

Elliptics

распределенная система хранения данных

Текущее положение дел

Для хранения средних и больших файлов используются собственные системы на выделенных серверах или облачных хранилищах, для раздачи - собственные кэшеры или популярные CDN (Nginx, Megafon, Akamai и др)

Плюсы: работает “из коробки”, не требует работы со стороны заказчика

Минусы: дорого, ненадежно, непроизводительно

Elliptics: хранение

- * создан для хранения средних и больших файлов
- * масштабируется от тысяч до десятков миллиардов ключей
- * р2р стриминг непосредственно с серверов хранения
- * географически распределенная репликация
- * работоспособность при отключении серверов и датацентров
- * отсутствие единых точек отказа
- * проверен **действительно** высокими нагрузками

Elliptics: хранение (пример)

“Бесконечное” хранилище.

В сутки записывается 2-2.5 Тб данных (файлы размером несколько мегабайт), данные не перезаписываются и редко удаляются.

Необходимо хранить минимум 2 копии данных на разных физических машинах, опционально хранить копии в разных датацентрах.

В среднем раз в месяц в хранилище добавляются 2 новых сервера с 24 дисками объемом 3 Тб. В месяц заменяется 1-2 диска.

Схема работает в production. Минус: нагрузка идет на серверы с новыми данными, старые серверы простаивают - нужна регулярная “перевеселенка”.

Elliptics: раздача тяжелых данных

Зачем использовать дорогой и ненадежный CDN, когда можно раздавать данные непосредственно из хранилища?

Хранилище должно уметь хранить копии в нужных точках раздачи, должно уметь раздавать тяжелые файлы. При сравнимой с CDN стоимостью объем ограничен только установленным оборудованием.

Elliptics: хранение и раздача (пример)

Аудио-видео стриминг.

~1 Pb уникальных данных, новые файлы записываются редко (единицы в день).

Необходимо хранить и раздавать данные. Канал в одном датацентре - 4 гигабит, нужно раздавать 7-8 гигабит. Клиенты подключаются с разной скоростью: от grgs до гигабита.

~30 серверов с 24 дисками по 3 Tb (~2 Pb) в 3 датацентрах. Стриминг (и обработка) данных идет непосредственно с дисков с серверов хранения через Nginx без проксирования - решена проблема узкого места с сетью и памятью на промежуточных прокси-серверах.

Схема работает в production.

Elliptics: встроенный кэш

Чем больше разных независимых технологий в проекте, тем выше вероятность возникновения проблем на стыках.

Мы создали встроенный в систему хранения данных распределенный SLRU cache, чтобы в одном пакете были сразу доступны все необходимые компоненты.

(тестовая нагрузка на кластер из 10 серверов, 1 Тб памяти - 1 миллион gps)

Elliptics: облачная инфраструктура

- * низкоуровневый доступ через C++/Python/Go API
- * высокоуровневый протокол на базе HTTP (ACL, группы реплик)
- * подробная статистика
- * репозиторий с пакетами
- * контейнеризация узлов хранения и прокси-серверов (Docker)

Elliptics

Используется для хранения (52+ миллиарда объектов, 100+ млрд запросов, 1-2 Тб новых данных в сутки):

- * Яндекс Карты, Фотки, Музыка, Маркет, вертикали, поиск по людям и др.
- * Coub.com
- * Sputnik.ru

Используется для раздачи данных (3+ Pб кластер, 8 Gbit/s раздача из нескольких датацентров):

- * Яндекс Диск
- * Яндекс Музыка

Elliptics

Open source community:

- * Открытая документация и tutorиалы
- * Бесплатная поддержка в списке рассылки

Reverbrain:

- * Администрирование сервисов (хранение, раздача, разработка)
- * Платная поддержка

Reverbrain.com