

Использование технологии OLAP при проектировании

многомерного хранилища данных

Капанов Н.А.; Снисаренко С.В.; Стасевич Н.А.

ИСИТ ИИТ, СУ

БГУИР

Минск, Республика Беларусь

e-mail: kafsu@bsuir.by

Аннотация — В докладе даётся краткое описание технологии OLAP, приведены основные этапы проектирования и создания многомерного хранилища данных с использованием технологии ROLAP, дано описание процедуры создания куба на основе разработанного алгоритма.

Ключевые слова: реляционная база данных, информационно-аналитическая система, компонент многомерного анализа, многомерное хранение данных, консолидация, OLAP – куб, ROLAP – технология.

I. ВВЕДЕНИЕ

Основная проблема заключалась в том, что большинство баз данных, используемых в настоящее время, имеют достаточно сложную структуру, насчитывают сотни таблиц. Из-за этого чаще всего они не могут относиться ни к одной из нормальных форм. Анализ таких баз данных превращается в серьёзную проблему для экспертов - аналитиков, и программистов, поскольку процесс создания аналитических структур ложится на их плечи. Существующие программные решения данной задачи либо слишком дороги, либо требуют больших временных затрат на каждом этапе работы, либо имеют ограничения по функционалу. Для решения этой проблемы существует технология OLAP, которая служит для проектирования многомерных хранилищ данных с использованием OLAP - кубов. Ещё одной проблемой является способ доступа к аналитическим данным. В настоящее время для создания любого отчёта аналитику требуется помощь программиста, знакомого с внутренней структурой аналитической системы, что влечёт за собой дополнительные временные и трудовые затраты.

II. КРАТКОЕ ОПИСАНИЕ ТЕХНОЛОГИИ OLAP

OLAP — технология обработки информации, включающая составление и динамическую публикацию отчётов и документов. Это ключевой компонент организации хранилищ данных. К основным особенностям OLAP систем относят [1]:

работу с многомерными структурами;
получение информации об аномалиях, закономерностях и трендах в базе данных без выяснения смыслового значения данных;

обеспечение быстрого доступа к результатам без SQL;

использование при анализе заранее рассчитанных сводных данных.

Основные понятия систем OLAP:

измерение — описывает элемент данных, по которому производится анализ. Например, распространённым элементом анализа является время;

атрибут — это полная коллекция элементов одного типа;

мера — количественная оценка одного факта;

куб — многомерная структура, включающая в себя набор измерений и групп мер [2].

Основные достоинства OLAP:

использование в запросах заранее рассчитанных данных;

запросы составляются на языке MDX. Особенностью данного языка является то, что нет необходимости знать структуру базы данных. Достаточно указать, что мы хотим из неё получить;

время выполнения запросов гораздо меньше, чем время выполнения эквивалентных на языке SQL. В ходе проведённых исследований было выяснено, что выигрыш во времени может достигать 150-200 раз;

возможность получения разнообразной информации на одних и тех же данных.

III. ЭТАПЫ ПРОЕКТИРОВАНИЯ МНОГОМЕРНОГО ХРАНИЛИЩА ДАННЫХ

Данная статья посвящена разработке многомерного хранилища данных для системы анализа деятельности предприятия, в частности созданию структуры базы данных, построению и моделированию многомерного куба, созданию запросов к базе данных для построения кросс-таблицы и диаграмм. Для проектирования многомерного хранилища данных выбрана ROLAP технология, так как она подразумевает смешанное хранение данных, т. е. разработана реляционная база данных, а агрегаты хранятся в той же базе данных в специально созданных служебных таблицах.

Этапы создания OLAP куба [3]:

1) Проектирование реляционной базы данных. Создание таблицы фактов - основной таблицы хранилища данных. Таковой будет являться таблица взаимодействия менеджеров. Она содержит уникальный составной ключ, объединяющий первичные ключи таблиц измерений. При этом как ключевые, так и некоторые не ключевые поля будут соответствовать будущим измерениям OLAP-куба. Помимо этого таблица фактов содержит поля, на основании которых в дальнейшем будут получены агрегатные данные.

2) Описание измерений OLAP – куба в таблице «измерения куба». Под измерением куба понимается последовательность значений одного из анализируемых параметров. В данной системе таких измерений принято четыре: время, менеджер, тип взаимодействия и статус взаимодействия. Представление такого количества измерения графически с разложением осей измерений невозможно, поэтому применим метод разложения куба в виде дерева.

3) Описание в таблице «иерархия измерения» структуры измерений.

Так как взят за основу ROLAP тип системы, то необходимые аналитические данные для формирования куба извлекаются из спроектированной реляционной базы данных при помощи функции содержащей SQL запросы [4].

После извлечения необходимых данных результат сохраняется во временной таблице. Клиентское приложение будет осуществлять все необходимые манипуляции с данными (разрез, вращение, детализация

и консолидация) и представлять их в виде отчета или диаграмм. Для этого разработан алгоритм, по которому реализован код процедуры формирования куба.

Данная функция позволяет сделать аналитическое агрегирование данных на основании заданных условий, которые определяются пользователем системы. После выполнения, процедура возвращает результат (количество взаимодействий) в ячейку согласно условиям.

IV. Выводы

В результате выбора и проектирования OLAP куба получили:

- многомерное концептуальное представление данных;
- интуитивное манипулирование данными;
- доступность;
- пакетное извлечение против интерпретации;
- OLAP-продукт поддерживает все модели анализа данных;
- прозрачность (к примеру, пользователь электронной таблицы способен получить любые необходимые данные из OLAP-машины, даже не подозревая, откуда они берутся);
- инструменты OLAP обеспечивают одновременный доступ (чтение и запись), интеграцию и конфиденциальность;
- обработка ненормализованных данных (т.е. модификация данных, выполненная в среде OLAP не должна приводить к изменениям данных, хранимых в исходных внешних системах);
- сохранение результатов OLAP (т.е. данные, модифицированные в OLAP, должны сохраняться отдельно от данных транзакций);

- исключение отсутствующих значений (отсутствующие значения отличны от нулей);

- обработка отсутствующих значений в OLAP (эти значения игнорируются OLAP-анализатором без учета их источника);

- гибкость формирования отчетов (измерения куба интерактивно размещаются так, как это нужно пользователю);

- производительность формирования отчетов существенно не падает с ростом количества измерений и размеров базы данных;

- неограниченные операции между размерностями.

В результате спроектирована многомерная модель данных.

Таким образом, решены следующие задачи:

- проведен анализ технологий создания OLAP-приложений;
- разработана многомерная структура хранилища данных с помощью ROLAP – технологии;
- разработан алгоритм, по которому реализован код процедуры формирования куба.

[1] Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Методы и модели анализа данных: OLAP и Data Mining. – СПб.: БХВ-Петербург, 2004. – 336 с.

[2] Кудрявцев Ю.А. OLAP технологии: обзор решаемых задач и исследований // Бизнес-информатика. – 2008. №1. – С. 66-70.

[3] Хоббс Л.В., Хилсон С.К. Разработка и эксплуатация хранилищ баз данных. Кудиц-образ Москва, 2004.

[4] Аллен К. Oracle PL/SQL. Лори Москва, 2005