

Библиотека транзакционного доступа к файлам из PostgreSQL

Семенов Александр Сергеевич

Научный руководитель:
доц. каф. СП, к.ф.-м.н. Д.В. Луцев

25.01.2024

Актуальность (1)

- Одна из составляющих любой информационной системы — данные
- Для хранения часто используются базы данных
- Один из типов хранимых данных — бинарные файлы
 - pdf-документы, изображения, архивы, медицинские данные...
- С ростом объемов бинарных данных становится необходимым их эффективное хранение

Актуальность (2)

Для многих информационных систем важна:

- Транзакционная поддержка операций над данными
- Возможность версионирования данных
- Размер базы данных
- Производительность

Постановка задачи

Целью работы является написание библиотеки для транзакционного доступа к бинарным данным

Задачи:

- Изучить существующие подходы к решению задачи хранения бинарных данных в PostgreSQL
- Определить подход к хранению бинарных данных
- Реализовать библиотеку для написания расширения PostgreSQL для транзакционного доступа к бинарным данным
- Провести тестирование полученного решения

PostgreSQL: тип bytea

- bytea — тип данных для хранения произвольного набора байтов
- Применяется механизм TOAST¹ для управления данными > 8Kб
- Недостатки TOAST:
 - Снижение производительности работы с большими объектами данных
 - Требуется дополнительное место для хранения TOAST таблиц
 - Избыточное обновление
- Не поддерживается версионирование

¹The Oversized-Attribute Storage Technique

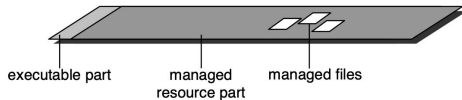
Система контроля версий Git

- Позволяет версионировать бинарные данные, но плохо
- Git умеет определить, что бинарный файл изменился, но не умеет определять, в каком именно месте произошло изменение
- Из-за этого любое изменение файла, даже самое небольшое, будет приводить к созданию копии файла
- Хотя для небольших данных может использоваться
 - AcidFS²

²<https://docs.pylonsproject.org/projects/acidfs/en/latest/>

Smart files³

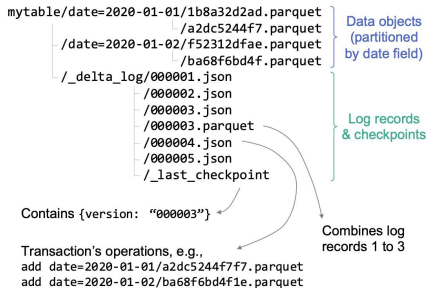
- Умные файлы являются исполняемыми, внутри есть своя файловая система, которая даёт возможность протоколировать все операции и поддерживает механизм версионирования
- Однако такой подход нельзя в полной мере назвать подходящим для решения задачи
 - требует пересмотра процессов работы с файлам
 - размер таких файлов будет больше



³Smart files: combining the advantages of DBMS and WfMS with the simplicity and flexibility of spreadsheets

Delta Lake⁴

- Слой хранения данных, который поддерживает ACID гарантии
- Использует формат файлов Parquet для хранения данных
- Помимо этого используется транзакционный журнал



⁴Delta Lake: High-Performance ACID Table Storage over Cloud Object Stores

Выбор файловой системы (1)

- Выбор между EXT4, Btrfs, ZFS
- Критерии отбора:
 - Наличие библиотеки для взаимодействия с файловой системой из программного кода
 - Поддержка транзакционного механизма взаимодействия через механизм copy-on-write
 - Поддержка создания снапшотов за константное время для реализации версионирования

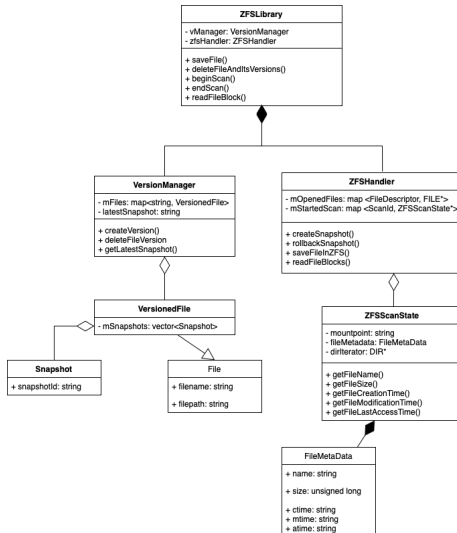
Выбор файловой системы (2)

- EXT4 не поддерживает механизм copy-on-write
- Btrfs показывает более низкую эффективность в сравнении с ZFS при работе с большими объёмами данных⁵
- Решено остановиться на ZFS и библиотеке OpenZFS⁶

⁵<https://www.enterprisedb.com/blog/postgres-vs-file-systems-performance-comparison>

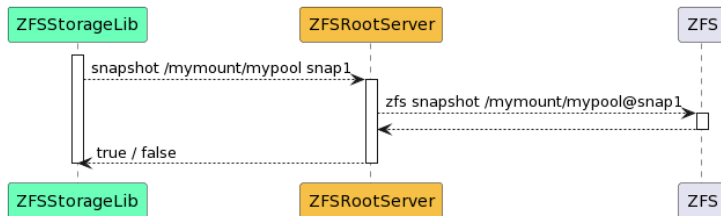
⁶<https://github.com/openzfs/zfs>

Диаграмма классов библиотеки



Разграничение адресного пространства

- Библиотека будет использоваться для написания расширения для СУБД PostgreSQL, которая работает в пользовательском адресном пространстве
- Однако выполнение некоторых функций файловой системы ZFS происходит на уровне ядра



Планы по тестированию

- Для тестирования функциональности библиотеки будет разработана отдельная программа, которая будет выполнять различные операции с файловым хранилищем в нескольких потоках
- Во время выполнения тестов будет проверяться согласованное состояния хранилища

Результаты

В результате были решены следующие задачи:

- Изучены существующие подходы к решению задачи хранения бинарных данных в PostgreSQL
- Определен подход к хранению бинарных данных
- Реализована библиотека для написания расширения PostgreSQL для транзакционного доступа к бинарным данным⁷

Планы для дальнейшей работы:

- Провести тестирование полученного решения

⁷Код проекта закрыт и принадлежит компании ООО “Датаджайл”