

Чек-лист по оптимизации ClickHouse® для DataLens

Юрий Красильников, BI-евангелист DataLens и ex-лид BI в Яндекс Маркете, делится лайфхаками, как ускорить дашборды DataLens с использованием СУБД ClickHouse



На уровне сервера ClickHouse

- ✓ По возможности увеличьте количество шардов
- ✓ Добавьте памяти (RAM)
- ✓ Обновляйте ClickHouse до LTS-версии
- ✓ Используйте SSD, а не HDD

На уровне настроек ClickHouse

- ✓ Включите query cache `use_query_cache = true`
- ✓ Включите новый анализатор `enable_analyzer = 1`
- ✓ Ограничьте длительность запросов, чтобы убрать самые плохие из них и не ухудшать время оставшихся `max_execution_time = 30`

На уровне витрин ClickHouse

- ✓ Сортируйте правильно
- ✓ Не используйте nullable-поля
- ✓ Собирайте широкие витрины, но смотрите на пользу полей в логах
- ✓ Удаляйте неактуальные данные
- ✓ Устанавливайте TTL на данные и колонки
- ✓ Используйте SKIP-индексы, но проверяйте их эффективность
- ✓ Если нужны справочники для JOIN — используйте словари ClickHouse
- ✓ Проверяйте скрипты из инспектора `EXPLAIN indexes= 1`

На уровне DataLens

- ✓ Увеличьте тайм-аут при подключении
- ✓ В датасетах минимизируйте JOIN'ы
- ✓ В дашбордах разрывайте связи между селекторами
- ✓ Детально смотрите логи Usage Analytics
- ✓ Используйте Editor для сложных расчётов
- ✓ Не делайте селекторы-списки на ID заказов, делайте ручной ввод
- ✓ В чартах используйте оптимальные для ClickHouse функции
- ✓ Фильтруйте данные по полям сортировки
- ✓ Минимизируйте CAST полей на уровне BI