

УДК 004.652.4.031.43

А.М. Павлов (6-й курс, каф.АиВТ), Е.В. Пышкин, к.т.н., доц.

МЕТОДЫ АНАЛИЗА И ПРИНЯТИЯ РЕШЕНИЙ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ OLAP-ТЕХНОЛОГИЙ

Технологию OLAP (Online Analytical Processing) можно определить как совокупность средств многомерного оперативного анализа данных, накопленных в хранилище данных. Хранилище данных – это предметно-ориентированное, привязанное ко времени и неизменяемое собрание данных для поддержки процесса принятия управляющих решений. Принципы организации OLAP были в основном сформулированы Е.Ф. Коддом в 1993 году. Позже его определение было переработано в так называемый тест FASMI, требующий, чтобы OLAP-приложение предоставляло возможности быстрого анализа разделяемой многомерной информации.

В OLAP-хранилище данные попадают из реляционной базы данных, при этом допускается одновременное использование разных СУБД. Важнейшим элементом хранилища являются метаданные, то есть информация о структуре, размещении и трансформации данных. Наличие метаданных обеспечивает эффективное взаимодействие различных компонентов хранилища. Следует также отметить, что модифицировать данные в хранилище напрямую нельзя.

OLAP предоставляет удобные быстродействующие средства доступа, просмотра и анализа деловой информации. Ориентированная на пользователя модель данных представляет собой множество многомерных кубов. Осями многомерной системы координат служат основные атрибуты анализируемого бизнес процесса. Например, процесс продаж может характеризоваться такими атрибутами как товар, регион, тип покупателя, производитель, магазин, оборудование – в зависимости от того, какие процессы необходимо анализировать. В качестве одного из измерений используется время. На пересечениях осей (dimensions) находятся данные, количественно характеризующие процесс. Эти данные называют мерами (measures). Примерами мер могут служить объемы продаж, остатки продукции на складе, издержки и т.п. Меры представляют собой агрегируемые данные. Меры рассчитываются при переносе данных из таблиц в куб. Пользователь, анализирующий информацию, может «разрезать» куб по разным направлениям, получать сводные или детальные сведения, осуществлять прочие манипуляции, которые могут потребоваться в процессе анализа. Срезом принято называть совокупность значений на осях.

Говоря об измерениях, следует упомянуть о том, что значения, наносимые на оси, могут иметь различные уровни детализации. Например, пользователя может интересовать суммарная стоимость заказов, сделанных клиентами в разных странах, или суммарная стоимость заказов, сделанных клиентами отдельной категории и т.д. Естественно, во втором случае результирующий набор агрегатных данных будет более детальным, чем в первом. Заметим, что возможность получения агрегатных данных с различной степенью детализации соответствует одному из требований, предъявляемых к хранилищам данных – требованию доступности различных срезов данных для сравнения и анализа.

В зависимости от того, как данные хранятся в кубе, существует 3 вида OLAP-кубов:

- MOLAP (Multidimensional OLAP) – исходные и агрегатные данные хранятся в многомерной базе данных. Такие кубы используют тогда, когда в первую очередь требуется добиться максимальной производительности при обработке запросов. Размер куба может быть довольно большим, если используется большое количество агрегатов. Этот тип кубов используется наиболее часто.

- ROLAP (Relational OLAP) – исходные данные остаются в той же реляционной базе данных, где они изначально и находились. Агрегатные же данные помещают в специально

созданные для их хранения служебные таблицы в той же базе данных. Такой тип хранилища наименее эффективен при обработке запросов, однако его использование позволяет модифицировать данные в кубе.

- **HOLAP (Hybrid OLAP)** – исходные данные остаются в той же реляционной базе данных, где они изначально находились, а агрегатные данные хранятся в многомерной базе данных.

Методика анализа данных с помощью OLAP является одной из областей профессиональных интересов автора. Примерами организации систем многомерного анализа данных, разработанных при непосредственном участии автора, являются недавние внедрения обзорных в докладе средств в сети магазинов «Дикси» и в гипермаркете «О`КЕЙ» в Санкт-Петербурге. Отметим, что базы данных, содержащие информацию, предназначенную для анализа, имеют размер порядка 10 миллионов записей. Время формирования выборки данных при таких объемах с помощью SQL-запросов составляет 10-15 минут на каждый запрос, а общее число запросов могло достигать 20. С помощью OLAP такой запрос выполнялся 5-10 секунд. Подобное быстродействие может оказаться очень существенным, если, скажем, директору компании необходимо быстро, за 20...30 минут получить представление о том, насколько была эффективна работы компании за месяц, как это соотносится с предыдущими месяцами, или какая ситуация была в этом же месяце год назад. Информация об этом может впоследствии использоваться для принятия решений на ближайшем совещании о политике компании.

В качестве OLAP-сервера в этих организациях использовался Microsoft Analysis Server. В качестве альтернативного варианта можно было бы рассматривать OLAP-сервер на базе Oracle, но он обладает меньшей функциональностью, не поддерживает перенос данных в куб из гетерогенных источников. При этом предполагается, что интерпретация данных осуществляется при использовании продукта Comshare Decision в качестве клиентского приложения. На основании сопоставительного анализа Comshare Decision с другими программами (рассматривались, в частности, MS Excel, MS Data Analyzer, Crystall Decision) были выявлены основные преимущества использования указанной компоненты. Во-первых, достигается высокая эффективность обработки с точки зрения быстродействия: выборка 15,000 строк из 10 миллионов выполняется в Comshare Decision около 1 минуты, что примерно в 8-10 раз быстрее по сравнению с Crystall Decision, в 15...20 раз быстрее по сравнению Data Analyzer, в 150-200 раз по сравнению с MS Excel. Во-вторых, обеспечиваются развитые возможности визуализации результатов анализа. Наконец, следует отметить возможность использования Comshare Decision через Web-браузер, то есть, нет необходимости устанавливать компоненту на компьютерах пользователя.