

hl⁺⁺

HighLoad⁺⁺

Анатомия баннерной системы

Артем Вольфтруб

www.gramant.ru



imho vi

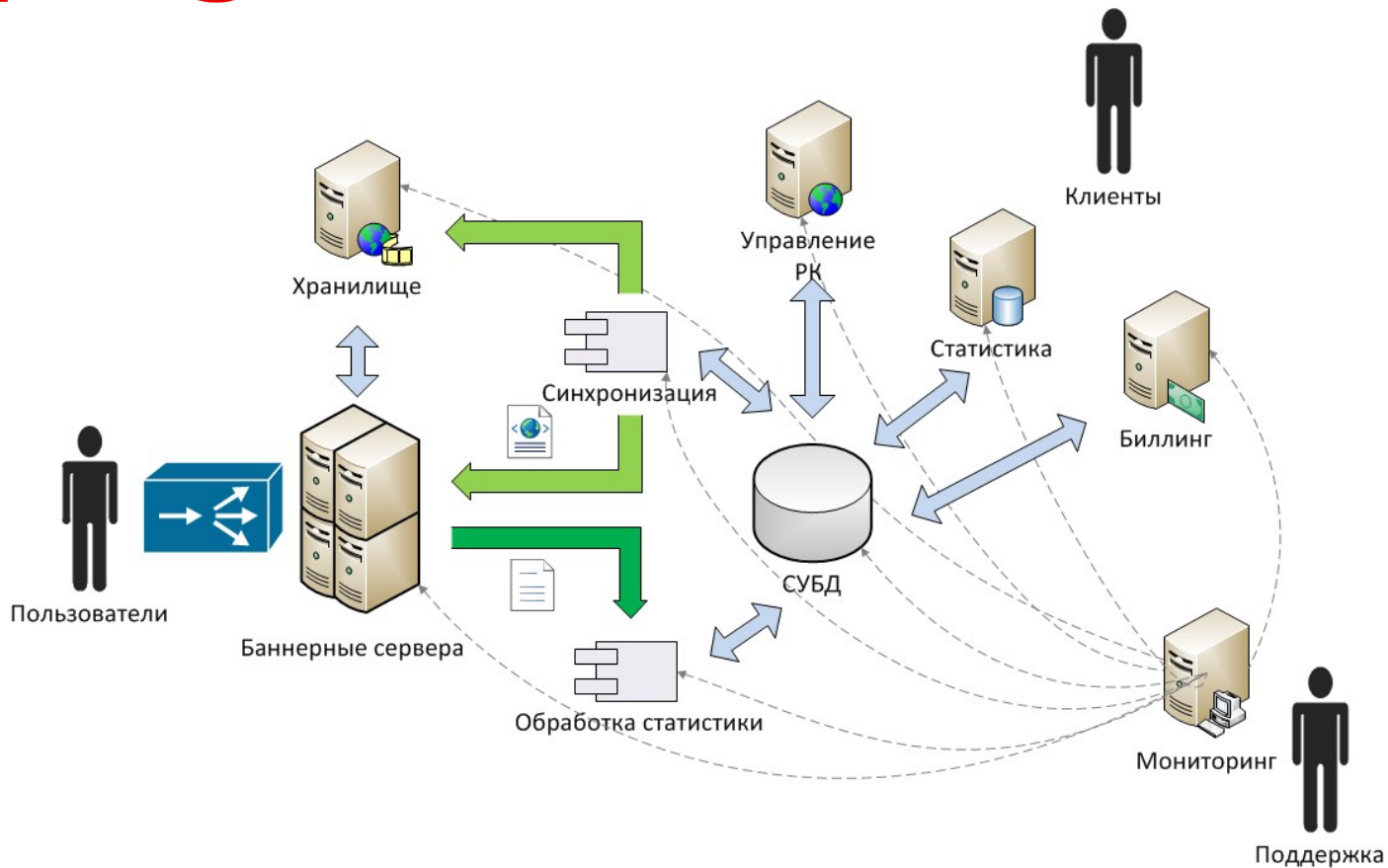
- 50 млн. показов в день
- Система на базе Dart Enterprise
- Расширенная статистика

ValueCommerce /

- 700 млн. показов в день
- Собственное решение
- Модель ориентированная на продажи

С чего начинается баннерка





На чем зарабатываем?

- На трафике
- На интересе
- На продажах



Ориентируемся на трафик

- Нагрузка на баннерные сервера
- Простая логика показов
- Простая статистика



Ориентируемся на интерес

- Сложный таргетинг
- Борьба с накрутками
- Сложная статистика
- Уникальные пользователи



Ориентируемся на продажи

- Мониторинг продаж
- Интеграция с рекламодателями
- Сложный биллинг



Баннерный сервер

- Mission critical компонент в любой баннерке
- Наиболее нагруженный компонент

-
- Простота
 - Слабая связанность
 - Независимость
 - Масштабируемость
 - Принцип «не молчать»



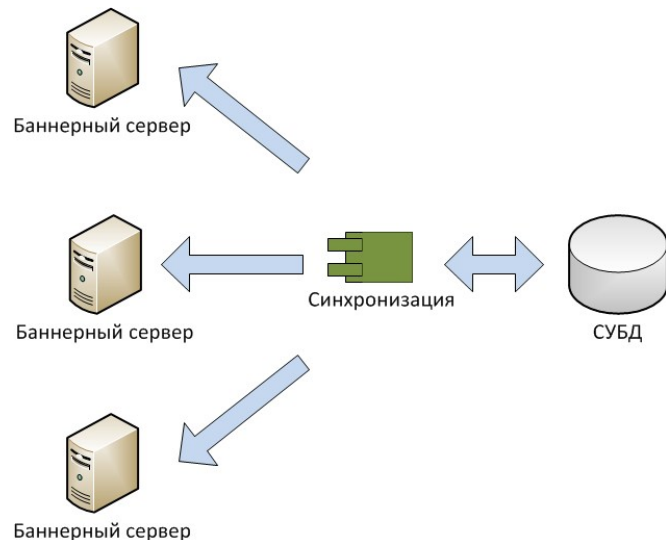
Стратегии обновления. Pull

- Убирает одно звено в цепи обновления
- Упрощает добавление нового сервера
- Позволяет иметь разные по содержимому сервера
- Усложняет код баннерного сервера
- Не позволяет обновить все сразу
- Не информирует о падении сервера
- Увеличивает неконсистентность данных



Стратегии обновления. Push

- Единая точка мониторинга
- Возможность обновить все сразу
- Обновления по расписанию
- Нужно регистрировать новый сервер
- Создается точка отказа
- Сложно поддерживать разный контент на баннерных серверах



Синхронизация

Инкрементальная

- Экономия трафика
- Уменьшение задержки обновления

Полная

- Одновременное переключение версии данных
- Простота реализации
- Необходима в любом случае



Хранение медиа данных

- Файловая система баннерных серверов
- Отдельное хранилище
- Распределенное хранилище (CDN, облако)



ЛИМИТЫ

- Одна из самых важных для бизнеса фич
- Бережливые менеджер и расточительные разработчики

Типы лимитов

- По рекламной кампании
- По пользователю



Решение. Обновление статистики

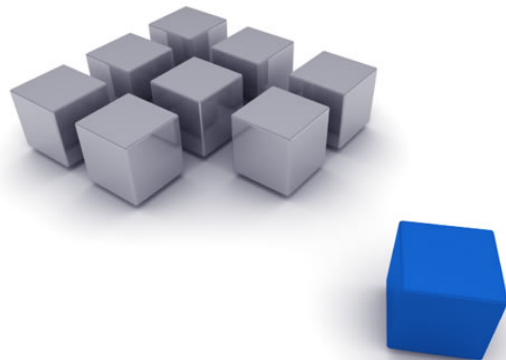
- Упрощает работу баннерного сервера
- Уменьшает время ответа
- Нельзя сделать без задержки, будут перекрутки.
- Требуется очень быстрая обработка статистики
- Сложно считать показы уникальным пользователям

Решение. Общий кеш

- Актуальные данные
- Отсутствие перекруток
- Дополнительная точка отказа
- Нужно синхронизировать со статистикой
- Данные могут вымываться

ЛИМИТЫ

- Не используйте без нужды
- По возможности используйте статистику
- Готовьтесь к падению кеша
- Разные стратегии для разных типов лимита



Таргетинг

- Занимает большую часть времени обработки запроса
- Откуда взять значение параметров?



Регистрация событий

- Память -> Файл -> Ротация
- Никакой предварительной обработки.



Стратегии показов

- Равномерное распределение по времени суток
- Динамический выбор лучшего баннера
- Принудительная открутка



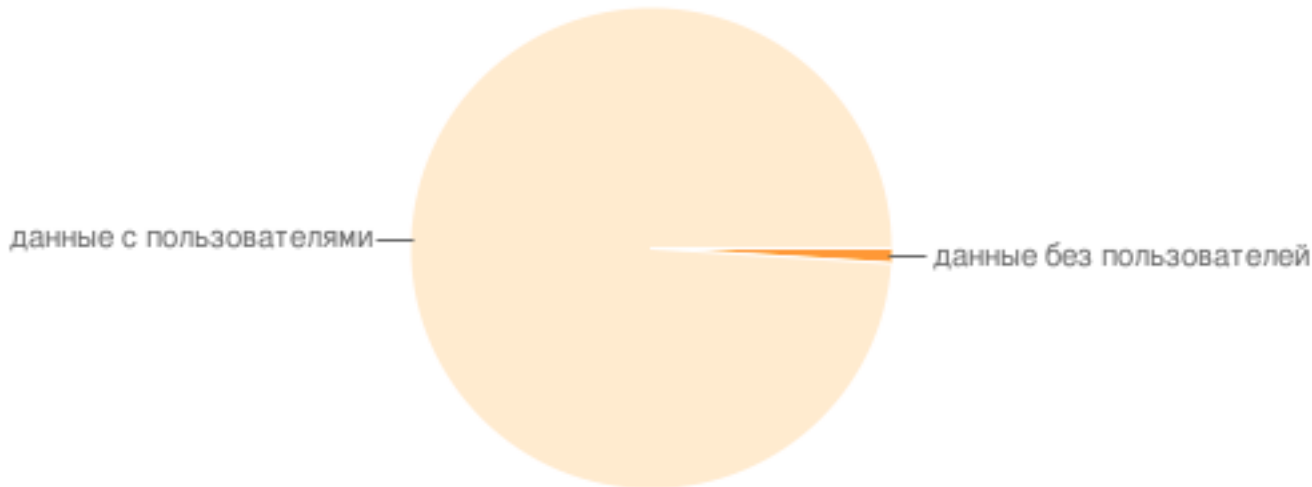
Статистика

Два основных типа данных

- Общая статистика
- Статистика по уникальным пользователям



Соотношение объемов данных



Ежедневное увеличение данных

Для 50 млн. показов в день

- Группировка по рекламным местам – 30-40 тыс. записей
- Группировка по пользователям 6-8 млн. записей

Что получаем взамен?

- Частотная характеристика
- Анализ пересечения аудитории
- Анализ параметров таргетинга
- Post-click анализ



Обработка статистики

- Отдельный компонент системы
- Узел суммирования и фильтрации данных
- Быстрая загрузка
- Разделение баз данных
- Денормализация данных

Данные по уникальным

- Расчет статистики в фоновом режиме
- Очередь отчетов
- Сохранение результатов на диск

hl⁺⁺

HighLoad⁺⁺



Мониторинг



Что дает мониторинг?

- Прогнозирование нагрузки
- Диагностика проблем на ранней стадии
- Выявление типовых проблем -> разработка универсальных решений

Виды мониторинга

- Физический уровень
- Уровень приложения
- Бизнес уровень

Мониторинг на физическом уровне

- Сеть
- Доступность сервера
- CPU
- Память
- IO
- Свободное место на дисках

Мониторинг на уровне приложения

- HTTP Errors
- Response Time
- Актуальность баннеров
- Актуальность статистики
- Размер очередей

Мониторинг на бизнес уровне

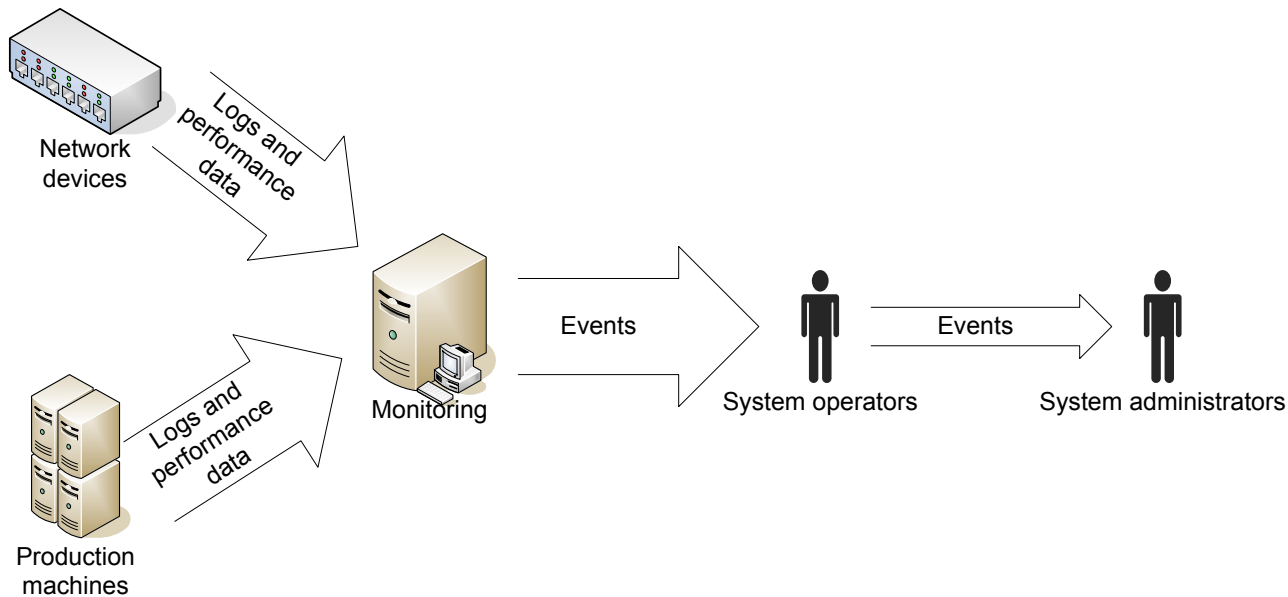
- Число зарегистрированных событий за период
- Динамика рекламных кампаний
- Количество отфильтрованных данных

Чем измеряем

- Критериальная система Nagios
- Тренды Cacti
- Журналы Tenshi



Система мониторинга



Поддержка системы

- Налаженная система мониторинга
- Наличие готовых лекарств
- Простота архитектуры
- Отлаженная процедура обновлений



hl⁺⁺

HighLoad++

Вопросы?

artem@gramant.ru

www.gramant.ru

