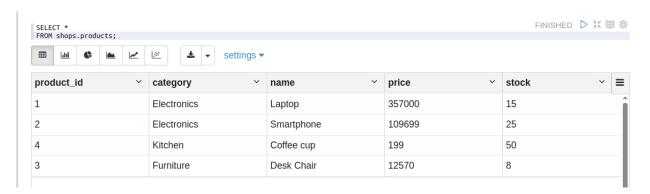
Добавим четыре продукта в таблицу:



#### 4. Выборка всех продуктов

Проверим, что данные успешно добавлены, выполнив запрос на выборку всех записей:

# SELECT \* FROM products;



# 5. Обновление количества товара на складе

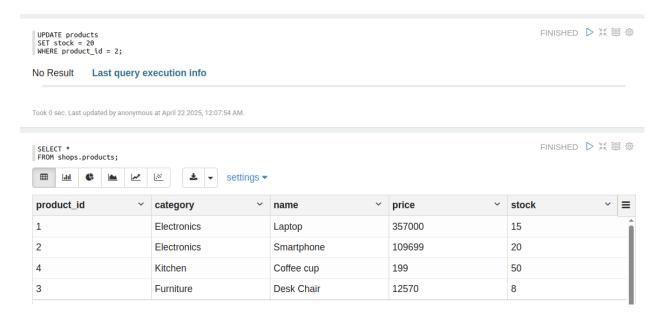
Обновим количество товара на складе для продукта с product\_id = 2:

UPDATE products SET stock = 30 WHERE product\_id = 2;

После выполнения этой команды количество товара "Smartphone" изменится с 25 до 30.

Убедимся, что обновление прошло успешно:

SELECT \* FROM products WHERE product id = 2;



#### Заключение

В теории удалось получить практические навыки работы с базой данных Cassandra, изучив основные операции по управлению данными, включая создание и использование ключспейсов, таблиц, выполнение запросов CQL, а также работу с различными инструментами подключения и администрирования.