**САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ**

**УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ,   
МЕХАНИКИ И ОПТИКИ**

Факультет информационных технологий и программирования

Кафедра информационных систем

Направление подготовки (специальность) 09.04.02

**О Т Ч Е Т**

**о производственной практике**

(наименование практики)

**Тема задания: Улучшение и внесение изменений в высоконагруженную платформу по хранению, обработке и предоставлению информации с возможностью детальной настройки правил обработки входных и выходных данных**

**Студент \_Трофимов В.А.\_ \_\_\_\_\_\_\_М4106\_\_\_\_\_\_\_**

(Фамилия И.О.) номер группы

**Руководитель практики от организации:** Гедзберг М.Ю., ООО "Люксофт Профешнл", старший

руководитель программы

(Фамилия И.О., должность и место работы)

**Ответственный за практику от университета:** Зубок Д.А., доц. каф. ИС

(Фамилия И.О., должность)

**Практика пройдена с оценкой \_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Подписи членов комиссии**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_( )**

(подпись)

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_( )**

(подпись)

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_( )**

(подпись)

**Дата \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Санкт-Петербург**

**2016**

Задачей практики является улучшение и внесение изменений в высоконагруженную платформу по хранению, обработке и предоставлению информации с возможностью детальной настройки правил обработки входных и выходных данных.

Задачи на период практики:

* Оптимизация использования платформой пространств таблиц в реляционной СУБД Oracle
* Модификация и написание утилит для автоматизации развертывания платформы на тестовые окружения
* Улучшение стабильности работы платформы при нестабильной работе внешних используемых компонент для авторизации, аутентификации и авторизации
* Написание новых правил обработки данных для нового формата входных данных

## **Описание рассматриваемой платформы**

Платформа представляет из себя высоконагруженное распределенное приложение, организующее автоматический забор данных из заранее сконфигурированных источников, проверку соответствию формата данных, трансформацию и их сохранение во внутреннюю базу данных и высокопроизводительный распределенный кэш в структурированном формате, позволяющем эффективно осуществлять поиск по хранимой информации. Впоследствии данные отдаются конечным пользователям через REST-интерфейс.

На данный момент поддерживается забор данных из CSV файлов и из реляционных баз данных через JDBC соединение.

Приложение является полностью распределенным — для изменения размера кластера необходимо указать в файлах настроек соответствующего узла информацию о всех узлах кластера.

Основные слои и компоненты платформы изображены на рисунке.

## 

Основные слои и компоненты:

* External Feeds – внешние источники данных, из которых платформа выкачивает данные по расписанию
* External Services – внешние системы для обеспечения идентификации, аутентификации и авторизации пользователей в системе
* Data Processing Service – управление забором данных из внешних источников и ее обработка
* Configuration Management Services – хранит информацию о забираемой и отдаваемой инофрмации, настройках соединений с источниками, расписаниях и правилах обработки данных
* Rule Engine – распределенный сервис по выполнению правил над поступающей информацией
* REST Services – предоставляет внешним пользователям доступ к хранимым данным
* Data Layer – слой хранения данных и настроек системы

Основные технологии и продукты, используемые в платформе:

* Hazelcast: распределенный кэш; предоставляет масштабируемый, высоконагруженный и отказоустойчивый доступ к данным
* Drools: ядро выполнения правил по обработке данных, позволяет с помощью высокоуровневых операций описывать принципы, по которым будут обработаны входные и выходные данные
* Spring: IoC контейнер
* Camel: координатор; позволяет управлять потоком данных и управления декларативно, используя язык разметки xml
* Tomcat: контейнер для веб-приложений
* JMX: Протокол управления и мониторинга для Java
* JDBC: протокол взаимодействия с БД
* Oracle DB: реляционная СУБД

## Работы, выполненнные в рамках практики

В рамках практики были выполнены следующие работы:

* Оптимизация использования платформой пространств таблиц в реляционной СУБД Oracle
* Модификация и написание утилит для автоматизации развертывания платформы на тестовые окружения
* Улучшение стабильности работы платформы при нестабильной работе внешних используемых компонент для авторизации, аутентификации и авторизации
* Написание новых правил обработки данных для нового формата входных данных

## Оптимизация использования платформой пространств таблиц в реляционной СУБД Oracle

Используемая СУБД имеет несколько пространств таблиц, обладающих различными параметрами производительности. До внесения изменений все сущности БД хранились в одном пространстве таблиц, который обладает наибольшим объемом и средними показателями производительности. Так как платформой подразумевается периодическое активное чтение и запись в БД индексируемых данных, было принято решение перенести индексируемые сущности в более высокопроизводительное пространство таблиц, обладающее лучшими характеристиками. Для автоматизации данного процесса были модифицированы миграционные скрипты на языке PL/SQL, которые осуществляют развертывание платформы «с нуля», а также процедуры на языке PL/SQL для оптимизации уже развернутых платформ.

## Модификация и написание утилит для автоматизации развертывания платформы на тестовые окружения

Частью процесса развертывания платформы является применение миграционных скриптов, которые приводят схему данных в актуальное состояние. В процессе написания данных скриптов легко допустить ошибку, которая приводит данные в неконсистентное состояние, что впоследствии крайне тяжело обнаружить во время работы платформы. Данный вид ошибок является критичным и требует устранения на ранних этапах разработки. Данные проверки до внесения изменений производились вручную уже после развертывания платформы. Для выявления данных ошибок был написал набор скриптов на PL/SQL, исполнение которых было интегрировано в набор утилит на языке Bash и Python. Данный набор скриптов автоматически вызывается при каждом развертывании платформы и сигнализирует об обнаруженных несоответствиях.

## Улучшение стабильности работы платформы при нестабильной работе внешних используемых компонент для авторизации, аутентификации и авторизации

Используемые платформой внешние системы для идентификации, аутентификации и авторизации могут выходить из строя на неопределенные промежутки времени. Во время выхода из строя данных систем, библиотеки, предоставляемые данными сервисами, приводят к системным ошибкам и нестабильному состоянию системы. В результате изучения декомпилированного кода данных библиотек было выявлено несколько узких мест, приводящих к данным ситуациям. Найденные ошибки были исправлены с помощью исправлений в коде платформы.

## Написание новых правил обработки данных для нового формата входных данных

С появлением новых источников данных возникла необходимость конфигурирования и написания правил для обработки новых данных.

## Вывод

По результатам практики были выполнены все поставленные задачи в полном объеме.