

М. А. Кита

(ГрГУ им. Я. Купалы, Гродно)

ИСПОЛЬЗОВАНИЕ ЛІТ-КОМПИЛЯЦІІ С ЦЕЛЬЮ ОПТИМИЗАЦІІ ОПЕРАЦІЙ GROUP BY В АНАЛИТИЧЕСКОЙ СУБД CLICKHOUSE

ClickHouse – это столбцовая система управления базами данных (СУБД) для онлайн обработки аналитических запросов (OLAP), разработанная в компании Яндекс [1, 2].

Одной из важнейшей функциональности для аналитических запросов является агрегация данных по ключу с использованием синтаксиса SQL операции GROUP BY. Примерами таких запросов являются

Материалы XXIV Республиканской научной конференции студентов и аспирантов «Новые математические методы и компьютерные технологии в проектировании, производстве и научных исследованиях», Гомель, 22–24 марта 2021 г.

получение суммы, минимального максимальных полей колонки, различные специализированные функции, которые предоставляет СУБД ClickHouse. Данная задача не может быть эффективно решена для всех возможных типов данных и сценариев, текущая реализация в СУБД ClickHouse старается использовать различные подходы и специализации максимально подходящая к выполнению конкретного запроса.

Цель работы – ускорить выполнение операции GROUP BY за счет использования JIT-компиляции. JIT-компиляция – это способ выполнения компьютерного кода, который включает в себя компиляцию кода и его исполнение во время выполнения программы.

В случае с операцией GROUP BY это позволит добиться того, что код, сгенерированный с помощью JIT максимально оптимизирован под конкретный CPU, дает возможность использовать максимально эффективно регистры и инструкции конкретного процессора такие как AVX-256, AVX-512. Так же при использовании JIT компиляции сокращается размер исполняемого бинарного файла СУБД ClickHouse, так как чтобы код был максимально эффективным, текущая реализация использует все возможные специализации на уровне компиляции. В качестве JIT компилятора был выбран LLVM проект программной инфраструктуры для создания компиляторов и сопутствующих им утилит, в частности JIT компилятор ORC.

Литература

- 1 The LLVM Compiler Infrastructure [Электронный ресурс]. – 2018. – Режим доступа: <https://llvm.org>. – Дата доступа: 02.01.2021.
- 2 ORC Design and Implementation [Электронный ресурс]. – 2003. – Режим доступа: <https://llvm.org/docs/ORCv2.html>. – Дата доступа: 12.01.2021.