|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования **«МИРЭА − Российский технологический университет»**  **РТУ МИРЭА** |

**Институт информационных технологий (ИИТ)**

**Кафедра прикладной математики (ПМ)**

**ОТЧЕТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ**

по дисциплине «Методы анализа данных»

**Практическая работа № 6**

|  |  |  |  |
| --- | --- | --- | --- |
| Студент группы ИНБО-01-17 | *ИНБО-05-22, Новиков Денис Владиславович* | (подпись) | |
| Преподаватель | *Шкерин Алексей Леонидович* | (подпись) | |
| Отчет представлен | «11» сентября 2024 г. | |  | |

Москва 2024 г.

## Постановка задачи:

Выполнить базовые операции с созданием, удалением и изменением таблицы в оболочке HBase. Использовать оболочку для размещения и получения данных в HBase. Использовать команды для вставки, извлечения, сканирования и удаления строк.

## Ход работы:

### Доступ к данным с помощью HBase

Запустим оболочку HBase. Выполним команду справки и просмотрим основную информацию об использовании оболочки HBase на Рисунке 1.

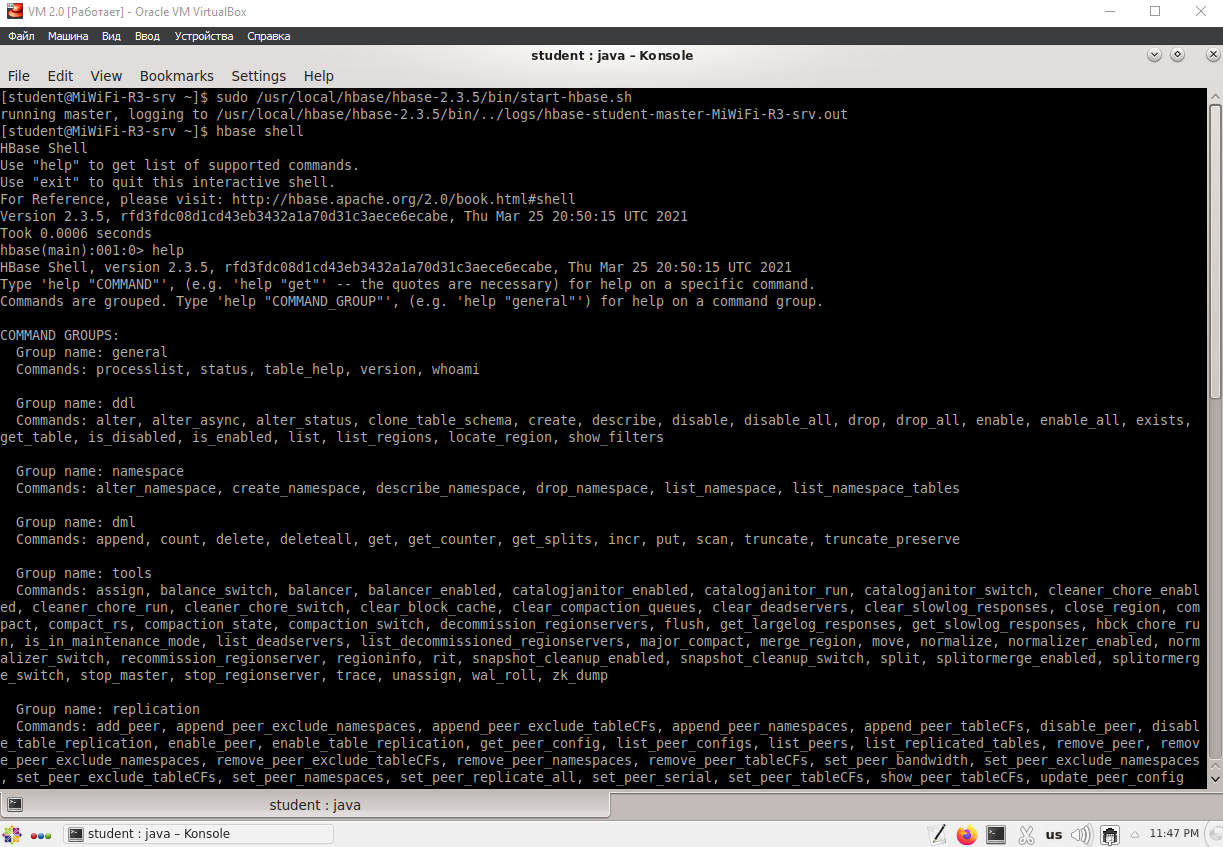


Рисунок 1 — Вход в HBase

Отобразим версию и состояние для базового использования на Рисунке 2.

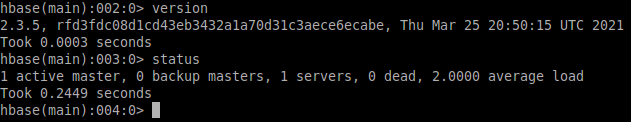


Рисунок 2 — Информация о HBase

Используни команду create для создания новой таблицы на Рисунке 3.

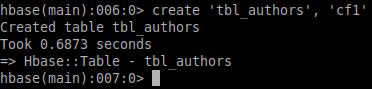


Рисунок 3 — Создание новой таблицы

Посмотрим список созданных таблиц с помощью команды list на Рисунке 4.

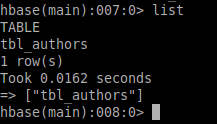


Рисунок 4 — Созданные таблицы

Используем команду describe для просмотра подробных сведений, включая настройки по умолчанию на Рисунке 5.

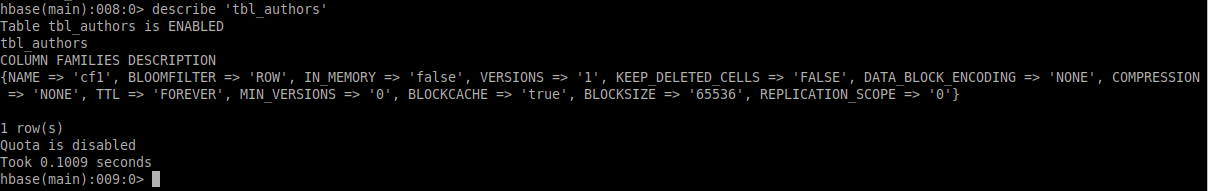


Рисунок 5 — Подробные сведения

Удалим только что созданную таблицу. Выведем все таблицы, чтобы убедиться, что таблица была успешно удалена на Рисунке 6.

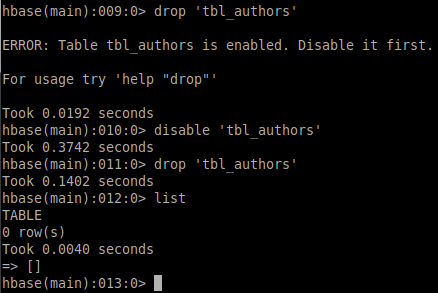


Рисунок 6 — Удаление таблицы и проверка удаления

Создаем тестовую таблицу и используйте команду put. Здесь мы вставляем три значения, по одному за раз на Рисунке 7.

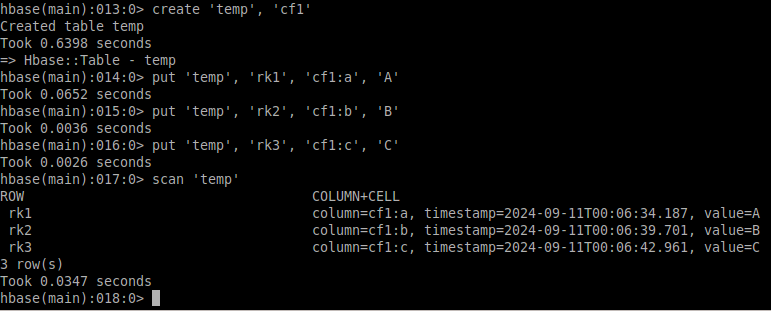


Рисунок 7 — Создание и заполнение новой таблицы

Подсчитаем количесто строк во временной таблице HBase, созданной в предыдущем шаге на Рисунке 8.

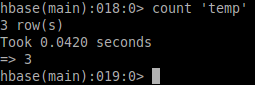


Рисунок 8 — Количество строк

Добавим еще значений в таблицу на Рисунке 9.

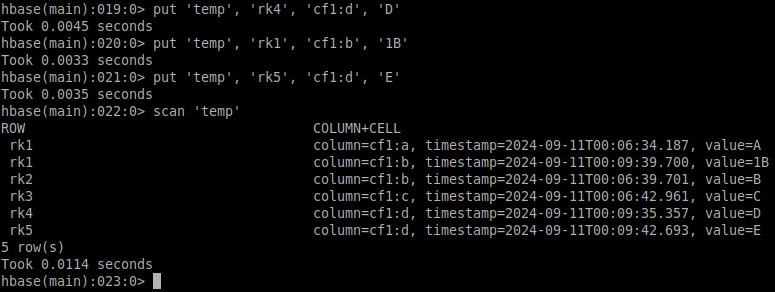


Рисунок 9 — Добавление значений в таблицу

Изменим атрибут version cf1 в таблице temp на 3 и проверим, правильно ли отражено изменение на Рисунке 10.

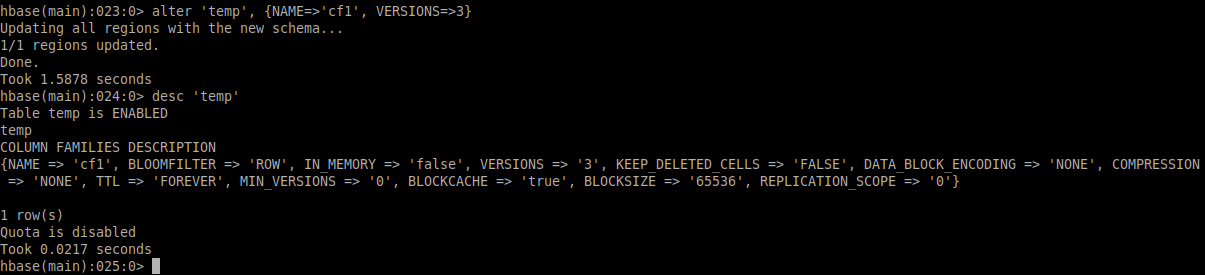


Рисунок 10 — Изменение атрибута version cf1

Измените временную таблицу, добавив семейство столбцов cf2 и cf3 с помощью команды ALTER. Выполним команду desc, чтобы проверить, добавлены ли cf2 и cf3 на Рисунке 11.

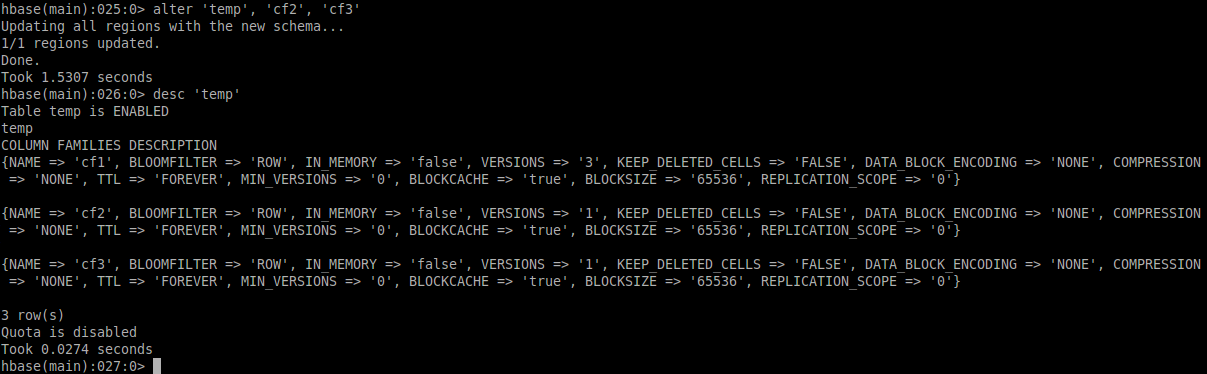


Рисунок 11 — Изменение временной таблицы

Удалим только что добавленный cf3 и проверим результат на Рисунке 12.

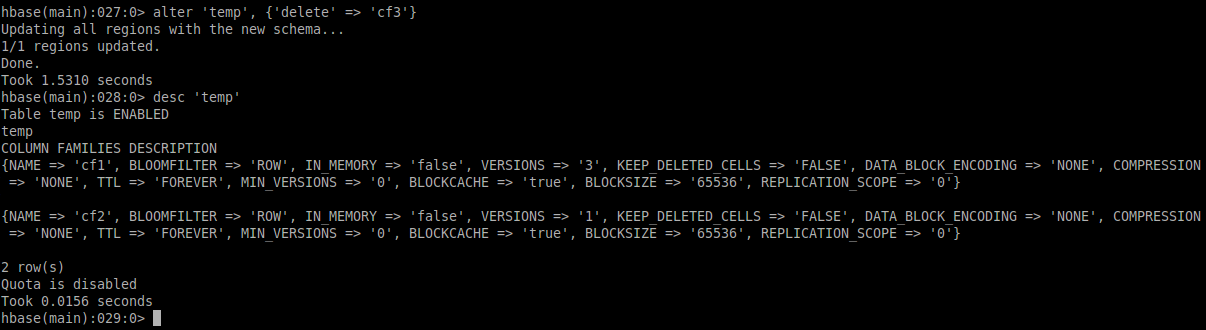


Рисунок 12 — Результат выполнения команды desc

Удалим временную таблицу. Выведем все таблицы, чтобы убедиться, что таблица была успешно удалена. Выйдем из оболочки HBase с помощью команды quit на Рисунке 13.

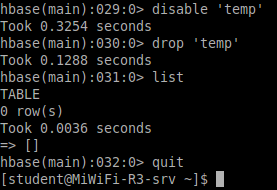


Рисунок 13 — Удаление временной таблицы и выход из оболочки HBase

Создадим файл rubyscript.rb с командами HBase на Рисунке 14.

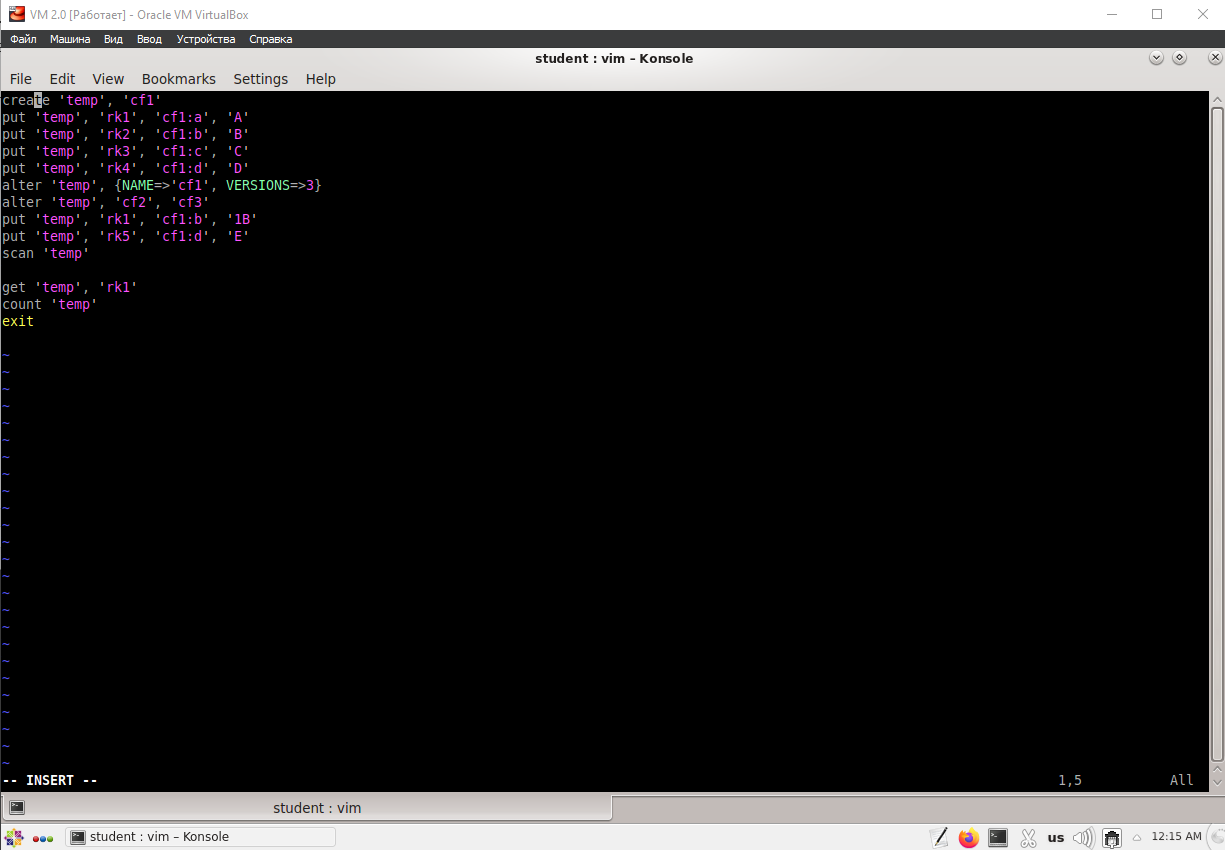


Рисунок 14 — Создаем файл rubyscript.rb

Команда, используемая для создания файла rubyscript.rb показана на Рисунке 15.



Рисунок 15 — Команда для создания и заполнения файла

Выполним этот файл из оболочки HBase на Рисунке 16.

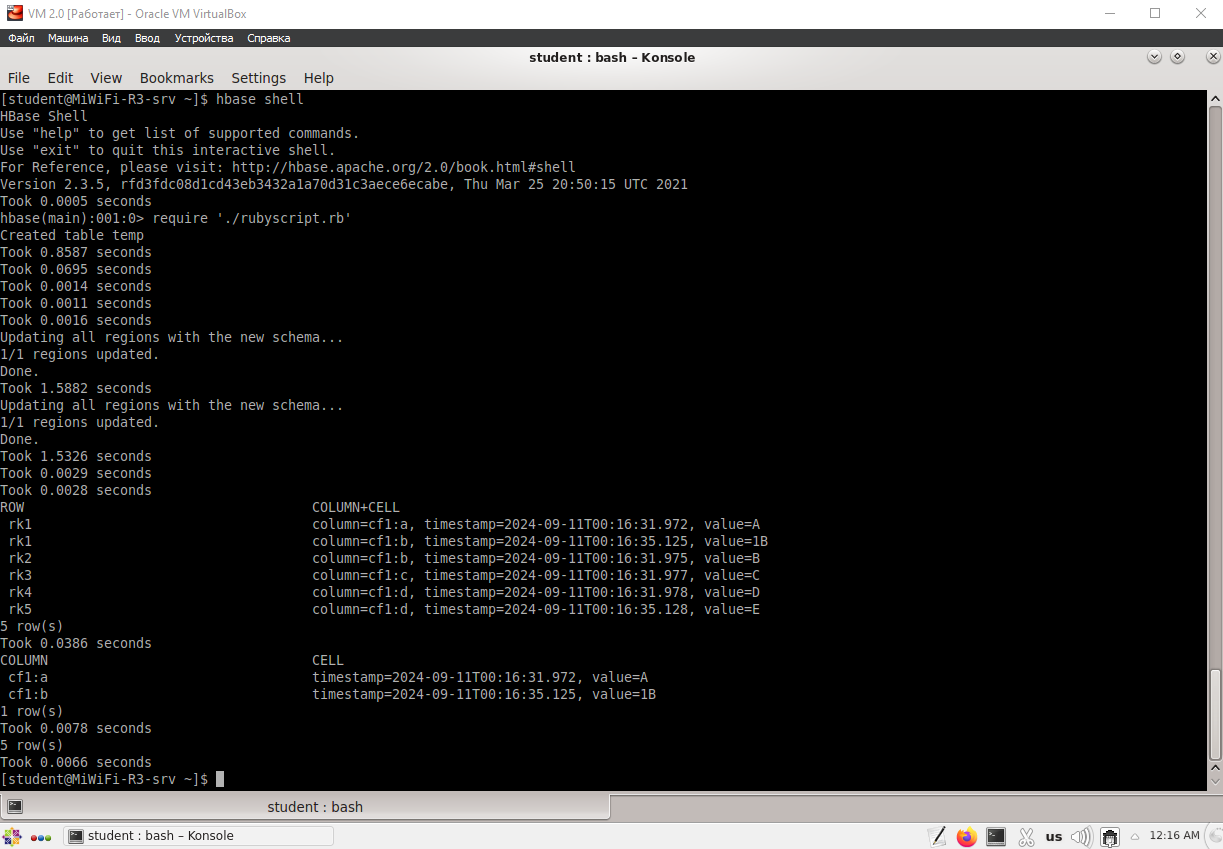


Рисунок 16 — Выполнение скрипта rubyscript.rb

Передадим команды в оболочку HBase в неинтерактивном режиме с помощью команды echo и оператора | (pipe) на Рисунке 17.

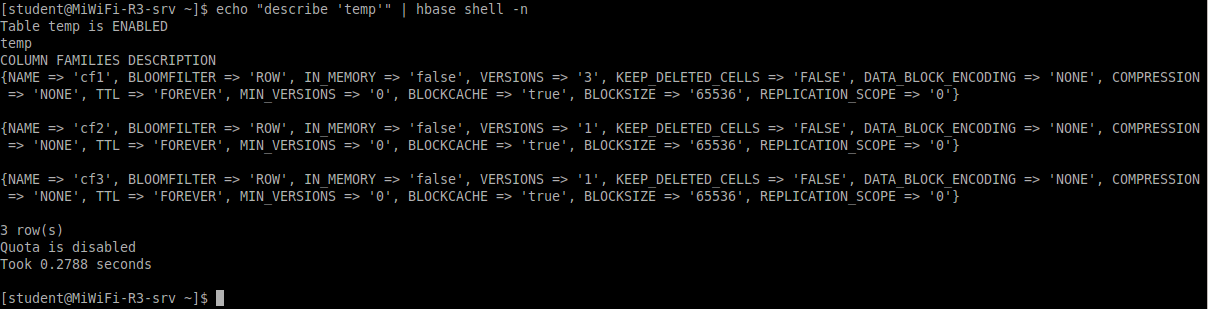


Рисунок 17 — Альтернативный запуск скрипта

### Доступ к данным с помощью команд DML

Введем новые данные и отобразим результат после запуск HBase на Рисунке 18.

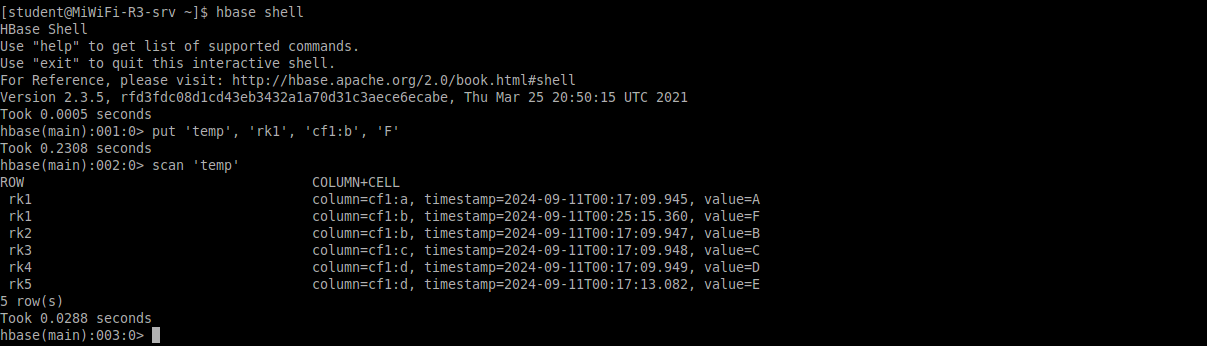


Рисунок 18 — Ввод и проверка новых значений

Получим строки с ключом rk1 из временной таблицы на Рисунке 19.

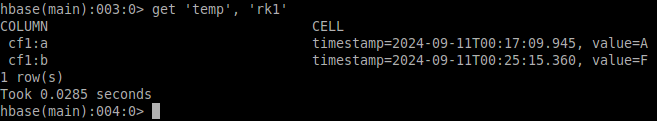


Рисунок 19 — Получение строки

Получим два значения предыдущих версий столбца 'b' в строке rk1 на Рисунке 20.

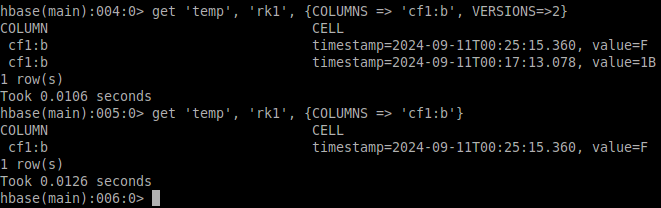


Рисунок 20 — Получение двух значений предыдущих версий столбца 'b' в строке rk1

Просмотрим всю таблицу на Рисунке 21.

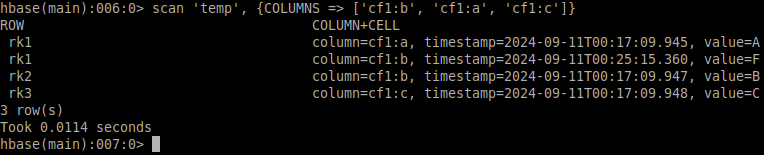


Рисунок 21 — Просмотр всей таблицы, но отображение только столбцов 'a', 'b', 'c'

Удалим ячейку из столбца 'b' из таблицы temp с ключом строки = rk1 на Рисунке 22.

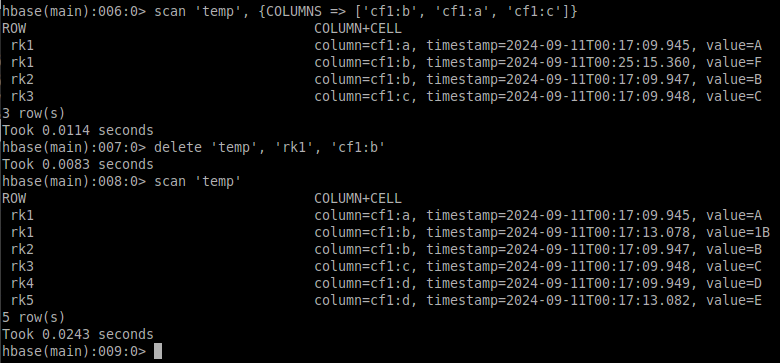


Рисунок 22 — Удаление ячейки из столбца 'b' из таблицы temp с ключом строки = rk1

Проверка на удаление на Рисунке 23.

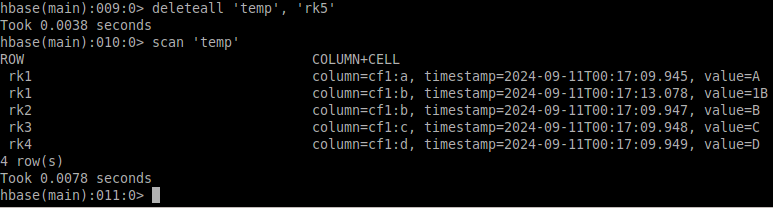


Рисунок 23 — Проверка

Выведите описание таблицы temp на Рисунке 24.

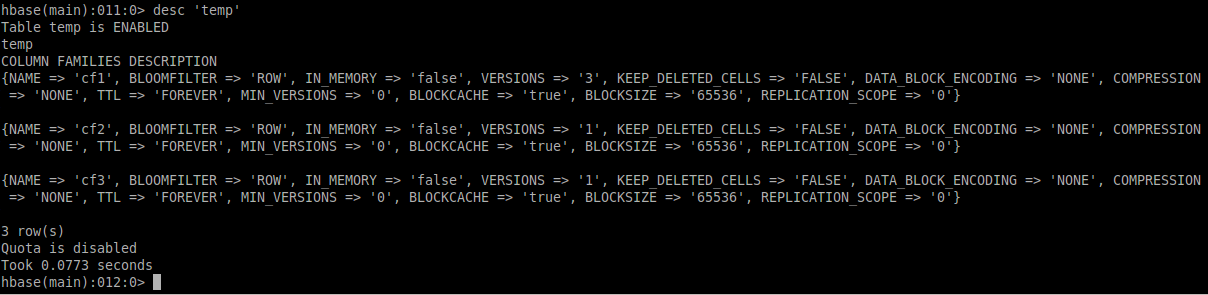


Рисунок 24 — Описание таблицы temp

Изменим значение с FOREVER на 30 секунд. Это означает, что версия будет удалена через 30 секунд после вставки данных в таблицу на Рисунке 25.

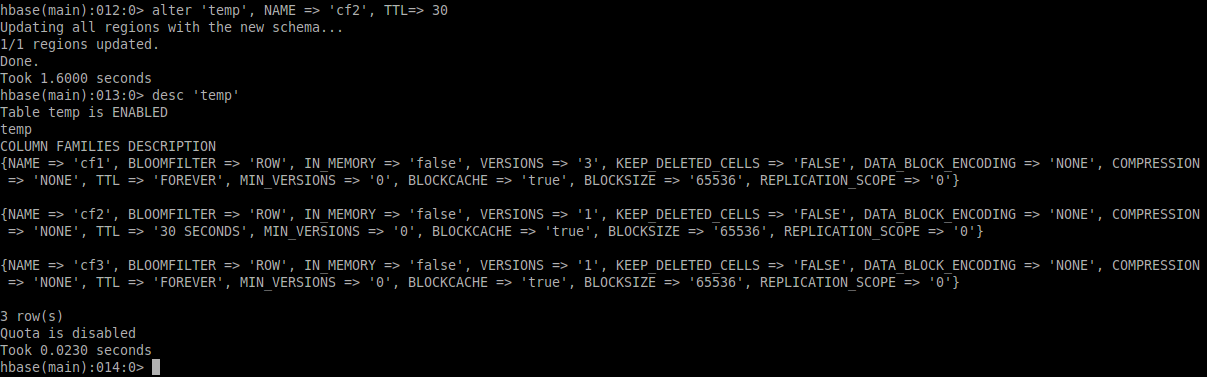


Рисунок 25 — Изменим значение

Вставим данные со следующими условиями и отобразим результаты на Рисунке 26.

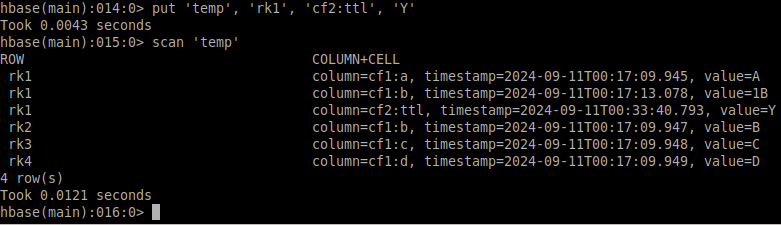


Рисунок 26 — Результаты после вставки

Подождав не менее 30 секунд, выполните команду сканирования еще раз, чтобы узнать, истек ли срок действия вставленной строки и была ли она удалена на Рисунке 27.

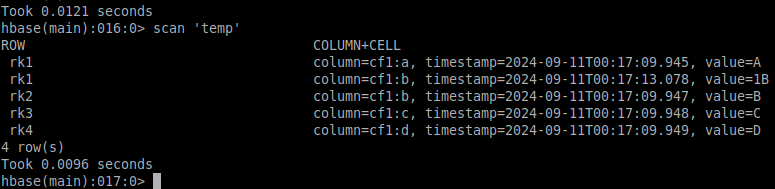


Рисунок 27 — Проверка удаления с задержкой в 30 секунд

Установим TTL на 10 секунд и установим MIN\_VERSIONS на 1 для cf2на Рисунке 28.

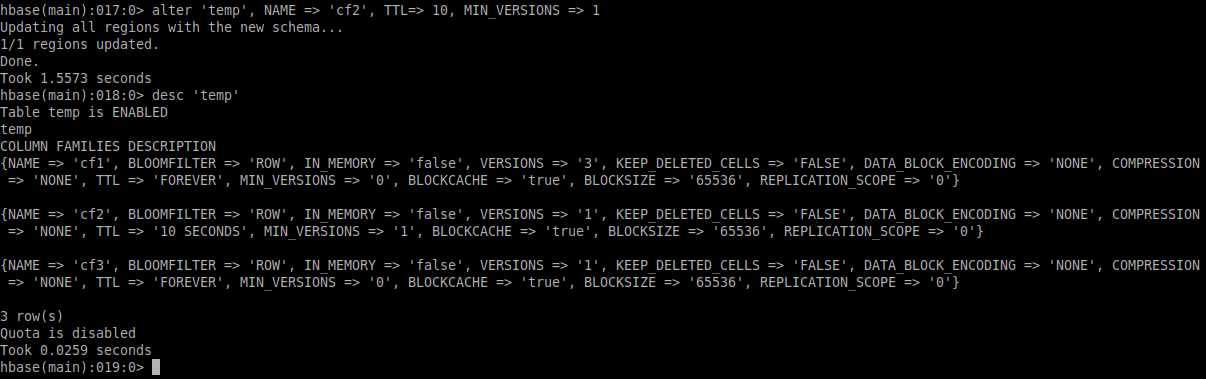


Рисунок 28 — Изменение параметров

Проверим значения TTL и MIN\_VERSIONS в cf2на Рисунке 29.

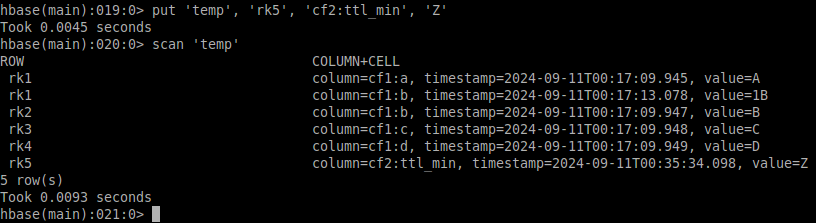


Рисунок 29 — Проверка значений

Вставим данные со следующими условиями и отобразим результаты. Срок действия вставленной строки неограничен на Рисунке 30.

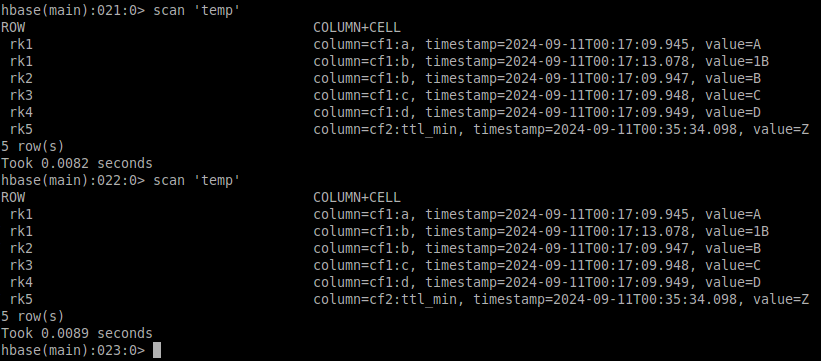


Рисунок 30 — Проверка вставленной строки через 10 секунд

### Самостоятельная работа с HBase

Создадим таблицу “movie” на Рисунке 31.

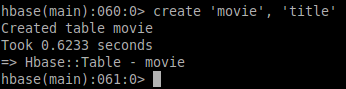


Рисунок 31 — Таблица “movie”

Добавим 3 фильма в эту таблицу на Рисунке 32.

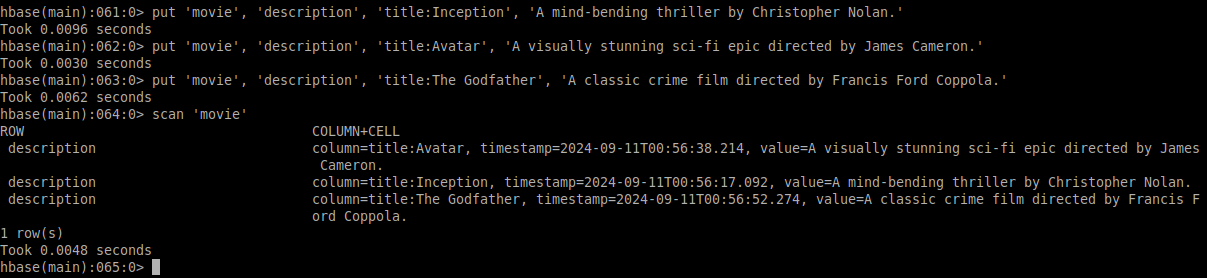


Рисунок 32 — Добавление фильмов

Создадим вторую таблицу для рейтингов фильмов и заполним ее на Рисунке 33.

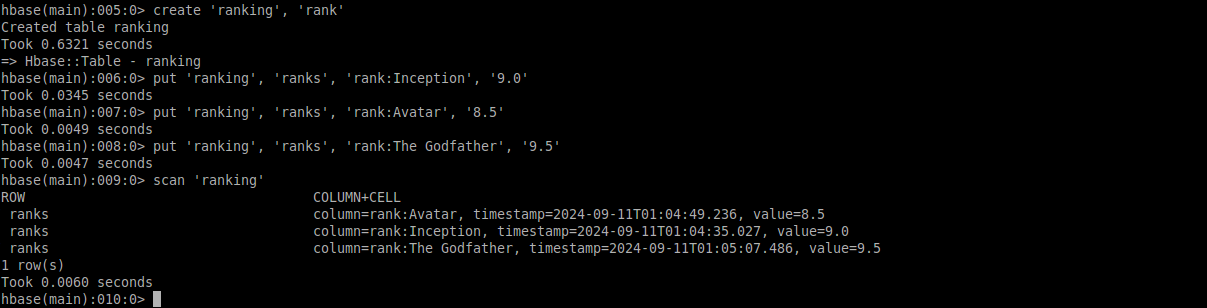


Рисунок 33 — Создание и заполнение таблицы ranking

Используем команду alter, чтобы добавить семейство столбцов fl\_title на Рисунке 34.

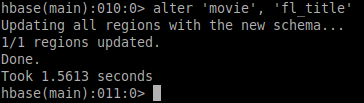


Рисунок 34 — Использование команды atler

Просмотрим описание таблицы “movie” на Рисунке 35.

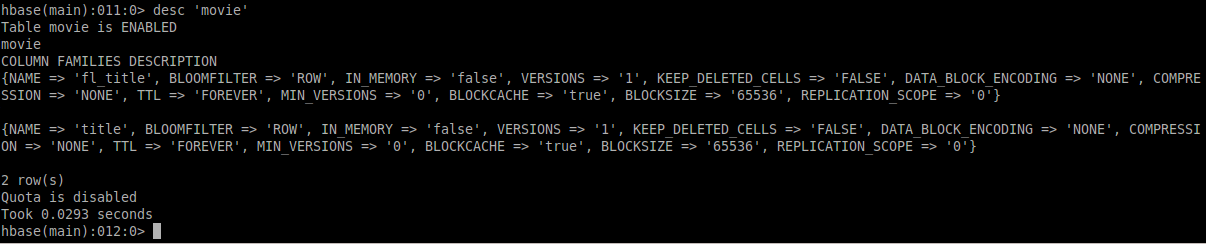


Рисунок 35 — Описание таблицы “movie”

Создадим третью таблицу под названием “user” и заполним ее тремя значениями на Рисунке 36.



Рисунок 36 — Создание и заполнение таблицы “user”

Используем команду alter, чтобы добавить семейство столбцов. После чего добавим ещё 2 значения возраста в таблицу “user” на Рисунке 37.

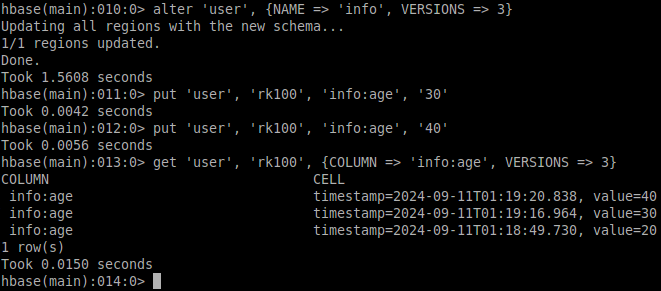


Рисунок 37 — Использование команды alter и добавление новых значений

Посмотрим содержимое таблицы “user” на Рисунке 38.

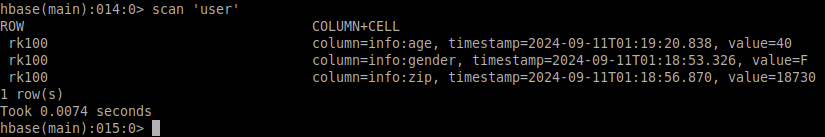


Рисунок 38 — Содержимое таблицы “user”

Удалим все вхождения возраста и проверим результат на Рисунке 39.

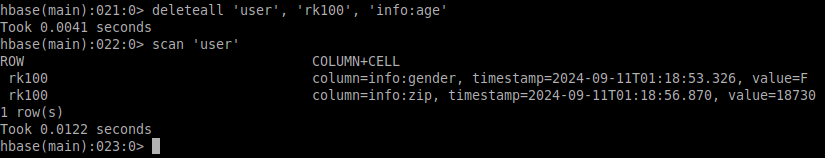


Рисунок 39 — Удаление всех вхождений возраста и проверка

Удалим оставшиеся значения в таблице, оставив ее пустой на Рисунке 40.

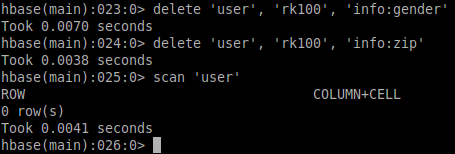


Рисунок 40 — Удаление оставшихся значений и проверка таблицы