**1+4инистерство науки и высшего образования**

**Российской Федерации**

**ФЕДЕРАЛЬНОЕ Государственное Автономное**

**образовательное учреждение высшего образования**

**«новосибирский национальный исследовательский государственный университет»**

**Факультет информационных технологий**

Кафедра cистем информатики НГУ

Направление подготовки 09.03.01 Информатика и вычислительная техника

Направленность (профиль) Компьютерные науки и системотехника

# ОТЧЕТ

**о прохождении** производственной практики, научно-исследовательской работы

**Обучающегося** Савченко Егора Владимировича **группы №** 20215**курса** 4

**Тема задания**: Релизация подсистем исполнения и алализа сложности планов SQL запросов.

**Место прохождения практики:** ФГБУН Институт систем информатики им. А.П. Ершова СО РАН (ИСИ СО РАН) 630090, Новосибирская обл., Новосибирск, пр. Академика Лаврентьева, 6

**Сроки прохождения практики:** с 18.03.2024 г. по 10.05.2024 г.

**Руководитель практики   
от профильной организации** Мигинский Денис Сергеевич \_\_\_\_\_\_\_\_\_\_\_\_

к.ф-м.н., доцент кафедры СИ ФИТ НГУ

(Ф.И.О. полностью, должность) (подпись)

**Руководитель практики от НГУ**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Ф.И.О. полностью, должность) (подпись)

**Руководитель ВКР** Мигинский Денис Сергеевич \_\_\_\_\_\_\_\_\_\_\_\_

к.ф-м.н., доцент кафедры СИ ФИТ НГУ

(Ф.И.О. полностью, должность) (подпись)

**Оценка по итогам защиты отчета:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(неудовлетворительно, удовлетворительно, хорошо, отлично)

**Отчет заслушан на заседании кафедры** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(наименование кафедры)

**протокол \_\_\_\_\_\_\_\_\_от** «\_\_\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_ 20\_\_\_\_\_г.

Новосибирск 2023

Содержание

[ОТЧЕТ 1](#__RefHeading___Toc14453_2114682792)

[1. ВВЕДЕНИЕ 3](#__RefHeading___Toc2168_2114682792)

[1.1 Цель практики 3](#__RefHeading___Toc930_3208514136)

[1.2 Актуальность темы задания 3](#__RefHeading___Toc950_3208514136)

[1.3 Предполагаемый результат 3](#__RefHeading___Toc932_3208514136)

[1.4 Описание организации 3](#__RefHeading___Toc967_3208514136)

[2. ОПИСАНИЕ ПРАКТИЧЕСКОЙ ДЕЯТЕЛЬНОСТИ 4](#__RefHeading___Toc1378_3208514136)

[2.1 Профессиональные задачи 4](#__RefHeading___Toc2259_3208514136)

[2.1.1 Разработка модуля чтения данных из промышленной реляционной базы даннных 4](#__RefHeading___Toc2244_3208514136)

[2.1.2 Разработка отображения данных в оперетивную память 4](#__RefHeading___Toc2248_3208514136)

[2.1.3 Разработка интерфейса управления данными 4](#__RefHeading___Toc2248_32085141361)

[2.1.4 Разработка операций по выборке, агрегации, композиции, фильтрации данных 5](#__RefHeading___Toc2248_320851413611)

[4. ЗАКЛЮЧЕНИЕ 6](#__RefHeading___Toc2172_2114682792)

[4.1 Результаты 6](#__RefHeading___Toc2250_32085141361)

[4.2 Выводы 6](#__RefHeading___Toc989_1255960384)

[4.3 Итог практики 6](#__RefHeading___Toc991_1255960384)

[Список литературы 7](#__RefHeading___Toc2174_2114682792)

# 1. ВВЕДЕНИЕ

Настоящий отчет представляет собой обзор результатов прохождения производственной практики, которая осуществлялась в ФГБУН Институте систем информатики им. А.П. Ершова СО РАН (ИСИ СО РАН).

## 1.1 Цель практики

Основной целью прохождения практики является реализация на языке Java подсистем исполнения и алализа сложности планов SQL запросов. Разработка моделей данных, операций над данными и интерфейсов взаимодействия между модулями системы.

## 1.2 Актуальность темы задания

Актуальность темы задания на практику исходит из технического задания на ВКР.

## 1.3 Предполагаемый результат

В качестве результата прохождения практики ожидается получить код на языке Java, который будет упрощенно иммитировать работу движка базы данных. Код должен поддержить основной функционал, описанный в техническом задании к ВКР, а именно: операции по выборке, агрегации, композиции, фильтрации реляционных данных.

## 1.4 Описание организации

ФГБУН систем информатики им. А.П. Ершова СО РАН (ИСИ СО РАН) - это научно-исследовательское учреждение, которое является одним из ведущих институтов в области информатики и вычислительной техники в России.

Институт имеет широкий спектр подразделений и научных лабораторий. Занимается разработкой и применением передовых информационных и коммуникационных технологий, решением комплексных научно-инженерных задач в области информатики, проведением исследований, разработкой программных продуктов, созданием интеллектуальных систем, разработкой алгоритмов и методов обработки данных, а также разработкой систем защиты информации и кибербезопасности.

# 2. ОПИСАНИЕ ПРАКТИЧЕСКОЙ ДЕЯТЕЛЬНОСТИ

В рамках прохождения производственной, научно-исследовательской практики я решал следующие профессиональные задачи:

## 2.1 Профессиональные задачи

### 2.1.1 Разработка модуля чтения данных из промышленной реляционной базы даннных

В первую очередь для реализации задуманных целей было необходимо прочитать данные из реляционного хранилища. Для этих целей я воспользовался существующим коннектором для подключения к системам баз данных для языка Java — JDBC. JDBC предоставляет интерфейс для чтения и модификации реляционных данных. Чтение происходит согласно конфигурируемым параметрам - создается DataSource, после чего из соответствующей схемы происходит чтенице всех таблиц и индексов этих таблиц (для предовтращения переполнения оперативной памяти предусмотрены лимиты). Прочитанные данные отображаются во внутреннюю структуру, описанную далее.

### 2.1.2 Разработка отображения данных в оперетивную память

Отображение данных — один из ключевых элементов системы. Отображение должно быть хорошо продумано для дальнейшено комфортного использования в рамках системы и не должно занимать много места относительно полезной нагрузки — объем чистых данных к объему отображения занимаемым этими данными (данные должны храниться наиболее оптимально, поскольку оперативная память ограничена). Было принято режение о выделении интерфейса «Value» - регламентирующего взаимодействие со значениями, хранящимися в отображении. Этот интерфейс имплементируют обертки над конкретными типами данных (BooleanValue, DoubleNumber, FloatNumber, IntegerNumber, LongNumber, NullValue, StringValue, TimestampValue, BigDecimalNumber). Обертки нужны для дальнейшего применения Visitor-а для валидации выражений, подстановки значений в выражения и просчитывания этих выражений.

Также был выделен основной интерфейс выражения - «Expression», который реализовывается (расширяется) следующими классами: Value, SubTable, Column, UnaryOperation, BinaryOperation, TernaryOperation, ExpressionList. Все это нужно для корректной и удобной реализации Join-ов разных видов и операции фильтрации (Where).

Все прочитанные значения из базы данных хранятся в отображении в строго типизированном виде, что позволяет использовать полиморфизм в полном объеме.

### 2.1.3 Разработка интерфейса управления данными

Помимо данных было необходимо реализовать операции над этими данными. Все операции (From, Where, Select, OrderBy, GroupBy, Offset, Limit, Distinct, InnerJoin, FullJoin, LeftJoin, RightJoin, Alias, CreateIndex, DropIndex) реализуют интерфейс «Command», обладающий единственным публичным методом, которых принимает на вход таблицу и возвращает таблицу. Таким образом, набор команд формирует таблицу (В любом наборе команд From стоит на первом месте, так так именно он создает начальную таблицу). Благодаря такому подходу можно довольно просто реализовать кеширование таблиц (нет необходимости сохранять ссылку на саму таблицу и сравнивать потом две таблицы (операция сравнения таблиц очень дорогостоящая), можно просто запомнить последовательность операций над данными, которая детерминорованно приведет к тому же результату). Кроме того, такой подход к организации управления данными позволяет относительно просто реализовать финкционал подзапросов.

### 2.1.4 Разработка операций по выборке, агрегации, композиции, фильтрации данных

Это наиболее сложная в реализации часть, так как необходимо реализовывать алгоритмы, которые будут приводить данные из одного состояния в другое точно так же как это делают промышленные базы данных. В рамках этой части были реализованы Join-ы разных видов (внутренный, полный, левый, правый) и внутренние стратегии (loop, hash, merge). Также были реализованы функции агрегации данных (Avg, Count, Identity, Max, Min, Sum), и другие важные части команд (groupBy, orderBy и др).

# 4. ЗАКЛЮЧЕНИЕ

## 4.1 Результаты

В ходе производственной практики по релизации подсистем исполнения и алализа сложности планов SQL запросов, был реализован программные код, удовлетворяющий требованиям технического задания на ВКР и поставленным целям практики.

В ходе практики я узнал о различных подходах и методах проектирования, изучил паттерн проектирования «visitor», научился декомпозировать сложные задачи на более мелкие.

## 4.2 Выводы

Отчетность по проделанной работе позволяет сделать следующие выводы:

Усвоение знаний и навыков: Практика позволила мне усвоить фундаментальные знания о предметной области, а тажке ознакомиться с ключевыми концепциями и инструментами.

Развитие компетенций: В процессе работы над проектом я улучшил свои навыки в области проектирования архитектуры систем, использования языка SQL, анализа производительности запросов. Полученные навыки являются важным активом для дальнейшей карьеры в сфере разработки программного обеспечения.

Самооценка сформированности компетенций и полученных на практике знаний и навыков является положительной. Я чувствую, что справился с поставленной задачей и сумел применить полученные знания при планировании и реализации системы.

## 4.3 Итог практики

В целом, производственная практика по релизации подсистем исполнения и алализа сложности планов SQL запросов была полезной и позволила мне расширить свой кругозор в области баз данных и проектирования систем. Я получил ценный опыт работы над реальным проектом и приобрел новые знания и навыки, которые будут полезны в моей дальнейшей карьере.

# Список литературы

1. Гарсиа-Молина Г., Ульман Д. Д., Уидом Д. Системы баз данных: Полный курс. – Вильямс, 2003. - 1083с.
2. PostgreSQL: Документация. [Электронный ресурс] //https://postgrespro.ru/: информ.- спавочный портал. URL: <https://postgrespro.ru/docs/postgresql> (дата обращения 30.10.2023)
3. Parboiled: Документация. [Электронный ресурс] <https://javadoc.io/doc/org.parboiled>: информ.-спавочный портал. URL: <https://javadoc.io/doc/org.parboiled/parboiled-java/latest/index.html> (дата обращения 17.11.2023)
4. ANTLR 4 Runtime 4.13.1: Документация. [Электронный ресурс] https://www.antlr.org/: информ.-спавочный портал. URL: <https://www.antlr.org/api/Java/index.html> (дата обращения 17.11.2023)