

Дисциплина
«Информационные технологии в
экономике»

**Лекция-28 Обработка данных и средства
создания отчетов**

Исучаемые вопросы

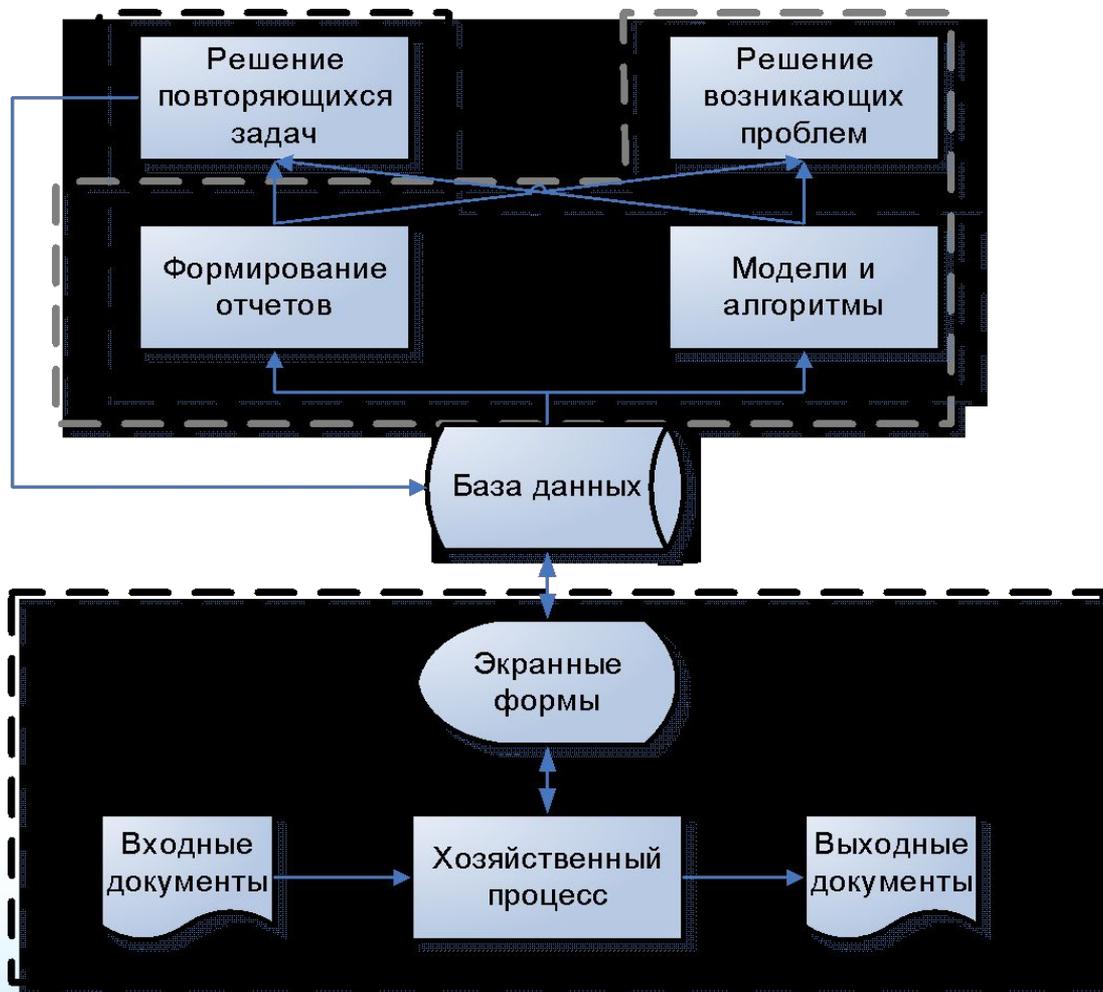
1. Режимы и виды аналитической обработки данных. Семантический слой для создания отчетов в терминах предметной области.
2. Отчеты и варианты их создания.
3. Средства создания отчетов.

Вопрос-1

Режимы и виды аналитической обработки данных. Семантический слой для создания отчетов в терминах предметной области.

Типы экономических информационных систем

Экономическая информационная система



Режимы обработки данных

- В области информационных технологий, всегда существовали два взаимодополняющих друг друга режима обработки данных:
 - режим, ориентированный на оперативную (операционную, транзакционную) обработку данных;
 - режим, ориентированный на анализ данных.
- На первых этапах автоматизации требуется упорядочить процессы повседневной рутинной обработки (переработки) данных, на что и ориентированы транзакционные информационные системы.
- СППР являются в определенном смысле вторичными, по отношению к ним. Здесь возможна аналогия с производством. Любая продукция, прежде чем попасть на склад и быть отгружена потребителю, должна быть сначала произведена. И прежде чем заниматься анализом данных, необходимо эти данные иметь (произвести).
- Наиболее популярные концепции построения аналитических систем:
 - Хранилища Данных (Data Warehouse);
 - Витрин данных (Data Mart),
 - On-Line Analytical Processing (OLAP).

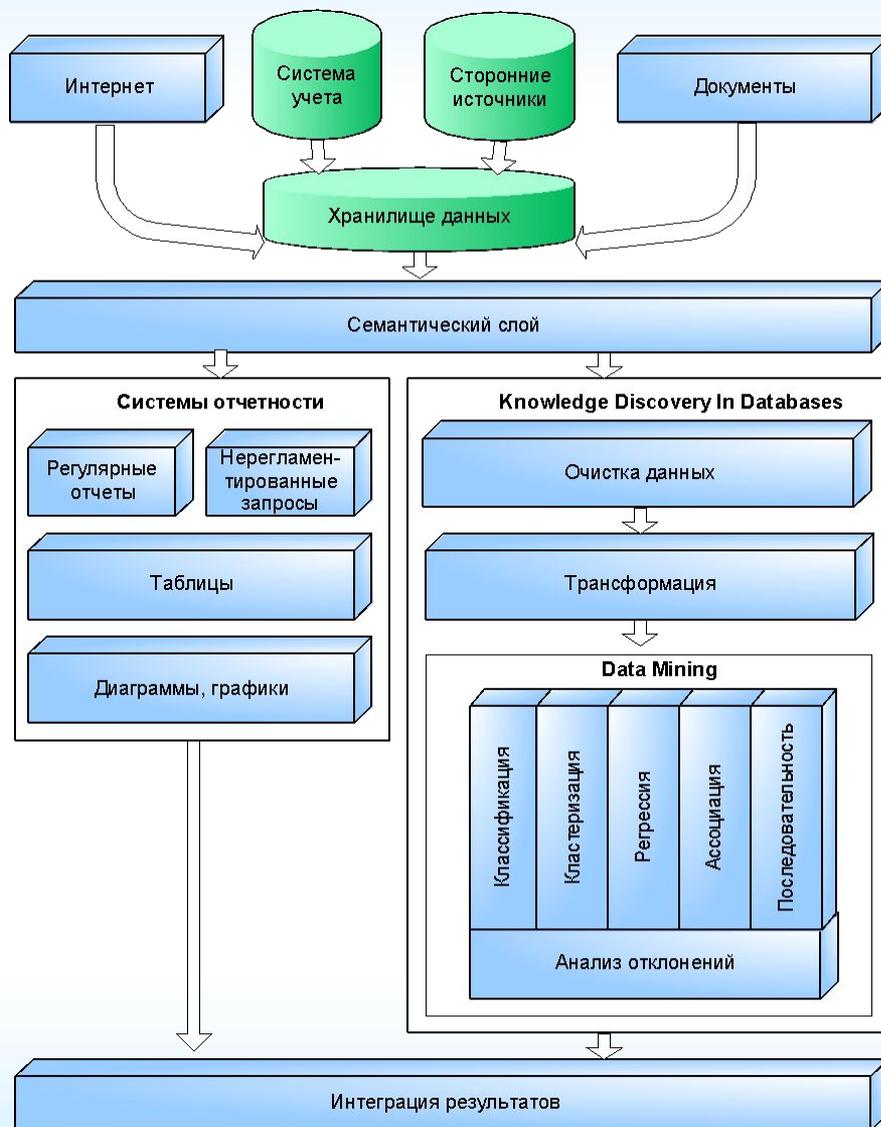
Режимы обработки данных

Режим оперативной обработки данных

Режим оперативной обработки данных или транзакций OLTP (On-Line Transaction Processing) применяется в информационных системах для сбора, регистрации, ввода исходных данных, относящихся к той или иной предметной области, первичной обработки данных, их хранения, адекватной визуализации, поиска, выдачи справок и отчетных материалов.

- Данные в информационную систему вводятся либо с документа, имеющего определенную правовую силу, либо непосредственно с места возникновения данных. В последнем случае документ, содержащий введенные данные, печатается системой и ему придается правовая сила.
- В OLTP системах над целевыми БД исполняются целевые транзакции (например, занесение в таблицу записи с параметрами выписанного счета, оприходованной фактуры или любого другого факта), которые изменяют состояние БД и приводят их в соответствие текущему состоянию того фрагмента реального мира, который моделирует БД.
- Обычно аналитические возможности OLTP-систем сильно ограничены, они используются для того, чтобы способствовать повседневной деятельности корпорации, и опираются на актуальные для текущего момента данные.

Режимы обработки данных



Режим анализа данных

- аналитической обработки управленческой информации (**системы отчетности**),
- подготовки аналитической отчетности (**Knowledge Discovery, Data Mining - добыча знаний**),
- экспертизы и анализа решений (**интеграция результатов**).

Сферы анализа

Поддержка принятия управленческих решений осуществляется в следующих сферах:

- сфера детализированных данных;
 - сфера агрегированных показателей;
 - сфера закономерностей.
-
- В *сфере детализированных данных* подсистемы ИАС или автономные ИС нацелены на поиск данных. Эту задачу отлично выполняют реляционные СУБД. В качестве языка манипулирования данными, ставшего стандартным, используется, как правило, SQL. Для поиска детализированной информации используются информационно-поисковые системы, которые могут работать с операционными, локальными или региональными базами и хранилищами данных, так и совместно с центральным ИХ.
 - *Сфера агрегированных данных* отличается агрегацией данных, оперативной аналитической обработкой, многомерным представлением в виде гиперкубов, многомерным анализом. В этой сфере используются специальные многомерные СУБД. Допустимо использование реляционных представлений данных. При правильном применении реляционных СУБД, как показано выше, показатели эффективности сопоставимы со специализированными многомерными. Агрегированные массивы при реляционном подходе представлены в виде схем: «звезды» и других. Агрегация может производиться также «на лету» при обработке запроса.
 - *Сфера закономерностей* связана или основана на интеллектуальной обработке данных. Главной задачей здесь является выявление закономерностей в исследуемых процессах, взаимосвязей и взаимовлияния различных факторов, поиск крупных «непривычных» отклонений, прогноз хода различных существенных процессов. Эта сфера относится к интеллектуальному анализу (Data mining).

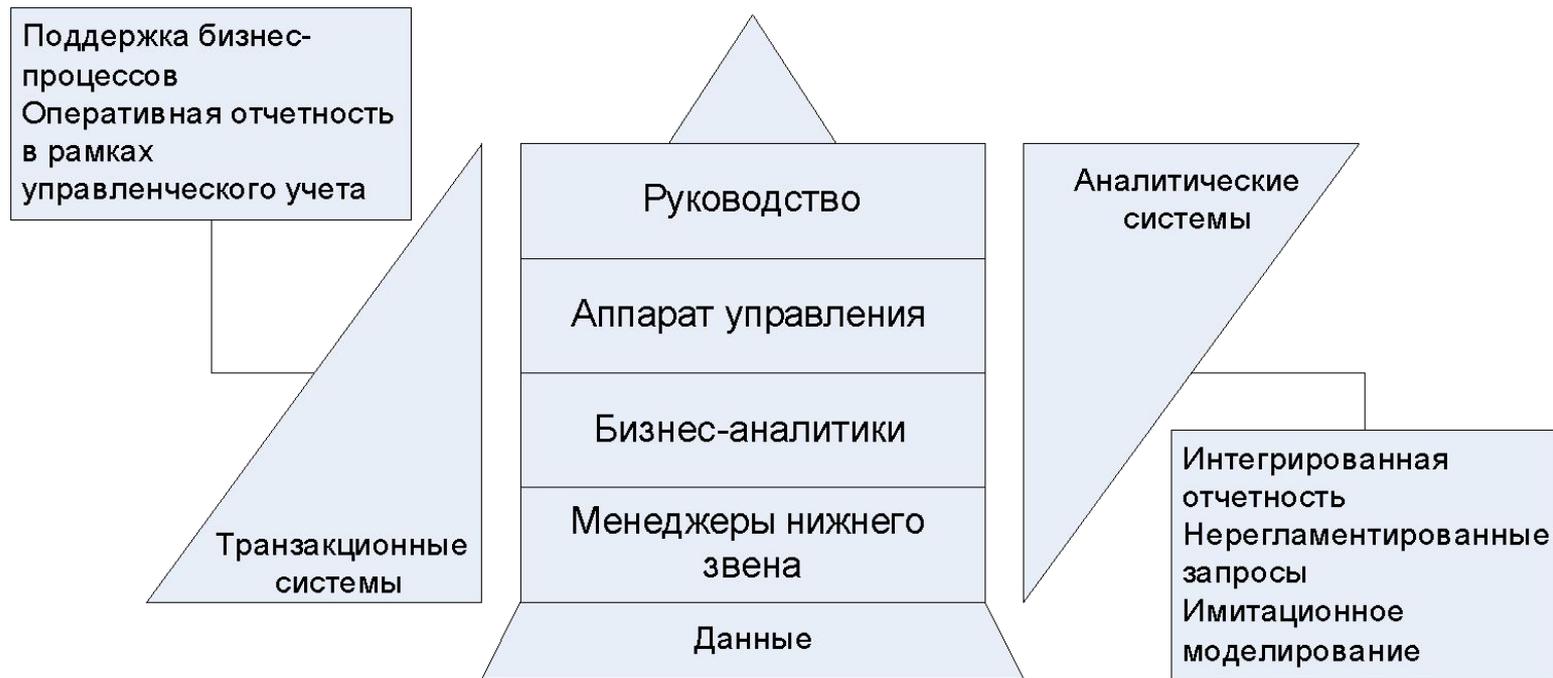
Режимы обработки данных

Сравнение характеристик данных в информационных системах ориентированных на операционную и аналитическую обработку данных

Характеристика	Операционные	Аналитические
Частота обновления	Высокая частота, маленькими порциями	Малая частота, большими порциями
Источники данных	В основном внутренние	В основном внешние
Объемы хранимых данных	Сотни мегабайт, гигабайты	Гигабайты и терабайты
Возраст данных	Текущие (за период от нескольких месяцев до одного года)	Текущие и исторические (за период в несколько лет, десятки лет)
Назначение	Фиксация, оперативный поиск и преобразование данных	Хранение детализированных и агрегированных исторических данных, аналитическая обработка, прогнозирование и моделирование

Режимы обработки данных

Сравнение характеристик данных в информационных системах ориентированных на операционную и аналитическую обработку данных



Виды аналитической обработки данных

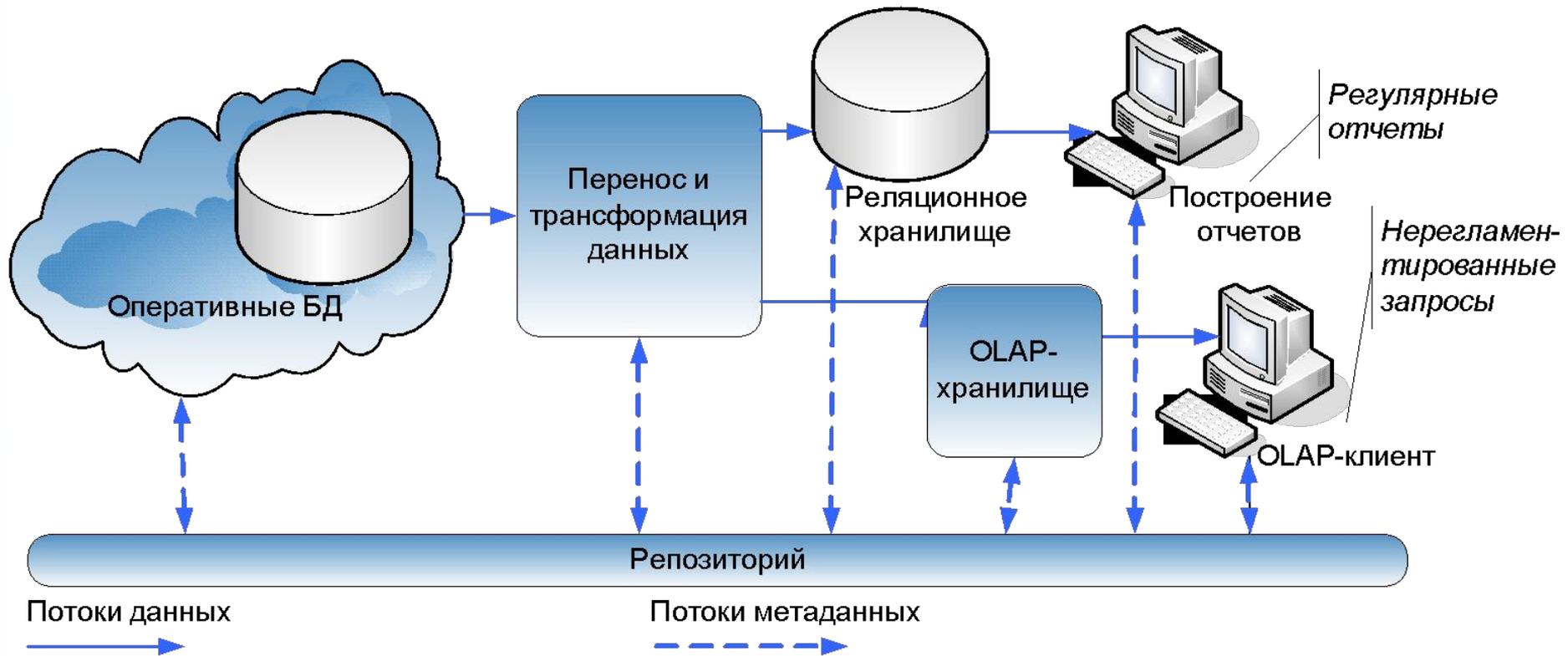
Выделяются два вида аналитической обработки данных

- статическая обработка данных;
- динамическая обработка данных.

Вопрос-2

Отчеты и варианты их создания.

Отчеты и варианты их создания

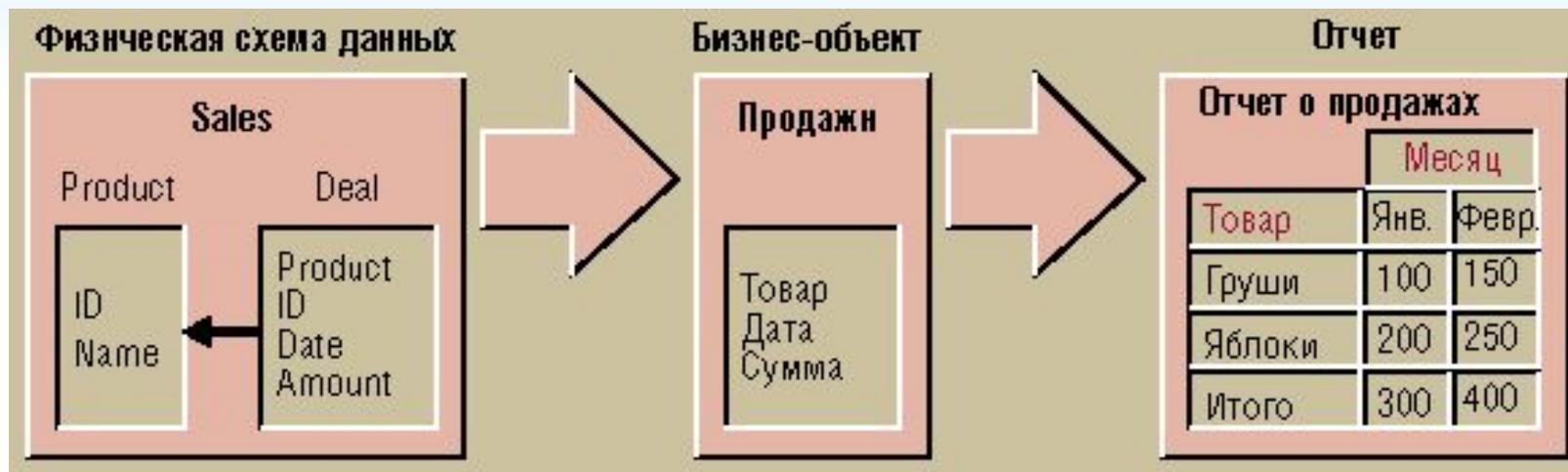


Виды аналитической обработки данных

Сравнение характеристик статического и динамического анализа

Характеристика	Статический анализ	Динамический анализ
Типы вопросов	Сколько? Как? Когда?	Почему? Что будет если?
Время отклика	Не регламентируется	Секунды
Типичные операции	Регламентированный отчет, диаграмма	Последовательность интерактивных отчетов, диаграмм, экранных форм. Динамическое изменение уровней агрегации и срезов данных.
Уровень аналитических требований	Средний	Высокий
Тип экранных форм	В основном определенный заранее, регламентированный	Определяемый пользователем
Уровень агрегации данных	Детализированные и суммарные	В основном суммарные
Возраст данных	Исторические и текущие и прогнозируемые	Исторические, текущие и прогнозируемые
Типы запросов	В основном предсказуемые	Непредсказуемые, от случая к случаю
Назначение	Регламентированная аналитическая обработка	Многопроходный анализ, моделирование и построение прогнозов

Семантический слой для создания отчетов в терминах предметной области



Общий принцип создания отчетов — 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 отчета. Программа-генератор отчета автоматически формирует SQL-запрос к реляционной базе данных 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.

Отчеты и варианты их создания

Понятие отчета

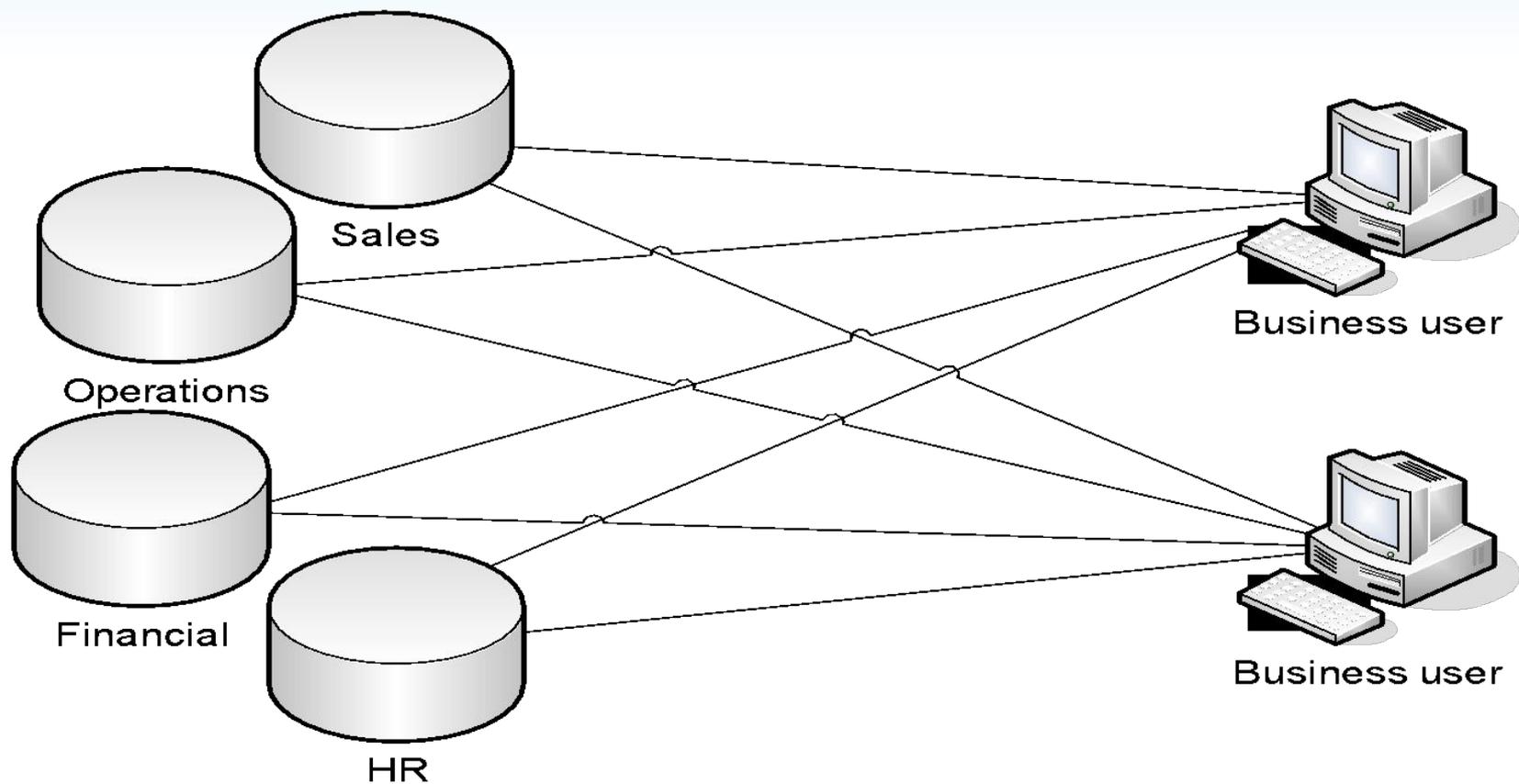
Отчет представляет собой документ, содержимое которого динамически формируется на основе информации, содержащейся в базе данных.

Такой документ может состоять из различных частей:

- колонтитулов, повторяющихся на каждой странице;
- заголовка, находящегося в начале отчета;
- областей, в которых располагаются фактические данные;
- областей, в которых расположены агрегатные данные (суммы, средние и т.д.);
- и др.

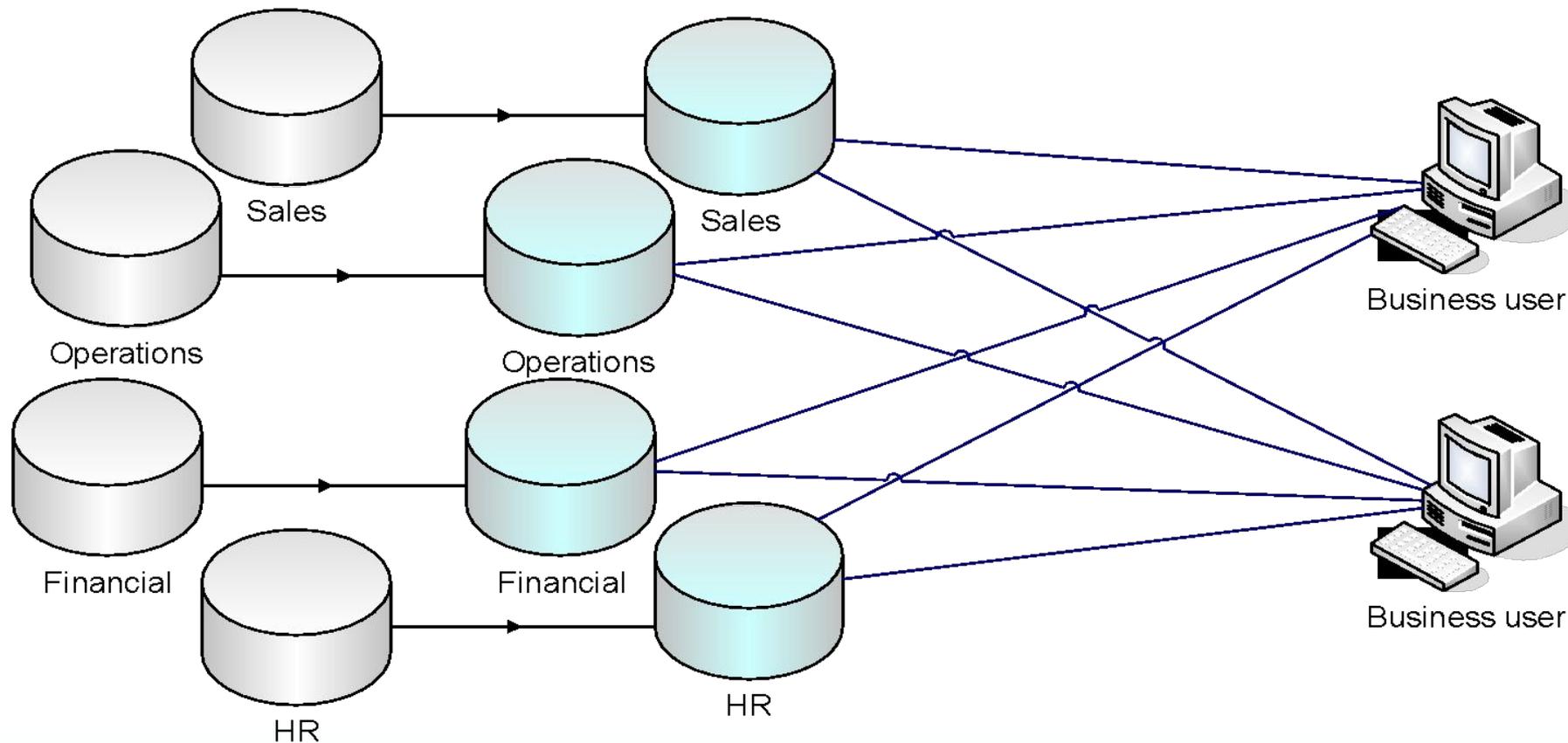
Архитектуры для создания отчетов

Отчетность на основе учетных систем (OLTP)



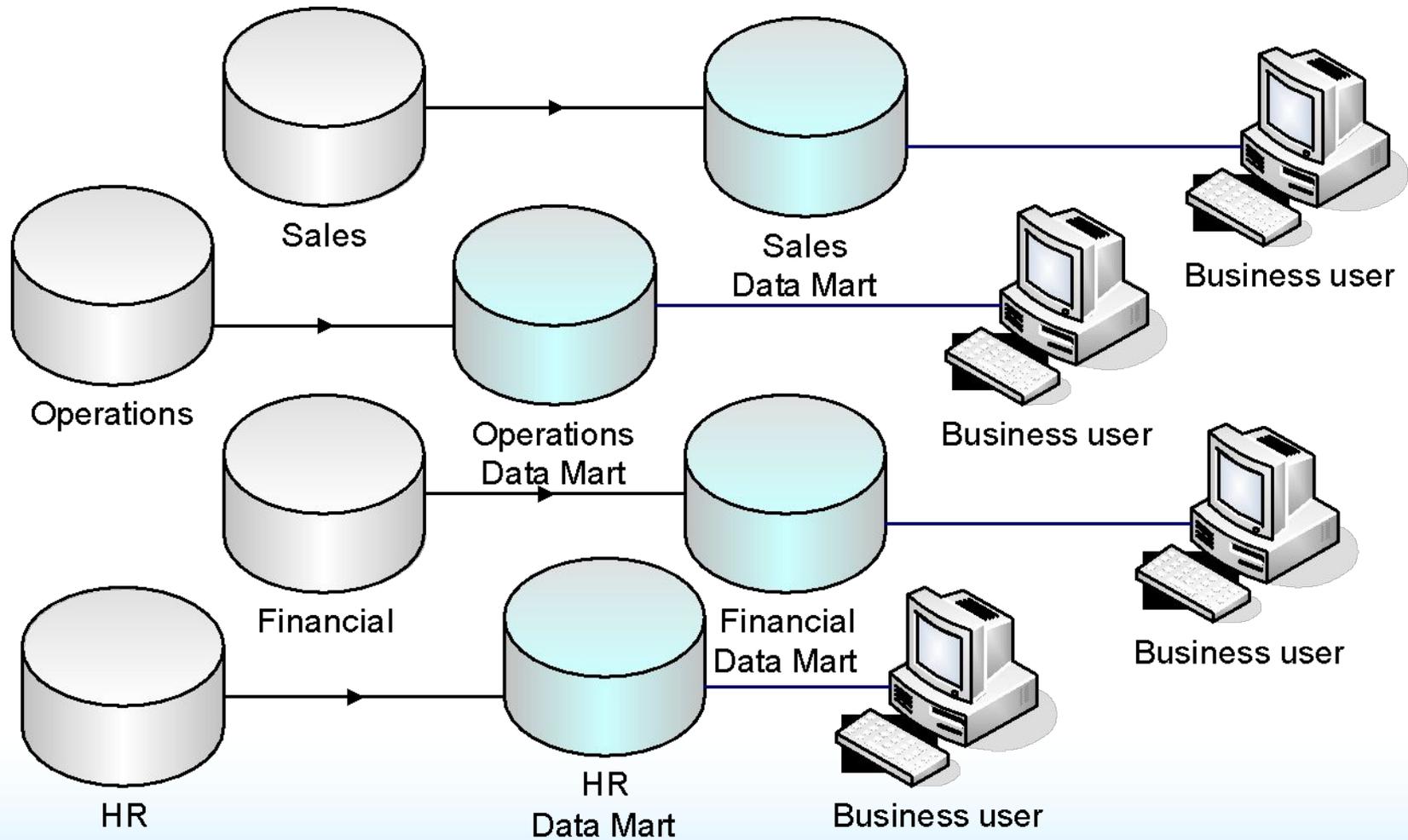
Архитектуры для создания ОТЧЕТОВ

Отчетность по дублированной базе учетной системы (репликация)



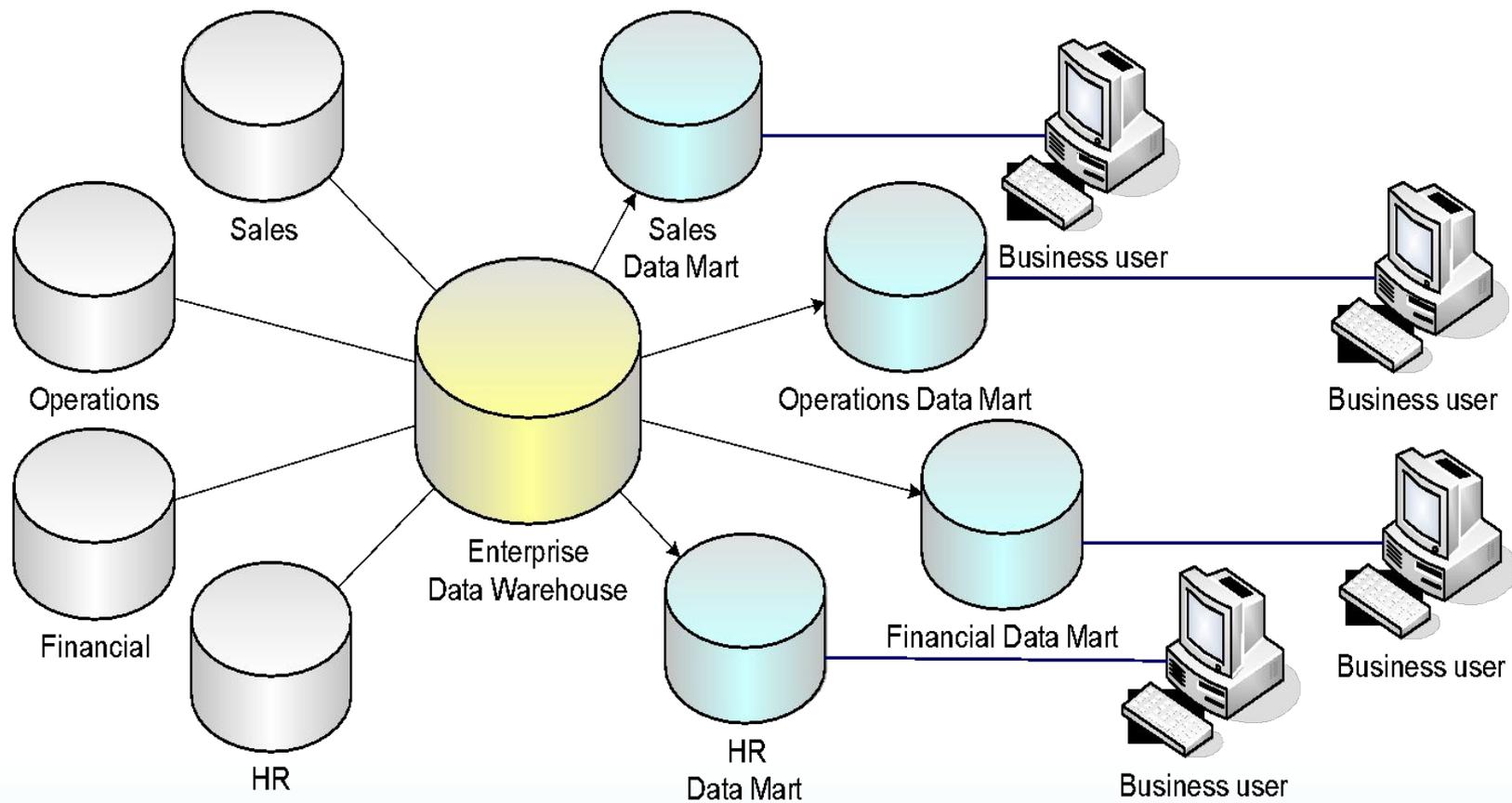
Архитектуры для создания отчетов

Отчетность с использованием витрин данных



Архитектуры для создания ОТЧЕТОВ

Отчетность по Хранилищу данных



Вопрос-3

Средства создания отчетов.

Средства создания отчетов

Специализированные средства создания отчетов

Отчеты создаются с помощью специализированных средств создания отчетов, которые характеризуются следующими особенностями:

- поддержка различных механизмов доступа к данным;
- наличие мастеров и визуальных инструментов, облегчающих рядовым пользователям подготовку отчетов, в том числе достаточно сложных;
- высокоточные инструменты позиционирования элементов отчета при печати;
- наличие средств деловой графики, а иногда простейших встроенных геоинформационных систем;
- интеграция с офисными приложениями. Как правило, поддерживается не только экспорт отчетов в документы этих приложений, но нередко и доступ к генератору отчетов из среды этих приложений с помощью модулей расширений;
- поддержка публикации данных в Internet. Обычно поддерживается не только экспорт отчета в формат HTML, но и почти точное воспроизведение в браузере всех элементов отчета (чаще с помощью DHTML и применения графики), а нередко и генерация отчетов на сервере с использованием ASP-кода;
- интеграция с наиболее популярными средствами разработки. В комплект поставки версий, предназначенных для разработчиков приложений, обычно входят библиотеки, предоставляющие интерфейс к Run-time-версиям генератора отчетов. В последнее время в комплект поставки принято также включать элементы управления ActiveX, использующие эти же библиотеки, иногда включаются и другие средства интеграции с приложениями, например серверные компоненты;
- встроенные языки программирования для создания формул;
- возможность создания аналитических отчетов (то есть отчетов, состоящих из агрегатных данных).

Средства создания отчетов

Генераторы отчетов, встроенные в средства разработки

- Как правило, инструменты подобного класса обладают гораздо меньшими возможностями, нежели специализированные средства создания отчетов. Они могут, например, поддерживать не все механизмы доступа к данным или не иметь высокоточных инструментов позиционирования элементов отчета, не поддерживать (или ограниченно поддерживать) экспорт отчета в файлы других форматов. Дело в том, что назначение этих генераторов отчетов несколько иное, чем у специализированных инструментов: они применяются, когда основными требованиями к печатному документу, генерируемому приложением или СУБД, является скорость и простота его получения, а не высокая точность печати или сложность алгоритма вычисления печатаемых данных.
- Специализированные генераторы отчетов, как правило, требуют отдельной установки и обычно нуждаются в соединении с базой данных, отличной от соединения, используемого приложением. Последнее требование нередко влечет за собой приобретение дополнительных клиентских лицензий. Встроенные же генераторы отчетов часто позволяют обойтись тем же самым соединением с базой данных, которое используется приложением; в этом случае дополнительная клиентская лицензия такому приложению не требуется.

Средства создания отчетов

«Нетрадиционные» средства создания отчетов

- Нередко в качестве средств создания отчетов применяются приложения, которые, с одной стороны, имеют высококачественные средства управления печатью документов или конвертирования их в различные форматы, а с другой стороны, являются серверами автоматизации, предоставляющими доступ к этим возможностям с помощью своих объектных моделей. Наиболее часто в качестве таких генераторов отчетов используются приложения Microsoft Office, обычно Microsoft Word и Microsoft Excel.
- К преимуществам создания отчетов с помощью Microsoft Office относится
 - определенная гибкость при выборе механизма доступа к данным: можно применять и механизмы доступа к данным самого Office (ODBC и OLE DB), и такие механизмы доступа к данным, которые поддерживаются средством разработки, применяемым при создании такого приложения;
 - при генерации печатного документа можно производить сколь угодно сложные вычисления и преобразования данных (так, как это делается внутри приложения, создающего такой документ) и к тому же записывать макросы, сохраняемые в дальнейшем в шаблоне будущего отчета и также позволяющие реализовать ту или иную последовательность действий.
- Наиболее часто для создания таких контроллеров автоматизации применяется Visual Basic или VBA; в справочной системе Microsoft Office имеются файлы, содержащие подробное описание объектных моделей приложений Office с примерами их использования.

Средства создания отчетов

OLAP-средства

- Информационные системы масштаба предприятия содержат приложения, применяемые менеджерами высшего звена и предназначенные для комплексного многомерного анализа данных, их динамики, тенденций и т.п. Такой анализ в конечном итоге призван способствовать принятию решений. Указанные приложения обычно обладают средствами предоставления пользователю агрегатных данных для различных выборок из исходного набора в удобном для восприятия и анализа виде. Чаще всего такие агрегатные функции образуют многомерный (а следовательно, нереляционный) набор данных (нередко называемый гиперкубом или метакубом), оси которого содержат параметры, а ячейки — зависящие от них агрегатные данные. Вдоль каждой оси данные могут быть организованы в виде иерархии, отражающей различные уровни их детализации. Благодаря такой модели данных пользователи могут формулировать сложные запросы, генерировать отчеты, получать подмножества данных.
- Технология комплексного многомерного анализа данных получила название OLAP (On-Line Analytical Processing).
- В настоящее время поддержка OLAP реализована во многих СУБД и инструментах, так как является оптимальным решением для большого класса приложений, где пользователи сталкиваются с многомерными данными (то есть с данными, зависящими от нескольких параметров, например от времени, местоположения и других характеристик).

Средства создания отчетов

Клиентские OLAP-средства

- Клиентские OLAP-средства представляют собой приложения, осуществляющие многомерный анализ данных, содержащихся в какой-либо СУБД, и вычисление агрегатных данных (сумм, средних величин, максимальных или минимальных значений) внутри своего адресного пространства. Если исходные данные содержатся в настольной СУБД, то вычисление агрегатных данных производится непосредственно самим OLAP-средством; если же источник исходных данных представляет собой серверную СУБД, то многие из клиентских OLAP-средств так или иначе используют возможности сервера баз данных для вычисления агрегатных данных (например, посылая на сервер SQL-запросы, содержащие оператор GROUP BY, — результат такого запроса содержит агрегатные данные, вычисленные на сервере).
- Также следует сказать, что клиентские OLAP-средства применяются обычно при небольшом числе параметров, используемых для анализа (обычно их не рекомендуется иметь более шести), и небольшом разнообразии значений этих параметров, поскольку полученные агрегатные данные должны размещаться в адресном пространстве такого средства и при увеличении числа параметров объем занимаемой данными оперативной памяти увеличивается экспоненциально. Именно поэтому даже самые примитивные клиентские OLAP-средства обычно содержат инструменты предварительного подсчета объема оперативной памяти, который потребуется при создании и отображении многомерного куба.

Средства создания отчетов

Серверные OLAP-средства

- Работа серверных OLAP-средств организована подобно работе реляционных серверных СУБД: хранение агрегатных данных и поддержка содержащего их хранилища осуществляются отдельным приложением или в ходе процесса, называемого OLAP-сервером. Одни клиентские приложения могут обращаться с запросами к такому хранилищу и в ответ на них получать либо те или иные подмножества из хранимой совокупности агрегатных данных, либо значения, вычисляемые на их основе. Другие приложения могут создавать указанные хранилища или обновлять их в соответствии с изменившимися исходными данными.
- Многие клиентские OLAP-средства позволяют обращаться к серверным OLAP-хранилищам, выступая в этом случае в роли клиентских приложений, выполняющих подобные запросы. Так, Microsoft Excel может обращаться к многомерному хранилищу данных (если для него имеется OLE DB-провайдер, удовлетворяющий спецификации OLE DB for OLAP), в частности к данным Microsoft SQL Server 7.0 OLAP Services и Microsoft SQL Server 2000 Analysis Services, и отображать подмножества данных, хранимых на этих OLAP-серверах (например, строить сечения многомерных кубов и отображать их в виде сводных таблиц и диаграмм).
- Преимущества применения серверных OLAP-средств по сравнению с клиентскими OLAP-средствами сходны с преимуществами применения серверных СУБД по сравнению с настольными, так как создание, пересчет и хранение агрегатных данных происходят на сервере, а клиентское приложение получает лишь результаты запросов к ним; трафик, время выполнения запроса и потребляемая оперативная память в общем случае оказываются намного ниже, чем в случае использования клиентских OLAP-средств с теми же исходными данными. Поэтому при большом числе параметров, используемых для анализа, при большом разнообразии их значений и большом объеме исходных данных следует применять серверные OLAP-средства. Как правило, средства анализа и обработки данных масштаба предприятия базируются именно на серверных OLAP-средствах.

Средства создания отчетов

Средства поиска закономерностей (Data Mining-средства)

- Корпоративная база данных предприятия обычно содержит набор таблиц, хранящих записи о тех или иных фактах или объектах. Однако совокупность таких записей, накопленных за несколько лет, может служить источником дополнительной информации, которую нельзя получить на основе одной конкретной записи, а именно сведений о закономерностях, тенденциях или взаимозависимости между какими-либо данными. Подобного рода информация обычно используется при прогнозировании и стратегическом планировании, и ценность ее для предприятия очень высока. Видимо, поэтому процесс ее поиска и получил название Data Mining (mining по-английски означает добычу полезных ископаемых, а поиск закономерностей в огромном наборе фактических данных действительно близок к этому). Другими словами, термин Data Mining реально обозначает не столько технологию, сколько сам процесс поиска корреляций, тенденций, взаимосвязей и закономерностей посредством различных математических и статистических алгоритмов — кластеризации, создания выборок, регрессионного и корреляционного анализа. Цель этого поиска — представить данные в таком виде, чтобы стало возможным понять поведение бизнес-процессов, а также построить модель, с помощью которой можно прогнозировать процессы, играющие роль при планировании бизнеса (например, динамику спроса на те или иные товары или услуги).
- Современные средства поиска закономерностей, корреляций и тенденций могут осуществлять его как в реляционных базах данных, так и в OLAP-кубах.

BI-средства масштаба предприятия и средства разработки BI-приложений

- К современным средствам Business Intelligence относятся не только генераторы отчетов и OLAP-средства, но и средства разработки BI-приложений (BI Platforms), а также средства анализа и обработки данных масштаба предприятия, позволяющие осуществлять комплекс действий, связанных с анализом данных и созданием отчетов (Enterprise BI Suites). Последние, как правило, содержат в своем составе и средства построения отчетов, и OLAP-средства, а нередко и Data Mining-средства.

Место средств отчетности в архитектуре информационно-аналитической системы на примере АК «Прогноз-5»

