

image not found or type unknown



В наше время без систем управления базами данных не обходится практически ни одна организация, особенно среди тех, которые традиционно ориентированы на взаимодействие с клиентами. Банки, страховые компании, авиа- и прочие транспортные компании, сети супермаркетов, телекоммуникационные и маркетинговые фирмы, организации, занятые в сфере услуг и другие – все они собирают и хранят в своих базах гигабайты данных о клиентах, продуктах и сервисах. Ценность подобных сведений несомненна. Такие базы данных называют операционными или транзакционными, поскольку они характеризуются огромным количеством небольших транзакций, или операций записи-чтения. Компьютерные системы, осуществляющие учет операций и собственно доступ к базам транзакций, принято называть системами оперативной обработки транзакций (OLTP – On-Line Transactional Processing) или учетными системами.

Учетные системы настраиваются и оптимизируются для выполнения максимального количества транзакций за короткие промежутки времени. Обычно отдельные операции очень малы и не связаны друг с другом. Однако каждую запись данных, характеризующую взаимодействие с клиентом (звонок в службу поддержки, кассовую операцию, заказ по каталогу, посещение Web-сайта компании и т.п.) можно использовать для получения качественно новой информации, а именно для создания отчетов и анализа деятельности фирмы.

Набор аналитических функций в учетных системах обычно весьма ограничен. Схемы, используемые в OLTP-приложениях, осложняют создание даже простых отчетов, так как данные чаще всего распределены по множеству таблиц, и для их агрегирования необходимо выполнять сложные операции объединения. Как правило, попытки создания комплексных отчетов требуют больших вычислительных мощностей и приводят к потере производительности.

Кроме того, в учетных системах хранятся постоянно изменяющиеся данные. По мере сбора транзакций суммарные значения меняются очень быстро, поэтому два анализа, проведенные с интервалом в несколько минут, могут дать разные результаты. Чаще всего, анализ выполняется по окончании отчетного периода, иначе картина может оказаться искаженной. Кроме того, необходимые для анализа данные могут храниться в нескольких системах.

Некоторые виды анализа требуют таких структурных изменений, которые недопустимы в текущей оперативной среде. Например, нужно выяснить, что произойдет, если у компании появятся новые продукты. На живой базе такое исследование провести нельзя. Следовательно, эффективный анализ редко удастся выполнить непосредственно в учетной системе.

Системы поддержки принятия решений обычно обладают средствами предоставления пользователю агрегатных данных для различных выборок из исходного набора в удобном для восприятия и анализа виде. Как правило, такие агрегатные функции образуют многомерный (и, следовательно, нереляционный) набор данных (нередко называемый гиперкубом или метакубом), оси которого содержат параметры, а ячейки — зависящие от них агрегатные данные — причем храниться такие данные могут и в реляционных таблицах. Вдоль каждой оси данные могут быть организованы в виде иерархии, представляющей различные уровни их детализации. Благодаря такой модели данных пользователи могут формулировать сложные запросы, генерировать отчеты, получать подмножества данных.

Именно это и обусловило интерес к системам поддержки принятия решений, ставших основной сферой применения OLAP (On-Line Analytical Processing, оперативная аналитическая обработка, оперативный анализ данных), превращающей “руду” OLTP-систем в готовое “изделие”, которое руководители и аналитики могут непосредственно использовать. Этот метод позволяет аналитикам, менеджерам и руководителям “проникнуть в суть” накопленных данных за счет быстрого и согласованного доступа к широкому спектру представлений информации.

Трудно найти в компьютерном мире человека, который хотя бы на интуитивном уровне не понимал, что такое базы данных и зачем они нужны. В отличие от традиционных реляционных СУБД, концепция OLAP не так широко известна, хотя загадочный термин “кубы OLAP” слышали, наверное, почти все. Что же такое Online Analytical Processing?

OLAP — это не отдельно взятый программный продукт, не язык программирования и даже не конкретная технология. Если постараться охватить OLAP во всех его проявлениях, то это совокупность концепций, принципов и требований, лежащих в основе программных продуктов, облегчающих аналитикам доступ к данным.

Для начала мы выясним, зачем аналитикам надо как-то специально облегчать доступ к данным. Дело в том, что аналитики – это особые потребители корпоративной информации. Задача аналитика – находить закономерности в больших массивах данных. Поэтому аналитик не будет обращать внимания на отдельно взятый факт, что в четверг четвертого числа контрагенту Чернову была продана партия черных чернил – ему нужна информация о сотнях и тысячах подобных событий. Одиночные факты в базе данных могут заинтересовать, к примеру, бухгалтера или начальника отдела продаж, в компетенции которого находится сделка. Аналитику одной записи мало – ему, к примеру, могут понадобиться все сделки данного филиала или представительства за месяц, год. Заодно аналитик отбрасывает ненужные ему подробности вроде ИНН покупателя, его точного адреса и номера телефона, индекса контракта и тому подобного. В то же время данные, которые требуются аналитику для работы, обязательно содержат числовые значения – это обусловлено самой сущностью его деятельности.

Централизация и удобное структурирование – это далеко не все, что нужно аналитику. Ему ведь еще требуется инструмент для просмотра, визуализации информации. Традиционные отчеты, даже построенные на основе единого хранилища, лишены одного – гибкости. Их нельзя “покрутить”, “развернуть” или “свернуть”, чтобы получить желаемое представление данных. Конечно, можно вызвать программиста, и он сделает новый отчет достаточно быстро – скажем, в течение часа. Получается, что аналитик может проверить за день не более двух идей. А ему (если он хороший аналитик) таких идей может приходить в голову по несколько в час. И чем больше “срезов” и “разрезов” данных аналитик видит, тем больше у него идей, которые, в свою очередь, для проверки требуют все новых и новых “срезов”. Вот бы ему такой инструмент, который позволил бы разворачивать и сворачивать данные просто и удобно! В качестве такого инструмента и выступает OLAP.

Хотя OLAP и не представляет собой необходимый атрибут хранилища данных, он все чаще и чаще применяется для анализа накопленных в этом хранилище сведений.

Оперативные данные собираются из различных источников, очищаются, интегрируются и складываются в реляционное хранилище. При этом они уже доступны для анализа при помощи различных средств построения отчетов. Затем данные (полностью или частично) подготавливаются для OLAP-анализа. Они могут быть загружены в специальную БД OLAP или оставлены в реляционном хранилище.

Важнейшим его элементом являются метаданные, т. е. информация о структуре, размещении и трансформации данных. Благодаря им обеспечивается эффективное взаимодействие различных компонентов хранилища.

Подытоживая, можно определить OLAP как совокупность средств многомерного анализа данных, накопленных в хранилище. Теоретически средства OLAP можно применять и непосредственно к оперативным данным или их точным копиям (чтобы не мешать оперативным пользователям). Но мы тем самым рискуем наступить на уже описанные выше грабли, то есть начать анализировать оперативные данные, которые напрямую для анализа непригодны.