PGMeetup.DBA

Способы сжатия данных в PostgreSQL

Жилин Михаил

PostgresPro

PGMeetup.DBA

О себе:

- Физтех
- 14 лет работы с системной производительностью
- 2 года в компании Postgres Professional
- Контрибьютор в open source проекты
- Счастливый муж и отец
- Волейбол, дайвинг, путешествия

Telegram: @mizhka
E-mail: mizhka@gmail.com

PostgresPro

План разговора

- Полезно ли сжатие для баз данных?
- Сжатие бывает разным
- Встроенные механизмы и не только
- Примеры и рекомендации

PGMeetup.DBA

PostgresPro

Полезно ли сжатие для баз данных?

PostgresPro

Полезно ли сжатие для баз данных?

Достоинства

- Экономия дискового пространства
- Экономия оперативной памяти
- Меньше нагрузка на диски

PostgresPro

Полезно ли сжатие для баз данных?

Достоинства

- Увеличение пропускной способности системы (больше tps)
- Улучшается время ответа (меньше мс)

PostgresPro

Полезно ли сжатие для баз данных?

Недостатки

Overhead	Shared Object	Symbol
63.82%	postgres	[.] pglz_decompress
16.04%	libc-2.31.so	[.] __memcpy_avx_unaligned_erms
1.99%	postgres	[.] LWLockAttemptLock
0.89%	postgres	[.] hash_search_with_hash_value
0.89%	postgres	[.] LWLockRelease
0.88%	postgres	[.] _bt_compare

PGLZ

Overhead	Shared Object	Symbol
46.70%	liblz4.so.1.9.3	[.] LZ4_decompress_safe
5.81%	postgres	[.] LWLockAttemptLock
2.59%	postgres	[.] LWLockRelease
2.19%	postgres	[.] _bt_compare
2.14%	postgres	[.] hash_search_with_hash_value

LZ4

PostgresPro

Сжатие бывает разным

PostgresPro

Сжатие бывает разным

Особенности

- Без потерь
- Максимально компактное
- За максимально короткое время
- Текстовые и бинарные данные
- А что такое сжатие?

PostgresPro

Сжатие бывает разным

Зависимости

- 1, 2, 3, 4, 5, 6, 7, 8, 9…
- 1, 2, 3, 5, 7, 11, 13, 17, 19, 23…
- 53363, 40550, 33191…

PostgresPro

Сжатие бывает разным

Теория информации

- Алгоритмическая теория информации
  Андрей Николаевич Колмогоров (1960-ые)
  - https://nautil.us/kolmogorov-complexity-and-our-search-for-meaning-237158/
  - Минимальная длина программы, которая выводит необходимый набор символов (данных)
- Парадокс неизмеримости: **всегда может существовать алгоритм с меньшей длиной программы**

PostgresPro

Сжатие бывает разным

Энтропийное сжатие

- Шеннон-Фано (1948-1949)
  - Вероятность символов
  - Чем чаще символ, тем меньше бит на символ
- Хаффман (1954)
  - Бинарное дерево частот и префиксы
  - Оптимальный набор

PostgresPro

Сжатие бывает разным

Сжатие по словарю

- Lempel-Ziv-Welch (1978-1984)
  - рекурсивная копия, повторяемость последовательностей символов
  - LZ1 (LZ77) / LZ2 (LZ78) / LZW
  - gif
- Deflate (1991) с добавление энтропийного сжатия
  - png, zip, gzip and others

PostgresPro

Сжатие бывает разным

Соревнование на скорость

- LZO (1996) - Markus F.X.J. Oberhumer
  - быстрая распаковка: read-only filesystems
  - https://github.com/markus-oberhumer

- LZ4 (2011) + Zstandard (2016) - Yann Collet
  - tradeoff и оптимизации под 32-/64-бит архитектуры
  - https://github.com/Cyan4973

PostgresPro

Сжатие бывает разным

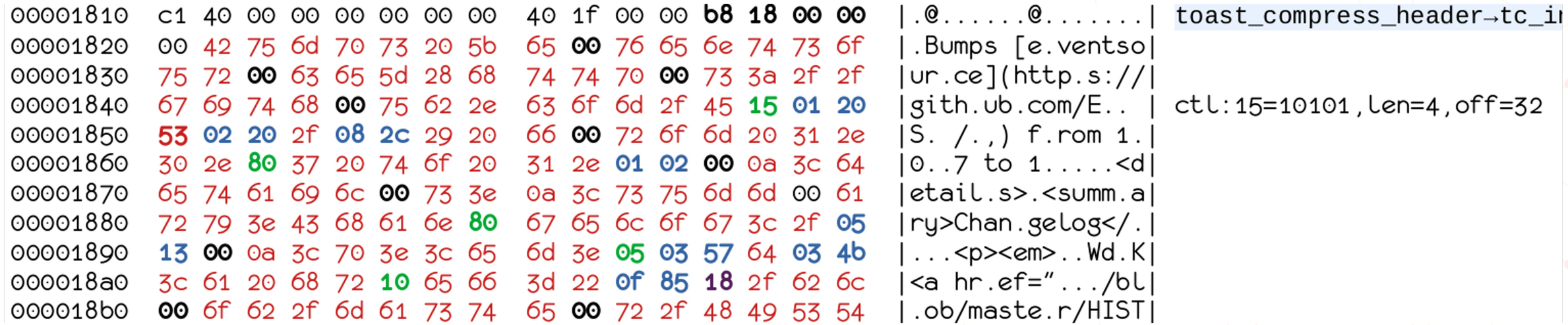
Современные реалии

- Смесь из энтропийного сжатия и сжатия по словарю
- Степень сжатия: от 1 до N (размер плавающего окна)
- Неэффективность на маленьких объемах
- LZ - словарь находится в сжатом поток

PostgresPro

Сжатие бывает разным

PGLZ internals



PostgresPro

Сжатие бывает разным

LZ4 vs PGLZ

- Нет побитовых операций
- Копирование по несколько байт вместо по одному

PostgresPro

Сжатие бывает разным

LZ4 vs PGLZ

Overhead	Shared Object	Symbol
63.82%	postgres	[.] pglz_decompress
16.04%	libc-2.31.so	[.] __memcpy_avx_unaligned_erms
1.99%	postgres	[.] LWLockAttemptLock
0.89%	postgres	[.] hash_search_with_hash_value
0.89%	postgres	[.] LWLockRelease
0.88%	postgres	[.] _bt_compare

PGLZ

Overhead	Shared Object	Symbol
46.70%	liblz4.so.1.9.3	[.] LZ4_decompress_safe
5.81%	postgres	[.] LWLockAttemptLock
2.59%	postgres	[.] LWLockRelease
2.19%	postgres	[.] _bt_compare
2.14%	postgres	[.] hash_search_with_hash_value

LZ4

PostgresPro

Сжатие бывает разным

zstd vs LZ4

- Добавлено энтропийное сжатие (Huffman + FSE)
- Многопоточный
- Отрицательные уровни сжатия (-7 - самый быстрый)
- Колоссальные размеры окон (до 128MiB)

PostgresPro

Сжатие бывает разным

Benchmarks

Алгоритм	Описание	Сжатие	Распаковка	Степень
deflate	добрый старый	5-100MiBps	10-200MiBps	**2.8**
lzo	быстрая распаковка	8MiBps	850MiBps	**2.8**
lz4	самый быстрый	**780MiBps**	**4500MiBps**	2.1
zstd	лучший баланс	480MiBps	1200MiBps	**2.8**

PGMeetup.DBA

PostgresPro

Встроенные механизмы сжатия данных

PostgresPro

Встроенное сжатие: inline и TOAST

Для строк/значений больше чем 2килобайта

- короткие в таблице
- длинные в TOAST

- PGLZ
- LZ4 (с 14ой версии)

PGMeetup.DBA

PostgresPro

Встроенное сжатие: Index Key Dedup

Группировка TID-ов в Posting List

- PostgreSQL 12+
- PostgresPro 10+

PostgresPro

Встроенное сжатие: WAL Full Page Image

Сжатие блоков в WAL файлах

- PGLZ
- LZ4 (с 15ой версии)

PostgresPro

Дополнительно…

Колоночное хранилище

- GreenPlum & ZedStore (fork)
- Citus Columnar & cstore_fdw (extension)
- Append-only optimizations
- lz4, zstd, zlib, rle

- Нет индексного сжатия
- Ряд ограничений

PostgresPro

Дополнительно…

Сжатие на уровне OS

- OpenZFS
- Прозрачно для базы данных
- lz4, zstd, snapshot-ы

- Пределы масштабирования
- Требует опыта в настройках

PostgresPro

Может быть завтра…

- Быстрого TOAST-а
- Дедупликации значений в таблицах
- Сжатие ключей в индексах

PostgresPro

Блочное сжатие: PostgresPro CFS

- Покрывает все типы данных (индексы, таблицы)
- Прозрачно для пользователей
- Простота конфигурации
  - задаётся на уровне tablespace-ов
- lz4, zstd, zlib, pglz, levels…
- Входит в PgPro Enterprise 9.6+

PGMeetup.DBA

PostgresPro

Типовые сценарии использования сжатия

PostgresPro

Типовые сценарии использования сжатия

Современные базы небольшое размера

- < 500GiB
- < 500tps
- Достаточное количество RAM для кэширования
- SSD/NVMe

Встроенного сжатия, скорее всего, достаточно

PostgresPro

Типовые сценарии использования сжатия

Современные базы большого размера

- > 1TiB

CFS или ZFS позволит сохранить терабайты

PostgresPro

Типовые сценарии использования сжатия

Медленные диски и/или мало RAM

- HDD / NAS
- Всего 2-4GiB RAM

CFS или ZFS могут значительно ускорить работу PostgreSQL

PostgresPro

Типовые сценарии использования сжатия

Файлохранилка (PDF, photos)

- Хранится в TOAST-е
- Медленно, но надёжно
- По копии на каждой реплике и в бекапах

Выносите хранение сжатых файлов из PostgreSQL на внешнее хранилище а-ля S3

PGMeetup.DBA

PostgresPro

Типовые сценарии использования сжатия

Аналитические базы данных, витрины

- Колонки без индексов
- Низкоэнтропийные данные

Попробуйте колоночное хранилище

PostgresPro

Типовые сценарии использования сжатия

Зашифрованные данные

- Энтропия
- Нет встроенного шифрования
- Не включайте сжатие TOAST для шифрованных данных

Сначала сжимать, потом шифровать

PostgresPro

Итого:

Use compression, Luke!

- Вопросы?

PGMeetup.DBA

СПАСИБО!

postgrespro.ru