



ВВЕДЕНИЕ

OLAP (On-Line Analytical Processing) -- оперативная аналитическая обработка данных. Иногда аналитическую обработку называют и онлайновой, и интерактивной, однако прилагательное «оперативная» как нельзя более точно отражает смысл технологии OLAP.

OLAP -- это не отдельно взятый программный продукт, а целая технология. Если постараться охватить OLAP во всех его проявлениях, то это совокупность концепций, принципов и требований, лежащих в основе программных продуктов, облегчающих доступ к данным.

Термин "OLAP" неразрывно связан с термином "хранилище данных" (Data Warehousing) -- местом OLAP в информационной структуре предприятия.

Эта технология применяется в хранилищах данных для повышения эффективности анализа данных. Причём анализ данных происходит в режиме реального времени.

1. ОСНОВА КОНЦЕПЦИИ OLAP

OLAP - представляет собой инструмент для анализа больших объёмов данных. Взаимодействуя с OLAP-системой, пользователь сможет осуществлять гибкий просмотр информации, получать произвольные срезы данных и выполнять аналитические операции детализации, свёртки, сквозного распределения, сравнения во времени. Вся работа с OLAP-системой происходит в терминах предметной области.

OLAP-системы являются часть более общего понятия Business Intelligence, которое включает в себя помимо традиционного OLAP-сервиса средства организации совместного использования документов, возникающих в процессе работы пользователей хранилища. Технология Business Intelligence обеспечивает электронный обмен отчётными документами, разграничение прав пользователей, доступ к аналитической информации из Internet/Intranet.

В основе концепции OLAP лежит принцип многомерного представления данных. По измерениям в многомерной модели выделяют факторы, влияющие на деятельность предприятия (например, время, продукты, отделения компании и т.п.) и получают гиперкуб, который затем наполняется показателями деятельности предприятия (цены, продажи, план, прибыли, убытки и т.п.). Наполнение это может вестись как реальными данными оперативных систем, так и прогнозируемыми на основе исторических данных, то есть данных, накопленных за определённый период времени.

Измерения гиперкуба могут носить сложный характер, быть иерархическими, между ними могут быть установлены отношения. В процессе анализа пользователь может менять точку зрения на данные (так называемая операция смены логического взгляда), тем самым, просматривая данные в различных разделах и разрешая конкретные задачи. Над кубами могут выполняться различные операции, включая прогнозирование и условное планирование (анализ типа «что, если»).

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, то есть информация о структуре, размещении и трансформации данных. Благодаря им обеспечивается эффективное взаимодействие различных компонентов хранилища.

2. ОБЩИЕ ТРЕБОВАНИЯ К OLAP-СИСТЕМАМ И СПОСОБЫ ХРАНЕНИЯ ДАННЫХ В НИХ.

В 1993 году Е.Ф. Коддом - создателем концепции реляционных СУБД и, по совместительству, OLAP -- были сформулированы критерии OLAP. Они заключаются в недостатках реляционной модели и, в первую очередь, указывают на невозможность «объединять, просматривать и анализировать данные с точки зрения множественности измерений, то есть самым понятным для корпоративных аналитиков способом». Общие требования к системам OLAP, расширяют функциональность реляционных СУБД и включают многомерный анализ как одну из своих характеристик.

Кодд определил 12 правил, которым должен удовлетворять программный продукт класса OLAP:

Многомерное концептуальное представление данных (Multi-Dimensional Conceptual View). Концептуальное представление модели данных в продукте OLAP должно быть многомерным по своей природе, то есть позволять аналитикам выполнять интуитивные операции «анализа вдоль и поперёк» («slice and dice»), вращения (rotate) и размещения (pivot) направлений консолидации.

Устойчивая производительность (Consistent Reporting Performance). С увеличением числа измерений и размеров базы данных аналитики не должны столкнуться с каким бы то ни было уменьшением производительности. Устойчивая производительность необходима для поддержания простоты использования и свободы от усложнений, которые требуются для доведения OLAP до конечного пользователя.

Клиент - серверная архитектура (Client-Server Architecture). Главная идея работы в среде клиент - сервер - это то, что серверный компонент инструмента OLAP должен быть достаточно интеллектуальным и обладать способностью строить общую концептуальную схему на основе обобщения и консолидации различных логических и физических схем корпоративных баз данных для обеспечения эффекта прозрачности.

Инструмент OLAP должен обеспечивать оптимальную обработку разреженных матриц. Скорость доступа должна сохраняться вне зависимости от расположения ячеек данных и быть постоянной величиной для моделей, имеющих разное число измерений и различную разреженность данных. Инструмент OLAP должен предоставлять им конкурентный доступ, обеспечивать целостность и защиту данных.

Интуитивное манипулирование данными (Intuitive Data Manipulation). Детализация данных в колонках и строках, агрегация и другие манипуляции, свойственные структуре иерархии, должны выполняться в максимально удобном, естественном и комфортном пользовательском интерфейсе.

Гибкий механизм генерации отчётов (Flexible Reporting). Должны поддерживаться различные способы визуализации данных, то есть отчёты должны представляться в любой возможности ориентации.

Набор этих требований, послуживших фактическим определением OLAP, следует рассматривать как рекомендательный, а конкретные продукты оценивать по степени приближения к идеально полному соответствию всем требованиям.

Позднее все эти требования были переработаны в так называемый тест FASMI, который также определяет требования к продуктам OLAP. FASMI -- это аббревиатура от названия каждого пункта теста:

Fast (Быстрый). Приложение OLAP должно обеспечивать минимальное время доступа к аналитическим данным -- в среднем порядка 5 секунд;

Analysis (Анализ). Приложение OLAP должно давать пользователю возможность осуществлять числовой и статистический анализ;

Shared (Разделяемый доступ). Приложение OLAP должно предоставлять возможность работы с информацией многим пользователям одновременно;

(Многомерность). Приложение должно обеспечивать многомерное концептуальное представление данных, включая полную поддержку для иерархий.

Information (Информация). Приложение OLAP должно давать пользователю возможность получать нужную информацию, в каком бы электронном хранилище данных она не находилась.

Данные могут храниться либо в реляционных, либо в многомерных структурах. Поэтому в настоящее время применяются три способа хранения данных:

MOLAP (Multidimensional OLAP) - исходные и агрегатные данные хранятся в многомерной базе данных. Хранение данных в многомерных структурах позволяет манипулировать данными как многомерным массивом, благодаря чему скорость вычисления агрегатных значений одинакова для любого из измерений. Однако в этом случае многомерная база данных оказывается избыточной, так как многомерные данные полностью содержат исходные реляционные данные.

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.

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

Отметим также, что подавляющее большинство современных OLAP-средств не хранит «пустых» значений (примером «пустого» значения может быть отсутствие продаж сезонного товара вне сезона).

Одним из популярных средств разработки OLAP-систем является семейство программных продуктов Oracle Express OLAP компании Oracle. Программное обеспечение Oracle Express предоставляет широкие возможности для создания аналитических систем на основе сервера многомерных баз данных - Oracle Express Server. В состав инструментальных средств Oracle Express входят средства создания и администрирования многомерных баз данных - Express Administrator, средство визуального создания "облегченных" клиентских приложений и презентаций - Express Analyzer, профессиональная инструментальная среда объектно-ориентированной разработки OLAP-приложений Express Objects, позволяющая создавать сложные интегрированные клиентские приложения, и другие средства, связанные с публикацией данных в Интернете.

3. OLAP НА КЛИЕНТЕ И НА СЕРВЕРЕ.

Многомерный анализ данных может быть произведен с помощью различных средств, которые условно можно разделить на клиентские и серверные OLAP-средства.

Клиентские OLAP-средства представляют собой приложения, осуществляющие вычисление агрегатных данных (сумм, средних величин, максимальных или минимальных значений) и их отображение, при этом сами агрегатные данные содержатся в кэше внутри адресного пространства такого OLAP-средства.

Если исходные данные содержатся в настольной СУБД, вычисление агрегатных данных производится самим OLAP-средством. Если же источник исходных данных - серверная СУБД, многие из клиентских OLAP-средств посылают на сервер запросы, содержащие оператор GROUP BY, и в результате получают агрегатные данные, вычисленные на сервере.

Многие средства разработки содержат библиотеки классов или компонентов, позволяющие создавать приложения, реализующие простейшую OLAP-функциональность. Помимо этого многие компании предлагают элементы управления ActiveX и другие библиотеки, реализующие подобную функциональность.

Отметим, что клиентские OLAP-средства применяются, как правило, при малом числе измерений (обычно рекомендуется не более шести) и небольшом разнообразии значений этих параметров, -- ведь полученные агрегатные данные должны уместиться в адресном пространстве подобного средства, а их количество растет экспоненциально при увеличении числа измерений. Поэтому даже самые

примитивные клиентские OLAP-средства, как правило, позволяют произвести предварительный подсчет объема требуемой оперативной памяти для создания в ней многомерного куба.

Многие клиентские OLAP-средства позволяют сохранить агрегатные данные в виде файла, что, в свою очередь, позволяет не производить их повторное вычисление. Отметим, что нередко такая возможность используется для отчуждения агрегатных данных с целью передачи их другим организациям или для публикации. Типичным примером таких отчуждаемых агрегатных данных является статистика заболеваемости в разных регионах и в различных возрастных группах, которая является открытой информацией, публикуемой министерствами здравоохранения различных стран и Всемирной организацией здравоохранения. При этом собственно исходные данные, представляющие собой сведения о конкретных случаях заболеваний, являются конфиденциальными данными медицинских учреждений, которые ни в коем случае не должны попадать в руки страховых компаний и тем более становиться достоянием гласности.

Идея сохранения агрегатных данных в файле получила свое дальнейшее развитие в серверных OLAP-средствах. В них сохранение и изменение агрегатных данных, а также поддержка содержащего их хранилища осуществляются отдельным приложением или процессом, называемым OLAP-сервером. Клиентские приложения могут запрашивать подобное многомерное хранилище и в ответ получать те или иные данные. Некоторые клиентские приложения могут также создавать такие хранилища или обновлять их в соответствии с изменившимися исходными данными.

Преимущества применения серверных OLAP-средств по сравнению с клиентскими OLAP-средствами: в случае применения серверных средств вычисление и хранение агрегатных данных происходят на сервере, а клиентское приложение получает лишь результаты запросов к ним, что позволяет в общем случае снизить сетевой трафик, время выполнения запросов и требования к ресурсам, потребляемым клиентским приложением. Поскольку все ведущие производители серверных СУБД производят (либо лицензировали у других компаний) те или иные серверные OLAP-средства, выбор их достаточно широк и почти во всех случаях можно приобрести OLAP-сервер того же производителя, что и у самого сервера баз данных.

, что многие клиентские OLAP-средства позволяют обращаться к серверным OLAP-хранилищам, выступая в этом случае в роли клиентских приложений, выполняющих подобные запросы. Помимо этого имеется немало продуктов,

представляющих собой клиентские приложения к OLAP-средствам различных производителей.

ЗАКЛЮЧЕНИЕ.

В последние годы аналитическая обработка данных привлекает все большее внимание в мире. Например, аналитические модули появились в составе всех основных западных и российских финансово-производственных приложений ведь в условиях рыночной экономики качество информационной поддержки деятельности руководителей и аналитиков является одним из факторов достижения успеха предприятия. OLAP и является той технологией, которая превращает "сырые" данные OLTP в информацию и знание для конечных пользователей.

В заключение, можно определить OLAP как совокупность средств анализа данных, накопленных в хранилище, а также его основные особенности:

Выделение из большого объёма исторических данных содержательной информации (знаний) с использованием средств обработки информации на основе методов искусственного интеллекта

Использование мощной вычислительной техники и специального хранилища данных, которое накапливает информацию из различных источников за большой период времени, а также обеспечение оперативного доступа к данным.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ:

1. Асеев Г.Г. Электронный документооборот /Г.Г. Асеев. -К.: Кондор, 2007. -- 500 с.
2. Асеев Г.Г. Электронный документооборот /Г.Г. Асеев. Х.: ХДАК, 2000. -- 470 с.
3. Заботнев М.С. Методы представления информации в разреженных гиперкубах данных [Электронный ресурс]. - Режим доступа: <http://www.olap.ru/basic/theory.asp>
4. Введение в OLAP. [Электронный ресурс]. - Режим доступа: <http://www.olap.ru/basic/oolap.asp>
5. Стариков. Ядро OLAP системы. [Электронный ресурс]. - Режим доступа: <http://www.masters.donntu.edu.ua/2004/kita/petrov/library/lec8.htm>
6. Альперович М. Введение в OLAP и многомерные базы данных. [Электронный ресурс]. - Режим доступа: <http://www.olap.ru/basic/alpero2i.asp>

