



ПОЛИТЕХ
Санкт-Петербургский
политехнический университет
Петра Великого

Санкт-Петербургский политехнический университет Петра Великого
Институт компьютерных наук и технологий
Высшая школа программной инженерии

Программный аналитический комплекс для образовательной платформы «Открытое образование»

Состав проекта:

Перескокова А.А., Кольцов А., Барсуков Н., Кочугова В., Сысоев И.

Руководитель:

к.т.н., доцент
Никифоров И.В.

Заинтересованное лицо (business owner):

Толпыгин С.

Существует множество образовательных платформ, но наиболее популярные из них в РФ:

- «*Открытое образование*» - современная образовательная платформа, предлагающая онлайн-курсы по базовым дисциплинам, изучаемым в российских университетах
- Moodle
- DonorsChoose

Все образовательные платформы имеют возможность **сохранять активность (действия) пользователей на сайте** во время прохождения обучения на курсе или выполнения тестовых и экзаменационных заданий (**ведение аудита** действий пользователя)

Но далеко не все образовательные платформы предоставляют возможность **анализировать активность пользователей** и преподавателей с целью совершенствования контента курса, ресурса и платформы.

Платформа «Открытое образование» одна из наиболее активно используемых систем образования всей России, в том числе СПбПУ, и к сожалению, имеет крайне ограниченный набор аналитических возможностей.

Поэтому актуальной является задача разработки программного аналитического комплекса для платформы «Открытое Образование»

- Платформа позволяет
 - Логгировать каждое действие пользователя (студента, преподавателя, администратора).
 - Формат логирования JSON.
 - Возможность вручную скачать аудит в виде лог-файла
 - Необходим специальный доступ. Не каждому пользователю доступно такое действие.

Сопутствующие ограничения:

- Большой объем файлов лога (например, 18Гб, 100 000 000 записей)
 - невозможность человеку вручную проанализировать лог файлы и собрать статистику
 - открыть файл для просмотра стандартными программными средствами
- Наличие «пустых», неинформативных строк
- Наличие более 200 разных типов событий пользователей

Необходимо предоставить возможность анализа логов за счет создания программного средства объединяющего в себе выгрузку, хранение и обработку

Цель: предоставить возможность проведения анализа логов портала «Открытое образование» за счет создания программного аналитического средства, реализующего в себе выгрузку, хранение, обработку и отображение результатов анализа.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Исследовать структуру и формат логов, предоставляемых образовательной платформой «Открытое образование».
2. Развернуть тестовый стенд с API платформы для сбора тестовых данных, реализации программного взаимодействия и проведения экспериментов.
3. Предложить программную структуру программного средства, позволяющего анализировать логи.
4. Реализовать программное средство (удобное для конечного пользователя).
5. Продемонстрировать работу приложения.

- Каждый пользователь имеет уникальный идентификатор
- Каждое событие имеет свой тип
 - Типы строго задокументированы платформой, например
 - `edx.course.enrollment.activated` - событие, при котором студент записался на курс;
 - `edx.course.enrollment.deactivated` - событие, при котором студент вышел из курса
 - Каждая строка лога представляет собой отдельное событие
 - Вложенные структуры различны в зависимости от события
- Обработывая и комбинируя анализ различных событий можно получать «интересные» результаты

Пример JSON'а события:

```

1  {
2    "username": "AAAAAAAAA",
3    "event_source": "server",
4    "name": "edx.course.enrollment.deactivated",
5    "referer": "http://localhost:8001/container/vi
6    "accept_language": "en-US,en;q=0.8",
7    "time": "2014-01-26T00:28:28.388782+00:00",
8    "agent": "Mozilla/5.0 (Windows NT 6.1; WOW64; Tr
9    "page": null
10   "host": "courses.edx.org",
11   "session": "a14j3ifhskngw0gfgn230g",
12   "context": {
13     "user_id": 9999999,
14     "org_id": "edX",
15     "course_id": "edX\\DemoX\\Demo_Course",
16     "path": "\\change_enrollment",
17   },
18   "ip": "NN.NN.NNN.NNN",
19   "event": {
20     "course_id": "edX\\DemoX\\Demo_Course",
21     "user_id": 9999999,
22     "mode": "honor"
23   },
24   "event_type": "edx.course.enrollment.deactivated"
25 }
    
```

`edx.cohort.user_removed`

When a course team member selects **Instructor** in the LMS to change the cohort assignment of a learner on the instructor dashboard, the server emits an `edx.cohort.user_removed` event.

History: Added 7 Oct 2014.

Event Source: Server

event Member Fields:

Field	Type	Details
<code>cohort_id</code>	number	The numeric ID of the cohort.
<code>cohort_name</code>	string	The display name of the cohort.
<code>user_id</code>	number	The numeric ID (from <code>auth_user.id</code>) of the removed user.

1. Анализ активности пользователя и преподавателя курса
2. Карта активности на курсе
 - Подразумевается как пользователь перемещался по разделам курса.
 - По каждому пользователю должен быть построен граф перемещения по разделам. В узлах графа раздел, переходы между узлами - перемещения. Переходы должны быть подписаны порядковым номером перехода.
 - Итоговый вид - табличка в csv, которая потом может быть подана на блок визуализации и построения графа.
3. Определение махинаций студентов
 - Например, время нахождения на тесте 1 мин, или время просмотра всех видеолекций 1 мин, или наоборот человек уснул на просмотре видео.
4. Прогнозирование поведения студента на курсе.
 - Сможет ли этот пользователь завершит курс или он на него зарегистрировался просто так.
5. Определение наиболее интересного материала курса и материала курса, который стоит переработать, доработать
 - Наиболее плохие вопросы в тестах, по которым у пользователей возникает больше всего вопросов.
6. Сравнение результатов запуска курса из года в год

- Представление активности пользователя на сайте в виде сводной многомерной таблицы позволяет оценивать зависимости показателей и строить выборки по срезам
- Таблица активности преподавателя отличается от таблицы активности студента

Идентификатор пользователя	Разделы курса						Тест					Экзамен		
	Название раздела			Название раздела			№ вопроса		№ вопроса					
	Время проведенное в разделе (мин)	Кол-во подходов просмотра видео, шт	Кол-во открытий текстового материала								
test_user_name														



Запуск
OpenAnalysis.exe

Ввод данных для программы:

- Ввод данных адреса OpenEdu (файл настроек)
- имя пользователя и пароль для доступа к логам
 - Идентификатор курса
 - Идентификатор пользователя (опционально)
- Выбор задачи анализа

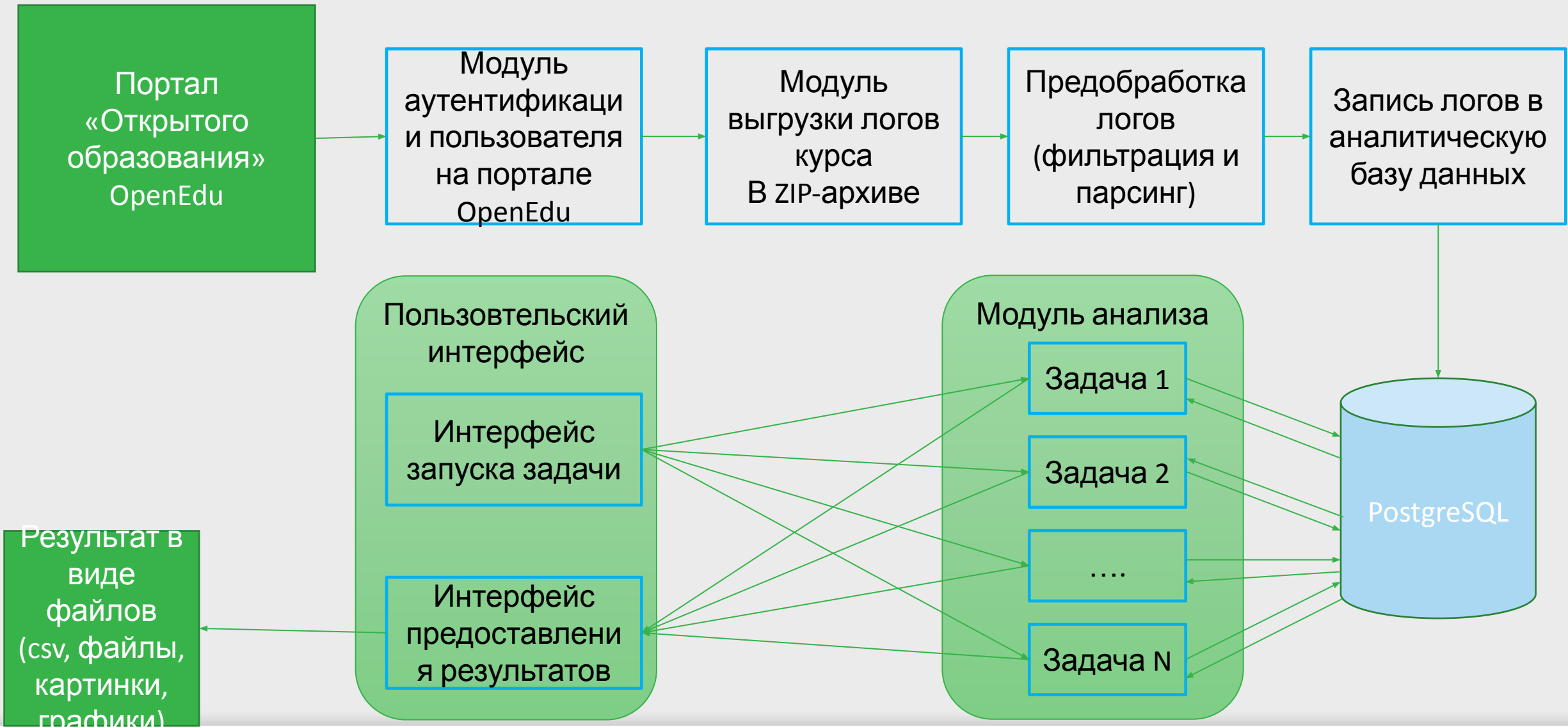
Внутренняя работа системы
незримая для конечного пользователя

Портал
«Открытого
образования»
OpenEdu

Результат
анализа
(понятный
пользовател
ю)

- Система может быть установлена
 - Локально – для демонстрации работоспособности
 - На удаленном сервере с мощными аппаратными возможностями для достижения наилучшей производительности
- Для установки системы требуется:
 - JDK 11
 - Настроенная БД PostgreSQL (лучше конечно сделать embedded BD, чтобы она незримо устанавливалась)
 - Что еще?

- Продумать **сводную табличку анализа активности** пользователей: студентов и преподавателей.
- Реализация на начальном этапе должна быть **максимально простой** и **расширяемой** (без Spark и Flink)
- Возможность **простой установки** и **использования**
- Наличие крайне **простой и интуитивной** документации
- **Наглядное** и **интуитивное** отображение результатов анализа
- Наличие тестового стенда



1. Развернуть тестовый стенд edx, доступный по внешнему IP
2. Реализовать модуль утентификации и выгрузки лога в ZIP-архиве
3. Реализовать несколько простых задач анализа
4. Реализовать модуль сохранения результатов анализа в файл csv
5. Обернуть «каркас» программного решения в конечный дистрибутив
6. Проверить работоспособность и интеграцию модулей
7. Провести демонстрацию и нагрузочное тестирование
8. Расширить и усложнить аналитические задачи
9. Расширить форматы отображения результатов
10. Применить современные программные средства для обработки: Spark и Flink

- Дистрибутив поставляется в виде ZIP-архива – openEduAnalytics.zip
- Скачать последний дистрибутив можно с репозитория
 - <https://github.com/mandarinSh/OpenPolyEdu>
- Дистрибутив self-contained и содержит все необходимое для установки

OpenEduAnalytics.zip	→	OpenEduAnalytics	
		- bin	
		- start.bat	- скрипт запуска
		- config	
		- application.yml	- основной файл настроек или любой другой формат
		- libs	
		- <list of jars>	- зависимости (для Java, для Python, и т.д.)
		- soft	
		- PostgreSQL_x.x.x.zip	- используемые ПО, чтобы пользователь никуда не лезть
		качать	
		- JDK_11.zip	
		- Python x.x.x	
		- workdir	- рабочая директория для временных файлов и БД
		- results	- директория с результатами запусков

- start.bat выполняет следующие действия
 - очищает “workdir”
 - распаковывает jdk в workdir/install_soft
 - распаковывает PostgreSQL в workdir/install_soft
 - распаковывает Python в workdir/install_soft
 - прописывает JAVA_HOME/Python/etc. для локального старта
 - осуществляет настройку PostgreSQL (копирует подготовленный заранее файл)
 - запускает сервисы PostgreSQL
 - запускает модуль аутентификации и выгрузки логов и ждет завершения его работы
 - запускает модуль фильтрации и загрузки логов в PostgreSQL и ждет завершения работы компонента
 - запускает модуль аналитики и предлагает пользователю выбрать требуемую задачу
 - генерирует файл с результатами
 - просит пользователя выбрать другую задачу и также генерирует результат
 - по завершении работы аналитического модуля осуществляется остановка сервисов PostgreSQL.
- Безде можно делать оптимизацию и смотреть на наличие уже выполненных шагов, пропуская, например, установку или выгрузку. Все это очень легко настраивается передаваемыми опциями bat

Демонстрация

Спасибо за внимание!