

Эволюция архитектуры хранения и отдачи фотографий в Badoo

Юрий Насретдинов



<http://www.devconf.ru>

План

- О сайте
- Хранение фотографий (до «облака»)
- «Облако»
- Отдача фотографий
- Заключение

План

- О сайте
- Хранение фотографий (до «облака»)
- «Облако»
- Отдача фотографий
- Заключение

О сайте

- 250 млн пользователей
- 70 тыс RPS на PHP-FPM
- 50 языков интерфейса
- 1 Пб фотографий и видео
- 2,5 ДЦ

Хочу сюда!



Следующие



Один человек хотел с вами встретиться, но он не соответствует вашим критериям. Сбросить фильтр



Ваша популярность
Оч. низкая



Повысить

Посетители 7

Кому вы нравитесь 24

Избранные 1

Заблокированные 7

Активировать Супер силу



Mari, 25

1 ОБЩИХ ЗНАКОМЫХ

0 ОБЩИХ ИНТЕРЕСОВ



2/4

Быстрые клавиши: 1 Да 2 Нет ПРОБЕЛ Следующее фото

Пожаловаться на фото?



Станьте заметнее в Знакомствах

Кивва, Дарья и 355 других девушек поблизости будут видеть вас чаще, если вы добавите показы своего фото.

Сегодня вас видели: 250 человек

Добавить показы

План

- О сайте
- **Хранение фотографий (до «облака»)**
- «Облако»
- Отдача фотографий
- Заключение

Общая архитектура v1

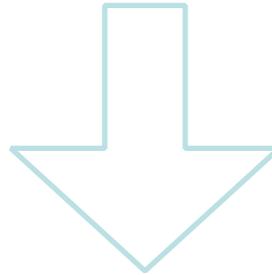


Photo servers

Общая архитектура v2



Photo servers

Fibre Channel



Storage Area Network

Upload (v2)

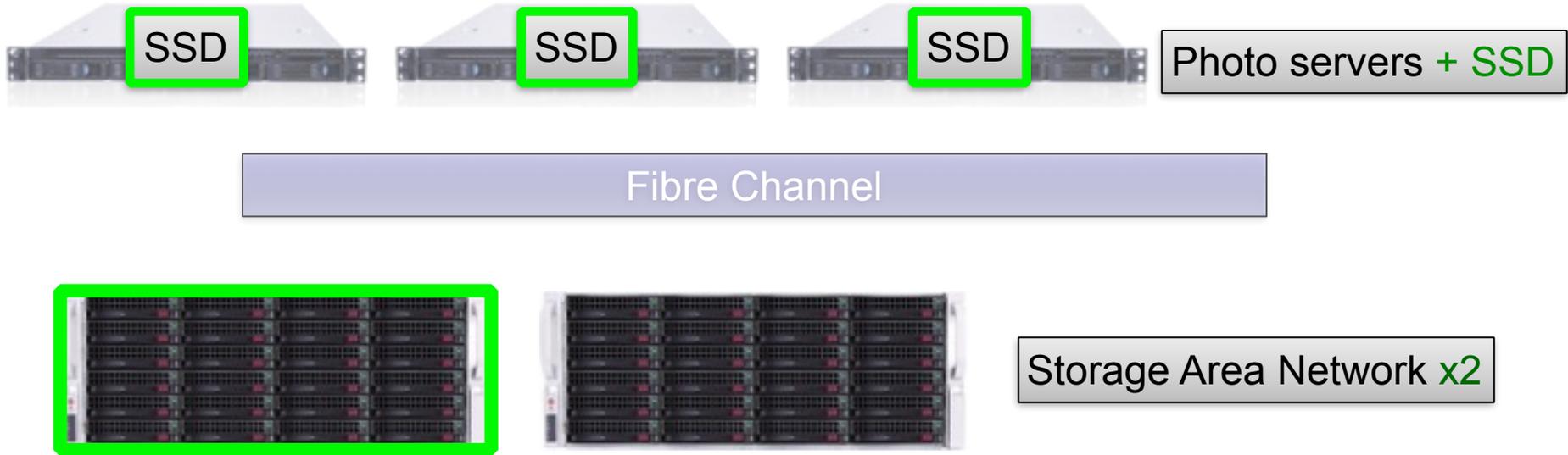


<http://www.devconf.ru>

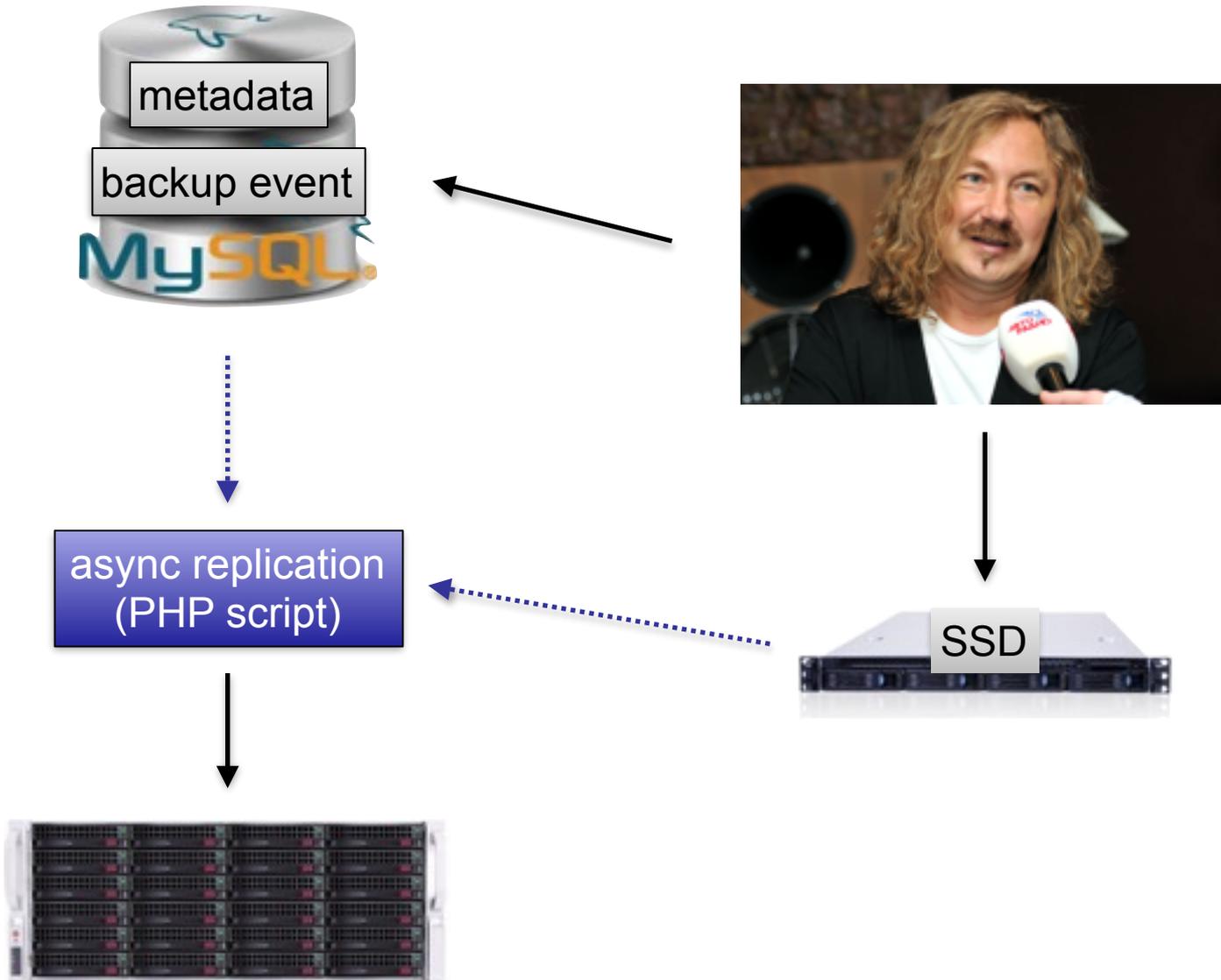
Общая архитектура v2

- + Стоимость (\$ / Тб)
- + Простота для программистов
- **SPoF в виде SAN**
- **Сложность для админов** (сборка SAN, установка FC-свичей и карт, сложные опции монтирования по сети)
- **Низкая производительность**
- **Низкая надежность** (драйвера FC-карт, сами FC-карты, непосредственно SAN)

Общая архитектура v3 (v2+replication)



Upload (v3)



Отдача (v3)

try_files (nginx):



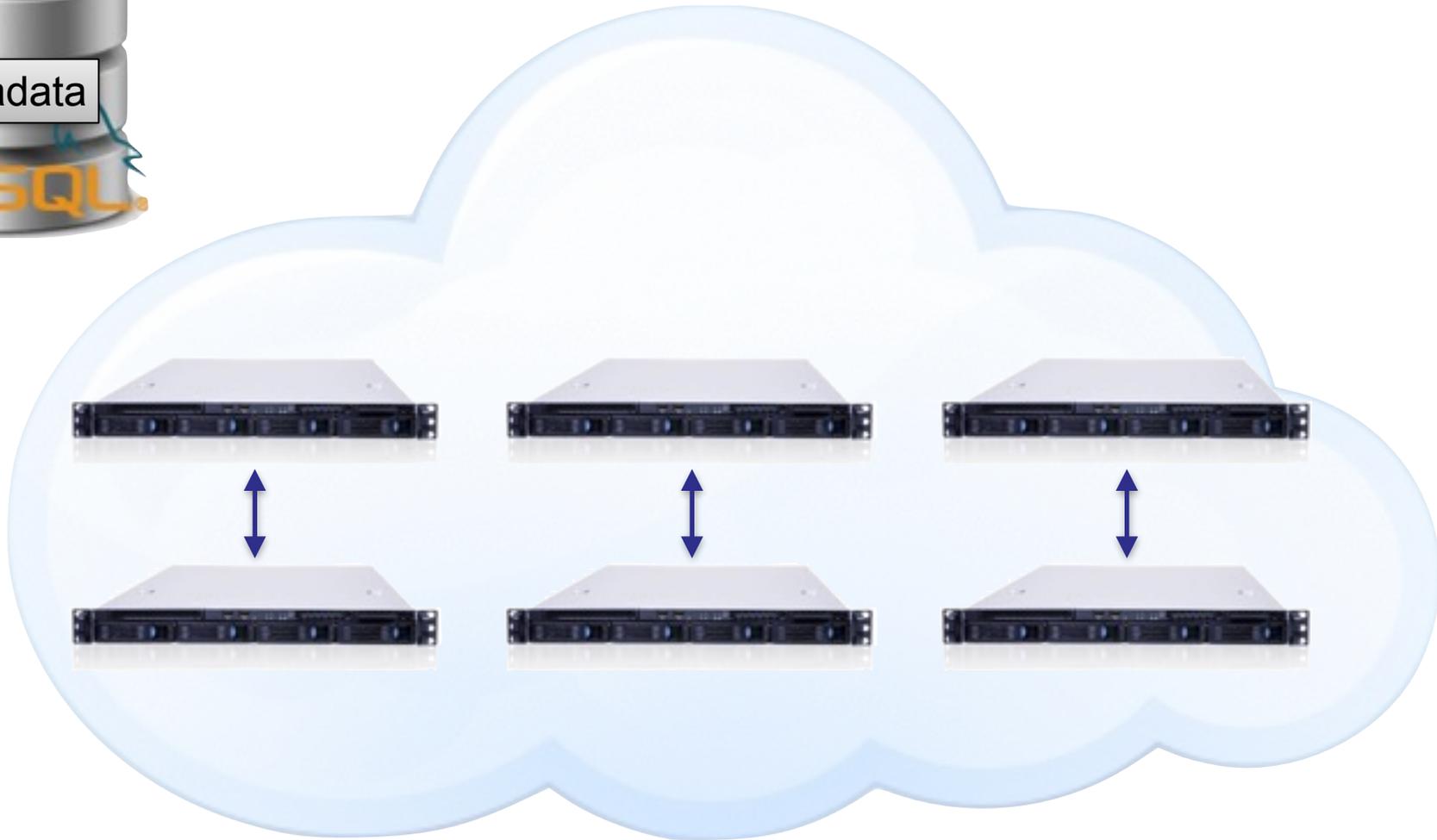
Общая архитектура v3

- + Стоимость (\$ / Тб)
- + ~~Простота для программистов~~
- ~~SPoF в виде SAN~~
- **Сложность для админов** (сборка SAN, установка FC-свичей и карт, сложные опции монтирования по сети)
- **Низкая производительность**
- **Низкая надежность** (драйвера FC-карт, сами FC-карты, непосредственно SAN)

План

- О сайте
- Хранение фотографий (до «облака»)
- **«Облако»**
- Отдача фотографий
- Заключение

Общая архитектура v4 («облако»)



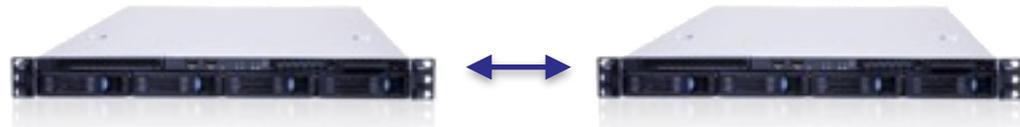
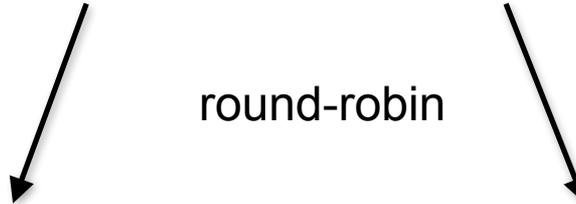
Upload (v4)



timestamp

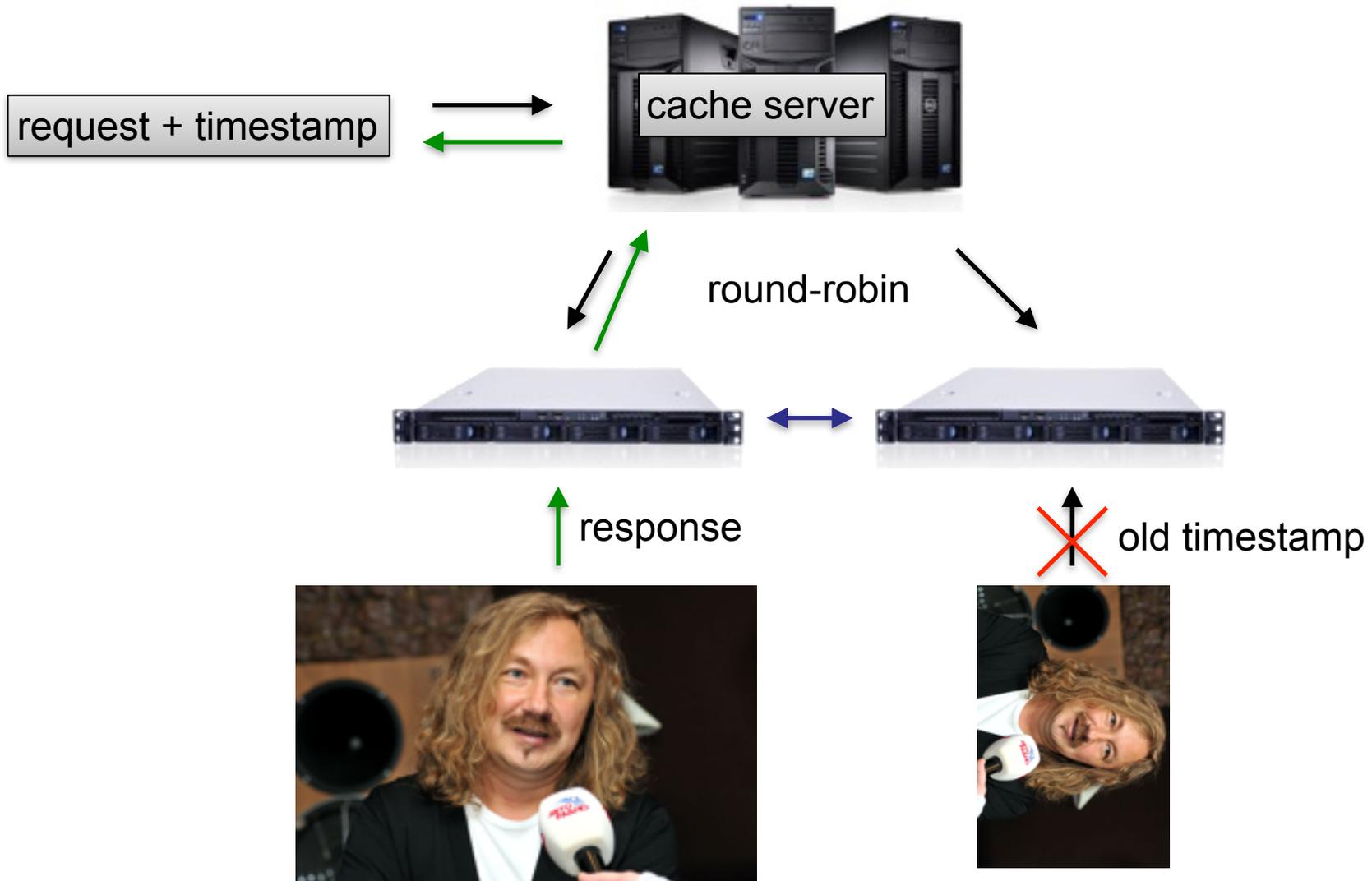


round-robin

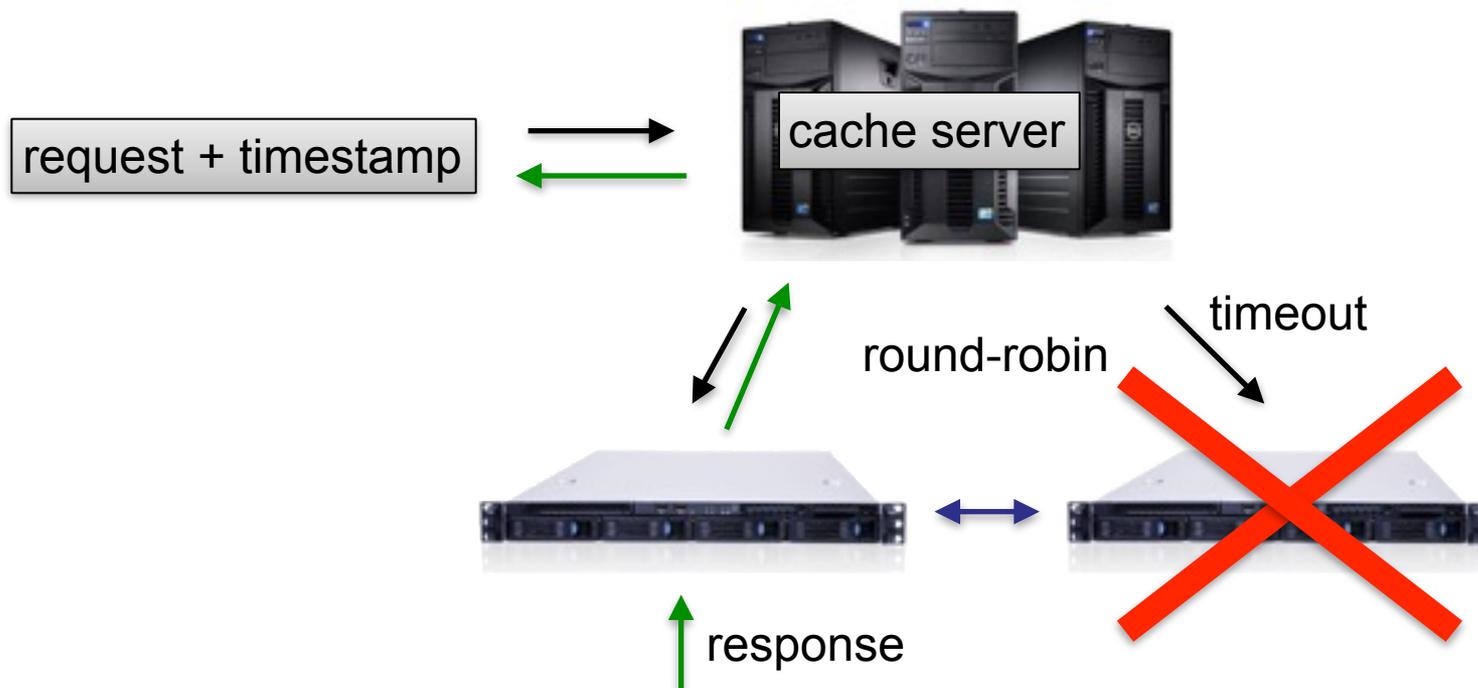


async replication (database queue)

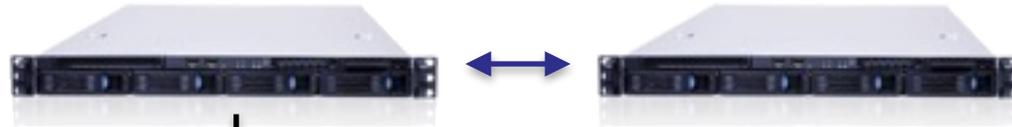
Отдача (v4)



Отдача (падение сервера)



Модификация



check timestamp:

local load
or
remote load

async replication

modify + save



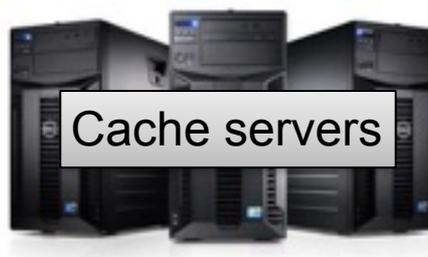
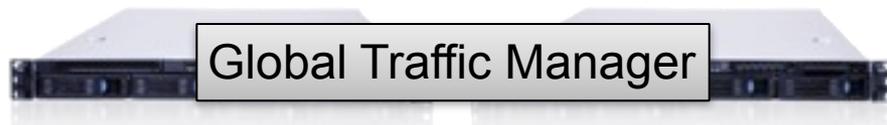
Общая архитектура v4 («облако»)

- + Стоимость (\$ / Тб)
- + Простота для админов
- + Отказоустойчивость
- + Производительность
- Высокая сложность разработки

План

- О сайте
- Хранение фотографий (до «облака»)
- «Облако»
- **Отдача фотографий**
- Заключение

Отдача фотографий

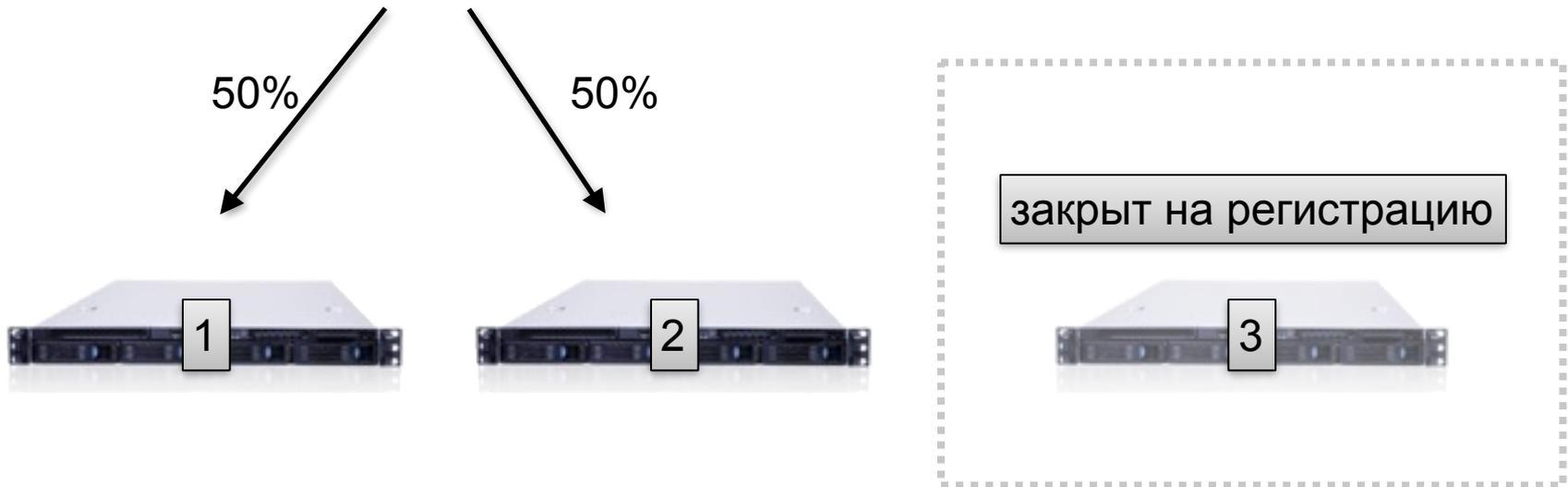


Кеширующие сервера

- Нужны для уменьшения нагрузки на SAN
- Своя логика: кеширование после нескольких ХИТОВ вместо LRU
- Cache hit ratio - 98%
- 100k RPS на фотокеши
- 2k RPS на фотосервера

Регистрация пользователя

Выбор фото-сервера (выбор «place_id»)



Регистрация пользователя

- у каждого пользователя есть свой `place_id`
- `place_id` определяет фото-сервер и базовый путь на диске
- `12345` => `/shared/photos/12/3/4/5/`

Размещение фотографий

- Соответствие «place_id => hostname» в конфиге приложения

Путь на диске:

photo_id/1000
/photos/**12**/**3**/**4**/**5**/**67890**/d**45**/**45678**_**920**.jpg
place_id user_id photo_id size

Фотографии на сайте

Запросы приходят на URI вида:

```
/<place_id>/<user_id>/<photo_id>_<size>.jpg
```

Пример:

```
/12345/67890123/45678_920.jpg
```

place_id

user_id

photo_id

size

Фотографии на сайте

Домен + начало URI:

```
http://pcache-pv-eu1.badoocdn.com/p23/
```

eu1 и **us1** - названия ДЦ

p23 - номер кеширующего сервера

URI до фотографии может быть зашифрован:

```
http://.../p23/hidden?euri=...
```

План

- О сайте
- Хранение фотографий (до «облака»)
- «Облако»
- Отдача фотографий
- Заключение

Заключение

- Храним 1 Пб фотографий
- Использовали SAN
- Сделали user-level репликацию
- Перешли на локальные диски
- Отдача фотографий через кеширующие сервера

Вопросы

Юрий Насретдинов, ведущий разработчик, Badoo

@YNasretdinov

y.nasretdinov@corp.badoo.com