

hl⁺⁺

HighLoad⁺⁺

Рconnect: граната в руках обезьяны

Сергей Аверин

The logo consists of the lowercase letters 'hl' in white on a red square background, followed by two plus signs '++' in white.

HighLoad ++

The logo for 'badoo' features the letters 'b', 'a', 'd', and 'o' in blue, green, and orange respectively, with a stylized orange circle containing a white dot for the second 'o'. This is followed by a black dash and the word 'ЭТО:' in black.

badoo — ЭТО:

Социальная сеть для знакомств с новыми людьми

В Top-200 Alexa с 2007 года

127 миллионов зарегистрированных пользователей

10 миллионов пользователей в день

2 миллиона фотографий загружаются ежедневно

The logo consists of the lowercase letters 'hl' in white on a red square background, followed by two plus signs '++' in white.

HighLoad++

badoo — ЭТО:

30 тыс. запросов/сек к PHP backends

MySQL, PHP, C/C++, Linux, nginx, PHP-FPM, memcached

Много своего

Pconnect — что это?

Persistent connection/keep-alive/connection reuse — по сути одно и то же

Pconnect в мире FastCGI-демонов

Плюсы:

- Экономия ресурсов сервера
- Меньшая нагрузка на сеть
- Экономия времени и ускорение отклика

Действительно Plug'n'Play?

<http://www.php.net/manual/en/features.persistent-connections.php>:

“Persistent connections were designed to have one-to-one mapping to regular connections. That means that you should always be able to replace persistent connections with non-persistent connections, and it won't change the way your script behaves.”

Это маркетинг. И не более того.

В реальности

Чаще всего не Plug'n'Play

Должны быть рассчитаны на rconnect:

- протокол
- клиентская библиотека
- серверное ПО

Не «серебряная пуля»

В больших/highload системах более ощутимо

Проблемы



Софт не рассчитан на большое кол-во открытых соединений

Сколько памяти съедает инициализированный коннект?

Выделяется ли в софте по треду/процессу на коннект?

Есть ли ограничение на кол-во коннектов с одного адреса?

Ограничения сетевой подсистемы Linux

Проблемы

2

Разберитесь в коде

Не закрывайте хорошие соединения функцией `fclose`

Проверьте, как работает ваш extension, особенно все его warning'и и error'ы

- <http://www.php.net/manual/en/memcached.addservers.php>

В php вам поможет `ftell($socket)`

Проблемы



Когда нужно закрывать соединения

При синтаксических ошибках

При тайм-аутах

Если вы не дочитали ответ до конца

Проблемы



Stateful протоколы

Вы всегда знаете состояние полученного соединения?

Транзакции/переменные/авторизация

Реинициализация соединения

Проблемы

Асинхронная природа сокетов

Должна быть возможность сопоставления запросов и ответов

- Единственное, где я это нашел, это в MongoDB и MessagePack-RPC

<http://www.mongodb.org/display/DOCS/Mongo+Wire+Protocol>

<http://wiki.msgpack.org/display/MSGPACK/Design+of+RPC>

Способ, которым мы пользуемся (не дает гарантии)

Проблемы



set_time_limit

Есть, было и будет в FastCGI-демонах

Терминация запроса пользователем

Shutdown обработчик

Проблемы



Ограниченный функционал FastCGI-демонов

Клиентские библиотеки зачастую бездумно написаны

Не всегда можно понять новое ли это соединение

Нет полноценных connection pool'ов

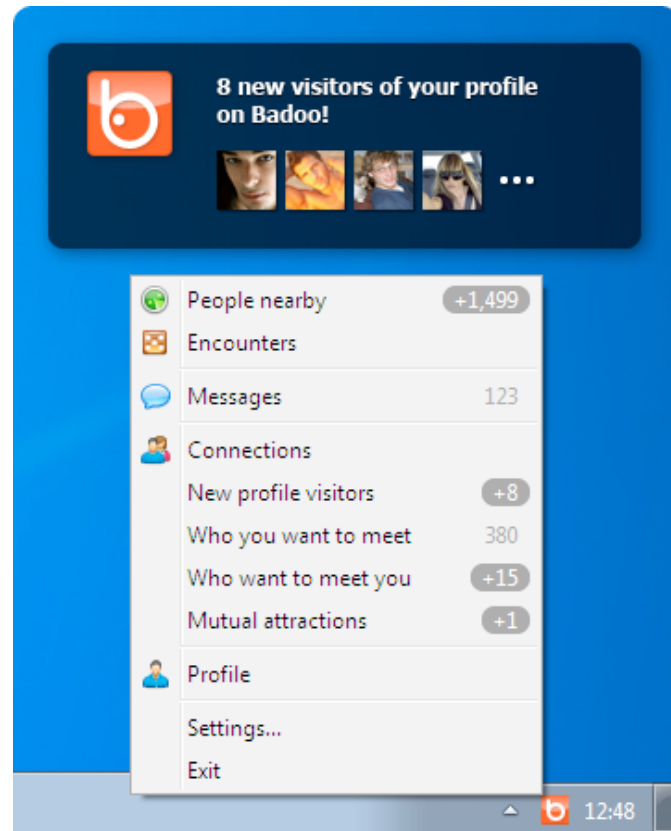
hl⁺⁺

HighLoad⁺⁺

От теории к практике

Badoo Desktop

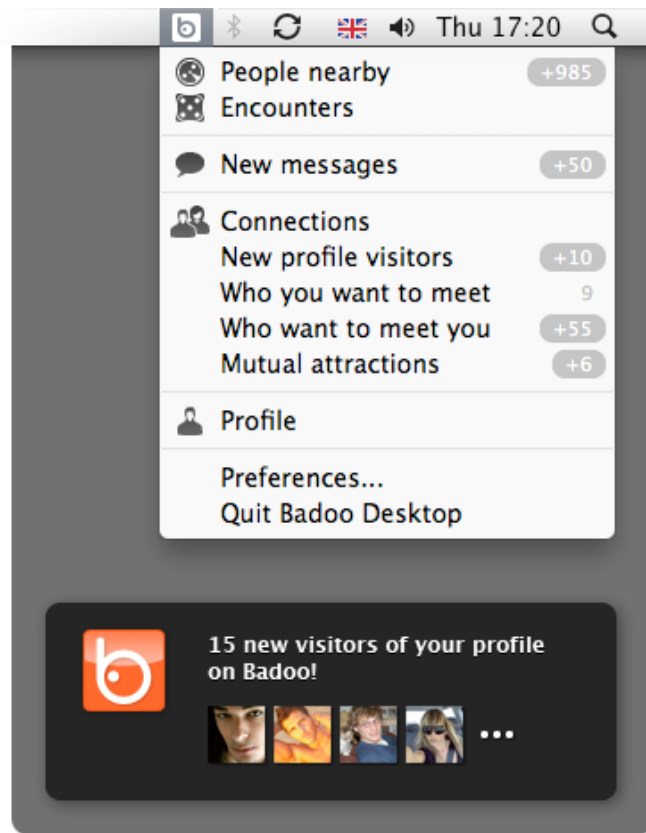
- Бесплатная Win/Мас программа
- Поддерживает ваш онлайн-статус на сайте
- Уведомления о новых событиях
- Дает нам местоположение пользователя
- Удобный доступ к разделам badoo.com



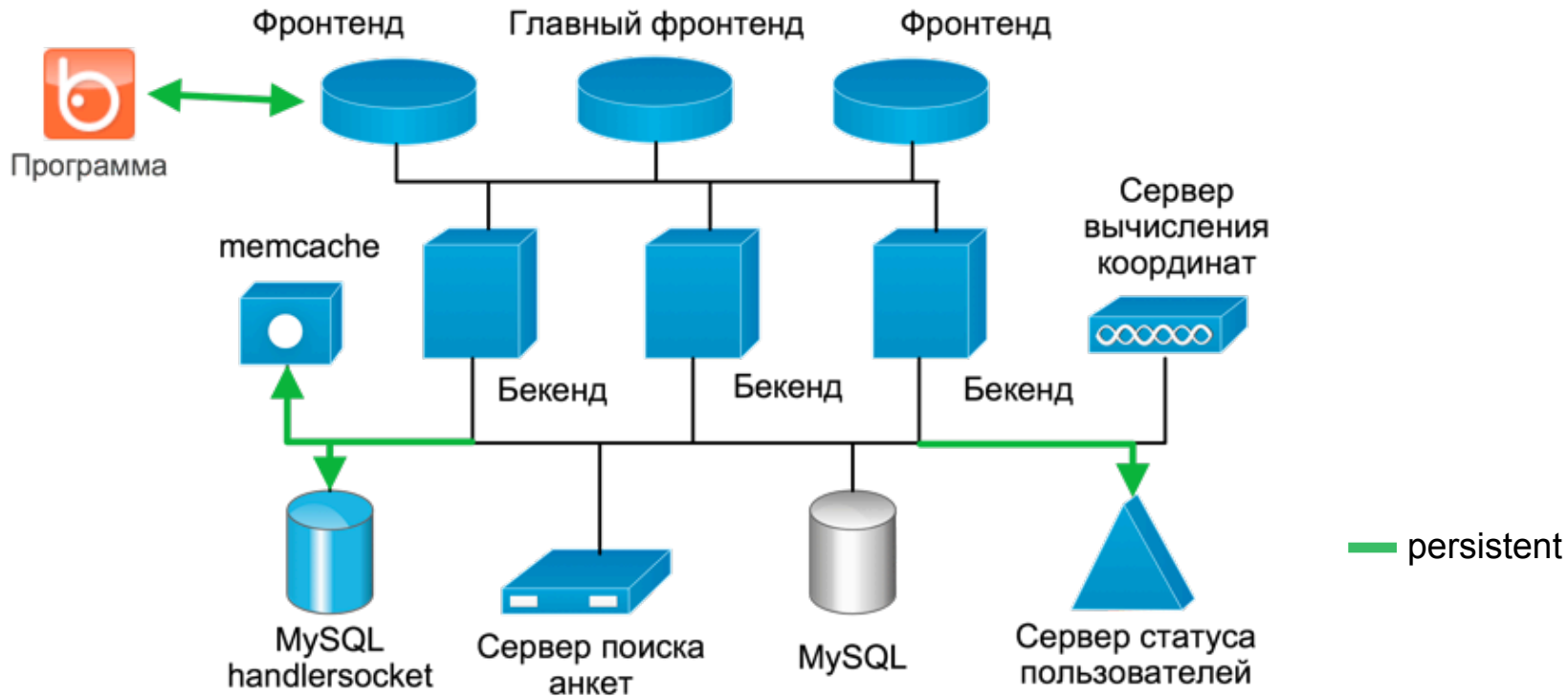
hl⁺⁺**HighLoad⁺⁺**

Badoo Desktop

- 3 млн. пользователей в месяц
- 1 млн. подключенных программ в пике
- 20 тыс. запросов/сек к PHP backends



Наш setup



Программы —> Frontend

Есть способ сделать, чтобы «просто работало»:

- Гарантии доставки нет
- Ответы являются отдельными командами
- Одна программа — одно соединение
- Есть сохранение порядка
- Нарезкой байт на отдельные команды занимается демон, держащий соединения

The logo consists of the letters 'hl' in white on a red square background, followed by two plus signs '++' in white.

HighLoad++

Frontend —> Backend

Отдельный http-запрос на каждую входящую команду

Рconnect пока нет, но планируем через HTTP Keep-Alive

http://en.wikipedia.org/wiki/HTTP_persistent_connection

Backend —> MySQL

Протокол на rconnect полностью не рассчитан

Есть packet number, но оно обнуляется с каждым новым запросом

Есть реинициализация соединения, используйте mysql_i_connect

Тред на коннект

Для нас малоинтересно, баз много, php-серверов много, запросы редки

http://forge.mysql.com/wiki/MySQL_Internals_ClientServer_Protocol

The logo consists of the letters 'hl' in white on a red square background, followed by two plus signs '++' in white.

HighLoad++

Backend —> HandlerSocket

Протокол на rconnect не рассчитан, зато коннект ест мало ресурсов

Используем rconnect

Написали свою библиотеку на php

Проверяем, что получили ключ, который просили

Получили прирост в 3—5 раз

<http://tinyurl.com/64vwn52>

Backend —> Memcached

Протокол на rconnect не рассчитан, зато коннект ест мало ресурсов

Используем rconnect

Extension нормально работать не смог, написали свою библиотеку на php

Тщательно ловим все ошибки

Проверяем, что получили ключ, который просили

Получили прирост в 2 раза

<https://github.com/memcached/memcached/blob/master/doc/protocol.txt>

<http://code.google.com/p/memcached/wiki/MemcacheBinaryProtocol>

The logo consists of the letters 'hl' in white on a red square background, followed by two plus signs '++' in white.

HighLoad++

Backend —> своя in-memory DB

Протокол на rconnect не рассчитан

Используем rconnect

Только команды на запись, ответ только либо OK либо ERROR

Гарантии записи нет и мы к этому готовы

Каждая отдельная ошибка нам не важна, они всегда идут только пачками

hl⁺⁺

HighLoad⁺⁺

Продвинутые подходы

Ленивая инициализация

В handlersocket-протоколе есть открытие индекса таблицы

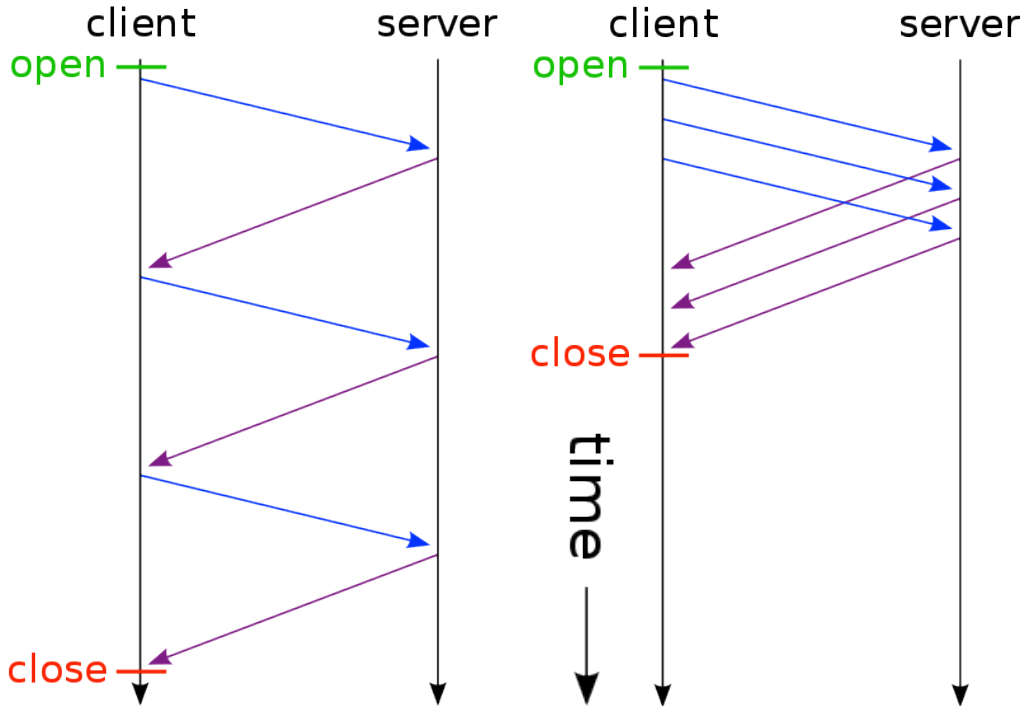
Без rconnect открывали каждый раз после коннекта

С rconnect ловим ошибку про неизвестный индекс

Теперь мы можем открывать его только если это нужно

Pipelining

Если в протоколе есть уникальный request-response id на каждый запрос, то будет надежно работать



Connection pooling

Плюсы:

- Уменьшение времени получения коннекта
- Упрощает код
- Дает контроль над использованием ресурсов

Бывает внутри процесса и в виде отдельного демона

- Внутри FastCGI — пишите сами, плюс у некоторых extension'ов есть свои пулы. В мире Python — SQLAlchemy
- Вне процесса сложнее и функциональнее, обычно есть доп. логика

Connection pooling

pgpool — PostgreSQL connection pool daemon

- <http://pgpool.projects.postgresql.org/>
- В Postgres процесс на коннект

MySQL Proxy — швейцарский нож

- http://forge.mysql.com/wiki/MySQL_Proxy_FAQ
- load balancing, failover, query analysis/modification, R/W splitting

SQL Relay — ODBC, Oracle, MySQL, PostgreSQL, Sybase, MS SQL Server, IBM DB2, Firebird, SQLite and MS Access

- <http://sqlrelay.sourceforge.net/sqlrelay/faq.html>

Чеклист

- 1) Что конкретно вы экономите и с какими последствиями?
- 2) Потянет ли это все ваша система?
- 3) Вы разобрались с кодом, ошибками, закрываете коннект когда нужно?
- 4) Рассчитан ли протокол на `rsconnect`?
- 5) Если нет, вы подстелили себе соломы?
- 6) Вы правильно работаете со `stateful` протоколом?
- 7) Вы используете `connection pool` для экономии ресурсов?
- 8) Вы готовы писать свои велосипеды?

hl⁺⁺

HighLoad++

Спасибо!

Вопросы?

Контакты:

http://twitter.com/ryba_xek

s@averin.ru

<http://averin.ru/slides/>

Vadoo ищет классных людей!

Нам нужны:

- умный и вменяемый PHP/MySQL программист
- релиз-менеджер
- QA специалисты

Что даём:

- интересную работу
- 2000 евро за успешную рекомендацию

```
$> cd pub  
$> more beer
```