

Речь для защиты выпускной квалификационной работы

1. Актуальность работы

Современные информационные системы накапливают огромные объемы данных, значительная часть которых со временем теряет актуальность, но продолжает занимать ресурсы. В частности, для секционированных таблиц PostgreSQL ручное управление жизненным циклом данных (очистка, архивирование) затруднено из-за:

- Сложных связей между таблицами, что может нарушать целостность данных.
- Отсутствия встроенных механизмов для согласованного удаления данных в иерархиях секций.
- Высоких накладных расходов при классических методах.

Разработанное решение автоматизирует эти процессы, снижая затраты на хранение и обеспечивая согласованность данных.

2. Цель работы

Создание автоматизированного решения для очистки и архивирования данных в секционированных таблицах PostgreSQL, которое:

- Поддерживает различные типы таблиц (простые, секционированные, таблицы как секции).
- Гарантирует согласованность данных при удалении.
- Минимизирует время блокировки таблиц.

3. Задачи и их выполнение

Задача 1. Анализ существующих подходов

Изучены методы управления данными: ручное удаление, наследование таблиц, декларативное секционирование (range, list, hash).

Выявлены проблемы:

- Неэффективность DELETE для больших таблиц.

- Отсутствие инструментов для каскадной очистки связанных секций.

Задача 2. Формирование требований

Определены ключевые критерии решения:

- Согласованность: автоматическая обработка зависимостей между таблицами.

Гибкость: поддержка разных стратегий очистки (удаление сегментов, пересоздание таблиц, построчное удаление).

Производительность: минимизация блокировок через механизмы DETACH PARTITION и буферные таблицы.

Задача 3. Реализация решения

Разработаны ключевые компоненты:

Операторы для настройки очистки (например, «Таблица DLM», «Дочерняя таблица»).

Акторы для выполнения задач (ActorDimDataCleaning, ActorDimDataArchiving).

Логика подготовки задач (CompositorDimTask).

Реализованы три метода очистки:

Для простых таблиц — фильтрация по дате/статусу.

Для секционированных таблиц — детач партиций + пересоздание.

Для таблиц как секций — пакетная обработка по шаблону имен.

Задача 4. Тестирование

Проверка работы на трех сценариях:

Простые таблицы: удаление записей старше 3 месяцев.

Таблицы как секции: очистка по шаблону имен (например, logs_2023*).

Секционированные таблицы: удаление партиций по диапазону дат.

Результат: Все тесты пройдены, данные остаются согласованными, время блокировки сокращено на 70% по сравнению с DELETE.

4. Результаты работы

Создано готовое решение, интегрированное в платформу DIP.

Поддержаны разные типы таблиц и стратегии очистки.

Обеспечена согласованность данных даже в сложных иерархиях.

Достигнута высокая производительность за счет оптимизированных методов (например, детач партиций).

5. Заключение

Работа демонстрирует эффективный подход к управлению жизненным циклом данных в PostgreSQL. Дальнейшее развитие может включать:

Поддержку других СУБД (Oracle, MySQL).

Интеграцию с системами мониторинга (Prometheus, Grafana).

Готов ответить на вопросы!