

КЛАССИФИКАЦИЯ OLAP ПО СПОСОБУ ОРГАНИЗАЦИИ ХРАНИЛИЩА ДАННЫХ. ДОСТОИНСТВА И НЕДОСТАТКИ

Файзрахманов Е.Г.

Томский политехнический университет
634050, Россия, г. Томск, пр-т Ленина, 30
E-mail: jeenefg@mail.ru

Введение

На сегодняшний день множество предприятий используют базы данных для хранения корпоративной информации. Такое повсеместное использование баз данных на разных уровнях управления привело к накоплению большого количества информации. Эти данные потенциально возможно использовать для анализа и построения статистической модели и стратегии развития предприятия, и тем самым облегчить процесс принятия управленческих решений. Но для этого информацию необходимо представить в нужной и удобной для аналитика форме. Здесь нам и помогут OLAP технологии.

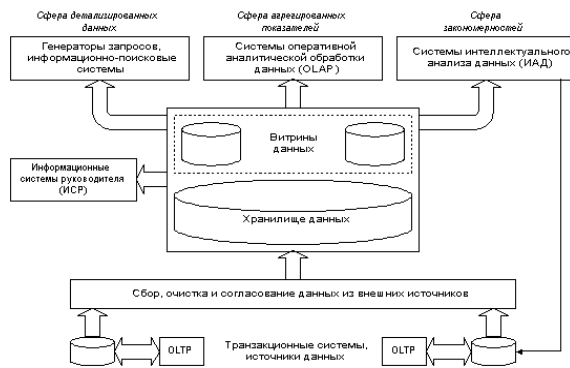


Рис. 1. Полная структура корпоративной информационно-аналитической системы [1]

В основе концепции OLAP лежит принцип многомерного представления данных. Однако во многих публикациях аббревиатурой OLAP обозначается не только многомерный взгляд на данные, но и хранение самих данных в многомерной БД. Однако Е.Ф. Codd, определивший общие требования к OLAP, отмечает, что "Реляционные БД были, есть и будут наиболее подходящей технологией для хранения корпоративных данных. Необходимость существует не в новой технологии БД, а, скорее, в средствах анализа, дополняющих функции существующих СУБД и достаточно гибких, чтобы предусмотреть и автоматизировать разные виды интеллектуального анализа, присущие OLAP" [1].

Модели OLAP

На данный момент существует три основных модели OLAP отличающихся по способу организации базы данных:

- многомерный OLAP (MOLAP, Multidimensional OLAP);
- реляционный OLAP (ROLAP, Relational OLAP);

- гибридный OLAP (HOLAP, Hybrid OLAP)[2].

Многомерный OLAP

В многомерных OLAP-системах структура куба хранится в многомерной базе данных вместе с предварительно обработанными агрегатами и копиями листовых значений. В связи с этим все запросы к данным удовлетворяются многомерной системой баз данных, что делает MOLAP-системы исключительно быстрыми. Архитектура MOLAP требует большого объема дискового пространства из-за хранения в многомерной базе копий листовых данных[2]. Несмотря на это, объем дополнительного пространства обычно не слишком велик, поскольку данные в MOLAP хранятся исключительно эффективно.

Достоинства MOLAP-систем:

- все данные хранятся в многомерных структурах, что существенно повышает скорость обработки запросов;
- доступны расширенные библиотеки для сложных функций оперативного анализа;
- обработка разреженных данных выполняется лучше, чем в ROLAP.

Недостатки:

- необходимы специальные инструменты для формирования кубов и их пересчета в случае изменения базовых значений;
- сложно изменять измерения без повторной агрегации

Реляционный OLAP

В реляционных OLAP-системах структура куба данных хранится в реляционной базе данных. Меры самого нижнего уровня остаются в реляционной витрине данных, служащей источником данных для куба. Предварительно обработанные агрегаты также хранятся в реляционной таблице[3]. Благодаря реляционным таблицам, архитектура ROLAP позволяет хранить большие объемы данных. ROLAP-системы лишены запаздывания в части листовых данных, Поскольку в архитектуре ROLAP листовые значения берутся непосредственно из витрины данных.



Рис. 2. Пример схемы «звезда» [4]



Рис. 3. Пример схемы «снежинка» [4]

Достоинствами ROLAP-систем являются:

- возможность использования ROLAP с хранилищами данных и различными OLTP-системами;
- возможность манипулирования большими объемами данных;
- безопасность и администрирование обеспечивается реляционными СУБД.

Недостатки:

- получение агрегатов и листовых данных происходит медленнее, чем в MOLAP и HOLAP;
- функциональность систем ограничивается возможностями SQL, так как аналитические запросы пользователя транслируются в SQL-операторы выборки;
- сложно пересчитывать агрегированные значения при изменениях начальных данных;
- сложно поддерживать таблицы агрегатов.

Гибридный OLAP

В гибридных OLAP сочетаются черты ROLAP и MOLAP. В моделях HOLAP используются преимущества и минимизируются недостатки обеих архитектур. В HOLAP-системах структура куба и предварительно обработанные агрегаты хранятся

в многомерной базе данных. Это позволяет обеспечить быстрое извлечение агрегатов из структур MOLAP. Значения нижнего уровня иерархии в HOLAP остаются в реляционной витрине данных, которая служит источником данных для куба.

К достоинствам подхода можно отнести комбинирование технологии ROLAP для разреженных данных и MOLAP для плотных областей, а к недостаткам – необходимость поддержания MOLAP и ROLAP.

Заключение

Можно сделать вывод, что выбор модели OLAP зависит от требований к скорости загрузки данных, требуемого дискового пространства и т.д. После выбора модели и настройки OLAP корпоративная информация предприятия будет представлена аналитику в необходимой форме, что позволит пересмотреть стратегию дальнейшего развития предприятия.

Литература

1. Способы аналитической обработки данных для поддержки принятия решений [Электронный ресурс]. – Режим доступа: <http://infovisor.ivanovo.ru/press/paper04.html>, свободный.
2. Кузьмин А. Н. Методы и модели обработки информации в хранилищах данных /А. Н. Кузьмин //Математическое моделирование, численные методы и комплексы программ. 2006. – С.193.
3. OLAP [Электронный ресурс]. Режим доступа: <http://ru.wikipedia.org/wiki/OLAP> свободный
4. Оперативная аналитическая обработка данных: концепции и технологии [Электронный ресурс]. – Режим доступа: http://www.olap.ru/basic/olap_and_ida.asp, свободный
5. OLAP и многомерные базы данных [Электронный ресурс]. – Режим доступа: <http://www.rae.ru/monographs/141-4638>, свободный

АВТОМАТИЧЕСКИЙ АНАЛИЗ ДОКУМЕНТОВ НА ЕСТЕСТВЕННЫХ ЯЗЫКАХ

Губин М.Ю.

Томский политехнический университет
634050, Россия, г. Томск, пр-т Ленина, 30
E-mail: gubin.m.u@gmail.com

Введение

Одной из фундаментальных проблем машинной обработки текстов является то, что естественные человеческие языки обладают большой выразительностью и сложностью, существенное влияние на смысл текста в них оказывает контекст и эмоциональная составляющая. Понимание есте-

ственного языка включает куда больше, чем разбор предложений на индивидуальные части речи и поиск значений слов в словаре. Оно базируется на обширном фоновом знании о предмете, идиомах, используемых в этой области, а также на способности применять общее контекстуальное знание для понимания недомолвок и неясностей, прису-