

Федеральное государственное автономное образовательное учреждение
высшего
образования «Национальный исследовательский университет ИТМО»

Факультет программной инженерии и компьютерной техники

Отчет

По лабораторной работе №4

По дисциплине «Базы данных»

Вариант 1341

Выполнил: Чураков Александр Алексеевич,
группа Р3131

Преподаватель: Харитоновна Анастасия
Евгеньевна

Санкт-Петербург

Задание

Составить запросы на языке SQL (пункты 1-2).

Для каждого запроса предложить индексы, добавление которых уменьшит время выполнения запроса (указать таблицы/атрибуты, для которых нужно добавить индексы, написать тип индекса; объяснить, почему добавление индекса будет полезным для данного запроса).

Для запросов 1-2 необходимо составить возможные планы выполнения запросов. Планы составляются на основании предположения, что в таблицах отсутствуют индексы. Из составленных планов необходимо выбрать оптимальный и объяснить свой выбор.

Изменяются ли планы при добавлении индекса и как?

Для запросов 1-2 необходимо добавить в отчет вывод команды EXPLAIN ANALYZE [запрос]

Подробные ответы на все вышеперечисленные вопросы должны присутствовать в отчете (планы выполнения запросов должны быть нарисованы, ответы на вопросы - представлены в текстовом виде).

1. Сделать запрос для получения атрибутов из указанных таблиц, применив фильтры по указанным условиям:
Н_ТИПЫ_ВЕДОМОСТЕЙ, Н_ВЕДОМОСТИ.
Вывести атрибуты: Н_ТИПЫ_ВЕДОМОСТЕЙ.НАИМЕНОВАНИЕ,
Н_ВЕДОМОСТИ.ЧЛВК_ИД.
Фильтры (AND):
а) Н_ТИПЫ_ВЕДОМОСТЕЙ.НАИМЕНОВАНИЕ < Ведомость.
б) Н_ВЕДОМОСТИ.ИД < 39921.
Вид соединения: LEFT JOIN.
2. Сделать запрос для получения атрибутов из указанных таблиц, применив фильтры по указанным условиям:
Таблицы: Н_ЛЮДИ, Н_ВЕДОМОСТИ, Н_СЕССИЯ.
Вывести атрибуты: Н_ЛЮДИ.ИД, Н_ВЕДОМОСТИ.ИД, Н_СЕССИЯ.ЧЛВК_ИД.
Фильтры (AND):
а) Н_ЛЮДИ.ИД = 152862.
б) Н_ВЕДОМОСТИ.ЧЛВК_ИД < 153285.
с) Н_СЕССИЯ.ДАТА = 2012-01-25.
Вид соединения: INNER JOIN.

Реализация на SQL

1)

```

explain analyse
select t."НАИМЕНОВАНИЕ", v."ЧЛВК_ИД"
from "Н_ТИПЫ_ВЕДОМОСТЕЙ" t
      left join "Н_ВЕДОМОСТИ" v on v."ТВ_ИД" = t."ИД"
where "НАИМЕНОВАНИЕ" < 'Ведомость'
      and v."ИД" < 39921
тип;

```

2)

```

explain analyse
select p."ИД" as people_id, v."ИД" as ved_id, s."ЧЛВК_ИД" as session_man_id
from "Н_ЛЮДИ" p
      inner join "Н_ВЕДОМОСТИ" v on v."ЧЛВК_ИД" = p."ИД"
      inner join "Н_СЕССИЯ" s on s."ЧЛВК_ИД" = p."ИД"
where p."ИД" = 152862
      and v."ЧЛВК_ИД" < 153285
      and s."ДАТА" = '2012-01-25'

```

Уменьшение времени выполнения запросов

Запрос №1

Индексы, добавление которых уменьшит время выполнения запроса

Таблица Н_ТИПЫ_ВЕДОМОСТИ:

Индекс на атрибуте НАИМЕНОВАНИЕ ускорит фильтрацию строк с условием
Н_ТИПЫ_ВЕДОМОСТЕЙ.НАИМЕНОВАНИЕ < Ведомость.

Возможные планы выполнения без индексов

План 1

1. Сначала происходит последовательное сканирование таблицы "Н_ТИПЫ_ВЕДОМОСТЕЙ" с применением фильтра Н_ТИПЫ_ВЕДОМОСТЕЙ.НАИМЕНОВАНИЕ < Ведомость.
2. Затем для каждой строки из таблицы Н_ТИПЫ_ВЕДОМОСТЕЙ происходит выполнение вложенного цикла.
3. Для каждой строки из таблицы Н_ТИПЫ_ВЕДОМОСТЕЙ происходит поиск соответствующих строк в таблице "Н_ВЕДОМОСТИ".
4. Применяется фильтр к результатам сканирования таблицы Н_ВЕДОМОСТИ: ""ИД"" < 39921.

План 2

1. Последовательное сканирование таблицы "Н_ВЕДОМОСТИ" с фильтром "ИД" < 39921.
2. Последовательное сканирование таблицы "Н_ТИПЫ_ВЕДОМОСТЕЙ"
3. Затем происходит хеширование результатов сканирования таблицы "Н_ТИПЫ_ВЕДОМОСТЕЙ" на основе значения столбца ""ИД".
4. Выполняется хеш-объединение между результатами сканирования таблицы "Н_ВЕДОМОСТИ" и хешированными результатами сканирования таблицы "Н_ТИПЫ_ВЕДОМОСТЕЙ". Объединение происходит по условию "v.""ТВ_ИД"" = t.""ИД"".

Оптимальным является план 2, так как фильтр по ИД выполняется до соединения таблиц, что уменьшает количество строк для обработки.

При добавлении индексов вместо полного скана таблиц будет использоваться индексный скан

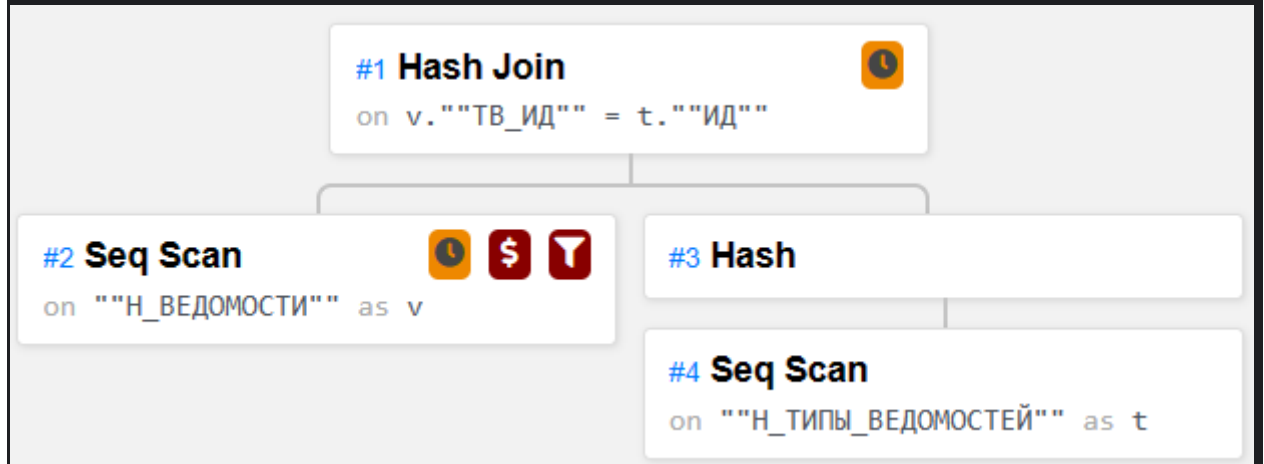
План выполнения

Удалим условие "НАИМЕНОВАНИЕ" < 'Ведомость' и увеличим максимальный ID иначе нет подходящих значений.

```

Hash Join (cost=1.07..7468.62 rows=98061 width=30) (actual
time=2.036..56.851 rows=98307 loops=1)
" Hash Cond: (v.""ТВ_ИД"" = t.""ИД"" )"
" -> Seq Scan on ""Н_ВЕДОМОСТИ"" v (cost=0.00..6846.50 rows=98061 width=8)
(actual time=2.015..30.363 rows=98307 loops=1)"
" Filter: (""ИД"" < 399220)"
" Rows Removed by Filter: 124133
-> Hash (cost=1.03..1.03 rows=3 width=30) (actual time=0.012..0.014
rows=3 loops=1)
Buckets: 1024 Batches: 1 Memory Usage: 9kB
" -> Seq Scan on ""Н_ТИПЫ_ВЕДОМОСТЕЙ"" t (cost=0.00..1.03 rows=3
width=30) (actual time=0.004..0.005 rows=3 loops=1)"
Planning Time: 0.300 ms
Execution Time: 61.568 ms

```



Запрос №2

Индексы, добавление которых уменьшит время выполнения запроса

Индекс на атрибуте ДАТА таблицы Н_СЕССИЯ ускорит фильтрацию строк с условием Н_СЕССИЯ.ДАТА = 2012-01-25.

Возможные планы выполнения без индексов

План 1

1. Полный скан таблицы Н_ЛЮДИ и поиск строки с ИД=1528623
2. Полный скан таблицы Н_ВЕДОМОСТИ с применением фильтра v."ЧЛВК_ИД" < 153285
3. Полный скан таблицы Н_СЕССИЯ с применением фильтра s."ДАТА" = '2012-01-25'
4. Соединение таблиц с использованием Hash Join по атрибутам

План 2

1. Сначала выполняется вложенная операция "Nested Loop", которая состоит из двух частей:

а) Сканирование таблицы "Н_ЛЮДИ" (обозначается как р). Поиск выполняется по ключу "ИД", равному 1528623.

б) Битовое сканирование кучи таблицы "Н_СЕССИЯ" (обозначается как s). Производится поиск записей, где значение поля "ЧЛВК_ИД" равно 1528623, и где дата равна '2012-01-25'.

2. Затем результаты вложенной операции "Nested Loop" объединяются с помощью еще одной операции "Nested Loop". Для каждой строки, полученной из предыдущей операции, выполняется сканирование таблицы "Н_ВЕДОМОСТИ" (обозначается как v). Поиск выполняется по ключу "ЧЛВК_ИД", который меньше 153285 и равен 1528623.

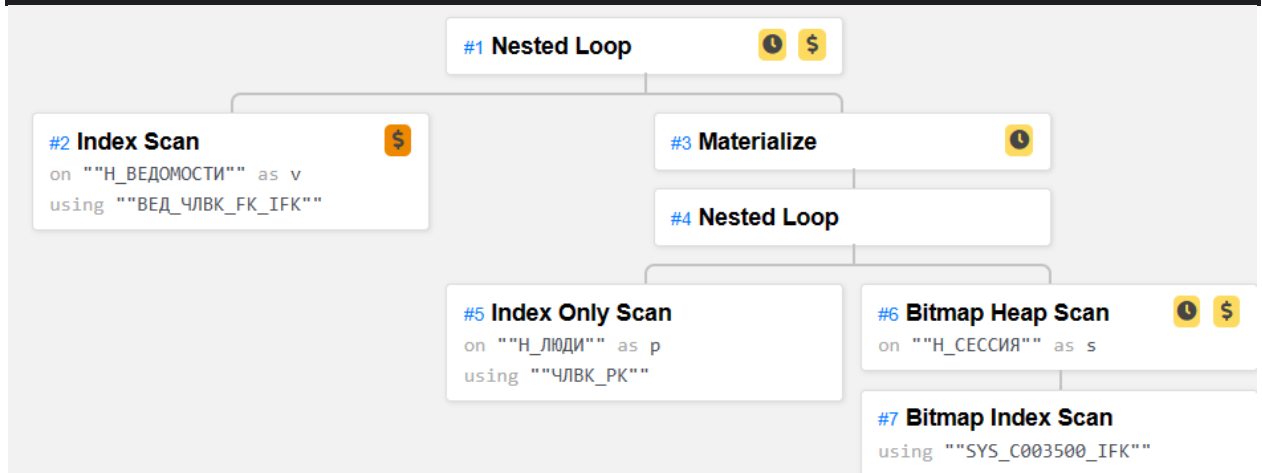
Оптимальным является план 2, потому что фильтрация данных выполняется до соединения таблиц, что уменьшает количество строк для обработки

При добавлении индексов вместо полного скана таблиц будет использоваться индексный скан

План выполнения

Удалим "ДАТА" = '2012-01-25' из фильтра и изменим условие на "ИД" = 110136, чтобы результат запроса был не пустым.

```
Nested Loop (cost=5.18..297.59 rows=2604 width=12) (actual time=0.063..0.294
rows=798 loops=1)
"  ->  Index Scan using ""ВЕД_ЧЛВК_FK_IFK"" on ""Н_ВЕДОМОСТИ"" v
(cost=0.29..190.58 rows=62 width=8) (actual time=0.022..0.030 rows=19
loops=1) "
"      Index Cond: ((""ЧЛВК_ИД"" < 153285) AND (""ЧЛВК_ИД"" = 110136)) "
"  ->  Materialize (cost=4.89..74.57 rows=42 width=8) (actual
time=0.002..0.008 rows=42 loops=19)
"      ->  Nested Loop (cost=4.89..74.36 rows=42 width=8) (actual
time=0.036..0.081 rows=42 loops=1)
"          ->  Index Only Scan using ""ЧЛВК_РК"" on ""Н_ЛЮДИ"" p
(cost=0.28..4.30 rows=1 width=4) (actual time=0.010..0.011 rows=1 loops=1) "
"              Index Cond: (""ИД"" = 110136) "
"                  Heap Fetches: 0
"          ->  Bitmap Heap Scan on ""Н_СЕССИЯ"" s (cost=4.61..69.64
rows=42 width=4) (actual time=0.020..0.058 rows=42 loops=1) "
"              Recheck Cond: (""ЧЛВК_ИД"" = 110136) "
"                  Heap Blocks: exact=21
"          ->  Bitmap Index Scan on ""SYS_C003500_IFK""
(cost=0.00..4.60 rows=42 width=0) (actual time=0.012..0.012 rows=42 loops=1) "
"              Index Cond: (""ЧЛВК_ИД"" = 110136) "
Planning Time: 0.241 ms
Execution Time: 0.376 ms
```



Вывод

В ходе выполнения лабораторной работы я освоил работу с реляционной алгеброй и научился строить планы выполнения запросов, а также их диаграммы. Я изучил различные виды индексов и узнал, как использовать их для оптимизации запросов