

Системы бизнес-аналитики

К.ф.-м.н Нурутдинова Алсу Рафаиловна

Лекции 1-3

О курсе

- Учимся любить данные и работать с ними
- Python
- SAS

Система оценивания

Теория

Контрольная работа

Практика:

Практические задания

Зачет

Допуск к зачету: работа на практике(30б) +контрольная работа (20б)

Зачет(Тестирование+задание) - 50 баллов

95% рабочего времени аналитик крупной компании тратит на сбор данных и подготовку типового отчета.

Менее 1% времени аналитик занимается анализом

Минимум 20% рабочего времени посвящать анализу

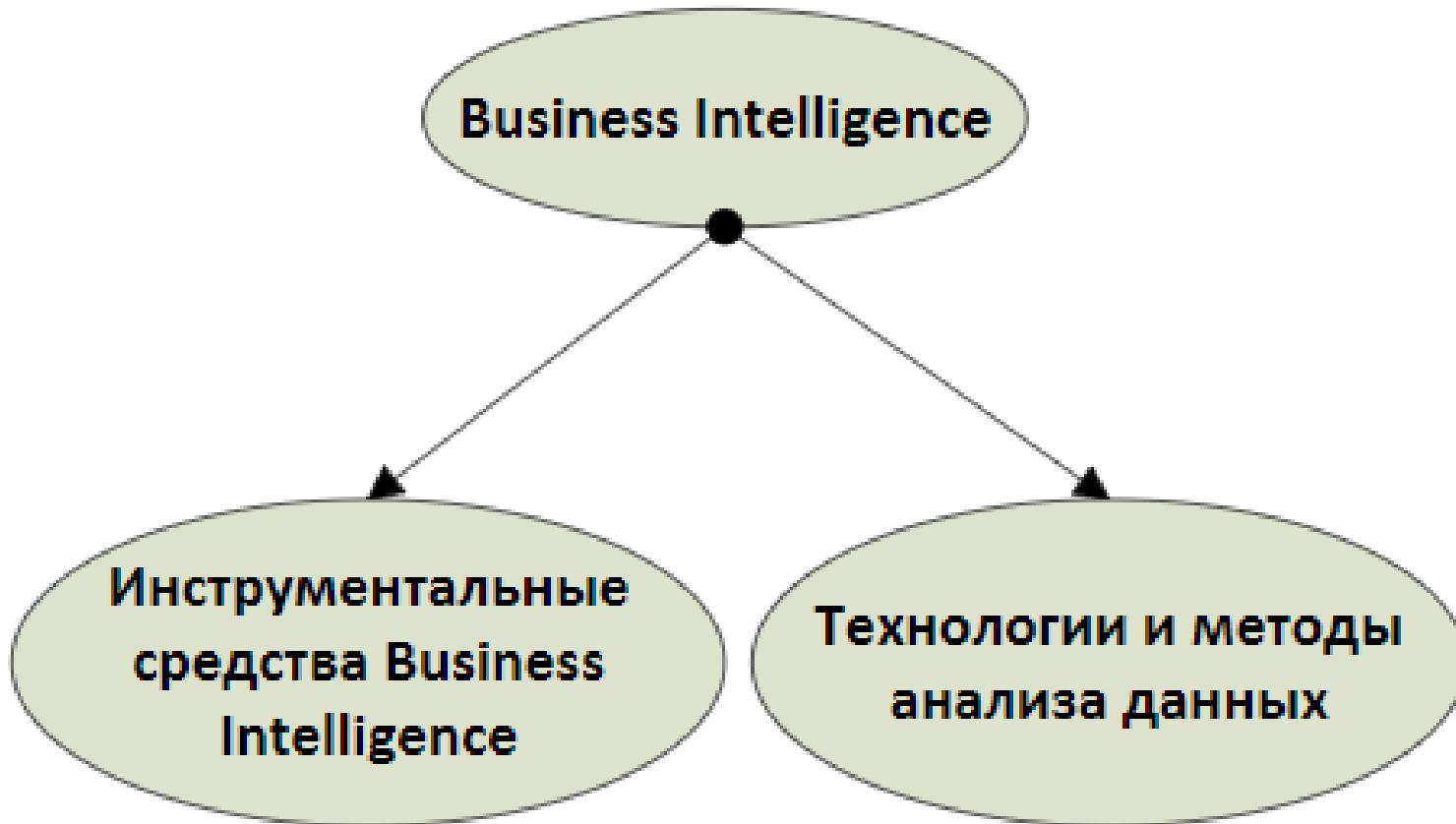


В процессе аналитической деятельности нужно создавать что-то новое



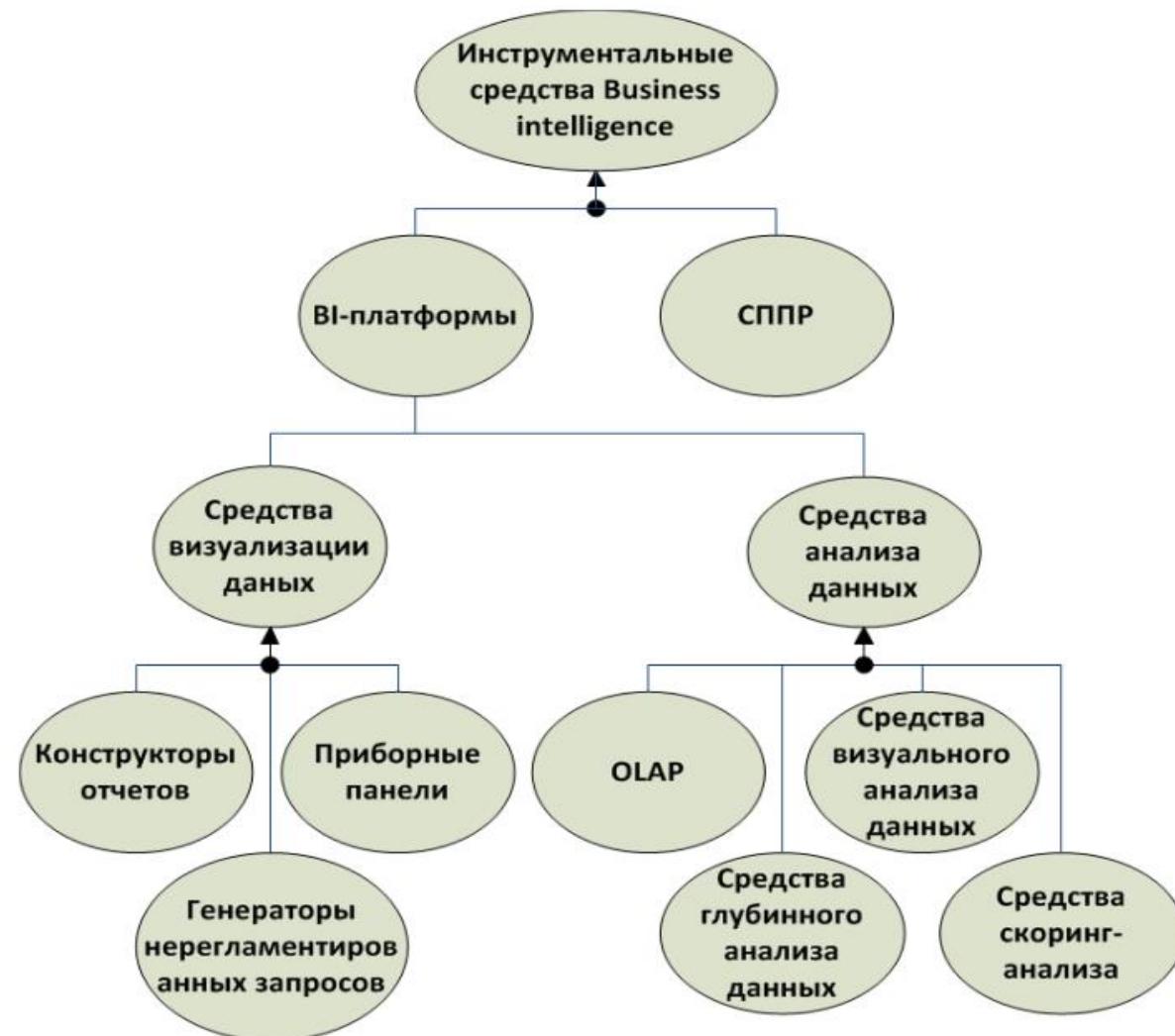
Создаваемое новое должно приносить реальную пользу

- Статистический анализ данных (Statistical Data Analysis)
- Искусственный интеллект (Artificial Intelligence) 1955
- Распознавание образов (Pattern Recognition)
- Машинное обучение (Machine Learning) 1959
- Статистическое обучение (Statistical Learning)
- Интеллектуальный анализ данных (Data Mining) 1989
- Машинный интеллект (Machine Intelligence) 2000
- Бизнес-аналитика (Business Intelligence, Business Analytics)
- Предсказательная аналитика (Predictive Analytics) 2007
- Большие данные (Big Data) 2008
- Аналитика больших данных (Big Data Analytics)
- Наука о данных (Data Science) 2011



Таким образом, под термином Business Intelligence будем понимать и комплекс инструментальных средств анализа больших массивов данных с целью поддержки принятия эффективных решений, и область знаний, изучающую технологии и методы анализа данных с целью поддержки принятия эффективных решений

В современном состоянии термин Business Intelligence понимается и как комплекс аналитического инструментария, и как область знаний, изучающая вопросы применения различных методов анализа данных.



- <https://compscicenter.ru/videos/intelligence/>

Лекция 1. Введение в бизнес-аналитику

- Информация
- Данные
- Знания
- Управление знаниями
- Информационные системы

«Кто владеет информацией, тот владеет миром» ©

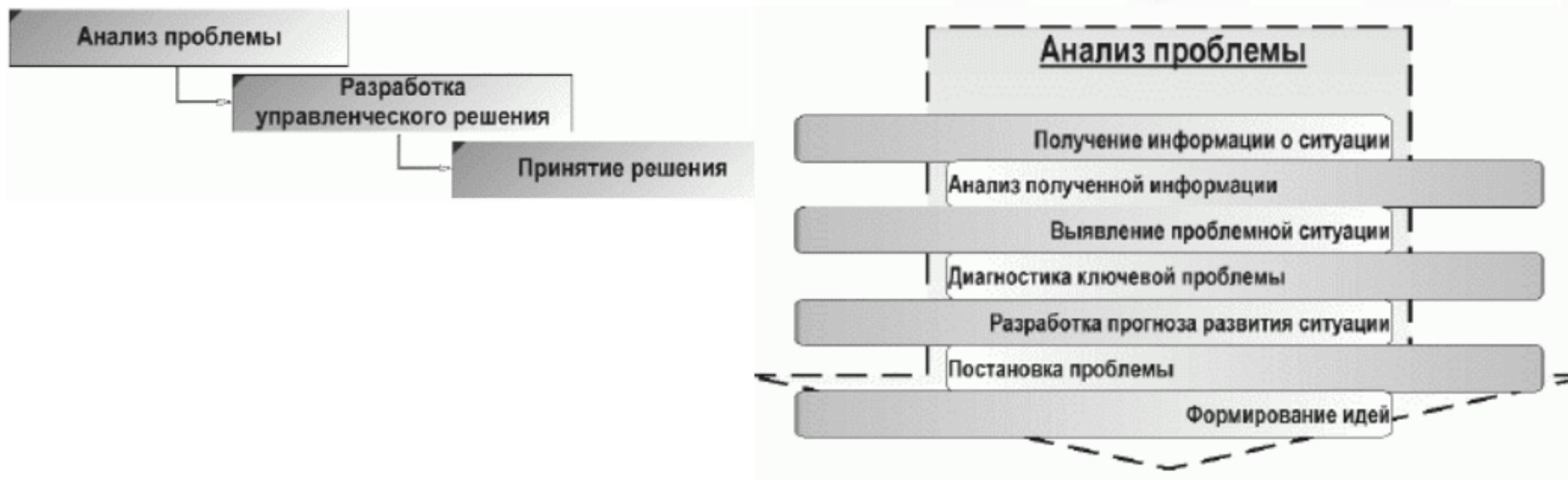
Успех компании на рынке напрямую зависит от того, как быстро менеджмент компании может распознать изменения динамики рынка и насколько своевременно может отреагировать на них с целью увеличения прибыли, исходя из существующих реалий рынка. Менеджеры компании должны отслеживать тенденции рынка, идентифицировать конкурентов и угрозы, оценивать риски, преобразовывать стратегию компании, оценивать свои ресурсы и т.д.

Ни одна компания более не может длительно оставаться конкурентоспособной на рынке и оперативно подстраиваться под переменчивую конъюнктуру без доступа к мощному инструменту для анализа данных.

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.

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, разрознены и в общем случае не готовы для анализа.

Возникает потребность в информационных системах аналитического анализа данных, способных 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.



Цифровая экономика — деятельность по созданию, распространению и использованию цифровых технологий и связанных с ними продуктов и услуг;

Цифровые технологии — технологии сбора, хранения, обработки, поиска, передачи и представления данных в электронном виде.

«Сквозные» цифровые технологии — технологии, применяемые для сбора, хранения, обработки, поиска, передачи и представления данных в электронном виде, в основе функционирования которых лежат программные и аппаратные средства и системы, востребованные во всех секторах экономики, создающие новые рынки и изменяющие бизнес-процессы.

Рис. 2. Переход к новым бизнес-моделям в условиях цифровой экономики

Классическая бизнес-модель vs Цифровая бизнес-модель

Классическая бизнес-модель



Цифровая бизнес-модель



Источник: [TPA Global, 2018].

ВСПОМИНАЕМ ОСНОВНЫЕ ТЕРМИНЫ И ТЕХНОЛОГИИ ИТ

Информация — сведения, воспринимаемые человеком или специальными устройствами как отражение фактов материального мира в процессе коммуникации (ГОСТ 7.0-99)

- знания о предметах, фактах, идеях и т. д., которыми могут обмениваться люди в рамках конкретного контекста (ISO/IEC 10746-2:1996)
- знания относительно фактов, событий, вещей, идей и понятий, которые в определённом контексте имеют конкретный смысл (ISO/IEC 2382-1:1993).

Первоначально «информация» — сведения, передаваемые людьми устным, письменным или другим способом (с помощью условных сигналов, технических средств и т. д.); с середины XX в. термин «информация» превратился в общенаучное понятие, включающее обмен сведениями между людьми, человеком и автоматом, автоматом и автоматом; обмен сигналами в животном и растительном мире; передачу признаков от клетки к клетке, от организма к организму (например, генетическая информация) и т.д.

Информация должна обрести некоторую форму представления (то есть превратиться в **данные**), чтобы ей можно было обмениваться.

Предметом изучения *информатики* являются именно **данные**: методы их создания, хранения, обработки и передачи.

К информации предъявляются следующие *требования*:

- достоверность;
- полнота;
- полезность;
- своевременность;
- релевантность (существенность).

Ценность информации является величиной, имеющей различные значения для различных людей, в соответствии с их способностью понять информацию и использовать ее в дальнейшем в различных областях своей деятельности. Информация ценна, поскольку она способствует достижению поставленной цели. Информация может действительно иметь различную ценность, если критерием оценки избрать ее использование для различных целей.

Одной из важнейших разновидностей информации является **информация экономическая**. Её отличительная черта - связь с процессами управления коллективами людей, организацией. Экономическая информация сопровождает процессы производства, распределения, обмена и потребления материальных благ и услуг. Значительная часть её связана с общественным производством и может быть названа производственной информацией.

Информация относительно независима от ее носителя, так как возможны ее преобразования и передача по различным физическим средам разными физическими сигналами.

Понятие информации предполагает наличие двух объектов: источника информации и потребителя; важно, чтобы информация для потребителя имела смысл, чтобы он мог оценивать ее для своих целей. Характерным носителем информации является сообщение, - все то, что подлежит передаче.

Верхним уровнем информации являются знания.

Наряду с энергией, полезными ископаемыми и т.д. информация является ресурсом общества. По мере продвижения технологического прогресса информационный ресурс становится наиболее важным национальным ресурсом. Эффективность промышленной эксплуатации информационных ресурсов определяет экономическую мощь страны.

Сейчас количество информации, поступающее в промышленность, управление, науку доходит до тревожных значений. Время удвоения объема информации, накопленных научных знаний - 2-3 года. В таком информационном состоянии общества очень важна эффективная эксплуатация информационных ресурсов.

Экономическая информация - это совокупность сведений о реальных социально-экономических процессах в обществе, которая служит основой для изучения и управления этими процессами и людьми, объединенными в социальные и производственные организации. Для экономической информации характерны следующие *параметры*:

- многообразие источников и потребителей;
- нарастающие значительные объемы и большой удельный вес рутинных процедур при их обработке;
- многократное повторение циклов получения и отправки в установленные временные периоды (декада, месяц, квартал, год);
- необходимость обеспечения конфиденциальности в использовании отдельных ее частей;
- чрезвычайная важность в подготовке и принятии деловых решений.

Информация осознана современным обществом как необходимое условие прогресса. Наибольший экономический успех сегодня сопутствует тем предприятиям, которые активно используют современные средства коммуникаций, информационные технологии и их приложения. Новые информационные технологии и связанные с ними прикладные задачи требуют создания новой среды - информационной, а также систем управления информационными ресурсами.

Любое управленческое решение имеет поставщика или источник информации, адресата решения и последствия. Есть три главных источника информации для управленческих решений - это:

- *Внешний мир* (клиенты, партнеры, конкуренты, государство), который трудно контролировать, - поставляет информацию, какую хочет и когда хочет. Информация эта может быть неполной, недостоверной или просто ложной.
- *Предприятие* (руководители всех уровней, исполнители). Эту информацию как будто можно контролировать и управлять ею. Однако люди имеют разные должности, образование и опыт, разные взгляды на жизнь и разные цели. Естественно, что информация, поставляемая ими, субъективна - она требует проверки и уточнения.
- *Руководитель* - его интеллект, опыт, интуиция, возможности, связи. Он аккумулирует, сортирует и анализирует информацию, отбирая ту, которая, как кажется ему, поможет принять нужное решение.

"Как работают с информацией в компании и какая информация нужна для принятия решения?"



Руководителю нужна совокупная, достоверная, взвешенная информация (собранные и обработанные данные), распределенная по основным направлениям финансово-экономической и производственной деятельности компании, пригодная для всестороннего анализа и достаточная для принятия решения.

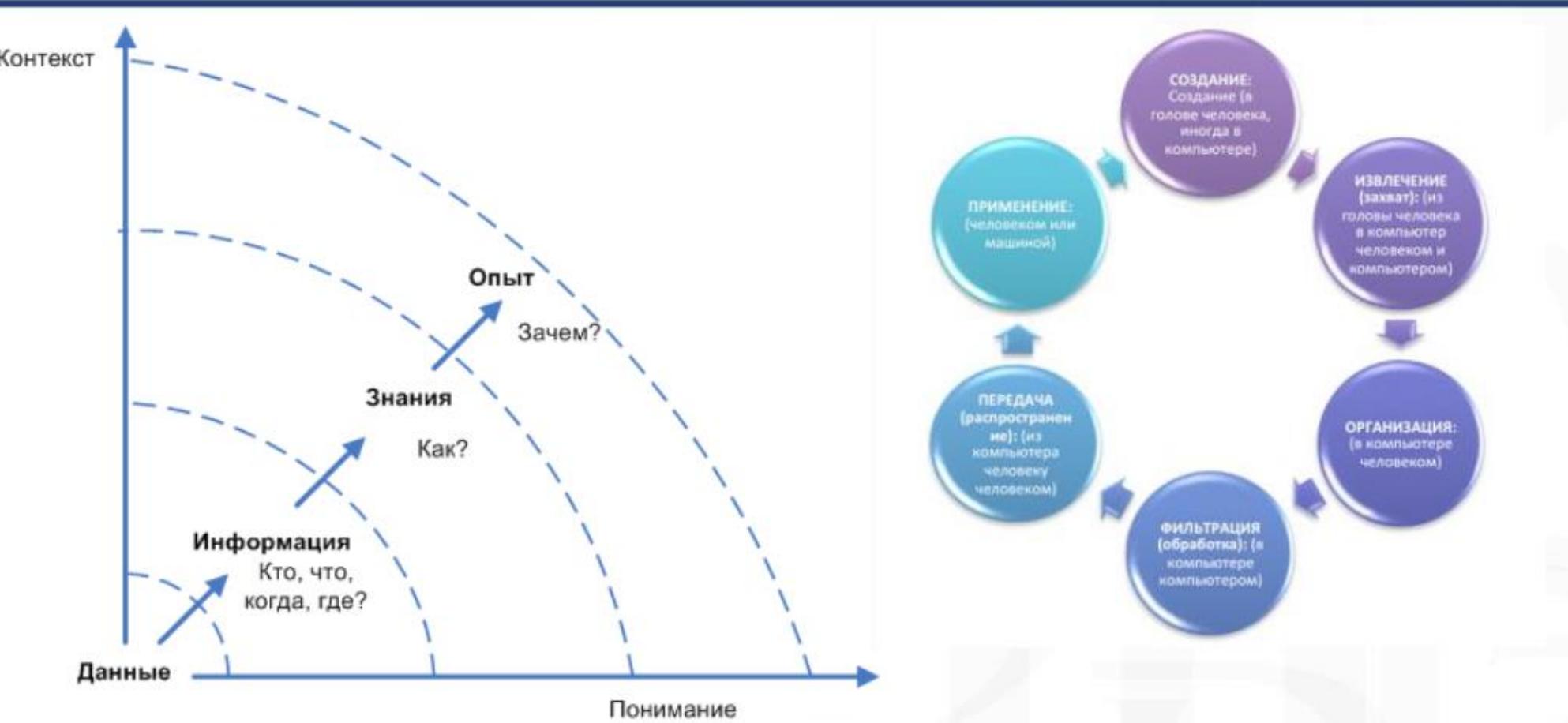
Данные - набор дискретных фактов о событиях. В рамках Управления знаниями над ними выполняются следующие действия:

- сбор точных данных;
- анализ, систематизация и преобразование в информацию;
- определение наиболее значимых данных и концентрация ресурсов на их сборе.

Информация - форматированные данные. Информация, как правило, хранится в слабоструктурированных формах - документах, e-mail, файлах. Управление знаниями предназначено для того, чтобы информацию можно было легко запросить, найти и использовать. Это необходимо для исключения ошибок, облегчения поиска и предотвращения избыточных работ.

Знания - комплект накопленных взглядов, опыта, идей, ценностей и суждений. Люди получают знания, опираясь на свой собственный опыт и на опыт других людей, а также путем анализа информации, поступающей из различных источников. Знания предоставляют и структурируют информацию таким образом, чтобы ее можно было легко использовать для принятия решений, что означает использование опыта предыдущих внедрений.

Мудрость - предоставляет ультимативный процесс понимания материального и способность принимать решения, основываясь на логических рассуждениях.



Знания рассматриваются в рамках модели Данные-Информация-Знания-Мудрость (Data-to-Information-to-Knowledge-to-Wisdom или DIKW) - способ представления взаимосвязей между данными, информацией, знанием и мудростью.

Знания не только представляют собой самостоятельную ценность, но и порождают мультипликативный эффект по отношению к другим факторам производства, воздействуя на уровень эффективности их применения. Таким образом, в современной экономике источником конкурентных преимуществ становится не выгодная рыночная позиция, а сложные для репликации знания как активы и способ их размещения. Причем в центре внимания здесь находится не создание знаний, а их движение и использование в организации.

Информация и знания, составляющие основу интеллектуального капитала, обладают рядом специфических характеристик в отличие от денежных, природных, трудовых и технических ресурсов организации:

- ценность знаний заключается в их изобилии, в то время как прочие ресурсы оцениваются исходя из понятия редкости;
- в структуре себестоимости «материализованного знания» (наукоемких товаров и услуг) преобладает тенденция к накоплению издержек на начальной стадии производства;
- между затратами знаний на входе и объемом знаний на выходе нет значимого экономического соответствия.

Знания это необходимая информация, используемая по определённым правилам и в соответствии с определенными процедурами и с учетом отношения (понимание, одобрение, игнорирование, согласие, отрицание и т. д.) использующих субъектов к этой информации. Сегодня организационные знания рассматриваются одновременно как информационный запас и как поток (движение этой информации) одновременно.

«Знания — это сочетание оформленного опыта, ценностей, контекстной информации и взглядов эксперта, которая дает схему для оценки и объединения нового опыта и информации. В организациях они зачастую попадают не только в документы или хранилища, но и в организационные процедуры, процессы, практику и нормы» [Давенпорт и Прусак. Рабочее знание: Как организации управляют тем, что они знают].

Сpirаль знаний — модель, объясняющая как при создании новых знаний явные и неявные знания взаимодействуют в организации благодаря четырем процессам их преобразования:

- социализации (превращению неявных знаний в неявные);
- комбинации (обращению явных знаний в явные);
- экстернализации (превращению неявных знаний в явные);
- интернализации (превращению явных знаний в неявные) [Икуджио Нонака].

В процессе социализации происходит невербальная передача скрытого знания от одного члена организации к другому, например, с помощью наблюдения одного человека за другим. Экстернализация представляет собой процесс превращения скрытого знания в явное при помощи необычного использования языка, различных метафор и аналогий. Комбинирование — это передача явных, кодифицированных знаний от одного человека другому при помощи книг, газет, лекций, компьютерных технологий, а интернализация — превращение явного знания в скрытую форму, например, посредством практического выполнения какой-то деятельности.

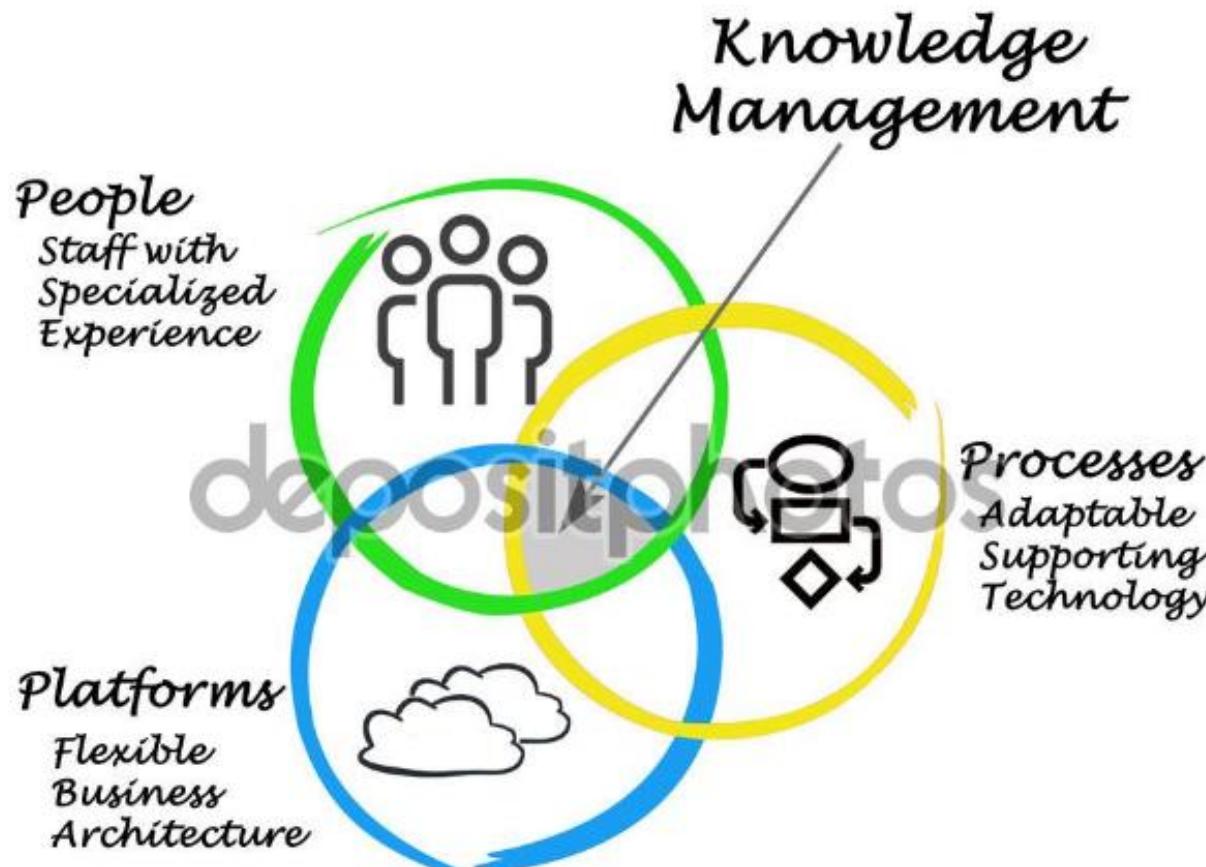
Для каждой фирмы характерна своя модель управления знаниями, учитывающая специфику деятельности, масштабы производства, организационные особенности, корпоративную культуру компании. Однако, вне зависимости от направления движения информационных потоков, менеджмент знаний должен обеспечивать контроль за осуществлением в организации следующих процессов:

- создание новых знаний;
- использование имеющихся знаний при принятии решений;
- воплощение знаний в продуктах и услугах;
- передача существующих знаний из одной части организации в другую;
- обеспечение доступа к необходимым знаниям, а также защита знаний.

«Технология делает знания единственным источником стойкого конкурентного преимущества.

Все остальное выпадает из уравнения конкуренции». *Лестер Тероу*

Управление знаниями (Knowledge Management) - целенаправленная организация деятельности всей компании, где знания рассматриваются, как главный стратегический фактор успеха; процесс, отвечающий за сбор, анализ, сохранение и предоставление знаний и информации в Организации. Первичная цель Управления знаниями - увеличение эффективности путем снижения необходимости в повторном поиске знаний.



Процесс управления знаниями отвечает за то, чтобы нужная информация поступала к компетентному лицу своевременно для поддержки в принятии решений.

Задачи процесса:

- повышение результативности поставщика услуг, качества услуг и удовлетворенности заказчиков, а также снижение затрат;
- обеспечение понимания персоналом ценности предоставляемых заказчикам услуг;
- обеспечение своевременного доступа персонала к следующей информации:
 - кто в настоящее время использует услуги;
 - текущие уровни потребления;
 - ограничения для предоставления услуги;
 - трудности, с которыми сталкиваются заказчики.



Общая схема коллективного управления знаниями



Наиболее характерной чертой современного управления организацией стали постоянные изменения, обусловленные, в свою очередь, изменениями внешней среды, которая стала чрезвычайно динамичной. Например: давление высоких технологий, культурная и социальная мобильность общества и т.п.

Ранее неоспоримые достоинства процветающих предприятий — устойчивая организационная структура и устоявшиеся процессы — ныне обрачиваются недостатком, поскольку чрезсчур инерционная организация не позволяет своевременно реагировать на изменение требований рынка. Управление знаниями может рассматриваться как мощное конкурентное преимущество только в фирме, ориентированной на постоянное развитие бизнеса и бизнес-процессов.

Кроме того, решения по управлению знаниями становятся тем актуальнее, чем меньше бизнес-процессы и люди привязаны к структуре организации и географическим границам. И только посредством сравнения корпорации осознают, что управление знаниями обеспечивает адаптацию, выживание и сохранение компетенций на фоне постоянно растущих глобальных изменений

Важность для предприятия задач управления знаниями определяется уровнем его управленческой зрелости и уровнем конкуренции на тех рынках на которых оно собирается работать.

«Менеджмент знаний» с точки зрения стратегической перспективы развития бизнеса всегда состоял в осознании и выделении в качестве «объекта управления» всего того, что знает и умеет компания до того, как это сделают другие, в получении выгоды посредством выявления и/или создания благоприятных возможностей, о которых еще и не думали другие.

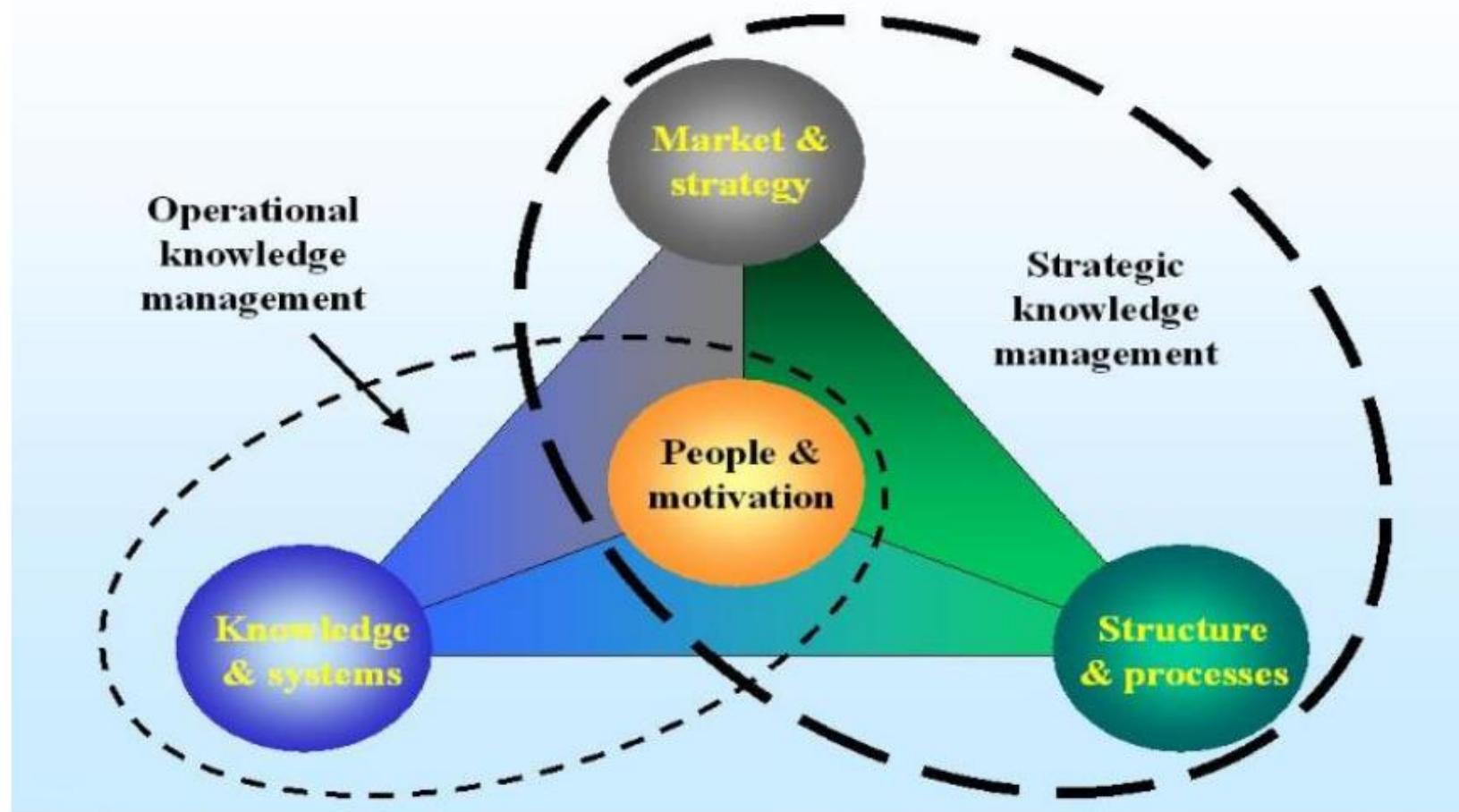
Основными объектами Knowledge Management являются объективированные и не объективированные знания, связи этих знаний с процессами компании, а также сотрудники компании, которые создают и используют эти знания и реализуют процессы, в которые они вовлечены. Эти знания могут использоваться для прогнозирования поведения, регулирования отношений и процессов, которые могут стать более эффективными после применения методов управления знаниями.

«Менеджмент знаний» (или, точнее, менеджмент на основе знаний) – это целенаправленная организация деятельности всей компании, где «знания» рассматриваются, как главный стратегический фактор успеха. «Управление знаниями», в узком смысле этого термина, является лишь составной частью этой глобальной задачи.

«Знание должно переходить в действия»: в эффективное представление информационных ресурсов для принятия решений, и в поддержке процессов выполнения принятых решений. Только такая стратегия управления знаниями и её воплощение в жизнь с помощью информационных и коммуникационных технологий обеспечивает наибольшую вероятность быть первым.



Knowledge management



Компоненты управления знаниями

- А) Люди Им передается опыт для создания новых идей (нововведений)
- Б) Процессы Нужны для совместного использования и распространения информации
- В) Технологии Необходимы для быстрой и эффективной работы людей и процессов

Культура является важнейшей проблемой в сфере знаний, поскольку именно человеческий фактор (ценности, уровень связей или изолированности в организации) создает или разрушает систему управления знаниями. Человеческие взаимодействия и отношения нередко называются «социальным капиталом», элементом общего капитала фирмы.

Технология не может одна разрешить проблемы знания или создать среду обмена знаниями, хотя и является очень важным элементом системы управления знаниями. Применение современных информационных технологий, ни в коем случае не должно устранять необходимые элементы обычного межличностного общения, ведь именно они делают процессы обмена знаниями в организации более интенсивными. В связи с этим необходимо уделять внимание не только материально-технической части, но и, главным образом, организационным моментам.

Структура **организационных** знаний состоит из практических, теоретических, стратегических, коммерческих и производственных знаний. Организация извлекает информацию, выстраивает умозаключения и генерирует новые знания с целью повышения качества выпускаемых изделий и оказываемых услуг и, следовательно, конкурентной позиции фирмы. Управление каждым из перечисленных элементов в составе системы управления знаниями основано на использовании уже рассмотренных процессов — создании, хранении, использовании и распространении знаний в рамках организации.

Таким образом, с точки зрения менеджмента знаний, суть фирмы заключается в ее потенциале создания, передачи, сбиивания воедино, интеграции и эксплуатации знаний как активов. В результате из знаний образуются компетенции, которые, в свою очередь, служат основой для создания продуктов и услуг, предлагаемых фирмой на рынке. Иными словами, управление знаниями, интегрируя в себе множество различных дисциплин, таких как управление персоналом, маркетинг, экономика, психология и информатика, является технологией XXI века, позволяющей организациям обеспечивать свою конкурентоспособность на рынке.

Преимущество в конкуренции обеспечивается за счет того, что корпорация, использующая системы управления знаниями, понимает, какая же информация у нее есть, где она находится, где и когда применять!

Информационные технологии (ИТ, также — информационно-коммуникационные технологии) (Information Technology - IT) — процессы, методы поиска, сбора, хранения, обработки, предоставления, распространения информации и способы осуществления таких процессов и методов (ФЗ № 149-ФЗ); приёмы, способы и методы применения средств вычислительной техники при выполнении функций сбора, хранения, обработки, передачи и использования данных (ГОСТ 34.003-90); ресурсы, необходимые для сбора, обработки, хранения и распространения информации (ISO/IEC 38500:2008).

Информационные технологии — совокупность методов и способов получения, обработки, представления информации, направленных на изменение ее состояния, свойств, формы, содержания и осуществляемых в интересах пользователей.

Информационные технологии - широкий класс дисциплин и областей деятельности, относящихся к технологиям формирования и управления процессами работы с данными и информацией, в том числе с применением вычислительной, компьютерной и коммуникационной техники.

В широком понимании ИТ охватывают все области создания, передачи, хранения и восприятия информации и не только компьютерные технологии. При этом ИТ часто ассоциируют именно с компьютерными технологиями.

Информационными технологиями является вся совокупность технических средств, оказывающих тем или иным образом посредничество между носителями информации и ее потребителями, при этом результатом этой посреднической деятельности является воспроизведение, переработка и хранение той или иной информации.

Отрасль информационных технологий занимается созданием, развитием и эксплуатацией информационных систем. Информационные технологии призваны, основываясь и рационально используя современные достижения в области компьютерной техники и иных высоких технологий, новейших средств коммуникации, программного обеспечения и практического опыта, решать задачи по эффективной организации информационного процесса для снижения затрат времени, труда, энергии и материальных ресурсов во всех сферах человеческой жизни и современного общества. Информационные технологии взаимодействуют и часто составляющей частью входят в сферы услуг, области управления, промышленного производства, социальных процессов.

Специалистов по компьютерной технике и программированию часто называют ИТ-специалистами.

Информационная система, ИС (Information System - IS) — система обработки информации и соответствующие организационные ресурсы (человеческие, технические, финансовые и т. д.), которые обеспечивают и распространяют информацию (ISO/IEC 2382-1:1993).

Информационная система предназначена для своевременного обеспечения надлежащих людей надлежащей информацией, то есть для удовлетворения конкретных информационных потребностей в рамках определенной предметной области, при этом результатом функционирования информационных систем является *информационная продукция* — документы, информационные массивы, базы данных и информационные услуги.

Информационная система - система, предназначенная для реализации и ведения информационной модели какой-либо области человеческой деятельности. Эта система должна обеспечивать следующие средства для протекания информационных процессов:

- сбор информации,
- преобразование и обработка,
- анализ,
- хранение и защита,
- передача для использования.

В самом общем случае: *информационная система* - это взаимосвязанная совокупность определенной идеологии работы с информацией, методов, технологий, технических средств, используемых для сбора, обработки, хранения и выдачи информации потребителю в интересах достижения поставленной цели. Современное понимание информационной системы предполагает применение компьютера в качестве основного технического средства переработки и использования информации.

В деятельности организации информационная система рассматривается как программное обеспечение, реализующее деловую стратегию организации. При этом хорошей практикой является создание и развертывание единой корпоративной информационной системы, удовлетворяющей информационные потребности всех сотрудников, служб и подразделений организации. Однако на практике создание такой всеобъемлющей информационной системы слишком затруднено или даже невозможно, вследствие чего на предприятии обычно функционируют несколько различных систем, решающих отдельные группы задач: управление производством, финансово-хозяйственная деятельность, электронный документооборот и т. д. Часть задач бывает «покрыта» одновременно несколькими информационными системами, часть задач — вовсе не автоматизирована. Такая ситуация получила название «лоскутной автоматизации» и является довольно типичной для многих предприятий

В самом общем случае: *информационная система* - это взаимосвязанная совокупность определенной идеологии работы с информацией, методов, технологий, технических средств, используемых для сбора, обработки, хранения и выдачи информации потребителю в интересах достижения поставленной цели. Современное понимание информационной системы предполагает применение компьютера в качестве основного технического средства переработки и использования информации.

В деятельности организации информационная система рассматривается как программное обеспечение, реализующее деловую стратегию организации. При этом хорошей практикой является создание и развертывание единой корпоративной информационной системы, удовлетворяющей информационные потребности всех сотрудников, служб и подразделений организации. Однако на практике создание такой всеобъемлющей информационной системы слишком затруднено или даже невозможно, вследствие чего на предприятии обычно функционируют несколько различных систем, решающих отдельные группы задач: управление производством, финансово-хозяйственная деятельность, электронный документооборот и т. д. Часть задач бывает «покрыта» одновременно несколькими информационными системами, часть задач — вовсе не автоматизирована. Такая ситуация получила название «лоскутной автоматизации» и является довольно типичной для многих предприятий.

Основными критериями успеха в бизнесе стали профессиональное управление, умение обеспечить эффективную работу персонала, правильно идентифицировать, проектировать, реализовывать и совершенствовать бизнес-процессы, эффективно вести организационно-административную и хозяйственную деятельность. В этих условиях современные информационные технологии и создаваемые на их основе интегрированные информационные системы становятся незаменимым инструментом в обеспечении достижения стратегических целей и устойчивого развития компаний и организаций.

В современных условиях информационные системы играют и будут играть все большую роль и в достижении стратегических целей компаний, обеспечении бизнесу конкурентного преимущества на рынке.

Информационные системы становятся стратегическим источником информации и применяются на всех уровнях организации любого профиля. Вовремя предоставляя нужную информацию, ИС помогают организации достичь успеха в своей деятельности, создавать новые товары и услуги, находить новые рынки сбыта, обеспечивать себе достойных партнеров, организовывать выпуск продукции по низкой цене и т.д.

Применяемые на предприятии информационные технологии поддерживают реализацию тех или иных решений менеджеров. Однако, в свою очередь, новые системы и технологии диктуют свои специфические условия ведения бизнеса, изменяют компании.

Лекция 2. Введение в бизнес-аналитику(BI, BA)

Термины BI, BA

Функции бизнес-аналитики

Технологии бизнес-аналитики

Business Intelligence — это процесс анализа информации, выработки интуиции и понимания для улучшенного и неформального принятия решений бизнес-пользователями, а также инструменты для извлечения из данных значимой для бизнеса информации; процесс, технологии, методы и средства извлечения и представления знаний.

BI - это совокупность технологий, программного обеспечения и практик, направленных на достижение целей бизнеса путём наилучшего использования имеющихся данных.

Цель BI — интерпретировать большое количество данных, заостряя внимание лишь на ключевых факторах эффективности, моделируя исход различных вариантов действий, отслеживая результаты принятия решений.

BI наиболее эффективен, когда он объединяет данные, полученные из рынка, на котором работает компания (внешние данные), с данными из источников внутри компании, таких как финансовые и операции с данными (внутренние данные). В сочетании, внешние и внутренние данные дают полную картину бизнеса, которая, создает "интеллект" — быстрое понимание, которое не получить из простого набора данных.

Рождение термина "Business Intelligence" датируется 1958 годом, когда американский ученый Ханс Петер Лун (1896-1964) опубликовал в IBM System Journal статью «A Business Intelligence System». В ней он представил бизнес как набор различных видов деятельности в науке, технологиях, коммерции, индустрии и даже в законодательной сфере, а обеспечивающие его системы – системами, поддерживающими разумную деятельность (intelligence system).

Словом *intelligence* Лун обозначал способность устанавливать взаимосвязь между представлениями отдельных фактов и действиями в интересах решения поставленных задач и достижения намеченных целей.

В 1989 году аналитик из Gartner Ховард Дреснер дал BI расширительную трактовку, предложив использовать BI в качестве общего термина для различных технологий, предназначенных для поддержки принятия решений - "пользователецентрический процесс, включающий доступ и исследование информации, ее анализ, выработку интуиции и понимания, которые ведут к улучшенному и неформальному принятию решений".

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".

- ***Business Intelligence*** - программные средства, функционирующие в рамках предприятия и обеспечивающие функции доступа и анализа информации, которая находится в хранилище *данных*, а также обеспечивающие принятие правильных и обоснованных управленческих решений.
- Понятие BI объединяет в себе различные средства и технологии анализа и обработки *данных* масштаба предприятия.
- На основе этих средств создаются BI-системы, цель которых - повысить качество информации для принятия управленческих решений.
- BI-системы также известны под названием **Систем Поддержки Принятия Решений (СППР, DSS, *Decision Support System*)**. Эти системы превращают данные в информацию, на основе которой можно принимать решения, т.е. поддерживающую принятие решений.

BI имеет отношение к процессу превращения данных в знания, добытые о бизнесе с использованием различных ИТ, а знаний - в действия для получения выгоды.

Инструменты BI — ИС, которое позволяют бизнес-пользователям видеть и использовать большое количество сложных данных. Знания, основанные на данных, (data-based knowledge) получаются из данных с использованием инструментов business intelligence и процесса создания и ведения хранилища данных (data warehousing).



Итак, Business Intelligence можно определить как:

- процесс превращения данных в информацию и знания о бизнесе для поддержки принятия улучшенных и неформальных решений;
- информационные технологии (методы и средства) сбора данных, консолидации информации и обеспечения доступа бизнес-пользователей к знаниям;
- знания о бизнесе, добытые в результате углубленного анализа детальных данных и консолидированной информации.

В русскоязычной литературе термин "Business Intelligence" переводится как "бизнес-аналитика", "бизнес-интеллект", "интеллектуальный анализ данных", "деловая осведомленность" или вводится просто как аббревиатура BI.

*«*intelligence*»: способность узнавать и понимать; готовность к пониманию; знания, переданные или приобретенные путем обучения, исследования или опыта; действие или состояние в процессе познания; разведка, разведывательные данные.

**Business
Analytics**



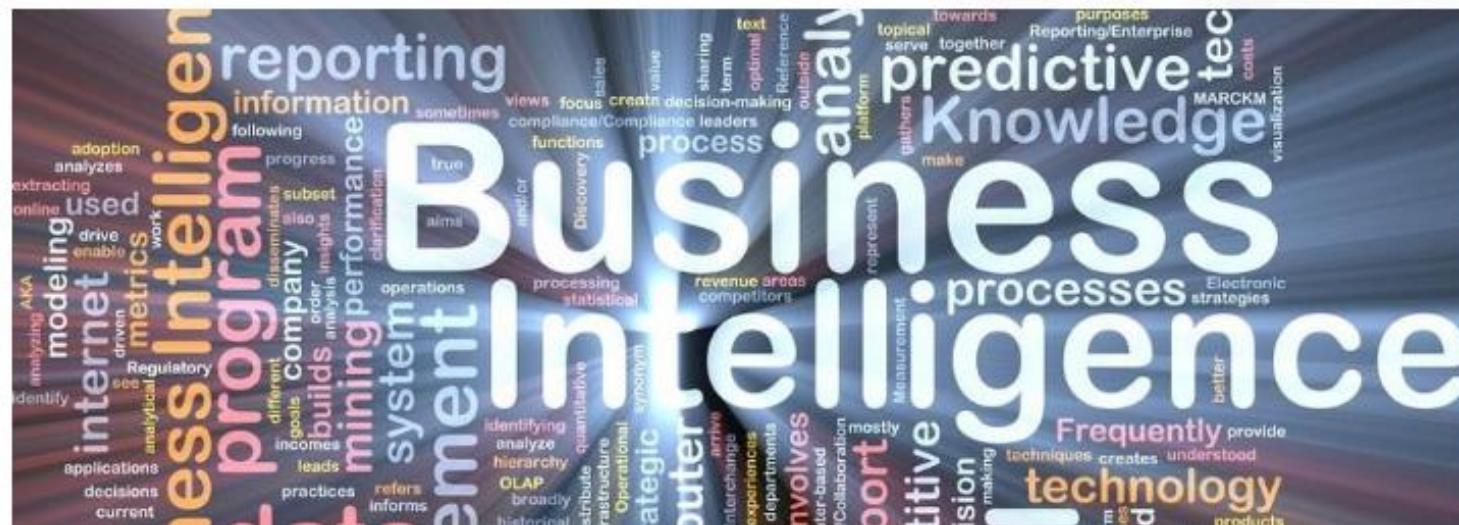
**Business
Intelligence**

citia.co.uk

What's the difference between Business Analytics and Business Intelligence? The correct answer is: everybody has an opinion, but nobody knows, and you shouldn't care.

BI-системы или системы бизнес-аналитики (Business Intelligence) - это аналитические системы, которые объединяют данные из любых различных источников информации, обрабатывают их и предоставляют удобный интерфейс для всестороннего изучения и оценки полученных сведений.

Данные полученные в результате такого анализа помогают достигать поставленных бизнес-целей с помощью оптимального использования имеющихся данных. Комплексный анализ данных по всем направлениям бизнеса позволяет повысить его эффективность и снизить издержки. BI-системы - это единый прозрачный источник данных о бизнесе компании для ее руководства.



Таким образом, основной целью систем бизнес-аналитики является обеспечение быстрого доступа к данным, выполнение анализа данных и информационная поддержка процесса принятия решений. Бизнес-аналитика помогает компании создавать знания из всей доступной информации для принятия эффективных управленческих решений и превращения этих решений в действие.

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*.

**Всевозможные типы
отчетов**

Номер	Наименование	Исполнитель	Компания	Компания-издатель	Файл, страница	Файл, страница	Файл, страница
1	Приложение к рапорту о продажах и маркетинге	Сергей	ОГРН 123456789012345678	ОГРН 123456789012345678	10	10	10
2	Приложение к рапорту о продажах и маркетинге	Сергей	ОГРН 123456789012345678	ОГРН 123456789012345678	10	10	10
3	Приложение к рапорту о продажах и маркетинге	Сергей	ОГРН 123456789012345678	ОГРН 123456789012345678	10	10	10
4	Приложение к рапорту о продажах и маркетинге	Сергей	ОГРН 123456789012345678	ОГРН 123456789012345678	10	10	10
5	Приложение к рапорту о продажах и маркетинге	Сергей	ОГРН 123456789012345678	ОГРН 123456789012345678	10	10	10
6	Приложение к рапорту о продажах и маркетинге	Сергей	ОГРН 123456789012345678	ОГРН 123456789012345678	10	10	10
7	Приложение к рапорту о продажах и маркетинге	Сергей	ОГРН 123456789012345678	ОГРН 123456789012345678	10	10	10
8	Приложение к рапорту о продажах и маркетинге	Сергей	ОГРН 123456789012345678	ОГРН 123456789012345678	10	10	10
9	Приложение к рапорту о продажах и маркетинге	Сергей	ОГРН 123456789012345678	ОГРН 123456789012345678	10	10	10
10	Приложение к рапорту о продажах и маркетинге	Сергей	ОГРН 123456789012345678	ОГРН 123456789012345678	10	10	10
11	Приложение к рапорту о продажах и маркетинге	Сергей	ОГРН 123456789012345678	ОГРН 123456789012345678	10	10	10
12	Приложение к рапорту о продажах и маркетинге	Сергей	ОГРН 123456789012345678	ОГРН 123456789012345678	10	10	10

Консолидация

Единая таблица данных

ID	Название	Описание	Бренд
0	Компьютер	Беспроводной Портативный	Бренд 1
1	Линейка уз	Линейка из пластика	Бренд 1
2	Скотч	Штукатурка	Бренд 1
3	Резак для бумаги	Для бумаги	Бренд 1
4	Бумага цветная	Для печати	Бренд 1
5	Печатная краска	На пластики	Бренд 1
6	Резак для яиц	Свежий яичек	Бренд 1
7	Бумага А4	Печати Стартап, подорожник	Бренд 1
8	Печать термий	Фотоальбом	Бренд 1
9	Печать на бумаге	Текст в принтер	Бренд 1
10	Бумага	Салфетки	Бренд 1
11	Ручка	Печать	Бренд 1
12	Печати	Несколько	Бренд 1

Очистка данных

ID	Название	Маркер	Бренд
0	Продукт 1	Имя продукта 1 в базе	Бренд 1
1	Продукт 2	Имя продукта 2 в базе	Бренд 2
2	Продукт 3	Имя продукта 3 в базе	Бренд 2
3	Продукт 4	Имя продукта 4 в базе	Бренд 2
4	Продукт 5	Имя продукта 5 в базе	Бренд 2
5	Продукт 6	Имя продукта 6 в базе	Бренд 2

В машинно-читаемый
формат

Очистка и
сопоставление данных

Выгрузка данных
в систему

Не-, полу-, разно-
структурированные данные

Структурированные данные
в едином формате

Агрегированные данные
в машинном формате

Запрос обновленных
данных

Принятие решения

Анализ данных

Изменение бизнес-процессов
компании

Аналитический
отчет

Для принятия управленческих решений в системах бизнес-аналитики реализованы следующие функциональные возможности:

- оперативный доступ к необходимой информации;
- интеграция, преобразование и хранение данных из различных информационных систем организации;
- конструирование многомерных хранилищ данных;
- углубленный анализ больших объемов информации;
- визуализация данных (построение наглядных отчетов);
- построение информативных отчетов различной сложности, в т.ч. регламентных отчетов;
- моделирование и прогнозирование ключевых показателей деятельности (KPI) для принятия решений (задачи «Что будет, если ..?» и «Что необходимо для ..?»).

Системы бизнес-аналитики позволяют руководителям принимать обоснованные управленческие решения при значительной экономии времени на поиск и анализ необходимой информации, а также при существенном снижении рутинной нагрузки на персонал.

Ключевыми преимуществами от внедрения BI-системы являются повышение скорости и точности составления отчетов, аналитики и планирования, повышение эффективности процессов принятия решений, увеличение удовлетворенности клиентов. Гораздо реже BI-системы позволяют добиться сокращения операционных расходов или увеличения объемом выручки.

Задачи, решаемые BI-системой

- ✓ поддержка развития бизнес-процессов и структурных изменений предприятия;
- ✓ возможность моделирования различных бизнес-ситуаций в единой информационной среде;
- ✓ проведение оперативного анализа по нестандартным запросам;
- ✓ снижение рутинной нагрузки на персонал и высвобождение времени для более глубокой аналитической работы;
- ✓ устойчивая работа при увеличении объема обрабатываемой информации, возможность масштабирования.

В части поддержки стратегического развития предприятия BI-системы обеспечивают

- ✓ оценку эффективности различных направлений бизнеса;
- ✓ оценку достижимости поставленных целей;
- ✓ оценку эффективности использования ресурсов, в том числе дочерними предприятиями;
- ✓ оценку эффективности операционной, инвестиционной и финансовой деятельности;
- ✓ бизнес-моделирование и оценку инвестиционных проектов;
- ✓ управление затратами, налоговое планирование, планирование капитальных вложений.



Аналитики из Gartner считают, что в платформе для бизнес-анализа должны быть реализованы 15 ключевых возможностей, сгруппированных по трем основным категориям.

Возможности интеграции

BI-инфраструктура – все инструменты BI-платформы должны быть реализованы «в едином ключе», в том числе должны использовать общие данные, единую объектную модель, сквозную модель безопасности, администрирования, порталной интеграции, общий движок исполнения запросов.

Управление данными — все инструменты в рамках единой платформы должна скреплять не только модель данных, но и единые средства для поиска, получения, хранения, повторного использования и публикации объектов данных.

Разработка — BI-платформа должна предоставлять как пользователям, так и разработчикам набор специфических инструментов для создания BI-приложений, которые могут интегрироваться между собой и выстраиваться в сквозные бизнес-процессы, в том числе внедряемые в сторонние (внешние) бизнес-приложения.

Взаимодействие — BI-платформа должна содержать средства для обмена как самой информацией, так и мнениями по поводу полученных результатов. BI-приложения также должны иметь возможность на основании определенных бизнес-правил назначать задания конкретным пользователям и отслеживать их выполнение.

Представление информации

Отчетность — возможность создания форматированных и интерактивных отчетов, с развитыми механизмами для их распространения и обновления.

Контрольные (информационные) панели (dashboards) — это один из видов представления отчетности, способный представлять данные в наглядном, интуитивно понятном виде, при помощи различных шкал, показателей, индикаторов и т.п. Посредством таких контрольных панелей пользователи могут следить за текущим состоянием ключевых показателей и процессов и сравнивать их с намеченными, целевыми значениями. Такие панели позволяют извлекать операционную информацию из бизнес-приложений и делают ее доступной в реальном времени.

Произвольные (ad hoc) запросы — доступная для пользователей возможность самостоятельно (без привлечения ИТ-специалистов) создавать и выполнять уникальные, нетиповые запросы. Для реализации таких возможностей в BI-платформе должен присутствовать развитый семантический слой, позволяющий находить и извлекать нужную информацию из имеющихся источников.

Интеграция с приложениями Microsoft Office — в некоторых случаях BI-платформа используется как промежуточный инструмент для выполнения аналитических задач с соблюдением правил корректности и безопасности данных. При этом в качестве клиентской части BI-системы могут выступать продукты семейства Microsoft Office (в частности, Excel).

Представление информации

Возможности поиска - добавление поискового индекса как к структурированным, так и к неструктурированным источникам данных и объединение их в классифицированную структуру с определенными измерениями (зачастую использующими семантический слой BI) с тем, чтобы пользователи могли с легкостью ориентироваться и изучать данные посредством поискового (как у Google) интерфейса.

Мобильные возможности – этот функциональный блок объединяет инструментарий, позволяющий доставлять отчеты и контент аналитических панелей на мобильные устройства (такие как смартфоны и планшеты), а также использовать интерактивные возможности девайсов (например, нажатие) и другие опции, обычно не доступные на десктопах и ноутбуках (геолокация и другие).

Анализ данных

Оперативная аналитическая обработка данных (OLAP, OnLine Analytical Processing) — поддержка OLAP значительно ускоряет процессы обработки запросов и выполнения расчетов, обеспечивая анализ данных в различных срезах (такой стиль анализа обозначается термином «*slicing and dicing*», что дословно переводится как «нарезание на кубики и ломтики», то есть в продольном и поперечном направлении). Такие инструменты также позволяют осуществлять планирование, использовать анализ по типу "что, если...". OLAP при этом может быть реализован посредством различных архитектур данных и хранилищ, как дисковых, так и *in-memory*.

Интерактивная визуализация — максимально наглядное представление данных с использованием различных интерактивных изображений, схем и графиков (вместо привычных таблиц с обычными строками и столбцами).

Предиктивное моделирование и *data mining* (интеллектуальный анализ данных) — эти инструменты призваны помочь компаниям классифицировать данные, формировать собственные номинальные и количественные шкалы, а также использовать для их анализа развитый математический аппарат.

Анализ данных

Карты показателей (scorecards) — позволяют на основании определенных ключевых показателей (отображаемых на контрольных панелях) контролировать ход выполнения стратегических планов и достижения ключевых показателей эффективности (KPI). Тем самым операционные показатели «привязываются» к целевым стратегическим индикаторам. Для дальнейшего, более детального анализа эти показатели могут расшифровываться при помощи дополнительных отчетов. Такие механизмы позволяют реализовывать различные методологии управления, такие как системы сбалансированных показателей (BSC).

Перспективное моделирование, симуляторы и оптимизация - инструментарий предназначен для поддержки принятия решений в условиях выбора правильного значения переменной как детерминированных, так и при моделировании случайных процессов.

Gartner Group определяет состав рынка систем *Business Intelligence* как набор программных продуктов следующих классов:

- средства построения хранилищ *данных* (*data warehousing*, ХД);
- системы оперативной аналитической обработки (*OLAP*);
- *информационно-аналитические системы* (*Enterprise Information Systems, EIS*);
- средства интеллектуального анализа *данных* (*data mining*);
- инструменты для выполнения запросов и построения отчетов (*query and reporting tools*).

Классификация Gartner базируется на методе функциональных задач, где программные продукты каждого класса выполняют определенный набор функций или операций с использованием специальных технологий.

Средства BI сегодня весьма разнообразны. Они включают:

- инструменты для построения реляционных хранилищ данных - специальным образом спроектированных баз данных, позволяющих с высокой скоростью выполнять запросы на выбор данных;
- серверные и клиентские средства построения многомерных хранилищ данных, содержащих в нереляционной структуре агрегатные данные (суммы, средние значения);
- клиентские приложения для предоставления пользовательских интерфейсов к реляционным и многомерным хранилищам данных;
- средства создания решений на основе подобных хранилищ, анализа многомерных и реляционных данных, генерации отчетов по многомерным и реляционным данным.

Основные возможности BI-систем развиваются по четырем основным направлениям:
хранение данных, интеграция данных, анализ данных и представление данных.

Данные, используемые для бизнес-анализа, организуются в специальные хранилища (data warehouse, DW). Эти данные должны отражать текущую, реальную и полную картину бизнеса. Информация в хранилище данных (включая исторические данные) собирается из различных операционных (транзакционных) систем и структурируется специальным образом для более эффективного анализа и обработки запросов (в отличие от обычных баз данных, где информация организована таким образом, чтобы оптимизировать время обработки текущих транзакций).

Хранилища данных содержат огромные объемы информации, охватывающей все доступные стороны деятельности предприятия и позволяющие рассматривать все аспекты функционирования бизнеса в совокупности. Для решения более узких, конкретных задач из общего хранилища могут вычленяться подмножества данных – так называемые витрины данных (data marts).

Хранилища данных - это предметно-ориентированные, интегрированные, стабильные, поддерживающие хронологию наборы данных, используемые для поддержки принятия управленческих решений.

Данные в хранилище попадают из оперативных систем, а также из внешних источников. По аналогии с "материальными" хранилищами, хранилища данных предусматривают такие операции, как сбор данных ("приход материалов на склад"), хранение данных ("складской запас"), перемещение в витрины данных ("отгрузка товаров в розничную сеть").

Акти

Основные требования к хранилищам данных:

- поддержка высокой скорости получения данных из хранилища;
- поддержка внутренней непротиворечивости данных;
- возможность получения и сравнения так называемых срезов данных (slice and dice);
- наличие удобных утилит просмотра данных в хранилище;
- полнота и достоверность хранимых данных;
- поддержка качественного процесса пополнения данных.

Выполнить все перечисленные требования в рамках одного и того же продукта зачастую не удается. Поэтому для реализации хранилищ данных обычно используется несколько продуктов, одни из которых представляют собой собственно средства хранения данных, другие - средства их извлечения и просмотра, третьи - средства их пополнения и т. д.

В отличие от так называемых оперативных баз данных, с которыми работают приложения, изменяющие данные, реляционные хранилища данных проектируются таким образом, чтобы добиться минимального времени выполнения запросов на чтение (у оперативных же баз данных чаще всего минимизируется время выполнения запросов на изменение данных). Обычно данные копируются в хранилище из оперативных баз данных согласно определенному расписанию. Типичная структура хранилища данных существенно отличается от структуры обычной реляционной СУБД. Как правило, эта структура денормализована (это повышает скорость выполнения запросов) и может допускать избыточность данных.

Для формирования и поддержания хранилищ данных используются так называемые **ETL**-средства – инструменты для извлечения данных (extract), преобразования данных (transform), то есть приведения их к необходимому формату, обработки в соответствии с определенными правилами, комбинировании с другими данными и т.п., а также для загрузки данных (load), записи данных в хранилище или в другую базу.

Инструменты для работы с **SQL** (structured query language), позволяющие пользователям напрямую обращаться к данным. В последнее время инструменты для формирования и обработки запросов стали более «дружественными», ориентированными на неподготовленных бизнес-пользователей (а не на квалифицированных ИТ-специалистов).

Корпоративные порталы, обеспечивающие взаимосвязь на уровне данных и бизнес-процессов. Такие порталы обеспечивают совместный доступ к информации. (**EAI**, enterprise application integration).

Для всестороннего анализа данных в современных BI используются **OLAP-инструменты** (online analytical processing). Они позволяют рассматривать различные срезы данных, в том числе временные, позволяющие выявлять различные тренды и зависимости (по регионам, продуктам, клиентам и т.п.).

В наиболее развитые BI-решения включены блоки для **глубокого исследования данных** (data mining). Иногда этот термин ошибочно используют для обозначения инструментов, позволяющих по-новому представить (отобразить) информацию, однако на самом деле эти инструменты призваны помочь в выявлении скрытых (неочевидных) закономерностей, моделей, составления прогнозов. Они основаны на сканировании и статистической обработке огромных массивов данных и в конечном итоге призваны облегчить принятие правильных и обоснованных стратегических решений благодаря анализу различных вариантов развития событий. В качестве инструментов используются нейронные сети, деревья решений.

Инструменты OLAP являются аналитическими инструментами, которые первоначально были основаны на многомерных базах данных (МБД). МБД — это базы данных, сконструированные специально для поддержки анализа количественных данных с множеством измерений, содержат данные в «чисто» многомерной форме. Большинство приложений включают измерение времени, другие измерения могут касаться географии, организационных единиц, клиентов, продуктов и др.

OLAP позволяет организовать измерения в виде иерархии. Данные представлены в виде гиперкубов (кубов) — логических и физических моделей показателей, коллективно использующих измерения, а также иерархии в этих измерениях. Некоторые данные предварительно агрегированы в БД, другие рассчитываются «на лету».

Средства OLAP позволяют исследовать данные по различным измерениям. Пользователи могут выбрать, какие показатели анализировать, какие измерения и как отображать в кросс-таблице, обменять строки и столбцы «pivoting», затем сделать срезы и вырезки («slice&dice»), чтобы сконцентрироваться на определенной комбинации размерностей. Можно изменять детальность данных, двигаясь по уровням с помощью детализации и укрупнения «drill down/ roll up», а также кросс-детализации «drill across» через другие измерения.

Многомерные хранилища обычно содержат агрегатные данные (например, суммы, средние значения, количество значений) для различных выборок. Чаще всего такие агрегатные функции образуют многомерный набор данных, называемый кубом, оси которого (называемые измерениями) содержат параметры, а ячейки - зависящие от них агрегатные данные (иногда их называют мерами). Вдоль каждой оси данные могут быть организованы в иерархии, отражающие различные уровни их детализации.

	2003	2004	2005
	VISA	Mastercard	Cash
Argentina	11806.28	9190.48	1263.9
Austria			4039.5
Belgium	1745.42	1207.28	14924.12
Brazil			5208.28
Canada	2952.4		
Denmark	1739.76	1376	
Finland	5470.98	3538.92	2328.46
France	11927.48	9823.43	11052.28
Germany	2208.62	1739.6	4681.16
Ireland		330.9	608
Italy	2139.1		1357.6
Mexico			786
Norway	459		
Poland	1268.3	716.72	285.12
Portugal	236.5	220.3	2235.8
Spain	3021.23	2380	1488.8
Sweden	2490.5		1628.32
Switzerland	5094.88	1520.8	901.2

Как правило, агрегатные данные получаются путем выполнения серии запросов на группировку данных при этом довольно часто в качестве источника данных для подобных запросов выступают реляционные хранилища данных.

Цель хранения агрегатных данных - сократить время выполнения запросов, поскольку в большинстве случаев для анализа и прогнозов интересны не детальные, а суммарные данные.

Как исходные, так и агрегатные данные могут храниться либо в реляционных, либо в многомерных структурах. Поэтому в настоящее время применяются три способа хранения данных:

- MOLAP (Multidimensional OLAP) - исходные и агрегатные данные хранятся в многомерной базе данных;
- ROLAP (Relational OLAP) - исходные данные остаются в той же реляционной базе данных, где они изначально находились, агрегатные же данные помещают в специально созданные для их хранения служебные таблицы в той же базе данных;
- HOLAP (Hybrid OLAP) - исходные данные остаются в той же реляционной базе данных, где они изначально находились, а агрегатные хранятся в многомерной базе данных.

Некоторые OLAP-средства поддерживают хранение данных только в реляционных структурах, некоторые - только в многомерных. Однако большинство современных серверных OLAP-средств поддерживают все три способа хранения данных. Выбор способа хранения зависит от объема и структуры исходных данных, требований к скорости выполнения запросов и частоты обновления OLAP-кубов.

Выпущенные в течение последних лет СУБД ведущих производителей - IBM, Microsoft, Oracle, содержат средства для создания многомерных хранилищ данных. Существуют и отдельные продукты для создания OLAP-хранилищ - их выпускают компании Hyperion, Sybase, Business Objects и некоторые другие.

Интеллектуальный анализ данных (data mining) представляет собой процесс обнаружения корреляции, тенденций, шаблонов, связей и категорий. Она выполняется путем тщательного исследования данных с использованием технологий распознавания шаблонов, а также статистических и математических методов. При этом многократно выполняются различные операции и преобразования над сырьими данными (отбор признаков, стратификация, кластеризация, визуализация и регрессия), которые предназначены:

- 1) для нахождения представлений, которые являются интуитивно понятными для людей, которые, в свою очередь, лучше понимают бизнес-процессы, лежащие в основе их деятельности;
- 2) для нахождения моделей, которые могут предсказать результат или значение определенных ситуаций, используя исторические или субъективные данные.

В отличие от использования OLAP разведка данных в значительно меньшей степени направляется пользователем, вместо этого полагается на специализированные алгоритмы, которые устанавливают соотношение информации и помогают распознать важные (и ранее неизвестные) тенденции, свободные от предвзятости и предположений пользователя.

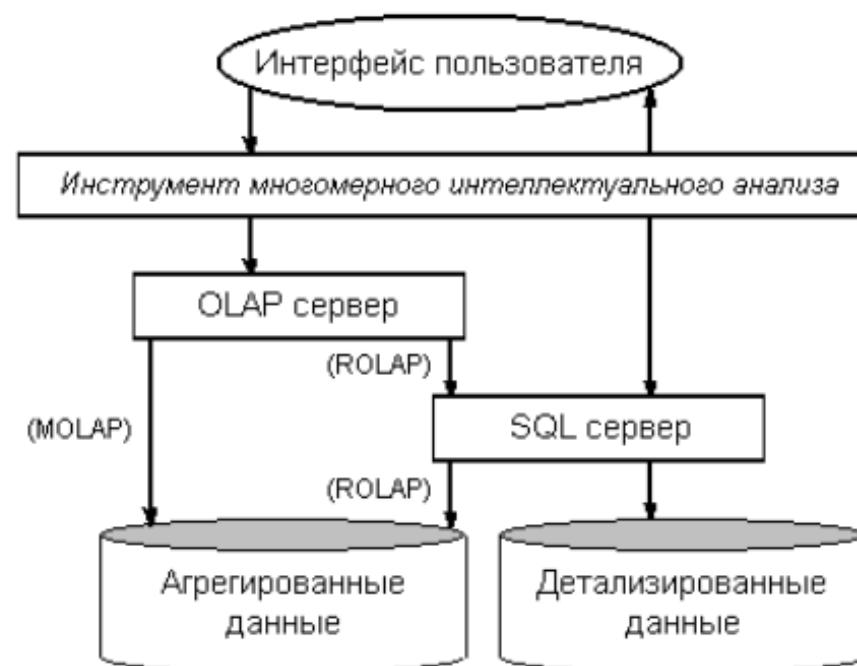
В общем случае процесс Data Mining состоит из трёх стадий :

- 1) выявление закономерностей (свободный поиск);
- 2) использование выявленных закономерностей для предсказания неизвестных значений (прогностическое моделирование);
- 3) анализ исключений, предназначенный для выявления и толкования аномалий в найденных закономерностях.

Иногда в явном виде выделяют промежуточную стадию проверки достоверности найденных закономерностей между их нахождением и использованием (стадия валидации).

Примерами искомой информации могут служить сведения о том, какие категории покупателей чаще всего приобретают тот или иной товар, какая часть покупателей одного конкретного товара приобретает другой конкретный товар, какая категория клиентов чаще всего вовремя не выплачивает предоставленный кредит. Подобного рода информация обычно используется при прогнозировании, стратегическом планировании, анализе рисков, и ценность ее для предприятия очень высока. Если при статистическом анализе или при применении OLAP обычно формулируются вопросы типа "Каково среднее число клиентов банка, не вернувших вовремя кредит, среди неженатых мужчин от 40 до 50 лет?", то применение Data Mining, как правило, подразумевает ответы на вопросы типа "Существует ли типичная категория клиентов, не возвращающих вовремя кредиты?". При этом именно ответ на второй вопрос нередко обеспечивает принятие успешного бизнес-решения.

Оперативная аналитическая обработка и интеллектуальный анализ данных - две составные части процесса поддержки принятия решений. Но сегодня большинство систем OLAP заостряет внимание только на обеспечении доступа к многомерным данным, а большинство средств data mining, работающих в сфере закономерностей, имеют дело с одномерными перспективами данных. Эти два вида анализа должны быть тесно объединены, то есть системы OLAP должны фокусироваться не только на доступе, но и на поиске закономерностей.



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

Генераторы запросов и отчетов — типично «настольные» инструменты, предоставляющие пользователям доступ к базам данных, выполняющие некоторый анализ и формирующие отчеты.

Запросы могут быть как незапланированными (ad hoc), так и иметь регламентный характер.

На рынке ПО сейчас представлено немало средств создания отчетов: как отдельных продуктов, так и входящих в состав средств разработки приложений или СУБД, и реализованных в виде либо серверных служб, либо клиентских приложений. Как правило, средства создания отчетов поддерживают широкий спектр универсальных механизмов доступа к данным (ODBC, OLE DB, ADO.NET), нередко - средства прямого доступа к наиболее популярным СУБД с помощью их клиентских API, содержат средства деловой графики, интегрируются с офисными приложениями, позволяют публиковать отчеты в Интернете, включают классы или компоненты, предназначенные для создания приложений, реализующих (наряду с другими возможностями) генерацию отчетов.

Для представления данных используются различные графические средства – отчеты, графики, диаграммы, настраиваемые при помощи параметров.

Общепринятым средством визуализации данных в современных BI-решениях являются информационные (контрольные, приборные) панели (dashboards), на которых результаты отображаются в виде шкал и индикаторов, позволяющих контролировать текущие значения выбранных показателей, сравнивать их с критическими (минимально\максимально допустимыми) значениями и таким образом выявлять потенциальные угрозы для бизнеса.

Контрольные панели считаются одним из наиболее удобных способов представления информации о «состоянии здоровья» бизнеса. Они позволяют уместить на экране всю важнейшую информацию о текущих операциях, выявленных и потенциальных проблемах.

Контрольные панели, как и карты показателей (scorecards), основаны на анализе ключевых показателей эффективности (KPIs). Однако, как правило, контрольные панели отображают текущее состояние общих показателей, а карты показателей предназначены для сравнения текущих показателей с плановыми, целевыми, и отображают динамику изменения этих показателей во времени. Карты показателей обычно бывают более персонализированными, настраиваются в зависимости от ролей и задач конкретного пользователя (финансовое управление, снабжение, продажи и т.п.). При необходимости все эти показатели могут быть детализированы при помощи дополнительных отчетов, графиков и диаграмм.

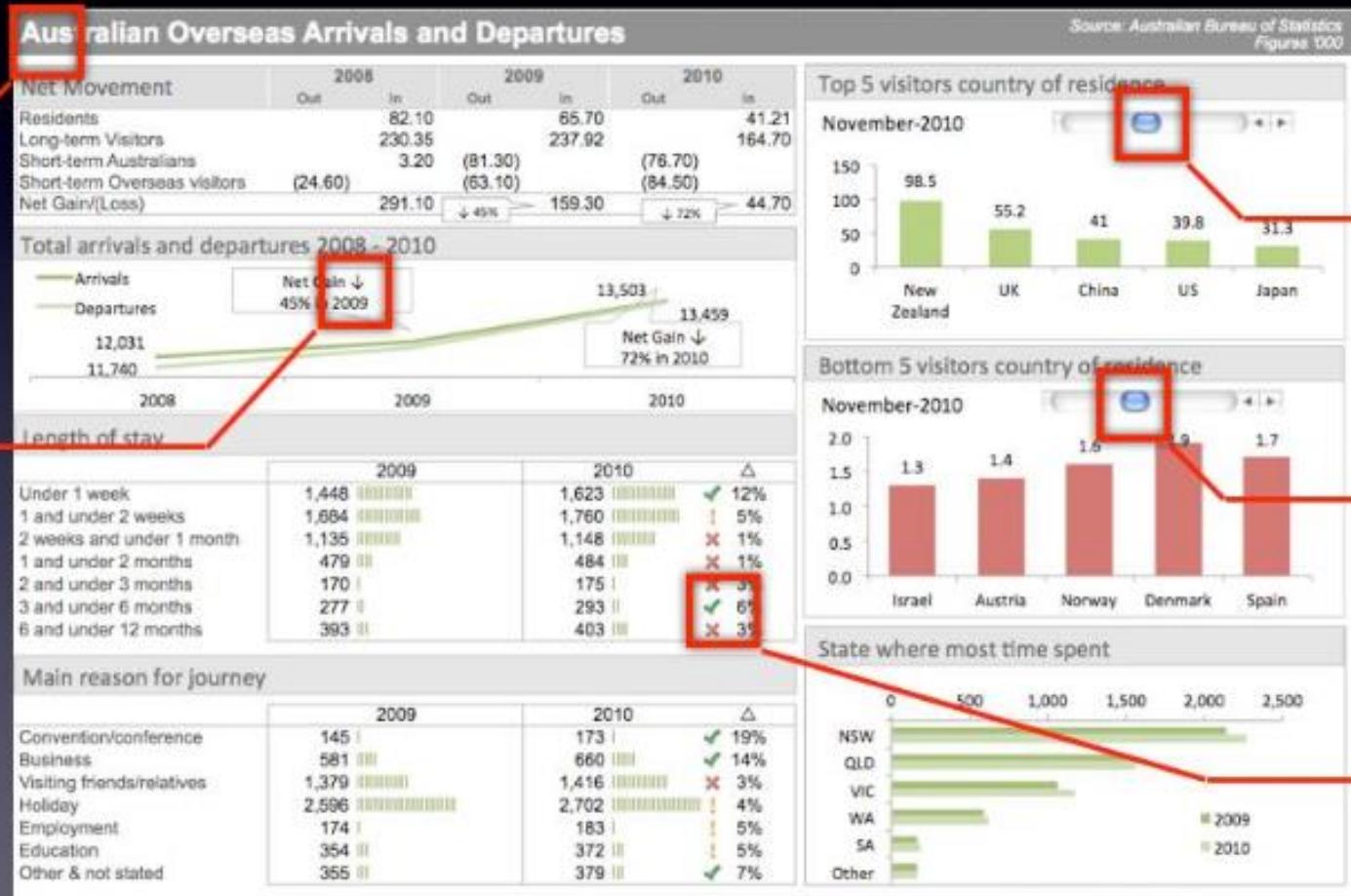
Dashboard Example

Fits Screen

- One page report
- Gives overview
- Inter-relationships

Charts

- Visual presentation
- In-chart comments
- Compare trends



Slider 1

- Interactive element
- Mouse controlled

Slider 2

- Interactive element
- Mouse controlled

Checkmarks

- Visual indicators

Finance **Healthcare** **Legal** **Manufacturing** **Marketing** **Projects** **Sales** **Service Management**

Add a dashboard item:

Hospital Aging Summary

Category	Value
1. Curr-30	3.5M
2. 31- 60	2.2M
3. 61- 90	1.2M
4. 91-120	1.1M
5. 121-150	1.1M
6. 151-180	633.4K
7. 181-360	2.5M
8. OVR 360	870K

Hospital Bill Lag Category Summary

Category	Value
1. 0 - 5	1.1M
2. 6 - 10	1.1M
3. OVR 10	10.9M

Hospital Payer Summary

Payer	Value
Adm Charity	17.8K
Blue Cross	537K
Blue Cross MC+	978.6K
Champus / Tricare	138.4K
Commercial	390K
Discount	2.9M
Family Health P...	1.4M
Firstguard MC+	1.3M
Healthcare USA MC+	220.4K
Healthnet	154.6K
Healthnet HMO /...	647.5K
HMO	372.3K
Medicare	1.5M
Medicare Part B	1.7M
Other MC+	26.8K
Special Programs	272.2K
TPL	535.3K

McKesson ATB Aging Summary (current period)

Category	Value
1. Curr-30	3.5M
2. 31- 60	2.2M
3. 61- 90	1.2M
4. 91-120	1.1M
5. 121-150	1.1M
6. 151-180	633.4K
7. 181-360	2.5M
8. OVR 360	870K

McKesson ATB Aging (Over 360 Days)

Value	Color
0.0	Green
0.5	Yellow
1.0	Orange
1.5	Red
2.0	Dark Red
2.3	Black

McKesson Occupancy Percentage

Location	Occupancy (%)
SIU Surgery ICU	95%
MIU Medical ICU	85%
CCU CARDIOVASC CARE UNIT	85%
IMT Intermed/Telemetry	80%
4GL Inpatient 4 Gold	80%
3BL Inpatient 3 Blue	80%
4RD Inpatient 4 Red	80%
3GL Inpatient 3 Gold	80%
4BL Inpatient 4 Blue	80%
LDR Labor/Del/Rec/Post	80%
GYN IP 6 GYN/Antepartum	40%
NSI NURSERY-ITR/ICU/CON	10%
NSI NURSERY-RTH/LDRP	10%
4GR Inpatient 4 Green	10%

McKesson Occupancy Trend

Date	Occupied	Total Beds
Apr 11 2009	~180	~350
Apr 12 2009	~150	~350
Apr 13 2009	~180	~350
Apr 14 2009	~150	~350
Apr 15 2009	~180	~350
Apr 16 2009	~150	~350
Apr 17 2009	~180	~350
Apr 18 2009	~150	~350
Apr 19 2009	~180	~350
Apr 20 2009	~150	~350

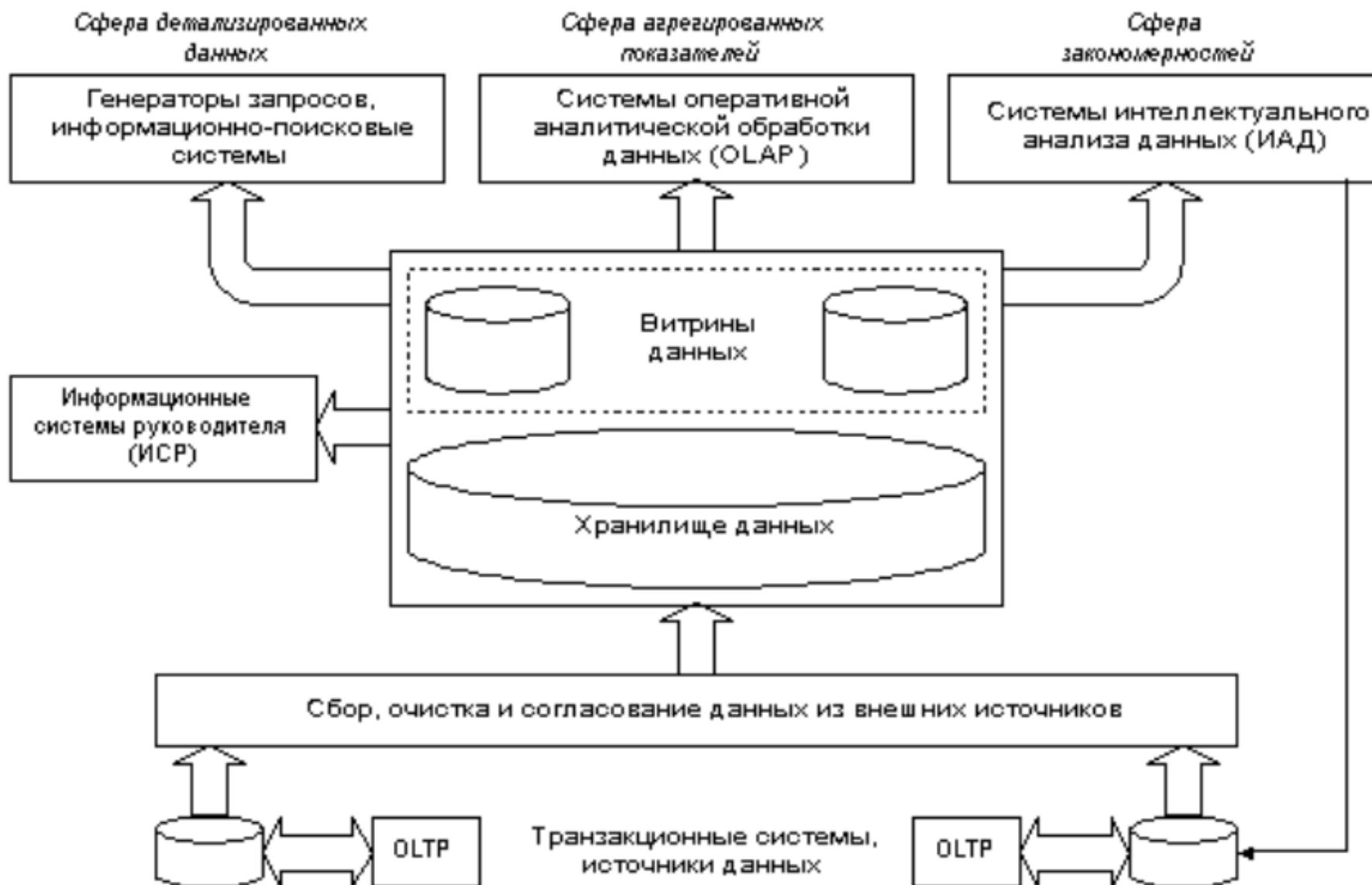
Meditech Monthly Overtime by Dept

Showing 1 to 88 of 88

Dept Number	Dept Name	Overtime Hours
16710	E.R.	1593.3
16721	LONGVIEW AMBULANCE	1486.4
158450	HOUSEKEEPING	1094.0
16730	GSMC - ALS EMS	895.6
16727	HARLETON EMS	714.4
17420	PULMONARY SERVICES	661.9
16745	ORE CITY EMS	661.4
16290	A 600 IMC	648.6
16732	LINDEN EMS	648.3
16011	REHAB UNIT	603.1
16009	SHORT STAY UNIT	585.5
16280	IMC CARDIAC	557.8
16729	LONE STAR - EMS	519.8
16728	DANGERFIELD EMS	434.8

Meditech MTD Revenue by Dept (Top 20)

Dept	Revenue
WOUND CARE CENTER	~4.8M
DELIVERY	~4.2M
A300-MEDICAL	~3.6M
LINDEN EMS	~3.2M
LONGVIEW AMBULANCE	~2.8M
S-ICU	~2.4M
LONE STAR - EMS	~2.0M
IMC CARDIAC	~1.6M
A 500 -M-ICU	~1.2M
NICU	~1.0M
TATUM AMBULANCE	~0.8M
ONCOLOGY	~0.6M
LACTATION CENTER	~0.4M
AMBULANCE	~0.2M
PEDI	~0.1M
PICC LINES	~0.05M
A300 - MED MONITOR	~0.02M
ELDERVILLE EMS	~0.01M
S 300-MEDICAL	~0.01M
GSMC - ALS EMS	~0.01M



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. Хранилище данных извлекает данные из множества транзакционных или оперативных систем, а затем интегрирует и хранит данные в специализированной БД. Например, в хранилище данных 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 хранилища данных и анализируют ее. Таким образом, определяются тенденции, структуры и исключения. Аналитические инструменты помогают пользователям преобразовать информацию в знания.

Business Performance Management (BPM) - это методология, направленная на оптимизацию реализации стратегии и состоящая из набора интегрированных циклических аналитических процессов, которые поддерживаются соответствующими технологиями и имеют отношение как к финансовой, так и к операционной информации. BPM позволяет предприятию определять, измерять и управлять эффективностью своей деятельности, направленной на достижение стратегических целей. Ключевые финансовые и операционные процессы BPM включают планирование, консолидацию и отчетность, анализ ключевых показателей эффективности и их распространение в рамках организации.

- управление эффективностью бизнеса (Business Performance Management, BPM);
- управление эффективностью деятельности предприятия (Enterprise Performance Management, EPM);
- управление эффективностью деятельности корпорации (Corporate Performance Management, CPM);
- стратегическое управление предприятием (Strategic Enterprise Management, SEM).

Методологию сбалансированных систем показателей (Balanced Scorecard, BSC), определяют как инструмент, позволяющий трансформировать миссию и стратегию организации в исчерпывающий набор показателей эффективности, которые служат основой для системы стратегического управления и контроля.

Система Balanced Scorecard позволяет выстроить сквозную связь между стратегией и тактикой организации, в результате чего задача трансформации стратегии в реальные действия оказывается решена. Кроме того, такая система позволяет не только формализовать стратегию, но и контролировать успешность ее реализации за счет измерителей и значений целевых показателей.

Balanced Scorecard по праву можно назвать наиболее популярной среди методик стратегического управления. Но это не означает отсутствия других методов и подходов, многие из которых также получили достаточно широкое распространение и признание. Примерами таких разработок могут служить методика управления стоимостью компании (Value Based Management, VBM).

Лекция 3. ВІ платформы

Важным инструментом Business Intelligence являются BI-платформы.

- Аналитики компании Gartner определили BI-платформу как инструмент, который дает организациям возможность строить приложения, позволяющие изучать и понимать бизнес. Аналитики Gartner в своем ежегодном отчете выделяют 12 основных функций современных платформ Business Intelligence, которые объединены в три группы:

1. Средства интеграции. В данную группу включены следующие функции:

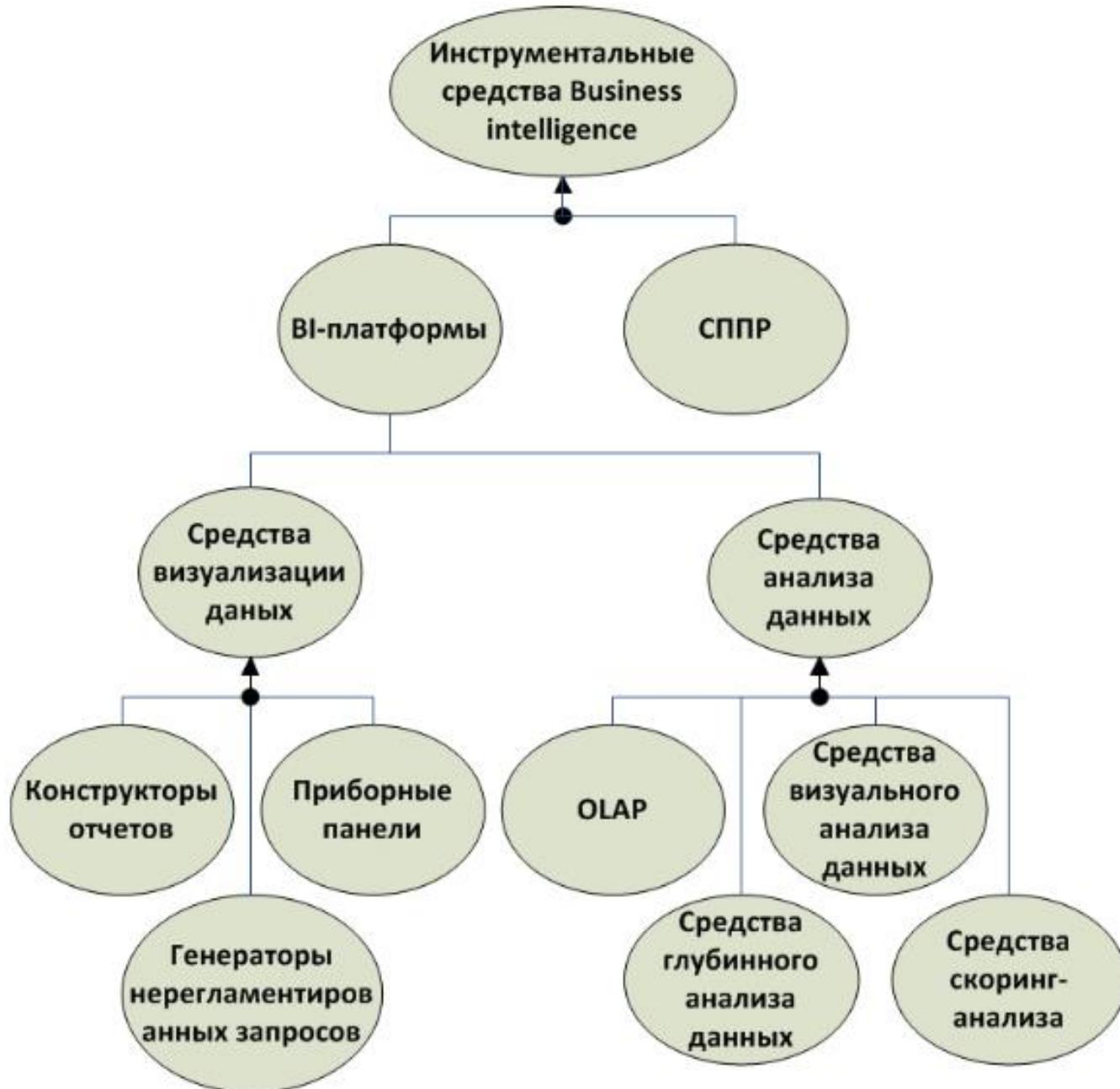
- Инфраструктурное единство. Все компоненты платформы должны использовать единые средства обеспечения безопасности, метаданные, средства администрирования, средства генерации запросов, а также иметь однотипные интерфейсы
- Управление метаданными. Все компоненты платформы должны не только опираться на одни и те же метаданные, но также должны обеспечивать быстрый поиск, хранение, использование и публикацию объектов метаданных. Средства разработки BI-платформа должна предоставлять средства программной разработки и интеграции BI-приложений. BI-платформа должна предоставлять разработчикам возможность создания приложений без кодирования, на основе применения мастеров.
- Совместная работа и управление рабочими процессами. Платформа должна предоставлять пользователям возможность разделять информацию и обсуждать её с помощью общих папок и средств ведения дискуссионных «тредов» (discussion threads).
- В дополнение BI-приложения могут давать возможность назначать и отслеживать события или задачи, возложенные на отдельных пользователей, на основе неких заранее определенных бизнес-правил. Обычно данная функциональность предоставляется на базе интеграции с отдельным Workflow-инструментом

2. Средства визуализации данных. К данной группе относят следующие функции:

- Средства создания отчётов. Современные BI-платформы предоставляют возможность создавать форматированные интерактивные отчеты. В дополнение к этому поставщики BI-платформ должны предоставлять широкий набор типов отчетов (финансовых, операционных и т.п.).
- Дэшборды, или приборные панели. Приборные панели – это специальное средство визуализации данных, представляющее информацию в виде интуитивно понятного графического изображения (диаграммы, круговые шкалы, светофоры ит.п.). Данные индикаторы показывают состояние анализируемых параметров на фоне их целевого назначения. Руководитель или аналитик, подобно пилоту самолета, видит перед собой «доску приборов» и управляет системой, ориентируясь на значения индикаторов. При этом ключевые факторы, необходимые для управления предприятием, должны быть так или иначе измерены и представлены в виде показателей.
- Генератор нерегламентированных запросов. Платформа должна предоставлять пользователям возможность полу-чать ответы на возникающие вопросы. Система предоставляет средства навигации по доступным ресурсам данных.
- Интеграция с офисными приложениями. Часто BI-платформы используются как промежуточное звено в цепочке анализа информации, а MicrosoftOffice(в частно-сти, Excel) выступает как BI-клиент. В этих случаях очень важно, чтобы BI-платформа обеспечивала интеграцию с MicrosoftOffice, включая поддержку форматов документов, формул и сводных таблиц

3. Средства анализа данных. В данную группу объединены следующие функции

- Оперативно-аналитическая обработка данных (OLAP). На сегодняшний день технология OLAP стала концептуальной основой при реализации средств обработки и манипуляции данными в информационно-аналитических системах и системах поддержки принятия решений.
- Продвинутая визуализация или средства визуального анализа. Инструменты продвинутой визуализации позволяют представлять данные для более эффективного их восприятия посредством использования интерактивных картинок и диаграмм вместо таблиц.
- Средства глубинного анализа данных. В последнее время пользователи проявляют всё больший интерес к средствам статистического и интеллектуального анализа данных. Это требует от BI-платформ включения в них данного инструментария.
- Средства скоринг-анализа. Используются контрольные показатели, отображаемые на информационной панели, для более глубокого анализа путем наложения их на некоторую стратегическую карту, которая увязывает ключевые параметры производительности со стратегическими задачами



Примеры систем

Tableau, Qlik, Microsoft

- Интуитивно понятный пользовательский интерфейс
- Предоставление некоторых версий своих продуктов в бесплатное и неограниченное использование
- Доступ к бесплатным учебникам и профессиональному форуму, поощрение совместного сотрудничества и обмена знаниями
- Упрощение процесса развёртывания



The screenshot shows the Tableau website homepage. At the top, there's a navigation bar with links for Products, Solutions, Learning, Community, Support, and About. On the right side of the header are buttons for PRICING, SIGN IN, TRY NOW, and a search icon.

Below the header, there are three main product sections:

- Tableau Desktop**: Analytics anyone can use. [LEARN MORE →](#)
- Tableau Server**: Analytics for organizations. [LEARN MORE →](#)
- Tableau Online**: Cloud analytics for organizations. [LEARN MORE →](#)

To the right of these sections is a sidebar titled "MORE TABLEAU PRODUCTS" with links to Tableau Public, Embedded Analytics, Developer Tools, and Pricing, each accompanied by a small circular icon.

The central part of the page features a large, colorful data visualization on a laptop screen. The visualization consists of multiple stacked area charts in shades of orange, yellow, grey, green, and blue, showing fluctuating data over time. A tooltip on the visualization displays the text: "United Kingdom Date: March 2014 UN-Acces: 99.52% Stat: Covid-19".

On the left side of the page, there's a large, semi-transparent image of a person working at a desk with a laptop, and a text overlay that reads: "Want your data make an impact". Below this text is a button labeled "SEE IT IN ACTION" with a small circular icon.

At the bottom left, there's a small URL: <http://www.tableau.com/products>.

Надежный | [Http://www.qlik.com/ru-ru](http://www.qlik.com/ru-ru)

Откройте полную картину, скрытую в ваших данных

С помощью Qlik®

вы можете создавать визуализации, интерактивные панели и приложения, которые помогут ответить на наиболее важные вопросы, возникающие в вашей компании. Таким образом, вы сможете увидеть [полную картину происходящего, скрытую в ваших данных](#).

CHAT LIVE

Простота использования. Высокая эффективность.

Анализируйте большие объемы информации и осуществляйте поиск по всем источникам. Qlik избавляет вас от устоявшихся представлений о взаимосвязях данных и помогает изучить подлинные корреляции.



Полная линейка продуктов Qlik®

Решайте любые бизнес-вопросы. Изучайте данные и получайте ценные сведения, которые отвечают изменяющимся потребностям вашей компании. Делитесь информацией и сотрудничайте с другими пользователями, где бы вы ни были.



Qlik Sense®

Обнаруживайте ценные сведения с помощью приложения для визуализации данных, доступного каждому. Благодаря Qlik Sense пользователи ваших организаций могут создавать любые интерактивные визуализации и принимать обоснованные решения.

[ПОДРОБНЕЕ](#)

QlikView®

Получите более ценные сведения, создавая многофункциональные приложения управления знаниями с помощья любого уровня бизнес-решений.

[ПОДРОБНЕЕ](#)

Qlik Sense Cloud

Создавайте визуализации и работайте с приложениями Qlik Sense в любой точке. Прайвилегии и другие возможности работы в базовой среде. Благодаря облаку вы можете делько начать данные, и мгновенно обновлять связи к ним и видеть перспективы.

[ПОДРОБНЕЕ](#)

Аналитическая платформа Qlik

С этой платформой мы можем с легкостью создавать как простые пабличные, так и комплексные интегрированные анал-приложения. Настоящая инто-заработка, воплощенная в реальность.

Qlik NPrinting™

Современное решение для создания и рассылки отчетов.

Qlik DataMarket

Принимайте обновленные деловые решения. С помощью платформы Qlik DataMarket вы можете находить данные на внешних источниках, управлять ими и добавлять в свою базу.

The screenshot shows the Microsoft Power BI website homepage. At the top, there is a navigation bar with links for Microsoft, Power BI, Products, Solutions, Partners, Learning, Sign In, and Free registration. A yellow sidebar on the left lists various Power BI products: Power BI, Power BI Desktop, Power BI Premium, Power BI Mobile, Power BI Embedded, Power BI Report Server, Overview, and Prices. Below the sidebar, there are two mobile device screens displaying Power BI reports. A central tablet screen shows a detailed dashboard with various charts and metrics. A yellow banner at the bottom of the main content area contains the text: "Решение Power BI Premium уже доступно. Помогите своим пользователям получать необходимую информацию." followed by a link "Дополнительные сведения".

Что такое Power BI?

Power BI — это набор средств бизнес-аналитики для получения важных данных в организации. Подключайтесь к сотням источников информации, упростите ее обработку и динамический анализ. Создавайте великолепные отчеты и публикуйте их для пользователей своей организации, которые смогут работать с ними в браузере и на мобильных устройствах.

Персонализированные панели мониторинга с уникальным полным обзором ситуации в компании теперь доступны каждому. Вы сможете легко управлять решением со встроенными средствами управления и

<https://powerbi.microsoft.com/ru-ru/what-is-power-bi/>

Самостоятельный обзор современных BI-платформ:

- Microsoft (<http://www.microsoft.com/en-us/bi/default.aspx>);
- Oracle (<http://www.oracle.com/us/solutions/business-analytics/business-intelligence/overview/index.html>);
- MicroStrategy (<http://www.microstrategy.com/software/business-intelligence/>);
- IBM (<http://www-142.ibm.com/software/products/us/en/category/SWQ20>);
- SAP (<http://www.sap.com/cis/platform/businessintelligence/index.epx>);
- SAS (<http://www.sas.com/technologies/bi/>);
- PROGNOZ (<http://www.prognoz.ru>);
- BaseGroup Labs (<http://basegroup.ru>)

- Спасибо за внимание!