

Как студенты и филолог писали технически сложный проект: акторы, методы разладки, nlp

Евгений Тюменцев
Генеральный директор
Hello World! Technologies

Тюменцев Евгений

19 лет преподаю в вузе (мат. факультет, ФКН)

**14 лет в ИТ (разработчик, архитектор,
руководитель проектов, тех. директор,
директор)**



#0 Задача

Жил-был сайт ...

Новости, недвижимость, авто, объявления



2012 год

11 лет проекту, сменилось 2-3 состава разработчиков

4,5 с – среднее время отклика страницы

70 тыс. 500-х ошибок на 1 млн. страниц

30 с - время поиска по тексту объявления (timeout)

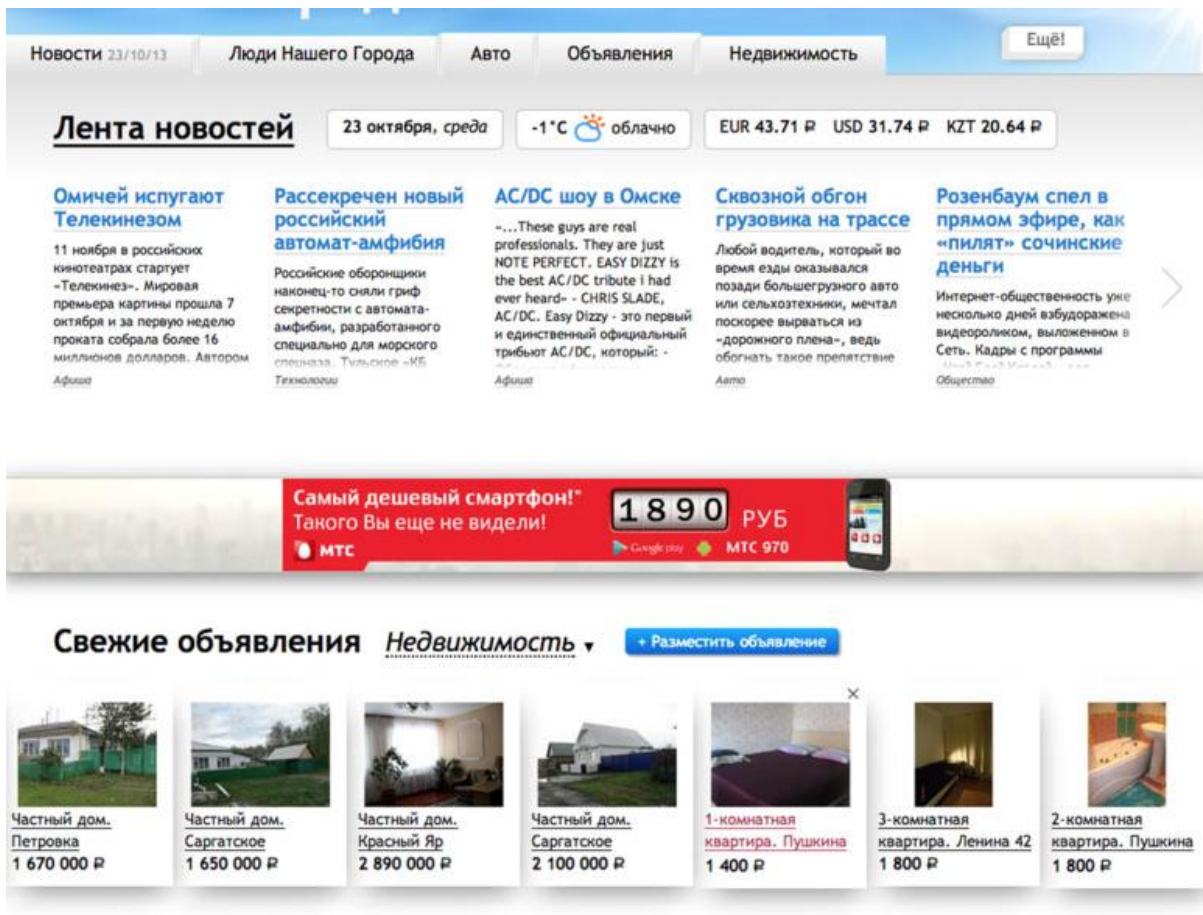
более 200 таблиц БД

1 запрос в неделю на срочные большие изменения

1 серьезная авария в месяц

И вот захотел сайт сделать
редизайн ...

Полный редизайн



Новый дизайн (в одной и ведущих веб-студий)

<1с время отклика страницы

“Полнотекстовый поиск” - основной поиск по сайту

Стабильная работа сайта

Соответствие ФЗ о рекламе

Microsoft

7 студентов и 2 опытных разработчика, филолог-аналитик

400 агентств недвижимости (интеграция)

“Полнотекстовый поиск”

The image shows a screenshot of a search engine interface. At the top, there are navigation links: "Покупка авто", "Прокат", "Каталог", "Сервис", "Работа", and "Продать авто". Below these is a search bar containing the text "Подержанный Ford Focus автомат за 800 000 руб." and a "найти" button. Below this is a blue car graphic with a search bar inside it containing "Toyota до 700 тыс ру" and a "Найти" button. Below the car is another search bar with the text "Я ищу" and a "Найти" button. At the bottom, there are five categories: "Компьютеры,", "Ремонт и", "Детский мир", "Мобильная связь", and "Уютный дом". Below these is a large blue search bar with a "Найти" button and the word "ПОИСК" in large letters.

Что люди ищут?

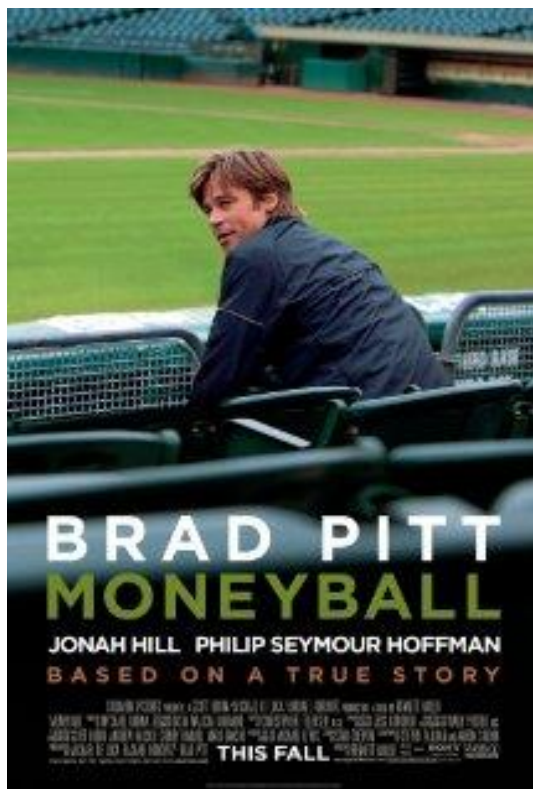
Двушка в нефтах

Однокомнатная у политеха

Хрущевка недорого

Матрешка

#1 Что делать?



Как сделать проект при таких исходных данных?

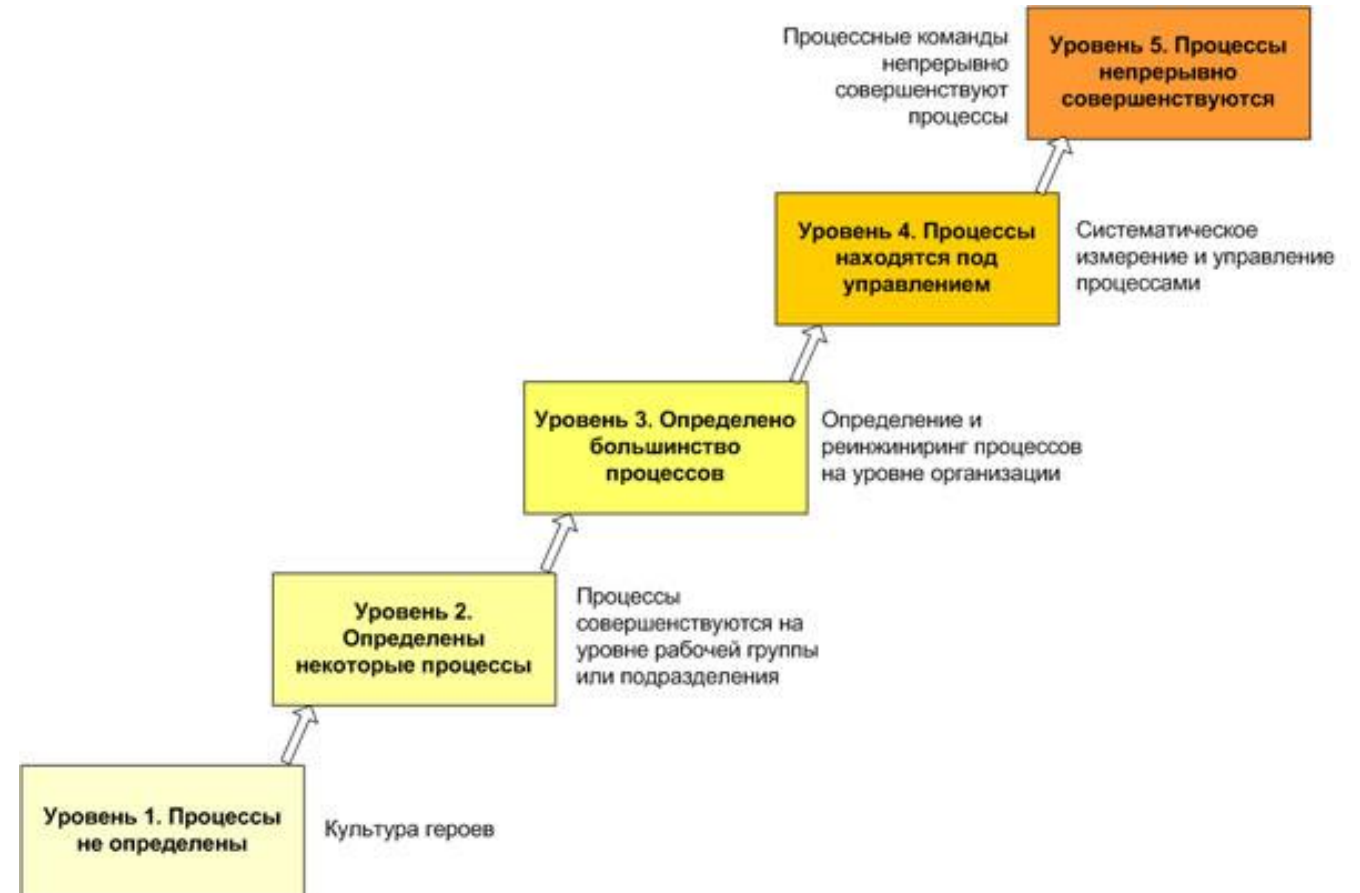
Надо, по-максимуму, исключить влияние квалификации отдельного программиста на результаты проекта

Человек, который изменил все

Идея



Гарольд Керцнер Модель зрелости управления процессами



Архитектура - акторы



Карл Хьюит

1973 г. A Universal Modular ACTOR
Formalism for Artificial Intelligence

Актор – вычислительная сущность

- Отправить конечное число сообщений другим акторам
- Создать конечное число акторов
- Выбрать поведение для приема следующего сообщения

Глобальное время vs акторы



Измерения – карты Шухарта



Уолтер Шухарт

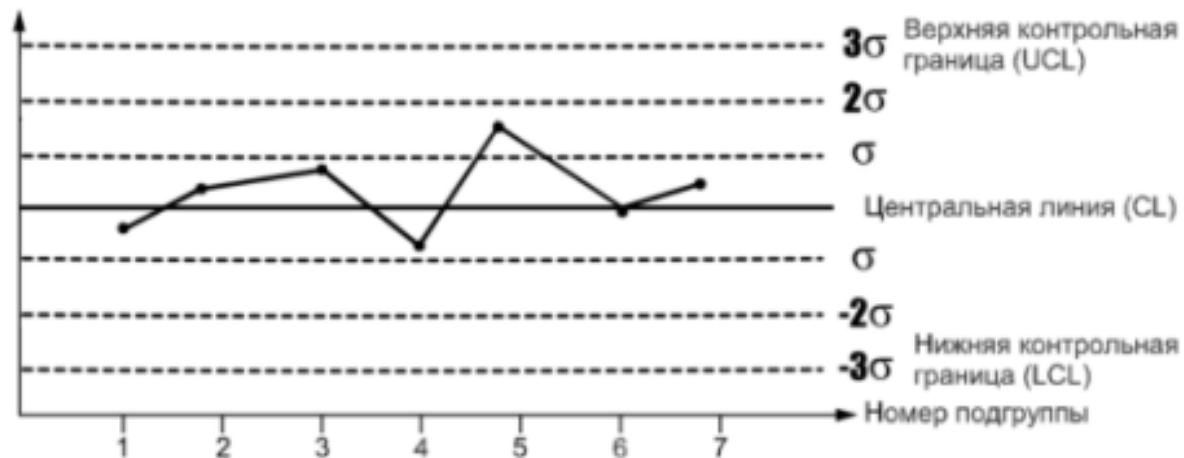
1923 г. AT&T

Ламповые усилители сигнала

Количество аварийных бригад и транспорта?

Применяются в различных сферах производства как средство постоянного улучшения качества выпускаемой продукции

Немного о картах Шухарта



$$\bar{\bar{X}} = \frac{1}{n} \sum X_i$$

для среднего

$$UCL = \bar{\bar{X}} + A_2 \bar{R}$$

$$LCL = \bar{\bar{X}} - A_2 \bar{R}$$

$$CL = \bar{\bar{X}}$$

$$R = \max_i X_i - \min_i X_i$$

для размаха

$$UCL = D_4 \bar{R}$$

$$LCL = D_3 \bar{R}$$

$$CL = \bar{R}$$

ГОСТ Р 50779.42-99 (ISO 8258-91)

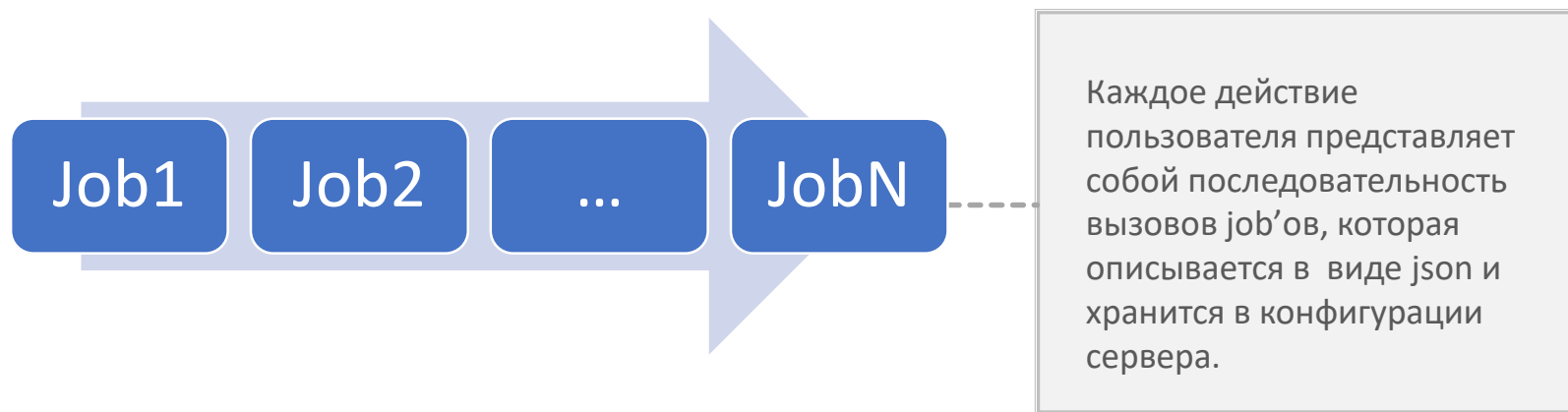
#2 Что сделано?

Как решить проблему зависимостей?



Карты сообщений и сервер конфигураций

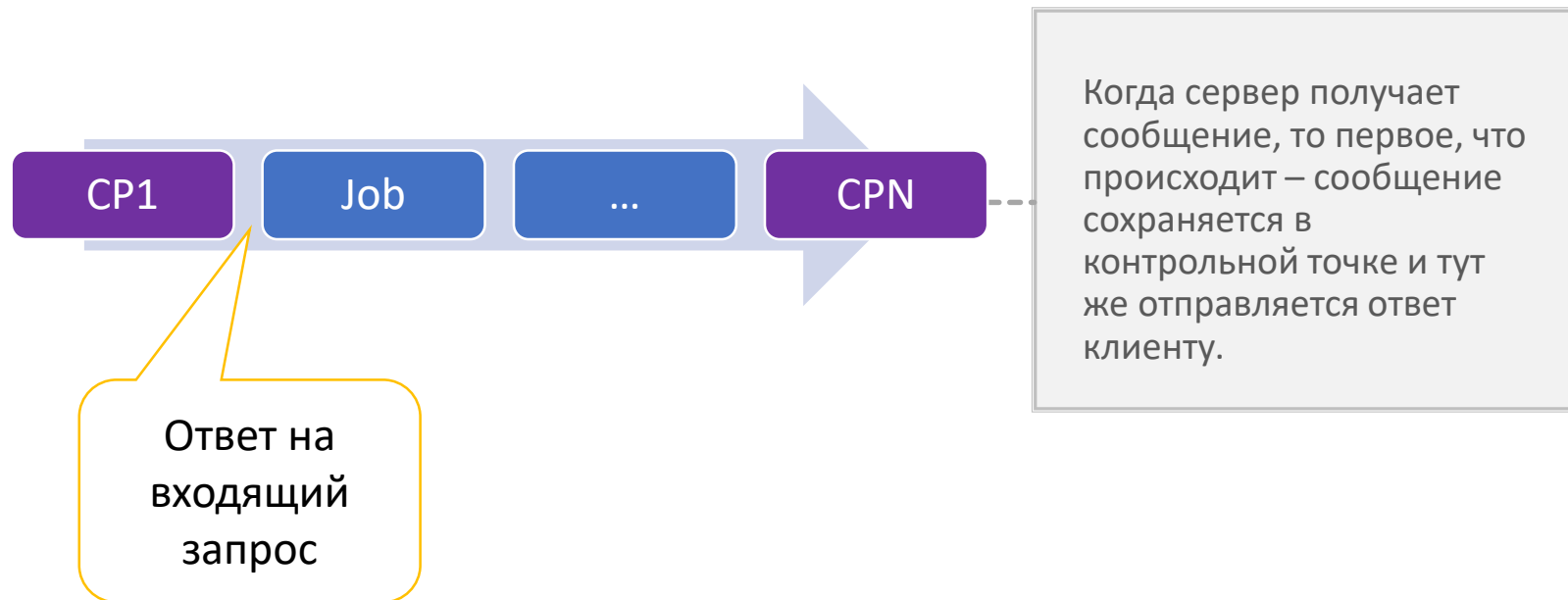
Можно менять алгоритм обработки действий пользователя без программирования или быстро вернуть последнюю рабочую версию.



Как сгладить ошибки для пользователей?

Контрольные точки

Если в момент обработки произойдет ошибка, то пользователь этого никогда не заметит, а разработчики всегда могут повторить данную операцию без участия пользователя, когда исправят ошибку.

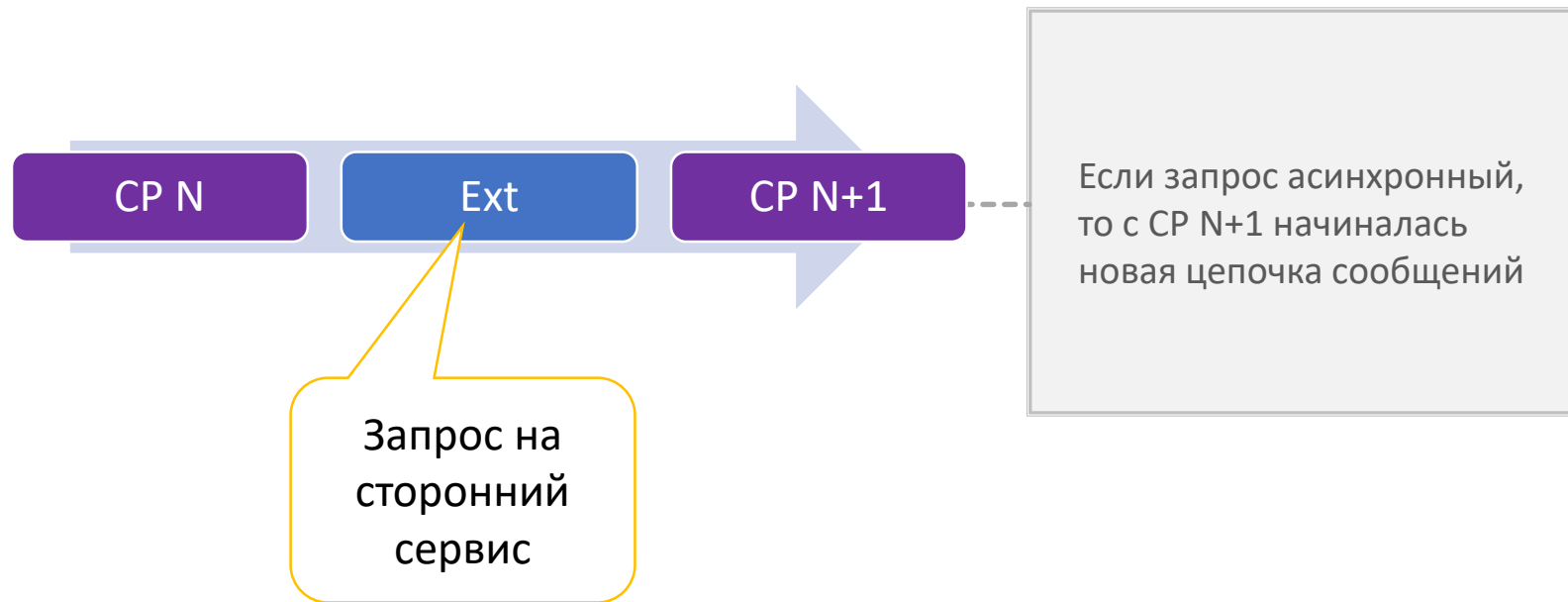


Взаимодействие с внешними сервисами

Если результат сервиса потерять нельзя

Перед вызовом сервиса и после ставим контрольную точку.

Перед – чтобы не бомбардировать сторонний сервис, если у нас пойдет что-нибудь не так, а после, чтобы не потерять ответ сервиса, если у нас, что-нибудь сломается после.



Взаимодействие с внешними сервисами

Proxy Actor

Как контролировать запросы в БД?

Документо-ориентированное хранилище (в MS SQL)

```
{  
  "firstName": "Иван",  
  "lastName": "Иванов",  
  "address": {  
    "streetAddress": "Московское ш., 101, кв.101",  
    "city": "Ленинград",  
    "postalCode": 101101  
  },  
  "phoneNumbers": [  
    "812 123-1234",  
    "916 123-4567"  
  ]  
}
```

Вся информации о
сущности хранится в
виде одной записи в
формате json.

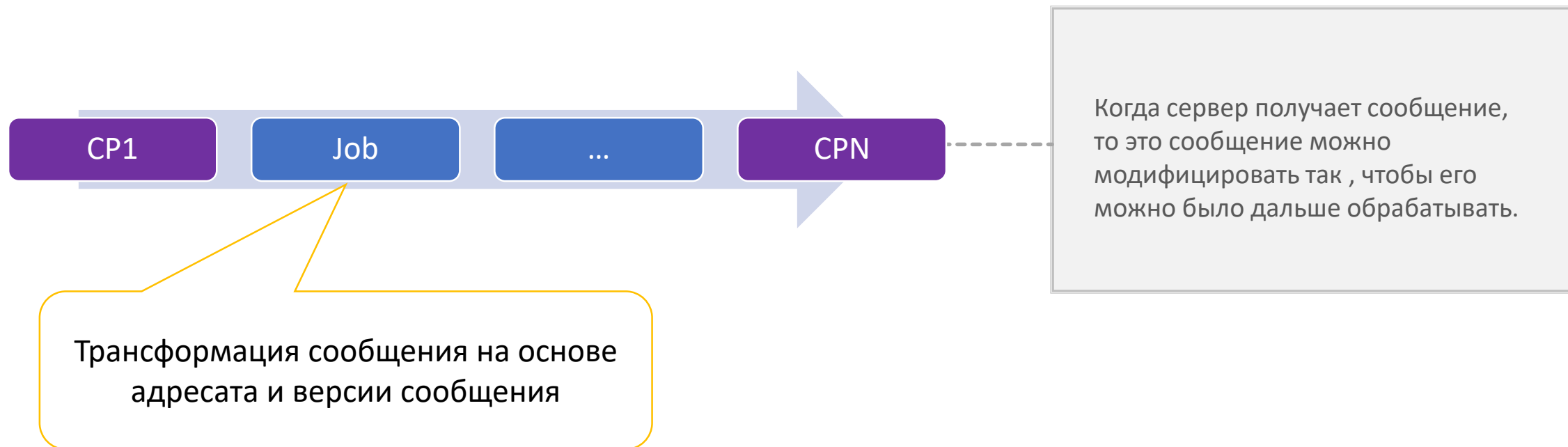
Пример поискового запроса

```
{  
  "FullTextSearch": "tag001Mazda tag0023",  
  "Ranges": {  
    "Price": {"left": 200000, right: 700000}  
  },  
  "Order":  
  [  
    {Field: "PublishDate", Reverse: true}  
  ],  
  "Count": 20,  
  "Page": 2  
}
```

Мы – не Yandex!

Версионирование сообщений

Проще самим приспособиться, чем добиваться полного соответствия



Пользователей не переделаешь

улица Осавиахимовская → улица Ос^оавиахимовская.

улица Мира → проспект Мира

Южная, 2 – целых три дома в разных частях города

Двушка → 2-комнатная квартира

Нефты → микрорайон городок нефтяников

Тойота, тайота, toyota, тоета, тоёта, таета – всего 15 вар.

Матрешка → Mazda 3

Буханка, таблетка, и т.д.

Таз → ВАЗ

Отказались от жестких форматов данных

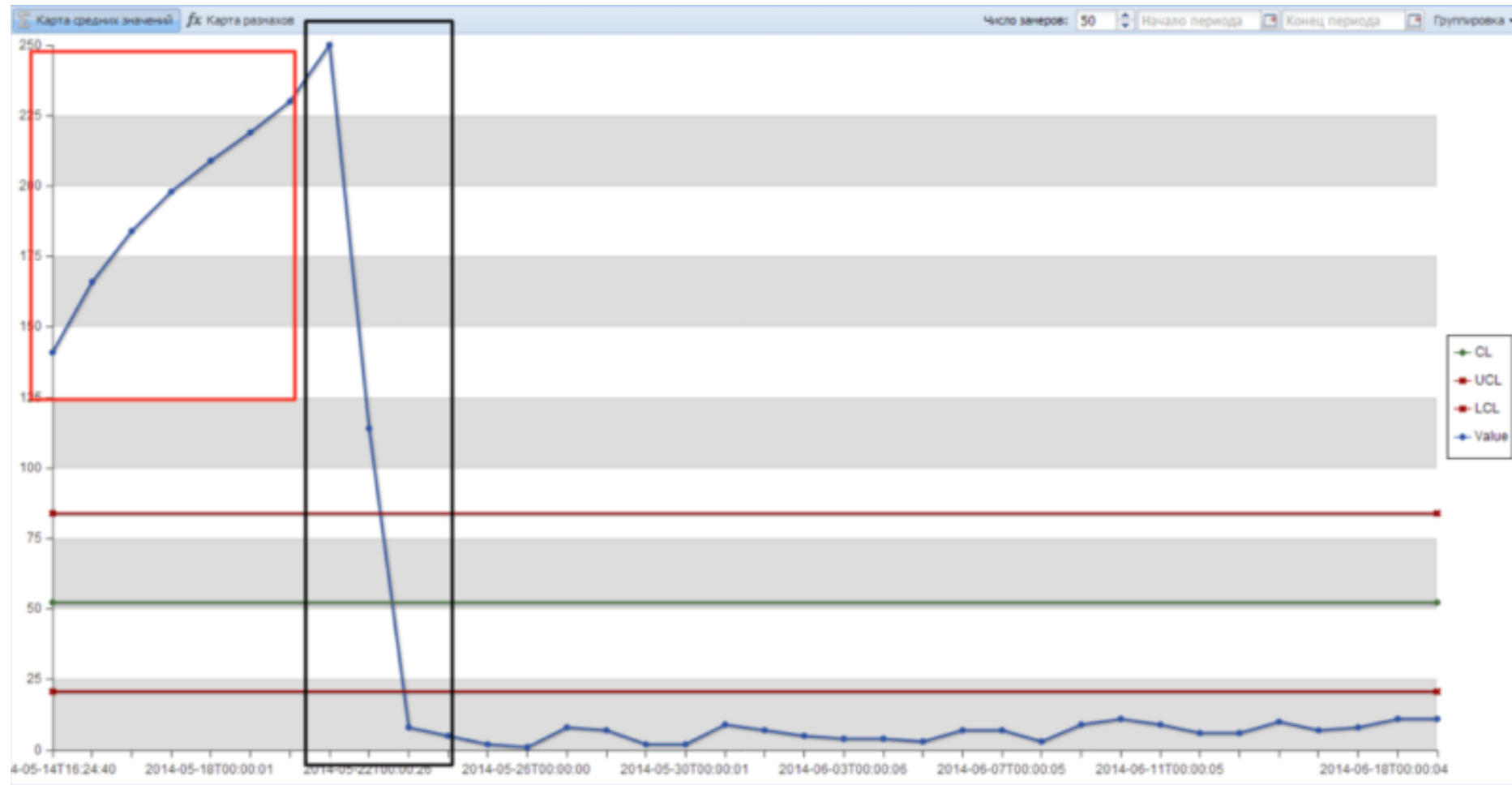
Характеристики объявления или запроса приходят в свободной форме и парсятся уже на сервере:

2-комн. квартира, САО, улица Мира, 64, рядом с остановкой Политех, в х/с, с/у/р, общая площадь: 50 кв. м, жилая – 34, кухня 8, на 3 этаже 5-этажного панельного дома, хрущевка

Двушка в Советском округе, проспект Мира, 64, в районе ост. Политех, 50/34/8, 3/5п, не требует ремонта, санузел раздельный, хрущевка.

#3 Что контролировалось?

Количество сообщений в Check Point



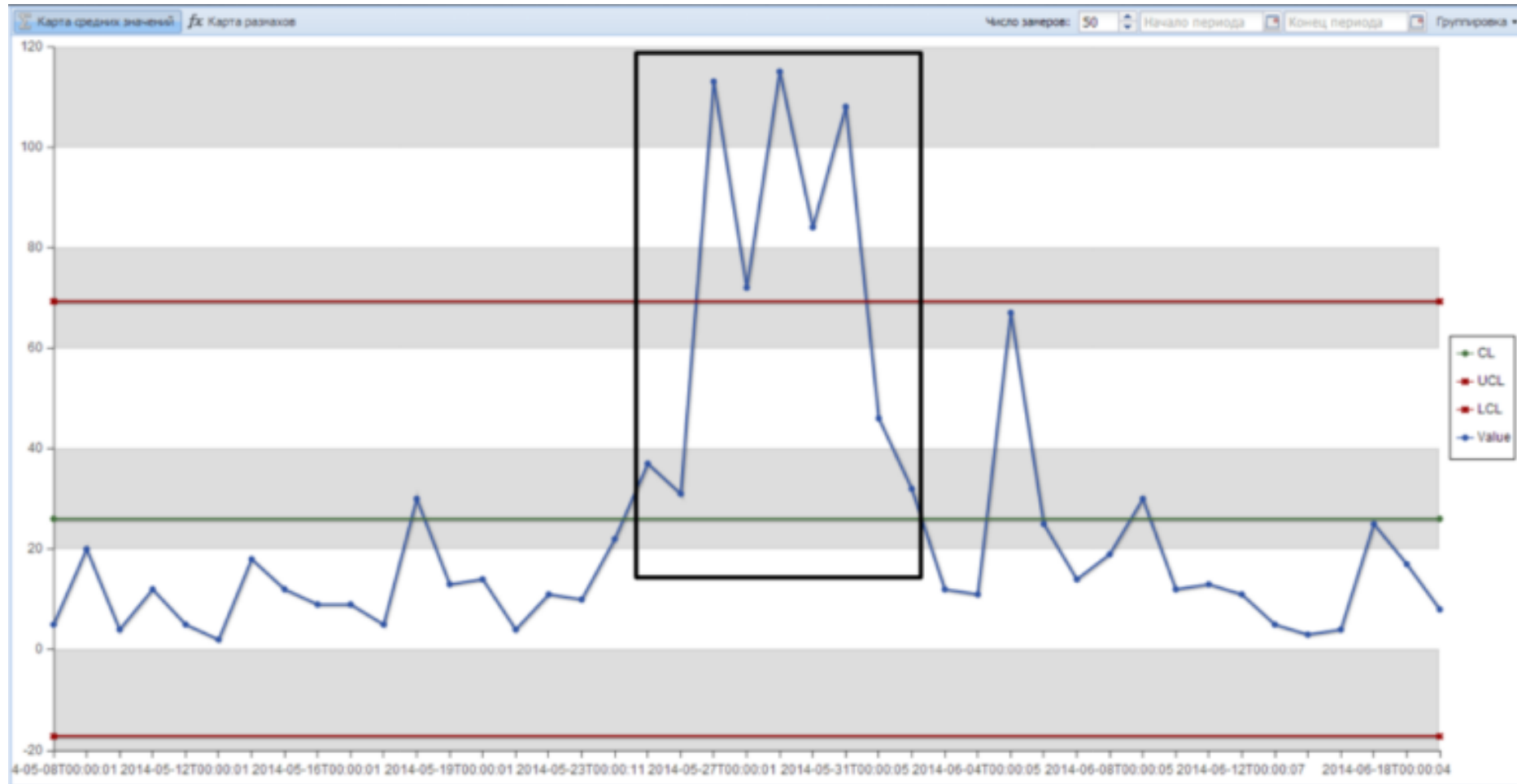
Где искать проблему?

Карта, которая измеряет среднюю скорость запросов к базе данных – если время запросов статистически стабильно, то проблема не в коде!

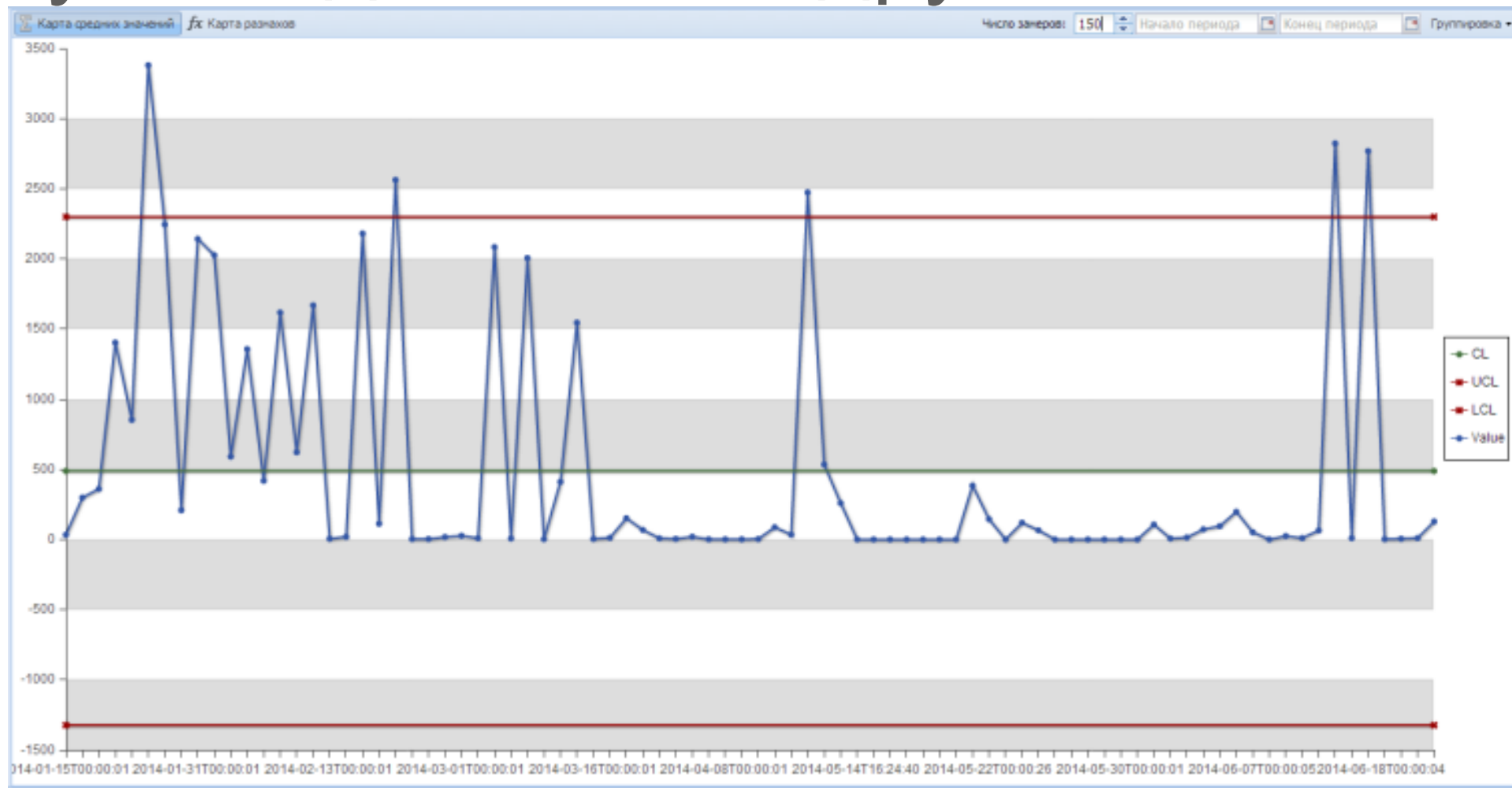
Результаты

- 1 ошибка на 10 тыс. операций

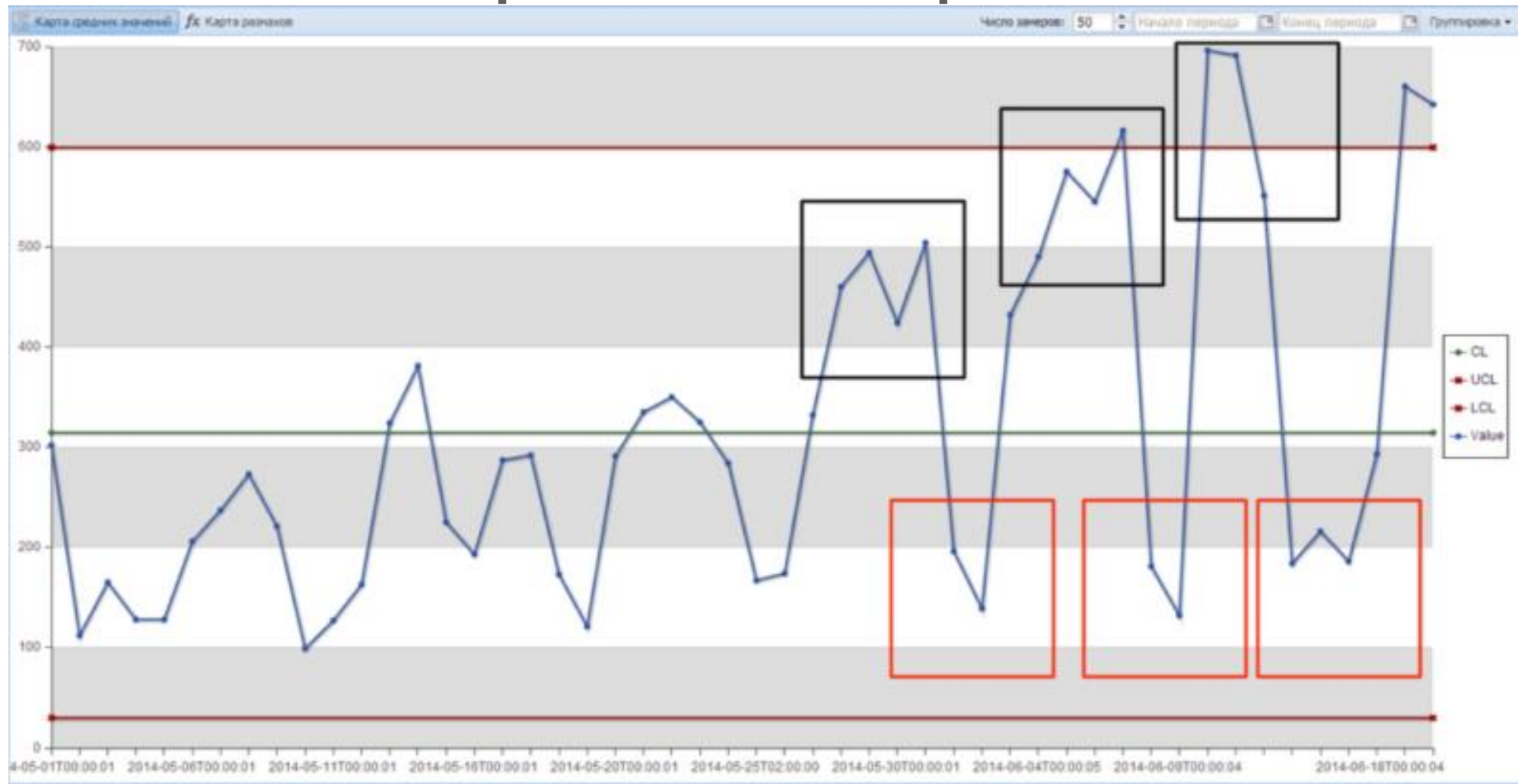
Обмен сообщениями м/у пользователями



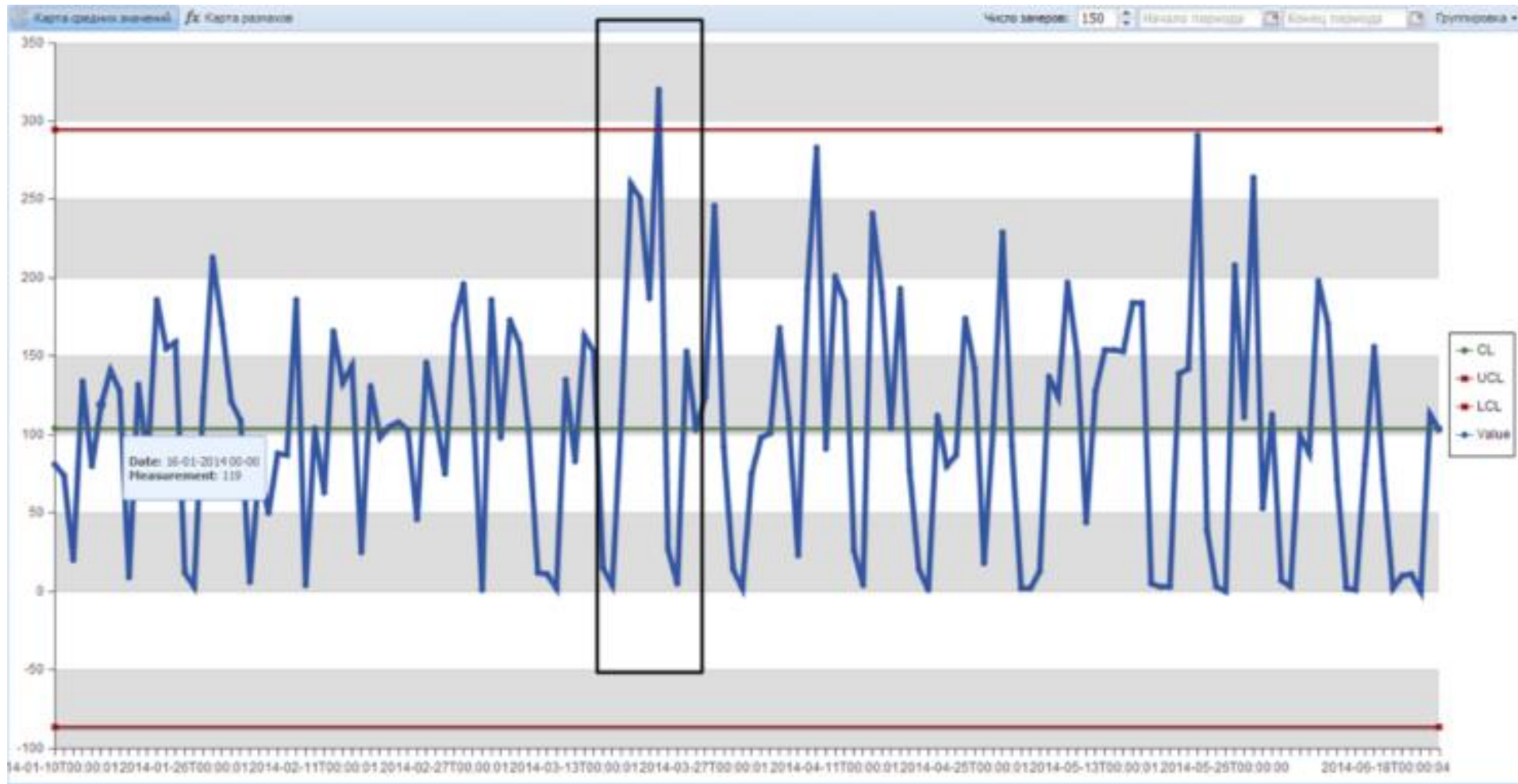
Аргумент для коллег других компаний



Докажи, что тролль – хороший!



Реклама не работает



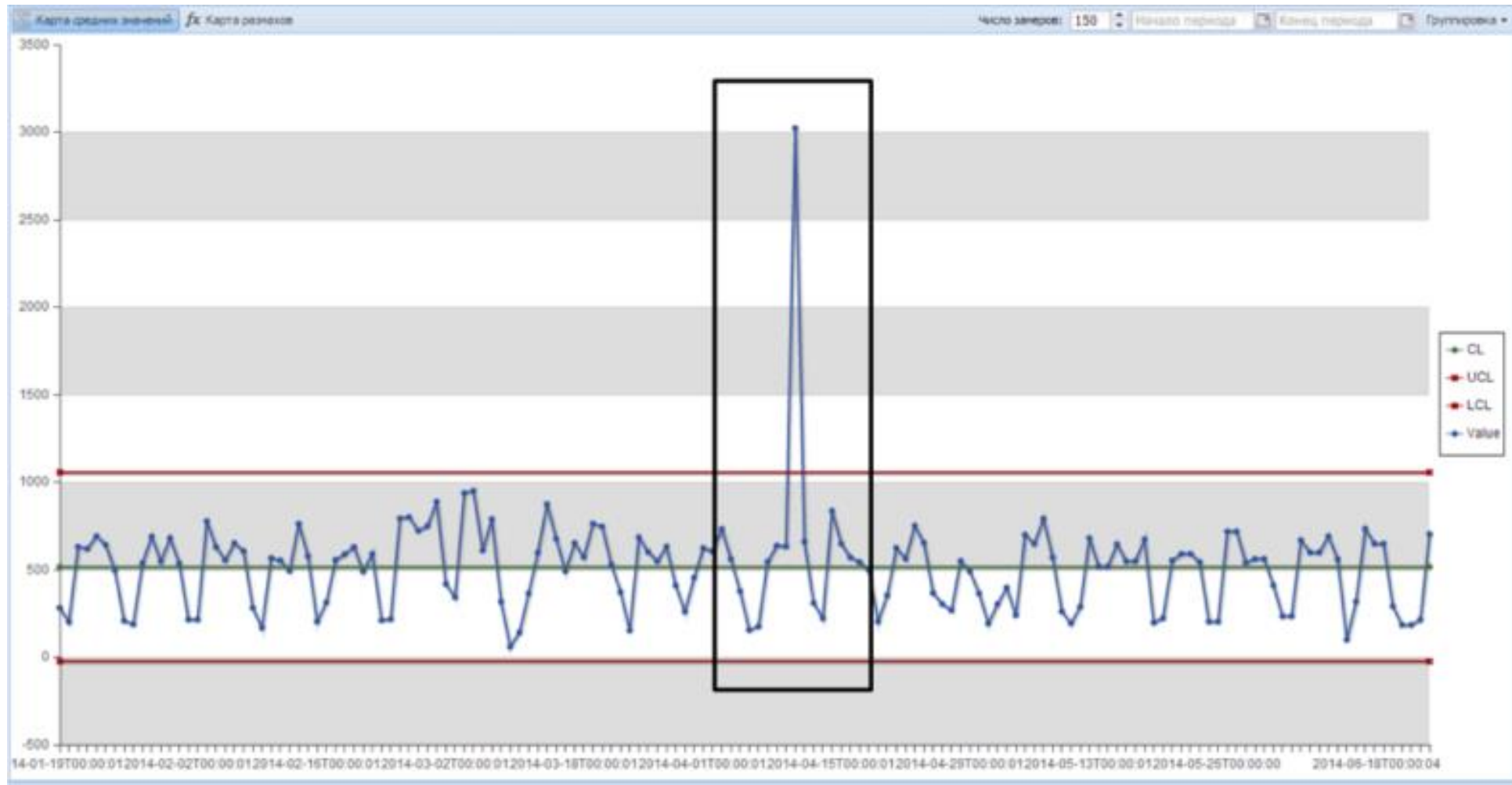
Гипотеза не работает



Гипотеза не работает



Точно? А если в недвижимости?



#4 Результаты?

Результаты

- 400-800 мс – время отклика страницы

Результаты

- 400-800 мс – время отклика страницы
- 3 сервера вместо 15

Результаты

- 400-800 мс – время отклика страницы
- 3 сервера вместо 15
- Модерация объявлений с 24 часов до 30 сек

О формализации процесса разработки
программного кода

[http://links.hwdtech.com/formal-definition-of-
software-development-process](http://links.hwdtech.com/formal-definition-of-software-development-process)

О формальном определении абстракции

<http://links.hwdtech.com/2018-abstraction-def>

Правила хорошего кода: вредные советы

https://vk.com/good_code_rules

Введение в теоритические основы
акторной модели

~август

#5 Вопросы?