

ОПТИМИЗАЦИЯ ЗАПРОСОВ С ПОМОЩЬЮ JIT-КОМПИЛЯЦИИ В АНАЛИТИЧЕСКОЙ СУБД CLICKHOUSE

Кита М. А., Рудикова Л. В.

*Гродненский государственный университет имени Янки Купалы, Гродно, Беларусь,
e-mail: rudikowa@gmail.com, kitaetoya@gmail.com*

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

Одним из важнейшей функциональностью для аналитических запросов является агрегация данных по ключу с использованием синтаксиса SQL операции GROUP BY. Примерами таких запросов являются 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 эффективный 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 [1-4].

Литература

1. Paul, J. Improving Execution Efficiency of Just-in-time Compilation based Query Processing on GPUs // Johns Paul, Bingsheng He Shengliang Lu, Chiew Tong Lau / [Electronic resource]. – 2015. – Mode of access: <https://llvm.org>. – Date of access: 14.03.2021.
2. The LLVM Compiler Infrastructure / [Electronic resource]. – 2020. – Mode of access: <https://llvm.org>. – Date of access: 14.03.2021.
3. ORC Design and Implementation / [Electronic resource]. – 2020. – Mode of access: <https://llvm.org/docs/ORCv2.html>. – Date of access: 14.03.2021.
4. ClickHouse documentation / [Electronic resource]. – 2020. – Mode of access: <https://clickhouse.tech>. – Date of access: 14.03.2021.