**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** учебной практики, ознакомительной практики

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

**Тема задания**: Изучение методов оценивания 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 25.09.2023 г. по 22.12.2023 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 Изучение методов оценивания сложности планов SQL запросов 4](#__RefHeading___Toc2244_3208514136)

[2.1.2 Анализ и сравнение методов оценивания сложности планов SQL запросов 4](#__RefHeading___Toc2248_3208514136)

[2.2 Промежуточный результат 4](#__RefHeading___Toc2259_32085141361)

[3 АРХИТЕКТУРНОЕ РЕШЕНИЕ 6](#__RefHeading___Toc1386_3208514136)

[3.1 Введение в архитектуру 6](#__RefHeading___Toc1592_3208514136)

[3.2 Архитектурные блоки 6](#__RefHeading___Toc1592_32085141361)

[3.2.1 Клиент 6](#__RefHeading___Toc1592_320851413611)

[3.2.2 Парсер 6](#__RefHeading___Toc1592_3208514136111)

[3.2.3 Дерево (граф) запроса 6](#__RefHeading___Toc769_1255960384)

[3.2.5 Внутренне представление данных 7](#__RefHeading___Toc769_1255960384112)

[3.2.4 Обработчик запроса и калькулятор сложности запроса 7](#__RefHeading___Toc769_12559603841)

[3.3 Краткое обоснование архитектурного решения 7](#__RefHeading___Toc943_1255960384)

[4. ЗАКЛЮЧЕНИЕ 8](#__RefHeading___Toc2172_2114682792)

[4.1 Результаты 8](#__RefHeading___Toc2250_32085141361)

[4.2 Выводы 8](#__RefHeading___Toc989_1255960384)

[4.3 Итог практики 8](#__RefHeading___Toc991_1255960384)

[Приложения 9](#__RefHeading___Toc651_1255960384)

[Приложение 1 9](#__RefHeading___Toc720_3946569409)

[Список литературы 10](#__RefHeading___Toc2174_2114682792)

# 1. ВВЕДЕНИЕ

Настоящий отчет представляет собой обзор результатов прохождения учебной, ознакомительной практики, которая осуществлялась в ФГБУН Институте систем информатики им. А.П. Ершова СО РАН (ИСИ СО РАН). Данная практика была организована с целью закрепления теоретических знаний, полученных в процессе изучения дисциплин, а также получения практического опыта в сфере разработки программного обеспечения. Помимо получения технического опыта, сопутствующей целью практики было улучшение коммуникационных навыков.

## 1.1 Цель практики

Основной целью прохождения практики является получение и освоение информации о методах оценивания сложности планов SQL запросов. В качестве вторичной цели можно выделить: «Формирование собственного представления об архитектурном решении поставленной задачи ВКР» (Тема ВКР: «Разработка учебной системы оценки сложности планов SQL запросов»).

## 1.2 Актуальность темы задания

Актуальность темы задания на практику исходит из необходимости наличия базы знаний по теме ВКР. Помимо получения информации, необходимо обладать экспертизой для возможности фильтрации, оценивания и преобразования полученных знаний, для дальнейшего их применения. В ходе разработки поставленной темы практики, оба вышеперечисленных утверждения выполняются.

## 1.3 Предполагаемый результат

В качестве результата прохождения практики ожидается, что студент (я) будет обладать достаточным объемом информации о методах оценивания сложности планов SQL запросов для дальнейшей разработки ВКР. Сможет обосновать выбор конкретного способа оценивания и, по возможности, представит собственное архитектурное решение поставленной задачи ВКР на основе выбранного метода.

## 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 учебной, ознакомительной практики, я решал следующие профессиональные задачи:

## 2.1 Профессиональные задачи

### 2.1.1 Изучение методов оценивания сложности планов SQL запросов

Мною были изучены различные подходы по оценке сложности запросов, включая методы, основанные на анализе структуры запроса, объеме данных и статистике использования индексов. Этот анализ помогает понять, какие факторы влияют на производительность запросов и каким образом можно улучшить их выполнение.

Метод анализа структуры запроса. Этот метод включает в себя анализ структуры SQL запроса и выделение ключевых элементов, таких как отображения, фильтры, условия и сортировка. Я использовал этот метод для определения сложности, основывыясь на внутренней структуре запроса.

Метод анализа объема данных. Этот метод включает в себя анализ объема данных, с которыми работает SQL запрос (количество записей в таблице, количество возвращаемых строк, наличие ограничений).

Метод анализа статистики использования индексов. Этот метод включает в себя анализ статистики использования индексов в SQL запросах. Я использовал этот метод для определения сложности запросов на основе статистики использования индексов (учитывая количество операций чтения и записи).

### 2.1.2 Анализ и сравнение методов оценивания сложности планов SQL запросов

Я выполнял реальные SQL запросы с использованием различных методов оценки сложности. Для этого я использовал следующие инструменты и платформы: MySQL, PostgreSQL. Это позволило мне применить теоретические знания и понять, как вышеперечисленные методы работают на практике. Также я проводил анализ полученных результатов, что помогло определить, какой метод оценки сложности лучше подходит для конкретного запроса и как можно улучшить выполнение данного запроса.

## 2.2 Промежуточный результат

После изучения области, совместно с научным руководителем было принято решение о применении проприетарного подхода в рамках решения поставленной задачи ВКР. Суть подхода заключается в прямом подсчете количества исполняемах операций, но не в настоящей базе данных, а над построенном отображением базы данных в оперативную память. Благодаря такому решению будет обеспечиваться детерминированность полученного результата все зависимости от внешних факторов (мощности конкретного пк, реализация внутреннего оптимизатора SQL запросов конкретной базы данных, внутреннего кеша и статистики базы данных), кроме того, можно будет конфигурировать вес конкретных операций исходя из типа данных и поводимых над ними операциями.

Однако есть ряд существенных сложностей и ограничений:

1. Необходимо самостоятельно реализовавать выборку данных (самостоятельно исполнять запросы, а не полагаться на базу данных);
2. Необходимо задать и реализовать обработку подмножества операций по выборке/фильтрации/агрегации данных языка SQL (учитывая индексы);
3. Невозможно обрабатывать большие объемы данных (все данные должны помещаться в оперативную память);
4. Возникает сложность в обработке и/или способе задания самого запроса (необходимо либо парсить SQL, либо предоставлять DSL с синтаксисом запоминающем SQL);

Исходя из вышеперечисленных премуществ, недостатков и ограничений в качестве результата прохождения учебной, ознакомительной практики хочу представить описание моего архитектурного решения для «Учебной системы оценки сложности планов SQL запросов».

# 3 АРХИТЕКТУРНОЕ РЕШЕНИЕ

## 3.1 Введение в архитектуру

Глобально система представлена двумя основными блоками: клиент и серсер. Взаимодействие между частями системы будет происходить посредствам Json-RPC.

Сервер представляет из себя программу на языке Java, которая будет исполнять запросы пользователя к «виртуальной» базе данных. Также задачей этой подсистемы является подсчет сложности исполнения запроса в абстрактных единицах измерения (сумма перемножений количества операций конкретного типа на вес операции этого типа). Сервер содержит максимально возможное число внутренних компонентов системы, чтобы сделать клиента как можно более «легким».

Клиент будет реализован на языке Haskell (может быть реализован на любом другом языке). Задача клиента имитировать форму задания запросов приближенную к классическому SQL синтаксису (чтобы система напоминала реальные инструменты выборки реляционных данных).

Пользователь взаимодействует только с клиентской частью системы. Подразумевается, что система будет целиком развертываться на персональном компьютере пользователя и будет включать в себя одного клиента, один сервер и настоящую базу данных.

## 3.2 Архитектурные блоки

Архитектурно система разделена на несколько функциональных блоков (см Приложение 1).

### 3.2.1 Клиент

Клиент — это программа, назначение которой формировать запросы к серверу и отрисовывать ответы. Планируется создание fluent interface для формирования запросов. Flient interface будет создавать SQL строку с допустимым подмножеством операций. С точки зрения пользователя запрос — фрагмент кода. Код транслируется в SQL строку, она обрабатывается на сервере. Взаимодействие происходит через Json-RPC. При инициализации клиент может запросить у сервера конфигурацию отображения, благодаря этому (при выборе подходящего языка программирования) можно будет реализовать клиента предоставляющего fluent interface со строго типизированной формой построения запросов.

### 3.2.2 Парсер

Парсер — часть сервера отвечающая за построение семантического дерева запроса из SQL строки. SQL формируется контекстно-зависимой грамматикой, поэтому парсер должен иметь возможность обращаться к текущему контексту исполнения. Парсер является одной из самых сложных подсистем сервера, спасает только отсутствие необходимости поддержки всего множества опереторов и токенов стандарта SQL.

### 3.2.3 Дерево (граф) запроса

Из пользовательского ввода формируется структура данных, для удобного итерирования и дальнейшей обработки запроса — дерево запроса. Оно необходимо для идентификации смысловых отношений и связей между различными элементами запроса.

### 3.2.5 Внутренне представление данных

Виртуальная база данных полностью строится в оперативной памяти. Данные берутся из инстанса реальной базы данных. Конфигурация отображения задается программно. Виртуальная база данных представляет из себя структуру «удобную для проведения выборки данных» и, по возможности, оптимизированную. Она содержит дескрипторы таблиц, индексов, а также сами реляционные записи.

### 3.2.4 Обработчик запроса и калькулятор сложности запроса

Части сервера ответственные за непосредственно исполнение запроса над виртуальной базой данных (внутренним представлением данных) и подсчет сложности запроса. Наиболее важные и сложные части системы. Важно, чтобы алгоритм исполнения запросов был максимально приближен к исполнению планов запросов в реальных базах данных. Также необходимо тщательно продумать обход структур данных, сложность и затраты памяти, чтобы на исполнение отдельного запроса уходило приемлиемое время ожидания. Подсчет сложности производится в количестве атомарных (с точки зрения создаваемой учебной системы) операций.

## 3.3 Краткое обоснование архитектурного решения

Использование клиент-серверной архитектуры позволяет разделить логику системы на пользовательскую и неизменяемую часть, обеспечивает гибкость и масштабируемость.

Использование протокола RPC в качестве протокола взаимодействия обуславливается удобством и простотой использования, а также возможностью (при необходимости) переноса сервера на другую удаленную машину.

Клиент планируется как можно более «тонким» чтобы иметь возможность относительно просто заменить его на другую реализацию (например на другом языке программирования, это важный аспект для учебной системы).

Необходимость наличия парсера исходит из предыдущего пункта (привязка к стандарту SQL, а не к конкретной реализации способствует стандартизации взаимодействия подсистем, кроме того облегчает написание клиентской части).

# 4. ЗАКЛЮЧЕНИЕ

## 4.1 Результаты

В ходе учебной практики по изучению методов оценивания сложности планов SQL запросов было проведено исследование и создано детальное описание архитектуры учебной системы по оцениванию сложности планов запросов.

В ходе практики я узнал о различных подходах и методах, что включало изучение оптимизаторов запросов, статистических методов оценки, оценки времени выполнения и других факторов, влияющих на сложность запроса.

## 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 ознакомился и освоил основные концепции и методы оценки сложности планов SQL запросов, и смог успешно применить полученные знания при планировании системы.

## 4.3 Итог практики

В целом, практика по изучению методов оценивания 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)

# Приложения

Приложение 1. Схематичное изображение архитектурного решения.

