**Лекция 5:**

**Системы деловой осведомленности (Business Intelligence Systems) и хранилища данных**

**Введение**

В этой лекции мы будем использовать многие понятия на интуитивном уровне, не давая им точных определений. Такие понятия будут определены в последующих лекциях.

**Определение систем деловой осведомленности**

**Понятие деловой осведомленности**

Многозначность английского слова "Intelligence" приводит к неопределенности трактовки термина "Business Intelligence" как в российских, так и в зарубежных литературных источниках, посвященных тематике использования информационных технологий для аналитической поддержки бизнеса. Английское слово "Intelligence" означает способность узнавать и понимать, готовность к пониманию, знания, переданные или приобретенные путем обучения, исследования или опыта, действие или состояние в процессе познания, разведку, разведывательные данные. В русском языке слово "интеллект" означает мыслительную способность человека.

**Термин "Business Intelligence"** получил широкое распространение, когда был введен в обращение аналитиками компании Gartner Group в конце 80-х годов прошлого века как **"пользователецентрический процесс, включающий доступ и исследование информации, ее анализ, выработку интуиции и понимания, которые ведут к улучшенному и неформальному принятию решений"**. Хотя ранее этот термин, например, использовался в компании IBM в качестве внутрикорпоративного термина.

К 1996 году содержание термина было уточнено, и "Business Intelligence" стал пониматься как **"инструменты для анализа данных, построения отчетов и запросов, которые могут помочь бизнес-пользователям преодолеть море данных для того, чтобы помочь синтезировать из них значимую информацию"**.

В русскоязычной литературе термин "Business Intelligence" переводится как "бизнес-интеллект", "*интеллектуальный анализ данных*", "деловая осведомленность" или вводится просто как аббревиатура BI. В настоящем курсе мы будем использовать термины "деловая осведомленность" и "бизнес-аналитика" как синонимы.

Однако все-таки в настоящее время не существует однозначного определения термина "деловая осведомленность" (BI). Отметим следующие важные аспекты трактовки содержания данного термина.

* *Деловая осведомленность* понимается как процесс, методы, технологии, средства извлечения и представления знаний.
* *Деловая осведомленность* понимается как знания о бизнесе и для бизнеса.

Таким образом, *деловая осведомленность* в широком смысле слова понимается как:

* **процесс превращения данных в информацию и знания о бизнесе для обеспечения принятия улучшенных и неформальных решений;**
* **информационные технологии сбора данных, консолидации информации и обеспечения доступа пользователей к бизнес-знаниям;**
* **знания о бизнесе, полученные в результате анализа данных и консолидированной информации**.

**Системы бизнес-аналитики**

В современном мире успех компании на рынке напрямую зависит от того, как быстро менеджмент компании может распознать изменения динамики рынка и насколько своевременно может отреагировать на них с целью увеличения прибыли, исходя из существующих реалий рынка. Менеджеры компании должны отслеживать тенденции рынка, идентифицировать конкурентов и угрозы, оценивать риски, преобразовывать стратегию компании, оценивать свои ресурсы и т.д. Информация является необходимым производственным ресурсом для принятия эффективных управленческих решений.

Компании накопили значительные объемы данных и имеют доступ к еще большим объемам внешних данных. Менеджерам необходимо, чтобы эта информация была преобразована, предварительно обработана и соответствующим образом организована для быстрого доступа, анализа и принятия решений. Такой подход к данным есть, с одной стороны, создание конкурентного преимущества, а с другой стороны – требование к публикации данных для менеджеров компании. Публикация данных для менеджеров, обеспечивающая быстрый доступ к данным, выполнение анализа данных и информационную поддержку процесса принятия решений, является основной целью систем бизнес-аналитики. Бизнес-аналитика помогает компании создавать знания из всей доступной информации для принятия эффективных управленческих решений и превращения этих решений в действие.

Таким образом, ключевую роль в управлении организацией в целом и ее отдельными производственными функциями играет информация. Данные, которые доступны менеджерам и аналитикам непосредственно из корпоративных информационных систем, не унифицированы, разрозненны и в общем случае не готовы для анализа. *Системы деловой осведомленности или бизнес-аналитики* являются тем классом информационных систем, который позволяет превратить данные корпоративных информационных систем и данные из внешних источников в полезные для бизнеса информацию и знания, используемые в управлении, на основе которых можно принимать решения.

Информационным фундаментом для бизнес-анализа и систем бизнес-аналитики является хранилище данных. Основное требование к хранилищу данных системы бизнес-анализа состоит в том, чтобы обеспечить структурированную и организованную для решения задач бизнеса информационную среду. Как правило, такую среду лаконично представляют в виде *информационной пирамиды*, как показано на [рис. 4.1](https://intuit.ru/studies/courses/599/455/lecture/10161?page=1#image.4.1).



**Рис. 4.1.**Информационная пирамида

*Информационная пирамида* формируется из нескольких уровней.

* **Уровень оперативной информации**. На этом уровне ИТ обеспечивают работу с данными на уровне бизнес-процедур компании. Данные в автоматизированных системах являются хорошо структурированными и детальными. С этими данными работают специалисты компании: бухгалтеры, менеджеры продаж, плановики и т.д.
* **Уровень тактической информации**. На этом уровне ИТ обеспечивают интеграцию данных на уровне бизнес-процессов оперативного управления производством в рамках подразделений компании. С этими данными работают руководители подразделений компании при выполнении ежедневных *производственных заданий*.
* **Уровень стратегической информации**. На этом уровне ИТ обеспечивают интеграцию данных на уровне бизнес-процессов по направлениям хозяйственной деятельности компании. С этими данными работают аналитики и руководители высшего звена компании, которые готовят стратегические решения развития и деятельности компании на рынке.
* **Уровень принятия решений**. На этом уровне ИТ обеспечивают интеграцию и агрегацию данных на уровне бизнес-процессов компании для руководителей высшего звена компании. Этот уровень обеспечивает информационную поддержку принятия решений.

*Информационная пирамида* описывает среду бизнес-аналитики, которую можно описать следующим образом. В информационную среду бизнес-аналитики поступает первичный материал — данные, которые затем перерабатываются в автоматизированных системах и информационных продуктах.

В процессе переработки происходит переход от данных к информации. ХД извлекает данные из множества транзакционных или оперативных систем, а затем интегрирует и хранит данные в специализированной БД. Например, в ХД могут приводиться в соответствие и объединяться пользовательские записи из четырех оперативных систем (приложений для обработки заказов, обслуживания, продаж и поставок). Такой процесс извлечения и интеграции преобразует данные в новый информационный продукт — информацию.

Затем пользователи, работающие с аналитическими инструментами (например, для создания запросов, отчетов, OLAP-анализа и выполнения операций *интеллектуального анализа данных*), обращаются к данным из ХД и анализируют ее. Таким образом, определяются тенденции, структуры и исключения. Аналитические инструменты помогают пользователям преобразовать информацию в знания.

Теперь дадим определение *систем бизнес-аналитики или систем деловой осведомленности*.

**Определение 4.1**. ***Система деловой осведомленности, или бизнес-аналитики (BI System)*** **, — это система управления базой знаний предприятия, которая предоставляет ряд новых возможностей в существующей информационной системе предприятия для анализа бизнеса и управления базой знаний предприятия**.

К основным функциям системы бизнес-аналитики, как правило, относят следующие.

* Управление в реальном времени бизнес-знаниями в рамках всего предприятия.
* Простой доступ к информации для сотрудников компании различных уровней.
* Рост объема перерабатываемой информации, повышение конкурентоспособности.
* Проведение более эффективного анализа доходов и расходов.
* Предоставление исполнительным директорам комплексной и более наглядной картины предприятия по всем направлениям бизнеса.
* Отслеживание ситуации как на всем предприятии в целом, так и на каких-то конкретных проблемных участках в частности.

К основным технологическим средствам реализации функциональности систем бизнес-аналитики относят:

* отчеты и средства их создания;
* специализированные средства создания отчетов;
* генераторы отчетов, встроенные в средства разработки;
* нетрадиционные средства создания отчетов;
* OLAP-средства;
* клиентские OLAP-средства;
* серверные OLAP-cредства;
* средства поиска закономерностей (Data Mining-средства) и т.д.

**Система бизнес-аналитики является стержнем, вокруг которого формируются потоки стратегической бизнес-информации**. Данный инструмент помогает компании принимать решения, которые будут основаны на корректной информации, полученной вовремя.

В условиях, когда рынок постоянно меняется, а конкуренция становится все жестче, руководителям крайне необходимо выявлять и анализировать имеющиеся у предприятия резервы, которые могут существенно расширить возможности бизнеса.

Предлагаемые решения в области бизнес-аналитики должны предоставлять возможность оперативно анализировать тенденции рынка, осознавать движущие силы бизнеса и, основываясь на объективной информации, быстро реагировать на изменения рыночной ситуации и принимать верные решения.

Например, одним из возможных решений может быть графический инструмент для экономического анализа, относящийся к категории OLAP-приложений (On-line Analytical Processing), который:

* позволяет регулярно проводить анализ выполнения производственной программы;
* позволяет проводить анализ отклонений финансовых показателей, таких как, например, анализ расходов по приобретению сырья и материалов;
* позволяет собирать, обобщать, анализировать и представлять данные в виде легко читаемых графиков и аналитических приложений. В сочетании с системой планирования и управления ресурсами предприятия ERP обеспечивает гибкий, многоплановый анализ бизнеса на основе операционных данных;
* позволяет создавать так называемые "информационные кубы" — виртуальные информационные центры, которые содержат аналитические данные, существенные для выявления тенденций рынка, анализа и принятия стратегических решений.

Эти многомерные "информационные кубы" собирают и хранят всю информацию о деятельности предприятия. С их помощью можно моделировать и анализировать критические аспекты бизнеса, учитывая информацию о продукции, поставщиках, потребителях, товарообороте, ценах и доходах. Анализ ведется интерактивно, в реальном времени, с помощью удобных визуальных инструментов, а не просто на основе многочисленных отчетов с тысячами страниц, таблиц и чисел.

"*Кубы данных*" должны быть настроены для решения ряда критически важных аспектов бизнеса, включая анализ продаж, запасов, финансов, каналов снабжения и производства.

Специальные возможности должны обеспечивать мгновенную детализацию данных и всестороннее исследование проблемы. Результирующее двух- или трехмерное представление удобно для быстрого изучения тенденций и анализа отклонений.

Типовой состав программного обеспечения систем бизнес-аналитики включает в себя не только саму систему, но и обучающие материалы, техническую документацию, а также возможность получения технической поддержки и профессиональных консультаций. Все это помогает быстро и в совершенстве освоить систему, получить максимум преимуществ ее использования:

* система для оперативного анализа;
* стандартный каталог;
* стандартные шаблоны и модели;
* стандартные отчеты;
* обучение, поддержка, документация, профессиональные консультации.

Система бизнес-аналитики должна:

* иметь единый графический интерфейс пользователя (GUI), который обеспечивает интуитивную навигацию и комфортность в работе;
* предоставлять гибкий анализ и отчетность сотрудникам различных подразделений для просмотра сводных данных, используя одни и те же разрезы деятельности в сопоставимых показателях;
* иметь открытую архитектуру и организационную масштабируемость, которые дают контролируемое, последовательное и быстрое развертывание аналитической системы во всех подразделениях предприятия;
* иметь мощную систему административного контроля, которая освобождает ИT-службу от необходимости составления многих форм отчетности, но в то же время обеспечивает контроль доступа к базам данных, конфиденциальность и мониторинг изменений;
* обеспечивать быстрое внедрение, которое способствует скорейшему получению практических результатов и ускоряет отдачу от инвестиций в программное обеспечение.

Таким образом, системы бизнес-аналитики позволяют:

* лучше анализировать структуру покупательского спроса и потребности ваших клиентов;
* использовать гибкую стратегию продаж и целевой маркетинг;
* по-новому представить на рынке возможности вашей продукции;
* выявить слабые места в цепочке поставок;
* исключить неоправданные производственные и коммерческие расходы;
* выявить стратегические тенденции в изменении ценовой структуры и номенклатуры выпускаемой продукции.

**Информационная безопасность систем бизнес-аналитики**

Следует отметить, что многие компании не придают значения вопросам безопасности, игнорируя тот факт, что архитектурные компоненты систем бизнес-аналитики таят в себе определенную опасность. Обеспечение безопасности среды бизнес-аналитики – не менее важная задача, чем защита оперативных приложений.

Необходимость безопасности систем оперативной обработки транзакций (On-Line *Transaction* *Processing*, *OLTP*) осознается большинством компаний. Особенность реализации этой задачи для *OLTP*-приложений заключается в том, что она хорошо поддается структуризации и является статичной (определенные приложения каждый раз одинаковым образом обращаются к определенным данным). Круг пользователей весьма ограничен — это работники с определенными бизнес-функциями, они работают с приложениями и данными, которые касаются только их поля деятельности. Кроме того, *физическая структура* этих приложений также остается довольно постоянной. Инструментальные средства и *базовая* *структура данных* меняются нечасто.

Среда бизнес-аналитики и ХД, наоборот, характеризуется значительной динамичностью вкупе с широкой и часто меняющейся пользовательской аудиторией, причем пользователи могут быть как внутренними, так внешними. В такой ситуации гораздо сложнее (а иногда и практически невозможно) распределить пользователей по подмножествам данных; особенно это касается аналитических приложений высокого уровня, таких как, например, решения управления эффективностью корпорации (*corporate* *performance management*), где окончательная *информация* формируется на основе изучения данных всего предприятия. Помимо этого, *физическая структура* этой среды часто является неясной: в нее устанавливается множество различных средств, а сами данные пребывают в постоянном движении (из ХД в витрины данных и на пользовательские машины в информационные панели). В результате мероприятия по обеспечению безопасности корпоративной информации обходят стороной приложения бизнес-аналитики и ХД.

Для того чтобы гарантировать защищенность среды бизнес-аналитики, компании должны в первую *очередь* выполнить задачи безопасности, возникающие на уровне отдельных ее компонентов (см. [рис. 4.2](https://intuit.ru/studies/courses/599/455/lecture/10161?page=1#image.4.2).).



**Рис. 4.2.**Оболочка решений бизнес-аналитики

Каждый из основных компонентов среды бизнес-аналитики имеет свою степень риска и для обеспечения безопасности каждого компонента потребуется реализовать различные подходы (и различные технологии). Это крайне непростая задача, и, пожалуй, наибольшую сложность представляют "пробелы" между компонентами. Ведь программная *оболочка* для бизнес-аналитики практически никогда не поставляется одним поставщиком или в форме одной ИТ-технологии. При этом бесшовная *интеграция* между компонентами невозможна. Более того, именно то, как компоненты работают друг с другом, и то, как *информация* проходит между ними, и образует "точки риска".

**Безопасность данных**

Сама суть бизнес-аналитики подталкивает бизнес-пользователей к расширению доступа к данным и контроля над ними. Поэтому необходима жесткая политика по защите информации, которая должна помочь "залатать дыры", созданные многочисленными, слабо интегрированными технологическими компонентами, а также минимизировать огромный риск, присущий человеческому фактору.

Данные в ХД, витринах данных и операционных складах данных создают условия для осуществления всей бизнес-аналитики и, как правило, включают гигантские объемы детальных, транзакционных данных. Поскольку они часто отображают длительный отрезок времени, относящейся к истории существования компании, как, например, финансовая информация, обеспечение защищенности таких данных чрезвычайно важно. При рассмотрении задач безопасности данных следует задаться следующими вопросами:

* Кто располагает доступом к ХД, витрине данных, кубам и так далее?
* Каковы рамки их доступа: одна предметная область, множество предметных областей или все предметные области?
* Каким типом доступа они обладают, например, только чтение или возможность модификации?

Топология данных в среде бизнес-аналитики влияет на возможности доступа к данным и обеспечение безопасности. Во многих компаниях результаты запросов часто загружаются на индивидуальные машины с целью дальнейшей детализации и использования. Эти данные оказываются в витринах данных, настольных БД, информационных панелях или крупноформатных электронных таблицах и быстро оказываются вне пределов *инфраструктуры безопасности* ИT-отдела, хотя по-прежнему сохраняют свою конфиденциальную сущность. При рассмотрении топологии данных с точки зрения безопасности необходимо изучить следующие вопросы:

* Насколько распределенной является архитектура данных, поддерживающая бизнес-аналитику?
* Имеется ли дополнительное распределение данных и если да, то каковы связанные с ним риски?
* Что делают пользователи с загружаемыми данными?
* Передают ли пользователи данные внешним партнерам?

**Процесс сбора данных**

Обычно процесс сбора и подготовки данных для среды бизнес-аналитики очень сложный и "непрочный". Огромное число источников данных и значительное разнообразие данных приводят к многоступенчатым процессам, в которых данные интерактивно собираются и преобразуются для загрузки в ХД. Данные, подвергающиеся как процессу сбора, так и преобразования, также образуют следующие "точки риска".

* Кто располагает доступом к средствам извлечения данных из операционных систем?
* Где находятся данные, пребывающие в процессе сбора, перед тем как оказаться в ХД, и кто имеет доступ к этой области?
* Какова логика преобразования, безопасность которой реализуется в средствах извлечения, преобразования и загрузки (*ETL*)?
* Если никакие средства не используются, то какова защита *ETL*-процессов, написанных пользователем, от несанкционированных модификаций?

**Пользовательские средства формирования запросов и аналитические приложения**

Программные инструменты бизнес-аналитики и аналитические приложения — это, в первую очередь, механизмы, предназначенные для доступа к данным в ХД. Такие средства часто приобретались в большом количестве с целью широкого и глубокого развертывания бизнес-аналитики по всему предприятию. Эти инструменты представляют особую ценность только для определенных пользователей и несут серьезную опасность, если попадают не в те руки.

* Кто располагает разрешением на использование средств формирования запроса и отчетности?
* Назначен ли каждому пользователю личный ID?

Появление и развитие аналитических приложений для электронной коммерции по схеме "бизнес-бизнес" (business-to-business) и "поставщик-покупатели" (business-to-consumer) усилили насущность вопросов безопасности.

* Насколько свободно ваши клиенты и поставщики обмениваются предоставленной им информацией в рамках своих предприятий?
* Предоставляют ли они ее своим внешним акционерам?
* Не может ли эта информация попасть в руки ваших конкурентов?

**Политика информационной безопасности**

Корпоративная политика информационной безопасности часто не затрагивает информации, которая хранится, анализируется и поставляется посредством аналитических приложений. Поскольку бизнес-аналитика расширяет доступ к информации, часто передавая ее в непосредственное распоряжение бизнес-пользователей, информация быстро оказывается вне пределов *инфраструктуры безопасности* ИT-отдела. Поэтому при формировании корпоративной политики информационной безопасности необходимо рассмотреть следующие вопросы:

* Учитывает ли корпоративная политика безопасности специфику ИT?
* Имеются ли области значительного риска, которые могут быть устранены посредством такой политики?
* Затрагивают ли правительственные постановления информацию, которая хранится, анализируется и представляется среде бизнес-аналитики?
* Не нарушают ли текущие или планируемые мероприятия по развертыванию среды бизнес-аналитики эти постановления?

**Технологии обработки данных**

**Технология диалоговой аналитической обработки (OLAP)**

Концепция многомерного представления данных предполагает, что элементы данных (фактическая информация) являются точками многомерного пространства, размерности которого представляют собой содержательное описание таких фактов (точку зрения на них). В приложениях обработки многомерных данных сохраняются все проблемы с визуализацией многомерных массивов данных. Самые продвинутые, дорогостоящие и элитарные решения пребывают в пределах своих узких предметных ниш.

Конечные пользователи не намерены каким-то образом продвигать за свой счет электронные таблицы в интерфейс с многомерными БД. Конечно, электронные таблицы, в силу их удобства и простоты, являются излюбленным инструментом конечных пользователей. Однако, как показывает опыт, электронные таблицы хороши лишь тогда, когда они "заточены" под многомерность конкретных предметных областей.

Приложения OLAP зачастую бывают весьма громоздкими, обычно их рентабельность отвечает использованию в составе корпоративных рабочих групп, например в аналитических службах. Вообще, для эффективного использования решений OLAP нужна поддержка корпоративной инфраструктуры.

Как показывает анализ, Web-архитектуры быстро вытесняют традиционные клиент-серверные приложения для целого ряда категорий программного обеспечения, и рынок корпоративных OLAP-решений здесь не исключение.

Это направление стремительно развивается за счет появления различных Web-OLAP инструментов на базе HTML- и Java-технологий от известных поставщиков и быстро растущих новых компаний. В связи с расширением контингента пользователей продукты Web-OLAP разрабатываются для выполнения несколько иного анализа, чем традиционные клиент-серверные инструменты. Происходит переход от инструментальных средств исследования данных, ориентированных на специалистов-аналитиков, к готовым аналитическим приложениям, доступным для более широкого круга пользователей.

В [табл. 4.1](https://intuit.ru/studies/courses/599/455/lecture/10161?page=2#table.4.1) приведены критерии, которые определяют успех Web-OLAP-продуктов.

|  |
| --- |
| Таблица 4.1. Критерии успеха Web-Olap-решения |
| **Критерий** | **Описание** |
| **Удобство использования** | Успешный BI-продукт должен быть достаточно прост для неопытного пользователя, не имеющего специальной подготовки |
| **Интерактивность** | Программное средство должно реализовать интерактивные возможности, в том числе:* просмотр статических документов;
* динамическое обновление существующих документов, обеспечивающее доступ к самой свежей информации;
* динамическое выполнение нерегламентируемых запросов к источникам данных;
* динамическое неограниченное "углубление в данные" (*drill*-down)
 |
| **Функциональность** | Web-BI-приложение должно обеспечивать такие же возможности, как и традиционные клиент-серверные аналоги, удовлетворяя при этом дополнительным требованиям. Генерирование SQL, выполнение динамических пользовательских расчетов, различные методы навигации - все это необходимо и в Web |
| **Доступность и переносимость**. | Главное преимущество Web - доступность и переносимость. Информация должна быть доступна для любого устройства, рабочего места, в любой точке земного шара, вне зависимости от того, находятся ли данные в главном управлении компании, в удаленных офисах или на портативном устройстве. Клиентская часть идеального BI-продукта должна быть небольшой, чтобы удовлетворить различным уровням пропускной способности сети пользователя, а также соответствовать стандартизованной технологии |
| **Архитектура** | Поскольку Web-среда принципиально отличается от традиционной клиент-серверной, здесь возникает множество новых технологических проблем. *Многозвенная архитектура*, допускающая наличие различных типов клиентов (Java, HTML и т.п.), а также "собственное" соединение с Web-сервером (*NSAPI*, ISAPI) и сервером базы данных, необходима для корпоративного программного продукта |
| **Интеграция. Независимость от источников данных** | Корпоративная вычислительная среда содержит различные виды аппаратных и программных ресурсов, пакетных приложений и баз данных. Хорошо разработанное BI-приложение должно давать доступ к статическим документам любого типа (а не только к тем, которые оно само создает), а также интерактивный доступ к реляционным и многомерным базам данных, приложениям и другим источникам |
| **Производительность и масштабируемость** | Для обеспечения производительности и масштабируемости в Web необходимо реализовать следующие возможности:* балансировку нагрузки сервера приложения;
* собственное соединение с web-сервером;
* собственный доступ к базе данных;
* кэширование сервером приложения (данных или соединений с базой);
* персистентность, устраняющую проблему хранения состояний в Web
 |
| **Обеспечение безопасности** | Возможность администрирования через web - одно из ключевых преимуществ. Так, для изменения прав конкретного пользователя администратору не нужно появляться на его рабочем месте. Используя модули администрирования, можно создавать профили для отдельных пользователей или групп, предоставляя доступ только к авторизованной информации |
| **Стоимость внедрения и администрирования** | Стоимость внедрения Web-OLAP-решения в расчете на одного пользователя должна быть существенно ниже, чем для традиционных продуктов. Поскольку поддержка клиента - очень сложная задача для традиционных клиент-серверных продуктов, Web-решения устраняют часть накладных расходов, не требуя специального клиентского ПО, кроме браузера. Расходы на администрирование становятся значительно меньше, если:* снижается стоимость поддержки клиентской части ПО;
* снижается стоимость поддержки серверного компонента;

программа может эффективно функционировать в Web-среде, где распространяются тысячи отчетов/документов и тысячи пользователей нуждаются в защищенном интерактивном доступе к разным базам |

**Интеллектуальная обработка данных (Data Mining)**

Возникновение Data Mining (добыча данных) связано с наличием противоречия между *теоретическими методами* прикладной статистики и практикой решения реальных задач. Синонимами этого понятия являются обнаружение знаний в БД и *интеллектуальный анализ данных*.

Стимулом развития технологии Data Mining явился прорыв в технологиях электронного хранения больших объемов данных — деятельность любого предприятия сопровождается регистрацией и записью на электронные носители всех подробностей его деятельности.

Очевидно, что без технологии переработки потока этих "сырых" данных (*raw data*), последние образуют просто большую свалку.

Требования к технологии переработки:

* данные имеют неограниченный объем;
* данные являются разнородными по типу (количественными, качественными, текстовыми);
* результаты должны быть понятны и конкретны;
* инструменты для обработки сырых данных должны быть просты в использовании.

Традиционная прикладная статистика не справляется с поставленными задачами. Главная причина — она работает с фиктивными, средними величинами (концепция усреднения по выборке). Ее методы полезны при проверке заранее сформулированных гипотез (verification-driven data mining) и для грубого предварительного анализа, составляющего основу OLAP (online analytical processing).

В основу технологии Data Mining (discovery-driven data mining) положена концепция шаблонов (паттернов), отражающих фрагменты многоаспектных взаимоотношений в данных. Эти шаблоны представляют закономерности, свойственные подвыборкам данных, которые могут быть выражены в понятной человеку форме. Поиск шаблонов проводится методами, не ограниченными рамками априорных предположений о структуре выборки и виде распределений значений анализируемых показателей.

Понятно, что такие шаблоны должны быть нетривиальны (unexpected — неожиданные регулярности в данных, которые составляют так называемые скрытые знания, Hidden knowledge).

Есть понимание того, что "сырые" данные содержат глубинный пласт знаний и его нужно раскопать.

Различия в формулировках задач диалоговой аналитической обработки данных и *интеллектуального анализа данных* приведены в [табл. 4.2](https://intuit.ru/studies/courses/599/455/lecture/10161?page=2#table.4.2).

|  |
| --- |
| Таблица 4.2. |
| **OLAP** | **Data Mining** |
| Каковы средние показатели травматизма для курящих и некурящих? | Встречаются ли точные шаблоны в описании людей, подверженных повышенному травматизму? |
| Каково среднее соотношение размеров телефонных счетов существующих клиентов со счетами бывших клиентов? | Имеются ли характерные портреты клиентов, которые, по всей видимости, собираются отказаться от услуг телефонной связи? |
| Какова средняя величина ежедневных покупок по украденной и не украденной карточке? | Существуют ли стереотипные схемы покупок для случая мошенничества с кредитными карточками? |

Основные бизнес-приложения Data Mining

1. Розничная торговля (предполагается, что собирается информация о каждой покупке)
	* Анализ покупательской корзины (анализ сходства) — выявление товаров, которые покупаются вместе. Необходимо для рекламы, выработки стратегии создания запасов товаров и способов их раскладки в торговых залах
	* Исследование временных шаблонов. Создание запасов
	* Создание *прогнозирующих моделей*. Характер потребностей различных категорий покупателей с определенным поведением (приобретение товаров известных марок или распродажа). Разработка направленных мероприятий по продвижению товаров
2. Банковское дело
	* Выявление мошенничества с кредитными карточками. Анализ транзакций, которые оказались мошенническими
	* Сегментация клиентов. Маркетинговая политика банка становится более целенаправленной
	* Прогнозирование изменения клиентуры, модели ценности своих клиентов и соответствующее обслуживание каждой категории
3. Телекоммуникации
	* Анализ записей о подробных характеристиках вызовов — выявление категорий клиентов с похожими стереотипами поведения и разработка привлекательных наборов цен и услуг
	* Выявление лояльности клиентов. Определение характеристик клиентов, которые однажды воспользовавшись услугами компании, и с большей долей вероятности воспользуются еще раз
4. Страхование
	* Выявление мошенничества. Определенные стереотипы в заявлениях о выплате страховки, характеризующие взаимоотношения между юристами, врачами и заявителями
	* Анализ риска. Путем выявления факторов, связанных с оплаченными заявлениями, можно уменьшить свои потери

Необходимость автоматизированного *интеллектуального анализа данных* стала очевидной, в первую очередь, из-за **огромных массивов исторической и вновь собираемой информации**. Трудно даже приблизительно оценить объем ежедневных данных, накапливаемых различными компаниями, государственными, научными и медицинскими организациями. По мнению исследовательского центра компании GTE, только научные институты собирают ежедневно около терабайта новых данных.

Другой причиной роста популярности *интеллектуального анализа данных* ( **data mining** ) является **объективность получаемых результатов**. Человеку-аналитику, в отличие от машины, всегда присущ субъективизм: он в той или иной степени является заложником уже сложившихся представлений. Иногда это полезно, но чаще приносит большой вред.

### И, наконец, *интеллектуальный анализ данных* **дешевле**. Оказывается, что выгоднее инвестировать деньги в решения *интеллектуального анализа данных*, чем постоянно содержать целую армию высококвалифицированных и дорогих профессиональных статистиков. *Интеллектуальный анализ данных* вовсе не исключает полностью человеческую роль, но значительно упрощает процесс поиска знаний, делая его доступным для более широкого круга аналитиков, не являющихся специалистами в статистике, ма**Архитектура систем бизнес-уровня**

**Состав архитектуры современной системы бизнес-аналитики**

Задачами любой системы бизнес-аналитики являются эффективное хранение, обработка и анализ данных. В настоящее время накоплен значительный опыт в этой области.

Эффективное хранение информации достигается наличием в составе системы бизнес-аналитики 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 системы бизнес-аналитики организации в обобщенном виде представлена на [рис. 4.3](https://intuit.ru/studies/courses/599/455/lecture/10161?page=3#image.4.3).



**Рис. 4.3.**Архитектура современной системы бизнес-аналитики

Приведенная архитектура демонстрирует длинный путь, который проходят данные, прежде чем попасть на стол аналитику.

Разнообразие источников данных и необходимость их использования в каждом конкретном случае объясняется потребностью по-разному хранить информацию в зависимости от стоящих перед организацией задач. Если попытаться классифицировать источники данных по типам и назначению, то каждый из них можно условно отнести к одной из трех групп: транзакционные источники данных, ХД, витрины данных, информационные панели.

Данные в систему могут заноситься как вручную, так и автоматически. На этапе первоначальной фиксации данные поступают через системы сбора и обработки информации в так называемые транзакционные БД. Транзакционных БД в организации может быть несколько.

Поскольку транзакционные источники данных, как правило, не согласованы друг с другом, для анализа таких данных требуется их объединение и преобразование. Поэтому на следующем этапе решается задача консолидации данных, их преобразования и очистки, в результате чего данные поступают в так называемые аналитические БД. Аналитические БД, будь то ХД или витрины данных, и есть те основные источники, из которых аналитик черпает информацию, используя соответствующие инструменты делового анализа.

При этом система бизнес-аналитики среднего и крупного предприятия или организации должна обеспечивать пользователям доступ к аналитической информации, защищенной от несанкционированного использования и открытой как через внутреннюю сеть организации, так и пользователям сетей Интранет и Интернет. Таким образом, архитектура современной системы бизнес-аналитики является многоуровневой и включает 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.

Рассмотрим перечисленные уровни архитектуры и остановимся на примерах типовых инструментов, которые могут служить основой для построения каждого из них.

**Сбор и первичная обработка данных**

К первому уровню архитектуры системы бизнес-аналитики относятся упоминавшиеся уже источники данных, обычно именуемые транзакционными или операционными источниками (базами) данных, являющиеся частью так называемых OLTP-систем (*online transactional processing*). Транзакционные БД включают в себя источники данных, ориентированные на фиксацию результатов повседневной деятельности организации. Требования, предъявляемые к транзакционным БД, обусловили их следующие отличительные особенности: способность быстро обрабатывать данные и поддерживать высокую частоту их изменения, ориентированность, как правило, на обслуживание одного процесса, а не всей деятельности организации в целом.

Примерами здесь могут служить БД, которые используются в биллинговых системах операторами сотовой связи, в автоматизированных банковских системах коммерческих и государственных банков, в интернет-магазинах.

Информация в таких БД ориентирована на конкретное приложение и управляется транзакциями, она сильно детализирована и часто корректируется.

Транзакционные БД отлично справляются с валом повседневной информации, которая должна рутинно обрабатываться каждый день, но не позволяют получить общую картину положения дел в организации в целом и редко могут служить источниками для проведения комплексного анализа.

Итак, совокупность транзакционных источников данных образует нижнее звено архитектуры системы бизнес-аналитики любой организации. В дальнейшем будем исходить из того, что такая система предприятия строится на основе уже имеющихся на вооружении систем сбора и первичной обработки данных, включающих транзакционные источники данных.

**Извлечение, преобразование и загрузка данных**

Процесс извлечения, преобразования и загрузки данных поддерживается так называемыми *ETL*-инструментами (extraction, transformation, loading), предназначенными для извлечения данных из различных транзакционных источников нижнего уровня, их преобразования и консолидации, а также загрузки в целевые аналитические БД — ХД и витрины данных. На этапе преобразования устраняется избыточность данных, проводятся необходимые вычисления и агрегирование данных. Трехступенчатый процесс извлечения, преобразования и загрузки должен осуществляться на основе установленного регламента.

**Складирование данных**

К третьему уровню архитектуры системы бизнес-аналитики относятся источники данных, которые называют ХД (от англ. *Data Warehouse*). ХД включают в себя источники данных, ориентированные на хранение и анализ информации. Такие источники могут объединять информацию из нескольких транзакционных систем и позволяют анализировать ее в комплексе с применением современных программных инструментов делового анализа данных.

Напомним, что по определению ХД является предметно-ориентированной, интегрированной, некорректируемой, зависимой от времени коллекцией данных, предназначенной для поддержки принятия управленческих решений.

Характерные особенности ХД: относительно редкая корректируемость большинства данных, обновляемость данных на периодической основе, единый подход к именованию и хранению данных вне зависимости от их организации в исходных источниках.

ХД, являясь одним из главных звеньев архитектуры системы бизнес-аналитики любой средней или крупной организации, выступает в качестве основного источника данных для всестороннего анализа всей имеющейся в организации информации.

**Представление данных в витринах данных**

К четвертому уровню архитектуры системы бизнес-аналитики относятся источники данных, называемые витринами данных или *киосками данных* (*data marts*) и предназначенные для проведения целевого делового анализа. Витрины данных строятся, как правило, на основе информации из ХД, но могут также формироваться из данных, взятых непосредственно из транзакционных систем, когда ХД в организации по каким-либо причинам не реализовано.

По типу хранения информации витрины подразделяются на реляционные и многомерные. Витрины первого типа организуются в виде реляционной БД со схемой "звезда", где центральная таблица, таблица фактов, предназначенная в основном для хранения количественной информации, связана с таблицами-справочниками.

Многомерные витрины организуются в виде многомерных БД OLAP (Online Analytical Processing), где справочная информация представляется в виде измерений, а количественная — в виде показателей (метрик). Информация в многомерной витрине данных представляется в терминах бизнеса в виде, максимально доступном конечным пользователям, что позволяет существенно снизить время на получение требуемой для принятия решений информации.

С точки зрения пользователя, отличие витрин данных от ХД заключается в том, что ХД соответствует уровню всей организации, а каждая витрина обычно обслуживает уровень не выше отдельного подразделения и иногда может создаваться для индивидуального использования, отличаясь достаточно узкой целевой специализацией.

Отличие витрин данных от транзакционных БД заключается в том, что первые служат для удовлетворения потребностей конечных пользователей, не являющихся профессиональными программистами: аналитиков, менеджеров разных уровней, решающих различные задачи бизнеса. Транзакционные же БД используются в основном операторами, отвечающими за ввод и обработку первичной информации, а не за ее анализ, нацеленный на поддержку принятия решений.

Применение витрин данных, многомерных и реляционных, в сочетании с современными инструментами делового анализа данных позволяет превратить просто данные в полезную информацию, на основе которой можно принимать эффективные решения.

**Анализ данных**

К следующему уровню архитектуры системы бизнес-аналитики организации относятся современные программные средства, именуемые инструментами интеллектуального или делового анализа данных (Business Intelligence Tools), или BI-инструменты.

BI-инструменты позволяют управленческому звену организации проводить всесторонний анализ информации, помогают успешно ориентироваться в больших объемах данных, анализировать информацию, делать на основе анализа объективные выводы и принимать обоснованные решения, строить прогнозы, сводя риски принятия неверных решений к допустимому минимуму.

Инструменты *интеллектуального анализа данных* используются конечными пользователями для доступа к информации, ее визуализации, многомерного анализа и формирования как предопределенных по форме и составу, так и произвольных отчетов, создаваемых управленцем или аналитиком (без программиста). Как уже было сказано, в качестве входной информации для делового анализа выступают не столько "сырые" данные из транзакционных систем, сколько заранее обработанные данные из хранилища или представленные в витринах данных.

**Web-портал**

В настоящее время российские компании, вслед за западными коллегами, все активнее начинают внедрять у себя различные интернет-технологии. Уже сегодня все больше специалистов, работающих не только в сфере информационных технологий, начинают понимать выгоду от использования этих решений в целях повышения эффективности своего бизнеса. Проведение *интеллектуального анализа данных* с применением программных решений не только в локальной среде, но и в средах Интранет и Интернет, открывает аналитикам новые возможности работы с данными.

Современные тенденции развития архитектуры системы бизнес-аналитики базируются на применении интернет-технологий. Традиционный вид архитектуры системы бизнес-аналитики в недавнем прошлом дополнился web-порталом, постепенно приобретающим все более весомую роль в ее архитектуре. Возможность доступа к информации через привычный web-браузер позволяет экономить на затратах, связанных с закупкой и поддержкой настольных аналитических приложений для большого числа клиентских мест. Реализация web-портала позволяет снабжать аналитической информацией как пользователей внутри офиса, так и мобильных пользователей-аналитиков в любой точке мира, подключенных к порталу через Интернет.

**Реализация архитектуры системы бизнес-аналитики**

Сегодня на рынке информационных технологий представлен широкий спектр инструментальных средств, предназначенных для быстрой реализации компонентов архитектуры системы бизнес-аналитики. Применение таких инструментов позволяет не разрабатывать аналитические приложения заново, а воспользоваться готовыми современными технологиями и, следовательно, сократить время и затраты на их создание.

Решение задачи обеспечения пользователей информацией в системе бизнес-аналитики определяется в основном правильным подбором инструментов делового анализа. Но немаловажным является и выбор инструментов поддержки процессов извлечения, преобразования, загрузки и хранения данных.

При реализации системы бизнес-аналитики предприятия могут быть использованы программные решения как разных фирм-производителей (смешанные решения), так и одного производителя (платформенно-базированные решения). И в первом, и во втором случае имеются свои преимущества и недостатки. Поэтому выбор инструментов для реализации архитектуры системы бизнес-аналитики, несмотря на их многообразие, – задача не из простых.

На рынке не существует одного производителя, предлагающего лучшие решения всех требуемых для построения системы бизнес-аналитики программных компонентов. Поэтому совместное использование наиболее подходящих решений от различных производителей позволяет повысить функциональную мощность системы бизнес-аналитики. Критериями оценки инструментов могут выступать как их технические и стоимостные характеристики, так и скорость внедрения, а также уместность использования в каждом конкретном случае.

Однако использование продуктов от разных производителей приводит к значительному усложнению архитектуры системы из-за разнородности инструментальных решений. Это усложнение объясняется необходимостью интегрирования не связанных друг с другом инструментальных решений. Кроме того, администрирование системы оказывается непростой задачей, учитывая несогласованность данных и метаданных, управляемых отдельными, не связанными друг с другом модулями платформ от разных производителей.

При реализации архитектуры системы бизнес-аналитики от одного производителя (если пользоваться терминологией исследовательского центра Gartner, платформенно-базированное решение) решение необходимо искать среди фирм-производителей так называемых BI-платформ (Business Intelligence Platforms).

Данный сегмент рынка информационных технологий представлен более чем 20-ю компаниями, такими как (в алфавитном порядке): AlphaBlox, Arcplan, CA, Comshare, Crystal, Hyperion, Info Builders, Microsoft, Microstrategy, Oracle, PeopleSoft, ProClarity, Sagent, SAP, SAS, Whitelight и др. Среди них выделяются следующие семь лидеров и претендентов на лидерство в данной области: Microsoft, SAS, Oracle, SAP, PeopleSoft, Info Builders, Hyperion

Двое из перечисленных производителей, Microsoft и Oracle, в состоянии реализовать все уровни системы бизнес-аналитики своими силами, не прибегая к инструментам третьих фирм. Решающий критерий, выделяющий этих производителей, — наличие собственной СУБД.

Рассмотрим пример реализации системы бизнес-аналитики организации инструментами Microsoft.

тематике или программировании.

**Решение от компании Microsoft**

Компания Microsoft предлагает комплексный набор средств бизнес-анализа (*Business Intelligence*, BI) на основе масштабируемой платформы для организации ХД, анализа данных и генерации отчетов. Эти простые и мощные средства позволяют конечным пользователям обращаться к бизнес-информации и анализировать ее. Основой комплексного предложения для BI от Microsoft является *СУБД* *SQL* *Server* 2008 — 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, чтобы в итоге им смогли пользоваться все сотрудники.

В [табл. 4.3](https://intuit.ru/studies/courses/599/455/lecture/10161?page=4#table.4.3) приводится описание технологий *SQL* *Server* 2008, формирующих основу мощного BI-инструментария

|  |
| --- |
| Таблица 4.3. |
| **Компонент** | **Описание** |
| СУБД SQL Server | Масштабируемый высокопроизводительный механизм для хранения больших объемов данных. SQL Server подходит для консолидации всех бизнес-данных предприятия в центральном ХД для анализа и генерации отчетов |
| SQL Server Integration Services | Комплексная платформа для извлечения, преобразования и загрузки (*ETL*), обеспечивающая заполнение ХД и его синхронизацию с данными из разнородных источников, с которыми работают бизнес-приложения, используемые в организации |
| SQL Server Analysis Services | Аналитический механизм для реализации OLAP-решений (Online Analytical Processing, онлайновая аналитическая обработка): агрегирования бизнес-показателей из множества таблиц-измерений и создания решений для анализа данных (data mining), использующих специализированные алгоритмы, чтобы выявить шаблоны, тенденции и связи в бизнес-информации |
| SQL Server Reporting Services | Решение для генерации отчетов, облегчающее создание, публикацию и распространение подробных бизнес-отчетов по предприятию и за его пределами |

*SQL* *Server* 2008 не только является комплексной BI-платформой, но и тесно интегрирован с офисными решениями, такими как 2007 Microsoft Office *System*, что делает эту платформу доступной для всех сотрудников предприятия и позволяет им получить сведения, служащие основой для эффективных действий.

*SQL* *Server* 2008 поддерживает два типовых подхода к унификации бизнес-данных для анализа и генерации отчетов.

* **Хранилище данных** — специализированное хранилище для всех корпоративных данных, заполняемое из разнородных источников данных со всего предприятия и синхронизируемое с ними. Преимущество этого подхода состоит в возможности создания ХД, обеспечивающего оптимальную производительность при анализе и генерации отчетов и не влияющего на работу бизнес-приложений, послуживших источниками данных. Еще одно преимущество заключается в возможности проверки и консолидации данных из разных источников в единый согласованный массив информации, отражающий реальное состояние бизнеса.
* **Абстрагирование источников данных**. Средство SQL Server 2008 Analysis Services поддерживает создание представлений источников данных, обеспечивающих уровень абстракции для одного или нескольких источников данных. Созданное представление можно затем использовать как единый источник данных для Analysis Services, Integration Services и Reporting Services. Данные из соответствующих систем-источников загружаются в представления во время анализа или генерации отчета. Преимущество этого метода — возможность выполнять анализ данных бизнес-приложений в реальном времени. Кроме того, уровень абстракции в виде представлений источников данных заменяет длинные или запутанные имена таблиц именами, понятными человеку.

Чтобы обеспечить максимально высокую *производительность* и корректную работу, в *SQL* *Server* 2008 включены функции среды разработки, которые помогают создавать эффективные решения для анализа. К ним относятся:

* унифицированная среда для разработки всех BI-решений, включая Analysis Services, OLAP и приложения для анализа данных;
* встроенная поддержка полного жизненного цикла разработки, включая этапы проектирования, реализации, отладки и развертывания; поддержка групповой разработки посредством встроенных функций управления исходным кодом;
* встроенные интуитивно понятные дизайнеры и мастера, упрощающие создание решений на основе Analysis Services;
* дизайнер связей между атрибутами со встроенными средствами проверки на допустимость, позволяющий создать оптимальную структуру измерений;
* изрядно "похудевший" редактор измерений, обеспечивающий более высокую продуктивность и автоматически обнаруживающий связи "родитель – потомок";
* дизайнер кубов, также упрощенный и усовершенствованный, эффективнее обнаруживающий и классифицирующий атрибуты и идентифицирующий свойства членов;
* агрегирование отдельных разделов, позволяющее оптимизировать показатели по периодам и областям;
* дизайнер агрегатов с новым алгоритмом для создания первоначальных агрегатов. Дизайнер агрегатов оптимизирован для использования имеющихся агрегатов. Стало возможно просматривать существующие агрегаты, выполнять добавление или удаление из них. Обеспечивается интеллектуальная поддержка слияния новых и существующих агрегатов.

Генерация отчетов — важный элемент любого BI-решения; бизнес-пользователям требуются все более сложные отчеты. В *SQL* *Server* Reporting Services входит ряд средств, облегчающих создание решений для генерации отчетов:

* интерфейс разработки отчетов в Business Intelligence Development Studio (основанный на Visual Studio), в котором разработчики могут создавать, отлаживать и развертывать отчеты;
* Report Builder — средство разработки бизнес-отчетов, позволяющее бизнес-пользователям создавать и развертывать отчеты;
* широкий спектр структур для отображения данных, включая таблицы, матрицы, списки и диаграммы.

Кроме того, в *SQL* *Server* 2008 Reporting Services внесены существенные усовершенствования в плане повышения производительности и гибкости форматирования и публикации отчетов.

Преимущество *OLAP* состоит в том, что при моментальном доступе к точной информации *конечные* пользователи могут немедленно получать ответы даже на самые сложные вопросы. Поэтому при разработке всех версий *SQL* *Server* *Analysis* Services ставилась задача непрерывного сокращения времени обработки запросов и повышения скорости актуализации данных. Естественно, те же цели стояли и перед создателями *SQL* *Server* 2008 *Analysis* Services.

Средство *Analysis* Services в составе *SQL* *Server* 2008 предоставляет более широкие возможности в плане анализа, включая сложные вычисления и *агрегирование*. *Производительность* корпоративного уровня обеспечивается за счет:

* **гибкой модели кэширования**. В Analysis Services возможно управление кэшированием данных и агрегатов с целью оптимизации обработки запросов без превышения допустимой задержки передачи данных между кэшем и нижележащим хранилищем;
* **декларативных связей атрибутов**. В таблице-измерении Analysis Services можно явно объявлять связи между атрибутами в иерархии. Это позволяет Analysis Services заблаговременно рассчитывать агрегаты при обработке куба или измерения, что повышает производительность обработки запросов в реальном времени;
* **блочных вычислений**. Блочные вычисления устраняют излишние вычисления при расчете итогов (например, при агрегировании NULL-значений), значительно повышая производительность при анализе кубов и позволяя увеличивать сложность иерархии и вычислений;
* **обратной записи в MOLAP**. В Analysis Services 2008 снято требование запроса секций *ROLAP* при обратной записи, что дает огромный выигрыш в производительности;
* **масштабирования Analysis Services "вширь"**. Посредством виртуального IP-адреса можно открыть нескольким серверам Analysis Server доступ только для чтения к одной копии Analysis Services. Это позволяет создать решение для развертывания Analysis Services, обеспечивающее высокую масштабируемость;
* **сохранения плана выполнения запроса**. SQL Server 2008 поддерживает блокировку планов выполнения запросов, в результате планы "переживают" перезапуск, обновление и развертывание серверов (если это не мешает корректной работе). Это обеспечивает оптимальную производительность запросов к данным SQL Server.

Преимущество *SQL* *Server* 2008 на рынке BI-решений основано на масштабируемой инфраструктуре, благодаря которой *информационные технологии* делают возможным внедрение бизнеc-анализа по всему предприятию и *доступ* к результатам анализа там, где это необходимо пользователям. *SQL* *Server* 2008 обеспечивает значительный прогресс в организации хранилищ данных, предоставляя комплексную масштабируемую платформу, с помощью которой организации смогут быстрее интегрировать данные в ХД и управлять ими, доставляя результаты анализа всем пользователям. За счет более высокой масштабируемости BI-*инфраструктура* *SQL* *Server* 2008 способна генерировать отчеты любых размеров и сложности, управлять ими и делать отчеты доступными пользователям посредством тесной интеграции с Microsoft Office. Кроме того, *SQL* *Server* 2008 демонстрирует более высокую *производительность* в таких областях, как обслуживание ХД, генерация отчетов и *анализ*.

Общая *архитектура* решения для систем бизнес аналитики-компании Microsoft показана на [рис. 4.4](https://intuit.ru/studies/courses/599/455/lecture/10161?page=4#image.4.4).



**Рис. 4.4.**Решение для систем бизнес аналитики-компании Microsof

**Построение систем бизнес-аналитики: проблемы и решения**

*Информационные технологии* обеспечивают поддержку технологической цепочки обработки данных:

* сбор и получение данных;
* преобразование данных;
* предоставление данных.

Получение данных обеспечивается автоматизированными системами оперативной обработки данных или транзакционными системами обработки данных. Основное назначение таких систем – это обеспечение развитой формы учета данных на низком уровне бизнес-процессов организации. Пользователями этих систем являются специалисты.

Чтобы использовать собранные данные для анализа, их нужно привести к единому формату, преобразовать, согласовать и предварительно обработать. Эту задачу предназначены решать системы извлечения, преобразования и загрузки данных. Это важное звено перехода к анализу данных.

Предоставление данных обеспечивается *информационно-аналитическими системами* обработки данных. Такие системы разрабатываются с использованием технологии ХД и методов бизнес-аналитики. Основное назначение таких систем – это обеспечение развитой формы публикации данных. Их пользователями являются менеджеры.

Каждый *менеджер* обучался аналитической работе, применял *компьютер* при обучении в школе и университете, в повседневной работе окружен компьютерами и требует данных для *принятия решений*.

Публикация данных для менеджеров является первостепенной задачей. Хорошо известно, что публикация является успешной, если она удовлетворяет потребности читателей. Своевременная и по возможности полная публикация данных является средой для поддержки и *принятия решений*.

Для менеджера важно, чтобы публикация была:

* существенной для решения текущих бизнес-задач;
* понятной и простой в использовании;
* быстрой;
* эффективной в соотношении "цена/качество".

Рассмотрим комплекс проблем и пути их возможного решения, с которыми приходится сталкиваться при построении систем бизнес-аналитики1.

**Данные, необходимые для принятия решений, являются недоступными**

Первая проблема при создании систем бизнес-аналитики заключается в том, что в ХД оказываются недоступными данные, необходимые для принятия решений. Если в хранилище данных оказываются недоступными необходимые данные, нужно восполнить эту недостачу путем сбора бизнес-требований от конечных пользователей; изучения того, какая информация необходима бизнес-пользователям в процессе принятия решений; регулярных дискуссий с лицами, принимающими решения, для понимания новых требований; систематического исследования новых источников данных и метрик.

В связи с этой проблемой Ральф Кимбалл отмечает, что нельзя относиться к построению корпоративного ХД как к **проекту**, у которого имеется начало и конец. В действительности, построение ХД для системы бизнес-аналитики — это непрерывный процесс, который может закончиться только после отказа от построения ХД.

Отметим также, что на этот факт неоднократно указывали ряд исследователей в области построения ХД. Причиной такой точки зрения, скорее всего, является простое обстоятельство: бизнес-среда в современных экономических условиях может меняться очень быстро и динамично, что существенно влияет на потребности в данных.

**Недостаток партнерских отношений между конечными пользователями и ИТ-специалистами**

Вторая проблема при создании систем бизнес-аналитики заключается в недостатке партнерских отношений между конечными пользователями и специалистами в области ИT. Симптомами этой проблемы являются разочарование конечных пользователей имеющимся уровнем обслуживания; осуждение специалистами ИT конечных пользователей за их жалобы, компьютерную безграмотность и пренебрежение чтением документации; недооценка использования современных ИT руководством организации.

Как следствие, ХД не удовлетворяет потребности пользователей или работает слишком медленно, фактически, не используется пользователями. При этом отсутствуют административные решения, направленные на достижение согласия и исправление ситуации.

Общая идея решения этой проблемы: ИТ-персоналу необходимо жить в окружении бизнес-пользователей, чтобы лучше узнать специфику бизнеса компании и потребности ее заказчиков и завоевать доверие конечных пользователей.

Как показывает опыт, возникновение этой проблемы тесно связано с тем, что ИТ-специалисты при разработке автоматизированных систем не соблюдают требования соответствующих ГОСТов и не уделяют должного внимания разработке лингвистического и организационного обеспечения.

**Отсутствие ясности у конечных пользователей**

Третья проблема при создании систем бизнес-аналитики состоит в отсутствии явной познавательной и концептуальной модели конечных пользователей. Симптомом этой проблемы является выбор IT-специалистами инструментальных средств на основе бесед с потенциальными продавцами и знакомства с демонстрационными версиями без учета реальных потребностей пользователей.

IT-специалисты иной раз стремятся к сложным решениям и подразумевают, что конечным пользователям нравится работать на компьютерах. Но пользователи зарабатывают свои деньги за счет решения стоящих перед ними задач, и, возможно, рассматривают компьютер как средство, помогающее им решать эти задачи. Изучение и освоение новых программных продуктов не является их основной производственной задачей. До появления в организации новых программных продуктов бизнес-пользователи справлялись с решением своих задач и без них.

В качестве решения предлагается уточнение уровня познавательной и компьютерной грамотности конечных пользователей; построение концептуальной модели поведения пользователей при решении задач и принятии решений; выбор или настройка средств доставки информации, наилучшим образом соответствующих особенностям конечных пользователей.

Самый простой подход состоит в том, чтобы разделить пользователей на две категории – те, которые используют Excel, и те, которые считают электронные таблицы слишком сложными. Для первой категории нужно обеспечить возможность формулировки произвольных запросов, а вторым предоставить заранее подготовленные, может быть, параметризуемые отчеты.

Ральф Кимбалл предлагает простую модель оценки сложности программных инструментов:



**Рис. 4.5.**Модель использования ХД в системах бизнес-аналитики для принятия решений

Правило применения этой модели очень просто. Оно исходит из двух логических предпосылок: "Каждое нажатие — это подцель при достижении цели" и "Каждое нажатие – это отвлечение, как неожиданный звонок телефона". Отсюда вытекает эмпирическое правило: "1-3 нажатия – хорошо; 4-8 нажатий – приемлемо; больше 8 нажатий – провал".

На [рис. 4.5](https://intuit.ru/studies/courses/599/455/lecture/10161?page=5#image.4.5) показана простая модель использования ХД в системах бизнес-аналитики для принятия решений.

Как видно из рисунка, модель включает в себя отражение следующих бизнес-процессов принятия решений:

* публикация "правильных" данных;
* сравнение, определение пороговых значений, предупреждение и визуализация для идентификации;
* исследование и поиск причинно-следственных связей;
* выдвижение гипотез и исследование альтернатив по схеме "Что будет, если…";
* аудит и отслеживание принимаемых решений.

**Данные, необходимые для принятия решения, поступают с задержкой**

Четвертая проблема при создании систем бизнес-аналитики заключается в запаздывании данных, требуемых для принятия решений. Симптомом является потребность в данных в реальном времени. Здесь под требованиями "реального времени" понимаются любые требования к временным характеристикам данных, которые не могут быть удовлетворены действующей процедурой *ETL*.

Одно из возможных решений заключается в изменении процедуры *ETL* (Extraction, Transformation, Loading) за счет использования готовых инструментов извлечения данных, например, сообщений *EAI* (Enterprise *Application Integration*). Для быстрого удовлетворения потребностей пользователей можно связывать "горячие" разделы таблицы фактов со статическим ХД, не дожидаясь обновления таблиц измерений.

**Несогласованные измерения и факты. Победите интеграцию**

Пятая проблема при создании систем бизнес-аналитики состоит в том, что интеграции корпоративных данных препятствуют не сведенные к единой форме факты и измерения. Топ-менеджерам требуется всестороннее представление данных, а его невозможно получить, потому что в разных подразделениях данные представляются по-разному. В качестве решения предлагается при проектировании витрин данных использовать шинную матрицу для согласования данных. Как подчеркивает Ральф Кимбалл, это решение является не столько техническим, сколько организационным.

**Недостаточно подробные данные: невыразительная система бизнес-аналитики**

Шестой проблемой при создании систем бизнес-аналитики является недостаточная подробность (*гранулированность) данных*, результатом чего становится невыразительная система бизнес-аналитики. Симптомом является недостаточное число атрибутов у данных измерений. Предлагается постоянно стремиться к повышению выразительности данных, а для создания содержательного контекста данных использовать вспомогательные источники данных.

**Данные в неудобных форматах**

Седьмую проблему при создании систем бизнес-аналитики представляют неудобные форматы данных. По Ральфу Кимбаллу, неудобной является *нормализованная форма* реляционных данных. Симптомами проблемы, кроме этого, могут быть запутанность и запуганность пользователей, сложность формулировки запросов, сложные процедуры *ETL*, потребность в специализированном оборудовании для достижения требуемой производительности.

Одним из возможных решений является представление данных в многомерной модели. Это представление соответствует пользовательской интуиции, облегчает формулировку запросов, упрощает процедуру *ETL* и позволяет добиться нужного уровня производительности на обычной аппаратуре.

**Медленная, не адаптированная для пользователей доставка данных**

Восьмая проблема при создании систем бизнес-аналитики состоит в слишком медленной доставке данных конечным пользователям. Данные не поступают в оперативном режиме, пользователи остерегаются задавать медленно выполняемые запросы, имеются количественные ограничения на использование данных.

Решением этой проблемы является тщательное проектирование БД, создание многомерных моделей данных, подбор качественных программных средств СУБД с развитыми механизмами индексации, оснащение компьютеров основной памятью большого объема, использование распараллеливания, применение компьютеров с быстрыми центральными процессорами.

**Данные закрыты в отчете или информационной панели**

Девятая проблема при создании систем бизнес-аналитики проявляется в том, что некоторые данные оказываются "запертыми" в каком-то приложении и их невозможно переместить оттуда в другое приложение простым образом. Выходом из положения является использование только таких приложений, для которых возможно копирование данных в электронную таблицу через буфер обмена с помощью одного нажатия на клавишу мыши.

**Низкое качество данных**

Десятая проблема при создании систем бизнес-аналитики связана с низким качеством данных. Симптомами проблемы являются отсутствие содержательных данных, наличие ненадежных или бессмысленных данных, присутствие дублирующих или несогласованных записей (чаще всего такие записи относятся к заказчикам компании). В качестве решения проблемы предлагается расширить используемые средства *ETL* системой экранов качества данных. В многомерной модели данных для фиксации ошибок в данных создается схемы событий с ошибками (Error Event Schema) — таблица фактов со своими измерениями. На основе этой таблицы порождаются измерения аудита данных для других таблиц фактов, и эти измерения могут использоваться при формировании отчетов, в которых учитываются ненадежные данные.

**Преждевременно агрегированные данные**

Одиннадцатая проблема при создании систем бизнес-аналитики состоит в преждевременной агрегации данных. Наличие в многомерной модели агрегированных данных без соответствующих атомарных данных не позволяет проводить детализацию данных. Рекомендуемым решением проблемы является поддержка для витрин данных физических структур хранения, содержащих атомарные данные. Детализация данных поддерживается за счет агрегатной навигации [[2](https://intuit.ru/studies/courses/599/455/literature#literature.2)].

**Отвлечение и фокусировка на показателе ROI для ХД**

Двенадцатой проблемой при создании систем бизнес-аналитики Ральф Кимбалл считает отвлечение внимания на оценку показателей возврата инвестиций (ROI) ХД. Симптомами этой проблемы является расчет показателей ROI до создания ХД с применением стандартных методов, основанных на периоде окупаемости, *чистой приведенной стоимости*, внутренней норме прибыли, системе сбалансированных показателей, экономической добавленной стоимости. По его мнению, во всех этих методах упускается основной смысл стоимости и в конечном счете – ценности ХД.

ХД поддерживает принятие решений. Рекомендуется после принятия решения отнести часть полученной прибыли на счет ХД, а затем сравнить ее с расходами на ХД. Ральф Кимбалл рекомендует считать, что 20% прибыли, полученной в результате принятия решения, получено благодаря использованию ХД. Такой подход соответствует той идее, что единственным осмысленным способом оценки эффективности ХД является оценка его возможности поддерживать принятие решений конечными пользователями.

**Отвлечение на создание корпоративной модели данных**

Тринадцатая проблема при создании систем бизнес-аналитики состоит в затрате сил и времени на создание *корпоративной модели данных*. Симптомом является появление большого количества сущностей, которые никогда не наполняются реальными данными. Ральф Кимбалл считает, что усилия, затрачиваемые на разработку *корпоративной модели данных*, только задерживают работу над ХД, и расчет делается на то, что при выполнении процедуры *ETL* будут выявляться ошибки и несоответствие данных.

Отметим, что решение о разработке *корпоративной модели данных* действительно требует больших интеллектуальных затрат и времени на ее создание. Может оказаться, что модель устареет к моменту ввода ее в эксплуатацию. Исследователями по ИТ предлагаются различные подходы к созданию актуальной корпоративной модели.

**Мандат на получение всех данных**

### Четырнадцатой проблемой при создании систем бизнес-аналитики Ральф Кимбалл считает возможное требование использовать все источники данных для наполнения ХД. В соответствии с его опытом если при построении ХД выдвигается требование использования трех или более источников данных, то ХД не заработает и через два года. На первом этапе построения ХД рекомендуется потратить шесть недель на полноценный аудит данных, а затем выбрать один источник данных, который, во-первых, влияет на наиболее важные решения конечных пользователей, и во-вторых, проще всего подключается к процедуре *ETL*. После з**Хранилища данных и системы бизнес-аналитики**

Одной из главных предназначений системы бизнес-аналитики является публикация данных в формате, удобном для *принятия решений*, и предоставление простых и удобных инструментов руководителям организации для манипулирования этими данными с целью их исследования и анализа.

Одно из главных предназначений ХД — аккумулирование данных из различных источников для аналитической обработки и отделение формирования отчетов и проведения *интеллектуального анализа данных* от систем оперативной обработки данных предприятия с целью увеличения производительности.

Таким образом, ХД представляют собой очень большие репозитории исторических данных, а системы бизнес-аналитики представляют собой взаимосвязанный (или нет) набор приложений для бизнес-анализа этих данных.

Ниже, на [рис. 4.6](https://intuit.ru/studies/courses/599/455/lecture/10161?page=6#image.4.6), показано, как система бизнес-аналитики взаимодействует с ХД.



**Рис. 4.6.**Взаимодействие системы бизнес-аналитики с хранилищем данных

Разработчики систем бизнес-аналитики являются по существу издателями. Они собирают данные из различных источников, редактируют их для обеспечения качества и согласованности, обеспечивают *доверие* к опубликованной информации. Их успех оценивается конечными бизнес-пользователями: аналитиками, менеджерами различных уровней и руководством организации.

ХД поддерживает системы бизнес-аналитики, являясь ее информационным фундаментом.

* Запросы и отчетность:
	+ Публикация правильных данных.
* Поиск исключений:
	+ Визуализация, определение границ, сравнение, предупреждение.
* Анализ причинно-следственных связей:
	+ Что взаимосвязано.
* Моделирование потенциальных решений:
	+ Что, если...
* Отслеживание результатов принятых решений.

С этой точки зрения ХД помогает решать основные задачи поддержки систем бизнес-аналитики.

* Максимизация ценности интеллектуального капитала:
	+ применение знаний в определенной предметной области;
	+ демонстрация интуиции в простых действиях;
	+ принятие решений;
	+ достижение понимания среди всех лиц, принимающих решения;
* и одновременная минимизация затрат:
	+ разработка;
	+ администрирование (стандартные операции, маленькие сюрпризы, большие сюрпризы);
	+ очевидные затраты (персонал, аппаратное и программное обеспечение);
	+ скрытые затраты (упущенные возможности, отклонения).

На [рис. 4.7](https://intuit.ru/studies/courses/599/455/lecture/10161?page=6#image.4.7) показано, как хранилища данных управляют системами бизнес-аналитики.



**Рис. 4.7.**Как хранилища данных управляют системами бизнес-аналитики

На практике ХД функционируют в соответствии с нижеприведенными условиями.

* Децентрализованная, инкрементальная разработка.
* Различные технологии, не совместимые на простейшем уровне.
* Быстрая разработка.
* Постоянно меняющаяся среда/постоянно меняющиеся приоритеты.
* Потенциально несовместимые витрины данных.
* Немедленная реакция.
* Атомарные данные, данные в режиме реального времени, непрерывная история.
* Всестороннее представление о заказчике.
* Отслеживание, хранение, предсказание поведения.

Таким образом, ХД через хранимые в них данные управляют системами бизнес-аналитики и влияют на качество и эффективность принимаемых решений. Чтобы данные способствовали принятию качественных решений, они должны быть хорошо организованы. Организацию данных в ХД обеспечивает *модель данных*. Разработке таких моделей данных будут посвящены следующие лекции, а сейчас кратко подведем итоги настоящей лекции.

**Резюме**

Сфера систем бизнес-аналитики развивается быстро и динамично. Это объясняется тем, что в современных условиях *информация* становится реальным производственным ресурсом. К настоящему времени можно выделить следующие основные типы систем бизнес-аналитики.

* **Аналитическая и управленческая отчетность**. Наиболее распространенные, универсальные и вместе с тем эффективные системы для получения информации на различных уровнях управления компанией. В отличие от стандартных систем отчетности включают в себя богатые возможности для построения запросов, создания отчетов, визуализации данных и простой обработки полученных результатов непосредственно менеджерами и аналитиками, не являющимися специалистами в области ИТ.
* **Оперативный анализ информации OLAP**. OLAP-системы предназначены для менеджеров и аналитиков, которым требуется постоянное интерактивное взаимодействие с информацией.
* **Информационные панели. KPI/BSC**. Предназначены для отображения и мониторинга ключевых показателей работы компании. Как правило, используются в системах управления предприятием на основе KPI/BSC.
* **Системы нетривиального анализа данных и получения знаний**. В основе лежат технологии data mining, которые могут использоваться для решения сложно формализуемых задач.

Системы бизнес-аналитики и хранилища данных обеспечивают полноту, *достоверность* и актуальность информации, необходимой для принятия управленческих решений, снижают нагрузку на транзакционные системы за счет перераспределения функций формирования отчетности.

Системы бизнес-аналитики дают возможность решать *целый* ряд актуальных для современного предприятия задач:

* консолидировать информацию из разнородных источников (внутренних систем оперативного учета данных, внешних источников) в хранилище данных, с предварительной очисткой, преобразованием данных и приведением информации к общей *корпоративной модели данных*;
* рассчитывать требуемые показатели и статистические характеристики на основании ретроспективной информации из хранилища данных; определять взаимосвязи показателей (производить проверку статистических гипотез, кластеризацию и пр.);
* формировать наглядные графические и табличные представления результатов вычислений и имеющейся информации (визуализация данных);
* получать в автоматическом режиме предварительно заданные виды отчетности, формировать произвольные отчеты на основе созданной модели данных;
* проводить эксперименты с математическими моделями, описывающими поведение сложных систем (задачи класса "Что, если ...?"), что позволяет оценивать обоснованность и эффективность для бизнеса тех или иных шагов. Применять имитационные, управленческие, оптимизационные и статистические методы моделирования и прогнозирования.