

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

Мошенина Елена Дмитриевна БД-241м

Инструменты хранения и анализа больших данных

**Практическая работа 2.1. Cassandra**

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

Москва  
2025

**Содержание**

**Введение..... 3**

**Основная часть..... 3**

**Заключение ..... 7**

## Введение

### Цель:

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.

### 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:

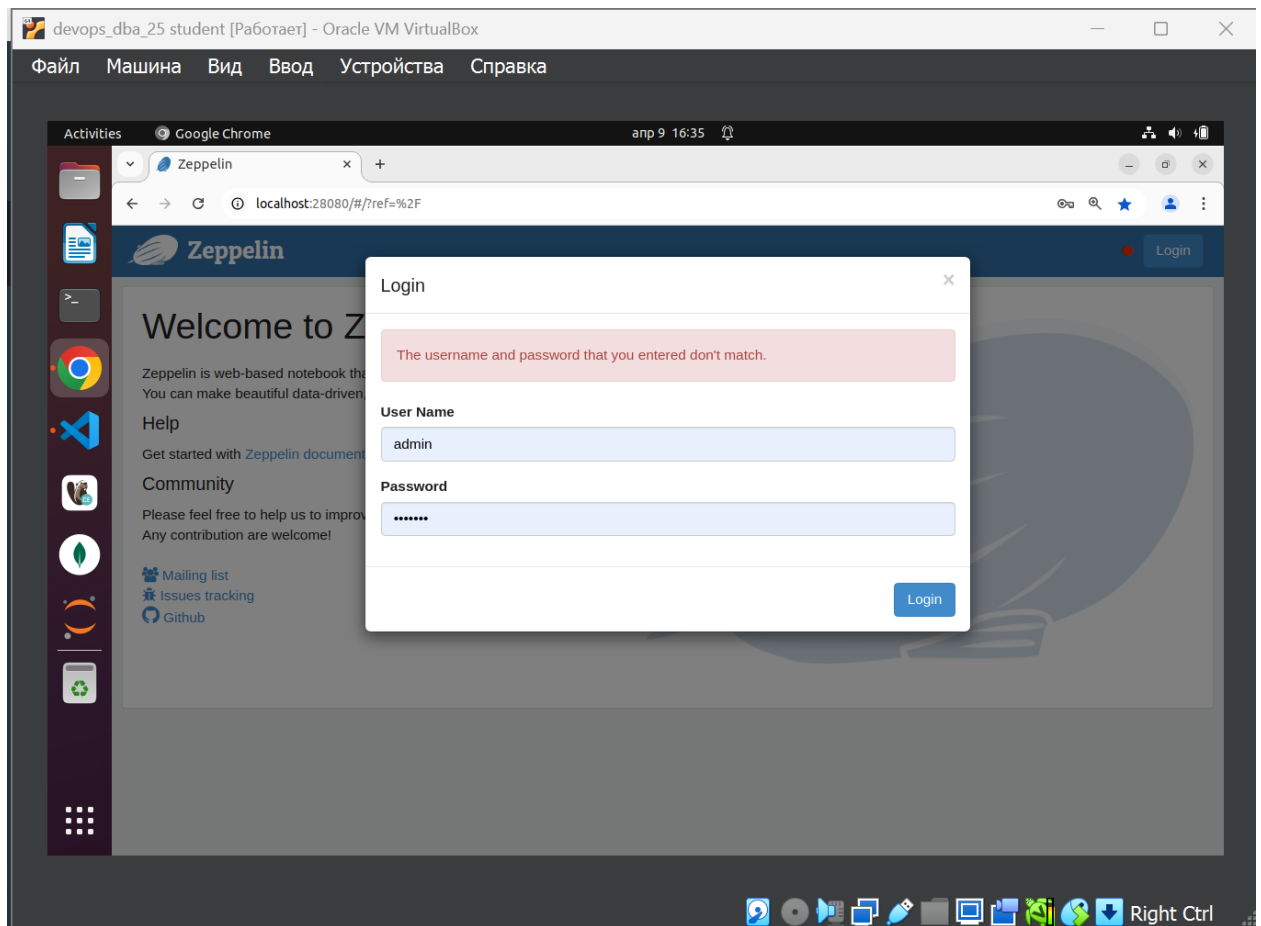
```
dev@dev-vm:~/Downloads/dba/nonrel/mongo$ cd ..
dev@dev-vm:~/Downloads/dba/nonrel$ cd cassandra
dev@dev-vm:~/Downloads/dba/nonrel/cassandra$ docker ps
```

CONTAINER ID	IMAGE	COMMAND	CREATED	STATUS
a5bb7206afbd	postgres:16	"docker-entrypoint.s..."	6 weeks ago	Up 27 minu
tes	0.0.0.0:5432->5432/tcp, [::]:5432->5432/tcp	postgres16		
ec4919b750da	dpage/pgadmin4:latest	"/entrypoint.sh"	6 weeks ago	Up 27 minu
tes	0.0.0.0:80->80/tcp, [::]:80->80/tcp, 443/tcp	pgadmin		

```
dev@dev-vm:~/Downloads/dba/nonrel/cassandra$ █
```

Убеждаемся, что запущены 3 контейнера

Заходим в Zeppelin, пароль не подходит.



## Индивидуальное задание **Вариант 15**

1. Создайте ключспейс movies с репликацией SimpleStrategy и коэффициентом репликации 1.
2. Создайте таблицу movies в ключспейсе movies с полями movie\_id (int), title (text), release\_year (int), running\_time (int), languages (set<text>), genres (set<text>), plot\_outline (text), cover\_url (text), top250\_rank (int) и первичным ключом movie\_id.
3. Вставьте три фильма в таблицу movies.
4. Выберите все фильмы из таблицы movies.
5. Обновите поле plot\_outline фильма с movie\_id = 0110912.

Судя по лекции, это задание выглядело бы так:

### *1. Создание ключспейса movies*

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

```
CREATE KEYSPACE movies WITH replication = {  
    'class': 'SimpleStrategy',  
    'replication_factor': 1  
};
```

Эта команда создает новое пространство ключей, где:

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

## ***2. Создание таблицы movies***

После создания ключспейса переходим в него и создаем таблицу movies:

```
USE movies;
```

```
CREATE TABLE movies (  
    movie_id int,  
    title text,  
    release_year int,  
    running_time int,  
    languages set<text>,  
    genres set<text>,  
    plot_outline text,  
    cover_url text,
```

```
top250_rank int  
PRIMARY KEY (movie_id)  
);
```

### ***3. Вставка данных в таблицу products***

Добавим три фильма в таблицу:

```
INSERT INTO movies (movie_id, title, release_year, running_time, languages,  
genres, plot_outline, cover_url, top250_rank)  
VALUES (?);
```

```
INSERT INTO movies (movie_id, title, release_year, running_time, languages,  
genres, plot_outline, cover_url, top250_rank)  
VALUES (?);
```

```
INSERT INTO movies (movie_id, title, release_year, running_time, languages,  
genres, plot_outline, cover_url, top250_rank)  
VALUES (?);
```

### ***4. Выборка всех фильмов***

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

```
SELECT * FROM movies;
```

Ожидаемый результат:

```
product_id | name      | price | category | stock  
-----+-----+-----+-----+-----
```

1	Laptop	999.99	Electronics	15
2	Smartphone	699.99	Electronics	25
3	Headphones	149.99	Accessories	50
4	Keyboard	49.99	Accessories	30

### ***5. Обновите поле `plot_outline` фильма с `movie_id = 0110912`***

```
UPDATE products SET plot_outline = 30 WHERE movie_id = 0110912;
```

Убедимся, что обновление прошло успешно:

```
SELECT * FROM movies WHERE movie_id = 2;
```

### **Заключение**

В результате выполнения поставленной цели должны были быть получены практические навыки работы с базой данных Cassandra. Были изучены основные операции по управлению данными, включая создание и использование ключспейсов, таблиц, выполнение запросов CQL, а также работа с различными инструментами подключения и администрирования.