

Кейс 5: Анализ производительности B-tree и B+-tree

Выполнили:

Шебордаев Андрей, Стрижов Тимофей, Олейник Владимир, Хани Юрий.

Задание 1: Сравнение производительности запросов

Условия тестирования

- Набор из 1000 чисел от 1 до 1000
- Сравниваются B-tree и B+-tree с одинаковым порядком (m=100)
- Деревья хорошо сбалансированы

Результаты производительности

Таблица сравнения:

| Метрика | Запрос 1 (10 эл-тов: 50-59) | Запрос 2 (100 эл-тов: 400-499) | Запрос 3 (500 эл-тов: 1-500) |
|----------------|-----------------------------|--------------------------------|------------------------------|
| B-tree | | | |
| Время | Малое | Среднее | Большое |
| Узлы | ~3-4 | ~13-19 | ~53-64 |
| B+-tree | | | |
| Время | Малое | Среднее | Среднее |
| Узлы | ~3-4 | ~5-7 | ~13-16 |

Общее время выполнения

- **B-tree:** Малое + Среднее + Большое = **Очень большое**
- **B+-tree:** Малое + Среднее + Среднее = **Умеренное**

Ответы на вопросы

1. Для каких запросов B+-tree выигрывает сильнее?

B+-tree выигрывает сильнее всего для **запросов на большие диапазоны (Запрос 3)**. С ростом размера диапазона преимущество становится подавляющим. Для маленьких диапазонов разница незначительна.

2. Почему B+-tree лучше для больших диапазонов?

1. **Последовательный доступ** - листовые узлы связаны указателями
2. **Высокая плотность ключей** - в листьях помещается больше записей
3. **Отсутствие обращений к внутренним узлам** - не требуется повторный спуск по дереву

3. В каких реальных системах это важно?

- **Базы данных** - PostgreSQL, MySQL, Oracle (запросы с BETWEEN, >, <)

- **Файловые системы** - NTFS, ReiserFS
- **Системы с временными рядами** - мониторинг, аналитика
- **Поисковые системы** - диапазонные запросы по датам

Задание 2: Выбор структуры для “Погодного сервиса”

Требования системы

- ☒ Частые запросы по датам (диапазоны)
- ☒ Редкие обновления данных
- ☒ Данные на диске
- ☒ Пример: “погода за последнюю неделю”

Топ-3 структуры данных

Таблица сравнения

| Критерий | B+-tree | LSM-дерево | Иммутабельное B-дерево |
|-----------------------------------|---|--|---------------------------------------|
| Скорость диапазонного запроса | ★★★★★ | ★★★★★ | ★★★★★ |
| Скорость обновления | ★★★ | ★★★★★ | ★★ |
| Эффективность использования диска | ★★★★★ | ★★★ | ★★ |
| Примечания | Золотой стандарт для диапазонных запросов | Оптимизирован под запись, хорош для чтения | Идеален для редких обновлений пачками |
| Итоговый приоритет | 1 | 2 | 3 |

Рекомендации

1. **B+-tree** - лучший выбор для текущих требований
2. **LSM-дерево** - хорошая альтернатива с перспективой на будущее
3. **Иммутабельное B-дерево** - специализированное решение для редких обновлений

Обоснование выбора

B+-tree обеспечивает оптимальный баланс между скоростью выполнения диапазонных запросов и простотой поддержки, что критично для погодного сервиса с его типичными сценариями использования.