

РАЗРАБОТКА ТЕХНОЛОГИИ И ПРОГРАММНОЙ СИСТЕМЫ АВТОМАТИЗИРОВАННОГО СБОРА И ИНТЕГРАЦИИ ГЕТЕРОГЕННЫХ ДАННЫХ

Специальность: 05.13.11 – “Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей”

Павликов Максим Константинович

Актуальность работы

Современные высоконагруженные веб-приложения строятся на микросервисной архитектуре, где количество микросервисов может превышать несколько сотен.

Микросервис - приложение, которое предоставляет достаточно небольшой набор функций. Каждый микросервис эластичный, легко меняющийся, элементарный и при этом законченный.

Микросервисные архитектуры используются в таких крупных проектах как ВКонтакте, Авито, 2ГИС, Badoo, HeadHunter.

Актуальность работы #2

Каждый микросервис может быть реализован, используя различные технологии, фреймворки и подходы к разработке программного обеспечения

Между микросервисами необходим обмен данными, который:

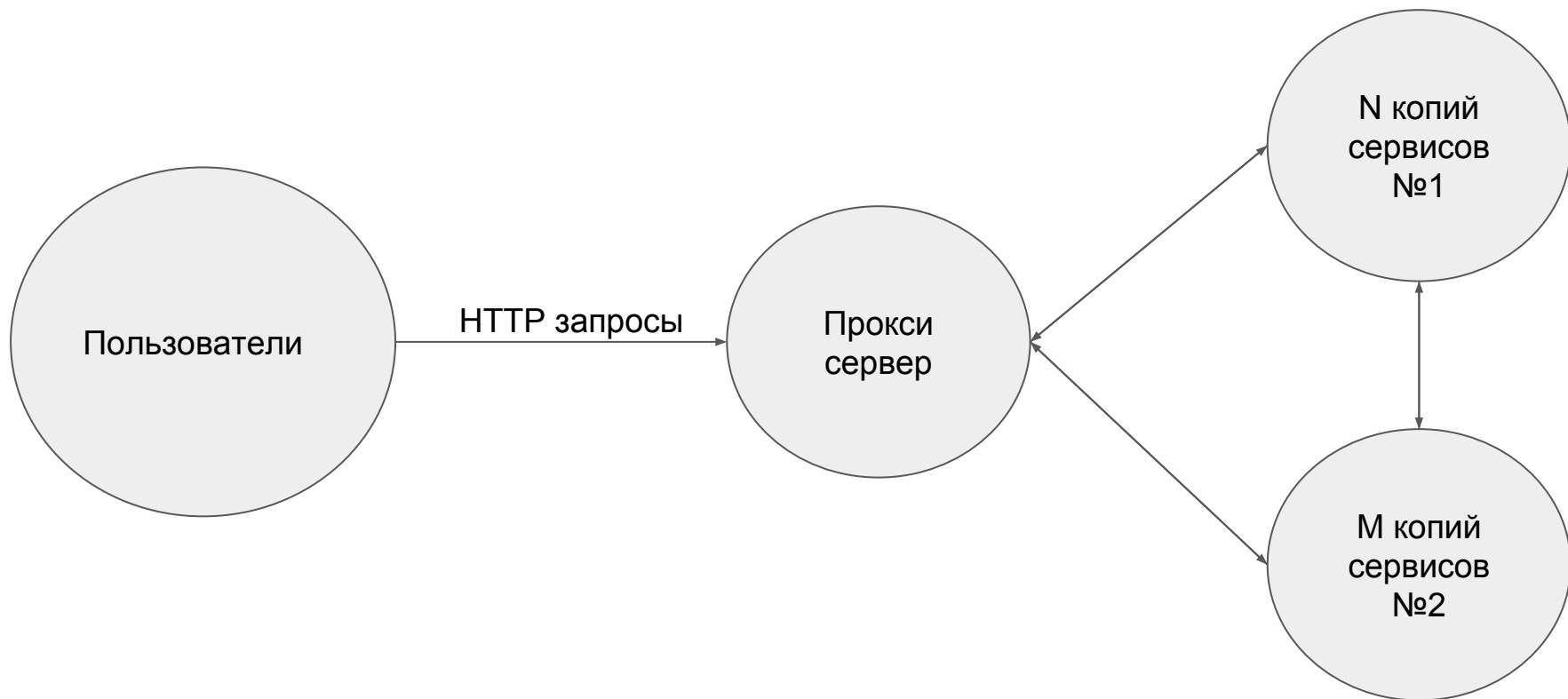
- не будет создавать дополнительной нагрузки на канал передачи данных
- легко адаптируем к добавлению новых микросервисов и изменению архитектуры существующих

Новизна

Основные научные результаты работы таковы:

- Предложен новый метод обмена данными, который позволяет автоматизировать сбор и интеграцию гетерогенных данных.
- Предложена спецификация протокола, в которой изложены основные положения по формату обмена данными, схемам представления и получения данных, а также порядок осуществления обмена.
- Создана программная реализация протокола, позволяющая установить автоматизированный сбор и интеграцию данными между произвольным числом программных систем.

Микросервисная архитектура



Технологии обмена данными - SOAP

SOAP - протокол для обмена структурированными сообщениями в распределённой вычислительной среде, используя XML в качестве формата преобразования данных

Особенности:

- Жесткий стандарт, не возникает никаких разногласий
- Избыточен из-за XML и создает нагрузку на канал передачи данных
- Трудности масштабирования
- Нет полной официальной поддержки современных скриптовых языков программирования таких как Go, Scala

Технологии обмена данными - REST

Архитектурный подход от Роя Филдинга, который описан в его докторской диссертации (<https://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>)

- не является протоколом, нет никаких стандартов (каждый трактует в силу собственных возможностей, что рождает разногласия и различные подходы к использованию технологии)
- никаких дополнительных нагрузок на канал из-за отсутствия ограничений на формат преобразования данных
- поставщик не должен подстраиваться под задачи клиента
- отсутствует концепция “wsdl” с описанием сервиса

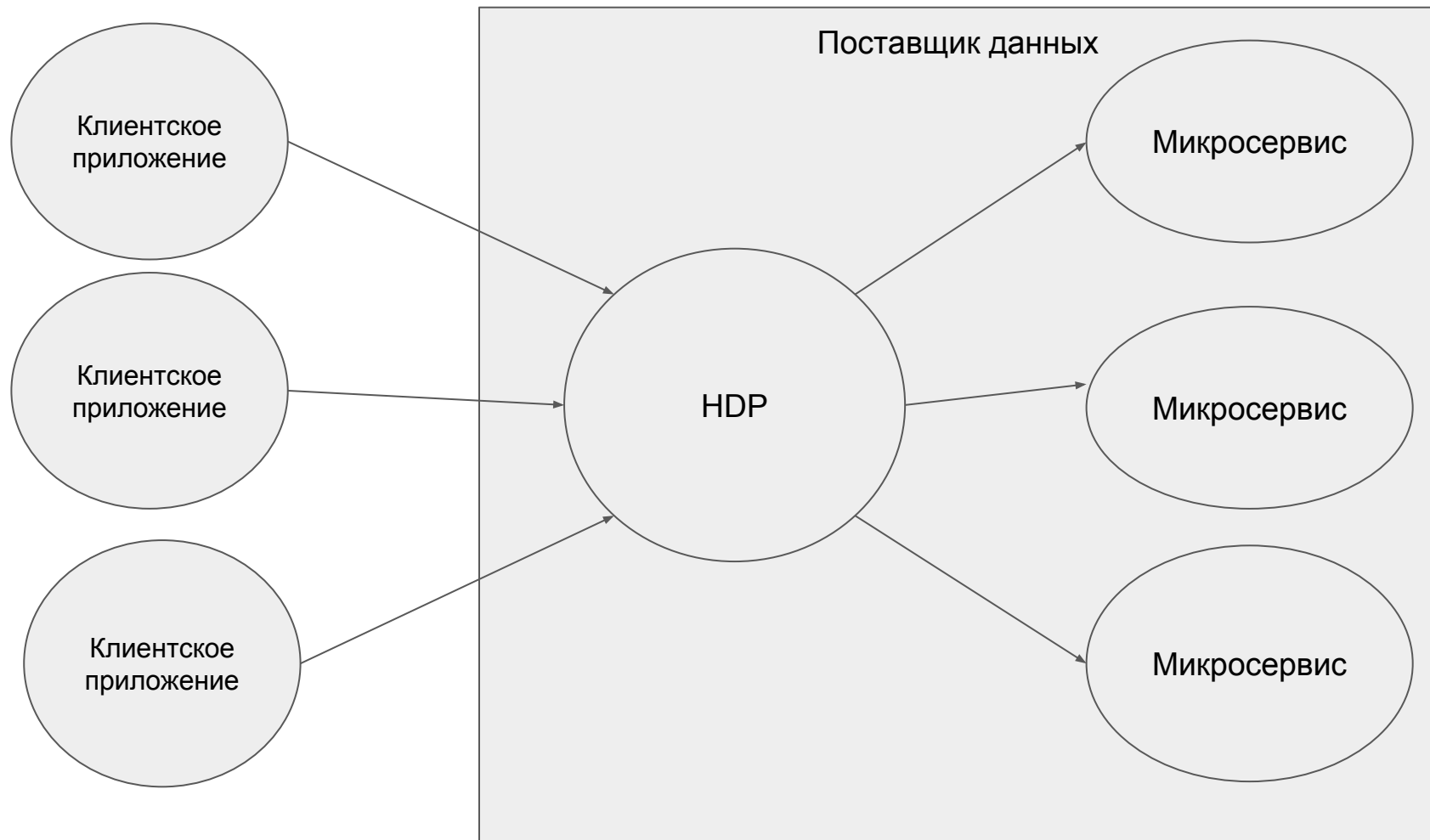
Протокол HDP (Heterogeneous Data Protocol)

Протокол для обмена гетерогенными данными в распределенной вычислительной среде.

При обмене данными по протоколу HDP все взаимодействующие узлы делятся на 2 типа - провайдер и клиент.

Провайдер - это узел, предоставляющий доступ к своим функциональным сервисам, которые в свою очередь предоставляют некие данные.

Клиент - это узел, запрашивающий данные доступные у провайдера.



Протокол HDP – сторона провайдера

Провайдер размещает:

- микросервисы в нескольких копиях
- входную точку перед своими микросервисами HDP Server

Провайдер настраивает конфигурационный файл, где указывает:

- предоставляемые данными, функции и параметры с указанием валидации
- внутренние адреса для доступа к микросервисам и их копиям
- указание балансировки нагрузки между копиями микросервисов
- настройки кеширования
- настройки запросов к прежним версиям сервисом

Этапы обмена по HDP

Этапы обмена данными по протоколу HDP:

1. Клиент запрашивает конфигурационный файл провайдера, в котором перечислены все доступные функции, возвращающие данные.
2. Клиент запрашивает данные провайдера согласно полученной на 1 этапе информации о доступных функциях: отправляет запрос провайдеру.
3. HDP Server поставщика получив запрос выполняет необходимые проверки, получает данные и возвращает их клиенту.

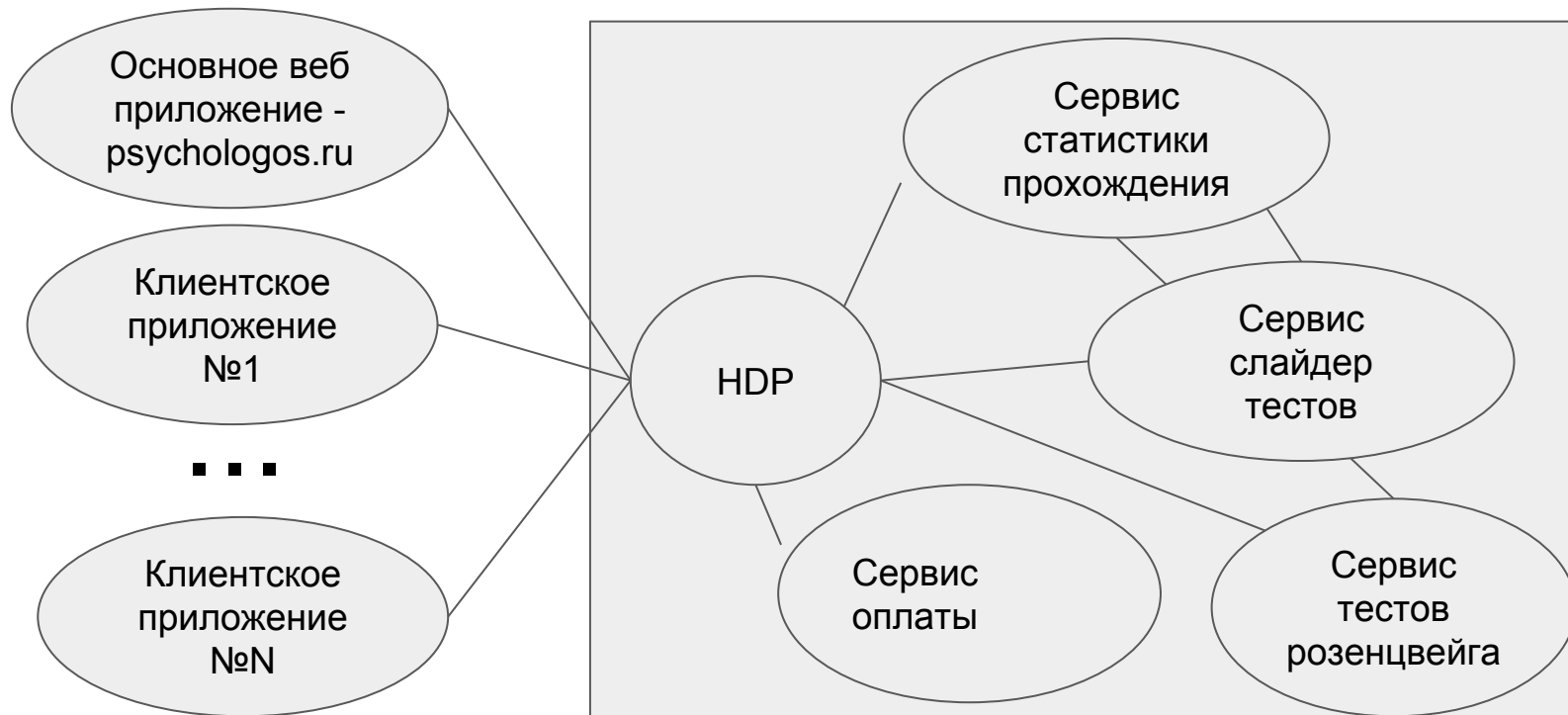
Примеры использования HDP

Микросервисы тестов (<http://psychologos.ru>)

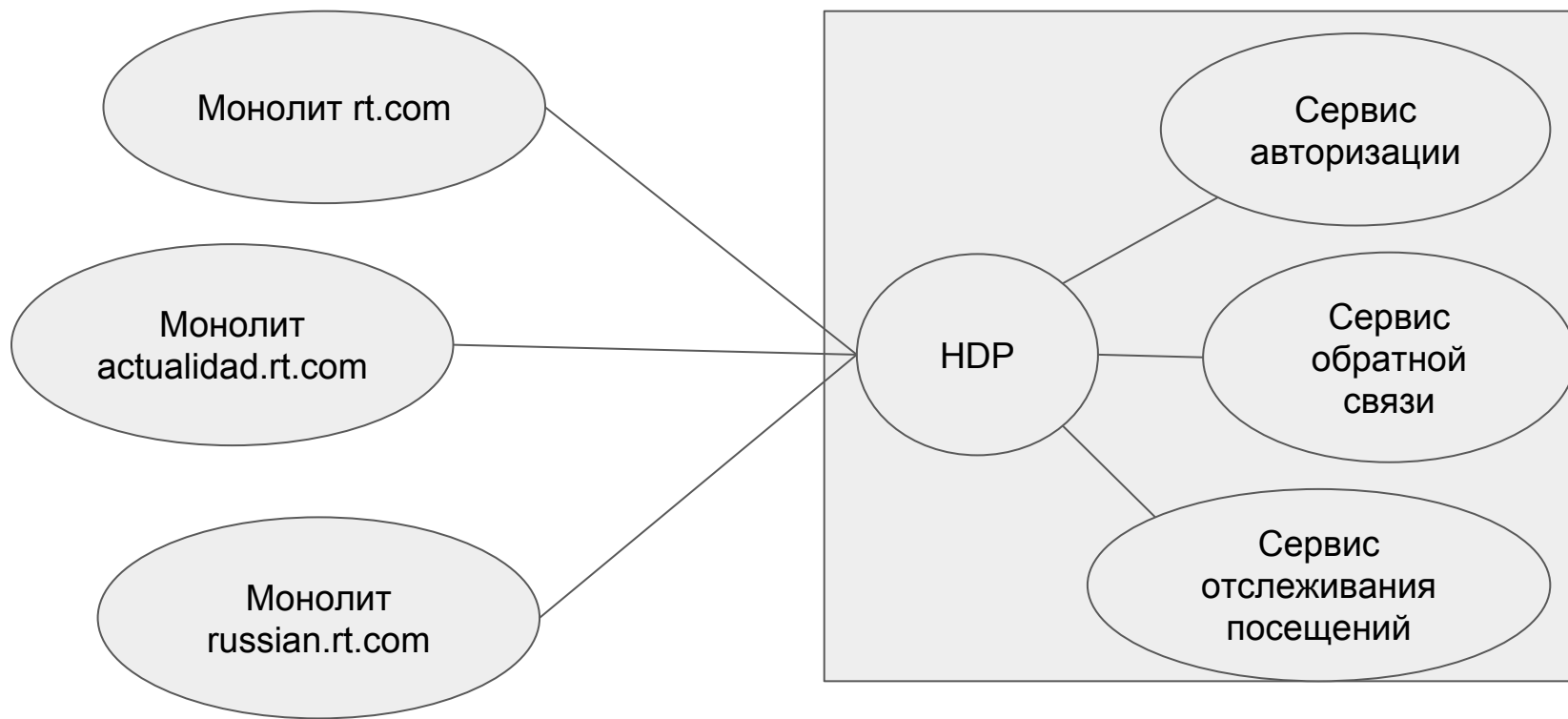
Микросервисы отслеживания посещений, авторизации, форм обратной связи, API мобильных приложений, API Telegram (<http://rt.com>)

Микросервисы скрининга ПСА и характеристик ДГПЖ - ЕМИАС (в разработке)

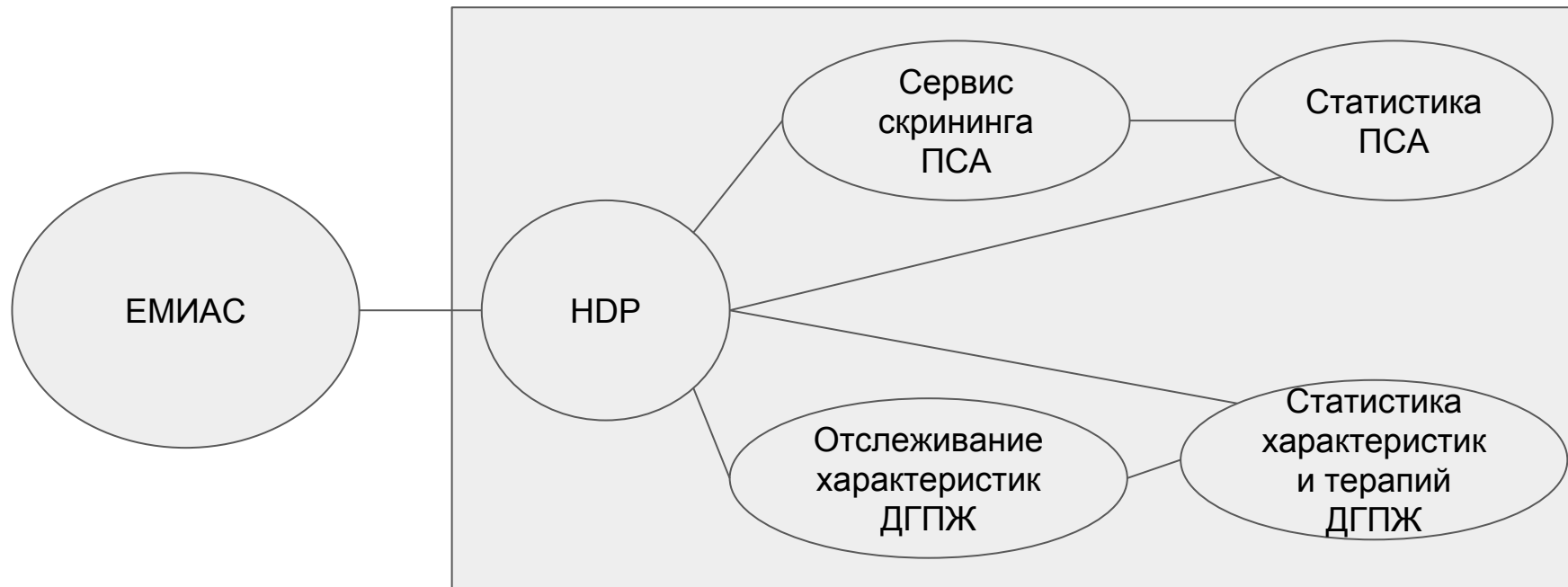
Микросервисы тестов (<http://psychologos.ru>)



Микросервисы трекинга посещений, авторизации, форм обратной связи



Микросевисы скрининга ПСА и характеристик ДГПЖ - ЕМИАС (в разработке)



REST vs SOAP vs HDP

	HDP	REST	SOAP
Наличие спецификации	Есть строгая спецификация	Есть только общее описание концепции	Есть строгая спецификация
Использование произвольного формата преобразования данных	Допустимо использовать любые форматы	Допустимо использовать любые форматы	Согласно спецификации допустим только XML
Экосистема и наличие библиотек под разные языки программирования	Хорошая; существуют реализации под большинство популярных языков программирования	Отличная; библиотеки доступны почти под все языки программирования	Удовлетворительная; для новых языков программирования официальных библиотек на данный момент нет

REST vs SOAP vs HDP #2

Сложность развертывания системы	Низкая; помимо сервисов достаточно развернуть лишь узел HDP Server	Низкая; необходимо разворачивать исключительно сами сервисы	Средняя; SOAP Server необходимо разворачивать под каждый отдельный сервис
Масштабируемость	Доступна (изменение конфигурационного файла)	Встроенной возможности масштабирования нет	Встроенной возможности масштабирования нет
Единая точка входа	Существует	Не существует	Существует
Накладные расходы (без учета самих данных)	Низкие	Низкие	Высокие из-за ограничения на формат
Обратная совместимость при изменении сервиса	Возможность существует	Возможность отсутствует	Возможность отсутствует

Взаимодействие веб приложения RT с микросервисом отслеживания посещений - результаты по скорости

Задачи микросервиса регистрация событий (посещение статьи) и выдача отчетов.

Время ответа от микросервиса на агрегацию данных в секундах; #n - номер запроса

	#1	#2	#3	#4	#5	#6	#10
SOAP	0.25	0.37	0.4	0.2	0.33	0.48	0.35
JSON	0.021	0.033	0.02	0.017	0.031	0.04	0.038
HDP	0.025	0.003	0.002	0.003	0.006	0.003	0.003

Программная реализация

HDP Server - сервис на Node.js - запатентован. Развертывание:

- установить Node.js и npm
- клонировать git репозиторий HDP Server и установить зависимости npm
`install --production`
- изменить конфигурационный файл согласно данным микросервисов:
`nano config/hdp.json`
- запустить `node index.js`

Клиентские библиотеки реализованы для языков программирования Python, PHP, Ruby, Go, Scala, Dart, Perl

Публикации ВАК

1. Павликов М.К. Технология HDP и программная система автоматизированной интеграции и сбора гетерогенных данных // "XXI век: итоги прошлого и проблемы настоящего плюс": Периодическое научное издание. - Пенза: Изд-во Пенз. гос. технол. ун-та, 2016. - №03(31). - 259 с. 2.

http://www.penzgtu.ru/fileadmin/filemounts/science/xxi_pub/2016_31.pdf

2. Павликов М.К. Протокол HDP // Научно-технический и производственный журнал "Вестник компьютерных и информационных технологий 8-2016". Москва: Изд-во "Спектр", 2016 - 56 стр.

<http://www.vkit.ru/index.php/current-issue/518-052-056>

3. Павликов М.К. HDP Server // Свидетельство о государственной регистрации программ для ЭВМ № 2016619454 от 19.08.2016

Публикации в других изданиях

4. Павликов М.К. Микросервисная архитектура веб-приложения на основе технологии HDP // Сборник научных трудов "Современные информационные технологии в управлении и образовании". Часть 2. В 3-х ч. - Москва: ФГБУ НИИ "Восход", 2016 - стр. 134 - 139.

5. Павликов М.К. Применение программной системы HDP в качестве прокси-сервера // Сборник материалов и докладов II Всероссийской научно-практической конференции (с международным участием) "Информационные технологии в прикладных исследованиях". Екатеринбург: Изд-во ИОЦ "Инфометод", 2016 - стр. 113-116.

6. Павликов М.К. Информационная система сбора и интеграции гетерогенных данных //Тезисы докладов XIX международной научной конференции "Системный анализ, управление и навигация", Анапа, 29 июня - 6 июля 2014 г.

Публикации в других изданиях #2

7. Павликов М.К. Технология автоматизированного сбора и интеграции гетерогенных данных для мониторинга учебного процесса //Материалы VIII международной научно-практической конференции “Новые информационные технологии в образовании”, Екатеринбург, 10-13 марта 2015г.

8. Павликов М.К. Технология автоматизированного сбора и интеграции гетерогенных данных //Тезисы докладов XIX международной конференции по вычислительной механике и современным прикладным программным системам, Алушта, 24-31 мая 2015 г.

9. Павликов М.К. Технология автоматизированного сбора и интеграции гетерогенных данных для взаимодействия программных систем различных аэропортов //Тезисы докладов XIX международной конференции "Авиация и космонавтика - 2015", Москва, 16-20 ноября 2015 г.

Спасибо за внимание