

АНАЛИТИЧЕСКАЯ СУБД CLICKHOUSE

В работе приводится описание причин использования аналитической СУБД ClickHouse при работе с большим объемом данных, сравнивается производительность строковой СУБД MySQL и столбцовой ClickHouse.

ВВЕДЕНИЕ

На сегодняшний день многие компании оперируют большими объемами данных. Часто необходим анализ данных в онлайн-режиме, но не все СУБД могут справиться с данной задачей. Решением является хранение «сырых» неагрегированных данных, производя все необходимые вычисления в момент запроса пользователя. Для этого была создана СУБД ClickHouse, которая обрабатывает неагрегированные данные (петабайты данных) с очень высокой эффективностью и в реальном времени, и при этом обладает приемлемой стоимостью.

I. ОСНОВНЫЕ ПОНЯТИЯ

ClickHouse — это столбцовая аналитическая СУБД с открытым кодом, позволяющая выполнять аналитические запросы в режиме реального времени на структурированных больших данных, разрабатываемая компанией Яндекс. Работа СУБД основана на сценарии OLAP.

OLAP (online analytical processing) - технология обработки данных, заключающаяся в подготовке суммарной (агрегированной) информации на основе больших массивов данных, структурированных по многомерному принципу.

II. ПРИЧИНЫ ИСПОЛЬЗОВАНИЯ CLICKHOUSE ДЛЯ АНАЛИЗА БОЛЬШОГО ОБЪЕМА ДАННЫХ

Причинами использования ClickHouse являются следующие преимущества: сжатие данных, хранение данных на диске, распределенная обработка запросов на многих серверах, поддержка SQL, векторный движок, наличие индекса, подходит для онлайн запросов, поддержка приближенных вычислений, репликация данных и поддержка целостности.

Но СУБД обладает также и недостатками: отсутствие транзакций; возможность изменять или удалять ранее записанные данные с низкими задержками и высокой частотой запросов не

предоставляется, есть массовое удаление и изменение данных.

III. СРАВНЕНИЕ ПРОИЗВОДИТЕЛЬНОСТИ MYSQL И CLICKHOUSE

По результатам внутреннего тестирования в Яндексе, ClickHouse обладает наиболее высокой производительностью. Чтобы подтвердить эти результаты, было проведено собственное тестирование.

В СУБД MySQL и ClickHouse были созданы таблицы с одинаковыми данными. К каждой таблице составлено по 3 одинаковых запроса. При поочередном запуске запросов были получены следующие результаты скорости выполнения каждого из них (табл.1.):

Таблица 1 – Результаты эксперимента

Запрос	MySQL, мс.	ClickHouse, мс
1	1050	19
2	4900	23
3	10600	74

IV. ВЫВОДЫ

ClickHouse позволяет выполнять аналитические запросы в интерактивном режиме по данным, обновляемым в реальном времени. Система способна масштабироваться до десятков триллионов записей и петабайт хранимых данных. Использование ClickHouse открывает возможности, которые раньше было даже трудно представить: можно сохранять весь поток данных без предварительной агрегации и быстро получать отчеты в любых разрезах.

Список литературы

1. Яндекс ClickHouse. Быстрее некуда. / Филиппов Олег // Системный администратор. — 2017. — № 1–2. — С. 56–58.
2. ClickHouse в системах сбора статистики / Календарев Александр // Системный администратор. — 2017. — № 3. — С. 56–59.

Зборовская Анна Андреевна, студентка 3 курса факультета информационных технологий и управления БГУИР, a.zborovskii@gmail.com

Коланда Вероника Евгеньевна, студентка 3 курса факультета информационных технологий и управления БГУИР

Садовская Вероника Олеговна, студентка 3 курса факультета информационных технологий и управления БГУИР, sadovskaya.veronika02@mail.ru

Научный руководитель: Трофимович Алексей Федорович, старший преподаватель кафедры информационных технологий автоматизированных систем БГУИР, trofimaf@bsuir.by