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

Институт информационных технологий (ИИТ)

Кафедра математического обеспечения и стандартизации информационных технологий (МОСИТ)

**ОТЧЕТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ**

по дисциплине «Распределенные системы управления базами данных»

**Практическое занятие № 3**

|  |  |  |  |
| --- | --- | --- | --- |
| Студент группы ИНБО-01-17 | *ИКБО-11-22, Гришин Андрей Валерьевич* | (подпись) | |
| Преподаватель | *Красников Степан Альбертович* | (подпись) | |
| Отчет представлен | «\_\_\_»\_\_\_\_\_\_\_\_202\_\_г. | |  | |

Москва 2025 г.

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

1.1 Что такое TTL

Столбцы и таблицы поддерживают необязательный срок действия,

называемый TTL (time-to-live); TTL не поддерживается для столбцов счетчиков.

Значение TTL задается в секундах. Срок действия данных истекает, когда они

превышают период TTL, после чего они помечаются меткой tombstone. Данные

с истекшим сроком действия по-прежнему доступны для запросов на чтение в

течение определенного периода. Обычные процессы сжатия данных и починки

автоматически удаляют tombstones.

1.2 Что такое Tombstones и для чего они нужны

В Cassandra удаленные данные не удаляются сразу с диска. Вместо этого

Cassandra записывает специальное значение, известное как tombstone, чтобы

указать, что данные были удалены.

Назначение:

Эти значения предотвращают возврат удаленных данных во время чтения

и в конечном итоге позволяют удалить данные посредством сжатия.

Tombstones являются частью реализации модели хранения данных (во

время записи существующие данные не читаются)

Tombstones — это записи — они проходят обычный путь записи,

занимают место на диске и используют механизмы согласованности Cassandra.

Если кластер управляется правильно, это гарантирует, что данные останутся

удаленными, даже если узел не работает, когда выполняется удаление.

Tombstones генерируются при помощи:

⎯ операции DELETE;

⎯ установки TTL;

⎯ вставки null значений;

⎯ вставки данных в части коллекций.

Tombstones написаны с отметкой времени. В идеальных условиях

они (и связанные с ними данные) будут удалены во время сжатия данных по

прошествии определенного времени.

Практическая часть:

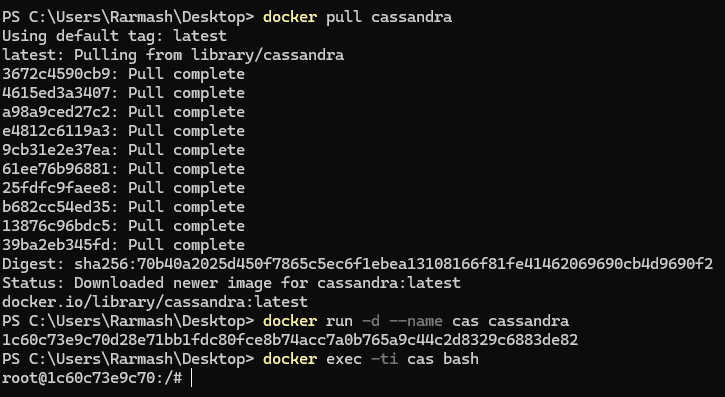


Рисунок 1 – Создание и вход в контейнер Cassandra

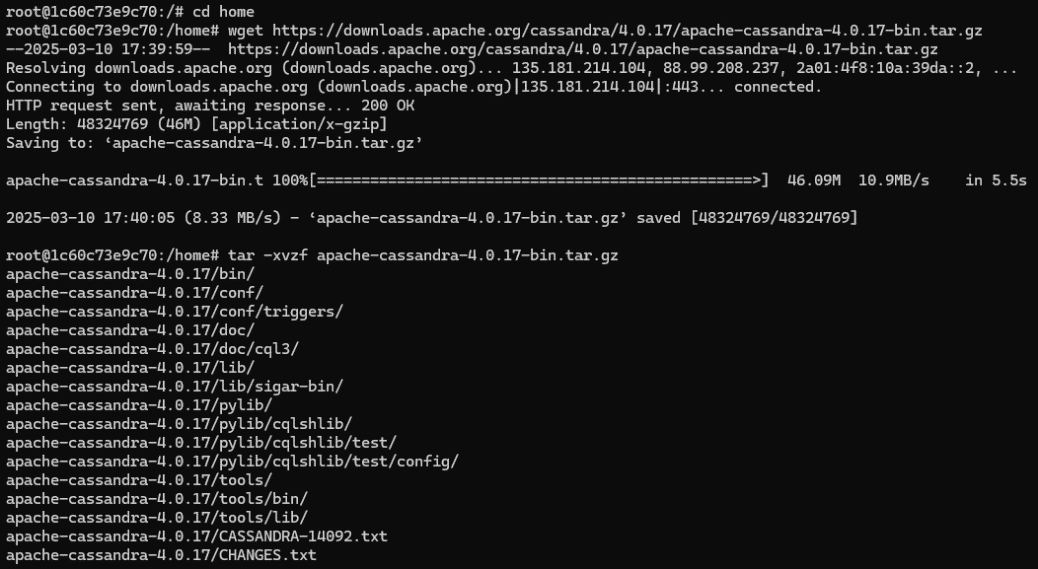
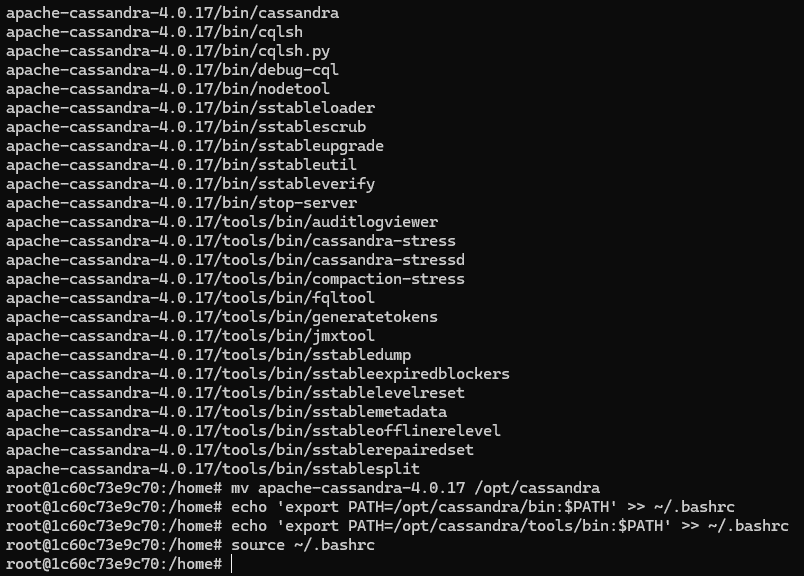
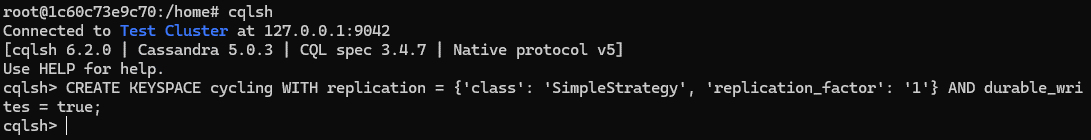
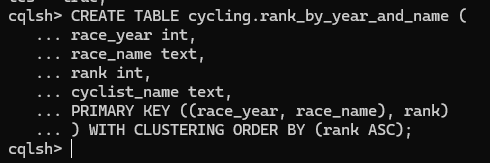
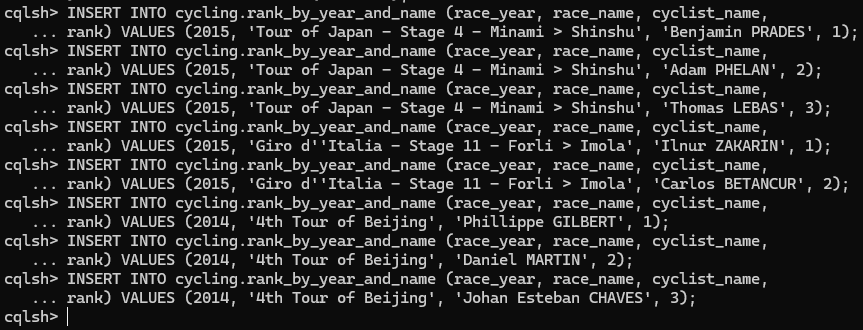


Рисунок 2 – Сборка инструментов Cassandra







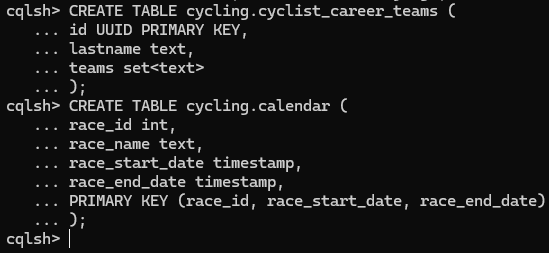
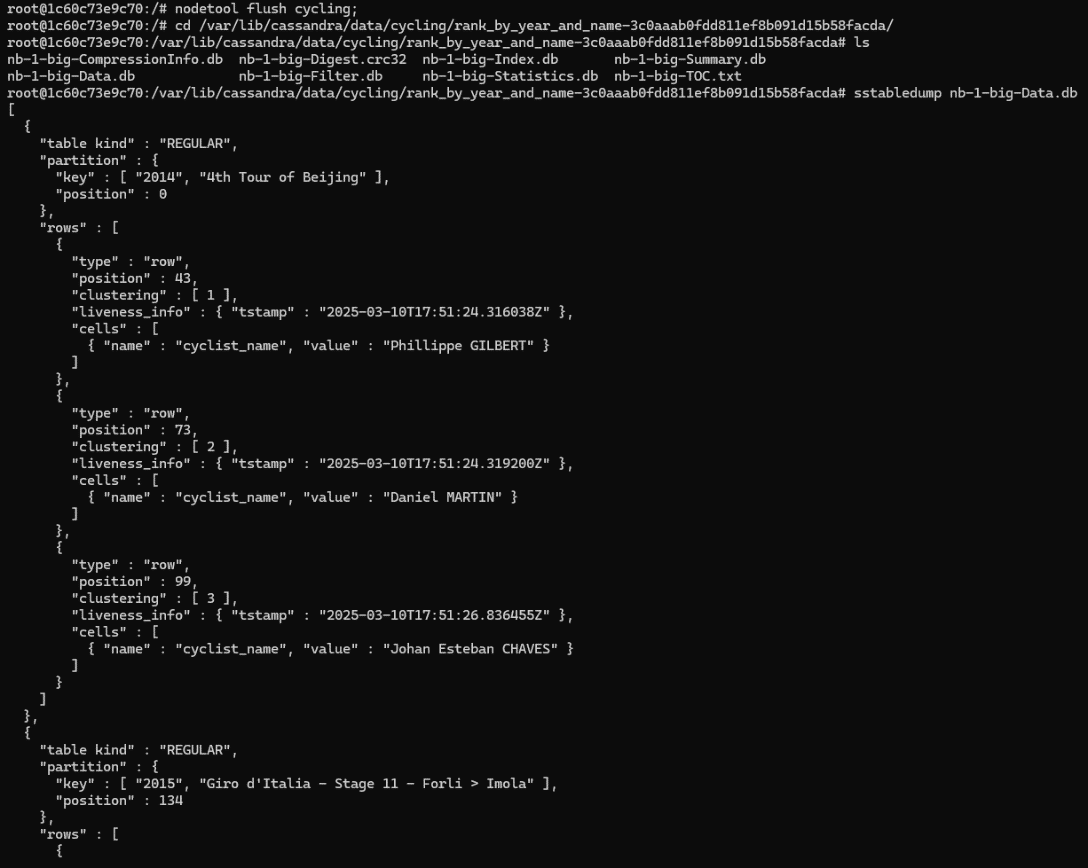


Рисунок 3 - Создание keyspace и таблиц



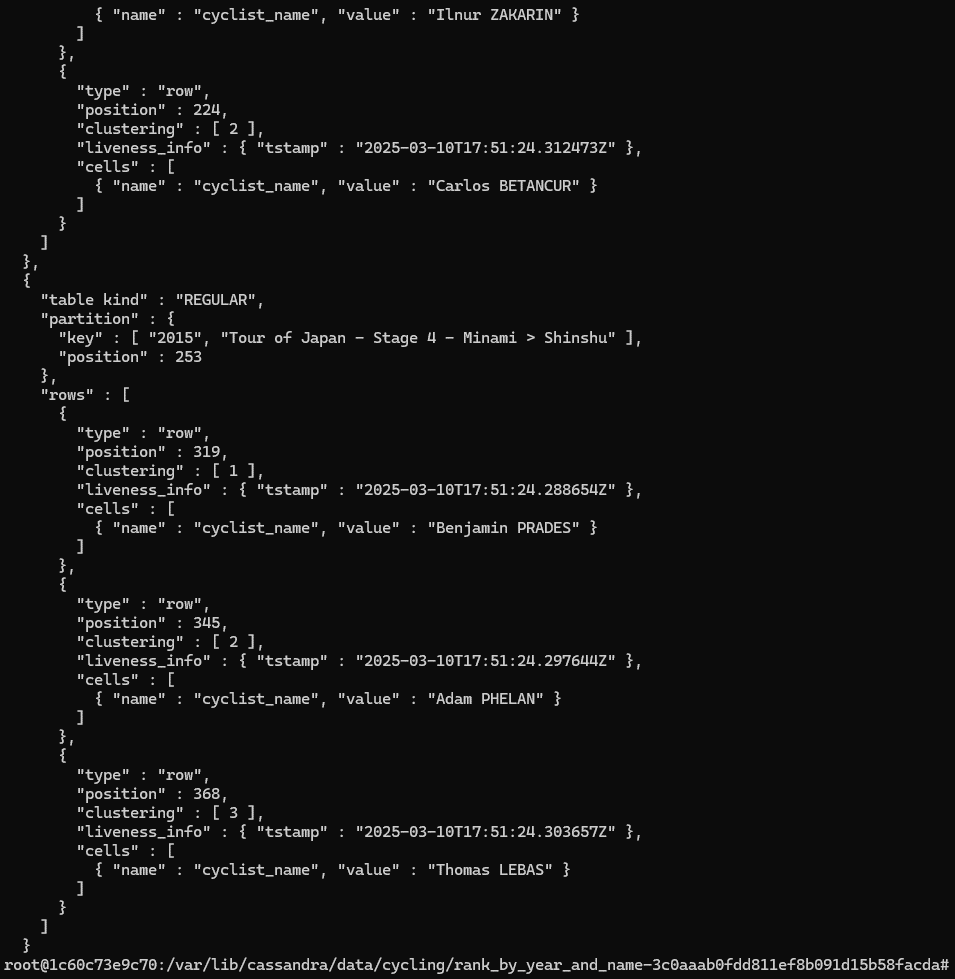


Рисунок 4 - Выгрузка данных в SSTable

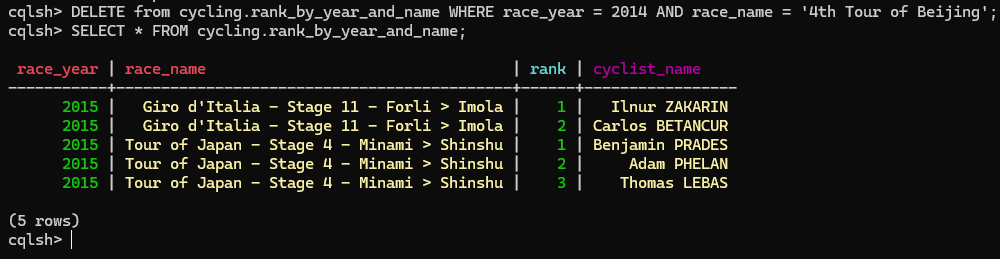


Рисунок 5 – Удаление записи из таблицы

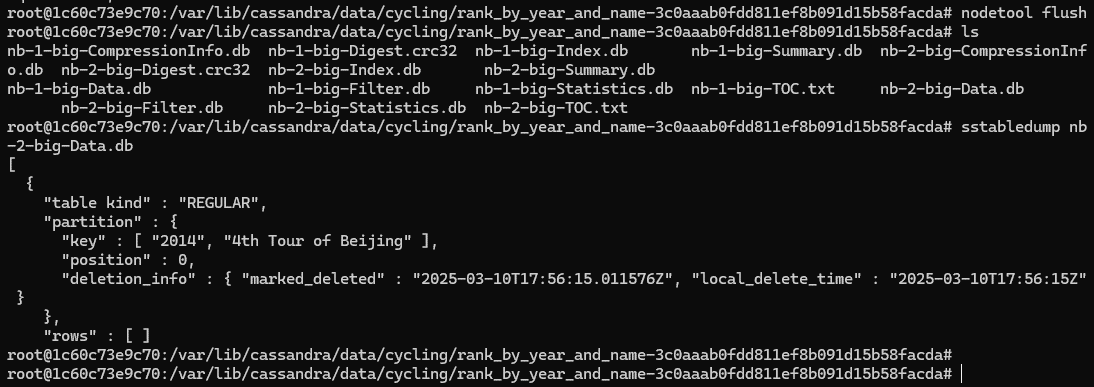
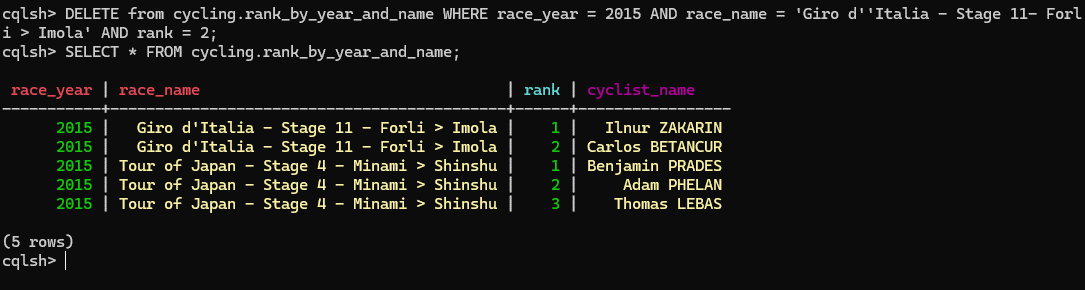


Рисунок 6 – Выгрузка Tombstone для разделов



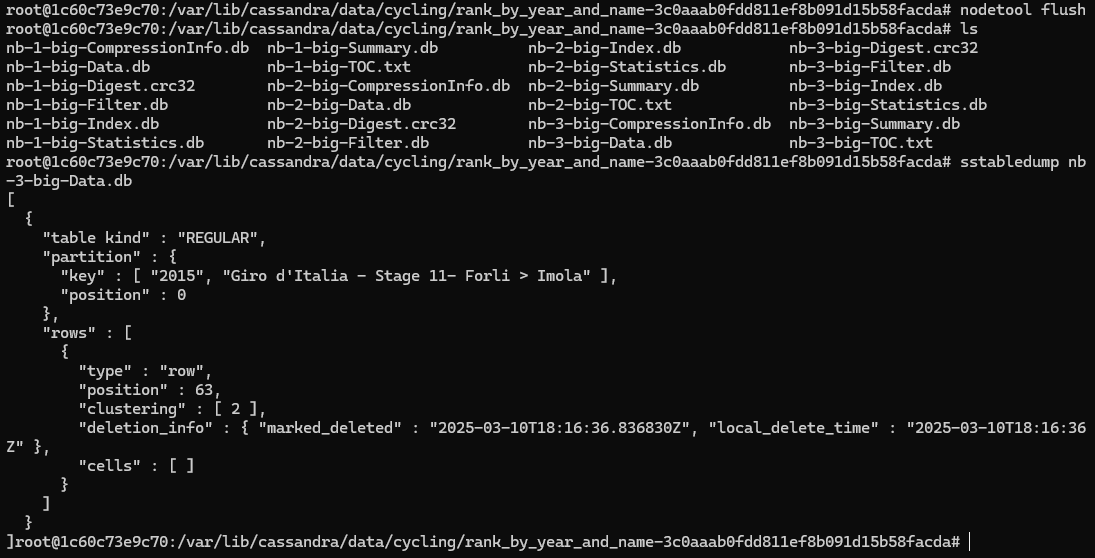


Рисунок 7 – Удаление строки и выгрузка Tombstone

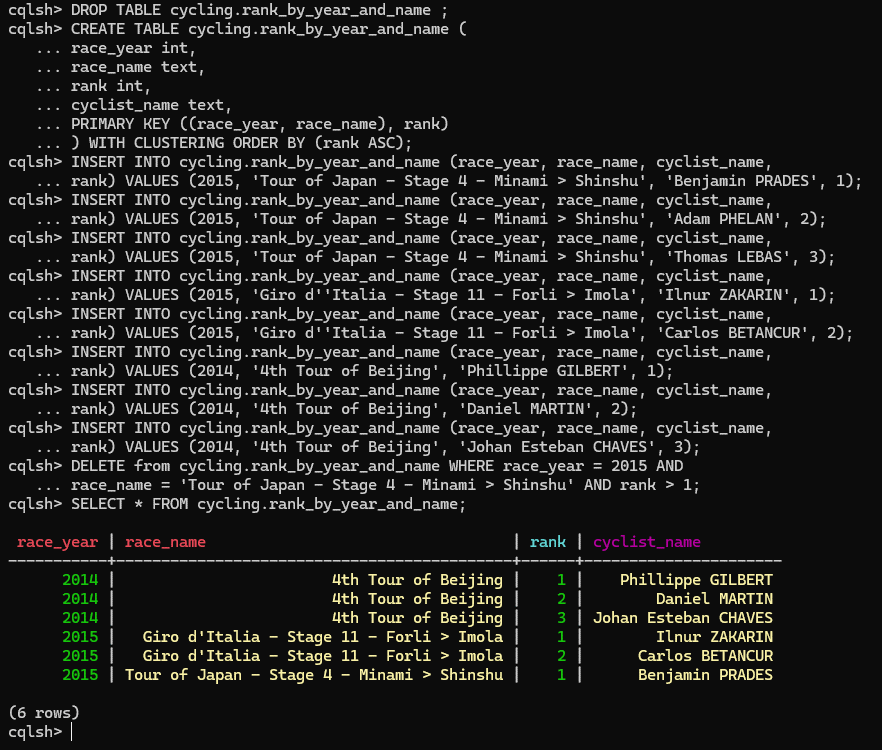
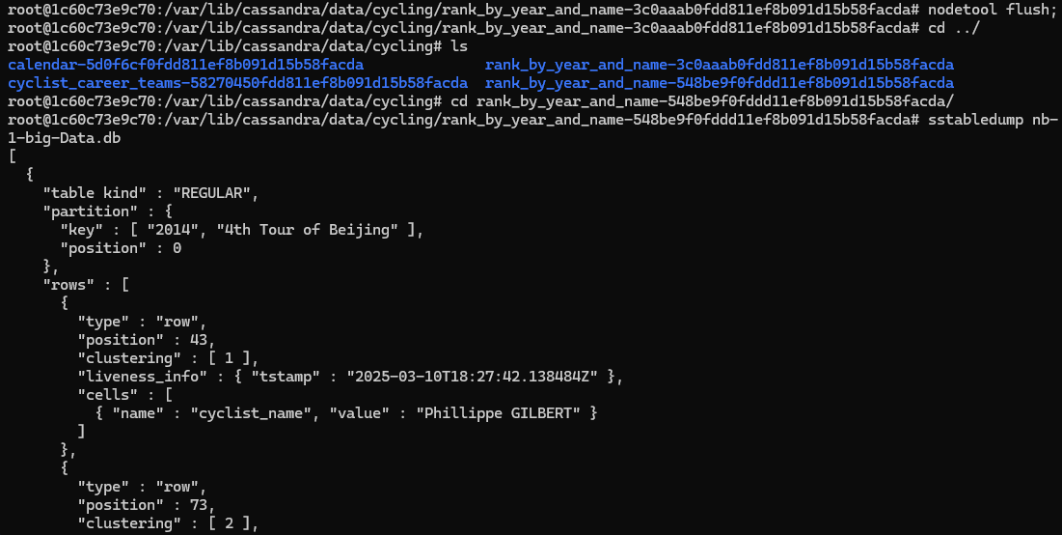
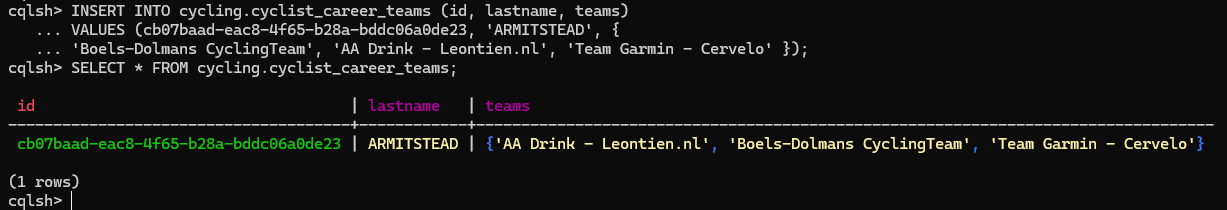


Рисунок 8 – Повторное создание и заполнение таблицы c последующим удалением записи

Рисунок 9 – Использование sstabledump для определения изменений в диапазоне



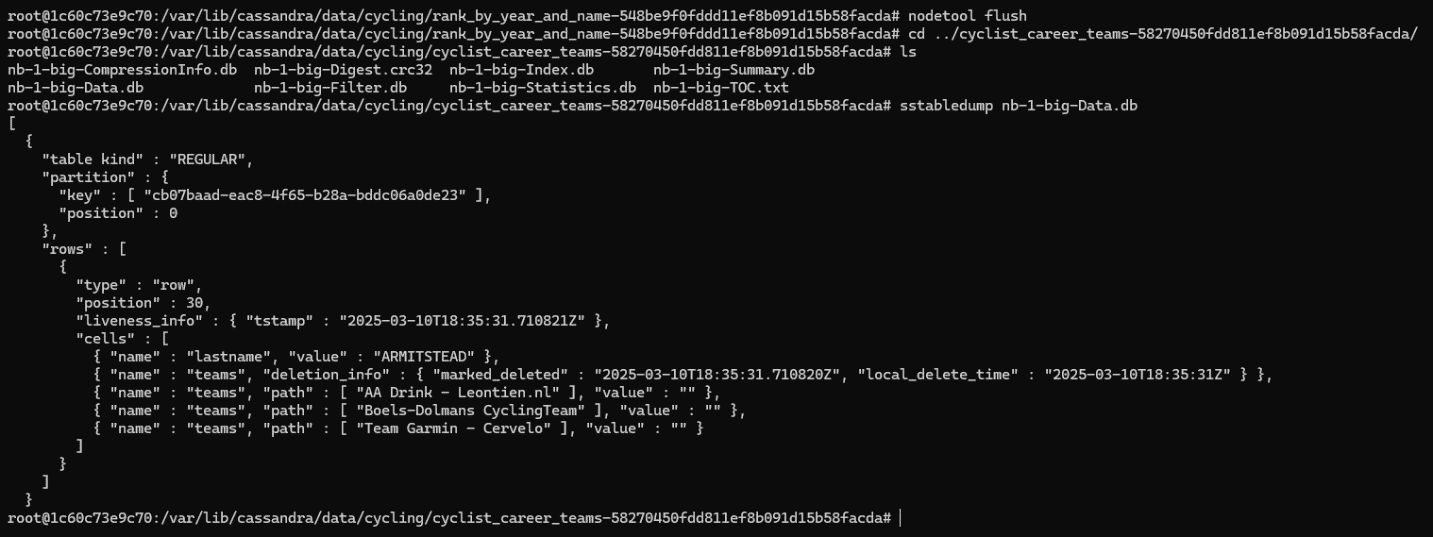
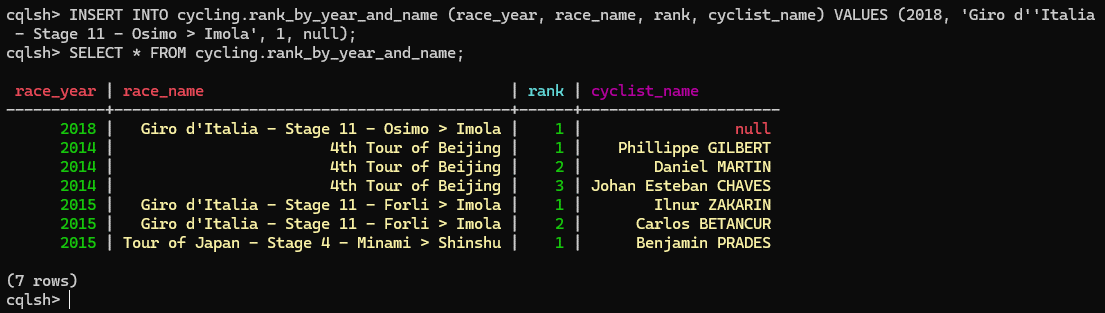


Рисунок 10 – Вставка данных в таблицу и определение изменений в столбцах

****

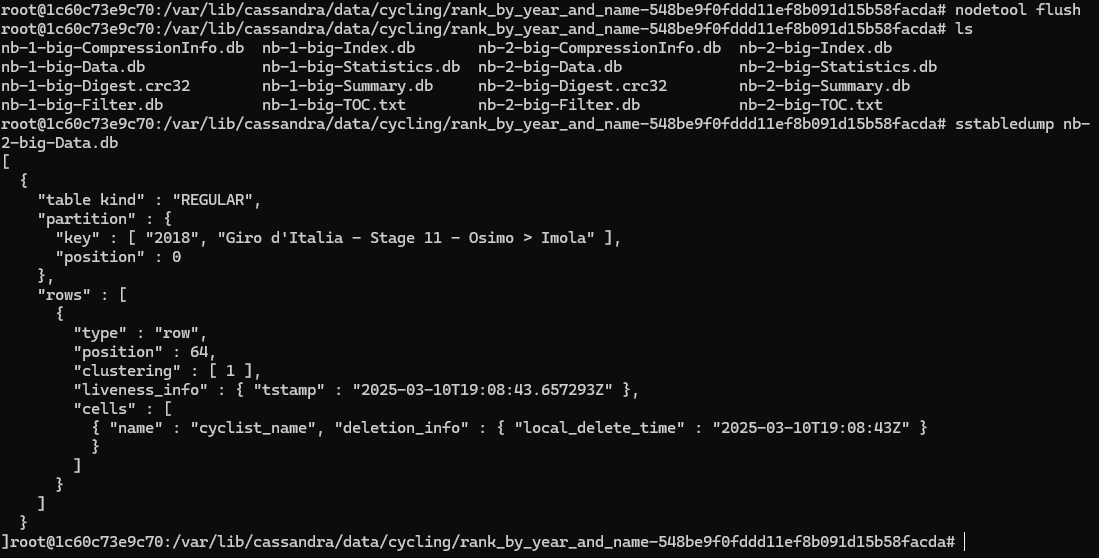


Рисунок 11 – Вставка null в ячейку и определение изменений для неё

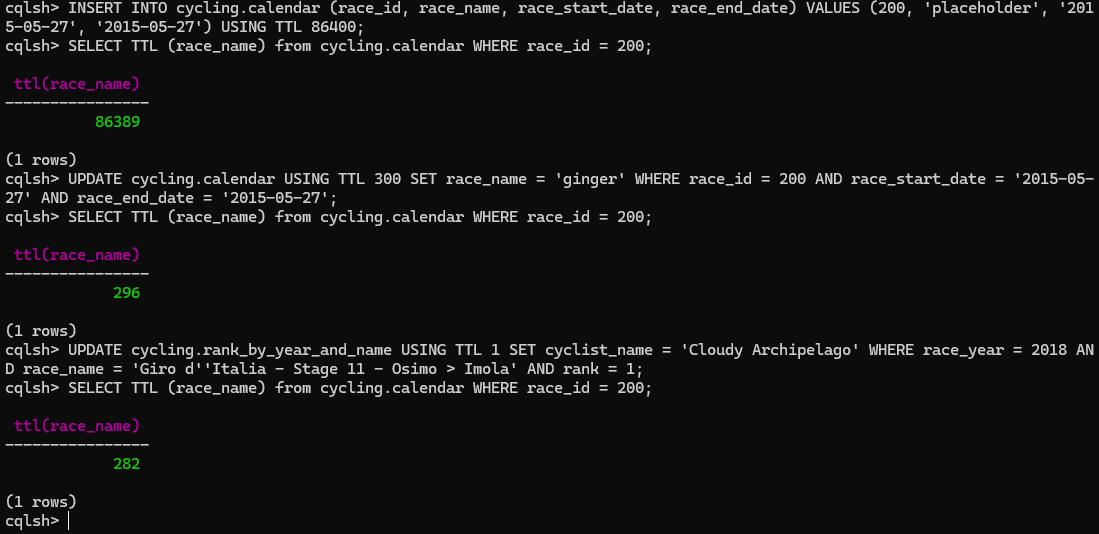
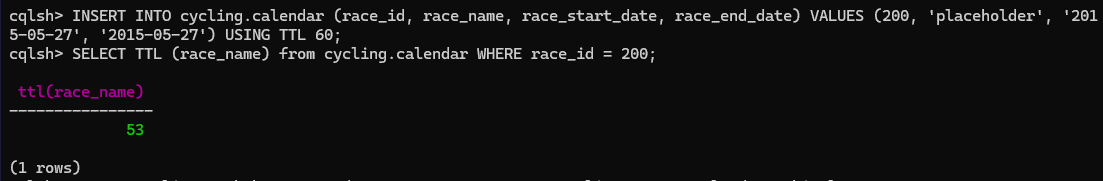


Рисунок 12 – Установка времени жизни записи



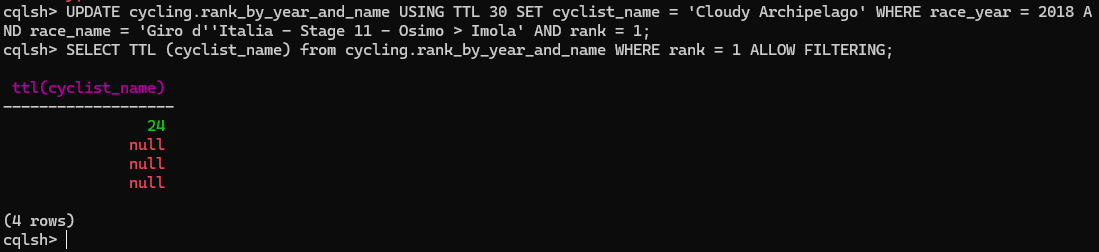


Рисунок 13 - Установка времени жизни записи

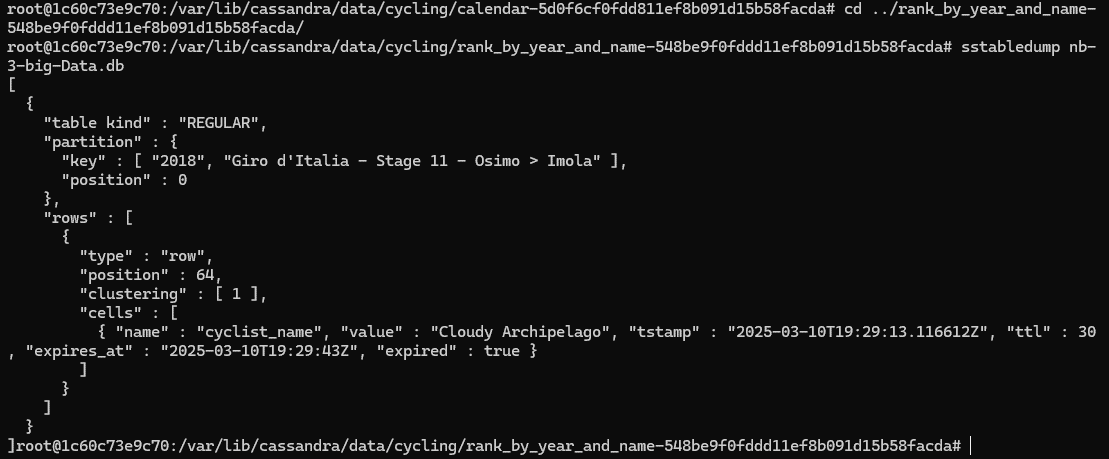


Рисунок 14 – Выгрузка TTL tombstones

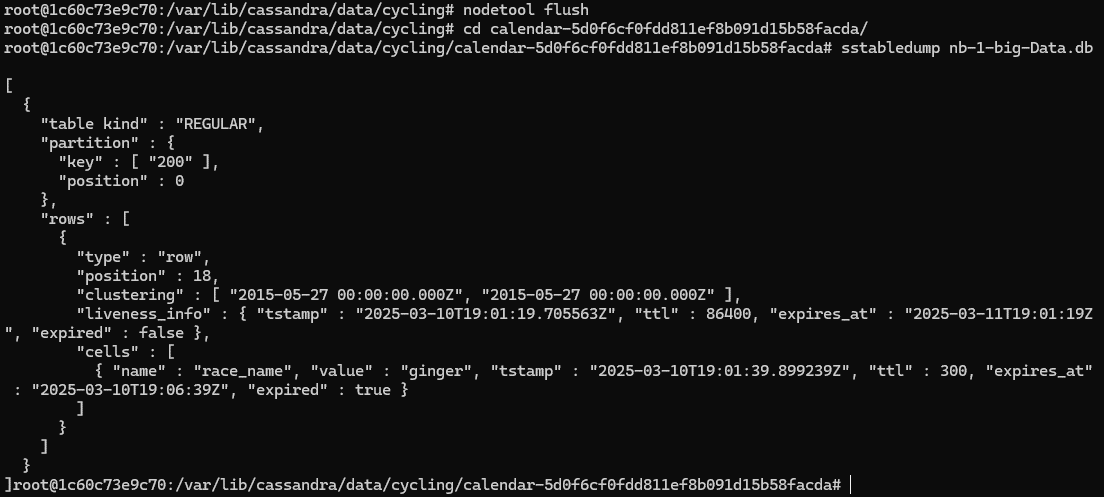


Рисунок 15 - Выгрузка TTL tombstones для отдельной ячейки

Заключение:

В ходе практического исследования были проанализированы механизмы TTL (Time-To-Live) и Tombstones в распределенных СУБД. Основное внимание уделено особенностям удаления данных и их связи с производительностью системы. Ключевые результаты:

1. **TTL-удаление**
   * Автоматическое удаление данных по истечении заданного времени демонстрирует эффективность, однако требует контроля за нагрузкой на сборку мусора (GC) для предотвращения снижения производительности.
2. **Типы Tombstones**
   * Изучены разновидности логического удаления: разделы, строки, диапазоны, ячейки и ComplexColumn-структуры. Каждый тип актуален для конкретных сценариев, но создает дополнительную нагрузку при обработке запросов.
3. **Производительность и оптимизация**
   * Неконтролируемое накопление Tombstones приводит к замедлению чтения из-за фильтрации "мусорных" записей и увеличению времени компактификации.
   * Для минимизации негативного влияния предложены меры: адаптация сроков хранения Tombstones под рабочие нагрузки, регулярная ручная или автоматическая компактификация данных.

TTL и Tombstones — мощные инструменты управления жизненным циклом данных, но их использование требует баланса между функциональностью и производительностью. Оптимальная настройка параметров (время хранения, частота обслуживания данных) и мониторинг системы критически важны для предотвращения деградации скорости обработки запросов.

Вопросы для самопроверки:

1. **Какая опция для таблицы устанавливает срок чтения удаленных данных?**  
   В Cassandra срок чтения удаленных данных (tombstones) управляется параметром таблицы gc\_grace\_seconds. Этот параметр определяет, сколько секунд Cassandra будет хранить tombstone (маркер удаления) перед тем, как он будет окончательно удален в процессе compaction. По умолчанию gc\_grace\_seconds равен 864000 секунд (10 дней).
2. **В чем заключается стратегия разрешения конфликтов last-write-win?**  
   Стратегия **Last-Write-Win (LWW)** используется в Cassandra для разрешения конфликтов при одновременной записи в одну и ту же ячейку данных. Суть стратегии:

* Cassandra использует временные метки (timestamps) для каждой операции записи.
* При конфликте (например, если два клиента одновременно обновляют одну и ту же ячейку) Cassandra выбирает запись с более поздней временной меткой (последнюю запись) как актуальную.
* Эта стратегия проста и эффективна, но может приводить к потере данных, если более ранняя запись не была учтена.

1. **Назвать 3 варианта применения tombstone.**

**Tombstone** — это специальный маркер, который Cassandra использует для обозначения удаленных данных. Вот три варианта их применения:

1. **Удаление строки или столбца**:
   * Когда данные удаляются (например, с помощью команды DELETE), Cassandra не удаляет их физически сразу, а создает tombstone, чтобы пометить данные как удаленные.
   * Это позволяет Cassandra распространить информацию об удалении на другие узлы в кластере.
2. **Истечение срока жизни (TTL)**:
   * Если для данных установлен TTL (Time-To-Live), по истечении этого срока Cassandra создает tombstone, чтобы пометить данные как устаревшие.
   * Tombstone будет храниться до тех пор, пока не истечет gc\_grace\_seconds.
3. **Обработка разрешения конфликтов**:
   * В распределенной системе могут возникать ситуации, когда одна реплика удаляет данные, а другая еще хранит их. Tombstone помогает синхронизировать данные между узлами, указывая, что данные были удалены.