

Домашнее задание №1

1. Отдельным файлом представлена схема в формате C4 двух вариантов использования ClickHouse в компании на последнем рабочем месте;
2. Нарисованная схема п. 1 отображает часть работы Платформы данных компании;
3. Схема представлена рядом в текущем каталоге;
4. ClickHouse относится к классу MPP колоночных систем управления данными, выполняющими OLAP - нагрузку;
5. Использование ClickHouse
 - 5.1. Проблемы, которые можно решать на ClickHouse я бы охарактеризовал задачами, реализующими свойственный для него профиль нагрузки:
 - 5.1.1. Работа ClickHouse внутри платформы данных. Загрузка данных в режиме online (иногда говорят об NRT – Near Real Time) из топиков Kafka, принадлежащих различным информационным системам в ClickHouse и предоставление этих данных в корпоративную систему отчетности;
 - 5.1.2. Работа ClickHouse внутри платформы данных. Загрузка данных в режиме пакетной загрузки (batch loaded) из подготовленных широких витрин Greenplum, который в свою очередь, загружая данные из разнородных информационных систем, на их основе генерирует большое число новых расчетных метрик, которые собираются в очень большие и широкие витрины. При формировании многочисленных расчетных метрик используется большое число различных сложных аналитических запросов и многочисленное число слоев. Помимо подготовки широких витрин на стороне Greenplum для передачи в ClickHouse, Greenplum берет на себя ad hoc запросы от пользователей. Таким образом нехарактерные для ClickHouse профили нагрузки берет на себя Greenplum, а ClickHouse забирает уже подготовленные широкие витрины;
 - 5.1.3. Работа ClickHouse внутри платформы данных. Обработка в ClickHouse данных, загруженных по п. 5.1.1 и 5.1.3 и сборка витрин для строго регламентных отчетов корпоративной системы отчетности компании;
 - 5.1.4. Работа ClickHouse внутри продуктовых информационных систем, которые находятся за периметром платформы данных. Получение потоковых контрактных данных из топиков Kafka, их быстрая агрегация в ClickHouse и использование в микросервисах текущей вертикали.
 - 5.2. Проблемы, которые не стоит решать на ClickHouse я бы охарактеризовал задачами, реализующими не свойственный для него профиль нагрузки:
 - 5.2.1. Выстраивать многочисленные слои (более 2-3) в ClickHouse как в Greenplum (более 10) и выстраивание длинного конвейера по обработке данных;
 - 5.2.2. Работа с ClickHouse с ad hoc – запросами пользователей, создающие случайные запросы на ClickHouse, содержащие многочисленные соединения различных витрин ClickHouse. Такая нагрузка может привести к общей деградации работы кластера, мешающей выдавать данные из заранее подготовленных витрин по также заранее оптимизированным запросам из корпоративной системы отчетности.
6. Информацию по ClickHouse можно получать:
 - 6.1. <https://clickhouse.com/docs/ru>
 - 6.2. <https://habr.com/ru/articles/>
 - 6.3. Форумы и Видеодоклады