**Лабораторная работа №1**

*Тема: хранилища данных и аналитическая обработка данных*

**Цель работы:** получение практических навыков проектирования, разработки и использования хранилищ данных.

**Задание:** спроектируйте БД в многомерной модели представления данных использую модель звезды или снежинки (в реляционной базе) согласно полученному варианту (используя программу Open System Architect или аналогичное CASE-средство, модель должна включать не менее 5 сущностей), реализуйте спроектированную базу в СУБД PostgreSQL.

Внесите в базу тестовые данные (не менее 10 строк в каждую таблицу).

Реализуйте аналитические запросы к базе, используя следующие конструкции секционирование (partitioning), упорядочивание (order by), кадрирование (с использованием rows и range), аналитических функций сведения (crosstab), ранжирования функций (row\_number, rank, dense\_rank), получения значения строк (first\_value, last\_value, lead, lag), статистические (var, varp, stdevp, stdev).

Для справки по синтаксису используйте ресурсы:

<https://postgrespro.ru/docs/postgrespro/9.5/tablefunc>,

<https://postgrespro.ru/docs/postgresql/9.5/tutorial-window>,

<http://www.sql-tutorial.ru/ru/book_crosstab.html>,

<https://postgrespro.ru/docs/postgrespro/9.5/functions-aggregate>.

**Отчет** по лабораторной работе должен содержать:

1. Фамилию и номер группы учащегося, задание
2. Описание многомерной модели (схема)
3. Физическую модель БД (sql-код)
4. Перечень тестовых данных (в виде таблиц)
5. Код запросов, задача (вопрос) для решения которых можно использовать полученные наборы данных (для каждого запроса), и результаты их выполнения (принскрин с базы).

# Варианты заданий

1. Погодные условия в регионе
2. Продажа комплектующих изделий
3. Демографическая ситуация в регионе
4. Продажа земельных участков
5. Рынок труда
6. Больница
7. Железнодорожный транспорт
8. Авиационные перевозки
9. Олимпиада
10. Футбол
11. Туристический бизнес
12. Социальные сети
13. Интернет-провайдер
14. Здравоохранение
15. Автострахование
16. Кредитование
17. Экология
18. Правонарушения
19. Литература
20. Компьютеры

# Теоретический материал

#### ROLAP (Relational OLAP)

ROLAP-системы позволяют представлять данные, хранимые в классической реляционной базе, в многомерной форме или в плоских локальных таблицах на файл-сервере, обеспечивая преобразование информации в многомерную модель через промежуточный слой метаданных. Агрегаты хранятся в той же БД в специально созданных служебных таблицах. В этом случае гиперкуб эмулируется СУБД на логическом уровне.

Преимущества ROLAP.

* Реляционные СУБД имеют реальный опыт работы с очень большими БД и развитые средства администрирования. При использовании ROLAP размер хранилища не является таким критичным параметром, как в случае MOLAP.
* При оперативной аналитической обработке содержимого хранилища данных инструменты ROLAP позволяют производить анализ непосредственно над хранилищем (потому что в подавляющем большинстве случаев корпоративные хранилища данных реализуются средствами реляционных СУБД).
* В случае переменной размерности задачи, когда изменения в структуру измерений приходится вносить достаточно часто, ROLAP системы с динамическим представлением размерности являются оптимальным решением, так как в них такие модификации не требуют физической реорганизации БД, как в случае MOLAP.
* Системы ROLAP могут функционировать на гораздо менее мощных клиентских станциях, чем системы MOLAP, поскольку основная вычислительная нагрузка в них ложится на сервер, где выполняются сложные аналитические SQL-запросы, формируемые системой.
* Реляционные СУБД обеспечивают значительно более высокий уровень защиты данных и хорошие возможности разграничения прав доступа.

Недостатки ROLAP.

* Ограниченные возможности с точки зрения расчета значений функционального типа.
* Меньшая производительность, чем у MOLAP. Для обеспечения сравнимой с MOLAP производительности реляционные системы требуют тщательной проработки схемы БД и специальной настройки индексов. Но в результате этих операций производительность хорошо настроенных реляционных систем при использовании схемы "звезда" сравнима с производительностью систем на основе многомерных БД.

### Моделирование многомерных кубов на реляционной модели данных

#### Схема звезда. Преимущества и недостатки

Схема типа звезды (Star Schema) - схема реляционной базы данных, служащая для поддержки многомерного представления содержащихся в ней данных.

\*Особенности ROLAP-схемы типа "звезда"\*

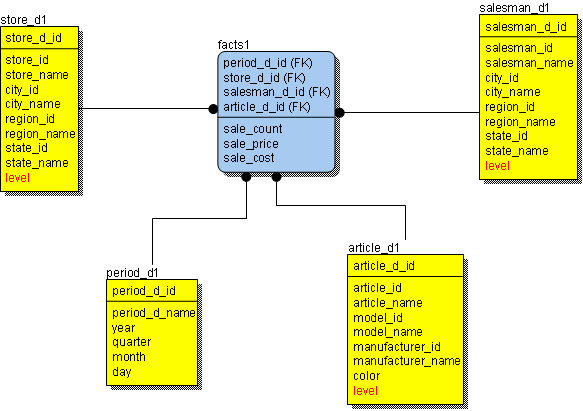
1. Одна таблица фактов (fact table), которая сильно денормализована. Является центральной в схеме, может состоять из миллионов строк и содержит суммируемые или фактические данные, с помощью которых можно ответить на различные вопросы.
2. Несколько денормализованных таблиц измерений (dimensional table). Имеют меньшее количество строк, чем таблицы фактов, и содержат описательную информацию. Эти таблицы позволяют пользователю быстро переходить от таблицы фактов к дополнительной информации.
3. Таблица фактов и таблицы размерности связаны идентифицирующими связями, при этом первичные ключи таблицы размерности мигрируют в таблицу фактов в качестве внешних ключей. Первичный ключ таблицы факта целиком состоит из первичных ключей всех таблиц размерности.
4. Агрегированные данные хранятся совместно с исходными.

https://www.intuit.ru/EDI/27_04_15_5/1430086749-14056/tutorial/320/objects/2/files/plus.gif**Преимущества**

Благодаря денормализации таблиц измерений упрощается восприятие структуры данных пользователем и формулировка запросов, уменьшается количество операций соединения таблиц при обработке запросов. Некоторые промышленные СУБД и инструменты класса OLAP / Reporting умеют использовать преимущества схемы "звезда" для сокращения времени выполнения запросов.

https://www.intuit.ru/EDI/27_04_15_5/1430086749-14056/tutorial/320/objects/2/files/minus.gif**Недостатки**

Денормализация таблиц измерений вносит избыточность данных, возрастает требуемый для их хранения объем памяти. Если агрегаты хранятся совместно с исходными данными, то в измерениях необходимо использовать дополнительный параметр - уровень иерархии.



#### Схема снежинка. Преимущества и недостатки

Схема типа снежинки (Snowflake Schema) - схема реляционной базы данных, служащая для поддержки многомерного представления содержащихся в ней данных, является разновидностью схемы типа "звезда" (Star Schema).

\*Особенности ROLAP-схемы типа "снежинка"\*

1. Одна таблица фактов (fact table), которая сильно денормализована. Является центральной в схеме, может состоять из миллионов строк и содержать суммируемые или фактические данные, с помощью которых можно ответить на различные вопросы.
2. Несколько таблиц измерений (dimensional table), которые нормализованы в отличие от схемы "звезда". Имеют меньшее количество строк, чем таблицы фактов, и содержат описательную информацию. Эти таблицы позволяют пользователю быстро переходить от таблицы фактов к дополнительной информации. Первичные ключи в них состоят из единственного атрибута (соответствуют единственному элементу измерения).
3. Таблица фактов и таблицы размерности связаны идентифицирующими связями, при этом первичные ключи таблицы размерности мигрируют в таблицу фактов в качестве внешних ключей. Первичный ключ таблицы факта целиком состоит из первичных ключей всех таблиц размерности.
4. В схеме "снежинка" агрегированные данные могут храниться отдельно от исходных.

https://www.intuit.ru/EDI/27_04_15_5/1430086749-14056/tutorial/320/objects/2/files/plus.gif**Преимущества**

Нормализация таблиц измерений в отличие от схемы "звезда" позволяет минимизировать избыточность данных и более эффективно выполнять запросы, связанные со структурой значений измерений.

https://www.intuit.ru/EDI/27_04_15_5/1430086749-14056/tutorial/320/objects/2/files/minus.gif**Недостатки**

За нормализацию таблиц измерений иногда приходится платить временем выполнения запросов.

