Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

“САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ”

Хранилища данных. Витрины. Архитектура хранилища данных

Кислюк Игорь Витальевич,

Группа К3421

Санкт-Петербург, 2015 г.

ВВЕДЕНИЕ

Принципы, лежащие в основе систем поддержки принятия решений, не позволяют эффективно обрабатывать транзакции, поэтому данные, применяемые для анализа, стали выделять в отдельные базы данных. Впоследствии эти базы стали называть хранилищами данных или информационными хранилищами. Отцом концепции использования хранилищ данных считают Билла Инмона.

Концепции хранилищ данных - это концепция подготовки данных для анализа. Она предполагает выполнение следующих положений:

1. Интеграции и согласования данных из различных источников

2. Разделения наборов данных

В тот же момент, развиваются системы OLTP для обеспечения высокого параллелизма работы пользователей, позволяя им одновременно обращаться к одному источнику данных и производить необходимую обработку данных

Системы OLAP принадлежат к категории систем поддержки принятия решений и управленческих информационных систем. Целью систем OLAP является анализ огромных объемов данных, генерирование резюме и агрегаций множеством различных способов с целью нахождения тенденций, конкурентных преимуществ и оптимизации в коммерческой деятельности.

Информация, хранимая в системах OLAP, изменяется редко. Данные здесь хранятся для использования в запросах, для создания отчётов, которые могут помочь людям планировать будущее своих предприятий.

База данных превращается в пространственную базу данных (dimensional database), отвечающую особой схеме либо структуре. Базы данных OLAP используются для построения кубов данных (data cube), являющихся многомерными представлениями данных, которые обеспечивают коммерческий анализ в реальном времени и высокую скорость обработки запросов.

АРХИТЕКТУРА БАЗ ДАННЫХ ДЛЯ ХРАНИЛИЩА

Структура базы данных для хранилища разрабатывается с целью максимально облегчить анализ информации.

Структура такой базы данных хранилища не будет реляционной, это будет пространственная база данных (dimensional database). Главная таблица пространственной базы данных называется таблицей фактов. Ее строки именуются фактами и являются оценочными показателями активности. Измерения помогают расположить факты соответствующим образом и представить такие атрибуты, как время, товар, покупатель и его географическое местоположение. Пространственные таблицы описывают данные в таблицах фактов.

В большинстве случаев информация в базе данных хранилища может быть в виде N-мерного куба фактов, отражающих деловую активность компании в течение определенного времени.

Каждая его ячейка представляет один факт, например, объем продаж в долларах. Вдоль одной грани куба располагаются месяцы, в течение которых выполнялись отражаемые кубом продажи. Второе измерение составляют категории товаров. Третье - регионы продаж. В каждой ячейке содержится объем продаж за соответствующей комбинации значений по всем трем измерениям.

ВИТРИНЫ ДАННЫХ

Многомерная модель позволяет производить быстрый анализ данных, но не позволяет хранить большие объемы информации. Реляционная модель, напротив, практически не имеет ограничений по объему накапливаемы данных

Ситуация, когда для анализа необходима вся информация, находящаяся в хранилище, возникает довольно редко. Обычно каждый аналитик или аналитический отдел обслуживает одно из направлений деятельности организации, поэтому в первую очередь ему необходимы данные, характеризующие именно это направление. Реальный объем этих данных не превосходит этих ограничений, присущих многомерным СУБД. Возникает идея выделить данные, которые реально нужны конкретным аналитическим приложениям, в отдельный набор. Такой набор мог бы быть реализован в многомерной БД. Источником данных для него должно быть центральное хранилище организации.

Если проводить аналогии с производством и реализацией продукции, то многомерные БД выполняют роль мелких складов. В концепции ХД их принято именовать витринами данных (Data Marts). Витрина данных - это специализированное тематическое хранилище, обслуживающее одно из направлений деятельности организации. Логическая схема СППР, использующей центральное ХД организации и витрины данных аналитических отделов.

Такая схема позволяет эффективно использовать возможности реляционных СУБД по хранению огромных объемов информации и способность многомерных СУБД обеспечивать высокую скорость выполнения аналитических запросов.

ЗАКЛЮЧЕНИЕ

Хранилища данных преследуют данные задачи - своевременное обеспечение аналитиков всей информацией и создание единой модели данных организации, а также создание интегрированного источника данных, предоставляющего удобный доступ к разнородной информации. Их актуальность повышается при увеличении данных, анализ которых требуется и производится. С одной стороны, они отходят от стандартной концепции реляционных баз данных и свое развитие получили в направлении больших данных - Big Data, с другой стороны - они так же упираются и базируются на базах данных.