

Отчет к ДЗ по теме Область применения и первое представление

1. Компания: *Okko*

- **Специфика проекта:** Okko до внедрения Clickhouse в качестве хранилища использовала PostgreSQL (плюс частично HDFS). Но вместе с развитием компании и технологий увеличивался и объем анализируемых данных и в итоге все дошло до того, что среднее время аналитического запроса составляло около минуты. Поэтому было принято решение искать альтернативные варианты хранения данных. Выбор пал на Clickhouse
- **Особенности внедрения:** При построении хранилища данных компания использует разные движки семейства MergeTree, так же активно применяет «словари», «преагрегацию данных» и проекции.
- **Итог:** После переезда, среднее время запроса в аналитическом кластере составляет около 2 с.
- **Ссылка на ресурс:** [Как небольшой команде переехать на ClickHouse: на какие грабли мы наступили и о каких фишках не знали](#)

2. Компания: *Маревен Фуд Сэнтрал*

- **Специфика проекта:** Обработка большого объема информации (миллиарды строк с ежегодным приростом 150-200 млн) большим количеством сотрудников (от 30 до 50 человек). Изначально все это было реализовано на обычной реляционной базе данных, построенной на базе Microsoft SQL Server. По мере роста данных компания стала сталкиваться со сложностями масштабирования и другими проблемами. Основные из них:
- Производительность системы отчетности
- Сложность масштабирования системы хранения данных
- Стоимость инфраструктуры
- **Особенности внедрения:** ClickHouse состоит из кластера трех серверов, все сервера работают на системе Ubuntu Server 20.04. Помимо самого ClickHouse, есть система управления репликациями и запросами, которая построена на Apache ZooKeeper и сервер управления фронтендом Apache Superset.
- **Итог:**
 - Сокращение времени выдачи отчета до 100 раз. Длительность снизилась с 1,5 часов, до 1,5 минут.
 - Сжатие данных сократило объем дискового пространства на 30%.
 - Увеличилась гибкость работы с данными для конечного пользователя
 - Повысилась отказоустойчивость системы
- **Ссылка на ресурс:** [ClickHouse внедрение в корпоративной среде компании Маревен Фуд Сэнтрал](#)
-

3. Компания: *ВКонтакте*

- **Специфика проекта:** Требуется:
 - Простота эксплуатации
 - Высокая скорость записи и чтения
 - Долговременное хранение (месяцы, годы)
 - Сжатие данных
 - Запись длинных строк
- **Особенности внедрения:** Из-за большого количества серверов (отсюда большое кол-во HTTP соединений и вставок) был написан промежуточный сервис, локальный проху «Kittenhouse», который принимает информацию от РНР, агрегирует информацию и делает вставку в Clickhouse
- **Итог:**
- Clickhouse подходит для логов
- Буферные таблицы имеют много кусков
- Много TCP соединений
- **Ссылка на ресурс:** [Как VK вставляет данные в ClickHouse с десятков тысяч серверов](#)

3. Ответ на дополнительные вопросы:

К каким классам систем относится ClickHouse?

ClickHouse относится к классу систем управления базами данных (СУБД) для аналитической обработки данных (OLAP).

Какую проблему вы бы решили, используя ClickHouse, а какую бы не стали?

Проблема, которую можно решить с помощью ClickHouse: Это обработка и анализ больших объемов данных в реальном времени, например, для аналитики веб-трафика, мониторинга систем или анализа пользовательского поведения. ClickHouse позволяет быстро выполнять сложные запросы и получать результаты в кратчайшие сроки.

Проблема, которую не стоит решать с помощью ClickHouse: ClickHouse не подходит для транзакционных систем (OLTP), где важна высокая скорость обработки отдельных транзакций и поддержка сложных операций с данными, таких как частые обновления и удаления.

Где можно получить помощь по ClickHouse и куда сообщать о багах?

[официальный чат ClickHouse](#), самое популярное и активное место для получения поддержки
[код ClickHouse](#), а так же [issue-трекер](#), с актуальными проблемами
[официальная документация](#)