

Оптимизация LSM-дерева на Go

СПБАУ, МИТ, 3 курс, октябрь 2016 -- февраль 2017

Автор: Елизавета Третьякова

Руководители: Кирилл Сергеевич Коротаев

Мотивация

Цель проекта:

- Улучшить производительность имеющейся реализации LSM-дерева на Go

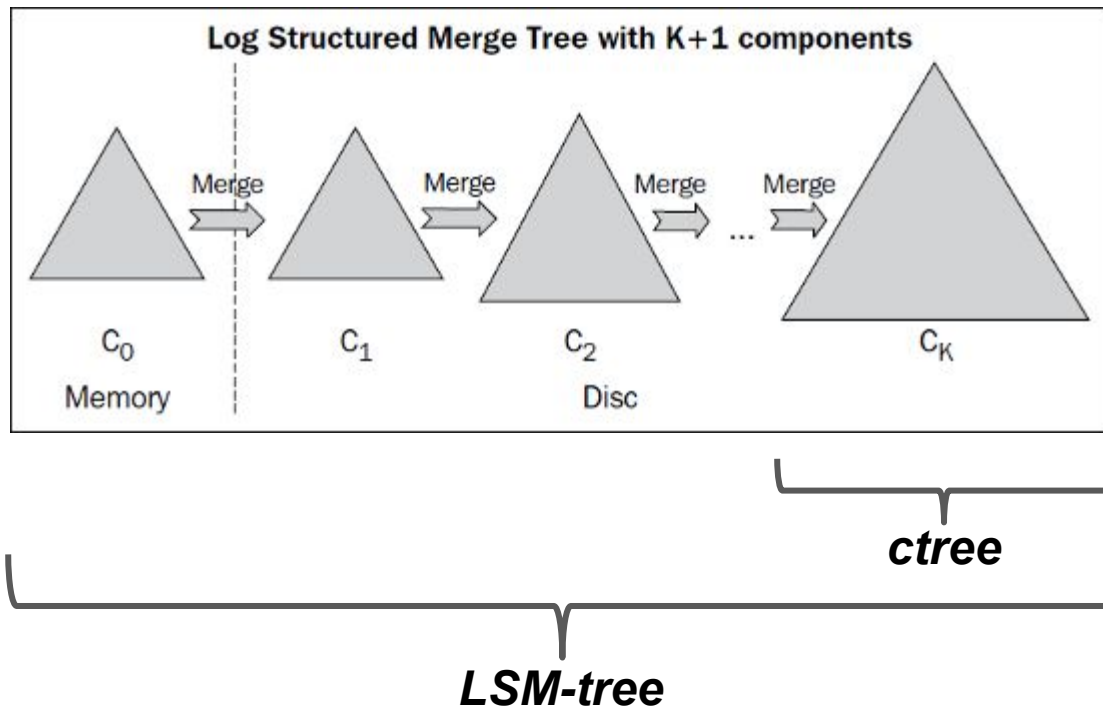
Актуальность:

- Эффективное хранение большого индекса (LSM в целом)
- Сравнение с текущей реализацией на C

Задачи

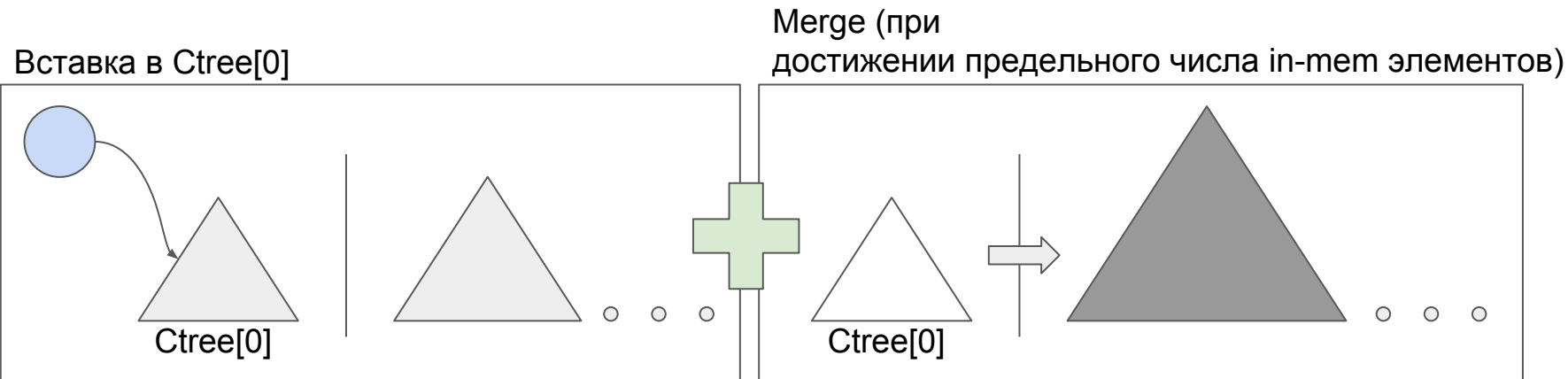
1. Изучить алгоритм работы LSM-дерева
2. Реализовать поддержку безопасной параллельной вставки элементов несколькими пользователями (синхронизация)
3. Реализовать поддержку перебалансировки дерева в background'е для минимизации задержки при вставке в дерево
4. Реализовать поддержку итератором операций $<$ и $<=$, а также возможность обхода в обратном направлении

LSM-дерево



Реализовано

1. Параллельная вставка
2. Перебалансировка дерева без задержки пользователя



Возникшие сложности:

1. Существующий в проекте код -- сложный
2. Создание и анализ бенчмарков

Дальнейшее развитие:

- Ограничение размера in-memory stree
- Профилирование и оптимизация времени работы
- Оптимизация компрессии данных на диске

Чему научилась

1. Много Go
2. Много вникания в существующий код
3. Много тестов, бенчмарков и отладки