



# АВТОМАТИЗАЦИЯ ПОДБОРА ИНДЕКСОВ ДЛЯ SQL ЗАПРОСОВ В СУБД MYSQL

## ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

Студент: Петухов Илья Сергеевич

Научный руководитель: Просуков Евгений Алексеевич

# 1. Цели и задачи работы

**Цель** - разработка инструмента для автоматизированного индексирования таблиц по заданному SQL запросу.

**Задачи:**

- изучить возможности современных СУБД и их инструментов для индексирования таблиц
- изучить правила использования индексов в выбранной СУБД
- разработать алгоритм, recommending индексы для SQL запросов
- провести эксперименты, проверяющие
  - применение рекомендуемых индексов
  - выигрыш во времени при использовании этих индексов.

Целью работы является ...

Были поставлены следующие задачи: ...

## 2. Анализ возможностей СУБД и их инструментов для индексирования

СУБД/возможность	Oracle	MS SQL Server (Azure)	PostgreSQL	MySQL
автоматический процесс создания индексов	да	да	нет	нет
автоматический процесс удаления индексов	да	да	нет	нет
отчет о рекомендуемых для создания индексов	да	да	да	нет
отчет о рекомендуемых для удаления индексов	да	да	да	да
отчет о найденных дубликатах индексов	да	да	да	да
анализ использования индекса для запроса	да	да	да	да

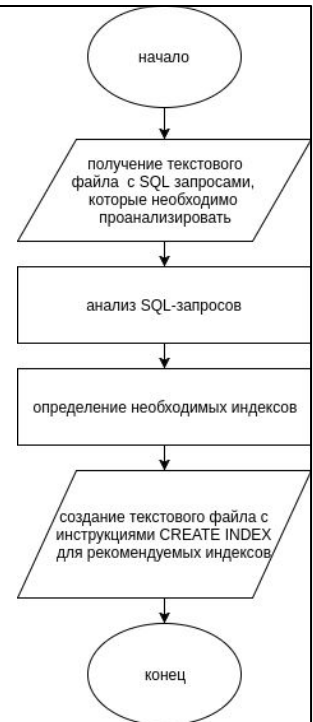
Из результатов анализа возможностей СУБД и их инструментов для индексирования видно, что в одной из самых популяр-ных СУБД MySQL отсутствует возможность получения отчетов о рекомендациях по созданию индексов, что становится причиной отсутствия в данной СУБД инстру-мента по автоматизированному управлению индексами (для создания, применения, уда-ления индексов)

Поэтому принято решение для СУБД MySQL разработать инструмент для определения индексов по заданному SQL запросу.

### 3. Аналог Index Advisor в PostgreSQL

Соответствующие требования к разрабатываемому ПО:

- поддержка Btree индексов
- считывание текстового файла с набором SQL запросов
- анализ простых и сложных запросов
- создание файла с инструкциями CREATE INDEX для создания рекомендуемых индексов



В качестве аналогичного инструмента рассмотрим Index Advisor для СУБД PostgreSQL. Процесс работы инструмента представлен на схеме.

По аналогии с этим инструментом можно предъявить следующие требования к разрабатываемому ПО: ....

Что подразумевается под сложным запросом рассмотрим на следующем слайде.

## 4. Вид сложного запроса

```
SELECT * FROM
t1 {INNER | LEFT | RIGHT} JOIN t2 ON conditional_definition
    [WHERE where_definition]
    [ORDER BY col_name [ASC | DESC]]

where_definition:
    where_expression or
    where_expression [AND] where_expression

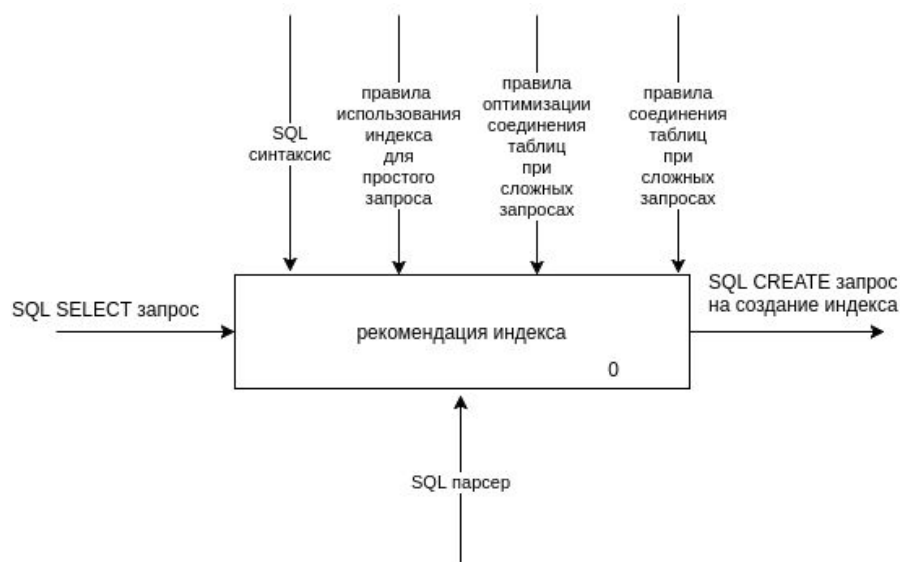
where_expression:
    column_name [> | >= | = | < | <= | < ] constant or
    column_name LIKE constant or
    where_definition

conditional_definition:
    conditional_expression or
    conditional_expression [AND] conditional_expression

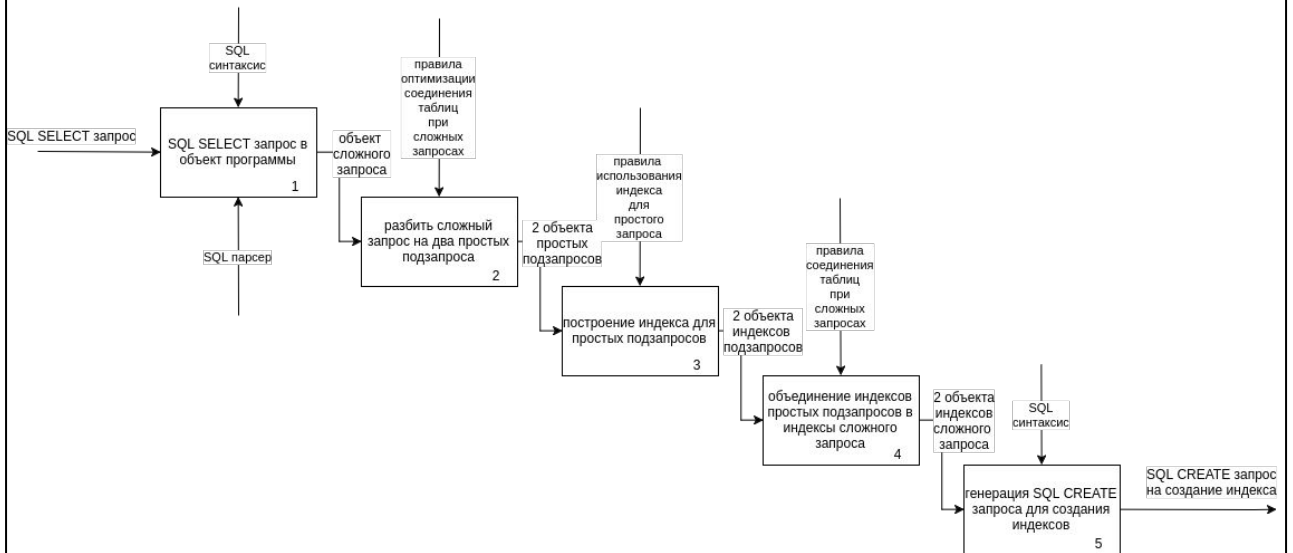
conditional_expression:
    column_name = column_name or
    conditional_expression
```

На данном слайде представлен вид сложного запроса, который поддерживается разработанным ПО.

## 5. Определение рекомендуемых индексов



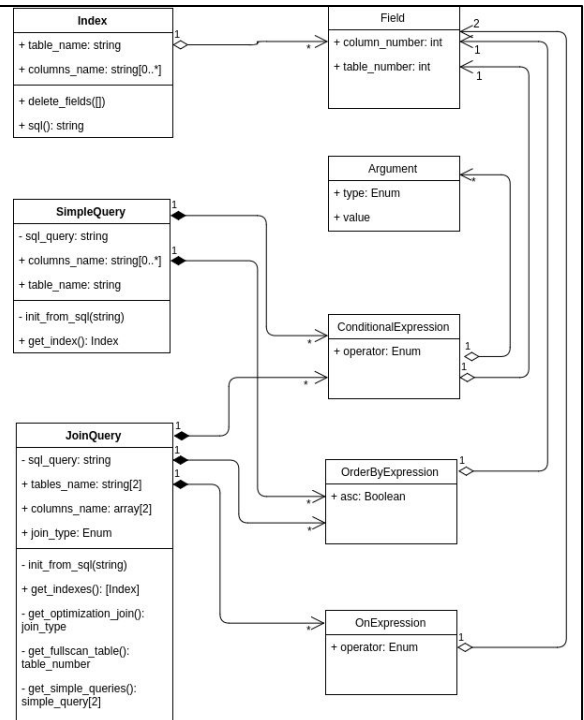
## 6. Определение индексов (декомпозиция)



## 7. Объектно-ориентированный анализ

Сущности:

- поле
- аргумент
- условное выражение
- выражение блока ON
- выражение блока ORDER BY
- простой запрос
- сложный запрос
- индекс



В результате объектно-ориентированного анализа можно выделить следующие сущности: ....

Диаграмма классов представлена на рисунке



## 8. Планирование эксперимента

Параметры и значения:

- join:
  - left, right, inner
- where:
  - -,
  - $t1.a = const, t2.a = const$ ,
  - $t1.a <> const, t2.a <> const$
- order by:
  - -,
  - $t1.a$ ,
  - $t2.a$

Кол-во экспериментов:  $3 * 5 * 3 = 45$

По методу всех пар: =15

Таблица 1

№	join	order by	where
1	inner	-	—
2	inner	$t1.c$	$t1.b = const$
3	inner	$t2.c$	$t2.b = const$
4	inner	-	$t1.b > const$
5	inner	$t1.c$	$t2.b > const$
6	left	$t2.c$	—
7	left	-	$t1.b = const$
8	left	$t1.c$	$t2.b = const$
9	left	$t2.c$	$t1.b > const$
10	left	-	$t2.b > const$
11	right	$t1.c$	—
12	right	$t2.c$	$t1.b = const$
13	right	-	$t2.b = const$
14	right	$t1.c$	$t1.b > const$
15	right	$t2.c$	$t2.b > const$

Когда было разработано ПО, нужно было протестировать его.

Для планирования эксперимента необходимо выделить параметры, которые могут изменяться и значения, которые они могут принимать (одно значения из одного класса эквивалентности). На слайде представлены параметры и их значения.

Получается что необходимо провести 45 экспериментов.

Однако по методу всех пар, т.к. большинство ошибок проявляются либо при конкретных значениях одного параметра, либо взаимным влиянием значений двух параметров, то получилось, что достаточно провести всего 15 экспериментов.

План экспериментов представлен в таблице

## 9. Результаты эксперимента

В 47% экспериментов, ускорение времени выполнения запроса было больше, чем на порядок (в среднем в 4777 раз)

Таблица 2

Результат	номера экспериментов	кол-во	%
Индексы уменьшили время больше, чем на порядок	2, 3, 5, 8, 12, 14, 15	7	47
Индексы уменьшили время меньше, чем на порядок	1, 6, 9, 10, 11, 13	6	40
Индексы увеличили время меньше, чем на порядок	4, 7	2	13
Индексы увеличили время больше, чем на порядок	-	0	0

Результаты эксперимента представлены в таблице 2.

## 10. Результаты

В результате работы удалось решить поставленные задачи и разработать соответствующее программное обеспечение.

Правильность работы ПО подтверждается серией поставленных экспериментов.

## 11. Направления дальнейшего развития

Развитие инструмента:

- увеличить поддержку возможного синтаксиса SQL запросов;
- учесть больше возможных оптимизаций запросов;
- добавить другие типы индексов, отличных от B+tree;
- поддерживать новые версии СУБД.

Использование инструмента в дальнейшей автоматизации процесса индексирования. Такое ПО должно:

- наблюдать за запросами;
- определять какие индексы необходимо построить
- наблюдать за построенными индексами;
- удалять невостребованные индексы

Хотелось бы выделить направления дальнейшего развития.

Во первых, это само развитие инструмента: ....

А также использование инструмента для дальнейшей автоматизации процесса индексирования.

**Спасибо за внимание!**