

Лекция 10 Архитектуры OLAP-клиентов

В лекции представлен обзор архитектур OLAP-клиентов на примере нескольких зарубежных продуктов, а также рассмотрены возможные варианты применения OLAP-клиентов для разработки аналитических приложений и систем.

Принципы работы OLAP-клиентов

OLAP-клиент предоставляет конечному пользователю интерфейс для выполнения произвольных запросов к данным, их многомерного анализа и формирования интерактивных, модифицируемых самим пользователем отчетов. При этом используются данные внешних источников.

Общий принцип работы OLAP-клиентов - предварительное описание семантического слоя, который "скрывает" физическую схему данных от пользователя. После создания этого слоя пользователь может самостоятельно манипулировать понятными ему объектами, используя термины предметной области. Например, для создания интерактивного отчета пользователь выбирает бизнес-объект "Продажи" и "перетаскивает" его атрибуты в области колонок или строк, далее он может так же просто, движениями мыши, задавать условия фильтрации или группировки данных (в этот момент OLAP-клиент генерирует SQL-запрос к реляционной базе данных (БД) или запрос к многомерной БД, например на языке MDX). Описание семантического слоя может храниться в выделенном репозитории метаданных, в приложении или в системном репозитории многомерной БД.

Разные OLAP-клиенты обладают различными архитектурами, и некоторые из OLAP-клиентов могут быть развернуты в нескольких вариантах архитектур. OLAP-клиенты способны работать в однопользовательском режиме с локальными данными отдельного

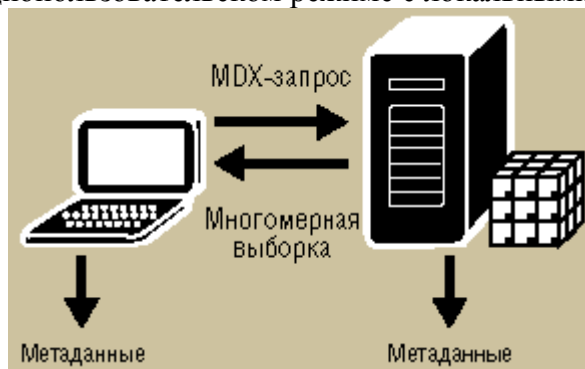


Рис. 1. OLAP-клиент MOLAP-сервера

ПК, в файл-серверной архитектуре, в классической двухуровневой клиент-серверной архитектуре, в трехуровневой клиент-серверной архитектуре с сервером приложения (в последнем случае OLAP-клиент может быть реализован как тонкий Web-клиент; эти Web-клиенты заслуживают отдельного обсуждения и здесь не рассматриваются). Точно так же в OLAP-клиентах организуется и работа с метаданными. Источниками данных могут быть локальные таблицы (например, в формате dbf), таблицы SQL-серверов, локальные многомерные кубы, многомерные базы (кубы) данных OLAP-серверов. Возможна реализация вычисляющей OLAP-машины в OLAP-клиенте, в

MOLAP-сервере (OLAP-сервере, который обрабатывает данные, хранимые в физической многомерной БД), в ROLAP-сервере (OLAP-сервере, обрабатывающем данные, хранимые в реляционной БД), в сервере приложений. Далее будут приведены несколько примеров архитектур OLAP-клиентов, предназначенных для использования в локальных сетях и на отдельных ПК.

Варианты архитектур OLAP-клиентов

OLAP-клиент MOLAP-сервера. Первичные данные из внешних источников периодически загружаются в многомерную БД MOLAP-сервера, в которой постоянно хранятся как первичные атомарные данные, так и вычисленные в момент загрузки агрегаты (см. рис. 1). Часть метаданных, а именно описания атрибутов куба, и OLAP-машина находятся на сервере; OLAP-клиент предоставляет средства организации диалога и отображает данные в виде таблицы.

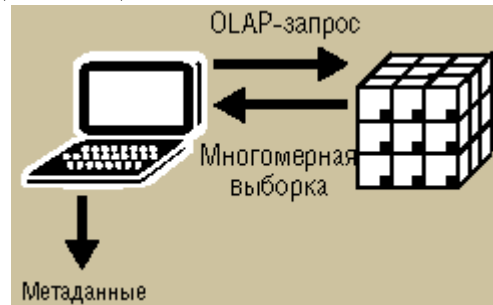


Рис. 2. OLAP-клиент с локальным кубом

При каждой манипуляции пользователя с этой таблицей генерируется запрос к OLAP-серверу на его языке, например MDX, и отображается вычисленный на сервере результат. Ограничения доступа пользователей к данным задаются на стороне MOLAP-сервера.

Наиболее простой OLAP-клиент этого типа - электронная таблица MS Excel, в которую встроен OLAP-компонент - "сводная таблица" (Pivot Table). Аналитическим приложением является книга MS Excel. Часть метаданных, описывающая свойства отчета, путь к серверу, имя базы данных и куба, хранится в книге MS Excel. Права доступа к отчету ограничиваются доступом к файлу MS Excel. Совместное использование настроек отчета группой пользователей обеспечивается копированием файла этих настроек на ПК каждого пользователя.

Существует несколько реализаций подобных OLAP-клиентов от третьих фирм, которые поставляются в виде дополнения (add-in) к MS Excel. Один из примеров - BusinessQuery компании BusinessObjects.

Достоинства и недостатки такого OLAP-клиента определяются свойствами MS Excel. Достоинства - простота освоения инструмента конечными пользователями и его относительно невысокая стоимость. Персональное владение файлом, в котором содержатся настройки отчета, удобно для пользователя, но усложняет сопровождение многопользовательских приложений. За счет предварительного вычисления агрегатов MOLAP-сервер обеспечивает наивысшую производительность обработки больших объемов данных и при этом предъявляются невысокие требования к ресурсам ПК пользователей. Но за это приходится платить необходимостью администрирования процессов загрузки данных из источников первичной информации.



Рис. 3. OLAP-клиент и ROLAP-сервер

Кроме приведенного здесь варианта, MS Pivot Table может работать и без OLAP-сервера (обрабатывая реально до 65 000 записей). OLAP-сервер MS OLAP Server

7/Analysis Services 2000 способен функционировать в режимах ROLAP, MOLAP или HOLAP (в последнем случае одновременно обрабатываются данные как многомерных, так и реляционных БД) - по выбору администратора.

OLAP-клиент с локальным кубом. Плоские данные могут быть перегружены в локальный многомерный куб/файл, который располагается на файл-сервере для общего использования или на локальном ПК - для персонального (см. рис. 2). Генерацию локальных кубов обеспечивают Cognos, MS Pivot Services, BusinessObjects.

Метаданные, описывающие физическую структуру куба, хранятся в нем самом, а метаданные, описывающие интерфейсы и отчеты, - в репозитории OLAP-клиента (BusinessObjects) или в приложении (Cognos, MS Excel).

Таким способом создаются "специальные редакции" OLAP-системы Cognos. Например, одна часть приложения Cognos Nasdaq Special Edition - генератор кубов - находится в бирже Nasdaq, а вторая часть - аналитическая - у пользователей. Пользователи подписываются на биржевые данные, рассылаемые им в виде локальных кубов.

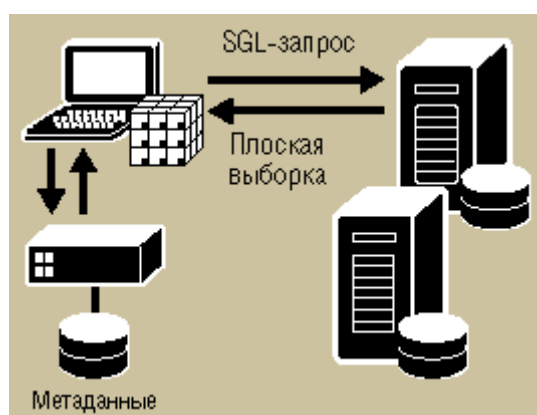


Рис.4 . OLAP-клиент с OLAP-машиной и центральным репозиторием метаданных

Эта технология удобна для автономной (off-line) работы на ПК, отключенном от сети, или для выполнения многократного анализа однажды подготовленных данных. И как показывают тесты, эта конфигурация до определенного объема данных показывает даже более высокое быстродействие, чем OLAP-сервер.

OLAP-клиент и ROLAP-сервер. Данные хранятся в реляционных СУБД внешних систем, все вычисления выполняются в масштабе реального времени (on-line) на специальном ROLAP-сервере (см. рис. 4). OLAP-клиент предоставляет интерактивный интерфейс пользователю, генерирует запросы на внутреннем языке системы к ROLAP-серверу, который на основе метаданных собственного репозитория преобразует их в SQL-запросы к источникам данных - реляционным СУБД.

Так реализована система MicroStrategy. MS OLAP Server 7/Analysis Services 2000 также может быть настроен для работы в такой архитектуре.

Трехуровневая конфигурация обеспечивает возможность обслуживания тысяч пользователей за счет масштабирования и кластеризации ROLAP-сервера. Требования к мощности ПК пользователей минимальны. В отличие от случая с MOLAP-сервером вычисления выполняются "на лету", что устраняет необходимость процедуры перезагрузки данных из реляционных БД в многомерные кубы. Такая система, как и OLAP-сервер, требует администрирования и не использует огромных возможностей современных ПК.



Рис. 5. OLAP-клиент с OLAP-машиной и множеством репозиториях метаданных в виде файлов

В MicroStrategy оригинально решена операция детализации, или углубления (drill-down). При каждом запросе пользователя на детализацию отчета генерируется SQL-запрос, результат которого отображается в новом окне. Это не кажется удобным интерфейсным решением для пользователя. Но за счет того, что одновременно обрабатываются только те данные, что отображаются на экране, система обладает уникальной возможностью анализировать данные из реляционных баз неограниченного размера, вплоть до терабайтов.

OLAP-клиент с OLAP-машиной. OLAP-клиент со встроенной OLAP-машиной не хранит агрегатные данные на диске, а в момент запроса загружает их в оперативную память и выполняет вычисления на ПК пользователя (см. рис. 4). Источниками данных могут быть таблицы SQL-серверов, локальные таблицы, ERP-системы.

Работа с метаданными реализуется по-разному. Например, может существовать выделенный центральный репозиторий метаданных. Он хранит описания источников данных, запросов и отчетов и права пользователей.

Построить такую архитектуру позволяет система BusinessObjects. Единый репозиторий метаданных дает всем пользователям возможность применять однажды настроенные отчеты. Администратор получает инструмент для разграничения прав пользователей на бизнес-объекты и отчеты.

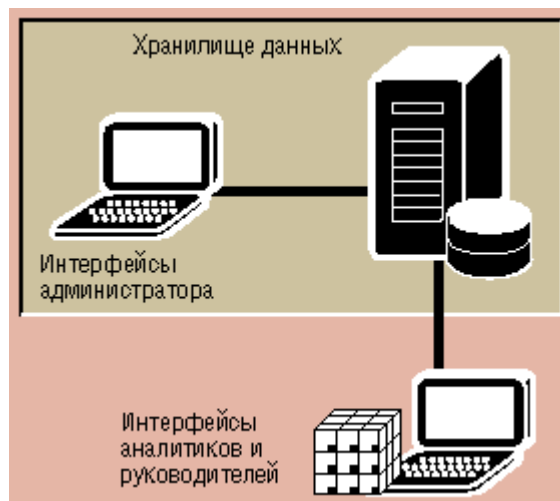


Рис. 6. OLAP-клиент с Хранилищем данных

В клиент-серверной архитектуре с OLAP-машиной на стороне клиента реализован и OLAP-клиент "Контур Стандарт" компании Intersoft Lab. Отличие от архитектуры, описанной выше, состоит в том, что репозиторий метаданных оформлен в виде файла. В терминологии системы он называется "файл приложения" и содержит описание семантического слоя, скрывающего расположение в сети и структуру базы данных, а также описание интерактивных отчетов. Этих файлов может быть один или много (см. рис. 5). Поэтому клиент может работать поочередно с несколькими репозиториями,

которые могут находиться как на его ПК, так и на файл-сервере. Разделение прав доступа на данные реализуется средствами СУБД, а на отчеты - средствами операционной системы: для каждой рабочей группы создается свой файл приложения, доступ к которому ограничивается членами этой группы.

Достоинство этого способа состоит в отчуждаемости репозитория, который можно просто скопировать, послать по почте в филиалы. Кроме того, репозиторий не требует обслуживания. Такая архитектура обеспечивает необычайную гибкость при решении прикладных задач. Этот OLAP-клиент может работать как DOLAP (Desktop OLAP), если данные находятся на локальном ПК, как файл-серверное приложение, если файлы БД лежат на сетевом диске, или как мощное многопользовательское клиент-серверное приложение, если данные предоставляет SQL-сервер.

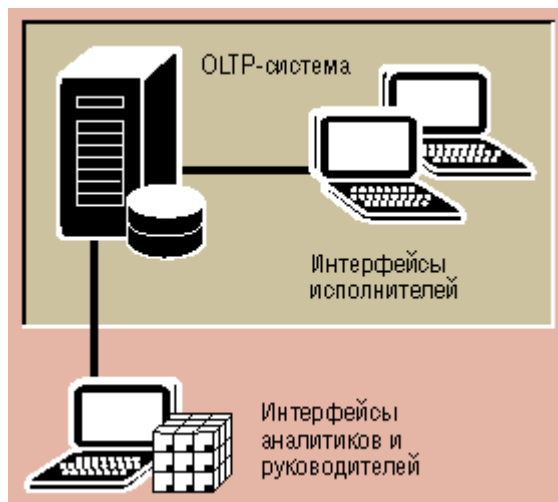


Рис. 7. OLAP-клиент, подключенный к БД транзакционной системы

Независимо от способа хранения метаданных, в этом случае клиент-серверная система выполняет распределенные вычисления, при этом благодаря расположению OLAP-машины на стороне клиента в полной мере используется мощность ПК пользователей. Кроме того, не требуется администрирования дополнительной многомерной БД или ROLAP-сервера. При этом максимальные объемы данных и скорость их обработки зависят от мощности ЦП и объемов оперативной памяти ПК. Как показало тестирование системы "Контур Стандарт", на ПК с тактовой частотой 400 МГц и 194 Мб оперативной памяти OLAP-операции над загруженным в оперативную память массивом из миллиона уникальных записей выполняются практически с нулевым временем отклика. При прогнозировании производительности системы следует учитывать важную особенность работы OLAP-клиента в клиент-серверной архитектуре с РСУБД. И Business Objects, и "Контур Стандарт" выполняют анализ запросов пользователей и генерируют SQL-запрос к СУБД, в котором описывается подходящий алгоритм предварительной группировки. Группирует сервер, а клиенту возвращается компактная выборка для дальнейших OLAP-вычислений. Размер этой выборки может быть в десятки и сотни раз меньше объема первичных данных. Таким образом, возможности OLAP-клиента также очень сильно зависят от мощности сервера БД. При достаточно быстром сервере и большом объеме реальных данных, в которых, как правило, существуют не уникальные записи, OLAP-клиент сможет обрабатывать десятки или даже сотни миллионов первичных, не агрегированных записей.

Применение OLAP-клиентов

OLAP-клиент может быть применен для решения нескольких бизнес-задач:

1. Создание аналитической инфраструктуры предприятия силами специалистов самого предприятия или внедренческой фирмы.

2. Создание тиражируемого аналитического приложения для популярной OLTP-системы силами независимого разработчика.
3. Создание тиражируемого Хранилища данных или высокоуровневого инструмента для построения Хранилищ данных.
4. Создание интегрированного аналитического модуля для OLTP-системы разработчиком этой системы.

Перечисленные бизнес-задачи могут быть решены с использованием различных конфигураций информационных систем, построенных с помощью OLAP-клиента.

Хранилище данных. Разработчик уникального Хранилища или универсального инструмента для построения Хранилищ может сконцентрироваться на проектировании БД и средств сбора и очистки данных. Выполнение запросов, анализ и выпуск отчетов делегируются готовым специализированным системам - OLAP-клиентам (см. рис. 6). Другими словами, программирование интерфейсов Хранилища может быть сведено только к созданию рабочего места администратора.

Транзакционная система. Транзакционные системы, от больших западных ERP-систем до компактных российских "бухгалтерий", наилучшим образом выполняют свои основные функции - автоматизацию учета и выпуск регламентированных отчетов, но не маркетинговый или финансовый анализ. Однако они накапливают совокупность данных, которые можно использовать для этого анализа.

При помощи OLAP-клиента эти данные можно получать в режиме реального времени непосредственно из БД OLTP-системы (см. рис. 7). Для этого структура БД должна быть прозрачной, а мощность сервера БД достаточной для одновременного выполнения "длинных" аналитических запросов и "коротких" транзакций операционистов.

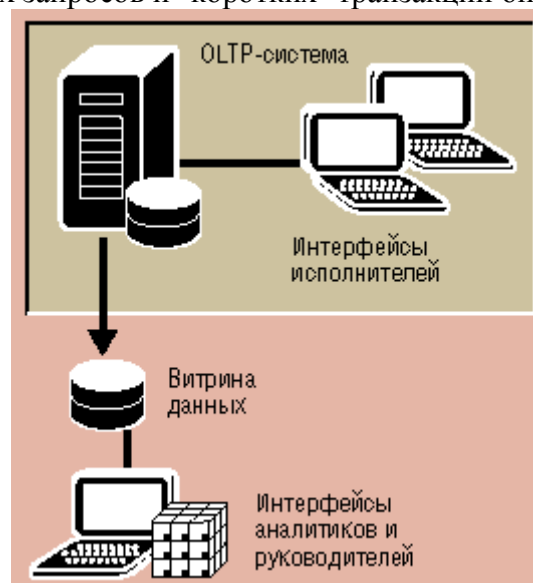


Рис. 8. Витрина данных OLTP-системы, созданная для анализа

Настройка OLAP-клиента на данные OLTP-системы дополняет ее оперативные и административные интерфейсы аналитическими, что дает большие экономические преимущества в сравнении с доработкой самих систем. Встроенные интерфейсы OLTP-системы будут продолжать служить для выполнения транзакций и, как правило, предназначаться для операционистов, а интерфейсы OLAP-клиента будут использоваться специалистами и руководителями.

Этот подход одинаково пригоден как для уникальных систем предприятия, так и для популярных тиражируемых систем. Так, BusinessObjects предоставляет "Шаблоны быстрого развертывания" (Rapid Deployment Templates, RDT). Они представляют собой настройки семантического слоя, переводящие схемы данных популярных на Западе ERP-систем, таких, как SAP R/3, PeopleSoft и другие, на язык бизнес-объектов. При помощи RDT

пользователь или поставщик конечных аналитических приложений может быстро настроить необходимые предприятию интерактивные интерфейсы и отчеты.

Настройки "Контур Стандарт" на тиражируемую бухгалтерскую систему, сохраненные в отдельном ресурсе - файле "Приложения", могут распространяться вместе с этой системой ее автором или, как дополнение к ней, независимым разработчиком.

Витрины данных. В случае, если БД оперативной системы закрыта для прямого доступа или ее структура не подходит для OLAP-анализа, можно создавать специальные наборы реляционных таблиц или многомерных кубов - Витрины данных (Data Mart) и периодически выгружать в них данные для анализа, например, из бухгалтерской системы (см. рис. 8).

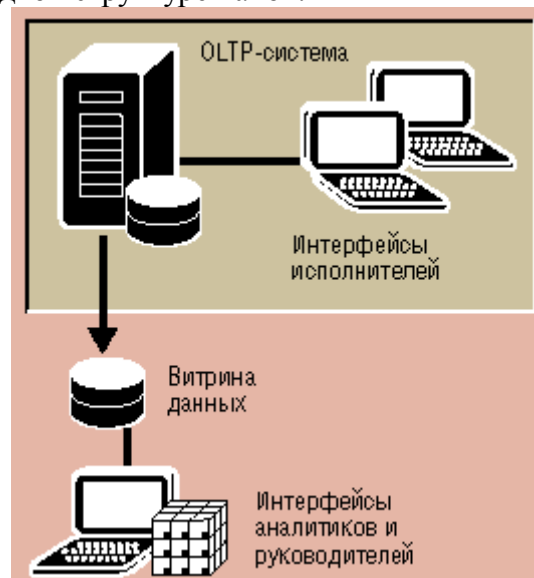
В этом случае аналитическое приложение для тиражируемой OLTP-системы состоит из трех элементов: программы выгрузки данных (она может быть разработана на языке программирования, встроенном в OLTP-систему), Витрины данных и OLAP-клиента, настроенного на Витрину данных.

Это решение может поставляться как самостоятельное приложение.

Витрины данных также создаются для уменьшения времени отклика при работе с очень большим Хранилищем данных. Для этого некоторые тематические данные регулярно выгружаются из него в многомерные или реляционные Витрины данных для использования группой специалистов.

Интеграция данных на рабочем столе пользователя. Важное достоинство большинства OLAP-клиентов - возможность интеграции данных из различных источников в одном приложении. Конечный пользователь будет работать с данными, физически расположенными в разных информационных системах предприятия, из единого интерфейса и единым образом, что очень удобно.

Допускается настройка одного пользовательского интерфейса OLAP-клиента сразу на несколько баз данных различных типов (SQL-серверы, локальные таблицы) и различных систем (бухгалтерская система, Хранилище данных, Витрина данных). Для этого в Business-Objects реализован специальный сервер BI-портал InfoView. MicroStrategy и "Контур Стандарт" могут отображать список отчетов независимо от их источников данных, в одной древовидной структуре папок.



"Архитектурная гибкость" OLAP-клиентов открывает неограниченные возможности для реализации идей профессиональных разработчиков, внедренческих компаний и ИТ-подразделений предприятий. Аккуратный анализ состава и структуры первичных данных, реальных запросов пользователей, имеющегося парка вычислительной техники и выбор на этой основе оптимального OLAP-клиента и архитектуры информационной системы может помочь сократить расходы и при этом удовлетворить потребности пользователей.

Для разработчиков будет полезно принять во внимание тенденции в OLAP-технологиях. Средний человек практически не в состоянии анализировать отчеты, имеющие свыше семи измерений и более двух-трех десятков элементов в одном измерении. Таким образом, объемы данных для "человеческого" анализа являются константой, а мощность компьютеров, в том числе персональных, постоянно растет. Уже сегодня OLAP-сервер исчезает как отдельный продукт и превращается в стандартный сервис SQL-сервера. Дальнейшее повышение мощности ПК может привести к тому, что ее станет достаточно для выполнения OLAP-анализа подавляющего большинства приложений баз данных.