

hl⁺⁺

HighLoad⁺⁺

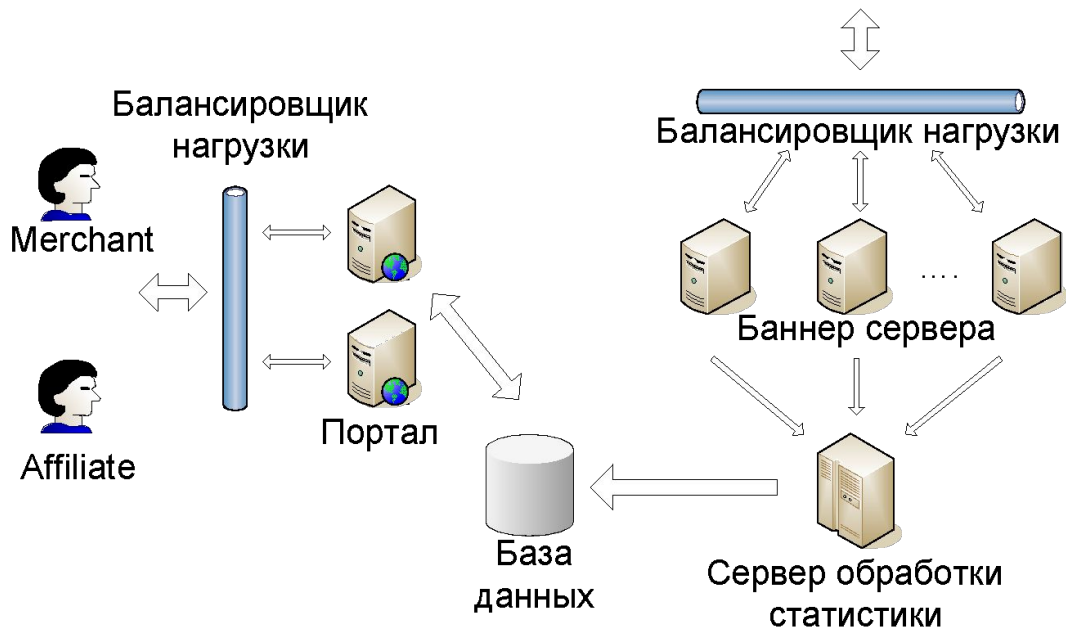
Обработка статистики в системе управления баннерной рекламой

Артём Вольфтруб

www.gramant.ru



Общая архитектура системы



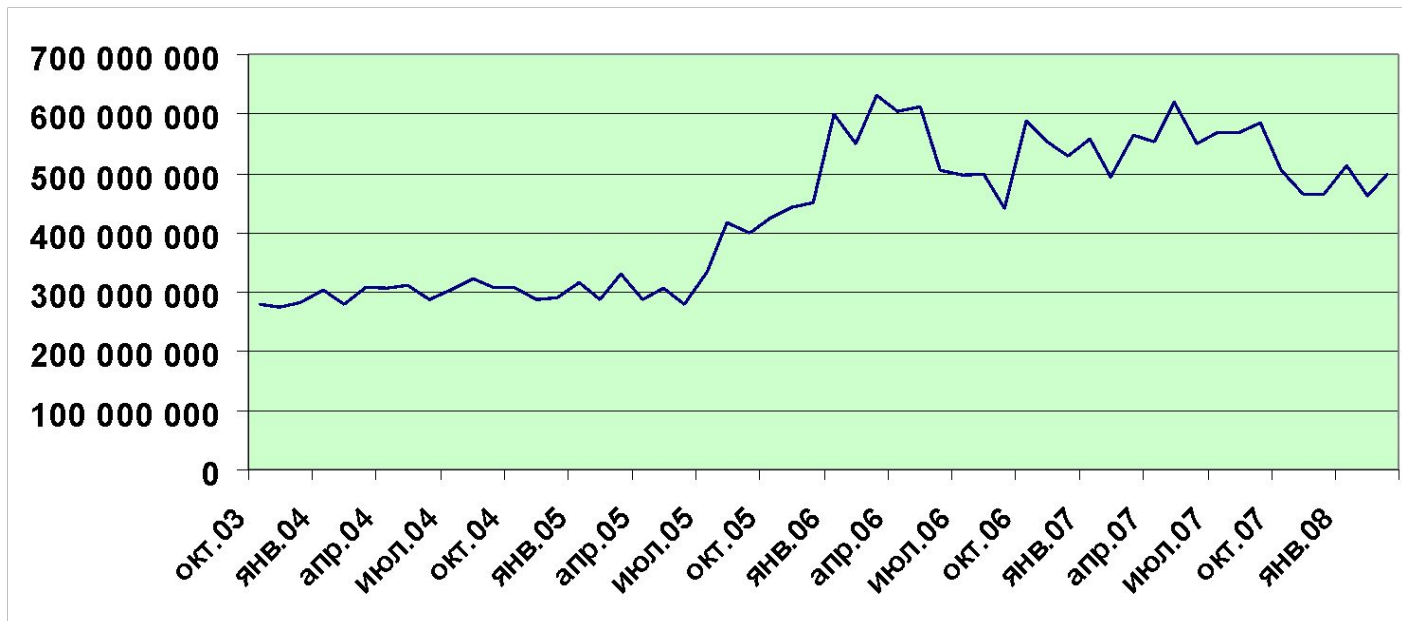
Размеры рекламной сети

- Более 200 000 рекламных площадок
- Более 20 000 рекламодателей
- Более 450 000 000 регистрируемых событий ежедневно

Нагрузка на систему

- 15 000 запросов к баннерным машинам в секунду
- Около 500 запросов к порталу в секунду

Число регистрируемых событий



- Распределение нагрузки
 - Балансировка запросов с помощью аппаратного LB
- Повышение отказоустойчивости
 - Взаимозаменяемость баннерных машин
 - Дублирование ключевых компонентов системы

Особенности бизнес модели

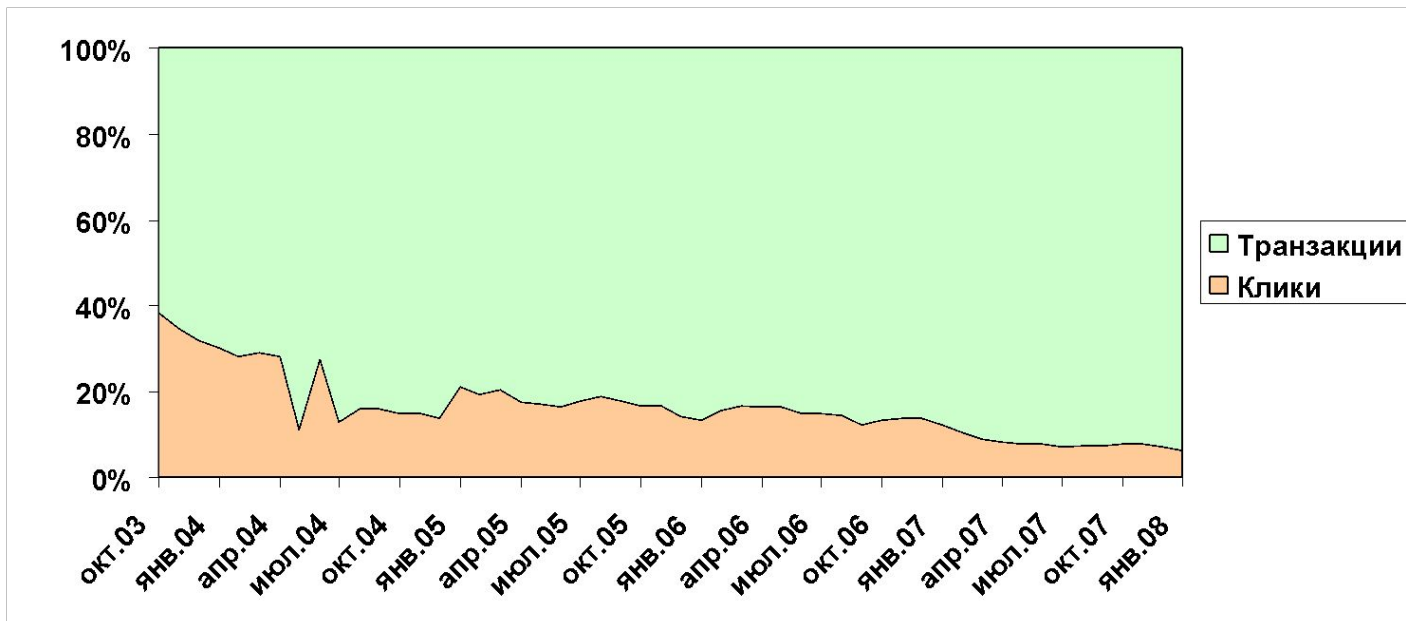
- Качество сервиса определяется количеством участников
- Отсутствует комиссия за показы
- Площадки заинтересованы в эффективном размещении
- Вовлеченность в процесс

Что регистрируем

- Показы
- Клики
- Транзакции



Комиссия за клики и транзакции



Основные проблемы

- Очень много данных
- Очень мало времени на обработку
- Система должна работать непрерывно

Способы уменьшения объема данных

- Предварительная группировка
- Проверка кликов

Способы ускорения обработки

- Ускорение загрузки там, где это возможно
- Задачи не должны блокировать друг друга
- Запуск процессов по расписанию
- Взаимосвязанность задач – это хорошо

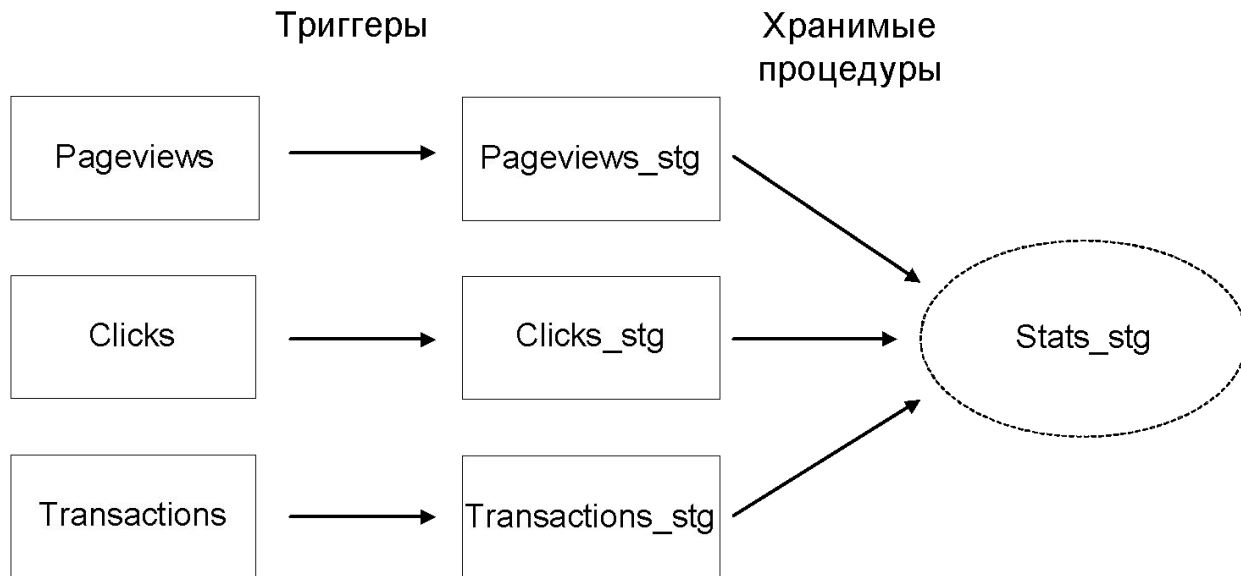
Этапы обработки статистики

- Загрузка логов с баннерных серверов
- Предварительная группировка данных
- Проверка кликов
- Загрузка данных в БД
- Расчет статистики

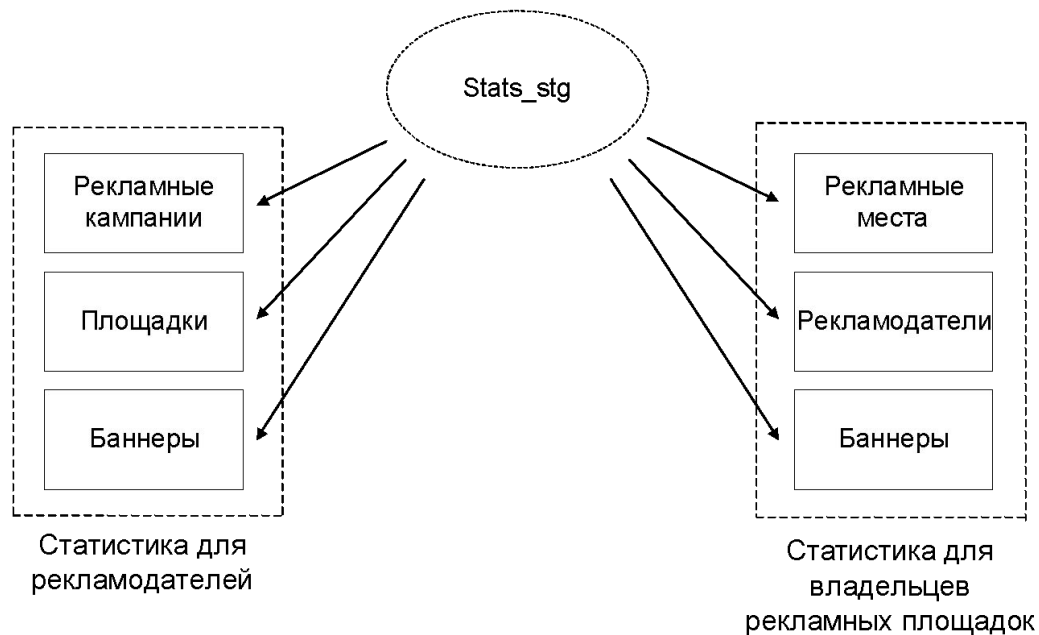
Проблемы базы данных

- Гигантский объем таблиц с показами и кликами
- Необходимо блокировать загрузку на время расчета статистики

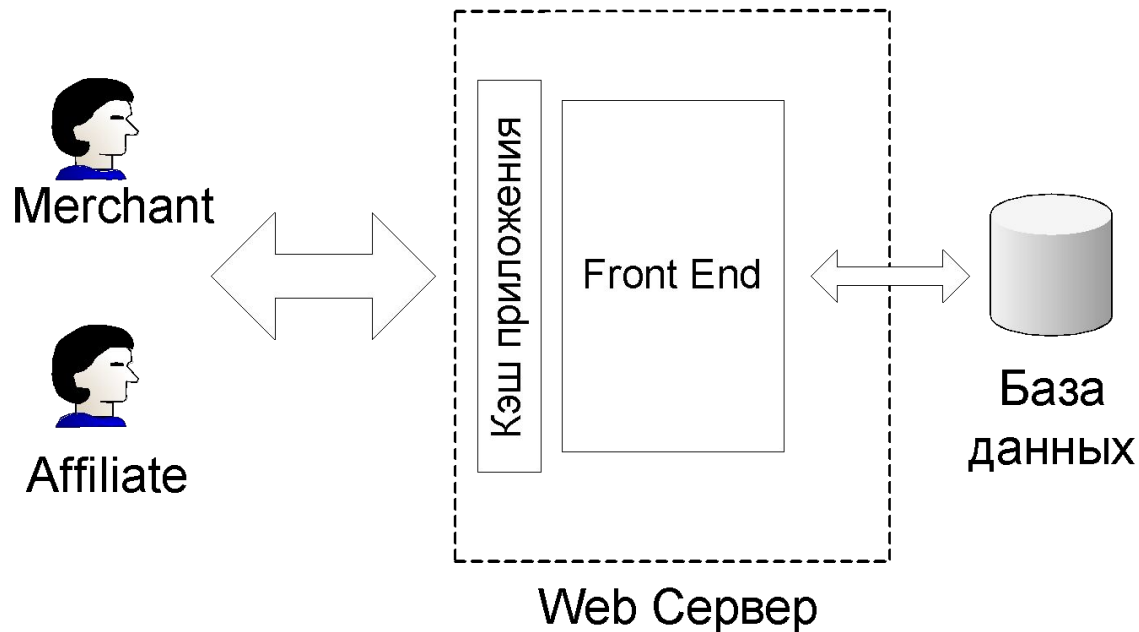
Предварительное заполнение БД



Денормализация данных



Отображение статистики



Почему мы кэшируем данные

- Запросы часто повторяются
- Статистика обновляется по расписанию. Если нет новых данных, зачем обращаться к БД?
- Сортировка средствами БД работает медленнее

Контроль работоспособности

- Остановка системы приводит к лавинообразному накоплению необработанной информации
- Клиенты хотят видеть свои данные
- Обработка статистики используется для биллинга

Основные индикаторы

- Число необработанных файлов на сервере
- Задержка загрузки файлов с баннер сервера
- Задержка обработки статистики

Что мы контролируем

- Баннерные сервера
- Сервер обработки статистики
- База данных

Как мы контролируем

- Готовые решения для контроля на аппаратном уровне (Nagios)
- Собственные разработки для контроля на уровне приложения

Как защититься от сбоев

- Раньше знаешь, больше спишь
- Проблему нужно предвидеть
- С проблемой нужно научиться бороться

hl⁺⁺

HighLoad⁺⁺

Вопросы?

Email: artem@gramant.ru

www.gramant.ru

