Министерство образования и науки РФ

Санкт-Петербургский Политехнический университет Петра Великого

Институт компьютерных наук и кибербезопасности

Высшая школа программной инженерии

ОТЧЕТ

по лабораторной работе вариант №3

Проектирование аналитической схемы базы данных

по дисциплине «Системы управления базы данных»

|  |  |
| --- | --- |
| Выполнил  студент гр. 5130904/10104  Преподаватель | Баженов Кирилл Максимович.  Прокофьев О. В. |

Санкт-Петербург

2024

**Практическое задание №3.1 Проектирование аналитической схемы**

**базы данных**

**Постановка задачи:**

Седьмое практическое задание связано с проектированием схемы базы данных для аналитики.

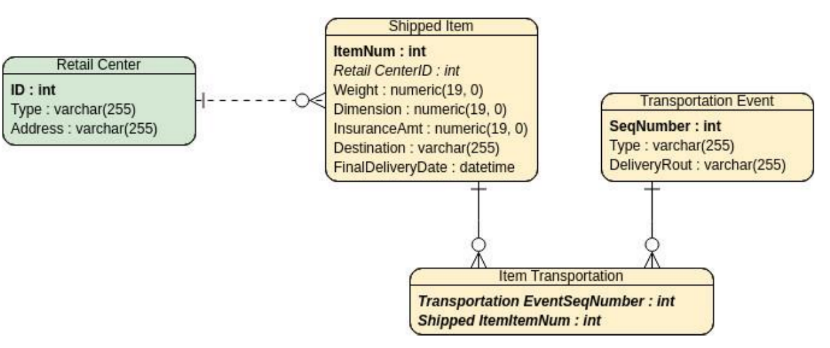
Будем исходить из того, что приложение, для которого была сделана база данных в задании стала

очень популярной и по ней каждый день можно собирать большой объем статистической

информации. Результатом данного практического задания являются: скрипты создания базы

данных, хранимая процедура (генератор) для ее заполнения, анализ плана выполнения запроса.

Пример используемой ER диаграммы (система доставки UPS)



Требования к БД

* Одна таблица (Item transportation) должна содержать не меньше 10 млн. записей, которые со

временем теряют актуальность.

* Другая таблица (Shipped Item), связанная с первой, должна содержать не меньше 1 млн.

записей.

* Остальные таблицы должны содержать по 5-10 записей.
* В одной из таблиц с количеством записей 1 млн. должна быть колонка с текстом, по которой

будет необходимо настроить полнотекстовый поиск.

Практическая часть включает:

1) наполнение таблицы, для этого нужно написать хранимую процедуру - генератор на языке

plpython3u, которая использует словари (для строковых типов), случайные значения (для

строковых, числовых типов).

2) оценку скорости выполнения запросов.

Для этого могут быть использованы механизмы секционирования, наследования и индексов.

Необходимо подготовить два запроса:

* Запрос к одной таблице, содержащий фильтрацию по нескольким полям.
* Запрос к нескольким связанным таблицам, содержащий фильтрацию по нескольким полям.

3) оценка скорости полнотекстового поиска (использование функций to\_tsvector to\_tsquery)

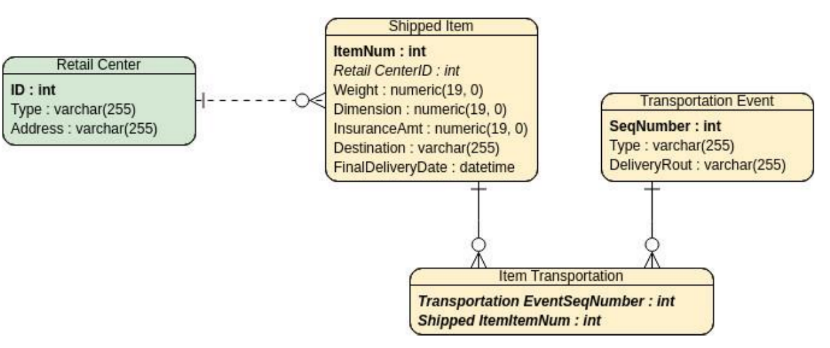
Для каждого из этих запросов необходимо провести следующие шаги:

* Получить план выполнения запроса без использования индексов (удаление индекса или

отключение его использования в плане запроса).

* Получить статистику (IO и Time) выполнения запроса без использования индексов.
* Создать нужные индексы, позволяющие ускорить запрос.
* Получить план выполнения запроса с использованием индексов и сравнить с первоначальным планом.
* Получить статистику выполнения запроса с использованием индексов и сравнить с первоначальной статистикой.
* Оценить эффективность выполнения оптимизированного запроса.

**Выбранная ER диаграмма**

****

**Скрипт DDL аналитической БД**

create table retail\_center (

id serial primary key,

type text,

address text

);

create table shipped\_item(

item\_num int primary key,

retail\_center\_id int references retail\_center(id),

weight numeric(19, 0),

dimension numeric(19, 0),

insurance\_amt numeric(19, 0),

destination text,

final\_delivery\_date timestamp

);

create table transportation\_event (

seq\_number int primary key,

type text,

delivery\_rout text

);

create table item\_transportation(

transportation\_event\_seq\_number int references transportation\_event(seq\_number),

shipped\_item\_item\_num int references shipped\_item(item\_num)

);