

© Алексеева Т.В.

© Московский финансово-промышленный университет «Синергия»

Содержание

Аннотация

Тема 1. Отчетность и анализ – функции распознавания ситуации в процессе принятия решения

Вопрос 1. Формирование аналитической отчетности и цикл принятия решения.

Вопрос 2. Показатели измерения результативности работы предприятия (KPI).

Вопрос 3. Единое информационное пространство предприятия – основа аналитической деятельности предприятия.

Вопрос 4. Отображение показателей отчетности в информационном пространстве предприятия.

Вопрос 5. Компонентная архитектура программных средств хранения данных и их аналитической обработки в Business Intelligens.

Вопросы для самопроверки:

Литература по теме:

Тема 2. Источники данных и хранение информации на предприятии

Вопрос 1. Способы хранения информации на предприятии и источники данных.

Вопрос 2. Принципы построения хранилищ и OLTP баз данных.

Вопрос 3. Свойства и структура хранилищ данных.

Вопрос 4. Виды хранилищ данных.

Вопрос 5. Технология работы хранилищ данных.

Вопрос 6. Рынок DWH.

Вопросы для самопроверки:

Литература по теме:

Тема 3. Инструменты интеграции данных из различных источников

Вопрос 1. Источники данных для формирования отчетности.

Вопрос 2. Технологии интеграции данных.

Вопрос 3. Рынок средств интеграции приложений.

Вопросы для самопроверки:

Литература по теме:

Тема 4. Методология оперативной аналитической обработки данных (OLAP)

Вопрос 1. Специфика оперативной аналитической обработки данных.

Вопрос 2. Требования Кодда к средствам оперативной аналитической обработки.

Вопрос 3. Сферы применения OLAP-технологий.

Вопросы для самопроверки:

Литература по теме:

Тема 5. Инструментальные средства оперативной аналитической обработки данных (OLAP)

Вопрос 1. Классификация и структура OLAP решений.

Вопрос 2. Принципы работы OLAP клиентов.

Вопрос 3. Выбор архитектуры OLAP-решения.

Вопрос 4. Рынок OLAP решений.

Вопросы для самопроверки:

Литература по теме:

Тема 6. Интеллектуальный анализ данных (Data mining)

Вопрос 1. Назначение интеллектуального анализа данных и примеры его применения в бизнесе.

Вопрос 2. Технологические этапы проведения интеллектуального анализа данных.

Вопрос 3. Методы интеллектуального анализа данных.

Вопрос 4. Программные средства интеллектуального анализа данных.

Вопрос 5. Интеграция оперативного и интеллектуального анализа данных.

Вопросы для самопроверки:

Литература по теме:

Тема 7. Информационные системы управления эффективностью бизнеса (BPM) и системы Бизнес-интеллекта (BI)

Вопрос 1. Информационные системы бизнес интеллекта и управления эффективностью бизнеса.

Вопрос 2. Определение BPM и BI.

Вопрос 3. Технологические этапы цикла управления эффективностью бизнеса (BPM).

Вопрос 4. KPI в контексте BPM.

Вопросы для самопроверки:

Литература по теме:

Тема 8. Информационно-методический комплекс управления эффективностью бизнеса (BPM) на предприятии

Вопрос 1. Архитектура, функциональные возможности и компоненты информационных систем управления эффективностью бизнеса.

Вопрос 2. Технология внедрения BPM.

Вопрос 3. Характеристика рынка инструментальных средств бизнес интеллекта, управления эффективностью бизнеса.

Вопросы для самопроверки:

Литература по теме:

Аннотация

Дисциплина «Информационно-аналитические системы» направлена на развитие информационных компетенций,

необходимых будущим специалистам в любой предметной области. Для эффективного принятия управленческих решений необходимо оперативно отслеживать быстро меняющиеся данные, рассматривать их с разных сторон, получать сводные данные, а при необходимости углубляться в содержимое этих данных для получения более детализированной информации, составлять различные отчеты буквально на лету. Для развития этих компетентностей аналитикам не только необходимо знать современные методы и средства анализа данных, но и уметь использовать возможности информационных технологий для поддержки аналитической работы на предприятии.

Настоящая дисциплина включена в учебные планы Университета по всем программам подготовки по направлениям Информационные системы и технологии, Математическое обеспечение и администрирование информационных систем, Экономика. Для изучения дисциплины требуются знания и навыки студентов по базовым экономическим курсам и дисциплинам специальности (предметная область), а также по курсам «Информатика» и «Информационные технологии». Знания по дисциплине «Информационно-аналитические системы» (ИАС) могут использоваться в курсах, связанных с обоснованием, поддержкой и принятием управленческих решений в процессе осуществления финансово-хозяйственной деятельности предприятия, а также в курсах «Технологии Больших данных (BigData)», «Управление данными», «Системы управления эффективностью бизнеса», «Менеджмент», «Маркетинг», «Финансовый анализ» и др.

Цель и задачи дисциплины.

Целью изучения дисциплины «Информационно-аналитические системы» (ИАС) является ознакомление студентов с информационными технологиями поддержки аналитической работы и их использованием на предприятии, освещение общих основ анализа информации и применения информационно-аналитических систем (как универсальных, так и специализированных).

Задачи дисциплины:

- сформировать общее представление о содержании и особенностях аналитической работы по подготовке и обоснованию принимаемых в процессе осуществления финансово-хозяйственной деятельности предприятия решений;
- передать знания о технологиях реализации информационного анализа и их применении;
- привить навыки применения работ в информационно-аналитических систем.

Входные компетенции:

- знать состав экономической и управленческой информации в предметной области, основные бизнес-процессы, характерные для конкретной предметной области, информационные технологии, используемые в различных предметных областях;
- уметь сформулировать требования к составу информации необходимой для принятия решений в конкретной предметной области;
- обладать навыками подготовки документов с использованием пакета Microsoft Office.

Выходные компетенции:

- знать состав и структуру информационного пространства предприятия, общую структуру и принципы работы информационно-аналитической системы, основные методы и технологии анализа экономической и управленческой информации: OLAP-технологии и технологии Data Mining;
- уметь выбирать и использовать необходимые методы и программные средства анализа и подготовки информации для поддержки принятия решений, составлять необходимые отчеты в ходе проведения OLAP-анализа;
- обладать навыками анализа информации с помощью инструментальных средств широкого применения (Excel, Mathcad и др.), получения информации (данных) и проведения аналитических работ с применением специализированных инструментальных средств Business Intelligence.

Тема 1. Отчетность и анализ – функции распознавания ситуации в процессе принятия решения

Цели:

Сформировать представление об аналитической деятельности и едином информационном пространстве на предприятии.

Задачи:

- Изучить цикл принятия решения и задачи анализа информации на каждом этапе цикла.
- Рассмотреть процесс формирования дерева целей и выбора показателей результативности.
- Изучить процесс формирования информационного пространства для анализа деятельности предприятия.

- Познакомиться с компонентной архитектурой программных средств Business Intelligens.

Вопросы темы:

1. Формирование аналитической отчетности и цикл принятия решения.
2. Показатели измерения результативности работы предприятия (KPI).
3. Единое информационное пространство предприятия – основа аналитической деятельности предприятия.
4. Отображение показателей отчетности в информационном пространстве предприятия.
5. Компонентная архитектура программных средств хранения данных и их аналитической обработки в Business Intelligens.

Основные понятия:

- *система поддержки принятия решений (СППР);*
- *DM-технологии;*
- *коэффициент приоритетности целей (КПЦ);*
- *ключевые показатели эффективности – KPI;*
- *информация;*
- *информационное пространство;*
- *информационные ресурсы (ИР);*
- *реквизит;*
- *составная единица информации;*
- *показатель;*
- *информационно-аналитическая система;*
- *бизнес-аналитика;*
- *Business Intelligens (BI).*

Вопрос 1. Формирование аналитической отчетности и цикл принятия решения.

Внедрение информационных технологий в процесс производства и управления изменяет традиционные взгляды на иерархические организационно-экономические структуры. Происходит становление новой модели управления, ориентированной на интеграционные процессы компаний, функционирующих на основе процессного и функционального управления.

Для эффективного принятия решения руководителю необходима тщательно подобранная и хорошо обработанная информация. Существуют различные методы обработки информации:

- методы поддержки хранения больших пополняющихся объемов информации (Big Data) и Систем хранения данных (СХД);
- методы предоставления компьютерных рассуждений;
- методы компьютерной аппроксимации антропоморфных аспектов умственной деятельности (когнитивная графика, эвристические методы, формализация поиска релевантного знания в процессе рассуждений и т.д.).

Цели, ради достижения которых принимается решение, можно разделить на три класса:

- 1) стратегические;
- 2) тактические;
- 3) оперативные.

Взаимосвязь между целями и типами решений и задачами можно представить с помощью рис. 1.



Рис. 1. Цели и типы решений

Оперативные решения периодические: одна и та же задача возникает снова и снова. Принятие *оперативных решений* ведет к вполне ожидаемым и прогнозируемым результатам. Например, если товары заказаны на склад, то существует высокая вероятность их поступления. Оперативные решения являются краткосрочными.

Тактические решения обычно принимаются управленцами среднего уровня, ответственными за обеспечение средствами для достижения целей и намерений, поставленных руководителями верхнего звена. «При каких условиях давать скидку заказчику?» – это пример тактического решения, принимаемого на среднем уровне управления.

Стратегические решения принимаются с учетом целей компании, определенных в ее уставе и уточненных высшим руководством предприятия. Эти цели определяют основу, на которой должно базироваться долгосрочное планирование, а также устанавливают критические факторы деятельности предприятия. Эти решения обеспечивают базу для принятия тактических и оперативных решений.

Процедуры поддержки принятия решения должны обеспечивать поддержку для выполнения следующих этапов:

1. Выработка стратегической цели.
2. Констатация ситуации.
3. Анализ (распознавание проблем (ситуации)).
4. Выработка целей и генерация альтернатив.
5. Выбор решения: выбор критерия (и его обоснование) и оценка альтернатив по критерию.
6. Анализ последствий принятия решения.
7. Принятие решения и его мониторинг.

Характеристика этапов.

Этап 1. Выработка стратегической цели. Компьютерная поддержка на этом этапе незначительна и возможна лишь в части консультаций или демонстрации аналогичных решений в иных организациях. Необходима стратегическая аналитическая информация.

Этап 2. Констатация ситуации. Этап формирования решения предназначен для выявления и констатации ситуации, в которой находится предприятие. Для этого, в зависимости от уровня принимаемого решения, используется обычная отчетность и аналитическая информация, поступающая с различных уровней систем поддержки принятия решений (СППР).

Этап 3. Анализ. Этот этап один из самых трудоемких и ответственных. Процедуры, выполняемые здесь, делятся на два класса: формальные и неформальные. *Формальные процедуры* (например, компьютерный анализ сложившейся ситуации) базируются на факторном прямом и факторном обратном анализах, а также на функционально-стоимостном и ситуационном анализе. Информационные технологии, применяемые на данном этапе – это

моделирующие технологии, технологии оперативного анализа данных (OLAP) и DM-технологии (data mining).

Неформальная часть этапа касается:

- выявления проблемы, образовавшейся в результате анализа ситуации;
- формулирования траекторной цели, ориентированной на достижение стратегической цели.

Этап 4. Выработка целей и генерализация альтернатив – одна из сложнейших задач в процессе формирования решения, требующая мощных инструментальных средств по извлечению знаний из большого объема данных (как внешних, так и внутренних). Существует достаточно много способов такого извлечения. Наиболее распространенным из них сегодня является метод извлечения знаний из баз данных – DM-технология.

Этап 5. Выбор решения: выбор критерия (и его обоснование) и оценка альтернатив по критерию. На этом этапе обосновывается критерий выбора, а на его основе оцениваются альтернативы и делается их выбор. И оценка, и выбор осуществляются системой на основе критерия, который был назван Лицом принимающим решения (ЛПР). Критерием отбора может служить любой признак, значение которого можно зафиксировать в некоторой шкале. Поскольку критерий служит для оценки вариантов решений, поэтому он должен быть измерим.

Этап 6. Анализ последствий принятия решения требует использования новейших инструментов, которые должны базироваться на прогнозных моделях и сценариях будущих событий. Поддерживать эти технологии способны DM-технология и OLAP.

Этап 7. Принятие решения и его мониторинг. Он предназначен для выполнения мониторинговых функций. В ряде СППР он выделяется в отдельную подсистему (блок), получившую название исполнительные расчетные системы. Они связаны с системами поддержки принятия решений и могут состоять из двух подсистем:

- 1) выдачи рекомендаций или инструкций, регламентирующих действий тех или иных сотрудников;
- 2) контроля (мониторинга) траекторных показателей.

Вопрос 2. Показатели измерения результативности работы предприятия (KPI).

Показатель служит для оценки объекта или процесса, поэтому он должен быть измерим.

В процессе управления предприятием используются показатели, которые оцениваются количественно, поскольку являются следствием структурированной (формализованной) информации. Но есть и качественные показатели, для обработки которых необходимо преобразование слабоструктурированной информации. Такое преобразование заключается во введении признаков, уточняющих, что мы понимаем под тем или иным качеством и использовании шкал бальной оценки.

Показатель должен измерять результат деятельности. Вопрос насколько избранный показатель измеряет то, что мы хотели бы измерить, очень важен, поскольку правильные измерители правильно передают смысл поставленной задачи и правильно мотивируют персонал. Если же показатели-ориентиры заданы неправильно, то и менеджмент вряд ли будет эффективен, а цели достигнуты. Поэтому необходимо четкое формальное отображение желаемых состояний предприятия (целей), для достижения которых будет разрабатываться стратегия.

При формулировании целей используются принципы – **SMART:**

- Specific – ясность и точность.
- Measurable – измеримость.
- Achievable – достижимость.
- Related – совместимость со стратегией и интересами.
- Time-bound – сроки.

Цель формулируется в результате констатации и анализа ситуации на основе миссии и видения. Стратегические цели связываются с показателями всех уровней управления, начиная с самого верхнего. Устанавливаются плановые значения показателей всех уровней, отслеживаются их фактические значения, выявляются отклонения фактических значений от плановых, осуществляется анализ сложившейся ситуации. Главная цель декомпозируется в дерево целей.

Декомпозиция главной цели в дерево целей – процесс неформальный, творческий, требующий определенных знаний и опыта. Для выбора пути достижения цели (подцели) задаются *коэффициенты приоритетности целей* (КПЦ). КПЦ – это инструмент управления выбором направления в достижении цели. При этом сумма КПЦ на одном уровне дерева, касающихся одного вышележащего узла, должна быть равна единице.

Управление на основе измерителей предполагает не только увязку стратегического и оперативного управления, но и создание многоступенчатой системы измерителей, соответствующих

вложенным циклам по уровням управления. В основе этого процесса лежит понятие целевого управления, которое предполагает дальнейшую трансформацию графа показателей в граф целей.

Итак, наличие измерителей позволяет реализовать целевое управление. Оно заключается в том, что от стратегических целей переходят к их декомпозиции. Затем превращают граф показателей в граф целей, задавая желаемые направления изменения показателей каждого уровня и числовые значения ожидаемых изменений. Далее: воздействие, мониторинг, анализ и изменение целей, если это необходимо. После этого «привязывают» цели к оргструктуре и назначают ответственных за их выполнение. При этом все сотрудники, вовлеченные в систему показателей, должны быть уверены в их правильности и точности, поскольку на них замкнута мотивация. Для каждого предприятия необходимо определить свои правильные показатели.

Количество ключевых показателей в системе измерения должно быть важным и достаточным для управления. Для стимулирования деятельности организации применяют ограниченное количество показателей (ключевых показателей эффективности – KPI), учитывающих параметры, которые действительно оказывают влияние на эффективность.

Внедрение системы измерителей – это инновационный проект. Предприятие должно оценить степень своей готовности к внедрению системы KPI. Внедрение информационных систем (ИС) изменит систему управления на предприятии, поэтому надо быть готовым к дальнейшим инновациям.

Эффективность внедренной системы KPI определяется степенью соответствия возможностей системы потребностям пользователей и системы управления.

Вопрос 3. Единое информационное пространство предприятия – основа аналитической деятельности предприятия.

Деятельность предприятия, любого объекта или даже самой маленькой системы неразрывно связана с информацией, которая окружает их повсюду. **Информация** – это сведения (сообщения, данные) независимо от формы их представления.

Если рассматривать современное предприятие как целостную сложную и открытую систему, то между его элементами идёт постоянный обмен информацией, которая на самом верхнем уровне представлена в виде человеческой речи, бумажных документов, данных в информационных системах.

Потоки информации, этапы обработки этой информации и те, кто участвует в обработке информации, входят в общее понятие «информационного пространства» предприятия, а деятельность по преобразованию потока информации в его сжатую характеристику представляет собой аналитическую деятельность и является неотъемлемым элементом совокупного информационного пространства.

Информационное пространство – совокупность банков и баз данных, технологий их сопровождения и использования, информационных телекоммуникационных систем, функционирующих на основе общих принципов и обеспечивающих информационное взаимодействие организаций и граждан и удовлетворение их информационных потребностей.

Структура информационного пространства. Основными компонентами информационного пространства являются:

- информационные ресурсы (ИР);
- средства и технологии информационного взаимодействия;
- информационная инфраструктура.

Информационные ресурсы часто определяют, как документы и массивы документов, хранимые в информационных системах. Все информационные ресурсы можно классифицировать по различным признакам:

- 1) отношению к предприятию: внутренние и внешние;
- 2) способу хранения (передачи): бездокументарные, документарные и электронные;
- 3) типу источника (пользователя): персональные, коллективные (группа, структурное подразделение) и общекорпоративные;
- 4) доступности: общедоступные, частично закрытые (только для группы лиц) и закрытые (для нескольких лиц).

Если же рассматривать информационные ресурсы, в рамках информационных систем, то можно выделить два типа ИР по содержанию: фактуальные и субъективные.

Фактуальные ресурсы включают различные виды учётной информации в информационных системах (в базах или хранилищах данных, файлах), а также электронные документы («текстовые» документы, хранящиеся в электронном виде).

Субъективные ресурсы представляют собой знания: вид информации, отражающей опыт специалиста (эксперта) в определенной предметной области, его понимание множества

текущих ситуаций и способы перехода от одного описания объекта к другому.

Все информационные ресурсы имеют различную степень своей структурированности. Под структурированностью понимается такое свойство информационного пространства, при котором его компоненты и взаимосвязи между ними выражены в явном виде. Различают пять степеней структурированности информационного пространства (ИП):

1. Неструктурированное информационное пространство (НИЦ). Примером такого пространства может служить разговор о составе договора.

2. Слабо структурированное информационное пространство (ССИП). Здесь структурирована только часть элементов ИП в письменной форме, например, письменно зафиксированы отдельные пункты договора.

3. Структурированное информационное пространство (СИП). Вся информация в таком пространстве задокументирована, понятия кодированы, имеются структурные компоненты, позволяющие обеспечить процессы загрузки данных в информационную систему.

4. Формально-структурированное информационное пространство (ФСИП). В таком пространстве определены связи между компонентами ИП и алгоритмы обработки элементов данных, например, сортировка данных, поиск, вызов данных и т.п.

5. Машинно-структурированное информационное пространство (МСИП). Вся информация, реализована в виде базы данных с возможностью её управления.

Комплексная автоматизация функции управления требует создания единого ИП на любом современном предприятии, в котором обычные сотрудники и руководство смогут осуществлять свою деятельность, следуя едиными правилами доступа, представления и обработки информации.

Начальным этапом создания такой системы является построение модели предметной области или другими словами модели ИП для конкретного бизнеса и позиционирование в ней своего предприятия. Исходя из составных элементов ИП, выделенных ранее, можно выделить три измерения любого полноценного ИП (рис. 2).

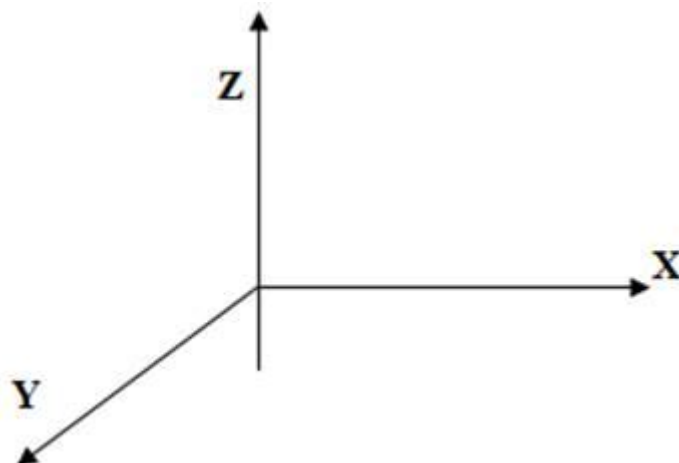


Рис. 2. Информационное пространство предприятия

Ось (Z) характеризует ресурсы ИП, которые накапливают фактографическую информацию, отражающую специфику деятельности организации. Это предметные базы данных, различные информационные системы, хранящие информацию, необходимую для оперативной работы сотрудников.

Ось (X) отражает специфику работы с полнотекстовыми документами, обеспечивающие внутреннее и внешнее взаимодействие организации с субъектами бизнеса. В документах содержится неструктурированная и слабоструктурированная информация, плохо поддающаяся аналитической обработке в автоматизированных системах.

Ось (Y) определяет атрибуты и процедуры процессов обработки информации, т.е. описывает последовательность и взаимосвязь всех процедур работы с информацией.

Развитие модели ИП напрямую связано с развитием компании. Сама модель ИП подходит для любой компании, любой сферы деятельности, т.к. не устанавливает конкретную технологию работы с информацией.

Вопрос 4. Отображение показателей отчетности в информационном пространстве предприятия.

В теории экономических информационных систем рассматривается понятие *единицы информации*. В зависимости от содержания (наполненности) различают несколько единиц информации:

- ***Реквизит*** или атрибут информационно отображает какое-либо свойство объекта.

- **Составная единица информации** состоит из набора реквизитов и информационно отображает сам объект или его часть.

- **Показатель** является разновидностью составной единицы информации, широко используется для отображения деятельности компании. Он обычно привязан к предметной области, определенным образом структурирован и успешно используется в информационных системах, в том числе в информационно-аналитических системах.

В экономическом словаре Лопатникова Л.И. ¹³ **показатель** (economic index figure, activity indicator) определяется как выраженная числом характеристика какого-либо свойства экономического объекта, процесса или решения.

В теории и практике автоматизированной обработки информации в показателе выявляется следующая структура: количественное значение (основание) и набор его содержательных признаков, называемый **идентификатором показателя**. Из определения следует, что показатель состоит из набора реквизитов:

- реквизитов оснований;
- реквизитов признаков.

Обе группы реквизитов должны обязательно присутствовать в показателе, иначе он теряет смысл. Такая совокупность реквизитов формирует высказывание, достаточное для передачи предметного смысла деятельности организации.

Для описания деятельности предприятия определяются категории показателей, разрабатываются системы показателей, отражающие эту деятельность. Затем разрабатывается структура показателей, которая может отображать показатель следующим образом:

$$P \Rightarrow (I, Q),$$

где

P – показатель деятельности, например, экономический;

I – набор реквизитов-признаков, определяющих предметный смысл показателя;

Q – количественное значение показателя.

Состав реквизитов-признаков при анализе показателя задается пользователем. Признаки могут детализировать предметную область,

раскрывать ту или иную сферу деятельности предприятия. При этом может формироваться иерархия признаков. Иерархии обычно отражают вертикальные и горизонтальные взаимосвязи в структуре организации.

Реквизиты-признаки могут быть единичными или составными, могут иметь определенные взаимосвязи. При разработке ИП эти взаимосвязи уточняются, строятся соответствующие иерархические схемы реквизитов-признаков, позволяющие детализировать рассматриваемый объект до необходимого уровня. Самый нижний уровень иерархии содержит единичные реквизиты, на верхних уровнях расположены множественные реквизиты.

Детализируя признаки анализа, получим следующую структурную формулу показателя:

$$P \Rightarrow ((F, B, O) (M, U, T, S)Q) ,$$

где

P – показатель деятельности, например, экономический;

F – формальная характеристика показателя;

B – технологический процесс;

O – объект измерения;

M – единица измерения;

U – уровень показателя;

T – интервал времени;

S – субъект (покупатель);

Q – количественное значение показателя.

Например, для фирмы, торгующей компьютерной техникой, показатель будет иметь вид:

Объем продаж \Rightarrow (жесткий диск, фирма изготовитель, объем диска) (менеджер, филиал, год, покупатель) значение объема продаж).

Для целей анализа структура показателя может быть еще расширена. Количество реквизитов-признаков показателя неограниченно.

Структура показателя может быть достаточно сложная. Для образного восприятия сложных понятий удобна их пространственная или геометрическая интерпретация. Представим все реквизиты-признаки показателя в координатном пространстве (рис. 3).

Число осей или число измерений этого пространства будет равно числу реквизитов-признаков. Так как реквизитов-признаков

может быть много, то ИП отображающее показатель будет многомерным.

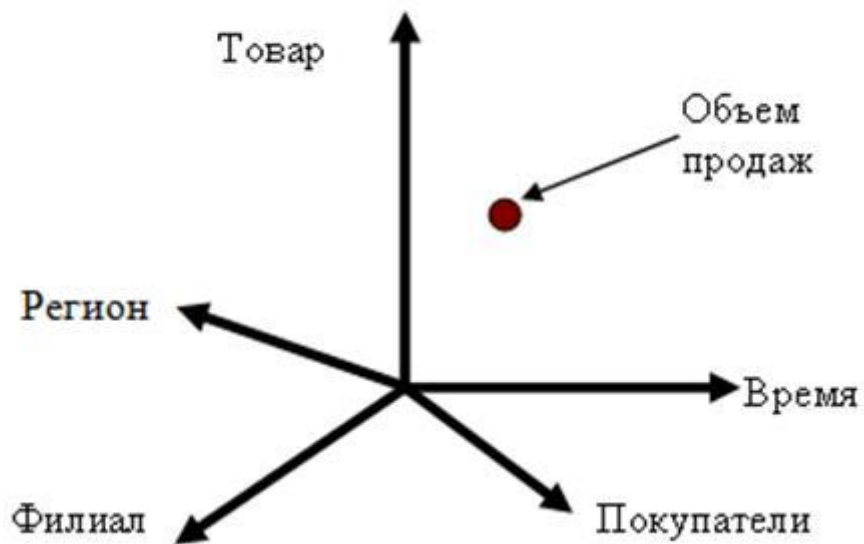


Рис. 3. Система координат многомерного информационного пространства показателя

Представить себе такое многомерное пространство достаточно сложно, поэтому обычно все стараются свести к более понятному трехмерному пространству. Пример такого пространства показателей приведён на рисунке 4.

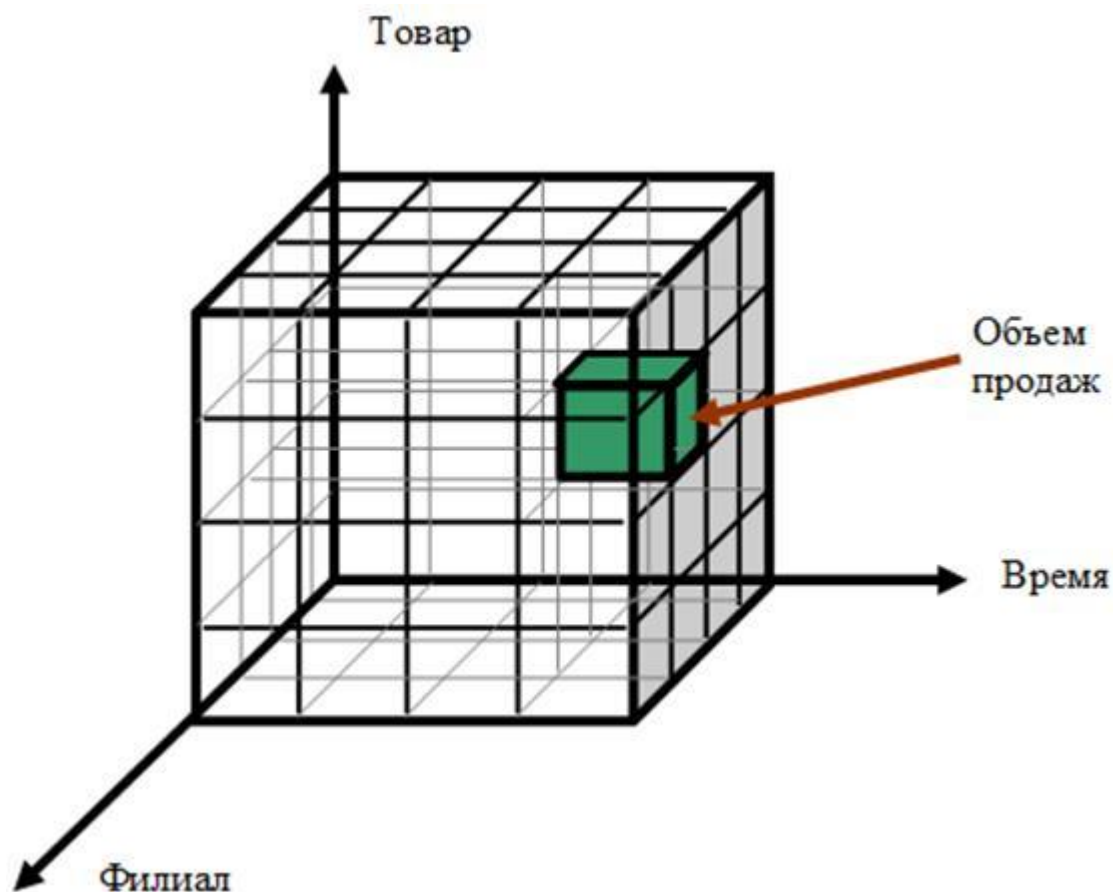


Рис. 4. Трёхмерное информационное пространство для отображения показателя «Объем продаж»

Значения реквизитов-признаков откладываются по осям, причем множественные признаки представляют собой отрезки на осях трехмерного пространства, внутри которых откладываются единичные признаки. Количественные значения или реквизиты-основания наполняют само ИП. Любая точка пространства представляет собой числовое значение показателя, которое идентифицируется проекцией на оси координат.

Рассмотрим множественное измерение «Филиал». В компании может быть несколько филиалов в разных городах или странах. В структуре каждого филиала есть подразделения и отделы, в которых работают сотрудники. На каждом из этих уровней можно проанализировать объем продаж. Эта зависимость может быть представлена в виде иерархической структуры (рис. 5).

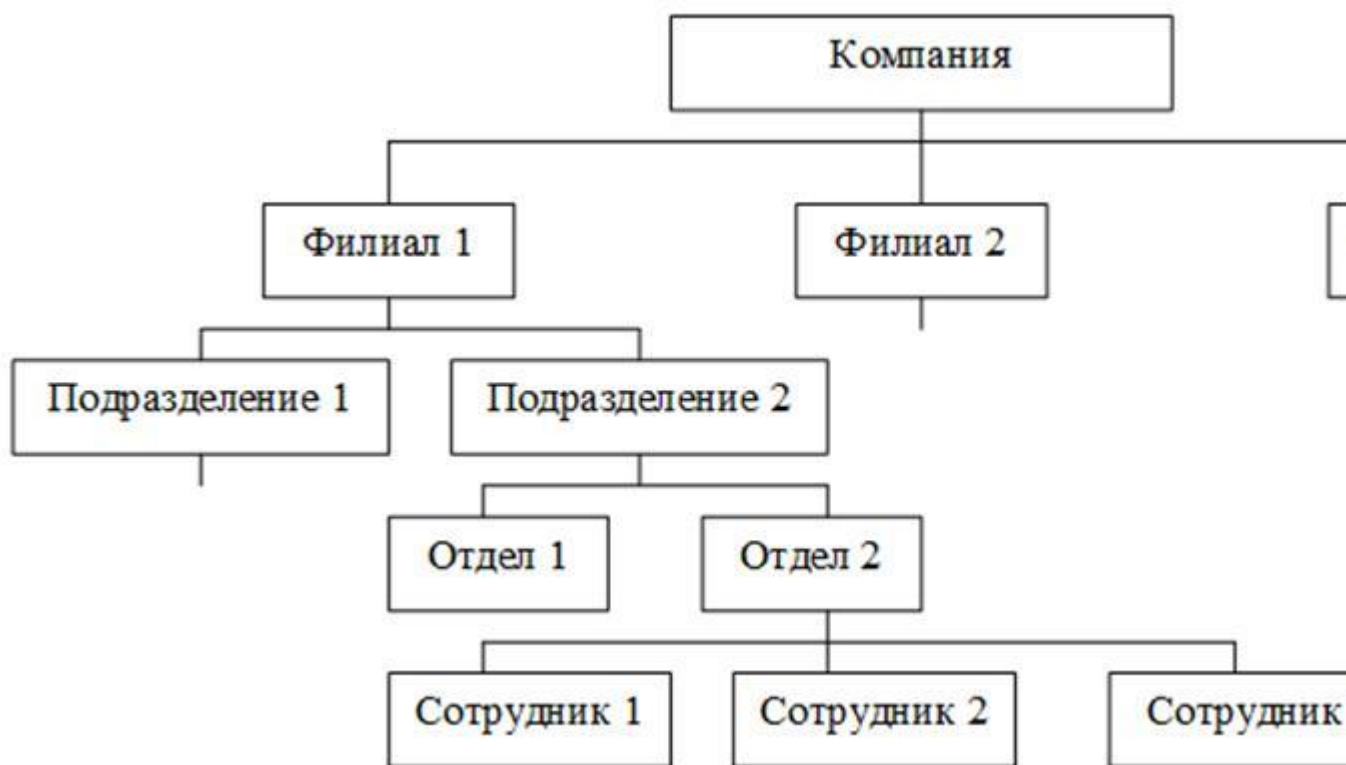


Рис. 5. Иерархия измерений

Вопрос 5. Компонентная архитектура программных средств хранения данных и их аналитической обработки в Business Intelligens.

Задачами любой информационно-аналитической системы (ИАС) являются эффективное хранение, обработка и анализ данных.

Эффективное хранение информации достигается наличием в составе ИАС целого ряда источников данных.

Обработка и объединение информации достигается применением инструментов извлечения, преобразования и загрузки данных. *Анализ* данных осуществляется при помощи современных инструментов делового анализа данных.

Архитектура современной ИАС организации в обобщенном виде представлена на рисунке 6.

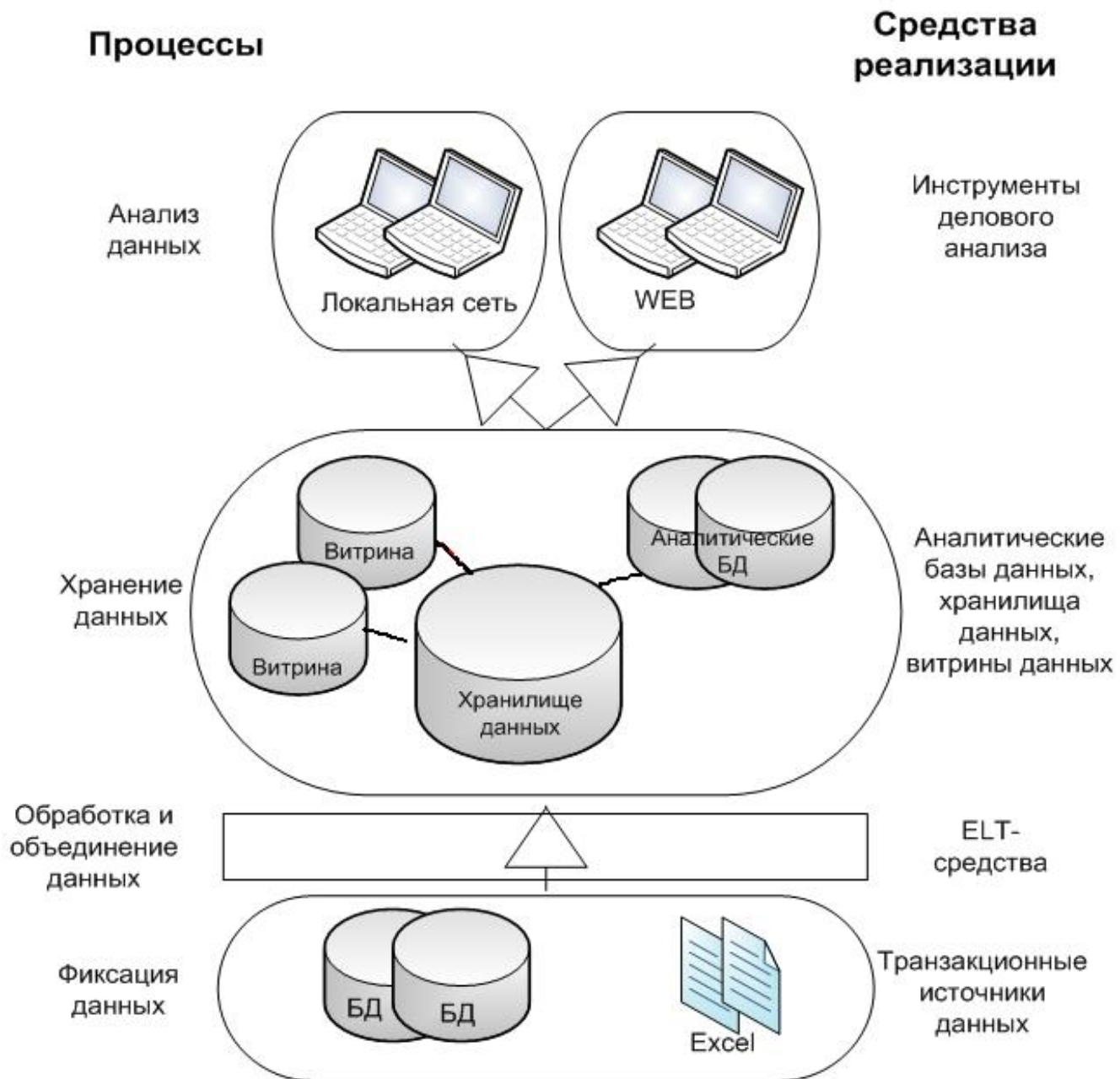


Рис. 6. Архитектура современной информационно-аналитической системы

Из рисунка 6 видно, что архитектура современной ИАС содержит следующие уровни:

- 1) сбор и первичная обработка данных;
- 2) извлечение, преобразование и загрузка данных;
- 3) складирование данных;
- 4) представление данных в витринах данных;
- 5) анализ данных;
- 6) Web-портал.

Сегодня на рынке информационных технологий представлен широкий спектр инструментальных средств, предназначенных для быстрой реализации компонентов архитектуры ИАС.

При реализации ИАС предприятия могут быть использованы программные решения как разных фирм-производителей – смешанные решения, так и одного производителя – платформенно-базированные решения.

Рассмотрим пример реализации системы бизнес-аналитики организации инструментами Microsoft. Общая архитектура решения для систем бизнес аналитики компании Microsoft показана на рисунке 7.

Основой комплексного предложения для BI от Microsoft является СУБД SQL Server – полнофункциональная платформа сервисов для работы с данными, позволяющая:

- унифицировать хранение и доступ к данным по всему предприятию;
- создавать сложные BI-решения и управлять ими;
- расширять круг пользователей BI-решения.

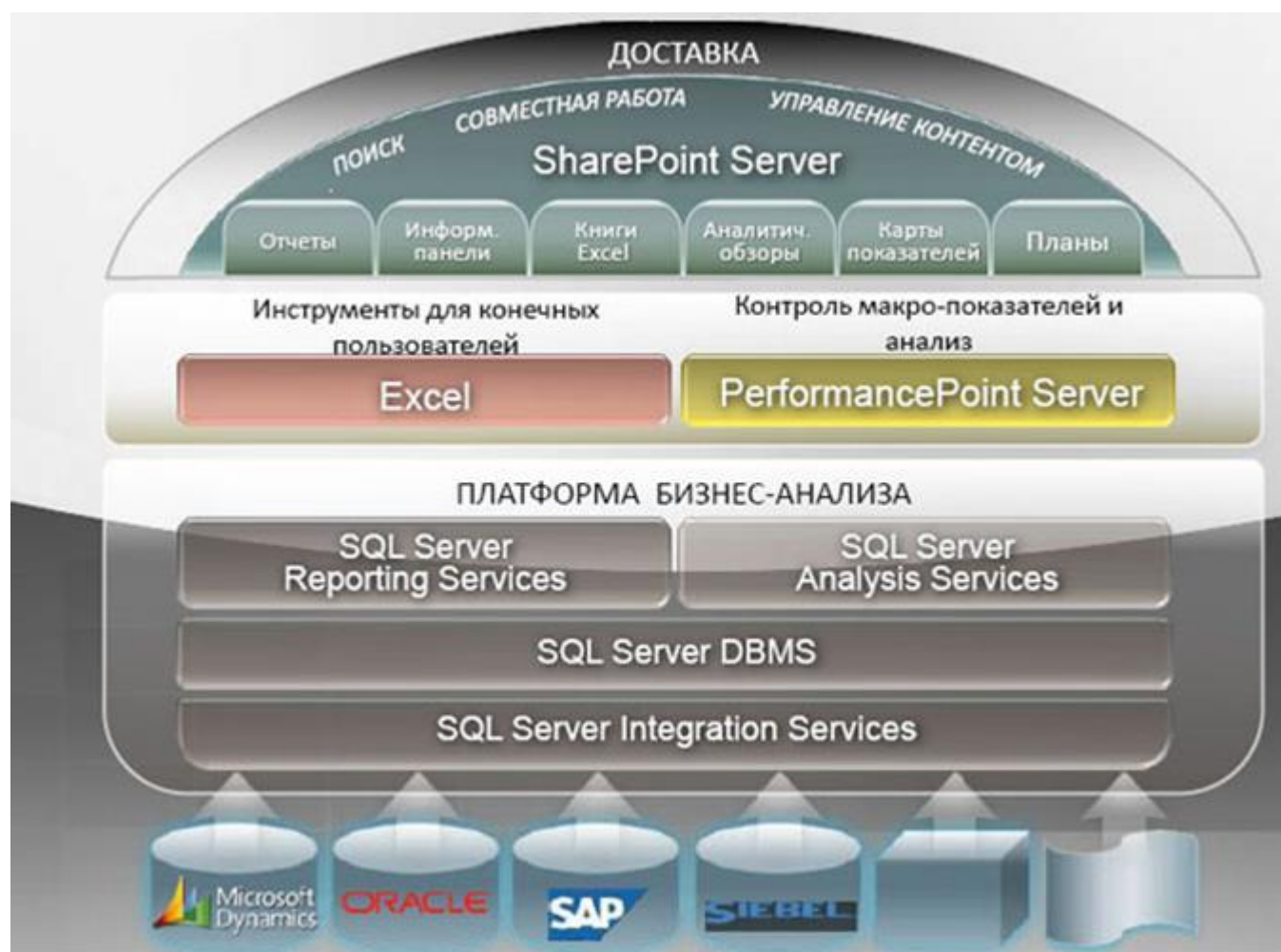


Рис. 7. Платформа бизнес аналитики-компании Microsoft

Вопросы для самопроверки:

1. Приведите классификацию целей предприятия по уровням управления.
2. Чем отличаются стратегические решения от оперативных?
3. Перечислите и охарактеризуйте этапы принятия решения?
4. В чем заключается этап анализа?
5. Какие шкалы бальной оценки используются для количественной оценки показателей?
6. Для чего используются коэффициенты приоритетности целей?
7. Дайте определение информационного ресурса.
8. Из каких элементов состоит информационное пространство?
9. Что такое показатель и какова его структура?
10. Перечислите основные компоненты современной информационно-аналитической системы.

Литература по теме:

1. Алдохина, О.И. Информационно-аналитические системы и сети: учеб. пособие / О.И. Алдохина, О.Г. Басалаева. – Ч. 1. Информационно-аналитические системы. – Кемерово: КемГУКИ, 2010. – Режим доступа: URL: <http://biblioclub.ru/>.
2. Белов В.С. Информационно-аналитические системы: основы проектирования и применения: учеб.-практ. пособие / В.С. Белов. – 2-е изд., перераб. и доп. – М.: Евразийский открытый институт, 2010. – Режим доступа: URL: <http://biblioclub.ru/>.
3. Информационные аналитические системы: учеб. / Под ред. В.В. Дика. – М.: Московский финансово-промышленный ун-т «Синергия», 2013.
4. Сайт компании «Microsoft». – Режим доступа: <http://www.microsoft.com/ru-ru/default.aspx>.

Тема 2. Источники данных и хранение информации на предприятии

Цели:

Сформировать представление об источниках хранения данных на предприятии.

Задачи:

- Изучить основные источники данных на предприятии.
- Рассмотреть принципы построения средств хранения данных.
- Изучить структуру и технологию работы хранилищ данных.
- Познакомиться с видами хранилищ данных и рынком DWH.

Вопросы темы:

1. Способы хранения информации на предприятии и источники данных.
2. Принципы построения хранилищ и OLTP баз данных.
3. Свойства и структура хранилищ данных.
4. Виды хранилищ данных.
5. Технология работы хранилищ данных.
6. Рынок DWH.

Основные понятия:

- *файловая система;*
- *оперативная база данных (OLTP);*
- *транзакция;*
- *хранилище данных;*
- *витрина данных;*
- *предметная ориентированность;*
- *интегрированность данных;*
- *неизменяемость;*
- *мера (measure);*
- *измерение (dimension);*
- *атрибут;*
- *иерархия;*
- *агрегация данных;*
- *«звезда» (star schema);*
- *«снежинка» (snowflake schema).*

Вопрос 1. Способы хранения информации на предприятии и источники данных.

В процессе деятельности предприятия накапливается большое количество информации.

Вся эта информация должна храниться на предприятии и быть в любой момент доступна для пользователя. Для хранения информации могут быть использованы различные средства: файловые системы, оперативные базы данных (OLTP) и хранилища данных (DWH).

Файловые системы. В современных условиях большинство рабочих мест сотрудников оснащены персональными компьютерами (АРМ – автоматизированное рабочее место). В процессе работы на каждом АРМ накапливается оперативная информация, документы, сопровождающие те или иные бизнес процессы. Эта информация хранится на компьютере в виде файлов.

Файл – это именованная область внешней памяти, в которую можно записывать и из которой можно считывать данные. Файлы бывают разных типов: обычные, специальные, файлы-каталоги.

Для организации хранения и управления файлами на компьютере используется файловая система, представляющая собой функциональную часть операционной системы. Файловые системы предназначены для обслуживания многих тысяч файлов и обеспечивают хранение слабо структурированной информации. Каждая файловая система поддерживает некоторую иерархическую файловую структуру (древовидную), включающую чаще всего неограниченное количество уровней иерархии (рис 8).

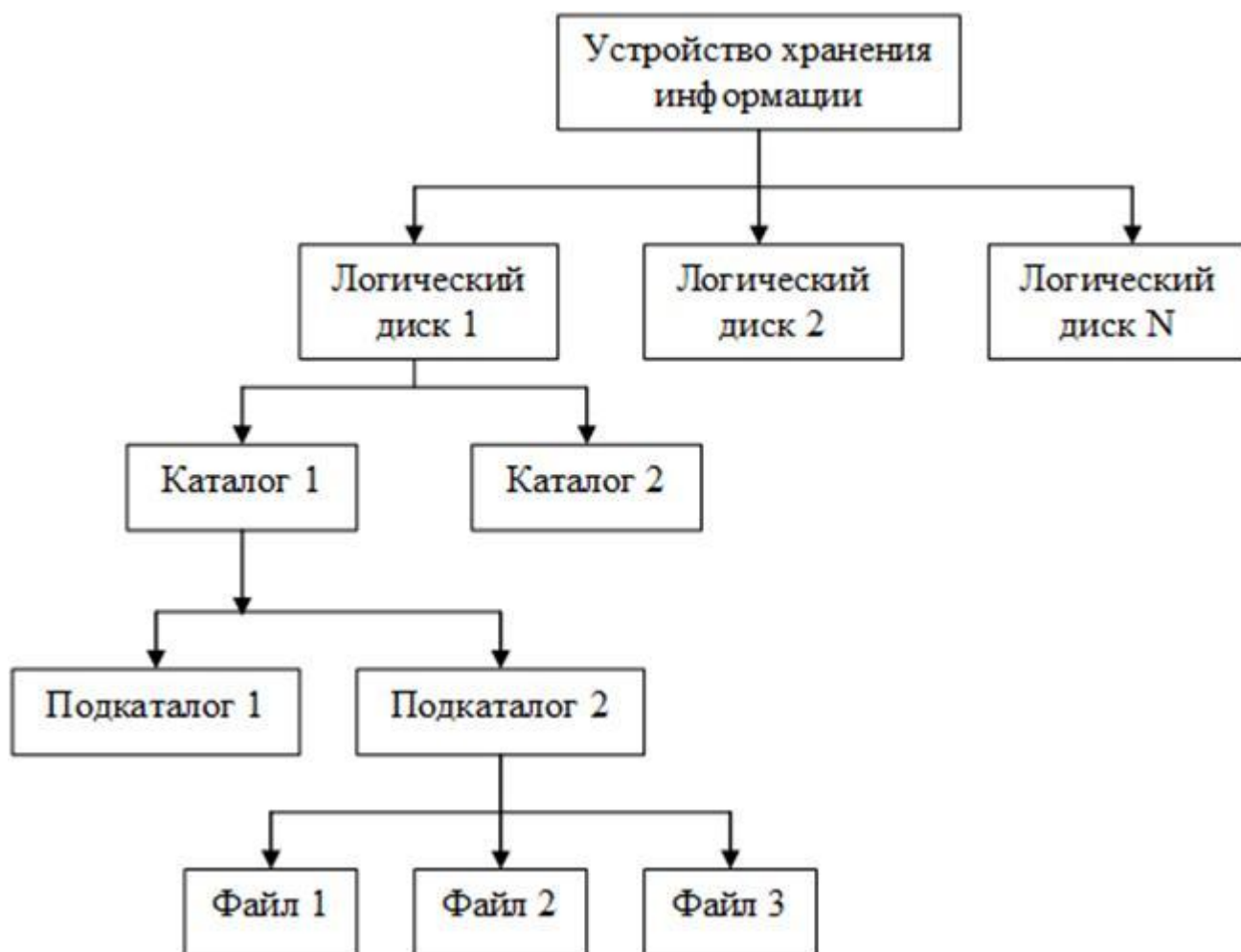


Рис. 8. Иерархическая организация файловой структуры хранения

Доступ к файлам организуется в файловых системах через каталоги. *Каталог* (от греч. *katalogos* – список) – справочник файлов и каталогов со ссылками на их расположение. При этом различают главный (корневой) каталог и подчиненные (вложенные) каталоги.

На современном предприятии рабочие места пользователей часто объединяются в локальные сети. Это позволяет организовать совместную эффективную работу с бизнес информацией группе пользователей, участвующих в определенном бизнес процессе. Часто хранение такой информации переносится на сервер, организуется так называемый файловый архив.

Оперативные базы данных (OLTP) используются предприятиями для поддержания их повседневной деятельности, для отслеживания информации, с которой они имеют дело в процессе решения оперативных задач. Это может быть информация о произведенных товарах, принятых заказах, оказанных услугах, выплатах, доходах и т.п.

Результатом фиксации указанной информации становится одна или несколько записей в оперативной базе данных. Сам процесс фиксации называют *бизнес-транзакцией*, а информацию – *данными транзакции*. **Транзакция** – это последовательность операторов манипулирования данными, выполняющаяся как единое целое и переводящая базу данных из одного целостного состояния в другое целостное состояние.

Системы оперативной обработки транзакций (OLTP от *англ.* OnLine Transaction Processing,) служат для хранения данных о выполняемых бизнес-транзакциях. Основная функция подобных систем заключается в одновременном выполнении большого количества коротких транзакций от большого числа пользователей. К числу транзакционных систем относятся:

- ERP-системы;
- автоматизированные банковские системы (АБС);
- биллинговые системы;
- учетные системы и некоторые др.

Данные в OLTP-системы поступают в основном из внутренних источников, причем это текущие данные за период от нескольких месяцев до одного года. Объемы хранимых данных могут составлять сотни мегабайт, гигабайты. Частота обновления данных высокая, обновления происходят маленькими порциями. Основное их назначение фиксация данных, оперативный поиск и преобразование данных. В основе таких систем лежат оперативные базы данных.

Хранилища данных. По определению Билла Инмона, основоположника хранилищ данных, «хранилище данных – это предметно-ориентированное, привязанное ко времени и неизменяемое собрание данных для поддержки процесса принятия управляющих решений». Задача хранилища – предоставить лицу, принимающему решения, информацию для анализа в одном месте в простой, понятной для восприятия структуре.

Данные в хранилище попадают из оперативных систем (OLTP-систем), которые предназначены для автоматизации бизнес-процессов, и других внутренних источников информации. Хранилище также может пополняться за счет внешних источников информации, например, статистических отчетов и т.п.

Хранилищу данных характерна малая частота изменений. Изменения производятся большими порциями и обычно по расписанию. Хранилище объединяет внутренние и внешние данные, в составе этих данных входят текущие данные и исторические за период до нескольких десятков лет. Объемы хранимых данных – гигабайты и терабайты. Основное назначение хранилищ данных – это хранение детализированных и агрегированных исторических данных, аналитическая обработка, прогнозирование и моделирование.

Можно выделить два типа хранилища данных:

- 1) корпоративные хранилища данных (enterprise data warehouses);
- 2) витрины или киоски данных (data marts).

Корпоративные хранилища данных содержат информацию, относящуюся к деятельности всей корпорации и собранную из множества оперативных источников данных. Их объем может достигать от десятков Гбайт до одного или нескольких терабайт.

Витрины данных (небольшие хранилища данных) содержат подмножество корпоративных данных и создаются для определенной группы пользователей, отделов или подразделений внутри организации. Они охватывают конкретный аспект, интересующий сотрудников данного отдела. Витрина данных может получать данные из корпоративного хранилища (зависимая) или данные могут поступать непосредственно из оперативных источников (независимая витрина).

Вопрос 2. Принципы построения хранилищ и OLTP баз данных.

Хранилище данных включает в себя несколько компонентов:

- подсистема хранения данных;
- подсистема метаданных или репозиторий;
- менеджер хранилища;
- менеджер загрузки;
- менеджер запросов.

Типичная архитектура хранилища данных приведена на рисунке 9.

Подсистема хранения данных предназначена непосредственно для хранения данных.

Подсистема метаданных (репозиторий). Метаданные – это данные о данных. Метаданные сами по себе не несут информации, но описывают атрибуты данных в хранилище.

Менеджер хранилища (warehouse manager) выполняет операции, связанные с управлением информацией, помещенной в хранилище данных.

Менеджер загрузки (load manager) выполняет операции, связанные с извлечением и загрузкой данных в хранилище.

Менеджер запросов (query manager) выполняет операции, связанные с управлением пользовательскими запросами.

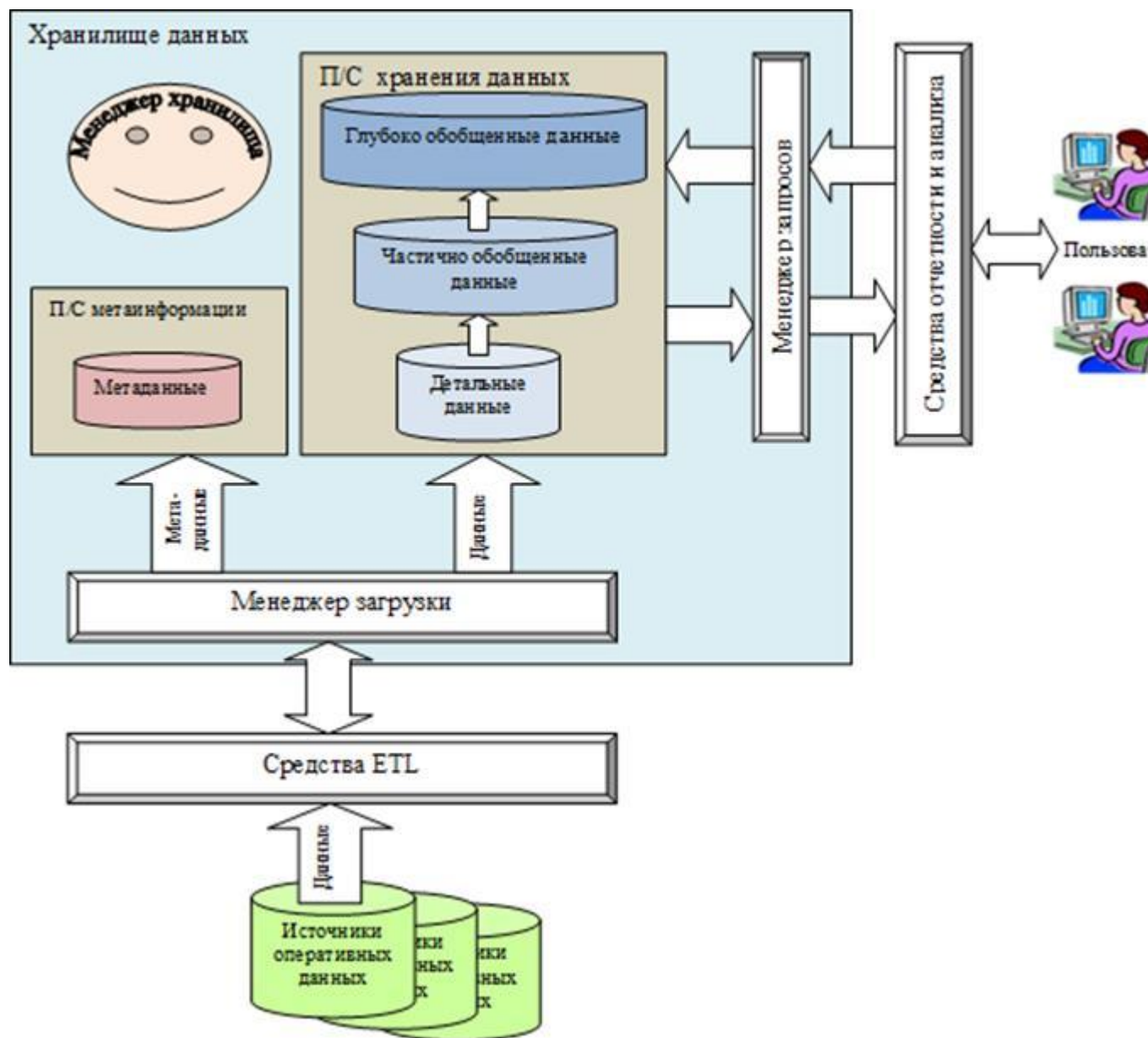


Рис. 9. Архитектура хранилища данных

Принципы построения формулируются исходя из задач, стоящих перед Хранилищем данных и OLTP-системой.

Принципы построения Хранилищ данных:

1. Оптимизация базы данных хранилища для быстрого выполнения объемных незапланированных запросов.
2. Пакетная загрузка данных, с предварительной обработкой данных.
3. Наличие пользовательских инструментов для создания сложных запросов и наглядных отчетов.
4. Возможность быстрого изменения структуры базы данных хранилища – это требование исходит от динамики бизнеса.

Принципы построения OLTP-систем отличаются от принципов построения Хранилищ данных и часто принципиально противоположны им. Для обеспечения своих ключевых функций OLTP-системы должны:

1. Быстро выполнять транзакции.
2. Защищать данные от пользовательских ошибок ввода.
3. Минимизировать изменения базы данных в процессе выполнения одной транзакции для ускорения ее выполнения.

Вопрос 3. Свойства и структура хранилищ данных.

Свойства информационных хранилищ. Уильям Инмон определил информационное хранилище данных как специальным образом администрируемую базу данных, содержимое которой имеет следующие свойства:

- Предметная ориентированность.
- Интегрированность данных.
- Привязка ко времени.
- Неизменяемость.
- Минимизация избыточности информации.

Предметная ориентированность: хранилище должно разрабатываться с учетом специфики предметной области.

Интегрированность данных: все данные о разных бизнес объектах, взаимно согласованы и хранятся в едином общекорпоративном Хранилище.

Привязка ко времени: данные хронологически структурированы и отражают историю за достаточный для выполнения задач бизнес анализа и прогнозирования период времени.

Неизменяемость: данные не обновляются в оперативном режиме, а лишь регулярно пополняются из систем оперативной обработки данных. Исходные (исторические) данные, остаются неизменными и используются исключительно в режиме чтения.

Минимизация избыточности информации в хранилище данных обеспечивается фильтрацией, сортировкой и очисткой данных при загрузке, приведением их к единому формату.

Структура хранилищ данных. Структуру хранилищ определяют хранимые в них данные. Выделяют четыре типа данных:

- 1) меры;
- 2) измерения;
- 3) атрибуты;

4) иерархии.

Мера (measure) – это численное значение показателя, выражающее определенный аспект деятельности организации. Это понятие соответствует такому понятию ИП как «реквизит-основание». Меры также называют фактическими значениями, или просто фактами. А таблицы, которые содержат данные значения, называются таблицами фактов.

Измерение (dimension) – это направление анализа, способ детализации данных, определяющий составляющие агрегированных мер.

Атрибуты – обычно понятные пользователю текстовые описания объектов.

Иерархия – это расположение частей или элементов целого в порядке от высшего к низшему. Иерархическая структура – это многоуровневая форма организации объектов. Графически эта структура представляется в виде дерева.

В хранилище данных меры хранятся в таблице фактов, а измерения – в таблицах измерений.

Таблица фактов является основной таблицей хранилища данных. Таблица фактов содержит:

- сведения об объектах или событиях, совокупность которых будет в дальнейшем анализироваться;
- уникальный составной ключ, объединяющий первичные ключи таблиц измерений;
- одно или несколько числовых полей, на основании которых в дальнейшем будут получены агрегатные данные.

Таблицы измерений содержат:

- неизменяемые либо редко изменяемые данные;
- ключевое поле для идентификации члена измерения, одно описательное поле, определяющее имя члена измерения.

Скорость роста таблиц измерений должна быть незначительной по сравнению со скоростью роста таблицы фактов.

Каждая таблица измерений должна находиться в отношении «один ко многим» с таблицей фактов. Если каждое измерение содержится в одной таблице, такая схема хранилища данных носит название «звезда» (star schema). Пример такой схемы приведен на рисунке 10.

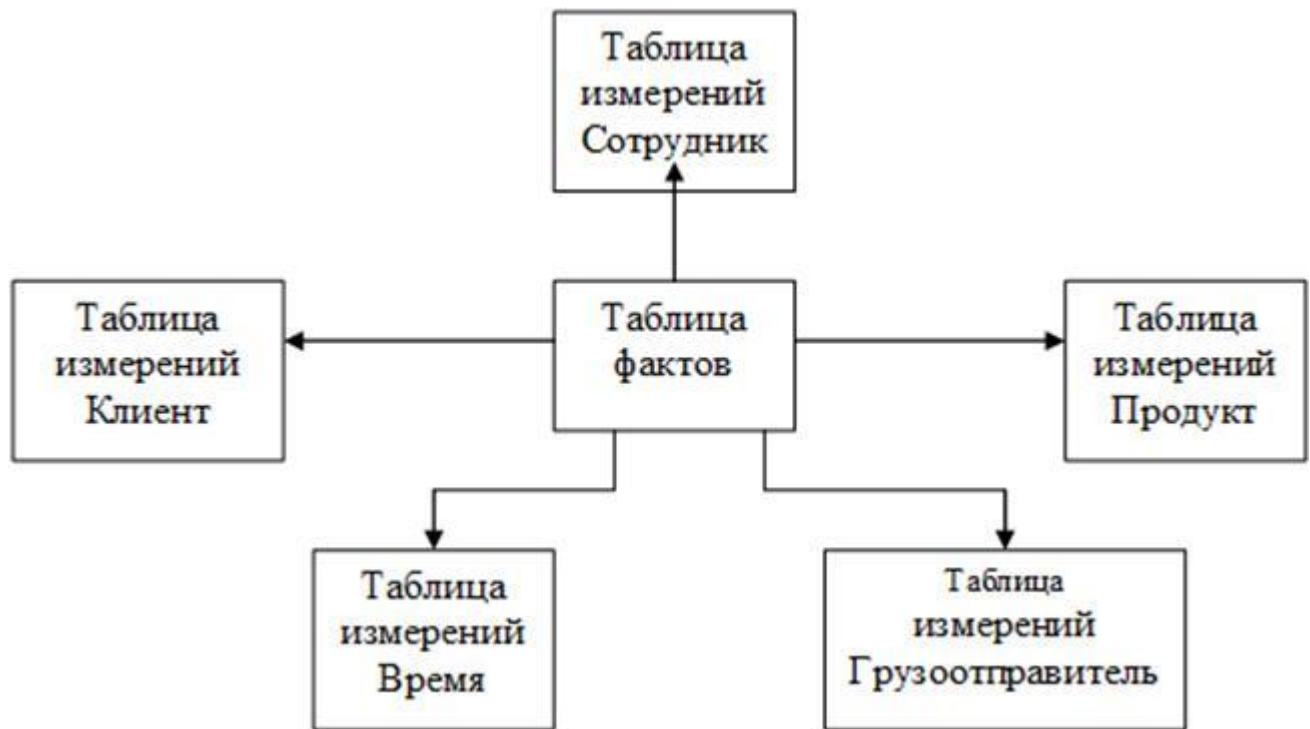


Рис. 10. Пример схемы «звезда»

Если же хотя бы одно измерение содержится в нескольких связанных таблицах, такая схема хранилища данных носит название «снежинка» (snowflake schema). Наиболее часто используется схема «звезда». Это вызвано требованием обеспечения высокой скорости выполнения запросов к хранилищу данных.

Вопрос 4. Виды хранилищ данных.

Реализация информационного хранилища может быть осуществлена несколькими способами:

1. Централизованное хранилище данных.
2. Распределенное хранилище данных.
3. Автономные витрины данных.
4. Шина взаимосвязанных витрин данных.
5. Единое интегрированное хранилище и много витрин данных.
6. Виртуальное хранилище данных.

1. *Централизованное хранилище данных* – единое централизованное хранилище данных объединяет информацию из различных источников: операционных баз данных.

2. *Распределенное хранилище данных* основано на распределении функций ИХ в соответствии с характером бизнеса или регионом в корпоративных структурах.

3. *Автономные витрины данных*: создаются небольшие предметно-ориентированные базы данных, в которых группируется информация, относящаяся к какому-либо достаточно самостоятельному направлению деятельности крупной корпоративной системы.

4. *Шина взаимосвязанных витрин данных* разрабатывается с использованием единых измерений, что в результате приводит к созданию логически интегрированных витрин.

5. *Единое интегрированное хранилище и много витрин данных* объединяет две концепции: единого интегрированного хранилища и связанных с ним и получающих из него информацию витрин данных.

6. *Виртуальное хранилище данных* – система, предоставляющая интерфейсы и методы доступа к OLTP-системе, которые эмулируют работу с данными в этой системе, как с хранилищем данных

Вопрос 5. Технология работы хранилищ данных.

Хранилище данных служит главным источником достоверной информации для руководителей и специалистов всех подразделений организации, что обеспечивает согласованность, своевременность и обоснованность принятия управленческих решений, облегчает выверку обязательной отчетности и обеспечивает выпуск управленческой отчетности.

Общий принцип работы Хранилища данных состоит в следующем: в OLTP системах выполняются учетные операции, затем с определенной периодичностью данные поступают в Хранилище, на основе которого осуществляется анализ информации, и выпускаются различные отчеты (рис. 11).



Рис. 11. Принцип работы Хранилища данных

Технология работы хранилища данных складывается из нескольких технологических процессов:

1. Технология сбора данных.
2. Технология очистки и загрузки данных.
3. Технология выполнения расчетов.
4. Технология отработки запросов.

Технология сбора данных обеспечивает регулярное и бесперебойное получение данных из удаленных филиалов, дополнительных офисов, из различных информационных систем. Эта технология включает в себя форматы данных, технологию их генерации, бизнес-правила, регламентирующие извлечение данных

из внешних источников, дистрибуцию метаданных (нормативно-справочной информации) и многое другое.

Технология очистки и загрузки данных обеспечивает входной контроль данных, автоматическое исправление ошибок, приведение данных к единым стандартам, загрузку больших массивов данных, многоуровневую журнализацию.

Технология выполнения расчетов. Специальный аппарат выполнения расчетов обеспечивает:

- агрегацию данных – расчет обобщенных показателей;
- консолидацию данных – суммирование данных по организационной иерархии;
- расчет производных показателей.

Технология отработки запросов предполагает выполнение сложных запросов к большим массивам данных.

Вопрос 6. Рынок DWH.

На рынке программного оборудования предлагается ряд продуктов, которые имеют принципиально разную функциональность, назначение, степень готовности к применению, однако все они позиционируются как Хранилища данных. Продукты, которые относят к категории Хранилищ данных, можно разделить на несколько групп, см. таблицу 1.

Таблица 1.

Название группы	Назначение	Продукты
Специальная СУБД	Предназначена для создания Хранилищ данных	СУБД SAP IQ MS Analysis Services Oracle Explorer
Инструмент программиста	Специальные CASE-средства, ориентированные на создание реляционных БД в идеологии Хранилищ	Sybase PowerDesigner IBM WebSphere DataStage
Отраслевые заготовки Хранилищ данных	Набор заготовок отраслевых приложений, применение которых сокращает сроки разработки Хранилища данных	Sybase PowerDesigner IBM WebSphere DataStage
Конструктор	<ul style="list-style-type: none"> • Обладает всеми свойствами конечных клиент-серверных продуктов. • Содержит готовые информационные объекты. • Позволяет создавать конечные управленческие приложения при помощи дизайнерских интерфейсов, ориентированных на аналитика. 	SAS Financial Management (SAS) Контур Корпорация (Intersoft Lab)
Специализированное	Реализуют одну или несколько задач на	Контур Корпорация Бюджет

приложение	платформе Хранилищ данных	холдинга (Intersoft Lab) Контур Корпорация Финансовое управление банком (Intersoft Lab)
Комплексная платформа разработки	Позволяет: <ul style="list-style-type: none"> • разработчикам софтверной компании создавать заказные или тиражные Хранилища данных; • программистам ИТ-подразделений разработать Хранилище собственными силами. 	средства генерации Хранилищ (data warehouse generation – DWG); средства управления Хранилищами (data warehouse management – DWM)

По данным исследования мирового рынка систем управления базами данных (СУБД), проведенного компанией Gartner в ноябре 2017 г., лидерами рынка СУБД являются компании Microsoft, Oracle, Amazon Web Services (AWS), SAP, IBM.

Вопросы для самопроверки:

1. Что такое хранилище данных?
2. Чем отличаются OLTP-системы от хранилищ данных?
3. Какие принципы лежат в основе построения Хранилищ данных?
4. Что такое витрина данных?
5. Перечислите свойства информационных хранилищ.
6. Что означает предметная ориентированность хранилища данных?
7. Какая информация хранится в таблице фактов?
8. Какая информация хранится в таблицах измерений?
9. Перечислите виды хранилищ данных.
10. Назовите основных игроков на рынке хранилищ данных.

Литература по теме:

1. Информационные аналитические системы: учеб. / Под ред. В.В. Дика. – М.: Московский финансово-промышленный ун-т «Синергия», 2013.
2. Выбор архитектуры хранилища данных. – [Электронный ресурс]. – Режим доступа: <http://iso.ru/ru/press-center/journal/2204.phtml>.
3. Александр Стулов. Хранилища данных: основные архитектуры и принципы построения в реляционных СУБД – [Электронный ресурс]. – Режим доступа: http://www.bipartner.ru/resources/dw_arch.html.

Тема 3. Инструменты интеграции данных из различных источников

Цели:

Сформировать представление о технологиях и инструментах интеграции данных на предприятии.

Задачи:

- Ознакомиться с источниками данных для формирования отчетности.
- Изучить технологии интеграции данных.
- Ознакомиться с рынком средств интеграции данных.

Вопросы темы:

1. Источники данных для формирования отчетности.
2. Технологии интеграции данных.
3. Рынок средств интеграции приложений.

Основные понятия:

- *децентрализованные источники данных;*
- *централизованный источник данных;*
- *интеграция данных;*
- *консолидация данных;*
- *федерализация данных;*
- *распространение данных;*
- *режим реального времени;*
- *пакетная интеграция данных;*
- *извлечение;*
- *преобразование;*
- *загрузка;*
- *интеграционная платформа.*

Вопрос 1. Источники данных для формирования отчетности.

Формирование разнообразной отчетности – важнейшая функция информационной системы предприятия. Формируемую на предприятии отчетность принято разделять на оперативную и аналитическую.

Оперативная отчетность. Задача оперативной отчетности – отражение текущего состояния предприятия и его различных подразделений. Такие отчеты требуются с высокой частотой и, как

правило, являются довольно узкоспециализированными, т.е. охватывают какое-то одно направление деятельности.

Оперативные отчеты можно получать непосредственно из оперативных (OLTP) систем. Так как используемые для отчетов данные хранятся в разрозненных источниках, такой способ хранения является децентрализованным (рис. 12).



Рис. 12. Децентрализованные источники данных

Оперативные системы должны обеспечивать приемлемую производительность при выполнении своей прямой задачи: обработки операций. Генерация отчетов ложится на них дополнительной нагрузкой и замедляет работу. Более того, если отчет охватывает зоны ответственности разных OLTP-систем или требует подключения внешних данных, то часто требуется какое-то программирование, экспорт в промежуточные форматы и дополнительные расчеты.

На корректность получаемой информации негативно влияет также и то, что порядок и формат ввода оперативных данных в OLTP-системы зачастую не отвечают требованиям их аналитической и статистической обработки. Данные могут повторяться или быть сохранены в виде, затрудняющем их машинную обработку. И наконец, оперативные системы накладывают существенные ограничения на хранение данных за большой период времени, т.к. попросту не предназначены для этого.

Попытки разделить отчетность и обработку транзакций предпринимались неоднократно. Одной из таких попыток является

создание копий («реплик») оперативных систем для получения отчетов. При этом данные из оригинальных оперативных систем реплицируются в системы-копии без каких бы то ни было преобразований. Эта мера позволяет решить только одну проблему: снятие с оперативных систем дополнительной нагрузки.

Другим подходом к решению этой проблемы является построение *витрин данных* (англ. Data Mart), поскольку витрина – это не просто копия исходной базы, а хранилище, оптимизированное для построения отчетов и анализа данных.

Решить проблемы хранения и анализа данных в масштабе всего предприятия призвана технология *хранилищ данных* (англ. Data Warehouse). Хранилище данных представляет собой единый *централизованный источник* информации (рис. 13).

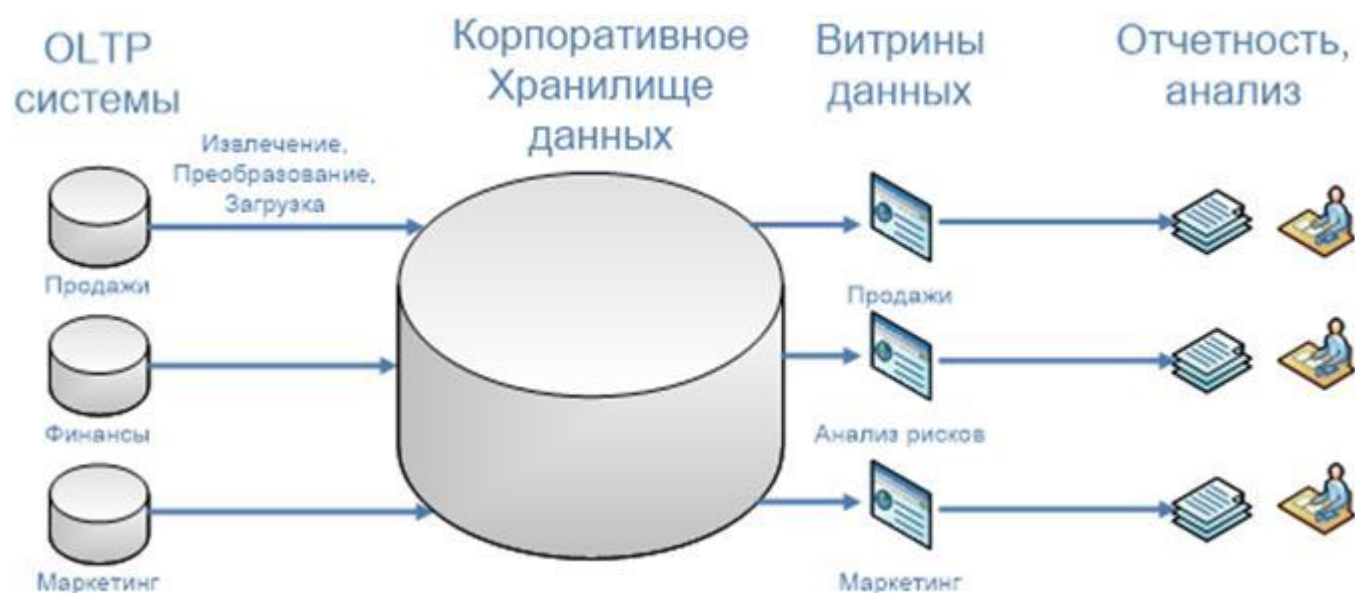


Рис. 13. Централизованный источник данных на основе ХД

Хранилище может содержать данные не только из OLTP-систем, но и из других разнообразных источников, таких как внешние базы данных, базы унаследованных (существовавших на предприятии раньше и впоследствии замененных) информационных систем или отдельные файлы (например, текстовые файлы или документы форматов Microsoft® Office).

Аналитическая отчетность. Хранилища данных обычно консолидируют информацию по всем аспектам деятельности предприятия и используются для углубленного анализа данных. Результатом аналитической деятельности являются итоговые аналитические отчеты, прогнозы, которые используются для принятия как стратегических, так и тактических решений.

Вопрос 2. Технологии интеграции данных.

Очевидно, что основой для принятия обоснованных решений должен быть единый источник достоверных данных в масштабах всей компании.

Интеграция данных позволяет представить корпоративные данные из разрозненных источников в виде цельной картины, обеспечивающей полноту и удобство доступа к ним. Интеграция данных может быть описана с помощью модели, которая включает приложения, продукты, технологии и методы.

Приложения – это конечные решения, созданные поставщиками информационных систем в соответствии с требованиями клиента. Приложения используют один или несколько продуктов интеграции данных.

Продукты – готовые к использованию решения, поддерживающие одну или несколько технологий интеграции данных.

Технологии представляют собой реализацию одного или нескольких методов интеграции данных.

Метод – это концепция интеграции данных, не привязанная к какому-либо конкретному способу доступа к данным. Выделяют три основных метода интеграции данных: консолидация, федерализация и распространение (рис. 14).

Рис. 14. Методы интеграции данных

Консолидация данных. При консолидации данные извлекаются из разрозненных источников и загружаются в единое постоянное хранилище, чаще всего, корпоративное хранилище данных. При этом обеспечивается единство структуры данных, что значительно упрощает их обработку и анализ.

При использовании этого метода обычно существует задержка между моментом обновления данных в первичных системах и временем, когда произошедшие изменения появляются в конечном месте хранения. Режим, когда данные в хранилище обновляются одновременно с источником, называется «режимом реального времени», но при консолидации данных достичь этого весьма проблематично. Для описания данных, обновление которых не

сильно отстает от источника (в зависимости от ситуации на несколько секунд, минут или часов), часто используется термин «режим, приближенный к реальному времени».

При наполнении хранилищ данными, которые допускают значительный временной интервал обновления (например, день и более), используются так называемые приложения пакетной интеграции данных. Это приложения, которые извлекают данные из первичных источников и перекачивают их в хранилище по определенному расписанию, например, раз в сутки ночью. Ключевое слово в этой технологии – «извлекают» («вытягивают», *англ.* pull): из исходной системы периодически извлекаются данные, отражающие её состояние на момент извлечения. В этом случае невозможно проследить, как именно менялись данные в промежутке между двумя извлечениями.

Оперативная интеграция запускается в момент изменения данных, т.е. происходит «по событию» (*англ.* event-driven). А пакетная интеграция – «по требованию» (*англ.* on demand).

Бизнес-приложения, которые обрабатывают консолидированное хранилище, могут генерировать запросы к данным, создавать на их основе требуемые отчеты и проводить анализ этих данных.

К сильным сторонам консолидации данных относят то, что этот подход позволяет осуществлять преобразование значительных объемов данных в процессе их передачи от первичных систем к конечным местам хранения.

Следует упомянуть и об определенных сложностях, связанных с данным подходом. В частности, это значительные вычислительные ресурсы, которые требуются для поддержки процесса консолидации данных, а также существенные ресурсы памяти, необходимые для поддержки конечного места хранения.

Консолидация данных – это основной подход, который используется для построения и поддержки оперативных складов данных и корпоративных Хранилищ данных.

Выделяют две технологии консолидации: «Извлечение, Преобразование, Загрузка» (*англ.* «Extract, Transform, Load», сокр. ETL) и «Управление контентом предприятия» (*англ.* Enterprise Content Management, сокр. ECM).

Областью применения ECM являются неструктурированные источники данных, такие как пользовательские документы и веб-страницы. Технология ETL наилучшим образом подходит для извлечения и обработки значительных объемов данных из структурированных источников.

ETL включает в себя этапы *извлечения данных* (Extract) из различных источников, их преобразования (Transform) и загрузки в целевое хранилище (Load). Данные обычно извлекаются («вытягиваются») из структурированных источников, таких как оперативные базы данные (OLTP), структурированные файлы (например, в формате XML) или любые другие.

Этап *преобразования* включает в себя:

- реструктурирование;
- выверку;
- очистку;
- агрегирование данных.

На этапе преобразования данные приводятся к единому стандарту представления, очищаются от ошибок, дубликатов и противоречивых значений. При необходимости данные также могут быть дополнены какими-то дополнительными значениями из других источников («обогащены»).

На конечном этапе проверенные и очищенные данные загружаются в целевое хранилище, на базе которого функционируют приложения оперативного и интеллектуального анализа данных (OLAP, Data Mining), строятся витрины данных, формируются отчеты.

Федерализация данных. Метод *федерализации* представляет требуемые в момент запроса данные из различных источников в виде единого виртуального (не хранящегося на постоянной основе) представления. В отличие от консолидации данные не перемещаются из источников в хранилище. Когда бизнес-приложение генерирует запрос, используя виртуальное представление, то процессор федерализации данных *извлекает* требуемые данные из соответствующих распределенных источников, представляет их таким образом, чтобы они отвечали требованиям запроса, и отправляет результаты бизнес-приложению, от которого пришел запрос.

Федерализация данных предоставляет доступ к данным в режиме реального времени. Процесс федерализации данных всегда заключается в извлечении данных из первичных систем «по требованию». Все необходимые преобразования данных осуществляются при их извлечении из первичных источников.

Один из ключевых элементов метода федерализации – это метаданные, которые используются процессором федерализации данных для доступа к первичным данным.

Считается, что основное преимущество федеративного подхода в том, что он обеспечивает доступ к актуальным на каждый

момент времени данным и избавляет от необходимости консолидировать первичные данные в новом складе данных. Однако федерализация не может служить заменой консолидации данных.

Этот метод не очень хорошо подходит для извлечения и преобразования больших массивов данных или для тех приложений, где существуют серьезные проблемы с качеством данных в первичных системах. Еще один существенный фактор – негативное влияние федеративных запросов на производительность первичных систем и дополнительные затраты на доступ к многочисленным источникам данных.

В связи с этим федерализацию целесообразно применять для расширения и усиления возможностей среды хранилища данных в соответствии со специфическими потребностями бизнеса. Федерализация данных также применяется в тех случаях, когда политика безопасности данных и лицензионные ограничения запрещают копирование данных первичных систем.

Технологией, которая поддерживает метод федерализации, является «*Интеграция корпоративной информации*» (англ. Enterprise Information Integration, сокр. EII).

Распространение данных. В основе метода *распространения* лежит копирование данных из системы-источника в одну или несколько систем-получателей. Такое копирование осуществляется посредством передачи изменений, произошедших в одной системе, в другие связанные с ней системы. Ключевое понятия метода распространения – «передают» («проталкивают»). Если при консолидации и федерализации данные извлекаются из исходных систем, то при распространении сами первичные системы передают произошедшие изменения системам-получателям.

Благодаря такому подходу реализуется главное преимущество метода – обеспечение актуальности данных в режиме реального времени или близком к нему. Так как передача данных инициируется при наступлении определенного события, то метод распространения является *событийным* (англ. event-driven).

Обновления в первичной системе могут передаваться в конечную систему синхронно или асинхронно. *Синхронная передача* требует, чтобы обновления в обеих системах происходили во время одной и той же физической транзакции. *При асинхронной* передаче обновления могут быть разнесены по нескольким транзакциям. Однако вне зависимости от используемого типа синхронизации метод распространения гарантирует доставку данных в систему назначения.

Метод распространения данных может использоваться для уравнивания рабочей нагрузки между системами, создания резервных копий и восстановления данных в случае аварийных ситуаций.

Среди технологий, поддерживающих распространение данных, выделяют:

- интеграцию корпоративных приложений (от *англ.* Enterprise Application Integration, сокр. EAI);
- технологию «Извлечение, Загрузка, Преобразование» (от *англ.* Extract, Load, Transform, сокр. E-LT);
- тиражирование корпоративных данных (от *англ.* Enterprise Data Replication, сокр. EDR).

Тиражирование корпоративных данных в основном реализуется в приложениях резервного копирования, архивирования и равномерного распределения нагрузки между различными системами.

В технологии E-LT, в отличие от ELT, преобразование данных происходит не до, а после их загрузки в целевое хранилище. При этом данные преобразуются не сразу после загрузки, а по пользовательскому запросу или определенному расписанию.

Интеграция корпоративных приложений позволяет различным приложениям «общаться» между собой посредством определенных стандартных интерфейсов. В технологии EAI объем данных, которыми обмениваются приложения, обычно невелик. EAI – это технология, обеспечивающая взаимодействие корпоративных приложений в режиме реального времени различными способами. В качестве *метода интеграции данных* здесь используется *метод распространения*.

Вопрос 3. Рынок средств интеграции приложений.

Комплексные системы интеграции для современных предприятий носят название интеграционных платформ. Ключевые понятия, на которые опираются интеграционные платформы на текущем этапе развития, это:

- сервис-ориентированная архитектура (*англ.* Service-oriented Architecture, сокр. SOA);
- сервисная шина предприятия (*англ.* Enterprise Service Bus, сокр. ESB);
- сервер приложений (*англ.* Application Server).

В основе *сервис-ориентированной архитектуры* лежит идея совокупности программных компонентов: сервисов, имеющих стандартные интерфейсы для использования этих компонентов и доступа к ним посредством сетевых протоколов.

Сервисная шина предприятия представляет собой инфраструктуру для реализации концепции SOA. Под ESB, как правило, подразумевается воплощение транспортного уровня, на котором происходит взаимодействие программных компонентов между собой.

Сервер приложений (также часто называется корпоративным сервером приложений (англ. Enterprise Application Server, сокр. EAS) – это системное программное обеспечение, которое служит контейнером, средой для выполнения всех остальных приложений. На базе этого контейнера строится сервисная шина ESB.

Основные поставщики интеграционных платформ.

Интеграционная платформа от компании **IBM** носит название WebSphere. Основой WebSphere является сервер приложений WebSphere Application Server (WAS), построенный на Java EE. Это масштабная платформа, которая включает в себя десятки компонентов и средств мониторинга и администрирования, функциональность которых может быть существенно расширена за счет дополнительных пакетов (англ. Feature Packs).

Линейка продуктов корпорации **Oracle** представлена двумя семействами:

- 1) перспективная Oracle WebLogic Server (развитие платформы, поглощенной в 2008 г. компании BEA Systems);
- 2) Oracle WebLogic Server (собственная разработка, которая поддерживается, но дальнейшее развитие не планируется).

Семейство WLS включает в себя различные версии сервера приложений и пакет WebLogic Suite, расширяющий функциональные возможности сервера.

Hat (JBoss). JBoss EAS – это основанный на Java EE сервер приложений, который имеет свободно распространяемую версию с открытым исходным кодом. Платформу интеграции JBoss можно дополнительно расширить средствами построения ESB, инструментами управления бизнес-процессами и другими модулями.

Microsoft. Технология .Net Framework в совокупности с Internet Information Services (сокр. IIS, проприетарный набор серверов для нескольких служб Интернета от компании Microsoft) обеспечивают полное покрытие функционала

серверов приложений, несмотря на то, что продукта с названием «сервер приложений» в линейки компании нет.

Функцию сервисной шины предприятия выполняет Microsoft BizTalk Server, который включает в себя:

- средства управления бизнес-процессами;
- интеграции приложений (EAI);
- адаптеры ко всем основным протоколам и системам.

Развитие SAP на рынке EAS основано на двух ключевых технологиях: ABAP и Java. Java EE сервер приложений SAP NetWeaver Application Server является основой инфраструктуры приложений, в состав которой входят:

- средства построения порталов SAP NetWeaver Portal;
- инструменты управления и интеграции бизнес-процессов SAP NetWeaver BPM и SAP NetWeaver Process Integration.

Вопросы для самопроверки:

1. Какие сложности возникают при использовании оперативных систем в качестве источников данных для аналитических отчетов?
2. Каковы цели интеграции данных?
3. Какие методы интеграции данных вы знаете?
4. Какие задачи решает федерализация данных, а какие консолидация?
5. Какие технологии реализуют метод консолидации? Федерализации?
6. В чем отличие ЕИ от ETL и EAI?
7. Какие этапы включает в себя технологии ETL и E-LT?
8. Что такое сервер приложений?
9. Перечислите известные вам интеграционные платформы.

Литература по теме:

1. Дубова Н. Интеграция приложений и бизнес-процессы – [Электронный ресурс]. – Режим доступа: <http://www.osp.ru/os/2009/10/11171290/>
2. Дубова Н. Краткий курс интеграции данных. – [Электронный ресурс]. – Режим доступа: <http://www.osp.ru/os/2007/09/4567212/> .
3. Информационные аналитические системы: учеб. / Под ред. В.В. Дика. – М.: Московский финансово-промышленный ун-т «Синергия», 2013.

4. Когаловский М.Р. Методы интеграции данных в информационных системах. – [Электронный ресурс]. – Режим доступа: <http://www.ipr-ras.ru/articles/kogalov10-05.pdf>

5. Туманов В.Е. Проектирование хранилищ данных для приложений систем деловой осведомленности (Business Intelligence Systems) // Интернет Университет Высоких Технологий, 2010 г. – [Электронный ресурс]. – Режим доступа: <http://www.intuit.ru/department/database/bispowerd>

Тема 4. Методология оперативной аналитической обработки данных (OLAP)

Цель: ознакомиться с оперативной аналитической обработкой данных.

Задачи:

- Изучить специфику оперативной аналитической обработки данных.
- Познакомиться с требованиями Кодда к средствам оперативной аналитической обработки данных.
- Изучить возможности применения OLAP-технологий в различных сферах бизнеса.

Вопросы темы:

1. Специфика оперативной аналитической обработки данных.
2. Требования Кодда к средствам оперативной аналитической обработки.
3. Сферы применения OLAP-технологий.

Основные понятия:

- *OLAP (On-Line Analytical Process);*
- *многомерный куб (Cube);*
- *операция «Вращение»;*
- *отношения и «Иерархические Отношения»;*
- *операция «Агрегации»;*
- *операция «Детализации»;*
- *формирование «Среза»;*
- *OLAP-отчет;*
- *область активных измерений;*
- *область неактивных измерений;*
- *область фактов;*
- *диаграмма;*

- *терм FASMI (Fast Analysis of Sha Multidimensional Information).*

Вопрос 1. Специфика оперативной аналитической обработки данных.

Оперативная аналитическая обработка данных выражается термином OLAP (On-Line Analytical Process) и очень часто понимается как интерактивная аналитическая обработка данных. Именно возможность работы в интерактивном режиме при анализе данных выгодно отличает OLAP-системы от любых других систем подготовки отчетности, в том числе и регламентированной.

Для эффективной работы аналитику требуется централизация всех данных и соответствующее структурирование информации, а также удобные инструменты для просмотра и визуализации информации. OLAP организует данные в виде многомерных кубов (Cubes). В качестве осей многомерной системы координат выступают основные атрибуты анализируемого бизнес-процесса.

Многомерный анализ определяется как одновременный анализ по нескольким измерениям. По каждому измерению производится консолидация данных. Любое направление консолидации включает серию последовательных уровней обобщения, где каждый вышестоящий уровень соответствует большей степени агрегации данных по соответствующему измерению. Средства оперативной аналитической обработки данных позволяют в любой момент перейти на нужный уровень иерархии.

Аналитик при анализе показателей деятельности, например, объема продаж, оперирует многомерными представлениями. Он рассматривает продажи в зависимости от потребительского спроса на конкретный товар, от региона продаж, от конкретного интервала времени. Исходя из этого, объем продаж можно представить в виде трехмерного куба (рис. 15), грани которого изображают массивы данных по товарам, регионам и времени, а внутри куба находятся значения объема продаж.

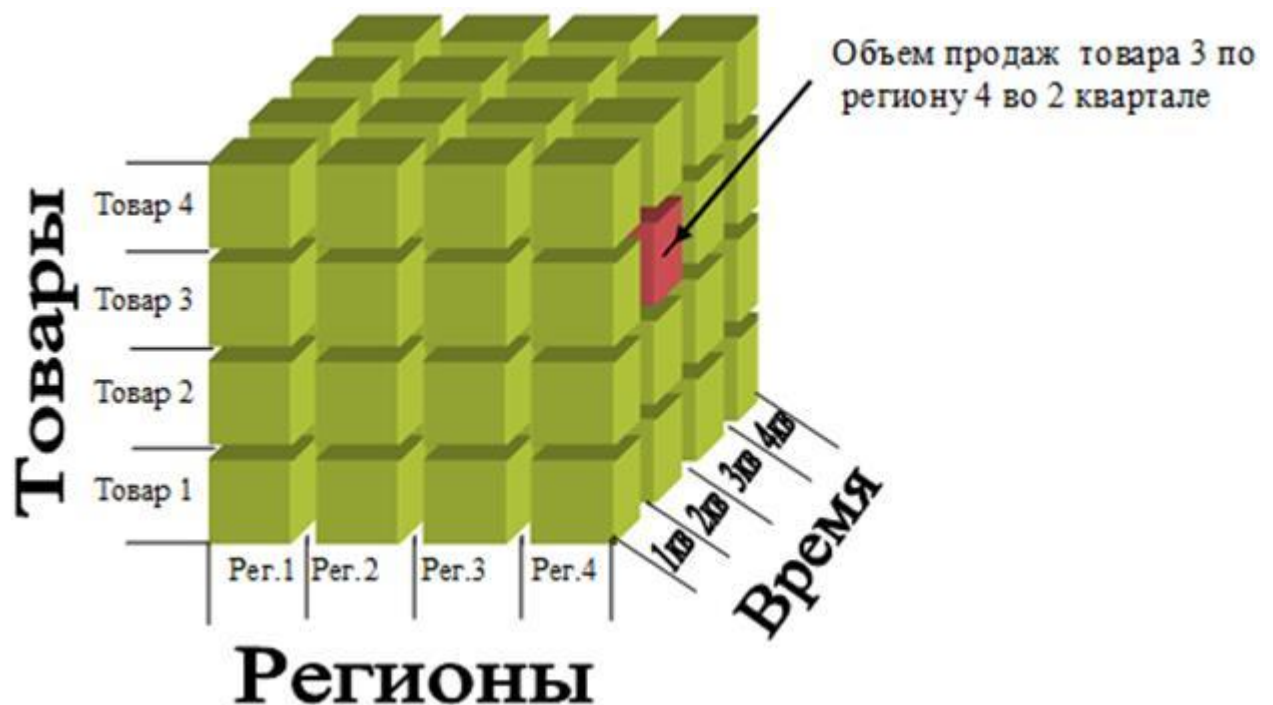


Рис. 15. Информационный куб «Объем продаж»

Однако сам информационный куб для анализа не пригоден. Представить или изобразить адекватно трехмерный куб пользователь еще в состоянии, а вот восьми- или двенадцатимерный – просто невозможно. Поэтому при анализе из многомерного куба извлекают обычные двумерные таблицы.

Методы извлечения информации из кубов данных.

Для извлечения информации из кубов данных используются различные операции манипулирования Измерениями:

1. Операция «Вращение».

Изменение порядка представления (визуализации) Измерений называется Вращением (Rotate). Эта операция обеспечивает возможность визуализации данных в форме, наиболее комфортной для их восприятия.

2. Отношения и Иерархические Отношения.

При рассмотрении информационного куба значения Показателей определяются только тремя измерениями. На самом деле их может быть гораздо больше и между их значениями обычно существуют множество различных Отношений (Relation). В свою очередь, множество Отношений может иметь иерархическую структуру – Иерархические Отношения.

3. Операция Агрегации.

В процессе анализа пользователь переходит от детализированных данных к агрегированным (обобщенным), т.е.

производит операцию Агрегации (Drill Up). При этом значения детальных показателей суммируются в агрегируемый показатель.

4. Операция Детализации.

Переход от более агрегированных к более детализированным данным называется операцией Детализации (Drill Down). При этом осуществляется декомпозиция признака агрегации на компоненты, например, признак года разбивается на кварталы. При этом автоматически детализуются числовые показатели.

5. Формирование «Среза».

Пользователя редко интересуют все потенциально возможные комбинации значений Измерений. Более того, он практически никогда не работает одновременно сразу со всем кубом данных. Подмножество куба, получившееся в результате фиксации значения одного или более Измерений, называется Срезом (Slice), а сама операция называется «разрезанием» куба. Аналитик как бы берет и «разрезает» измерения куба по интересующим его меткам. Этим способом аналитик получает двумерный срез куба (отчет) и с ним работает. Структура отчета представлена на рисунке 16.

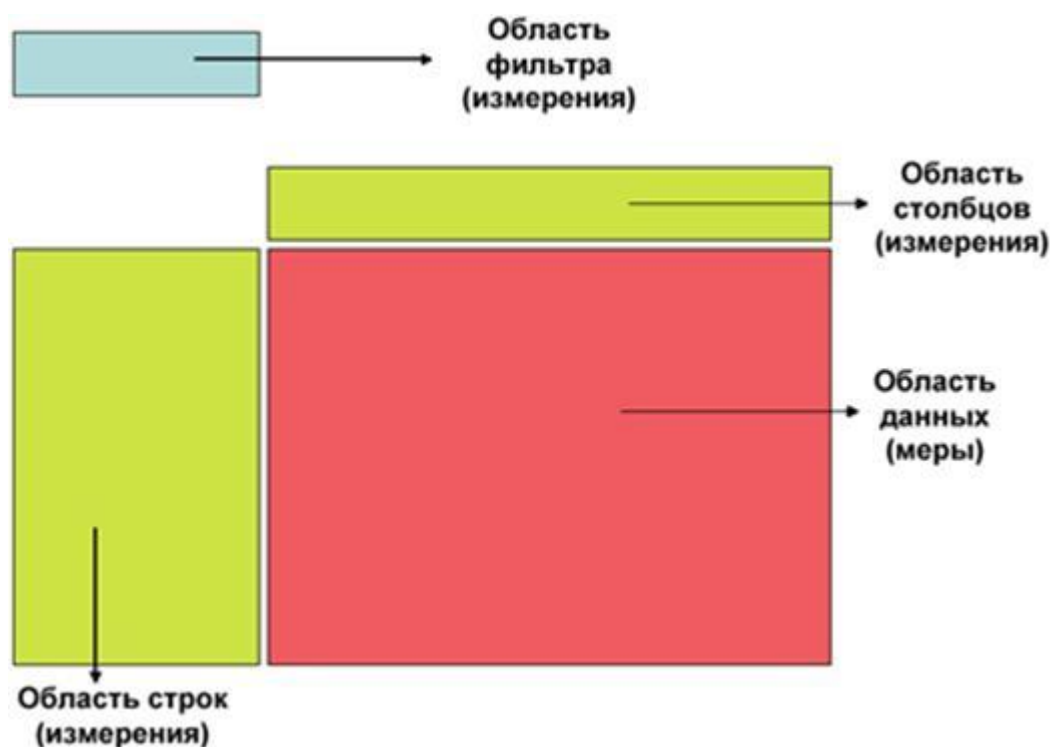


Рис. 16. Структура аналитического отчета

Например, если мы ограничим значение Измерения Товар – Товаром 2 (рис. 15), то получим подмножество куба (в данном случае – двумерную таблицу), содержащее информацию об истории продаж данного товара 2 в разные регионы в различные

временные периоды (рис. 17). Объем продаж здесь выражается одной мерой, например, количеством проданного товара.



Рис. 17. Поквартальный отчет о продажах товара 2

С точки зрения конечного пользователя суть OLAP-технологии состоит в том, что данные ему предоставляются в динамической таблице, автоматически суммирующей их в различных разрезах, и позволяющей интерактивно управлять вычислениями и формой отчета. Инструментами управления отчетом являются элементы самой таблицы.

На рисунке 18 представлен OLAP – отчет, созданный в системе Contour Reporter (компании Contour Components).

OLAP-отчет – управляемая динамическая OLAP-таблица, которая сопровождается синхронной диаграммой (графиком).

Измерения отображаются в названиях строк и столбцов таблицы, соответствующие им факты и итоги (агрегированные факты) – в ячейках таблицы. Колонки и строки являются инструментами управления таблицей. Пользователь может перемещать их, фильтровать, сортировать, детализировать/обобщать и выполнять другие OLAP-операции. При этом таблица автоматически вычисляет новые итоги (агрегаты). Управляя OLAP-таблицей, можно из одного набора данных сформировать множество отчетов.

В OLAP-таблице можно условно выделить несколько рабочих областей (рис. 18):

Область активных измерений – строки и столбцы OLAP-таблицы.

Область неактивных измерений (фильтры) – область, содержащая измерения, не отображенные в OLAP-таблице, но влияющие на представленные в ней данные.

Область фактов – таблица с числовыми данными, над которыми выполняются вычисления.

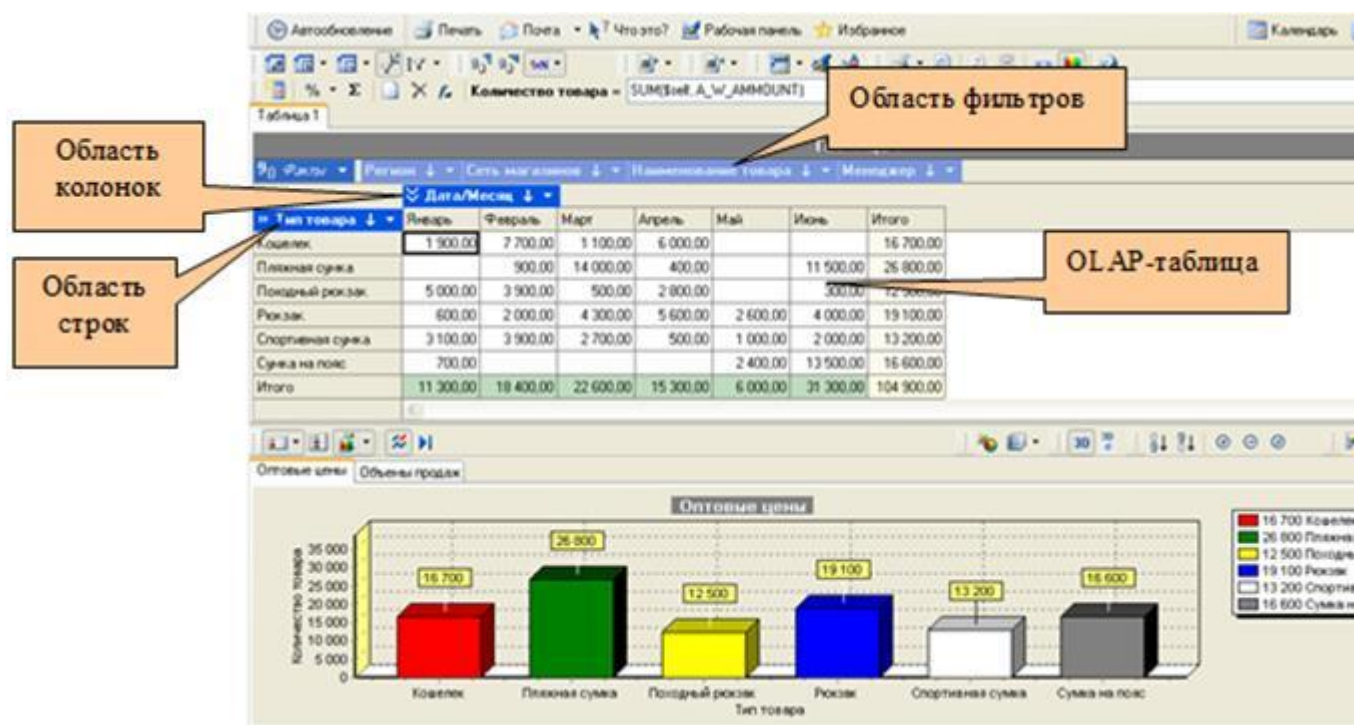


Рис. 18. OLAP-отчет

Диаграмма – синхронное с OLAP-таблицей графическое представление данных. Диаграмма строится по расположенным в крайнем левом положении элементам динамической таблицы, т.е. по крайнему левому измерению-строке, измерению-колонке и фактам (крайне левому или по всем).

Меняя колонки и строки, пользователь самостоятельно меняет вид отчета и группировки данных, система вычисляет новые итоги, суммируя тысячи, а то и миллионы строк.

Один из базовых принципов OLAP – способность интуитивно манипулировать данными, требует быстрого извлечения информации. Средства OLAP как раз и являются такими инструментами, обеспечивающими оперативность и гибкость в получении различных отчетов, существенно расширяя возможности участников процесса подготовки и принятия решений.

Вопрос 2. Требования Кодда к средствам оперативной аналитической обработки.

Аббревиатура **OLAP** была впервые введена Э.Ф. Коддом (E.F. Kodd), известным ученым в области реляционных БД, создателем широко распространенной реляционной модели данных. Кодд сформулировал концепцию комплексного многомерного анализа данных, накопленных в хранилище, в виде 12 основных правил, которым должны удовлетворять OLAP-системы, как продукты, предоставляющие возможность выполнения оперативной аналитической обработки (табл. 2). В 1995 г. к 12 основным требованиям были добавлены еще шесть (которые известны в значительно меньшей степени).

Таблица 2.

Основные требования Кодда к средствам OLAP

№	Название требования	Пояснения
1.	Многомерное представление данных	Средства должны поддерживать многомерный на концептуальном уровне взгляд на данные.
2.	Прозрачность	Пользователь не должен знать о том, какие конкретные средства используются для хранения и обработки данных, как данные организованы и откуда они берутся.
3.	Доступность	Средства должны сами выбирать и связываться с наилучшим для формирования ответа на данный запрос источником данных. Средства должны обеспечивать автоматическое отображение их собственной логической схемы в различные гетерогенные источники данных.
4.	Согласованная производительность	Производительность практически не должна зависеть от количества Измерений в запросе.
5.	Поддержка архитектуры клиент-сервер	Средства должны работать в архитектуре клиент-сервер.
6.	Равноправность всех измерений	Ни одно из измерений не должно быть базовым, все они должны быть равноправными (симметричными).
7.	Динамическая обработка разреженных матриц	Неопределенные значения должны храниться и обрабатываться наиболее эффективным способом.
8.	Поддержка многопользовательского режима работы с данными	Средства должны обеспечивать работу более чем одному пользователю.
9.	Поддержка операций на основе различных измерений	Все многомерные операции (например, Агрегация) должны единообразно и согласованно применяться к любому числу любых измерений.
10.	Простота манипулирования данными	Средства должны иметь максимально удобный, естественный и комфортный пользовательский интерфейс.
11.	Развитые средства представления данных	Средства должны поддерживать различные способы визуализации (представления) данных.
12.	Неограниченное число измерений и уровней агрегации данных	Не должно быть ограничений на число поддерживаемых Измерений.

Все правила Кодда разделены на четыре группы и названы «характеристиками» или особенностями (features). К ним относятся:

1. **Основные характеристики:** многомерность модели данных, интуитивные механизмы манипулирования данными, доступность данных, пакетное извлечение данных, архитектура «клиент-сервер», прозрачность, многопользовательская работа.

2. **Специальные характеристики:** обработка ненормализованных данных, хранение результатов отдельно от исходных данных, выделение отсутствующих данных, обработка отсутствующих значений.

3. **Характеристики построения отчетов:** гибкое построение отчетов, стабильная производительность при построении отчетов, автоматическое регулирование физического уровня.

4. **Управление размерностью:** общая функциональность, неограниченное число измерений и уровней агрегирования, неограниченные операции между данными различных измерений.

В начале 1995 г. Найджел Пендс (Nigel Pendse) и Ричард Крит (Richard Creeth), считая, что для большинства людей слишком обременительно помнить все правила Кодда (18 правил), переработали их в так называемый тест **FASMI** (Fast Analysis of Sha Multidimensional Information). В тесте OLAP-определение резюмировано только пятью ключевыми словами: Быстрый Анализ Разделяемой Многомерной Информации, т.е. характеристики OLAP-средства определены специфическим образом без указания на то, каким образом само средство должно быть реализовано. Это определение по достоинству было оценено и до сих пор используется для характеристики OLAP-средств.

Тест FASMI расшифровывается следующим образом:

- **FAST (Быстрый)** означает, что система должна обеспечивать выдачу большинства ответов пользователям в пределах приблизительно пяти секунд.
- **ANALYSIS (Анализ)** – система может справляться с любым логическим и статистическим анализом.
- **SHA (Разделяемой)** означает, что система осуществляет все требования защиты конфиденциальности.
- **MULTIDIMENSIONAL (Многомерной)** – это ключевое требование. Система должна обеспечить многомерное концептуальное представление данных, включая полную поддержку для иерархий и множественных иерархий.
- **INFORMATION (Информации)** – необходимая информация должна быть получена там, где она необходима.

В настоящее время технология OLAP получила широкое признание. Рост популярности OLAP приводит к дальнейшему расширению сферы ее применения.

Вопрос 3. Сферы применения OLAP-технологий.

OLAP-технологии применимы везде, где есть задача анализа многофакторных данных. Вообще, при наличии некоторой таблицы с данными, в которой есть хотя бы одна описательная колонка и одна колонка с цифрами OLAP-инструмент, как правило, будет эффективным средством анализа и генерации отчетов.

Рассмотрим некоторые сферы применения OLAP-технологий, взятые из реальной жизни. Этот набор идей при желании может быть расширен на другие сферы бизнеса.

Продажи. Основное направление анализа в этой сфере – это рассмотрение объемов продаж в различных разрезах: Время, Категория товара, Товар, Регион, Покупатель, Продавец и пр. Эта информация необходима для принятия управленческих решений об изменении ассортимента, цен, закрытии и открытии магазинов, филиалов, расторжении и подписании договоров с дилерами, проведения или прекращения рекламных кампаний и т.д.

Закупки. Многие предприятия закупают комплектующие и материалы у поставщиков. Торговые предприятия закупают товары для перепродажи. Возможных задач при анализе закупок множество: от планирования денежных средств на основе прошлого опыта, до контроля за менеджерами, выбирающими поставщиков.

Маркетинг. Маркетинговый анализ необходим для планирования развития бизнеса в любой сфере. Например, задачами этого анализа в розничной торговле являются: правильное позиционирование товара, выявление групп покупателей для целевой рекламы, оптимизация ассортимента. Данные для такого анализа получить сложно. В тех случаях, когда покупатель анонимен, а это для розничной торговли правило, для получения данных используются социологические опросы, периодическая регистрация и прочие не прямые способы получения информации. Фактами при этом служат: объемы покупок, количество купленных товаров, количество самих покупателей или клиентов.

Движение денежных средств – это целая область анализа, имеющая множество школ и методик. OLAP-технология может служить инструментом реализации или усовершенствования данных методик, но никак не их заменой. Анализируются денежные обороты безналичных и наличных средств в разрезе бизнес-

операций, контрагентов, валют и времени с целью оптимизации потоков, обеспечения ликвидности и т.д. Состав измерений сильно зависит от особенностей бизнеса, отрасли, методики.

Бюджет – одна из самых благодатных областей применения OLAP-технологий. Недаром ни одна современная система бюджетирования не считается завершенной без наличия в ее составе OLAP-инструментария для анализа бюджета. Большинство бюджетных отчетов легко строятся на основе OLAP-систем. При этом отчеты отвечают на очень широкую гамму вопросов:

- анализ структуры расходов и доходов;
- сравнение расходов по определенным статьям у разных подразделений;
- анализ динамики и тенденций расходов на определенные статьи;
- анализ себестоимости и прибыли.

Большинство современных методик предлагают для крупных организаций вести бюджет в разрезе организационной структуры (по подразделениям) и в разрезе финансовой структуры (по центрам учета: центрам прибыли и центрам расходов). Самый мощный анализ можно провести, если бюджет ведется в разрезе всех аналитических признаков, интересующих руководство предприятия.

Посещаемость сайта. Лог-файл Интернет-сервера многомерен по природе, а значит подходит для OLAP-анализа. Фактами являются: количество посещений, количество хитов, время, проведенное на странице и другая информация, имеющаяся в логе. Кроме того, анализ можно расширить за счет дополнительных справочников. Например: Хост, Ссылающийся сервер, Поисковая система, Посетитель.

Использование помещений – один из видов статистического анализа. Примеры: анализ загруженности учебных аудиторий, сдаваемых в аренду зданий и помещений, использования залов для конференций и пр.

Заработная плата – анализ расходов на зарплату, сравнение расходов по специальностям, филиалам, людям, динамика заработной платы.

Текучесть кадров на предприятии – анализ текущесть кадров на предприятии в разрезе филиалов, отделов, профессий, уровня образования, пола, возраста, времени.

Грузовые перевозки – анализ объемов перевозок, платы в разрезе сезонов, направлений, видов вагонов, грузов, грузоотправителей, грузополучателей, станций отправления, станций получения.

Рассмотрим более подробно проведение OLAP анализа в некоторых сферах бизнеса.

Анализ продаж. Представьте себе многофилиальную компанию с главным офисом в Москве. В состав компании входят филиалы, распределенные по регионам. В крупных городах этих регионов интересы компании представляют дилеры. С дилерами установлены давние тесные связи, позволяющие регулярно получать от них информацию по объемам реализации товаров в различных разрезах. Номенклатура товаров компании достаточно широка.

Анализ продаж такой компании может преследовать различные цели:

- подготовка стратегических управленческих решений в области продаж с перспективой на полугодие или год;
- формирование плана продаж на ближайший квартал и его анализ;
- ежемесячный мониторинг состояния продаж;
- еженедельная подготовка общих отчетов о деятельности компании и частных отчетов по отдельным товарным группам и филиалам, требующим более пристального контроля.

Основное различие между перечисленными задачами анализа продаж заключается в степени детализации исследуемых данных и размере анализируемых периодов времени.

1. Определим доходность товарных групп в целом по компании (рис. 19).

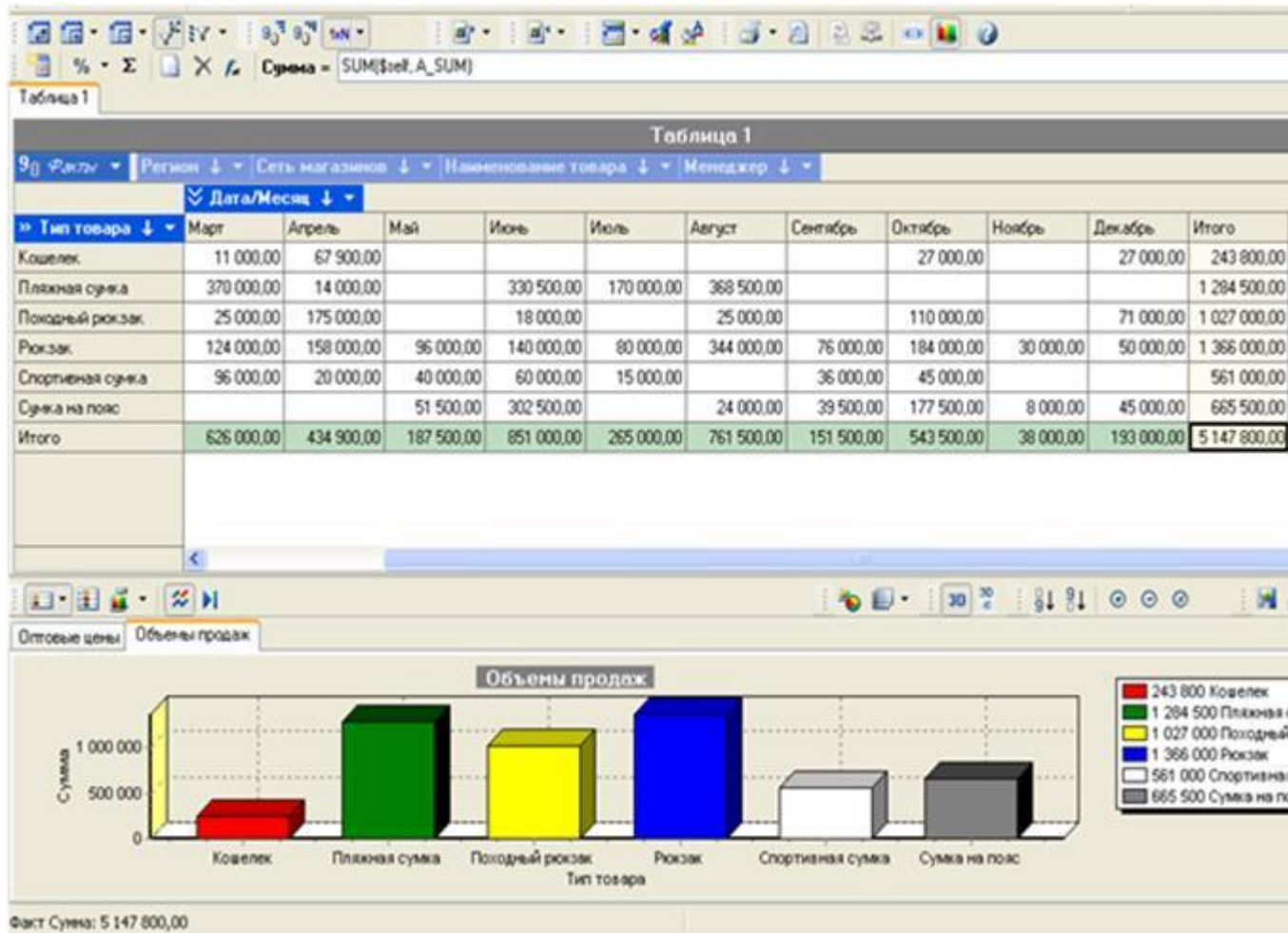


Рис. 19. Доходность товарных групп

Если мы хотим посмотреть динамику продаж групп товаров в зависимости от времени, то нужно поменять тип диаграммы (рис. 20).



Рис. 20. Динамика продаж групп товаров

2. Проанализируем поквартальные объемы продаж по регионам.

Обобщив все данные по измерению «Регион» и оставив в области колонок измерение по времени, соответствующее анализируемому периоду, мы получим интересующую нас информацию (рис. 21).

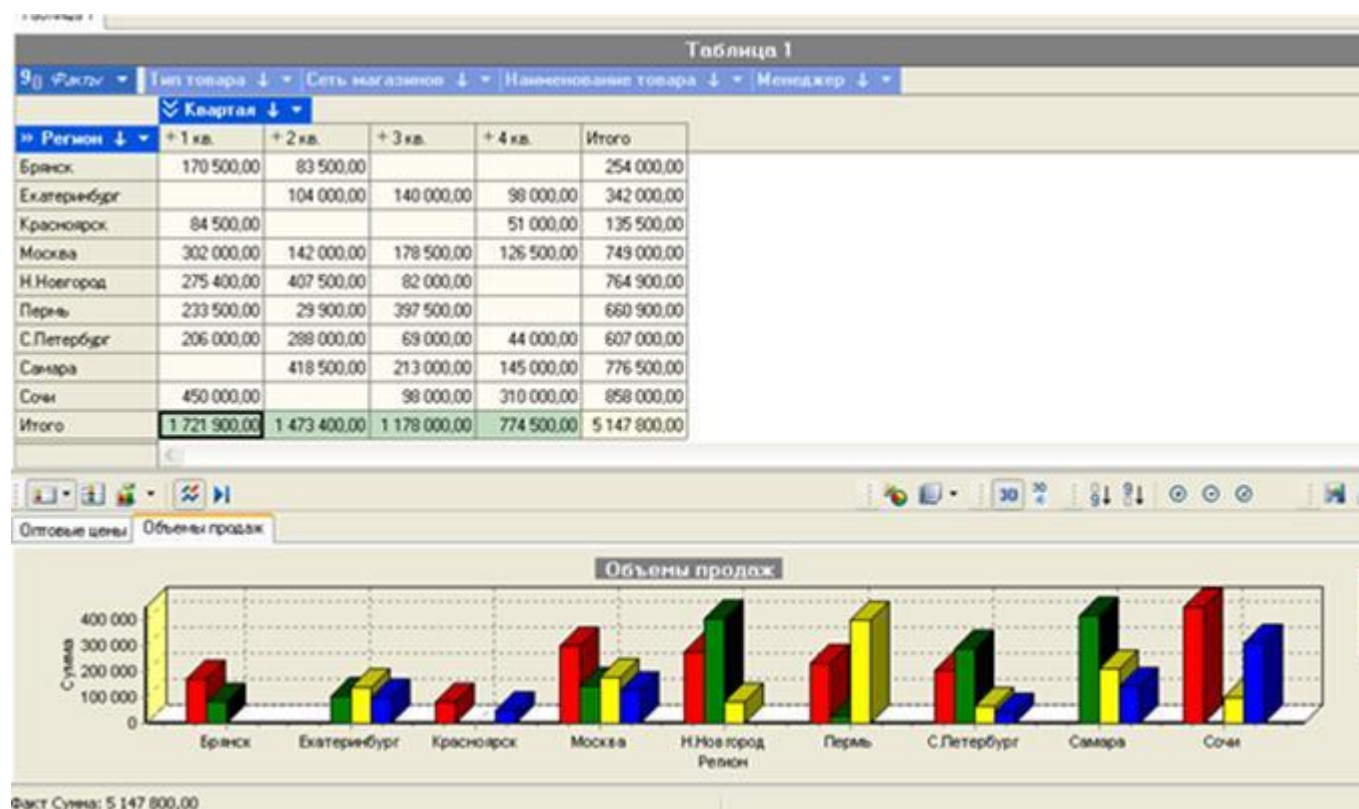


Рис. 21. Поквартальный отчет о продажах по регионам

Вопросы для самопроверки:

1. Что такое тест FASMI?
2. Опишите структуру OLAP-отчета.
3. Что такое агрегация данных?
4. Можно ли поворотом информационного куба получить новый состав данных?
5. Возможен ли одновременный анализ данных по нескольким измерениям?
6. Чем отличаются базовые операции оперативного анализа данных: поворот и срез?
7. Каким образом из информационного куба можно получить отчет в виде двумерной таблицы?
8. В чем заключается интерактивность пользовательского отчета?

9. Охарактеризуйте ключевое требование к OLAP-средствам.
10. Какие виды анализа, можно проводить с помощью OLAP-технологии?

Литература по теме:

1. Барсегян А.А. Технологии анализа данных. Data Mining, Visual Mining, Text Mining, OLAP. / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. – СПб.: БХВ-Петербург, 2009.
2. Белов, В.С. Информационно-аналитические системы: основы проектирования и применения: учеб.-прак. пособие / В.С. Белов. – 2-е изд., перераб. и доп. – М.: Евразийский открытый институт, 2010. – [Электронный ресурс]. – Режим доступа: <http://biblioclub.ru/index.php?page=book&id=90540>.
3. Информационные аналитические системы: учеб. / Под ред. В.В. Дика. – М.: Московский финансово-промышленный ун-т «Синергия», 2013.
4. Найгель Пендс. Что следует понимать под термином OLAP? / перевод Шамиля Абушаева. – [Электронный ресурс]. – Режим доступа: www.corportal.ru/Articles/DataTech/OLAP.
5. Пивоваров А. Что такое OLAP? – [Электронный ресурс]. – Режим доступа: www.corportal.ru/Articles/DataTech/OLAP.
6. Щавелёв Л.В. Оперативная аналитическая обработка данных: концепции и технологии. – [Электронный ресурс]. – Режим доступа: www.corportal.ru/Articles/DataTech/OLAP.

Тема 5. Инструментальные средства оперативной аналитической обработки данных (OLAP)

Цель:

Сформировать представление об инструментальных средствах оперативной аналитической обработки данных (OLAP).

Задачи:

- Ознакомиться с классификацией OLAP решений.
- Изучить структуру OLAP решений.
- Освоить принципы работы OLAP клиентов.
- Ознакомиться с инструментальными OLAP средствами, присутствующими на рынке ПО.

Вопросы темы:

1. Классификация и структура OLAP решений.
2. Принципы работы OLAP клиентов.

3. Выбор архитектуры OLAP решения.
4. Рынок OLAP решений.

Основные понятия:

- *MOLAP (Multidimensional OLAP);*
- *ROLAP (Relational OLAP);*
- *HOLAP (Hybrid OLAP);*
- *OLAP-машины;*
- *OLAP-серверы;*
- *OLAP-клиенты;*
- *OLAP-продукты;*
- *OLAP-компоненты;*
- *семантический слой;*
- *бизнес-объект;*
- *Analyze then query;*
- *Query then analyze.*

Вопрос 1. Классификация и структура OLAP решений.

На сегодняшний день в мире разработано множество продуктов, реализующих OLAP-технологии. Чтобы легче было ориентироваться среди них, используют классификации OLAP-продуктов по:

- способу хранения данных;
- месту нахождения OLAP-машины;
- степени готовности к применению.

Классификация по способу хранения данных.

OLAP-продукты по способу хранения данных делятся на три аналогичные категории:

1. MOLAP (Multidimensional OLAP).
2. ROLAP (Relational OLAP).
3. HOLAP (Hybrid OLAP).

Multidimensional OLAP. В случае MOLAP, исходные и агрегированные данные хранятся в многомерной БД или в многомерном локальном кубе: специальном формате, определенном OLAP-приложением. Данные передаются от источника данных в многомерную БД, а затем БД подвергается агрегации. Предварительный расчет агрегатов ускоряет OLAP-запросы. Такой способ хранения обеспечивает высокую скорость выполнения OLAP-операций.

Relational OLAP. В ROLAP-продуктах исходные данные хранятся в реляционных БД или в плоских локальных таблицах на файл-сервере. Агрегатные данные могут помещаться в служебные таблицы в той же БД. Преобразование данных из реляционной БД в многомерные кубы происходит по запросу OLAP-средства. При этом скорость построения куба будет сильно зависеть от типа источника данных и порой приводит к неприемлемому времени отклика системы. Главный недостаток ROLAP по сравнению с многомерными СУБД – меньшая производительность.

Hybrid OLAP. При использовании Гибридной архитектуры исходные данные остаются в реляционной базе, а агрегаты размещаются в многомерной. Построение OLAP-куба выполняется по запросу OLAP-средства на основе реляционных и многомерных данных. Такой подход позволяет избежать взрывного роста данных. При этом можно достичь оптимального времени исполнения клиентских запросов.

Классификация по месту размещения OLAP-машины.

Существует два варианта расположения OLAP-машины – вычислительного ядра любой OLAP-системы – на центральном сервере или на стороне клиента. По этому признаку OLAP-продукты делятся на OLAP-серверы и OLAP-клиенты.

OLAP-серверы. В серверных OLAP-средствах вычисления и хранение агрегатных данных выполняются на сервере. Сервер в OLAP-системах является источником, поставляющий данные для анализа. Он выполняет всю работу (в зависимости от модели системы), и хранит в себе всю информацию, к которой обеспечивается активный доступ. Клиентское приложение получает только результаты запросов к многомерным кубам, которые хранятся на сервере.

OLAP-клиенты. OLAP-клиент устроен по-другому. Построение многомерного куба и OLAP-вычисления выполняются в памяти клиентского компьютера.

Классификация по степени готовности к применению. Различают:

1. OLAP-компоненты.
2. Инструментальные OLAP – системы.
3. Конечные OLAP-приложения.

OLAP-компоненты. OLAP-компонент – это инструмент разработчика. С его помощью разрабатываются клиентские OLAP-программы. Различают MOLAP и ROLAP-компоненты.

MOlap-компоненты являются инструментами генерации запросов к OLAP-серверу. Они также обеспечивают визуализацию полученных данных.

ROLAP-компоненты содержат собственную OLAP-машину, которая обеспечивает построение OLAP-кубов в оперативной памяти и отображает их на экране. Одна из наиболее доступных OLAP-компонент – это Decision Cube в составе Borland Delphi.

Инструментальные OLAP-системы – это программные продукты, предназначенные для создания аналитических приложений. Различают две категории инструментальных OLAP-систем:

1) системы для программирования – это среда разработчика аналитических систем;

2) системы для быстрой настройки – это средства, которые предоставляют визуальный интерфейс для создания OLAP-приложений без программирования.

Конечные OLAP-приложения. Это готовые прикладные решения для конечного пользователя. Они требуют только установки, и, не всегда, настройки под специфику пользователя. Пример такого решения – это OLAP-приложения Аналитической платформы Контур, подготовленные для анализа данных в различных отраслях и для решения различных аналитических задач.

Структура OLAP решений.

OLAP-система состоит из следующих крупных блоков:

1. Внешний источник данных.
2. OLAP-машина.
3. Многомерный куб.
4. Инструменты отображения и управления.

1. **Источник данных.** В качестве источника данных выступает реляционная или не реляционная БД, текстовый файл, электронная таблица и др. Система может быть постоянно связанной с источником данных, чтобы получать новые порции данных в процессе работы пользователя по его командам или, получив данные однократно, в дальнейшем работать в режиме off-line. Источником данных может служить любая внешняя информационная система: АБС, бэк-офисная система, БД отдела, Хранилище данных.

2. **OLAP-машина** – это программный модуль, который производит вычисления:

- Преобразует плоские данные в многомерные в момент создания куба.

- Выполняет запросы пользователя в режиме реального времени в момент работы с отчетом.

3. Многомерная база данных (куб). Куб – это многомерная БД, хранящая два вида полей: измерения и факты в иерархической структуре, оптимизированной для мгновенного получения агрегатов фактов на пересечении значений (или членов) измерений.

Куб может храниться на диске или в оперативной памяти компьютера. Он может содержать предварительно вычисленные агрегаты или предоставлять специальные выборки данных для расчета этих агрегатов в режиме реального времени. Его структура обеспечивает мгновенное выполнение манипуляций над данными: фильтрацию, поиск, изменение правил агрегации (взаимного расположения измерений) и т.д.

4. Инструменты отображения и управления. Для проектирования куба, отображения данных и интерактивного управления отчетом применяются различные визуальные компоненты. Средства отображения и управления могут быть отдельными или интегрированными, например, пользователь может давать команду на выполнение расчетов по новым правилам агрегации, перемещая колонки таблицы или используя кнопки специальной инструментальной панели.

Вопрос 2. Принципы работы OLAP клиентов.

Во время настройки куба проектировщик описывает запрос к исходной БД и то, как поля плоского источника будут преобразованы в поля многомерного куба (измерения и факты).

Во время наполнения куба OLAP-машина получает из источника набор данных и загружает их в куб, перекладывая плоские данные в иерархическую, многомерную структуру хранения, создавая специальные индексы и вычисляя агрегаты (не всегда). При этом многомерный куб физически может сохраняться на диске или создаваться «на лету» в оперативной памяти, в зависимости от реализации системы.

Во время настройки отчетов визуальные компоненты – средства отображения и управления пользователя – настраиваются на поля многомерного куба, эти поля переименовываются, задается их формат и другие параметры отображения, расположение полей в областях таблицы или осях диаграммы.

Во время работы с отчетом, OLAP-машина получает команды пользователя, производит вычисления и возвращает иерархические наборы данных. При этом OLAP-машина может работать в

режиме online с источником данных, по мере надобности генерируя новые запросы к реляционному источнику данных (ROLAP-режим), или в режиме offline, используя только данные куба, сохраненные на диске (MOLAP-режим).

Работа OLAP-системы на примере работы OLAP-клиента.

OLAP-клиент – это система, предоставляющая конечному пользователю интерфейс для выполнения произвольных запросов, многомерного анализа и выпуска интерактивных отчетов.

Общий принцип работы OLAP-клиентов – предварительное описание семантического слоя, за которым скрывается физическая структура исходных схем данных, используются данные внешних источников. Семантический слой – это набор каталогов бизнес-терминов, характерных для предметной области и соответствующих определенным данным, а также их параметров и методов извлечения. После его создания пользователь может самостоятельно манипулировать понятными ему объектами в терминах предметной области для создания кубов и аналитических интерфейсов.

Разработка аналитических приложений с помощью клиентских OLAP-средств – процесс быстрый, не требующий специальной подготовки исполнителя. Пользователь, знающий физическую реализацию БД, может разработать аналитическое приложение самостоятельно, без привлечения ИТ-специалиста.

Процесс создания OLAP-приложения с помощью клиентского инструментального средства (рис. 22).

Источниками данных могут быть: локальные таблицы, реляционные базы данных (РСУБД) и др. Список поддерживаемых источников данных определяется конкретным программным продуктом. При создании семантического слоя источники данных – таблицы Sales и Deal – описываются понятными конечному пользователю терминами и превращаются в «Продукты» и «Сделки». Поле «ID» из таблицы «Product» переименовывается в «Код», а поле «Name» – в «Товар» и т.д. Затем создается бизнес-объект «Продажи» – выборка (result set или query), в которой настраиваются: алгоритм объединения таблиц по ключевым полям, условия фильтрации и набор возвращаемых полей.

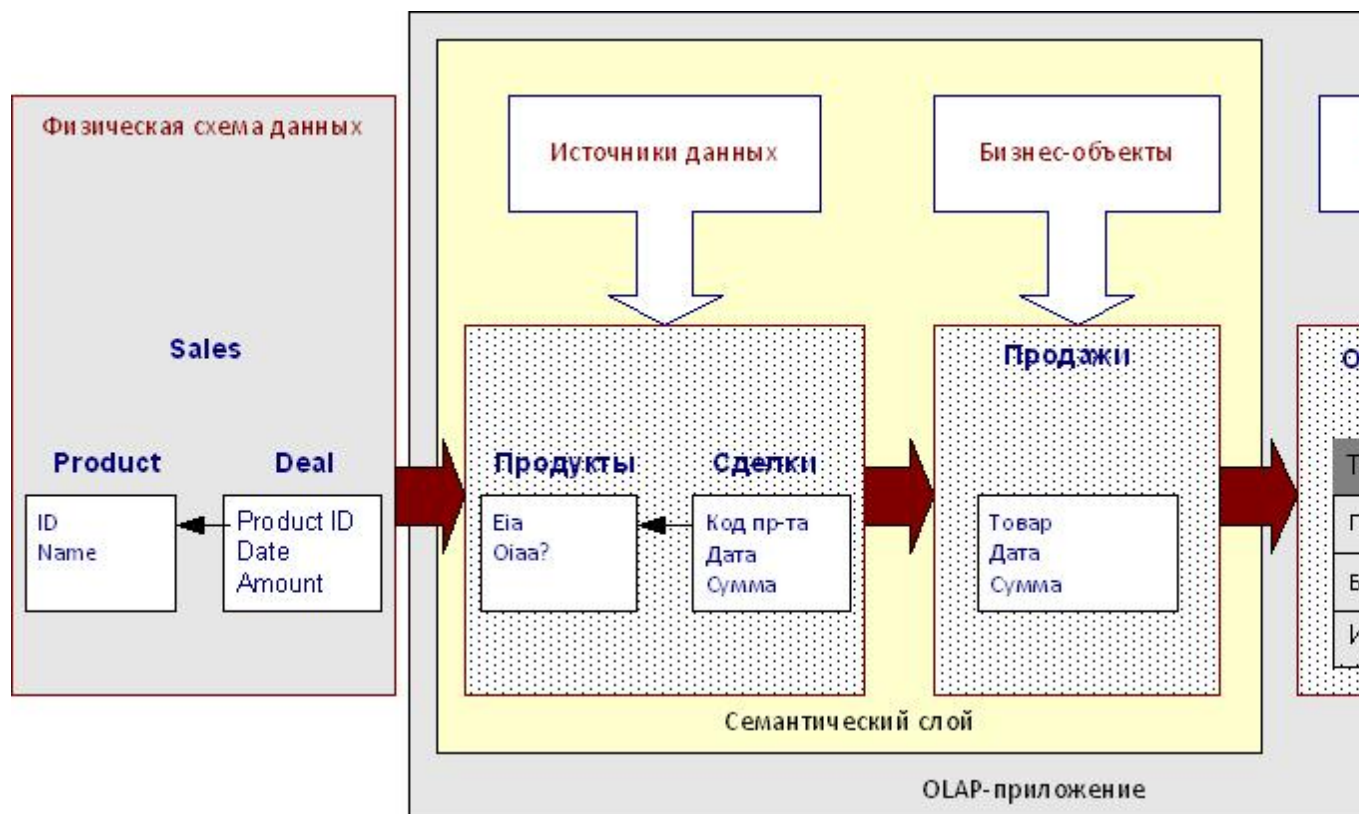


Рис. 22. Создание OLAP-приложения с помощью клиентского ROLAP-средства

Бизнес-объект – это плоская таблица, на основе которой формируется многомерный куб. При создании бизнес-объекта таблицы «Продукты» и «Сделки» объединяются по полю «Код» товара. Поскольку для отображения в отчете не потребуются все поля таблиц, бизнес-объект использует только необходимые, в данном случае, поля: «Товар», «Дата» и «Сумма». Бизнес-объектов может быть несколько.

Далее на базе бизнес-объектов создаются OLAP-отчеты. Пользователь выбирает бизнес-объект и располагает его атрибуты в области колонок или строк таблицы отчета, формируя пользовательский интерфейс. В нашем примере на базе бизнес-объекта «Продажи» создан отчет по продажам товаров по месяцам.

При работе с интерактивным отчетом пользователь может задавать условия фильтрации и группировки данных. В этот момент ROLAP-клиент обращается к данным в памяти. Например, применив в отчете о продажах фильтр по товарам, можно получить отчет о продажах интересующих нас товаров.

Все настройки OLAP-приложения могут храниться в выделенном репозитории метаданных, в приложении или в

системном репозитории многомерной БД. Реализация зависит от конкретного программного продукта.

Разные OLAP-клиенты имеют различную архитектуру. Некоторые из них могут быть развернуты в нескольких вариантах архитектур. OLAP-клиенты могут работать в:

- однопользовательском режиме с локальными данными;
- файл-серверной архитектуре;
- клиент-серверной архитектуре;
- трехуровневой клиент-серверной архитектуре с сервером приложения.

Точно также в OLAP-клиентах может быть организована и работа с метаданными.

Источники данных могут быть локальными таблицами, РСУБД, локальными многомерными кубами или OLAP-серверами.

Вычисляющая OLAP-машина может располагаться на клиенте, в OLAP-сервере, в ROLAP-сервере (сервере приложения).

Может быть реализован «тонкий» Web-клиент, который работает как минимум в трехуровневой архитектуре, где сервером приложения служит интернет-сервер.

OLAP-клиент может быть применен для решения различных бизнес-задач, например, создания:

- аналитической инфраструктуры предприятия силами предприятия или внедренческой фирмы;
- тиражируемого аналитического приложения для популярной OLTP-системы силами независимого разработчика;
- тиражируемого Хранилища данных, или высокоуровневого инструмента для построения Хранилищ данных;
- 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.

Вопрос 3. Выбор архитектуры OLAP-решения.

При реализации информационно-аналитической системы важно не ошибиться в выборе архитектуры OLAP-приложения. Оперативная аналитическая обработка подразумевает, что аналитик интерактивно работает с актуальными данными. При этом

актуальность данных определяется частотой обновления данных в хранилищах, моментом наполнения гиперкуба новыми данными.

Время формирования многомерной БД (гиперкуба) существенно зависит от объема загружаемых в нее данных, поэтому разумно ограничить этот объем. Но как при этом не сузить возможности анализа и не лишить пользователя доступа ко всей интересующей информации?

Существует два альтернативных пути:

1) *Analyze then query* («Сначала проанализируй – затем запроси дополнительную информацию»).

2) *Query then analyze* («Сначала запроси данные – затем анализируй»).

Оба подхода отличаются на концептуальном уровне и имеют свои достоинства и недостатки.

В первом случае (*Analyze then query*) в многомерную БД загружается обобщенная информация, например, месячные, квартальные, годовые итоги по подразделениям. А при необходимости детализации данных пользователю предлагается сформировать запрос к реляционной базе, содержащий требуемую выборку, например, по дням для данного подразделения или по месяцам и сотрудникам выбранного подразделения.

При подходе *Analyze then query* объем данных, загружаемых в многомерную БД, может быть достаточно велик, наполнение должно выполняться по регламенту и может занимать достаточно много времени. Однако все эти недостатки окупаются впоследствии, когда пользователь имеет доступ практически ко всем необходимым данным в любой комбинации. Обращение к исходным данным в реляционной базе данных осуществляется лишь в крайнем случае, когда необходима детальная информация.

Наиболее яркими представителями подхода «*Analyze then query*» являются инструментальные IBM Cognos Transformer и Powerplay.

Во втором случае (*Query then analyze*) пользователь, прежде всего, должен определиться с данными, которые он собирается анализировать и именно их загружать в микрокуб (небольшую многомерную БД).

К достоинствам второго подхода следует отнести «свежесть» информации, которую пользователь получает в виде многомерного отчета – «микрокуба». Микрокуб формируется на основе только что запрошенной информации из актуальной реляционной БД. Работа с микрокубом осуществляется в интерактивном режиме, т.е.

получение срезов информации и ее детализация в рамках микрокуба осуществляется моментально.

Другой положительный момент – проектирование структуры и наполнение микрокуба осуществляется пользователем «на лету», без участия администратора БД.

Однако подход страдает и серьезными недостатками. Пользователь, не видит общей картины и должен заранее определяться с направлением своего исследования. Подход Query then analyze реализует инструментальное средство SAP BusinessObjects.

Выбор и подхода, и инструмента его реализующего, зависит в первую очередь от преследуемой цели: всегда приходится балансировать между экономией бюджета и повышением качества обслуживания конечных пользователей.

Вопрос 4. Рынок OLAP решений.

Анализируя рынок с целью приобретения аналитических средств, специалисты руководствуются определенными требованиями. Одним из наиболее важных критериев является технология хранения данных. По этому признаку продукты можно разделить на следующие категории:

- Клиентский OLAP.
- ROLAP-сервер.
- MOLAP-сервер.

Клиентский OLAP. При данной реализации OLAP-средства вычисления выполняются на клиентской станции.

PivotTable (Microsoft) – это самый распространенный в мире OLAP-продукт, поскольку он входит в состав MS Office. PivotTable или «Сводная таблица» может использоваться в Excel для отображения данных электронной таблицы или внешней БД, а также как клиент для MS Analysis Services. PivotTable может применяться на web-страницах для создания решений в Интернет/Интранет сетях. Достоинством продукта является его простота и удобство, недостатками – ограниченная производительность (без Analysis Services) и относительно небогатая функциональность.

Сводная таблица Excel (PivotTable) представляет собой интерактивную таблицу, применяемую для суммирования или статистического анализа большого количества исходных данных, обычно содержащихся в одном из диапазонов ячеек Excel, либо являющихся результатом запроса к какой-либо БД.

Структура сводной таблицы изображена на рисунке 23.

Поля фильтров	
	Поля столбцов
Поля строк	Поля данных и итогов

Рис. 23. Структура сводной таблицы

При необходимости в Excel можно построить сводную диаграмму, синхронизированную со сводной таблицей.

IBM Cognos Analytics (ранее **IBM Cognos Business Intelligence**) – это единое решение, обеспечивающее полный спектр возможностей Business Intelligence (BI) в одном продукте. Позволяет легко и быстро анализировать любые тенденции развития бизнеса. Этот продукт работает в режиме MOLAP и позволяет получать отчеты, построенные на локальных многомерных кубах. Продукт отличают богатые вычислительные возможности.

SAP BusinessObjects. BusinessObjects – это целостное, интерактивное и открытое решение для бизнес-аналитики в любой организации. BusinessObjects – платформа бизнес-аналитики для управления работой организации, создания отчетов, запросов и анализа, а также управления корпоративной информацией.

Contour Business Intelligence (Contour Components). Contour Business Intelligence (Contour BI) – это программная платформа для регулярного выпуска, актуализации и распространения отчетов, обеспечивающая интерактивное проектирование отчетов по данным разных информационных систем, и предоставляющая широкий набор инструментов для оперативного анализа данных. Программные продукты Contour BI реализуют технологию оперативной аналитической обработки данных (OLAP) и могут быть рекомендованы в качестве базового инструмента для формирования отчетности по произвольным и регулярным запросам.

Программы Contour BI обеспечивают выпуск отчетов для публикации и распространения по сети Интернет/Инtranet или локальной сети. При этом используется уникальная технология микрокубов, позволяющая передавать большие объемы информации в экономичном формате (10–1000 кратное сжатие данных).

Технология микрокубов. Contour микрокуб – это новая концепция многомерной БД, представленной в виде мобильного интерактивного отчета. Contour микрокуб – это файл, который может быть размещен в локальной сети, на персональном компьютере, web-сервере, FTP-сервере, а также передан по e-mail. В файле микрокуба хранятся не только сильно сжатые многомерные данные, но и описание отчетов. В зависимости от состава исходных данных они сжимаются в микрокубе на 90–99.9 %. В результате огромная выборка становится небольшим файлом микрокуба. Микрокуб содержит данные, выгруженные из источника, описание структуры многомерной БД, алгоритмы расчета вычисляемых полей, а также набор отчетов для просмотра и анализа данных (рис. 24).



Рис. 24. Содержание микрокуба Contour

Структура микрокуба включает несколько обязательных элементов: Куб, Срез, Таблица, Диаграмма. В структуре микрокуба может быть настроено произвольное количество срезов, содержащих любое количество интерактивных таблиц и диаграмм (рис. 25).

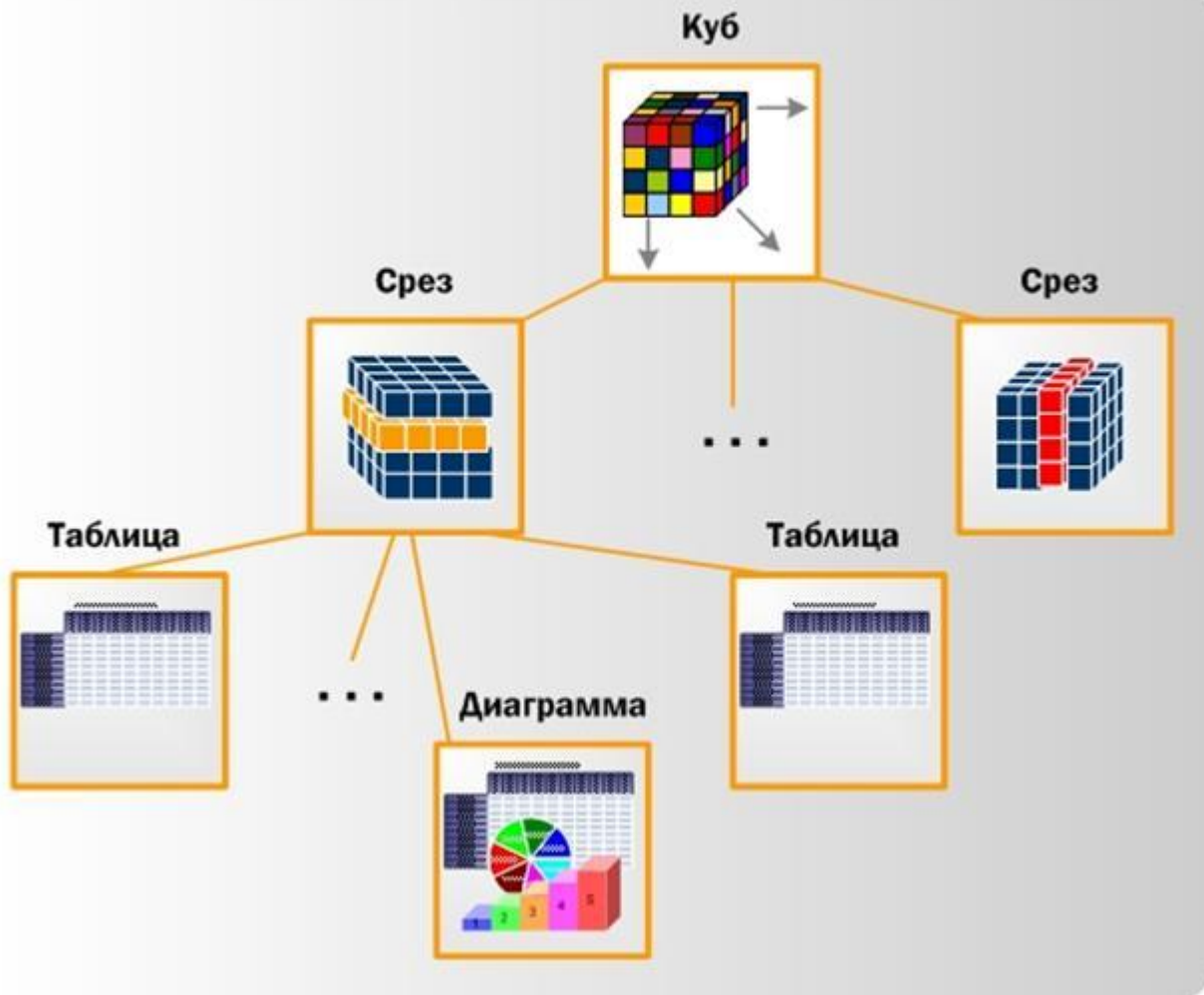


Рис. 25. Структура микрокуба Contour

В состав Contour BI включены следующие программные продукты:

Contour Reporter – это универсальная система для доступа к данным, выпуска интерактивных отчетов, просмотра и анализа данных, а также описания регламентов обновления и рассылки отчетов.

Contour Publisher – программа для автоматического создания и обновления отчетов по заданному регламенту и доставки их до конечных пользователей.

Contour BI Portal. Web – сервер для публикации интерактивных отчетов в интернет/интранет для руководителей, специалистов и широких кругов пользователей.

Contour DW. Декларативное хранилище данных для сбора, хранения данных и предоставления их аналитическим и репортировым программам.

ROLAP-сервер. Продукты этого класса основаны на сервере, который получает запросы от клиента, генерирует SQL-запросы к реляционным БД, выполняет вычисления и передает клиенту данные для отображения.

MicroStrategy (Microstrategy) – набор программных продуктов с широким диапазоном функций, построенный на унифицированной серверной архитектуре. Архитектурным базисом платформы MicroStrategy является MicroStrategy Intelligence Server™. Intelligence Server динамически собирает объекты метаданных для создания многопроходного запроса SQL, оптимизированного для различных реляционных СУБД. Сервер получает данные, производит дополнительные аналитические вычисления недоступные в БД, форматирует отчет и отправляет отчет бизнес пользователям через MicroStrategy Web, MicroStrategy Office, Desktop, или Narrowcast Server.

WebFocus (Information Builder) представляет собой интегрированный набор средств анализа данных масштаба предприятия, позволяющий использовать данные из различных СУБД и предоставлять результаты анализа пользователям, в том числе через Internet. Система состоит из сервера, который выполняет запросы к источникам данных, и тонкого клиента, реализованного как динамический html и add-ins к Excel. Особенность системы – упрощенный пользовательский интерфейс, в котором всегда отображается плоская таблица.

Новая версия WebFocus Business User Edition (BUE) позволяет легко генерировать отчеты, дашборды и визуализации данных без содействия ИТ-отдела или разработчиков BI и делиться своими результатами. WebFocus BUE включает:

- авторский инструмент аналитики InfoAssist+ с самообслуживанием, предназначенный для пользователей без специальных знаний;
- BUE Portal для управления контентом и сборки страниц с аналитикой;
- столбцовое хранилище, обеспечивающее высокую скорость обнаружения данных.

MOLAP-сервер. В случае MOLAP-сервера, исходные и многомерные данные хранятся в многомерной БД или в многомерном локальном кубе. Вычисления и хранение агрегатных данных выполняет сервер. Клиентское приложение получает только результаты запросов к многомерным кубам, которые хранятся на сервере.

Analysis Services (Microsoft). OLAP-сервер компании Microsoft входит в состав Microsoft SQL Server и является одним из лидеров рынка. Система может получать данные из произвольных источников, выполнять сложные вычисления и запросы от клиентских приложений, основанные на языке MDX.

Кубы служб Analysis Services – это многомерные структуры, обеспечивающие высокоскоростной доступ к большим объемам предварительно объединенных данных, и позволяющие конечным пользователям воспринимать интересующие их бизнес-данные со скоростью мысли. В службах Analysis Services хранятся бизнес-данные в формате с высокой степенью оптимизации и сжатия, носящем название многомерного OLAP (MOLAP).

Вопросы для самопроверки:

1. Какие OLAP продукты включает в себя классификация по способу хранения данных?
2. Опишите принцип работы ROLAP-клиентов.
3. Назовите преимущества OLAP-клиентов.
4. В каких архитектурах могут работать OLAP-клиенты?
5. Что представляет собой Микрокуб Контур?
6. Изобразите структуру сводного отчета в Excel.
7. Для чего предназначены Инструментальные OLAP-системы?
8. Что такое OLAP-машина?
9. Какие программные продукты входят в состав Contour BI?
10. Опишите последовательность создания OLAP-приложения.

Литература по теме:

1. Информационные аналитические системы: учеб. / Т.В. Алексеева, Ю.В. Амириди, В.В. Дик. – М.: Московский финансово-промышленный университет «Синергия», 2013.
2. Туманов В.Е. Проектирование хранилищ данных для систем бизнес-аналитики: учеб. пособие / В.Е. Туманов. – М.: Интернет-Университет Информационных Технологий, 2010. – Режим доступа: <http://biblioclub.ru> – Лекция 18.

Интернет-ресурсы:

1. Сайт компании «Contour Components». – Режим доступа: www.contourcomponents.ru
2. Сайт компании «IBM». – Режим доступа: <http://www-01.ibm.com/software/ru/data/cognos/>
3. Сайт компании «SAP». – Режим доступа: www.sap.com/

4. Сайт компании «Strategy Companion». – Режим доступа: <http://strategycompanion.com/>
5. Сайт компании «Microsoft». – Режим доступа: www.microsoft.com/Rus
6. Сайт компании «Information Builders». – Режим доступа: <http://www.informationbuilders.com/products/webfocus>

Тема 6. Интеллектуальный анализ данных (Data mining)

Цель: сформировать представление об интеллектуальном анализе данных на предприятии.

Задачи:

- Изучить задачи интеллектуального анализа данных.
- Ознакомиться со сферами применения интеллектуального анализа данных.
- Освоить технологию проведения интеллектуального анализа данных.
- Изучить программные средства интеллектуального анализа данных.

Вопросы темы:

1. Назначение интеллектуального анализа данных и примеры его применения в бизнесе.
2. Технологические этапы проведения интеллектуального анализа данных.
3. Методы интеллектуального анализа данных.
4. Программные средства интеллектуального анализа данных.
5. Интеграция оперативного и интеллектуального анализа данных.

Основные понятия:

- *Knowledge Discovery in Databases;*
- *Data Mining;*
- *интеллектуальный анализ данных;*
- *ассоциация;*
- *последовательность;*
- *классификация;*
- *кластеризация;*
- *прогнозирование;*
- *нейронные сети;*
- *генетические алгоритмы;*

- *дерево решений;*
- *эволюционное моделирование.*

Вопрос 1. Назначение интеллектуального анализ данных и примеры его применение в бизнесе.

Понятие «интеллектуальный анализ данных» (ИАД) соответствует англоязычному «Knowledge Discovery in Databases» (KDD), что буквально означает «обнаружение знаний в базах данных». Очень часто эти два понятия приравнивают к более популярному термину «Data Mining» (DM). Последнее принято переводить как «добыча (или раскопка) данных». Ряд авторов предлагают рассматривать эти три понятия как синонимы.

Data Mining – это процесс обнаружения в «сырых» данных ранее неизвестных нетривиальных практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности». Более кратко это можно сформулировать как «технологии выявления скрытых взаимосвязей внутри больших баз данных.

В основе методов технологии data mining лежит «концепция шаблонов» (паттернов), отражающих фрагменты многоаспектных взаимоотношений в данных. На основе подвыборок исходных данных выявляются закономерности (шаблоны), которые формулируются в понятной обычному человеку форме. Особенностью data mining является «нетривиальность разыскиваемых шаблонов». Под этим понимается то, что с помощью методов data mining должны выявляться неочевидные, неожиданные (unexpected) регулярности в данных, составляющие так называемые скрытые знания (hidden knowledge). Связано это с идеей о том, что в сырых данных о конкретных фактах из деятельности компании при грамотной раскопке можно обнаружить некоторый глубинный пласт знаний (рис. 26).

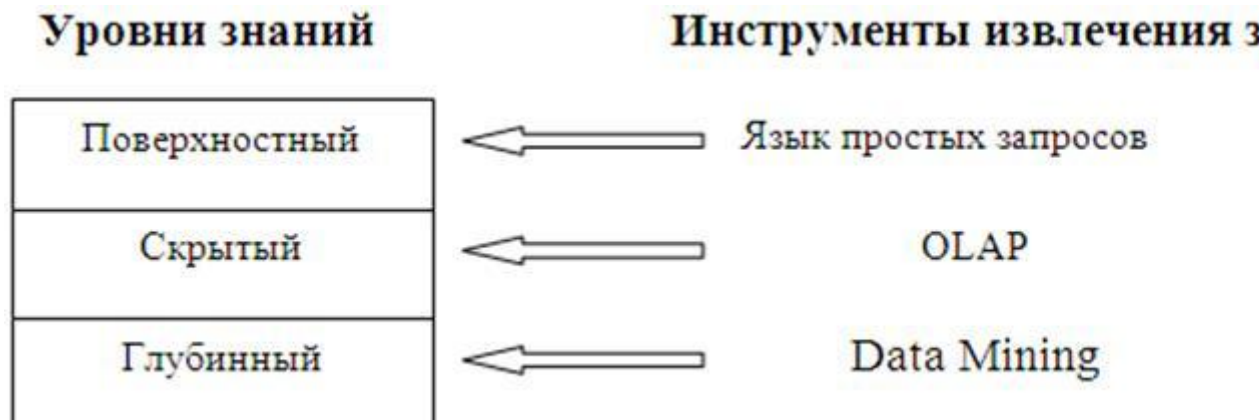


Рис. 26. Уровни знаний, извлекаемых из данных

Выделяют пять стандартных типов закономерностей, которые можно выявить с помощью методов data mining:

1. *Ассоциация* означает, что несколько событий связаны друг с другом, т.е. определяется наличие высокой вероятности связи между событиями.

2. *Последовательность* означает существование цепочки событий, связанных между собой во времени.

3. *Классификация* помогает выявить признаки, характеризующие группу, к которой принадлежит тот или иной объект.

4. *Кластеризация* позволяет в отличие от классификации выделять различные однородные группы данных, когда классификационные группы заранее не известны (они выявляются автоматически в процессе обработки данных).

5. *Прогнозирование* позволяет находить в исторической информации, представленной в виде временных рядов, такие шаблоны, которые отражают динамику поведения целевых показателей.

Информационные системы, реализующие методы data mining, в последние годы становятся все более и более популярны как инструменты для проведения анализа экономической информации. Это становится особенно актуальным в случаях, когда из имеющихся больших объемов ретроспективных данных можно извлечь знания для принятия управленческих решений в условиях неопределенности.

По сути, сфера применения ИАД не имеет ограничений. Методы data mining можно использовать в любой сфере, подразумевающей наличие каких-либо объемных массивов данных. Применение методов data mining имеет смысл, когда в компании накоплено очень большое количество данных. При этом крайне желательно, чтобы эти данные находились в грамотно спроектированном хранилище данных (Data Warehousing).

В таблице 3 представлены наиболее популярные примеры практического применения методов data mining в бизнесе.

Таблица 3.

Примеры практических применений методов data mining в бизнесе

№	Сфера применения	Типичные задачи, решаемые методами data mining
1.	Маркетинг	Рыночная сегментация, идентификация целевых групп, построение профиля клиента.
2.	Банковское дело:	Анализ кредитных рисков, привлечение и удержание клиентов, управление ресурсами.
3.	Страховые компании	Привлечение и удержание клиентов, прогнозирование финансовых показателей.
4.	Розничная торговля	Анализ деятельности торговых точек, построение профиля покупателя, управление ресурсами.
5.	Биржевые трейдеры	Выработка оптимальной торговой стратегии, контроль рисков.

Вопрос 2. Технологические этапы проведения интеллектуального анализа данных.

Процедура выполнения ИАД не зависит от предметной области и может считаться универсальной. Эта процедура выражается в определенной последовательности действий, которые требуется выполнить пользователю, чтобы извлечь из сырых данных значимую информацию (т.е. знания). При этом не имеет значения, какие именно методы будут применяться для обработки данных и получения результата. Процесс ИАД можно разбить на пять этапов (рис. 27).



Рис. 27. Этапы процесса ИАД

1. **Подготовка исходного набора данных.** На этом этапе необходимо собрать исходные данные (в том числе из разных источников) и сформировать из них выборку, на основе которой будет происходить применение какого-либо метода data mining. Объем этой выборки зависит от того метода, который планируется применять для получения нужного результата. Эту выборку из данных также называют «обучающей». Однако следует учесть, что помимо обучающей выборки, т.е. по которой строится модель, потребуется еще и выборка «тестовая», чтобы проверить качество полученной модели. Сбор данных из разных источников – задача непростая, она требует наличия программных средств доступа к различным источникам данных. Эта задача существенно упрощается при наличии в компании централизованного хранилища данных (data warehouse).

2. **Предобработка данных.** Собранные на первом этапе данные являются «сырыми» и нуждаются в специальной обработке прежде, чем к ним будут применены какие-либо методы. Связано это с тем, что в данных могут оказаться пропуски значений, аномальные значения, ошибочно введенные значения и т.д. В некоторых случаях исходные данные следует дополнить какими-либо показателями. При подготовке данных следует учитывать требования конкретного метода, с помощью которого будет проходить их обработка.

3. **Трансформация, нормализация данных.** На этом этапе отобранные ранее данные нужно привести к виду, пригодному для последующего анализа. Содержание этого этапа зависит от применяемого метода анализа. Так, например, метод нейронных сетей требует, чтобы все данные были числовыми и нормализованными. Для выполнения трансформации данных существуют такие алгоритмы, как приведение типов, квантование, приведение к «скользящему окну» и др.

4. **Data Mining.** На этом этапе происходит применение метода анализа к исходным данным и получение «знаний» в виде некоторой модели. Этот этап является ключевым в процедуре ИАД.

5. **Постобработка данных.** На этом этапе выполняется интерпретация результатов, полученных на предыдущем этапе, и разработка механизмов применения результатов в бизнесе (например, разработка программных приложений).

Вопрос 3. Методы интеллектуального анализа данных.

Рассмотрим теперь методы анализа, которые могут применяться на этапе «data mining». Специалисты компании BaseGroup подразделяют методы ИАД на три группы (рис. 28):

- 1) методы математической статистики;
- 2) методы эволюционного моделирования;
- 3) методы машинного обучения.



Рис. 28. Классификация методов ИАД

Методы математической статистики уже долгое время успешно применяются в различных сферах человеческой деятельности, позволяя при этом получать оптимальные решения задач анализа данных в условиях случайных и непредсказуемых воздействий на исследуемый объект. Наиболее популярные из этих методов: регрессионный, дисперсионный и корреляционный анализ.

Под **методами эволюционного моделирования** сегодня подразумевают искусственные нейронные сети и генетические алгоритмы.

Нейронные сети работают по аналогии с нервной системой живых существ: в зависимости от набора исходных сигналов на

входных рецепторах нейроны формируют некоторый результирующий сигнал. Нейронная сеть подвергается обучению, в результате которого происходит запоминание эмпирической зависимости между значениями входных факторов и результирующим. Значения входных факторов играют роль исходных сигналов, а для получения результата в нейронной сети в процессе обучения происходит выработка определенных связей (весов) между отдельными нейронами. Для этого при обучении необходима выборка данных большого объема, содержащая как исходные данные, так и правильные ответы результирующего показателя.

В основе *генетических алгоритмов* лежат принципы генетической эволюции, т.е. если особь определенного вида обладает в высокой мере каким-то свойством приспособленности, то велика вероятность, что у потомков этой особи данное свойство приспособленности закрепится и проявится еще сильнее.

При помощи методов эволюционного моделирования решают такие аналитические задачи, как классификация и кластеризация объектов, прогнозирование, управление динамическими объектами. По сравнению с методами математической статистики нейронные сети способны выполнить указанные задачи даже в том случае, когда формализовать исходную задачу сложно.

Методы машинного обучения основаны на алгоритмах построения «деревьев решений» и зарекомендовали себя при выполнении несложных классифицирующих задачах. При этом результаты выполненной ими классификации представляют собой набор выражений, доступных для понимания человеком. *Дерево решений* представляет собой иерархическую структуру, в которой переход с одного уровня на другой основан на ответе на некоторый вопрос (обычно это логический вопрос, подразумевающий ответ «да» или «нет»).

По сравнению с методами нейронных сетей деревья решений строятся заметно быстрее. В отличие от методов математической статистики, которые применимы только для числовых данных, деревья решений позволяют анализировать как числовые, так и символичные данные.

Вопрос 4. Программные средства интеллектуального анализа данных.

Программные средства ИАД условно можно разделить на два вида:

1. *Универсальные системы ИАД* – программные средства, которые поддерживают практически всю процедуру ИАД и автоматизируют большой набор методов анализа данных.

2. *Специализированные* – средства, созданные для применения в узкой предметной области или поддерживающие только определенный метод или набор методов анализа.

Универсальные программные средства. На рынке универсальных программных средств ИАД представлены следующие *зарубежные разработки*:

- IBM SPSS Modeler.
- SAS Interprise Miner.
- Qlik Analytics Platform.
- Knowledge Studio (Angoss Software).
- Oracle Advanced Analytics.

Конкуренцию им составляют такие *российские разработки*, как:

- PolyAnalyst (Megaputer Intelligence).
- Deductor (BaseGroup Labs).
- Prognoz Platform (Прогноз).

Практически все эти продукты имеют удобный графический интерфейс, широкие возможности в визуализации и манипулировании с данными, а также организуют доступ к различным источникам данных. Однако стоимость таких систем довольно высока и требует от покупателя серьезных работ по интеграции их в существующую корпоративную информационную систему.

Специализированные программные средства. Среди специализированных программных средств ИАД выделяют различные классы:

- предметно-ориентированные аналитические системы;
- статистические пакеты;
- нейросетевые пакеты;
- деревья решений;
- системы рассуждений на основе аналогичных случаев;
- генетические алгоритмы;
- алгоритмы ограниченного перебора;
- системы для визуализации многомерных данных.

Как правило, стоимость таких специализированных систем заметно меньше по сравнению с универсальными.

Предметно-ориентированные аналитические системы решают узкий класс специализированных задач. Они не требуют от пользователя каких-либо серьезных настроек после установки и практически сразу же готовы к использованию. Эти системы, как правило, реализуют несложные статистические методы, но при этом максимально учитывают специфику своей предметной области. Популярным примером таких систем могут служить программные средства «технического анализа» исследования фондовых рынков или средства анализа финансового состояния предприятия.

Статистические пакеты изначально не позиционировались разработчиками как средства для проведения ИАД: это мощные математические системы, предназначенные для статистической обработки данных любой природы. Они включают многочисленные инструменты статистического анализа, имеют развитые графические средства. Однако сейчас почти все серьезные статистические пакеты помимо обычных функций реализуют еще и некоторые методы data mining. Ограничением широкого распространения этих систем является их немалая цена, а также необходимость пользователям обладать глубокими знаниями в области статистики (иначе они не смогут эффективно проводить анализ данных). Примеры статистических систем: SAS/STAT (SAS), SPSS (SAP), Statgraphics Centurion 18 (StatPoint, Inc).

Нейросетевые пакеты – это класс разнообразных программных средств, реализующих алгоритм построения нейронных сетей, т.е. иерархических сетевых структур, в узлах которых находятся так называемые нейроны. Как уже отмечалось при описании метода нейронных сетей, для создания такой сети ее необходимо «обучить» на примере большого объема обучающей выборки из исходных данных. При этом сети демонстрируют очень хорошие результаты при прогнозировании и классифицировании, однако невозможно интерпретировать полученные с ее помощью результаты. Поскольку тренированная нейронная сеть представляет собой «умный черный ящик», работу которого невозможно понять и контролировать. Примеры нейросетевых пакетов:

- BrainMaker (CSS).
- NeuroShell (Ward Systems Group).
- OWL (Hyperlogic).
- Statistica Neural Networks (StatSoft).

Программные средства, реализующие методы деревьев решений (decision trees), представляют собой специализированные продукты для решения только задач классификации. Примерами таких систем являются:

- See5/C5.0 (RuleQuest).
- SIPINA (University of Lyon).
- IDIS (Information Discovery).
- KnowledgeSeeker (Angoss).

Системы рассуждений на основе аналогичных случаев (case based reasoning, CBR) так же, как и предыдущий класс, реализуют только один метод, который еще называют методом «ближайшего соседа» (nearest neighbour). У этого метода есть свои недостатки, однако в ряде задач он способен выдавать неплохие результаты. Примеры таких систем:

- KATE tools (Acknosoft).
- Pattern Recognition Workbench (Unica).

Другими примерами систем, реализующим только одну группу методов анализа, можно назвать системы *генетических алгоритмов* (программный продукт GeneHunter компании Ward Systems Group) и *системы алгоритмов ограниченного перебора* (система WizWhy от компании WizSoft).

Отдельный класс специализированных средств ИАД представляют собой *системы для визуализации многомерных данных*. Несмотря на то, что средствами графического отображения данных оснащены сегодня почти все продукты ИАД, на рынке присутствуют продукты, специализирующиеся только на этой функции, например, DataMiner от разработчика Skyline Communications. Системы визуализации предлагают пользователю дружелюбный пользовательский интерфейс, с помощью которого можно ассоциировать анализируемые показатели с такими параметрами диаграмм, как цвет, форма, ориентация относительно осей, размеры и другие свойства графических элементов изображения.

Вопрос 5. Интеграция оперативного и интеллектуального анализа данных.

Технологии ИАД тесно связаны с технологиями построения хранилищ данных (Data Warehouse), а также с технологиями

оперативной аналитической обработки (OLAP). По своей сути и OLAP, и ИАД являются составными частями процесса поддержки принятия решений.

На сегодняшний день большинство программных OLAP-средств акцентированы в основном на обеспечение доступа к многомерным данным (кубам), в то время как средства ИАД для поиска закономерностей оперируют одномерными массивами данных. Чтобы объединить эти два вида анализа, нужно сфокусировать OLAP-средства не только на способах доступа к данным, но и на выявлении закономерностей в них.

Перспективным направлением интеграции OLAP и ИАД считается объединение этих технологий в рамках корпоративной информационной системы на основе единого хранилища данных.

На рис. 29 изображен подход к подобной интеграции OLAP и ИАД.

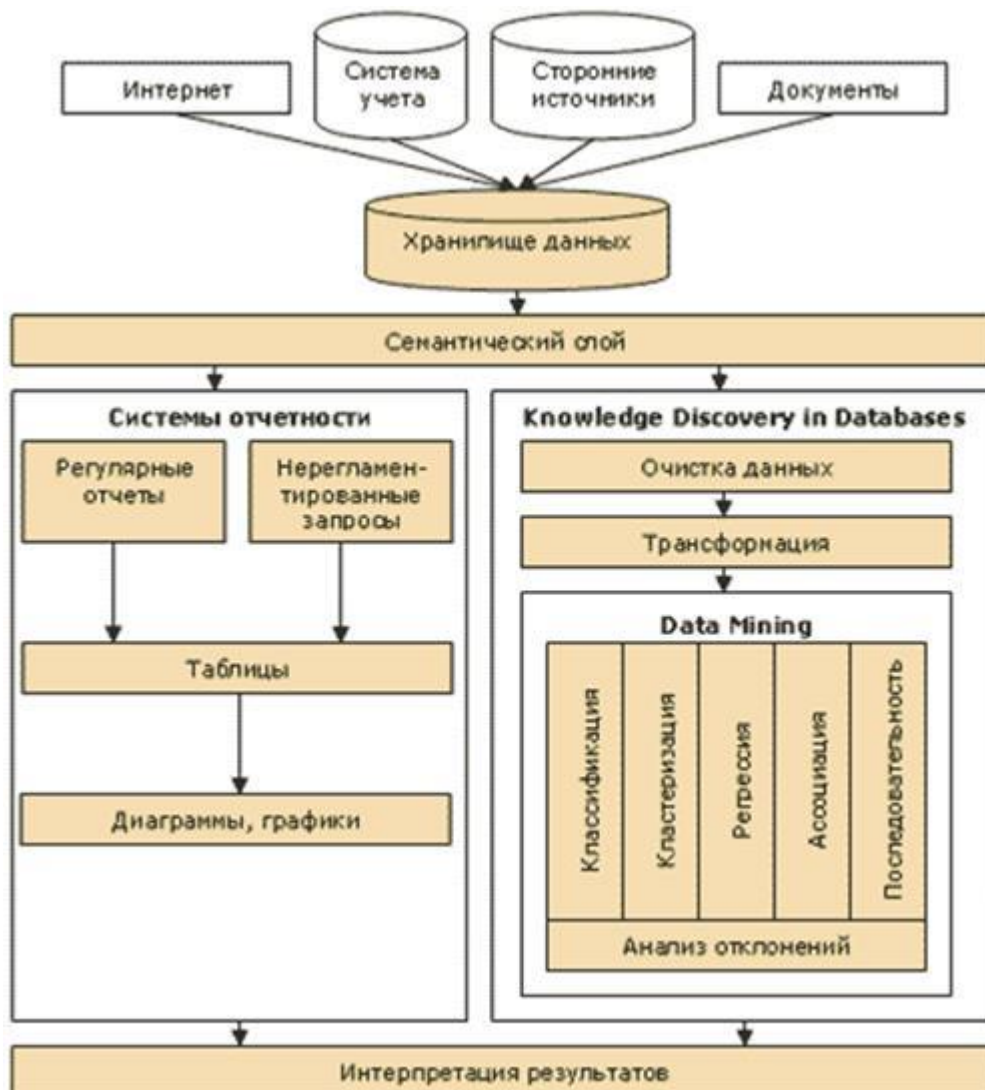


Рис. 29. Схема интеграции OLAP и ИАД в корпоративной информационной системе

В идеальном случае корпоративная система должна содержать в себе одновременно и средства отчетности, и средства ИАД. В этом случае такая информационная система покрывает практически все потребности бизнес-пользователей в проведении анализа данных.

Рассмотрим подробнее элементы схемы интеграции OLAP и ИАД, представленной на рис. 29. *Источником исходным данных* для любого вида анализа выступают различные учетные базы данных организации, документы. В качестве дополнительного источника можно использовать данные, публикуемые в сети Интернет. Для решения стратегических задач управления организацией важно иметь не только внутреннюю информацию, но и внешнюю (например, макроэкономические показатели, демографические данные, сведения о конкурентной среде и т.п.).

Базой для построения аналитической системы является *хранилище данных*, которое само по себе никаких функций анализа не реализует. Исходные данные собираются из различных источников (как внутри, так и вне организации) и загружаются в хранилище.

Семантический слой обеспечивает лицо, принимающее решение (ЛПР) «понятным» ему механизмом доступа к данным, т.к. он трансформирует термины экономической предметной области в процедуры вызова запросов к БД. Таким образом, ЛПР может запрашивать необходимые ему данные на почти естественном для него языке.

Назначение *систем отчетности* – обеспечить ЛПР ответом на вопрос «что происходит». Самый очевидный способ реализации этого – формирование регулярных отчетов для контроля текущей ситуации и выявления отклонений от нормы. Обычно элементы такого вида анализа реализованы в каждой организации. Системы отчетности в таком случае значительно ускоряют процесс получения отчета, но как единственный механизм анализа в организации этот подход недостаточен.

Другим способом использования систем отчетности является *обработка нерегламентированных запросов пользователя*. Периодически перед ЛПР возникает необходимость проверить правильность какой-либо идеи (гипотезы), но для этого ему необходимо иметь фактические данные, подтверждающие или опровергающие его идею. Поскольку такие идеи-гипотезы могут возникать неожиданно, и ЛПР заранее не знает, какого рода фактическая информация ему понадобится, то реализовать все

возможные запросы к данным в системах регулярных отчетов невозможно. Поэтому необходимо обеспечить ЛПР инструментом для оперативного сбора данных и представления их в удобной для восприятия форме. Обычно результаты нерегламентированных запросов оформляются в виде *таблиц* или *графиков* и *диаграмм* (хотя возможны и другие формы визуализации).

Для создания систем отчетности разработчики информационных систем могут применять различные подходы, но самой популярной и доказавшей свою эффективность на сегодняшний день является технология OLAP. Программные OLAP-средства позволяют легко извлекать запрашиваемые пользователем данные и визуализировать их в виде таблиц и диаграмм, но проводить более глубокий анализ (прогнозирование, кластеризация и др.) они, как правило, не могут, так как не позволяют пользователю строить математические и иные модели.

Для проведения более глубокого анализа данных, подразумевающего выбор адекватной модели и ее пошаговое улучшение от более грубой к приемлемо точной, пользователю необходимо обратиться к программным средствам, поддерживающим технологии ИАД.

Результаты обработки данных, как системами отчетности, так и системами ИАД всего лишь обеспечивают ЛПР информацией для принятия решений: в первом случае это данные в форме таблиц и диаграмм, а во втором – в виде моделей и правил.

Информационные технологии ИАД позволяют современной организации осуществлять так называемое «тиражирование знаний», когда один сотрудник на основе анализа получает определенные выводы и формулирует их в виде набора правил или математической модели. Эта модель (или правила) могут быть оформлены организацией как рабочие инструкции для других сотрудников, но могут быть заложены в бизнес-логику программных средств. В любом случае, остальные сотрудники компании могут уже не проводить самостоятельный анализ, а воспользоваться ранее полученными выводами. Это сокращает время на выполнение различных бизнес-процессов, а также позволяет организациям снизить требования к уровню профессиональной компетенции отдельных сотрудников без риска потери качества выполнения бизнес-процессов.

Вопросы для самопроверки:

1. Что такое «интеллектуальный анализ данных» (ИАД)? Как этот термин соотносится с понятиями Knowledge Discovery in Databases (KDD) и Data Mining (DM)?
2. В чем отличие ИАД от оперативной аналитической обработки данных (OLAP)?
3. Какие стандартные типы закономерностей (взаимосвязей) можно выявить с помощью методов ИАД?
4. В каких сферах человеческой деятельности можно применять ИАД?
5. Опишите этапы технологического процесса ИАД.
6. Какие группы методов применяются в ИАД?
7. Охарактеризуйте схему интеграции программных средств ИАД и OLAP в рамках одной корпоративной информационной системы
8. Чем отличается классификация от кластеризации?
9. Опишите процесс предобработки данных.
10. Охарактеризуйте методы эволюционного моделирования.

Литература по теме:

1. Информационные аналитические системы: учеб. / Т.В. Алексеева, Ю.В. Амириди, В.В. Дик. – М.: Московский финансово-промышленный ун-т «Синергия», 2013.
2. Белов В.С. Информационно-аналитические системы: основы проектирования и применения: учеб.-прак. пособие / В.С. Белов. – 2-е изд., перераб. и доп. – М.: Евразийский открытый институт, 2010. – Режим доступа: <http://biblioclub.ru> – Тема 4.
3. Нестеров С.А. Интеллектуальный анализ данных средствами MS SQLServer. – 2-е изд., испр. – М.: Национальный Открытый Университет «ИНТУИТ», 2016. – Режим доступа: <http://biblioclub.ru> – Лекции 1–3.
4. Дюк В. Data mining – интеллектуальный анализ данных. – [электронный ресурс] – Режим доступа: <http://www.interface.ru/fset.asp?Url=/oracle/dmiad.htm>
5. StatSoft. Методы добычи данных: глава из электронного учебника по статистике компании StatSoft. – [электронный ресурс] – Режим доступа: http://www.olap.ru/basic/data_mining.asp
6. Knowledge Discovery in Databases – обнаружение знаний в базах данных. – [электронный ресурс] – Режим доступа: <http://www.basegroup.ru/library/methodology/kdd/>

Тема 7. Информационные системы управления эффективностью бизнеса (BPM) и системы Бизнес-интеллекта (BI)

Цель: сформировать представление об информационных системах управления эффективностью бизнеса.

Задачи:

- Познакомиться с информационными системами управления эффективностью бизнеса.
- Рассмотреть связь ВРМ и ВІ.
- Изучить этапы цикла управления эффективностью бизнеса.
- Рассмотреть КРІ в контексте ВРМ.

Вопросы темы:

1. Информационные системы бизнес интеллекта и управления эффективностью бизнеса.
2. Определение ВРМ и ВІ.
3. Технологические этапы цикла управления эффективностью бизнеса (ВРМ).
4. КРІ в контексте ВРМ.

Основные понятия:

- *Business Intelligence (BI).*
- *Enterprise Resource Planning (ERP).*
- *Business Performance Management (BPM).*
- *Key Performance Indicators (KPI).*
- *Balanced ScoreCard (BSC).*
- *Система сбалансированных показателей (ССП).*
- *Коэффициентами приоритетности целей (КПЦ).*

Вопрос 1. Информационные системы бизнес интеллекта и управления эффективностью бизнеса.

История развития информационной системы (ИС) наиболее ярко отражается в их классификации по двум направлениям. В основе одних ИС лежит поддержка производственного цикла. Это системы:

- MRP (*англ. Material Requirements Planning*).
- MRPII (*англ. Manufacturing Resource Planning*).
- ERP (*англ. Enterprise Resource Planning*).
- APS (*англ. Advanced Planning/Scheduling*).
- SCM (*англ. Supply Chain Management*).
- CRM (*англ. Customer Relationship Management*).
- PLM (*англ. Product Lifecycle Management*).

- E-commerce.

Развитие ИС этого направления вплоть до APS систем заключалось в постепенном поглощении каждым следующим классом ИС предыдущего класса. Но APS системы так и не стали ожидаемым классом ИС, передав свой функционал BPM системам. А развитие этой ветви ИС продолжилось за счет таких систем, как SCM, CRM, PLM. Каждая из них стала чрезвычайно популярна и производители ERP систем стали включать в функционал своих продуктов эти новые инструменты. Так они появились на рынке как самостоятельный продукт и как часть ERP систем одновременно. Заговорили о новом классе ИС: ERP2.

ERP стали неотъемлемой частью системы управления предприятием. Поэтому они преподносятся производителем как всеобъемлющая корпоративная информационная система (КИС). Однако, в действительности, это не так. Несмотря на мощный объем функционала ERP-система фактически используется как отчетная система. Кроме того, ERP не обеспечивает полноценной поддержки принятия решений.

В основе другого класса ИС лежит *поддержка цикла принятия решения*: распознавание ситуации, выработка множества альтернативных решений и выбор решения по критерию для исполнения. История развития этого класса систем может быть представлена следующими названиями:

- TPS (*англ.* transaction processing systems – системы обработки транзакций).
- MIS (*англ.* management information systems – информационная система управления).
- DSS (*англ.* decision support systems – системы поддержки принятия решений).
- IPSS (*англ.* integrated performance support systems – интегрированные системы поддержки производительности).
- EIS (*англ.* executive information systems – исполнительные информационные системы).
- BI (*англ.* business intelligence – бизнес-аналитика/бизнес-интеллекта).

TPS, MIS, DSS развиваются по пути поглощения предыдущих классов, вплоть до IPSS. Далее появляется идея EIS, которая так и не состоялась как распространенный самостоятельный класс. Ожидаемой трансформации DSS и IPSS в EIS не произошло. EIS осталась концепцией, вместо которой возникла идея бизнес-

интеллекта. Эффективность Business Intelligence зависит от совершенства входящих в него инструментов. Ранние версии BI представляли собой набор отдельных, не связанных между собой инструментов бизнес-анализа, которые аккумулировали:

- СППР (DSS).
- систему запросов и отчётов (Q&R).
- OLAP систему.
- систему прогнозирования.
- инструменты извлечения данных (Data Mining).

В последние годы на первый план вышли аналитические платформы: наборы интегрированных между собой инструментов (компонентов) бизнес-анализа.

Системы, поддерживающие принятие решения, также постоянно расширяют свой функционал и методологию. Поэтому бизнес-интеллект просуществовал как лидирующее направление недолго. Функционал BI, стал быстро расширяться. Методология также потребовала изменений. Возник новый класс ИС – Business Performance Management (BPM). Этот термин стала применять аналитическая компания International Data Center (IDC). И, если при создании и внедрении сложных корпоративных систем OLTP класса предприятия возникает потребность в реинжиниринге бизнес-процессов, то при создании и внедрении Business Performance Management необходимы моделирование бизнес процессов и бизнес-инжиниринг, которые позволяют связать стратегический и оперативный уровни управления. Здесь мы часто сталкиваемся с омонимией: управление бизнес процессами – это Business Process Management.

Вопрос 2. Определение BPM и BI.

Business Performance Management (BPM) как класс возник в 2000 г., а в 2003 г. этот термин пришел в нашу страну. Этот термин, в принципе, устоялся, но в качестве синонимов до сих пор используются различные синонимы:

- Corporate Performance Management (CPM).
- Enterprise Performance Management (EPM).
- Strategic Enterprise Management (SEM).
- Performance Management (в узком смысле).

Business Performance Management (BPM) переводится как управление эффективностью бизнеса.

Понятие BPM-система может употребляться в двух значениях:

- 1) концепция управления, т.е. определенный подход к принятию управленческих решений и их практической реализации;
- 2) информационная система, т.е. комплекс программных средств, обеспечивающих практическую реализацию концепции BPM.

Управление эффективностью деятельности – это набор управленческих процессов (планирования, организации выполнения, контроля и анализа), которые позволяют бизнесу определить стратегические цели и затем оценивать и управлять деятельностью по достижению поставленных целей при оптимальном использовании имеющихся ресурсов. Это система управления, построенная на принципах управления стоимостью бизнеса.

Увлечение BPM привело даже к созданию в марте 2004 г. группы по разработке стандартов BPM (BPM Standards Group) несколькими ведущими в то время производителями: IBM, Hyperion Solutions Corp., IDC, Meta Group и др. Ее целью была разработка стандартов, которые могли бы использовать все производители ПО, поставщики услуг, консультанты и конечные пользователи.

Чем же отличается BI от BPM кроме функционала? И каковы причины консолидации BI и BPM? Отвечая на поставленные вопросы необходимо отметить, что в BPM, укрупнено, с точки зрения ИТ, входит BI, функционал BPM и ERP как один из источников информации. С позиций технической реализации BI остается базисом, на котором реализуется методологическая составляющая BPM.

Теперь поставщики BI стремятся представить свои продукты как ИС класса BPM, рассматривая BPM как следующую ступень в развитии BI.

Проблема состоит в том, что возможности BI ограничены. BI предлагает средства аналитики, но не может помочь изменить бизнес-процессы, когда это необходимо, или объяснить, как скорректировать планы и использовать прогнозы. Для этого предназначены средства моделирования ресурсов и другие инструменты в составе BPM. Поэтому BI, в этом смысле, уже пройденный этап. Можно сказать, что BPM – это следующая ступень в развитии Business Intelligence, как, впрочем, и ERP.

Приложения BPM – это готовые прикладные решения. Они поддерживают ключевые управленческие процессы:

- стратегическое и оперативное планирование;
- мониторинг доходов и расходов и подготовку финансовой и регламентной отчетности;

- бюджетное планирование и прогнозирование;
- управленческий учет и подготовку отчетности;
- функционально-стоимостной анализ (ФСА), опираясь на технологии BI;
- хранилища данных (DataWarehouse);
- инструменты репортинга (Query and Reporting Tools);
- OLAP и др.

Очень важно, что под BPM понимается не только ПО, но и ряд методик управленческого характера и математических инструментов.

К инструментам предметной области, входящим в BPM относятся:

- модели Key Performance Indicators (KPI) или ключевые показатели результативности работы предприятия;
- Balanced ScoreCard (BSC) или система сбалансированных показателей;
- бюджетирование;
- модели корпоративной мотивации;
- модели мониторинга и контроля исполнения;
- элементы управленческого учета;
- консолидация финансовых и нефинансовых данных.

А к информационным технологиям BPM относится его специфический функционал:

- технологии BI (Весь функционал DSS, OLAP технологии и технологии DM);
- ERP система;
- ИТ приложения бюджетирования;
- технологии моделирования и управления бизнес процессами.

ИС позволяют участникам процесса управления реализовать методики и бизнес-процессы управления на практике. Такая ИС как BPM должна охватывать всю компанию и предусматривать совместный доступ к данным, чтобы все участники процесса управления имели возможность обмениваться необходимой информацией. Важно отметить, что без взаимодействия информационной BPM-системы с транзакционными системами (ERP, CRM, PLM, SCM) и другими унаследованными системами (либо через хранилище данных через технологию ETL, либо напрямую) невозможна интеграция стратегического управления с оперативным. Можно говорить о разных источниках информации, но наиболее

результативный вариант ВРМ – это использование единого хранилища данных.

Платформы для построения хранилищ поставляются с уже готовыми структурами данных, механизмами создания витрин и вычисления показателей. Концепция отраслевых моделей данных предполагает применение в качестве основы для построения ХД и ВРМ-систем логической модели, которая описывает все области деятельности организации. *Отраслевая логическая модель предприятия* – это визуальное представление объектов предметной области, их атрибутов и взаимосвязей. Она является прототипом базы универсального хранилища. На основе логической модели разрабатывается физическая модель структур хранения первичных данных и расчетных показателей (витрин данных). Модель является каркасом для создания индивидуального ВРМ-решения предприятия, который можно наполнять данными, получая на каждом этапе полезный для потребителя результат.

Информация из ХД поступает на обработку в слой ВІ и далее используется функционалом ВРМ. ВРМ приложения включают в себя инструменты:

- прогнозирования;
- моделирования;
- планирования;
- бюджетирования;
- формирования консолидированной отчетности.

Над ВРМ приложениями расположены инструменты системы сбалансированных показателей и КРІ, которые обеспечивают стратегическое планирование и его связь с оперативным управлением.

ИС класса ВРМ в силу своей сложности имеют большую совокупную стоимость владения, чем ERP или ВІ отдельно взятые. Теоретически, цена информационной системы – интегрированный показатель ее качества, отражающий общую сумму затрат для компании-потребителя на различных стадиях жизненного цикла ИС. При выборе альтернатив ИС необходимо оценить совокупную стоимость владения.

Под *совокупной стоимостью владения* понимается сумма прямых и косвенных затрат, которые несет владелец системы за период жизненного цикла системы, которые могут включать затраты на:

- приобретение ИС;

- оборудование (сервера и клиентские места, периферийное и сетевое оборудование);
- программное обеспечение;
- обновление.

Затраты на эксплуатацию включают:

- управление задачами (ИС и сетью);
- поддержку работоспособности системы (персонал, справочная служба, обучение, закупки, подготовка контрактов на поддержку системы);
- разработку инфраструктуры и бизнес приложений.

Последнее время активно развиваются *облачные услуги*, появились услуги SaaS аутсорсинга ВРМ. Эта модель аутсорсинга расшифровывается как *Soft as a Service*, т.е. программное обеспечение как услуга. Установка ИС при ее использовании в организации не требуется, т.е. меняется трактовка жизненного цикла ИС и, следовательно, совокупная стоимость владения, ее структура.

К новым свойствам ВРМ можно отнести: мобильность, интероперабельность, кастомизацию, простоту и удобство.

До сих пор ВРМ остается модной ИС. Поскольку приобретение и использование серьезного программного обеспечения или продуктов престижного вендора является одним из средств создания имиджа в конкурентном окружении. ВРМ влияет на менеджмент, обеспечивая обратную связь в управлении и предъявляя новые требования к процессу.

Среди общих ИТ тенденций отмечаются рост объемов мобильной информации и рост информации вообще. Учитывая то, что ВІ осуществляет «разведку бизнес данных» в том числе и в банках мобильной информации, может появиться термин *мобильный ВІ*.

Еще одна тенденция – это *преднастроенные* и *предустановленные* ВРМ системы. Услуга заключается в конфигурировании железа и софта под конкретику внедрения ВРМ. Встроенная отраслевая бизнес логика дает некоторые конкурентные преимущества и предприятию потребителю и, то же время, поставщику. Можно настроить отчетность, интерфейсы, аналитику, ключевые показатели результативности работы предприятия и др. Ценность ВРМ растет с каждым отраслевым внедрением, да и просто с каждым внедрением.

Для использования ВРМ есть три преграды:

- 1) стоимость;

- 2) неготовность предприятия: его организационная неготовность;
- 3) отсутствие необходимых навыков у пользователей.

Одна из причин отсутствия навыков связана со степенью свободы пользователей, с возможностями самостоятельного выставления целей. Пользователи должны иметь соответствующие компетенции, чтобы осмысленно формулировать ВІ запрос. И, конечно, последние тесно связаны со степенью организационной зрелости предприятия.

Таким образом, факторами успеха внедрения и последующего использования ИС являются:

- зрелость и внутренняя культура предприятия;
- понимание эффективности ВРМ инструментов, качество консалтинга, наличие стратегии применения ВРМ;
- заинтересованность потребителя, готовность к изменениям, поэтапное развитие проекта.

Вопрос 3. Технологические этапы цикла управления эффективностью бизнеса (ВРМ).

В основе концепции ВРМ лежат цикл управления эффективностью бизнеса компании и процесс принятия решения. В цикле управления эффективностью бизнеса компании выделяются несколько этапов.

Этап 1. Разработка стратегии. Цель этого этапа – выделить стратегически важные показатели развития бизнеса и запланировать целевые количественные значения их метрик – *ключевых показателей эффективности* (Key Performance Indicators, сокр. КРІ). Фактически КРІ служат измерителями эффективности достижения стратегических целей компании во времени. Стратегическое планирование опирается на одну из методологий ВРМ, известную как *Система сбалансированных показателей* (Balanced Scorecard, сокр. BSC).

Этап 2. Планирование. На втором этапе разрабатываются оперативные планы для поддержки выработанной бизнес-стратегии. Ориентирами для их создания являются запланированные значения КРІ. Основным инструментом оперативного планирования является бюджет.

Этап 3. Мониторинг и анализ. Третий этап в цикле управления эффективностью – контроль исполнения оперативных бюджетов и КРІ. Источником информации для получения фактических значений является *управленческий учет*. «План-факт»

сравнение намеченных и достигнутых показателей бюджетов и КРІ позволяет обнаружить отклонения и выяснить причины их возникновения.

Этап 4. Регулирование. На этом этапе стратегия и планы приводятся в соответствие с реальными условиями деятельности и возможностями организации. По сути, этот этап не является заключительным, а выполняется параллельно остальным.

Регулирование включает:

- сценарное моделирование;
- прогнозирование;
- «скользящее» корректирование оперативных планов;
- КРІ.

В результате обеспечивается непрерывность управленческого цикла и его адаптивность к изменениям внутренней среды компании и ее внешнего окружения.

Модель ВРМ охватывает комплекс технологий управления компанией и обеспечивает взаимосвязь управленческих процессов на стратегическом и тактическом уровне.

Цикл управления эффективностью – это цикл управления с обратной связью, подразумевающий «движение» информации «сверху-вниз» и обратно, т.е. классический кибернетический контур из прямой и обратной связей. *Прямая связь* – это определение стратегии, которую необходимо декомпозировать до операционного уровня с помощью бюджетирования. *Обратная связь* – публикация отчетности, отражающей результаты, анализ и переход на внешний цикл стратегии.

Выделяют три функционально различных уровня управления эффективностью для прямой и обратной связи.

Уровни прямой связи.

Уровень 1. Моделирование стратегии и коммуникация:

- Определение целей деятельности (карта стратегии) и ключевых показателей эффективности функционирования организации (финансовых и нефинансовых показателей).
- Моделирование бизнеса (карта процессов), выявление факторов прибыльности, имеющихся ресурсов и ограничений.
- Декомпозиция задач верхнего уровня в целевые уровни нижестоящих звеньев.
- Целевые установки: стратегические цели, выраженные в конкретных числах.

Уровень 2. Процессно-ориентированное планирование:

- Определение способов достижения целей: формирование сценариев деятельности, расчёт объёмов необходимых ресурсов (материальных, кадровых, финансовых), расчёт плановой себестоимости и накладных издержек.

- Выравнивание операционного и финансового баланса ресурсов.

- Учёт использования ресурсов (нехватка/избыток), выявление «узких мест», не позволяющих наращивать обороты.

- Подключение аналитических модулей для решения задач формирования прогнозов, оптимизационных задач.

Уровень 3. Бюджетирование:

- Планирование конкретных шагов по их достижению: документооборот бюджетных форм, ведение классификаторов аналитики, описание финансовой структуры и принципов взаимодействия, исторические тренды, анализ отклонений.

- Организационные функции (процесс согласования бюджетов) и функции формирования свода бюджетов по отдельным подразделениям, бизнес-единицам, сегментам.

- Версионность бюджетов, сценарный анализ.

Уровни обратной связи.

Уровень 1. Консолидация, отчётность и анализ:

- Сбор фактических данных, формирование регулярной отчётности для внешних и внутренних пользователей, трансформация отчётности в различные стандарты.

- Мониторинг: отслеживание исполнения бюджета, фиксирование отклонений и выяснение их причин.

- Детальный анализ финансовых результатов и состояния баланса, сегментная отчётность, отчётность по центрам ответственности.

Уровень 2. Функционально-стоимостной анализ:

- Разнесение затрат с помощью функционально-стоимостного анализа по центрам ответственности, перенос затрат на основные и обеспечивающие процессы, формирование затрат по продуктам, категориям клиентов, каналам продаж.

- Анализ прибыльности в разрезе продуктов и услуг, филиалов, центров ответственности.

- Анализ трансфертных операций, обслуживающих затрат и взаиморасчётов.

- Выявление неэффективных процессов, сравнение затратных показателей с историческими и эталонными.

Уровень 3. Карты балльных оценок и обратная связь:

- Представление фактических результатов деятельности в сжатом, агрегированном виде, необходимом для сравнения запланированных целевых значений ключевых показателей эффективности с реально достигнутыми.
- Расчёт ключевых показателей эффективности, нормализация значений, расчёт сводных показателей.

Вопрос 4. KPI в контексте BPM.

KPI от *англ.* Key Performance Indicators (ключевые показатели эффективности или результативности) являются метриками бизнеса, рассчитанными на достижение планируемых результатов. С помощью KPI облегчается достижение необходимого результата за счет реализации цикла планирование-учет-анализ (план-факт анализ). Но на практике разработка результативных показателей измерения результативности бизнеса – достаточно непростое дело. Этап разработки KPI может оказаться длительным и продолжаться от нескольких недель, до нескольких месяцев.

Один из способов определения показателей результативности бизнеса – это «стратегические карты» в системе сбалансированных показателей (ССП) (или Balanced ScoreCard (BSC)).

ССП позволяет руководству определять показатели бизнеса и строить причинно-следственные отношения для любой организационно-штатной единицы компании. Следует помнить, что KPI декомпозируются с верхнего уровня управления до нижнего. Такая декомпозиция может быть реализована либо в BI с помощью OLAP-инструментов, реализующих не только многомерное, но и многоуровневое представление данных, либо в ХД. В нем должны присутствовать и стратегические значения KPI, и KPI других уровней декомпозиции. Разработка KPI заключается в определении их состава (состава метрик) и численных значений, которые обеспечивали бы достижение выбранных целей.

Сегодня концепция системы сбалансированных показателей является одной из наиболее востребованных. Это наиболее популярное инструментальное средство, позволяющее добиться сопряжения и синхронизации стратегических целей с тактическими и оперативными, а также осуществлять управление на основе согласованных измерителей.

В рамках СПП говорят о четырех перспективах:

- 1) финансы;
- 2) клиенты;
- 3) бизнес-процессы;
- 4) обучение и развитие.

Обычная технология реализации концепции предполагает следующие действия:

- декомпозицию установок от миссии к дереву целей (на основе измерителей);
- установление ответственных за достижение целей на всех уровнях декомпозиции;
- распределение стратегических целей между «перспективами»;
- определение причинно-следственных связей;
- определение измерителей для разных уровней организационной структуры управления (ОСУ);
- расстановка приоритетов лицом, принимающим решение, что позволяет учесть более и менее перспективные направления действий;
- агрегация: связь стратегии с процессами.

Что касается ИАС, то технология ССП влияет на ее проектирование и использование. Поскольку при разработке ССП сначала определяются бизнес-аспекты в дальнейшем анализируемых данных, появляется возможность:

- проектировать основной инструмент ВРМ-ИАС сверху-вниз параллельно с разработкой дерева целей;
- обрабатывать не только ретроспективные метрики, но и планируемые.

Связывая ретроспективные измерения и рассчитанные на их основе прогнозные, а также плановые, созданные в ССП, появляется возможность прогнозировать изменения в бизнесе.

Проектирование ИАС на основе BSC начинается с проектирования карты стратегии: ее графического описания в виде набора причинно-следственных связей. Для каждой перспективы (финансы, клиенты, бизнес процессы, обучение и развитие) должны быть определены стратегические цели и построено дерево целей.

Декомпозиция главной цели в дерево целей – это процесс неформальный, творческий, требующий определенных знаний и опыта. Для коррекции ошибок, допущенных пользователем,

необходимо введение коэффициентов коррекции проведенной декомпозиции, задаваемых пользователем. Коррекция может выполняться также и автоматически. Эти коэффициенты, называемые *коэффициентами приоритетности целей* (КПЦ), указывают предпочтение пользователем одного пути достижения цели (подцели) другим. КПЦ – это инструмент управления выбором направления в достижении цели (сумма КПЦ на одном уровне дерева, касающихся одного вышележащего узла, должна быть равна единице).

Образно процесс декомпозиции можно представить следующим образом. Сначала описываются основная цель и ближайшие подчиненные цели (рис. 30). Дальнейшее уточнение целей заключается в указании ожидаемого состояния корневого показателя (R), а также задании коэффициентов приоритетности целей (α и β). Очевидными условиями являются: $\alpha + \beta = 1$. Далее аналогичным образом идет декомпозиция каждой цели. Каждый из узлов дерева снабжен знаком «плюс» или «минус», указывающим на желаемое направление изменения (увеличение или уменьшение) соответствующего показателя в процессе достижения главной цели.

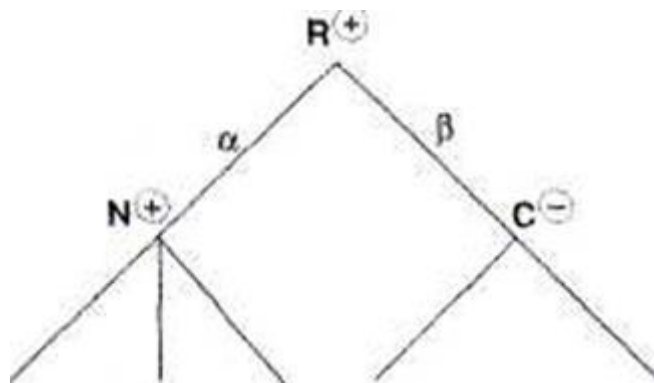


Рис. 30. Дерево целей

Для реального использования показателей в процессе управления необходимо превратить дерево показателей в дерево целей, а дерево целей – в дерево решений.

Вопросы для самопроверки:

1. К какому классу систем относятся ERP системы?
2. Дайте определение BPM системам, чем они отличаются от BI?
3. Какие ключевые управленческие процессы поддерживают системы BPM?

4. Перечислите инструменты предметной области, входящие в ВРМ.
5. Какие инструменты включают в себя ВРМ приложения?
6. Какие тенденции влияют на развитие и применение ВРМ?
7. Какие процессы входят в цикл управления в ВРМ?
8. Охарактеризуйте уровни прямой и обратной связи в процессе управления эффективностью.
9. Что такое КРІ и для чего они используются в организации?
10. Какие перспективы включает в себя сбалансированная система показателей, охарактеризуйте каждую из перспектив.
11. Для чего применяются коэффициенты приоритетности целей? Кто их устанавливает?

Литература по теме:

1. Внедрение сбалансированной системы показателей: пер. с нем. В. Толкача, С. Данишевича, М. Гавриша. – М.: Альпина Бизнес Букс, 2008.
2. Кандалинцев В.Г. Инновационный бизнес: применение сбалансированной системы показателей. - Издательство: Дело, 2010.

Интернет ресурсы:

1. Сайт компании «Агентство маркетинговых коммуникаций». – Режим доступа: <http://events.cnews.ru>
2. Сайт компании «АйТи». – Режим доступа: www.it.ru

Тема 8. Информационно-методический комплекс управления эффективностью бизнеса (ВРМ) на предприятии

Цели: сформировать представление об информационно-методическом комплексе управления эффективностью бизнеса и технологии внедрения ВРМ.

Задачи:

- Изучить состав информационно-методического комплекса ВРМ.
- Познакомиться с технологией внедрения ВРМ систем на основе информационно-методического комплекса.
- Ознакомиться с архитектурой ВРМ систем.
- Изучить предложения ведущих вендоров ВРМ систем на рынке ПО.

Вопросы темы:

1. Архитектура, функциональные возможности и компоненты информационных систем управления эффективностью бизнеса.
2. Технология внедрения ВРМ.
3. Характеристика рынка инструментальных средств бизнес интеллекта, управления эффективностью бизнеса.

Основные понятия:

- методологическая модель;
- информационная модель;
- информационно-методический комплекс;
- информационный комплекс;
- технологическая платформа;
- *Corporate performance management (CPM)*.

Вопрос 1. Архитектура, функциональные возможности и компоненты информационных систем управления эффективностью бизнеса.

ВРМ-система представляет собой совокупность четырех основных элементов:

- 1) участники процесса управления;
- 2) интерактивные бизнес-процессы управления;
- 3) методы управления;
- 4) информационные системы и технологии.

Все эти элементы взаимосвязаны. *Система управления* – это менеджеры, участвующие в управлении предприятием и реализующие бизнес-процессы управления или, в какой-то степени, влияющие на эти процессы в соответствии с установленными правилами и разработанными методиками при помощи тех или иных ИС.

Архитектура ВРМ должна поддерживать все процессы управления с помощью соответствующих технологий. Из идеи непрерывного цикла управления логически вытекает функциональность ИС класса ВРМ. Следовательно, архитектура ВРМ системы принимает следующий вид (рис. 31):



Рис. 31. Типовая архитектура BPM-системы

Архитектура включает платформу «бизнес-интеллекта» (Business Intelligence, BI), а также BPM-приложения, непосредственно осуществляющая различные функции управления. Среди приложений следует выделить подсистему, реализующую методы целевого управления организацией (Balanced Scorecard и др.), именно эта подсистема обеспечивает нацеленность системы управления на четыре ключевые области, которые определяют общее состояние компании (финансы, организация бизнеса, продукты, клиенты). Важно отметить интеграционные возможности: BPM-система обладает способностью интеграции с другими источниками корпоративной информации, включая хранилища данных, ERP- и CRM-системы, системы собственной разработки.

Таким образом, BPM-система состоит из двух компонентов:

- 1) методология управления компанией;
- 2) информационные технологии.

Это не противоречит классическому определению BPM, в котором BPM рассматривается как единство методологической и информационной составляющих.

Следовательно, качественное решение для управления эффективностью бизнеса должно представлять собой платформу,

которая базируется на двух моделях: методологической и информационной.

Методологическая модель обобщает и систематизирует отраслевые методики реализации ключевых ВРМ-процессов:

- планирование и бюджетирование;
- управление доходностью бизнеса;
- управление рисками;
- отчетность;
- финансовая консолидация.

Методология ВРМ охватывает весь комплекс технологий управления предприятием и обеспечивает единство управленческих процессов на стратегическом и тактическом уровне.

Методология базируется на идеи непрерывного управленческого цикла:

- разработка стратегии;
- планирование;
- мониторинг и контроль;
- анализ и регулирование.

В основе методологической модели лежит единый отраслевой словарь данных (модель данных), с помощью которого описываются все бизнес-объекты. Ведущие мировые поставщики ВРМ-решений (IBM, Sybase, Oracle) в составе своих ВРМ-платформ предлагают отраслевые логические модели данных для финансовой, энергетической, торговой и др. отраслей.

Эти модели содержат описание взаимосвязанных бизнес-объектов конкретной отрасли, необходимых для автоматизации комплекса управленческих задач. Наличие в составе ВРМ-решения методологической модели подтверждает, что данное решение относится к классу ВРМ.

Информационная модель представляет собой совокупность информационных технологий и программных продуктов, предназначенных для автоматизации методической модели.

В России ВРМ-технологии сегодня успешно развиваются, наиболее востребованы они в банковской отрасли. Среди поставщиков ВРМ-платформ хорошо известна российская компания Intersoft Lab, основной продукт которой ВРМ-система «Контур» поддерживается информационно-методическим комплексом. В состав информационно-методического комплекса входит методическая модель и информационный комплекс (рис. 32).

Методическая модель используется в проектах автоматизации для быстрого создания индивидуальной модели BPM-решения банка.

Информационный комплекс – это совокупность информационных технологий и программных продуктов, предназначенных для физической реализации заданной методической модели конкретного банка в архитектуре BPM. На рисунке 33 представлена архитектура BPM-решения компании Intersoft Lab на основе хранилища данных «Контур».



Рис. 32. Информационно-методический комплекс платформы «Контур». Источник: Intersoft Lab

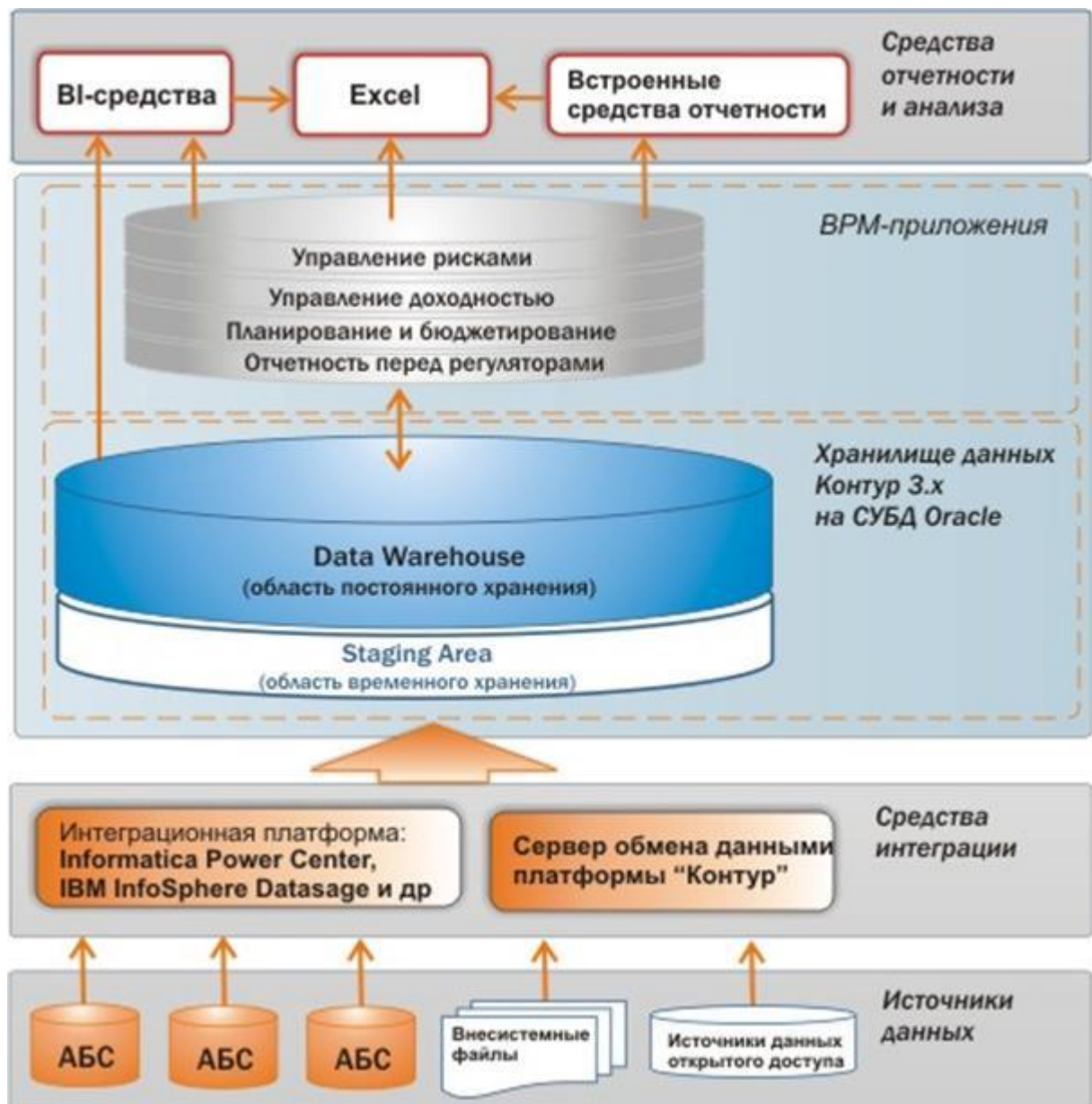


Рис. 33. Архитектура BPM-решения на основе хранилища данных «Контур». Источник: Intersoft Lab

Хранилища данных «Контур» построено на СУБД Oracle Database Server, в качестве СУБД можно использовать также Microsoft SQL Server.

Система BPM-приложений платформы «Контур» реализует методическую модель управления эффективностью банка от компании Intersoft Lab. В ее состав входят следующие модули:

- Планирование и бюджетирование.
- Управление доходностью.
- Управление рисками.
- Отчетность перед регуляторами.

- Финансовая консолидация.

В структуре BPM-решения от компании Intersoft Lab используются средства интеграции, как собственной разработки, так и средства известных производителей:

- Informatica PowerCenter – базовый продукт компании Informatica.
- InfoSphere DataStage – продукт компании IBM.

В проектах построения систем управления эффективностью бизнеса для автоматизации подготовки отчетности и анализа данных компания Intersoft Lab использует встроенные собственные средства анализа, широко распространенные средства анализа пакета Microsoft Office, программную платформу Contour Business Intelligence (Contour BI), аналитическую платформу Oracle Business Intelligence.

Вопрос 2. Технология внедрения BPM.

Процесс внедрения BPM-системы как и любой ИС состоит из нескольких этапов:

1. Предпроектное обследование, формирование бизнес требований к BPM-системе.
2. Разработка технического задания на внедрение BPM-системы.
3. Формирование команды проекта внедрения.
4. Адаптация системы на основе информации, полученной на этапе обследования.
5. Установка и настройка программного обеспечения.
6. Опытная эксплуатация системы.
7. Окончательная настройка системы с учетом недочетов, выявленных во время опытной эксплуатации.
8. Обучение персонала организации.

1. Предпроектное обследование. Цель предпроектного обследования – обоснование необходимости реализации проекта внедрения BPM-системы и формирования бизнес-требований к BPM-системе.

Предпроектное обследование включает в себя:

- сбор информации на основе анализа документальных источников и интервьюирования;
- изучение целей и задач компании, ее структуры, нормативной, деловой и политической среды, критических факторов

успеха, достоинств и недостатков, связанных с управлением бизнесом.

Обследование проводится с методологической и технической точек зрения.

Методологическое обследование включает:

- анализ бизнес-процессов организации, их выделения и систематизации;
- определение достоинства и недостатка существующей системы управления;
- анализ текущего состояния методологии управления и ее готовности к автоматизации.

При необходимости принимаются решения об изменении существующих практик учета или функциональных моделей, предлагаются меры по совершенствованию методологии.

Техническое обследование включает в себя анализ состава и состояния источников первичных данных, возможностей аппаратных средств. Анализируются объемы первичных данных, проверяются состав и качество первичных данных в учетных системах, их достаточность для решения поставленных задач управления. Рассматриваются информационные системы организации, оценивается их готовность к интеграции с системой управления эффективностью.

Срок проведения предпроектного обследования зависит от масштабов компании и предстоящего ВРМ-проекта, а также от готовности заказчика сотрудничать со специалистами, проводящими обследование.

По итогам предпроектного обследования формируются документы, в которых фиксируются результаты обследования, формируются бизнес-требования к составляющим ВРМ-проекта, определяются цели, задачи и приоритеты проекта, определяется состав предстоящих работ по разработке методик и их автоматизации.

2. Разработка технического задания на внедрение ВРМ-системы. Техническое задание (ТЗ) – это технический документ (спецификация), оговаривающий набор требований к системе и утверждённый как заказчиком/пользователем, так и исполнителем/производителем системы. В общем случае ТЗ может состоять из следующих разделов:

1. Назначение и цели внедрения ВРМ-системы.
2. Характеристики объекта автоматизации.
3. Требования к системе.

ТЗ составляется бизнес-консультантами или специалистами компании поставщика совместно с ИТ-менеджерами и представителями пользователей компании-заказчика. Чем подробнее будет составлено ТЗ, тем лучше. На этом этапе очень важен процесс документирования, чтобы потом не возникали проблемы расхождения взглядов компании-заказчика и внедряющей стороны.

3. Формирование команды проекта внедрения ВРМ-системы. Условием успеха проекта внедрения ВРМ-системы является тесное сотрудничество между заказчиком и разработчиком. Сотрудничество должно происходить на всех фазах жизненного цикла проекта.

Для реализации проекта, как со стороны заказчика, так и со стороны разработчика, должны быть созданы проектные команды, которые будут тесно взаимодействовать друг с другом. Состав команд определяется характером и масштабом проекта.

4. Адаптация системы на основе информации, полученной на этапе обследования. Приступать к развертыванию ВРМ-системы можно только после детальной проработки методической модели управления. Моделирование должно проводиться хорошо обученными сотрудниками предприятия с привлечением высококвалифицированных консультантов и с привязкой созданной модели к стандартам бизнеса и будущей системе.

В результате моделирования создаются:

- документы, описывающие все бизнес-объекты и бизнес-процессы организации;
- расчетные алгоритмы методической модели управления;
- состав отчетных форм.

Документы становятся основой для дальнейшей автоматизации методики.

На этом этапе происходит доработка и утверждение ТЗ. И здесь большую роль должны играть ключевые пользователи системы: владельцы бизнес-процессов. Пользователи – специалисты в своей области, они знают, чего они хотят.

5. Установка и настройка программного обеспечения. План установки системы зависит от выбранной стратегии внедрения ВРМ-системы. В зависимости от специфики решаемой задачи при внедрении ВРМ-системы компании обычно используют один следующих подходов:

1. Полная автоматизация или метод «Большого взрыва».
2. Автоматизация по направлениям или «Перекрестный подход» (внедрение в функционально связанных отделах).

3. Автоматизация по участкам или «Функциональный подход».

Установка и настройка программного обеспечения производится в соответствии с требованиями ТЗ.

6. Опытная эксплуатация системы. Эксплуатация системы подразумевает непосредственное использование ВРМ-системы для выполнения ею тех функций, для которых она предназначена. На этом этапе можно объективно оценить все сделанное ранее.

Запуск системы в эксплуатацию может производиться различными способами. Здесь тоже существует своя стратегия, с которой предприятию нужно определиться. Известны три способа начала использования новой системы:

1. *Параллельная стратегия* – в ходе эксплуатации работают как старая, так и новая системы, сравниваются результаты работы обеих систем и, если их выходные документы согласуются длительное время, можно переходить на новую систему.

2. «Скачок». Суть этой стратегии в том, что старая система резко прекращает работать и тут же начинает работать новая.

3. *Опытная эксплуатация «пилотного» проекта* – это тактика «скачка», но применяемая к ограниченному числу процессов. Область применения стратегии – малый участок деятельности.

Начало рабочей эксплуатации является самым критическим моментом в проекте.

7. Окончательная настройка системы с учетом недочетов, выявленных во время опытной эксплуатации. Потребность доработки системы возникает в случаях, когда:

- эксплуатация системы показала, что планируемые цели не достигнуты;
- выявлены ошибки в настройке системы;
- изменились потребности бизнеса.

Таким образом, инициация процесса доработки опирается на информацию об объективных показателях эффективности системы, измеряемых в процессе ее эксплуатации.

8. Обучение персонала организации. Одна из главных задач руководства организации заказчика и разработчика – активное обучение будущих пользователей ВРМ-системы, повышение уровня их квалификации как пользователей.

В дальнейшем стоит предусмотреть постоянную переподготовку сотрудников в связи с изменением технологий и стандартов, развитием ВРМ-системы.

Методология внедрения.

Важность методологии внедрения признают все разработчики автоматизированных систем, у большинства из них имеются свои собственные методические рекомендации. Подходы разных производителей ВРМ-платформ имеют много общего, но при этом каждый из них учитывает специфику предметной области, а также особенности того или иного программного продукта.

Конкретную методологию, относящуюся к внедрению ВРМ-систем в банках, предлагает компания Intersoft Lab. Методология основана на использовании *информационно-методического комплекса* (ИМК). Она включает четыре обязательных этапа:

1. Конструирование индивидуального комплекса управленческих методик банка на основе готовой методической модели.
2. Выбор компонент информационного комплекса для построения ВРМ-системы банка.
3. Разработка технического задания на настройку ВРМ-системы для поддержки индивидуальной методической модели.
4. Настройка ВРМ-системы в соответствии с техническим заданием и ввод ее в эксплуатацию.

Использование ИМК позволяет:

- повысить прозрачность и управляемость процесса внедрения ВРМ-системы;
- разграничить роли и зоны ответственности участников проекта внедрения системы на каждом технологическом этапе.

На этапе предпроектного обследования консультанты методисты проводят необходимые исследования, выясняют требования банка и готовят описание индивидуальной методической модели.

На этапе выбора компонент ИТ-консультанты выполняют аудит результатов методической постановки ВРМ-решения на предмет реализуемости на выбранной программной платформе и определяют состав информационного комплекса.

На следующем этапе ИТ-консультанты разрабатывают техническое задание на настройку ВРМ-системы.

На этапе настройки специалисты по внедрению выполняют установку и настройку готового решения.

Ролевое разграничение дает возможность банку-заказчику согласовывать и контролировать результаты каждого исполнителя в проекте внедрения ВРМ.

Вопрос 3. Характеристика рынка инструментальных средств бизнес интеллекта, управления эффективностью бизнеса.

По данным статистики основными игроками на рынке BPM-систем являются компании:

- Oracle Hyperion.
- SAP (Business Object).
- IBM (Cognos).
- SAS.
- российская компания Intersoft Lab.

Oracle Hyperion Performance Management Applications представляет собой комплекс прикладных бизнес приложений для поддержки процессов управления эффективностью (EPM), комплекс отраслевых аналитических приложений, общую информационно-аналитическую платформу, объединяющую инструменты и сервисы бизнес-анализа, а также различные источники данных и интеграционную платформу Oracle Fusion Middleware.

Комплекс решений поддерживает весь цикл управления, охватывающий процессы формирования целей, стратегических карт, ключевых показателей эффективности, моделирование сценарных прогнозов, планирование, мониторинг, анализ и формирование отчетов финансовых и операционных результатов.

Информационно-аналитическая платформа Oracle включает в себя средства:

- хранения данных Сервер Oracle Database;
- извлечения, преобразования и загрузки данных Oracle

Warehouse Builder;

- для анализа данных Oracle Business Intelligence.

Платформа обеспечивает полное интегрированное решение для создания хранилищ данных и эффективного анализа информации.

SAP NetWeaver Composition Environment – эта открытая технологическая платформа, на которой базируются все решения SAP. Она позволяет компаниям создать единое информационное пространство, объединив всех пользователей в единый процесс эффективного управления бизнесом. Она не просто связывает

информационные системы между собой, а обеспечивает полную интеграцию бизнеса.

Платформа содержит две составляющие:

1. Решение для интеграции.
2. Инструмент композиции.

Разделение платформы на два решения позволяет гибко использовать предлагаемый функционал.

В интеграционной платформе SAP NetWeaver реализованы инновационные технологии интеграции бизнес-среды предприятий и организации широкого, эффективного сотрудничества, охватывающего всех участников бизнеса. Она содержит несколько уровней интеграции:

- 1) уровень «интеграции приложений»;
- 2) уровень «интеграции процессов»;
- 3) уровень «интеграции информации»;
- 4) уровень «интеграции пользователей».

Пакетные композитные приложения SAP – это новое поколение программных решений. С их помощью можно оперативно создавать новые решения для бизнес-процессов путем комбинирования существующих функциональных возможностей и добавления новых функциональных модулей, т.е. строить композитные приложения над уже существующими на предприятии приложениями.

Композитные приложения SAP xApp Analytics – набор инновационных аналитических приложений, построенных в соответствии с концепцией сервисно-ориентированной архитектуры (SOA).

ВРМ-решения IBM Cognos. Решение IBM Cognos обеспечивает возможности планирования деятельности компании с помощью широкого набора интегрированных функций: управление по системе сбалансированных показателей, моделирование и постановка задач, планирование, бюджетирование, прогнозирование. Системы СРМ представляют собой набор мощных возможностей многомерного анализа и бизнес-планирования и бюджетирования в масштабах всей компании.

В состав СРМ-решения входят три продукта:

- 1) IBM Cognos Planning.
- 2) IBM Cognos Business Intelligence.
- 3) IBM Cognos Controller.

Продукты тесно интегрированы между собой и позволяют решать корпоративные задачи управления эффективностью бизнеса.

ВРМ-решения компании SAS. Системы управления эффективностью (*англ.* Corporate performance management (СРМ-системы)) дают целостное представление об эффективности деятельности организации. Они объединяют решения в области бизнес-аналитики с задачами консолидации управленческой отчётности, стратегического планирования, бюджетирования, анализа прибыльности, оптимизации затрат, управленческой отчётности и мониторинга деятельности организации.

Комплекс объединяет целый ряд технологий и бизнес-решений в единый пакет SAS Financial Intelligence. СРМ-решение базируется на Единой Аналитической Платформе (SAS Enterprise Intelligence Platform).

Вопросы для самопроверки:

1. Перечислите основные процессы в цикле ВРМ.
2. Из каких элементов состоит типовая архитектура ВРМ-системы?
3. Для чего предназначена методологическая модель ВРМ-решения?
4. Что собой представляет информационная модель ВРМ-решения?
5. Что выступает в качестве ключевого элемента методической модели ВРМ-решения компании Intersoft Lab?
6. Какие модули входят в состав ВРМ-приложений платформы «Контур»?
7. Из каких этапов состоит процесс внедрения ВРМ-системы?
8. Охарактеризуйте этап предпроектного обследования.
9. Зачем требуется проводить обучение персонала?
10. Назовите основных игроков на рынке ВРМ-систем.

Литература по теме:

1. Информационные аналитические системы: учеб. / Т.В. Алексеева, Ю.В. Амириди, В.В. Дик. – М.: Московский финансово-промышленный ун-т «Синергия», 2013.
2. Амириди Ю.В. Информационные системы в экономике. Управление эффективностью банковского бизнеса: учеб. пособие / Ю.В. Амириди, Кочанова Е.Р., Морозова О.А. – М.: Кнорус, 2009 г.
3. Духонин Е.Ю., Исаев Д.В. Управление эффективностью бизнеса. Концепция Business Performance Management / Под ред. Г.В. Генса. – М.: Альпина Бизнес Букс, 2005.

4. Амириди Ю., Иванова Н. Внедрение ВРМ-систем: мировые тенденции и отечественный опыт. – [Электронный ресурс]. – Режим доступа: <http://www.iso.ru/publication/document8630.phtml>.

Интернет ресурсы:

1. Сайт компании «Intersoft Lab». – Режим доступа: <http://www.iso.ru/ru>