

Технические рекомендации

Процессор

Fastboard использует для обработки данных СУБД российской разработки Clickhouse. ClickHouse реализует параллельную обработку данных и использует все доступные аппаратные ресурсы. При выборе процессора учитывайте, что ClickHouse работает более эффективно в конфигурациях с большим количеством ядер, но с более низкой тактовой частотой, чем в конфигурациях с меньшим количеством ядер и более высокой тактовой частотой. Например, 16 ядер с 2600 MHz предпочтительнее, чем 8 ядер с 3600 MHz.

Рекомендуется использовать технологии Turbo Boost и hyper-threading. Их использование существенно улучшает производительность при типичной нагрузке.

RAM

Мы рекомендуем использовать как минимум 16 ГБ оперативной памяти, чтобы иметь возможность выполнять нетривиальные запросы. Сервер ClickHouse может работать с гораздо меньшим объёмом RAM, память требуется для обработки запросов.

Необходимый объём RAM зависит от:

- Сложности запросов.
- Объёма данных, обрабатываемых в запросах.

Для расчета объёма RAM необходимо оценить размер промежуточных данных для операций GROUP BY, DISTINCT, JOIN а также других операций, которыми вы пользуетесь.

ClickHouse может использовать внешнюю память для промежуточных данных.

Дисковое пространство

Для установки ClickHouse необходимо 2ГБ свободного места на диске.

Объём дискового пространства, необходимый для хранения ваших данных, необходимо рассчитывать отдельно. Расчёт должен включать:

- Приблизительную оценку объёма данных. Можно взять образец данных и получить из него средний размер строки. Затем умножьте полученное значение на количество строк, которое вы планируете хранить.
- Оценку коэффициента сжатия данных. Чтобы оценить коэффициент сжатия данных, загрузите некоторую выборку данных в ClickHouse и сравните действительный размер данных с размером сохранённой таблицы. Например, данные типа clickstream обычно сжимаются в 6-10 раз.

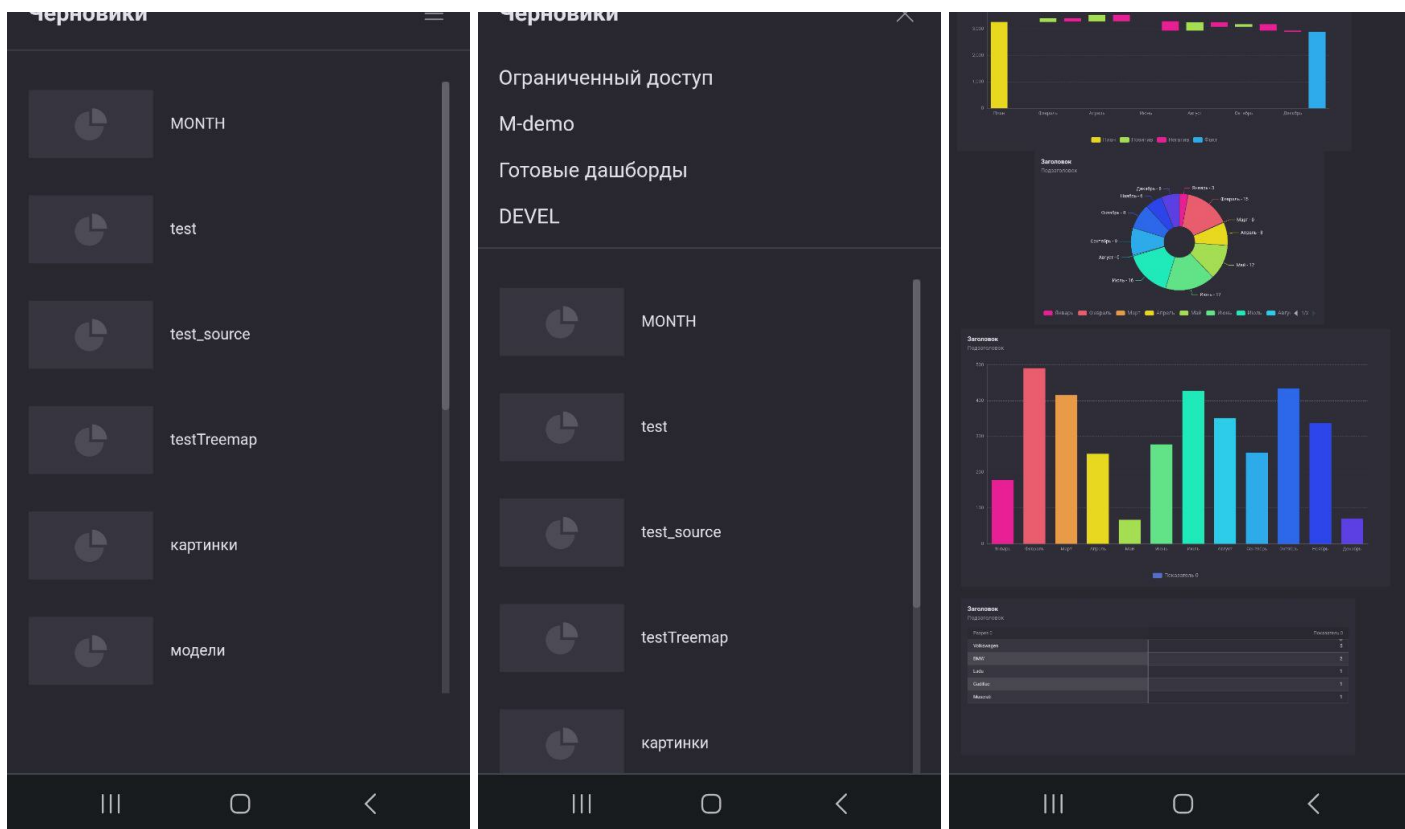
Для оценки объёма хранилища, примените коэффициент сжатия к размеру данных. Если вы планируете хранить данные в нескольких репликах, то необходимо полученный объём умножить на количество реплик.

Файл подкачки

Отключайте файл подкачки в продуктовых средах.

Поддержка мобильных устройств

Интерфейс Fastboard спроектирован с учетом поддержки работы на мобильных устройствах. Используемые элементы управления позволяют выбирать и просматривать существующие группы, проекты, страницы дашбордов.



Сеть

По возможности, используйте сети 10G и более высокого класса.

Пропускная способность сети критически важна для обработки распределенных запросов с большим количеством промежуточных данных. Также, скорость сети влияет на задержки в процессах репликации.

Продукт может быть развернут внутри закрытого контура без доступа к интернету.

Распределение нагрузки

Системное устройство Fastboard состоит из следующих компонентов:

- **5 сервисов:**

1. HTTP-сервис
2. Планировщик заданий (CRON)
3. Загрузчик данных из внешних источников
4. Блок трансформации файлов
5. Сервис трансформации данных для визуализаций

- **Сервисное хранилище PostgreSQL для настроек проектов и системных параметров**
- **Хранилище и расчеты данных для визуализаций ClickHouse**
- **Для обмена сообщений между сервисами и кеширование данных используется REDIS**
- **Менеджер процессов PM2 обеспечивает балансировку нагрузки между сервисами в рамках одного сервера**

Каждый из этих компонентов может быть запущен на разных серверах для оптимизации нагрузки

На одном сервере может быть запущено несколько экземпляров одного или нескольких компонентов

В мультисерверных решениях для работы системы можно использовать любые системы кластеризации и балансировщики нагрузки

Данная архитектура обеспечивает следующие преимущества:

Каждый узел (компонент/сервис) выполняет свою задачу. Например, расчет данных, в том числе многопоточный, выполняет ClickHouse, а преобразование данных в формат для визуализаций выполняет сервис трансформации. Таким образом достигается оптимальное распределение нагрузки между всеми компонентами системы.

Менеджер процессов PM2 обеспечивает балансировку нагрузки выделенных для системы ресурсов. Также система готова к работе с любой существующей системой управления кластерами. Установка максимально допустимых параметров потребления происходит на уровне операционной системы или выбранной пользователем системы кластеризации.

Revision #6
Created 15 August 2023 07:19:26 by Станислав
Updated 8 April 2024 14:13:45 by Станислав