

Учебное пособие
по дисциплине
«Системы поддержки принятия решений»
(курс лекций)

Оглавление

Введение.....	5
Тема №1. Принятие решений, решение и выбор, процесс принятия решений ..	6
Тема №2. СППР, концепция, основные термины и определения, решаемые задачи.....	17
Тема №3. Формирование баз моделей и систем управления моделями в СППР	25
Тема №4. Работа с оптимизационными моделями применимыми в процессах принятия решений: оптимальное управление запасами	31
№5. Концептуальные основы СППР, архитектура СППР	40
Тема №6. Пример реализации СППР – «Монитор руководителя»	53
Тема №7. Информационное пространство предприятия, показатели отчетности	59
Тема №8. Интеграция данных в рамках СППР из различных источников	70
Тема №9. «Оперативная аналитическая обработка данных в СППР»	86
Тема №10. «Хранилища данных, архитектура, классификация, примеры проектных решений»	96
Тема №11. «Области применения хранилищ данных»	106
Тема №12. «Реализация СППР в банковской сфере»	121
Тема №13. Интеллектуальный анализ данных – технология Data Mining	133
Тема №13. «Адаптация корпоративного портала для принятий решений (на примере 1С: Битрикс24)»	140
Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	149

Тема №1. Принятие решений, решение и выбор, процесс принятия решений

Каждый человек в жизни сталкивается с необходимостью принятия важных и не очень важных решений. Но не всякий задумывается, как сделать свой индивидуальный выбор наилучшим образом, получить наибольшую пользу и уменьшить возможные негативные последствия от реализованного решения.

В современных условиях резко повысилась цена, которую приходится платить обществу за недостаточно обоснованные экономические или социальные решения. Одновременно увеличилась и мера ответственности руководителей, принимающих решение. Как никогда ранее, усилилась взаимная зависимость всех лиц, участвующих в подготовке и принятии решения. Каждый руководитель, решая конкретные вопросы на своем уровне управления, должен увязывать интересы разных сторон, учитывать сложившиеся связи и последствия их нарушения.

Возрастающие требования к качеству управления в разных сферах человеческой деятельности диктуют необходимость выполнения специальной аналитической работы при формировании и принятии решения. Современный руководитель должен принимать решение не интуитивно, а используя соответствующий инструментарий для поиска лучшего варианта и обоснования сделанного выбора.

Для подготовки решения привлекаются специалисты-эксперты, консультанты, системные аналитики, а в сложных и уникальных ситуациях выбора их участие обязательно. Основная задача экспертов состоит в разработке альтернативных вариантов, выявлении достоинств и недостатков каждого из них, оценке последствий выбора того или иного варианта. Для эффективного выполнения своих функций эти специалисты должны обладать знаниями о существующих методах и средствах поддержки принятия решений, а также уметь применять такой инструментарий на практике.

Теория принятия решений - комплексная научная дисциплина, направленная на разработку методов и средств одному или нескольким лицам сделать обоснованный выбор наилучшего из имеющихся вариантов.

Решение и выбор

Слово «решение» имеет в русском языке несколько значений, которые для наших целей важно различать.

Во-первых, под решением понимается совокупность рассматриваемых возможностей, которые тем или иным образом выделены человеком, делающим выбор.

Во-вторых, решение представляет собой процесс поиска наиболее предпочтительных вариантов, включающий в себя обдумывание, изучение какого-либо вопроса или задачи, нахождение правильного ответа.

В-третьих, решением является и сам полученный в ходе поиска ответ, например один или несколько выбранных вариантов, результат анализа некоторой проблемы или математической задачи.

Наконец, решениями называют указы, постановления, распоряжения, приказы, акты органов законодательной и исполнительной власти, судебные и иные решения. В английском языке для этих понятий в первом случае говорят *alternative, decision*, во втором — *solving, choice*, в третьем — *solution, resolution*, в четвертом — *decree, order*.

Принятие решений в профессиональном отношении представляет собой особый вид человеческой деятельности, который состоит в обоснованном выборе наилучшего в некотором смысле варианта или нескольких предпочтительных вариантов из имеющихся возможных.

По-английски этот термин звучит, как *decisionmaking*, т.е. буквально означает «делание» или создание решения, что более адекватно смыслу этого словосочетания. Здесь слово «решение» соответствует своему первому и третьему значениям.

Задачи принятия решений часто отождествляются с задачами выбора, являющимися одними из самых распространенных задач, с которыми человек

сталкивается в своей деятельности. В повседневной жизни нам постоянно приходится делать выбор того или иного товара, покупаемого в магазине, блюда, заказываемого в кафе или ресторане, маршрута поездки или вида транспорта и т. п. В силу повторяемости, стереотипности ситуаций выбора человек принимает решение, почти не задумываясь, часто интуитивно или по аналогии. Лучший вариант обычно находится без какого-то особого анализа.

В более сложных и соответственно более редких, уникальных ситуациях, например при выборе места отдыха, учебы или работы, покупке квартиры или дорогостоящей вещи (автомобиля, мебели и т.д.), голосовании за того или иного кандидата или партию, человек более тщательно подходит к своему выбору. Прежде чем принять решение, он старается детально рассмотреть, оценить и сопоставить различные варианты, учесть разные точки зрения.

Еще более сложные задачи выбора возникают в профессиональной деятельности каждого руководителя, ученого, конструктора, врача, экономиста, финансиста, бизнесмена, военачальника. При принятии политических, экономических, производственных, военных решений требуется учитывать различные и зачастую не совпадающие интересы действующих сторон, нужно отыскивать и анализировать разнообразную информацию. Для сравнения различных вариантов действий приходится проводить всесторонний, иногда достаточно сложный анализ проблемной ситуации, разрабатывать для этого специальные модели, привлекать к выработке вариантов решения специалистов, экспертов, консультантов, аналитиков, использовать средства вычислительной техники, строить компьютерные системы поддержки принятия решений.

Подобные проблемы возникают у людей, занятых управлением сложными техническими объектами (энергетическими системами и установками, самолетами, кораблями и т.п.). Но здесь ситуации осложняются тем, что решение требуется принять оперативно, в реальном масштабе времени, практически не имея возможности для детального анализа всех альтернативных вариантов и возникающих последствий их реализации.

Необходимость обоснования выбора присутствует во всех сферах человеческой деятельности и обоснованный выбор особенно важен в управлении организационными и техническими системами.

В ситуациях принятия сложных решений всегда существует недостаток информации. Часть нужной информации нередко отсутствует, а имеющаяся информация может быть противоречивой. Опытный руководитель или специалист покрывает неполноту информации своими знаниями, умением и интуицией. Принятие верных решений в сложных ситуациях является своего рода искусством, которым владеют немногие.

Однако одного искусства в принятии решений в современных условиях мало. Возросла динамичность жизни; сократился период времени, в течение которого принятые ранее решения остаются верными; повысилась сложность вариантов принимаемых решений, их взаимозависимость и взаимосвязь; существенно увеличились возможные риски и неопределенность последствий, масштабы и размеры потерь в случае принятия недостаточно обоснованных решений. Как следствие, существенно возросла ответственность человека за принятие наилучшего, «самого правильного» решения, увеличились трудности, связанные с его нахождением, преодолеть которые невозможно без использования всего арсенала средств, накопленных современной теорией принятия решений.

Теория принятия решений

Изучением того, как человек принимает решения, и созданием методов выбора занимаются многие научные дисциплины, которые возникли и исторически развивались независимо друг от друга. К ним относятся теория принятия решений, системный анализ, исследование операций, теория статистических решений, теория игр, теория оптимального управления, экономическая кибернетика, теория организаций, информатика, искусственный интеллект, когнитивная психология, теория поведения и др.

Эти дисциплины с разных точек зрения анализируют механизмы, процессы и правила выбора применительно к объектам различной природы и в

различных условиях их существования. Все вместе они образуют междисциплинарную науку, помогающую человеку сделать обоснованный выбор.

Теория принятия решений, как самостоятельное научное направление, стала складываться в середине XX века в рамках методологии системного анализа, хотя самые первые работы по исследованию голосования как способа коллективного выбора появились еще в конце XVII века. Основное назначение теории принятия решений состоит в разработке методов и средств, позволяющих человеку или группе лиц сформулировать множество возможных вариантов решения проблемы, сравнить их между собой, найти среди них лучшие или допустимые варианты, которые удовлетворяют тем или иным требованиям (критериям), и при необходимости объяснить сделанный выбор.

Теория принятия решений может оказать существенную помощь в анализе и решении сложных проблем, но лишь тогда, когда ее методологические и математические средства применяются «правильно», соответственно их возможностям, не преувеличивая и не умаляя их роли в процессе нахождения решения.

Поэтому теорию принятия решений правильнее было бы назвать теорией поиска и обоснованного выбора наиболее предпочтительных для человека вариантов решения проблемы.

Существуют две противоположные точки зрения на роль формальных методов при решении практических проблем выбора. Люди, профессионально не владеющие математическими методами, нередко считают, что любая проблема может быть формально переведена на язык математики и потом решена ее средствами. Другие полностью отвергают такие возможности. Действительность же гораздо сложнее этих крайних утверждений.

Любые ситуации, требующие принятия решения, содержат, как правило, большое число неопределенных факторов, которые оказывают влияние, как на формальную постановку задачи, так и на средства ее решения. Эти неопределенные факторы можно в самом общем виде разбить на три группы.

Прежде всего, это так называемая *неопределенность природы*, т.е. факторы людям попросту неизвестные или от них не зависящие. Затем — *неопределенность человека*, который может вести себя непоследовательно, противоречиво, допускать ошибки, зависеть от других лиц (партнеров, противников и т. д.), чьи действия он не может полностью учесть или предвидеть. И наконец, *неопределенность целей*, которые могут различаться и не совпадать друг с другом. Например, авиаконструкторы, проектируя самолет, должны учитывать его целевое назначение, заданные показатели скорости, грузоподъемности и дальности полета, условия безопасности и комфортности для экипажа и пассажиров, факторы экономичности и технологичности производства и эксплуатации самолета, экологические требования и многие другие обстоятельства.

Полностью свести подобные задачи с неопределенностью к корректно поставленным математическим задачам нельзя в принципе. Чтобы сделать возможным их решение, надо как-то ограничить, уменьшить или, как говорят, «снять» неопределенность.

Для этого проводится *содержательный анализ проблемной ситуации*, делаются какие-либо предположения и вводятся упрощения в постановку задачи. И именно средства, входящие в состав тех или иных методов принятия решений, очень часто позволяют получить дополнительную информацию, нужную для формализации реальной проблемной ситуации и приведения ее к виду, пригодному для использования математических методов и получения приемлемого результата.

Говоря о практической применимости методов принятия решений, следует особенно подчеркнуть, что должны существовать как объективные внешние обстоятельства, так и субъективные внутренние условия, которые побуждали бы человека — руководителя, ответственного за решение стоящей проблемы, специалиста, аналитика — искать лучшие варианты ее решения. Без такой потребности спрос на научно обоснованные методы выбора будет невелик.

Участники процесса принятия решения

Принятие решений, как уже отмечалось, есть особый вид человеческой деятельности, направленный на нахождение наилучших из возможных вариантов. Конечный результат решения проблемы определяется многими участниками, имеющими различные функции. Главное место принадлежит человеку или группе людей, которые фактически осуществляют выбор предпочтительного решения.

В теории принятия решений такого человека или группу таких людей называют лицом, принимающим решение (ЛПР), или действующим, лицом, по-английски decisionmaker (DM), actor. Обычно в роли ЛПР выступает руководитель или группа компетентных в своей области специалистов, обладающих соответствующими знаниями и опытом деятельности, наделенных необходимыми полномочиями для принятия решения и несущих ответственность за реализацию принятого решения.

Иногда целесообразно специально выделить владельца проблемы (ВП) — человека или группу лиц, имеющих основания и мотивы для постановки проблемы, осознающих необходимость ее решения, инициирующих тем или иным образом принятие и выполнение нужного решения.

В ряде случаев ВП и ЛПР могут быть одним и тем же человеком, но могут быть и разными людьми.

Важную роль в процессе принятия решений, особенно тех из них, которые затрагивают политические, социальные, экономические и другие интересы различных общественных институтов, социальных групп, крупных организаций, играют так называемые активные группы (АГ). Эти группы объединяют людей, которые имеют общие интересы по отношению к проблеме, требующей решения, и стремятся оказать влияние на процесс выбора с тем, чтобы добиться нужного им результата. Активные группы — окружение, в котором протекает процесс решения проблемы и действует ЛПР. Обычно владелец проблемы принадлежит к одной из основных активных групп. Очевидно, что

интересы разных активных групп могут, как совпадать, так и отличаться друг от друга, а также от интересов и ЛПР, и владельца проблемы.

В сложных ситуациях выбора на разных этапах процесса подготовки и принятия решения могут привлекаться эксперты (Э) и консультанты по принятию решений (К).

Эксперты (от латинского *expertus*, опытный) — компетентные специалисты, профессионально разбирающиеся в решаемой проблеме, обладающие необходимой информацией о проблеме и об отдельных ее аспектах, но не несущие ответственности за принятое решение и его реализацию.

Консультанты по принятию решений оказывают помощь ЛПР и владельцу проблемы в организации процесса ее решения, в правильной постановке задачи принятия решения, обеспечивают сбор необходимой информации, разрабатывают модель проблемы, процедуры и методы принятия решения.

Те или иные участники имеют определяющее значение на разных этапах жизненного цикла процесса решения проблемы.

Процесс принятия решения. Теория принятия решений применима к объектам различной природы и в различных условиях их существования. Вместе с тем процессы принятия решений в разных сферах человеческой деятельности имеют много общего.

Формальные методы принятия решения могут оказаться полезными в следующих случаях:

- Существует некоторая проблема или проблемная ситуация, требующая своего разрешения. Нередко желаемый результат отождествляется с одной или несколькими целями, которые должны быть достигнуты при разрешении проблемной ситуации;

- Имеется несколько вариантов решения проблемы, способов достижения цели, действий, объектов, среди которых производится выбор. Эти варианты в теории принятия решений обычно называют альтернативами. Если существует одна возможность и выбор отсутствует, то нет и задачи принятия решения;

- присутствуют факторы, накладывающие определенные ограничения на возможные пути решения проблемы, достижения цели. Эти факторы определяются контекстом решаемой проблемы и могут иметь различную природу: физическую, техническую, экономическую, социальную, персональную и иную;
- имеется человек или группа лиц, которые заинтересованы в разрешении проблемы, имеют полномочия для выбора того или иного варианта решения и несут ответственность за выполнение принятого решения.

Приведем типовую схему процесса принятия решения, устанавливающую набор и последовательность этапов при принятии решения, и обозначим основных действующих лиц этого процесса и их роли. Жизненный цикл решения проблемы состоит из нескольких стадий (рис. 1.) и представляет собой многоэтапную итеративную процедуру.

Необходимость принятия решения возникает при появлении проблемной ситуации (этап 0). В этом случае проводится выявление проблемы (этапы 1 — 3), т.е. дается содержательное описание проблемы, определяется желательный результат ее разрешения, оцениваются имеющиеся ограничения.

На следующей стадии осуществляется постановка задачи принятия решения (этапы 4 — 7). Для этого требуется определить совокупность возможных вариантов решения (альтернатив).

В зависимости от рассматриваемой проблемы число возможных вариантов решения может составлять и несколько единиц и достигать десятков, сотен и даже тысяч. Теоретически число рассматриваемых вариантов может быть и бесконечным.

Чтобы полностью описать все возможные варианты решения, обычно приходится собирать и анализировать различную информацию, относящуюся к проблеме и альтернативным способам ее решения.

Отсутствие или невозможность получения нужных сведений может сделать проблему неразрешимой. В таких случаях приходится возвращаться к исходной постановке проблемы и изменять ее описание. Вторая стадия заверша-

ется формулировкой задачи принятия решения (более подробно она будет рассмотрена в следующем разделе). Следует отметить, что детальное содержательное описание разрешаемой проблемы уже на первом этапе во многом определяет возможные подходы к ее решению и может сразу привести к постановке задачи принятия решения, минуя все или многие из последующих этапов.

Сформулировав задачу принятия решений, переходят к поиску решения (этапы 8 — 10). Эта стадия включает в себя:

- подбор известного метода решения задачи или разработка нового метода;
- во-вторых, собственно сам процесс решения, состоящий в оценке и анализе различных вариантов решения и выборе среди них наиболее предпочтительного.

В ряде задач получение окончательного результата не представляет больших трудностей. Однако чаще это достаточно сложные и трудоемкие процедуры, требующие привлечения знаний и умения многих людей и возможностей современной вычислительной техники.

Вместе с тем, даже пройдя все этапы процесса решения проблемы, не всегда оказывается возможным сделать окончательный выбор. Встречаются ситуации, когда не удастся найти лучшее решение.

Нужного варианта может просто не быть в наличии. Тогда можно либо изменить формулировку исходной проблемы (этап 11), либо возвратиться на предыдущие этапы и собрать необходимую дополнительную информацию, внести изменения в формальную постановку задачи или модель проблемной ситуации. В любом случае проделанный поиск лучшего варианта решения, даже если он не привел к положительному результату, не будет бесполезным.

Он может натолкнуть на новое понимание рассматриваемой проблемы, обратить внимание на какие-то новые аспекты, которые необходимо учесть, указать на иные пути решения задачи. Если приемлемый вариант найден, наступает стадия исполнения решения (этапы 12, 13), на которой происходит



реализация принятого решения, осуществляется контроль над процессом реализации и оценивается результат разрешения проблемной ситуации.

Рис.1. Жизненный цикл решения проблемы

Стадия исполнения решения не относится к процедуре принятия решения, тем не менее, ее включение в общую схему важно с практической точки зрения, так как эта стадия замыкает жизненный цикл процесса возникновения, разрешения и исчезновения проблемной ситуации.

Тема №2. СППР, концепция, основные термины и определения, решаемые задачи

Системы поддержки принятия решений (СППР) или DecisionSupportSystems(DSS) — это компьютерные системы, которые путем сбора и анализа большого объема информации могут влиять на процесс принятия решений организационного плана в бизнесе и предпринимательстве. Интерактивные системы позволяют руководителям получить полезную информацию из первоисточников, анализировать ее, а также выявлять существующие бизнес-модели для решения определенных задач. С помощью СППР можно проследить за всеми доступными информационными активами, получить сравнительные значения объемов продаж, спрогнозировать доход организации при предполагаемом внедрении новой технологии, а также рассмотреть возможные альтернативные решения.

СППР решает две основные задачи:

- выбор наилучшего решения из множества возможных (оптимизация);
- упорядочение возможных решений по предпочтительности (ранжирование).

В обеих задачах первым и наиболее принципиальным моментом является выбор совокупности критериев, на основе которых в дальнейшем будут оцениваться и сопоставляться возможные решения (будем называть их также альтернативами). СППР помогает пользователю сделать такой выбор.

Для анализа и выработки предложений в СППР используются разные методы. Это могут быть: информационный поиск, интеллектуальный анализ данных, поиск знаний в базах данных, рассуждение на основе прецедентов, имитационное моделирование, эволюционные вычисления и генетические алгоритмы, нейронные сети, ситуационный анализ, когнитивное моделирование и др. Некоторые из этих методов были разработаны в рамках искусственного интеллекта. Если в основе работы СППР лежат методы искусственного интеллекта, то говорят о интеллектуальной СППР или ИСППР.

Близкие к СППР классы систем — это «Экспертные системы» и автоматизированные системы управления (АСУ). Пользователь взаимодействует с СППР через пользовательский интерфейс, выбирая частную модель и набор данных, которые нужно использовать, а затем СППР представляет результаты пользователю через тот же пользовательский интерфейс.

Модель управления и управления данными действуют, в основном, незаметно и варьируются от простой модели до сложной комплексной модели планирования, основанной на математическом программировании. Наиболее широкой сферой практического применения СППР являются планирование и прогнозирование для различных видов управленческой деятельности.

Пример популярного типа СППР - СППР в виде генератора финансового отчета. С помощью электронной таблицы можно сгенерировать модели, чтобы спрогнозировать различные элементы организации или финансового состояния. В качестве данных используются предыдущие финансовые отчеты организации. Начальная модель включает различные предположения относительно будущих трендов в категориях расхода и дохода. После рассмотрения результатов базовой модели менеджер проводит ряд исследований типа «Что, если...?», изменяя одно или большее количество предположений, чтобы определить их влияние на исходное состояние.

Характеристика систем поддержки принятия решений

Системы поддержки принятия решений:

- предполагают гибкость пользователей, адаптируемость и быструю реакцию;
- допускают, чтобы пользователи управляли входом и выходом;
- функционируют с небольшой помощью профессиональных программистов или без нее;
- обеспечивают поддержку для решений и проблем, которые не могут быть определены заранее;
- используют сложный анализ и инструментальные средства моделирования.

СППР имеют большую аналитическую мощьность, чем другие системы, поскольку построены с учетом применения моделей последующего анализа данных. СППР интерактивны; пользователь может изменять исходные предположения и использовать новые данные.

Процесс принятия решений человеком, как блоком принятия решений в СППР, включает четыре стадии:

- распознавание - состоит из идентификации и понимания проблем, встречающихся в организации: почему проблемы возникают, где и с каким результатом. Традиционные управляющие информационные системы (УИС), которые поставляют широкое многообразие детальной информации, могут помогать опознавать проблемы, особенно если системы сообщают об исключениях;
- проект - в течение принятия решений лицо, принимающее решение, продумывает возможные варианты решения проблем;
- выбор - заключается в подборе решений среди альтернатив;
- реализация - в процессе реализации решения менеджеры могут использовать систему сообщений, которая формирует информационные блоки относительно прогресса определенного решения.

СППР помогают находить ответы на следующие типовые вопросы:

1. Анализ примеров (caseanalysis) - оценка значений выходных величин для заданного набора значений входных переменных.
2. Параметрический анализ {«Что, если... ?»)- оценка поведения выходных величин при изменении значений входных переменных.
3. Анализ чувствительности - исследование поведения результирующих переменных в зависимости от изменения значений одной или нескольких входных переменных.
4. Анализ возможностей - нахождение значений входной переменной, которые обеспечивают желаемый результат (известен также под названием «поиск целевых решений», «анализ значений целей», «управление по целям»).

5. Анализ влияния - выявление для выбранной результирующей переменной всех входных переменных, влияющих на ее значение, и оценка величины изменения результирующей переменной при заданном изменении входной переменной.
6. Анализ данных - прямой ввод в модель ранее имевшихся данных и манипулирование ими при прогнозировании.
7. Сравнение и агрегирование - сравнение результатов двух или более прогнозов, сделанных при различных входных предположениях, или сравнение предсказанных результатов с действительными, или объединение результатов, полученных при различных прогнозах или для разных моделей.
8. Командные последовательности (sequences) - возможность записывать, исполнять, сохранять для последующего использования регулярно выполняемых команд и сообщений.
9. Анализ риска - оценка изменения выходных переменных при случайных изменениях входных величин.
10. Оптимизация - поиск значений управляемых входных переменных, обеспечивающих наилучшее значение одной или нескольких результирующих переменных.

История. Теоретические исследования в области разработки систем поддержки принятия решений проводились в технологическом институте Карнеги в конце 50-х начале 60-х годов XX века. Объединить теорию с практикой удалось специалистам из Массачусетского технологического института в 60-х годах. В середине и конце 80-х годов XX столетия стали появляться такие системы, как EIS, GDSS, ODSS. В частности, компания Texas Instruments разработала для United Airlines Gate Assignment Display System, что позволило значительно снизить убытки от полетов и отрегулировать управление различными аэропортами. Сфера возможностей СППР расширялась благодаря внедрению концепции хранилищ данных и инструментов OLAP.

Появление новых технологий отчетности сделало СППР незаменимой в менеджменте.

Классификации. По взаимодействию с пользователем выделяют три вида СППР:

- пассивные - помогают в процессе принятия решений, но не могут выдвинуть конкретного предложения;
- активные - непосредственно участвуют в разработке правильного решения;
- кооперативные - предполагают взаимодействие СППР с пользователем, при этом выдвинутое системой предложение пользователь может доработать, усовершенствовать, а затем отправить обратно в систему для проверки. После этого предложение вновь представляется пользователю, и так до тех пор, пока он не одобрит решение.

По способу поддержки различают:

- модельно-ориентированные СППР, используют в работе доступ к статистическим, финансовым или иным моделям;
- СППР, основанные на коммуникациях, поддерживают работу двух и более пользователей, занимающихся общей задачей;
- СППР, ориентированные на данные, имеют доступ к временным рядам организации. Они используют в работе не только внутренние, но и внешние данные;
- СППР, ориентированные на документы, манипулируют неструктурированной информацией, заключенной в различных электронных форматах;
- СППР, ориентированные на знания, предоставляют специализированные решения проблем, основанные на фактах.

По сфере использования выделяют общесистемные и настольные СППР.

- Общесистемные работают с большими системами хранилищ данных и применяются многими пользователями.

- Настольные являются небольшими системами и подходят для управления с персонального компьютера одного пользователя.

В архитектуре СППР выделяют четыре основных компонента:

— **Информационные хранилища данных**

Хранилище данных (*Data Warehouse*) — предметно-ориентированная информационная база данных, специально разработанная и предназначенная для подготовки отчётов и бизнес-анализа с целью поддержки принятия решений в организации. Строится на базисе систем управления базами данных и систем поддержки принятия решений. Данные, поступающие в хранилище данных, как правило, доступны только для чтения. Основу хранилищ данных на первых этапах разработки составляют данные из OLTP (*Online Transaction Processing*), транзакционных систем.

OLTP-системы предназначены для ввода, структурированного хранения и обработки информации (операций, документов) в режиме реального времени. Информационные ресурсы OLTP-систем копируются в хранилище данных таким образом, чтобы построение отчётов и OLAP-анализ не использовал ресурсы транзакционной системы и не нарушал её целостность. Как правило, данные загружаются в хранилище с определённой периодичностью, поэтому актуальность данных может несколько отставать от OLTP-системы.

— **Средства и методы извлечения, обработки и загрузки данных (ETL);**

ETL (от англ. *Extract, Transform, Load* — дословно «извлечение, преобразование, загрузка») — один из основных процессов в управлении хранилищами данных, который включает в себя:

- извлечение данных из внешних источников;
- их трансформация и очистка, чтобы они соответствовали нуждам бизнес-модели;
- загрузка в хранилище данных.

С точки зрения процесса ETL, архитектуру хранилища данных можно представить в виде трёх компонентов:

- источник данных: содержит структурированные данные в виде таблиц, совокупности таблиц или просто файла (данные в котором разделены символами-разделителями);
- промежуточная область: содержит вспомогательные таблицы, создаваемые временно, и, исключительно для организации процесса выгрузки.
- получатель данных: хранилище данных или база данных, в которую должны быть помещены извлечённые данные.

— **Многомерная база данных и средства анализа OLAP;**

OLAP(*online analytical processing*, аналитическая обработка в реальном времени) — технология обработки данных, заключающаяся в подготовке суммарной (агрегированной) информации на основе больших массивов данных, структурированных по многомерному принципу. Реализации технологии OLAP являются компонентами программных решений класса Business Intelligence.

— **Средства Data Mining.**

Data Mining (*добыча данных, интеллектуальный анализ данных, глубокий анализ данных*) — собирательное название, используемое для обозначения совокупности методов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

Английское словосочетание «*Data Mining*» пока не имеет устоявшегося перевода на русский язык. При передаче на русском языке используются следующие словосочетания: *просев информации, добыча данных, извлечение данных*, а также, **интеллектуальный анализ данных**. Более полным и точным является словосочетание «*обнаружение знаний в базах данных*» (англ. *knowledge discovery in databases*, KDD).

Основу методов Data Mining составляют всевозможные методы классификации, моделирования и прогнозирования, основанные на применении деревьев решений, искусственных нейронных сетей, генетических алгоритмов,

эволюционного программирования, ассоциативной памяти, нечёткой логики. К методам Data Mining также часто относят *статистические методы* (описательный анализ, корреляционный и регрессионный анализ, факторный анализ, дисперсионный анализ, компонентный анализ, дискриминантный анализ, анализ временных рядов, анализ выживаемости, анализ связей). Такие методы, однако, предполагают некоторые априорные представления об анализируемых данных, что несколько расходится с целями *Data Mining* (обнаружение ранее неизвестных нетривиальных и практически полезных знаний).

Одно из важнейших назначений методов Data Mining состоит в наглядном представлении результатов вычислений (визуализация), что позволяет использовать инструментарий Data Mining людьми, не имеющими специальной математической подготовки. В то же время, применение статистических методов анализа данных требует хорошего владения теорией вероятностей и математической статистикой.

Преимущества. СППР позволяет облегчить работу руководителям предприятий и повысить ее эффективность. Они значительно ускоряют решение проблем в бизнесе. СППР способствуют налаживанию коммуникаций. На их основе можно проводить обучение и подготовку кадров. Данные информационные системы позволяют повысить контроль над деятельностью организации. Наличие СППР дает большие преимущества по сравнению с конкурирующими структурами. Благодаря предложениям, выдвигаемым СППР, открываются новые подходы к решению повседневных и нестандартных задач.

Тема №3. Формирование баз моделей и систем управления моделями в СППР

- метод линейной оптимизации,
- транспортные задачи и логистика;
- задачи о назначениях и отборе;

В рамках данной темы рассмотрим технологии использования количественных моделей и методов, цель которых — найти оптимальную стратегию управления в условиях, когда все параметры и правила функционирования управляемой системы четко определены и не подвержены случайным воздействиям. В некоторых ситуациях случайные воздействия на процесс управления не учитываются, потому что они малы и несущественны, в других — случайные факторы, которые могут оказать влияние на нашу деятельность (поломки оборудования, катастрофы, социальные потрясения и т. п.) проявляются достаточно редко. Для начала рассмотрим метод линейной оптимизации, позволяющий в рамках ее моделей рассматривать задачи, цель которых — составление оптимальных планов. Речь может идти об оптимальных планах производства, продаж, закупок, перевозок, об оптимальном финансовом планировании, оптимальной организации рекламной кампании или об оптимальном плане инвестиционного портфеля фирмы.

Как известно, планирование — одна из основных функций менеджмента. При постановке любой задачи оптимизации необходимо определить количественную характеристику цели — целевую функцию. Это может быть максимум прибыли или минимум издержек (в денежном, временном или каком-либо другом выражении). Целевая функция показывает, почему одно рассматриваемое решение лучше или хуже другого, зависит от величин, называемых переменными решения. Эти величины можно изменять, разыскивая оптимальное решение. Цель оптимизации — найти такие значения переменных решения, при которых целевая функция максимальна или минимальна.

Любая оптимизация всегда проводится при наличии некоторых ограничений — условий, ограничивающих изменение переменных решения при поиске максимальной или минимальной целевой функции. Эти ограничения могут диктоваться вторичными целями (например, минимизируя риск инвестиционного портфеля, необходимо одновременно добиться ожидаемой прибыли не хуже заданной), ограниченностью ресурсов, установленными “правилами игры” (рыночные ограничения, нормативные акты, лимитирующие ту или иную характеристику или любые требования субъекта, принимающего решения).

Линейная оптимизация имеет дело с моделями, в которых целевая функция линейно зависит от переменных решения и ограничения представляются линейными уравнениями или неравенствами относительно переменных решения. Фактически это означает, что целевая функция и ограничения могут представлять собой только суммы произведений постоянных коэффициентов на переменные решения в первой степени, т. е. выражения типа:

$$c_1x_1 + c_2x_2 + \dots + c_nx_n.$$

Важность моделей линейной оптимизации обусловлена тем, что очень много важных для практики проблем, относящихся к самым разным сферам деятельности, могут быть проанализированы с помощью моделей линейного программирования; существуют эффективные и универсальные алгоритмы решения задач линейной оптимизации, реализованные в общедоступном программном обеспечении; методы анализа моделей линейной оптимизации не только позволяют получить оптимальное решение, но и дают информацию о том, как может измениться это решение при изменении параметров модели. Именно эта информация, позволяющая получить ответы на вопросы типа “что, если”, представляет особую ценность для лица, принимающего решение.

Конечно, модели с нелинейными соотношениями между переменными также важны для практики. Однако в отличие от моделей линейной оптимизации не существует универсального алгоритма, который бы во всех случаях гарантированно приводил к искомому оптимуму. Поэтому для проведе-

ния нелинейной оптимизации требуется уделить больше внимания деталям алгоритма и его реализации, чем обычно может уделить менеджер.

С точки зрения менеджера, транспортные задачи — это любые задачи, связанные оптимизацией перевозок. С точки зрения специалиста по исследованию операций, транспортная задача — это специальный тип задачи линейной оптимизации, для которой существуют эффективные алгоритмы решения.

Классическая транспортная задача имеет цель - минимизация транспортных издержек при перевозках однотипных грузов от нескольких поставщиков, расположенных в разных местах, к нескольким потребителям. При этом в транспортной задаче принимают в расчет только переменные транспортные издержки, т. е. считают, что суммарные издержки пропорциональны количеству перевезенных единиц груза. При постановке транспортной задачи задается таблица транспортных издержек для перевозки единицы груза C_{ij} от i -го поставщика к j -му потребителю. Эта таблица имеет m строк (число поставщиков) и n столбцов (число потребителей). Таблица перевозок имеет те же размеры ($m \times n$) и содержит переменные решения - X_{ij} . При этом задаются объемы запасов поставщиков и величины заказов потребителей. В транспортной задаче предполагается, что необходимо вывести запасы каждого i -го поставщика и удовлетворить заказ каждого j -го потребителя. Это возможно, если сумма запасов всех поставщиков равна сумме заказов всех потребителей, что является условием сбалансированности.

Ограничения транспортной задачи имеют вид: сумма переменных решения вдоль каждой i -й строки должна быть равна запасу поставщика S_i , а сумма переменных решения вдоль каждого j -го столбца — заказу соответствующего потребителя D_j . Для получения целевой функции (суммарные издержки), необходимо рассчитать суммы произведений каждой строки таблицы транспортных издержек на соответствующую строку таблицы перевозок и сложить их, суммируя по i от 1 до m . При этом номер поставщика $1 < i < m$, номер потребителя $1 < j < n$.

$$(C_{ij}) = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \dots & \dots & \dots & \dots \\ c_{m1} & c_{m2} & \dots & c_{mn} \end{pmatrix}, \quad (x_{ij}) = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix}, \quad (S_j) = \begin{pmatrix} S_1 \\ S_2 \\ \dots \\ S_m \end{pmatrix}$$

$$D_j = (D_1, D_2, \dots, D_n)$$

$$\sum_{j=1}^n (x_{ij} - S_j) = 0, \quad \sum_{i=1}^m (x_{ij} - D_j) = 0, \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n$$

$$\text{Min: } = C = \sum_{i=1}^m \sum_{j=1}^n x_{ij} c_{ij}$$

Если задача сбалансирована и никаких других ограничений, кроме упомянутых выше, нет, то количество ненулевых перевозок X_j не будет превышать $(m+n-1)$.

Несбалансированность в транспортной задаче

Если сумма запасов превышает сумму заказов (излишек запасов) или, наоборот, сумма запасов меньше, чем сумма заказов (дефицит запасов), необходимо сбалансировать задачу. В первом случае, когда

$$\sum_{i=1}^m S_i > \sum_{j=1}^n D_j,$$

нужно добавить в таблицу транспортных издержек и в таблицу перевозок по одному лишнему столбцу. Это можно трактовать так, как если бы появился еще один, “фиктивный” потребитель. Если потребовать, чтобы заказ этого “потребителя” в точности равнялся разности между суммой всех запасов и суммой всех заказов

$$D_{\text{фict}} = \sum_{i=1}^m S_i - \sum_{j=1}^n D_j,$$

а издержки перевозок грузов от него к любому поставщику равны нулю. Вновь имеем сбалансированную транспортную задачу. При этом переменные решения в лишней строке — это тот объем грузов, который не получит каждый потребитель.

Еще одно возможное осложнение транспортной задачи — это запрещение определенной перевозки от i -го поставщика к j -му потребителю для составляемого плана перевозок. В этом случае, естественно, можно просто вве-

сти ограничение $X_{ij}=0$. Однако это вновь означает невозможность использования эффективных “транспортных” алгоритмов решения.

Чтобы сохранить форму транспортной задачи и учесть этот запрет, достаточно в таблице транспортных издержек заменить C_{ij} на очень большое число (на порядок большее, чем максимальная цена перевозки в таблице транспортных издержек). Это фактически будет означать, что оптимизационный алгоритм наверняка положит соответствующее значение перевозки $X_{ij}=0$, поскольку перевозка по этому маршруту просто крайне невыгодна.

Во втором случае, когда

$$\sum_{i=1}^m S_i \neq \sum_{j=1}^n D_j,$$

нужно добавить в таблицу транспортных издержек и в таблицу перевозок по одной лишней строке. Это можно трактовать так, как если бы появился еще один, “фиктивный” поставщик. Потребуем, чтобы запас этого “поставщика” в точности равнялся бы разности между суммой всех заказов и суммой всех запасов

$$D_{\text{фict}} = \sum_{j=1}^n D_j - \sum_{i=1}^m S_i$$

Задача о назначениях. Задача о назначениях — это модель для количественного анализа ситуаций, когда менеджер должен назначить рабочих для выполнения различных производственных операций, распределить ряд производственных заданий по различным машинам (которые могут эти задания выполнить с различной эффективностью) или решить, какого торгового агента в какую область послать для продвижения продукции фирмы. Это распределение или назначение должно быть сделано из соображений либо наибольшей эффективности, либо наименьших затрат. С математической точки зрения задача о назначениях — это частный случай транспортной задачи, в которой число поставщиков (например, число рабочих или, иначе, поставщиков рабочей силы) в точности равно числу потребителей (работ, различных технологических операций). Поэтому таблица “транспортных издержек” (аналогом которых

может выступать любая мера эффективности выполнения той или иной операции данным работником) должна быть квадратной.

Кроме того, в задаче о назначениях от каждого поставщика к каждому потребителю поставляется только одна единица “груза” (например, только одного рабочего можно назначить для выполнения данной работы) или ни одной. Поэтому все “запасы” и все “заказы” равны 1. Все переменные решения в задаче о назначениях могут принимать только значения 1 или 0. Такие значения получаются при решении автоматически. При этом, разумеется, “транспортные” алгоритмы решения гораздо более эффективны, чем алгоритмы решения задач целочисленной линейной оптимизации.

Задача о назначениях также может быть несбалансированной, если количество рабочих (претендентов на работы) не равно количеству работ. Так же как и в случае транспортной задачи, это осложнение разрешается добавлением дополнительного столбца и строки (фиктивной работы, если претендентов больше, чем работ, или фиктивного рабочего, если наоборот).

Задачи оптимизации логистики и цепочек поставок. Задачи, возникающие в деятельности отдела логистики, часто гораздо сложнее и разнообразнее, чем простая транспортная задача, хотя последняя очень часто может входить в них как составная часть. Часто алгоритмы решения логистических задач вообще не имеют ничего общего с транспортной задачей. Например, популярная практическая задача о выборе оптимального маршрута объезда нескольких клиентов сводится к весьма сложной задаче целочисленной линейной оптимизации.

Аналогично задачи о выборе оптимального поставщика или задачи о назначениях с дополнительными условиями потребуют явного введения условия целочисленности.

Тема №4. Работа с оптимизационными моделями применимыми в процессах принятия решений: оптимальное управление запасами

В рамках данной лекции введем необходимые обозначения и формулы, которые будут нами применяться в процессах вычислений:

Q — объем заказа, количество единиц;

EOQ — экономичный размер заказа (economic order quantity);

n — число заказов в год;

D, Di — годовой спрос, количество единиц;

S — затраты переналадки или издержки заказа;

C — стоимость единицы товара, изделия;

h — затраты хранения в год, процентов от стоимости

H — затраты хранения на единицу в год, денежных единиц;

p — скорость производства, штук в единицу времени;

d — скорость потребления, штук в единицу времени

L — время выполнения заказа, доставки и т.п.

T — время выполнения заказа, доставки и т.п.

I — наличие товара на складе, количество единиц;

ROP — точка перезаказа (reorder point);

SS — страховой запас, безопасный резерв (safety stock);

Экономичный размер заказа:

$$EOQ = \sqrt{\frac{2DS}{H}}$$

Годовые издержки хранения:

$$TH = H \frac{Q}{2}$$

Годовые издержки заказа:

$$TS = \frac{D}{Q} \cdot S$$

Полные годовые издержки:

$$T_{\text{полные}} = TH + TS + D \cdot S$$

Экономичный размер партии продукции:

$$EBQ = \sqrt{\frac{2DS}{H} \cdot \frac{p}{p-d}}$$

Годовые издержки хранения:

$$TH = H \frac{Q}{2} \cdot \frac{p-d}{p}$$

Оптимальная частота заказов для группы товаров из m наименований^

$$n = \sqrt{\frac{\sum_{i=1}^m D_i h_i C_i}{2S}}, \text{ где } i - \text{ номер товара в списке.}$$

Оптимальные размеры заказов для товаров из группы:

$$Q_i = \frac{D_i}{n}$$

Точка перезаказа:

$$ROP = dL + SS$$

Введем небольшие теоретические замечания.

Задачи управления запасами, на первый взгляд, достаточно просты: какова должна быть величина товарного запаса на складе, чтобы минимизировать издержки по управлению запасами и обеспечить достойный уровень обслуживания клиента? При этом задачи разделяется на две части:

Как сделать издержки управления запасами минимальными при заданном (постоянном или непостоянном, но известном) спросе?

Как оценить риск возникновения дефицита на складе с учетом случайных вариаций реального спроса? Сколько нужно платить за содержание необходимого резервного запаса для того, чтобы снизить риск возникновения дефицита до приемлемого уровня и обеспечить достойный уровень обслуживания клиентов?

Первая часть вопроса относится к задачам принятия решений в условиях полной определенности. Вторая часть, предусматривает детальный анализ характеристик случайного спроса.

Важнейшая функция запасов состоит в том, что они играют роль буфера, смягчающего удары, испытываемые фирмой в результате нестабильных поставок сырья или товаров от поставщиков или сильных вариаций потребительского спроса на тот или иной продукт. В производственном процессе, запасы незавершенной продукции (полуфабрикатов) необходимы для обеспечения независимости различных производственных операций. Поддержание большого уровня запасов позволяет реже их восполнять, тратить меньше времени менеджеров на формирование заказа, его оформление, контроля доставки новой партии товара. Все это толкает менеджеров, непосредственно отвечающих за наличие запасов продукции на складах фирмы, увеличивать уро-

вень этих запасов. С запасами связаны специфические издержки хранения, размер которых может составлять 20-40% в год от стоимости среднегодового уровня запаса, при чем, что особенно существенно, большая их часть не проходит через бухгалтерию, как прямые затраты, а является, так называемыми, «упущенными возможностями». Несмотря на то, что их «сразу не видно», большие упущенные возможности, приводят фирму к банкротству так же быстро, как и большие прямые затраты.

Основная идея теории оптимального управления запасами состоит в том, чтобы разделить издержки на переменные и постоянные. Оказывается, что эти две группы издержек по-разному зависят от размера заказа и уровня запаса товара на складе.

Рассмотрим природу этих издержек и пути их оптимизации.

Переменные издержки- издержки хранения должны быть прямо пропорциональны количеству единиц хранимых запасов и стоимости единицы запаса. Основную часть этих издержек составляют упущенные возможности при альтернативном использовании капитала, «замороженного» в запасах. Каждая область бизнеса характеризуется своей требуемой нормой доходности. Капитал, вложенный в этот бизнес, в среднем (по стране, региону, городу) должен давать определенный процент дохода ежегодно. Капитал, вложенный в запасы, такого процента не дает.

Следовательно, неполученный процент – это издержка хранения. Если товар приобретен в кредит, то за этот кредит нужно платить проценты, что опять-таки составляет издержки хранения. При цивилизованном ведении бизнеса, товар должен быть застрахован и подлежит налогообложению. Страховка и налог на запас также составляет определенный процент от стоимости товара и также входит в издержки хранения. Перечисленные издержки строго пропорциональны стоимости запасов. Поэтому их удобно задавать в расчете на единицу запаса в год. Мы будем использовать для обозначения таких удельных издержек хранения либо большую букву Н (от английского термина Holdingcost), полагая, что размерность этой величины

[H] - денежная единица/(единица запаса * в год)

либо маленькую букву h , полагая, что это процент от стоимости единицы запаса C при хранении этой единицы в течение года.

Тогда, суммарные предельные издержки хранения всегда будут пропорциональны количеству хранимых единиц запаса и времени хранения, а коэффициентом пропорциональности, как раз, будет H .

Разумеется, можно относить издержки хранения к любому временному интервалу (неделя, квартал, год). В практике торговых складов чаще в качестве базового временного интервала используется именно год. На производстве, это могут быть и другие, более короткие интервалы (неделя).

Главное из чего надо исходить при решении включать те или иные складские затраты в величину H или не включать - это условие (хотя бы приблизительное) пропорциональности суммарных издержек хранения количеству хранимых единиц данного запаса и времени их хранения.

Например, в издержки хранения можно включить потери от распродажи «залежалого товара» по сниженным ценам. Правда, оценить вклад этих издержек в величину H сложнее, поскольку потери от снижения цены продаж «залежавшейся» части купленной партии, нужно распределить на всю партию (чтобы сохранилась пропорциональность издержек хранения количеству хранимых единиц запаса). Однако, при длительном ведении бизнеса, средний процент от стоимости купленной партии, соответствующий этому виду издержек может быть оценен более или менее определенно.

Постоянные издержки - издержки по запуску новой партии продукции - (производство) или затраты на формирование и оформление заказа - (торговля). Эти издержки не зависят от величины предполагаемой партии продукции (заказа). В торговле их чаще всего связывают, с оплатой труда менеджеров, «ведущих» этот заказ, с возможными затратами на сопровождение заказа сотрудником фирмы (контроль погрузки – разгрузки, ускорение прохождения оформления документов на таможне и т.п.), с офисными расходами при

оформлении и размещении заявки поставщику на новый заказ и другими сопутствующими расходами.

В производстве постоянным издержкам соответствуют затраты на переналадку оборудования для выпуска данной партии продукции (устойчивый английский термин – Setupcost). Величину этих издержек, в расчете на один заказ (или на одну переналадку производственной линии) принято обозначать буквой S .

Размерность этой величины $[S]$ – денежная единица/на один заказ.

Эти издержки постоянные в том смысле, что S не зависит от размера партии продукции данного наименования. Однако, чем больше размер заказа, тем реже приходится оплачивать расходы на его оформление, тем меньше затраты на оформление заказов (или на переналадку оборудования) за выбранный базовый период (год, неделя и т.п.).

При ведении бухгалтерского учета, в издержки хранения включают прямые расходы на содержание склада: амортизация здания (или аренда), оплата персонала, охрана и т.п. Хотя перечисленные издержки, несомненно, относятся к категории складских издержек, при анализе оптимизационных моделей управления запасами, их не следует включать ни в величину H , ни в величину S . Дело в том, что все эти издержки являются интегральными. Они совершенно не зависят от размера закупленной и хранимой партии данного товара. Даже если склад пустой, фирма все равно несет эти издержки. Они не зависят от того, какие именно товары, и в каком количестве хранятся на складе.

Разумеется, если принято решение существенно снизить размер товарных запасов, то для их хранения понадобится меньше складских площадей, и следовательно уменьшатся затраты на их содержание. Прямые складские издержки, таким образом, влияют на рентабельность торговой фирмы и должны учитываться при оценке эффективности работы склада. Они определяют решения, связанные с определением размеров складов, ассортимента продуктов, с которым должна работать фирма и т.п. Однако, такие решения принимаются гораздо реже, чем решение о размере покупаемой партии продукции данного

наименования, которое и является предметом рассмотрения оптимизационных моделей управления запасами. Размер склада и ассортимент продуктов в таких задачах не могут рассматриваться как переменные решения. Поэтому на результаты оптимизации уровней запасов и размеров заказа, указанные выше оказывают лишь косвенное влияние, и прямо в них фигурировать не могут.

Рассмотрим модель экономичного размера заказа, которая является одной из первых моделей количественного менеджмента. До сих пор данная модель остается практически инструментом при управлении запасами. Приведем основные допущения и параметры модели.

Модель отвечает на вопрос—какой должен быть размер заказа (и как часто его нужно делать) для данного вида товара («артикула»), что минимизировать издержки его хранения, при условии, что:

- спрос на запас постоянен и составляет D единиц в год;
- закупочная цена единицы запаса постоянна и равна C ;
- издержки хранения единицы запаса в год равны H (или $h\%$ от стоимости единицы запаса C);
- стоимость оформления одного заказа (или стоимость переналадки оборудования для начала новой партии продукции) равна S .

Допущения, сформулированные в первом и втором пунктах, являются сильным упрощением по сравнению с реальным бизнесом, мы их примем, чтобы получить ответ на поставленный вопрос в виде простой формулы, которая может служить полезным ориентиром и в более реальных ситуациях. Затем эти упрощения можно отбросить и проанализировать более реальные ситуаций с помощью тех или иных вычислительных инструментов.

На Рис. 777 показано как меняется в принятой модели товарный запас данного артикула. Если в начальный момент времени на склад приходит новая партия данного товара Q , то с течением времени, его товарный запас уменьшается с постоянной скоростью на d единиц в день, и через некоторое время обращается в ноль.

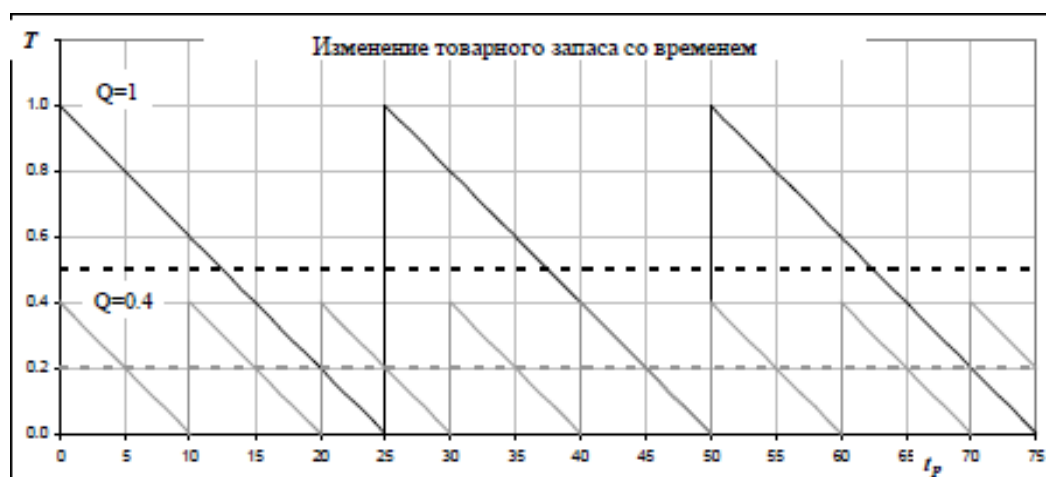


Рис. 777. Изменения товарного спроса во времени

Однако, если заблаговременно сделать заявку на такую же по величине новую партию товара, и при этом «подгадать» так, чтобы она пришла как раз тогда, когда весь запас этого артикула на складе исчерпан, товарный запас снова подскочит до величины Q , снова будет уменьшаться с постоянной скоростью и т.д. Если ежедневный спрос на данный товар d , а время выполнения заявки поставщиком L (от английского термина *Leadtime*), то новую заявку нужно делать, очевидно, тогда, когда на складе осталось $d \cdot L$ единиц запаса данного артикула. Если каждый раз заказывать партию одного и того же размера, то при годовом спросе D нужно повторить этот цикл D/Q раз. Важно понять, что годовой спрос отнюдь не определяет размер закупаемой партии Q . Можно закупать редко и большими партиями, а можно часто и малыми. В сумме за отраженный на графике период и в первом и во втором случае куплено одно и то же количество товара. Так, что за год и та, и другая стратегия удовлетворят потребность клиентов в этом товаре. Однако оказывается, что складские издержки при этом будут разными. Действительно, средний уровень товарного запаса на складе в первом случае составляет 0,5 единиц, а во втором – 0,2 условные единицы (Рис. 777).

Издержки хранения этого товара за год будут различны. В общем случае можно, очевидно, написать, что если закупается партия товара величиной Q , и этот запас линейно уменьшается до нуля, то его средний уровень равен $Q/2$. Тогда, годовые издержки хранения равны TH (см. формулу в обозначениях).

Чем меньше заказываемая партия товара Q , тем меньше издержки хранения за год. При $Q \rightarrow 0$, издержки хранения нулевые. Однако, чем меньше размер партии, тем чаще нужно делать заказ, и, следовательно, тем больше издержки, связанные с оформлением заказа. Нетрудно понять, что поскольку для удовлетворения годового спроса D на данный товар с помощью заказов по Q единиц необходимо D/Q заказов, годовые издержки на оформление заказов составят TS (см. формулу в обозначениях).

Соответственно, полные складские издержки за год составят:

$$TH = H \frac{Q}{2} + \frac{DS}{Q}$$

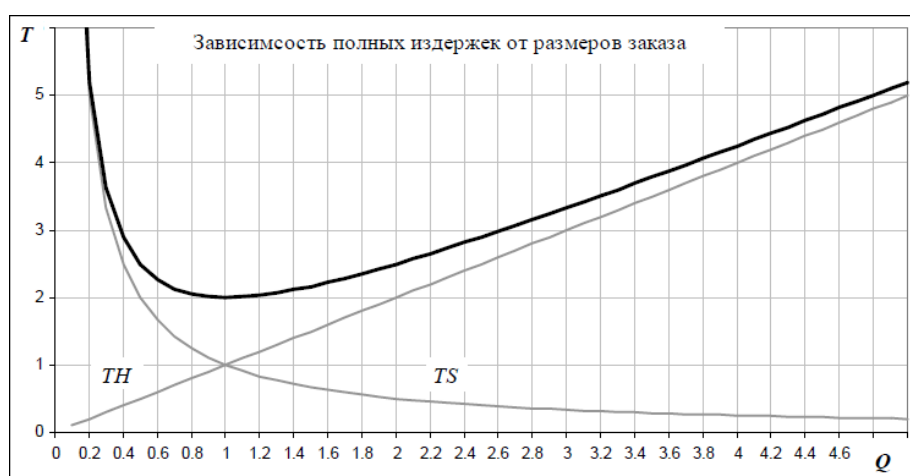


Рис.778. Зависимость полных издержек от размера заказ

На Рис. 778 показан график зависимости этих издержек T от величины заказа Q (а также показано, как изменяются величины TH и TS). Видно, что первое слагаемое в сумме T (издержки хранения за год) линейно растет с ростом величины заказа Q , в то время как второе слагаемое убывает обратно пропорционально Q . Понятно, что сумма T имеет минимум. Величину заказа, соответствующего этому минимуму обозначают как EOQ (сокращение от английского термина *EconomicOrderQuantity*). Это и есть оптимальный (или экономичный) размер заказа, обеспечивающий минимум полных складских издержек.

Подставив значение EOQ в выражение для годовых издержек хранения TH , оформления заказа TS и полных издержек T_{min} , получим:

$$TH=TS=\frac{\sqrt{2DSH}}{2} \Rightarrow T_{min}=\sqrt{2DSH}.$$

Таким образом, при экономичном размере заказа годовые издержки хранения и оформления заказа равны друг другу, а полные издержки – в два раза больше.

Краткую справку по определению оптимальной частоты заказа для группы товаров с помощью ЭВМ, использованию предлагаемой модели производства оптимальной партии продукции, имеющимся при этом ограничениям модели экономичного размера заказа (партии продукции) и возможностях их преодоления можно получить использованной нами источнике[///].

№5. Концептуальные основы СППР, архитектура СППР

В процессе своей деятельности промышленные предприятия, корпорации, ведомственные структуры, органы государственной власти и местного самоуправления накопили большие объемы данных в базах данных АИС. Эти данные хранят в себе большие потенциальные возможности по извлечению полезной аналитической информации, на основе которой можно выявлять скрытые тенденции, строить стратегию развития, находить новые решения.

Реализация потенциальных возможностей существующих АИС привело к формированию полноценного рынка технологий бизнес-анализа (Business Intelligence, BI) - категория технологий сбора, хранения, анализа и публикации данных, позволяющая корпоративным пользователям извлекать аналитическую информацию и принимать обоснованные решения. В русскоязычной терминологии подобные системы называются системами поддержки принятия решений (СППР).

Современные СППР (Decision Support System, DSS), возникшие как естественное развитие автоматизированных систем управления и систем управления базами данных, представляют собой системы, приспособленные к решению задач управленческой деятельности, являются инструментом, призванным оказать помощь ЛПР в решении неструктурированных и слабоструктурированных многокритериальных задач.

Развитие систем информационно-аналитической поддержки принятия решений, ориентированных на формирование процессов управления в информационных системах промышленных предприятия, корпораций, ведомственных структур, органов государственной власти и местного самоуправления показано на рисунке 1.

В настоящее время нет общепринятого определения СППР, поскольку архитектура СППР существенно зависит от вида решаемых задач, от доступных данных и правил их использования (знаний), а также от пользователей системы.

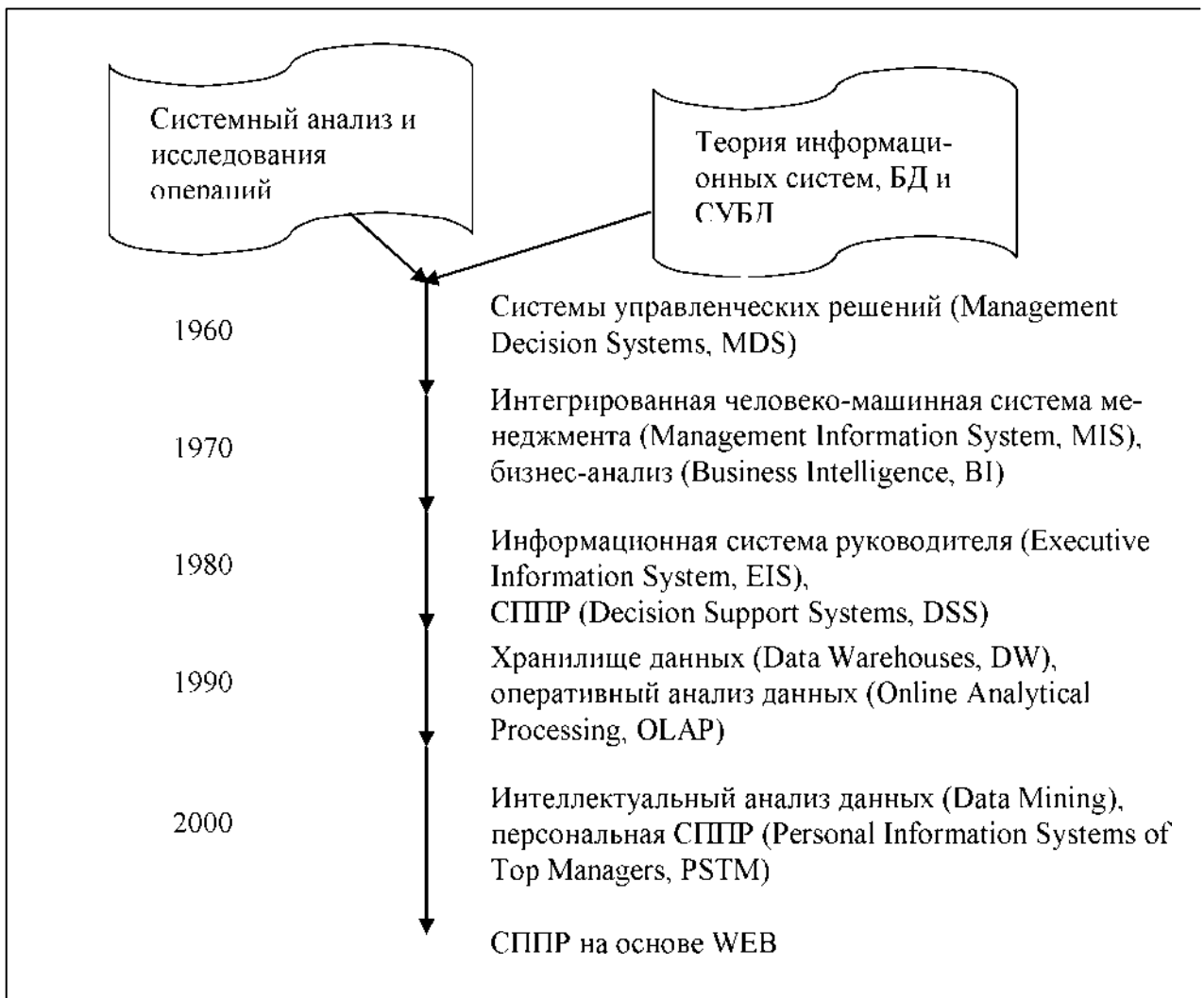


Рис.1. Развитие систем информационно-аналитической поддержки управленческих решений

Приведем некоторые свойства, общепризнанные для СППР:

- использование данных и моделей;
- решение слабоструктурированных и неструктурированных задач;
- поддерживают, а не заменяют, выработку решений ЛПР;
- целенаправленны на повышение эффективности (оперативность и обоснованность и др.) решений, обеспечивающих потенциальные возможности объекта управления.

Определение. СППР - диалоговая автоматизированная система, использующая правила принятия решений на основе моделей и баз данных, а также интерактивный компьютерный процесс их взаимодействия.

На рисунке 2 приведена архитектурно-технологическая схема информационно-аналитической поддержки принятия решений.

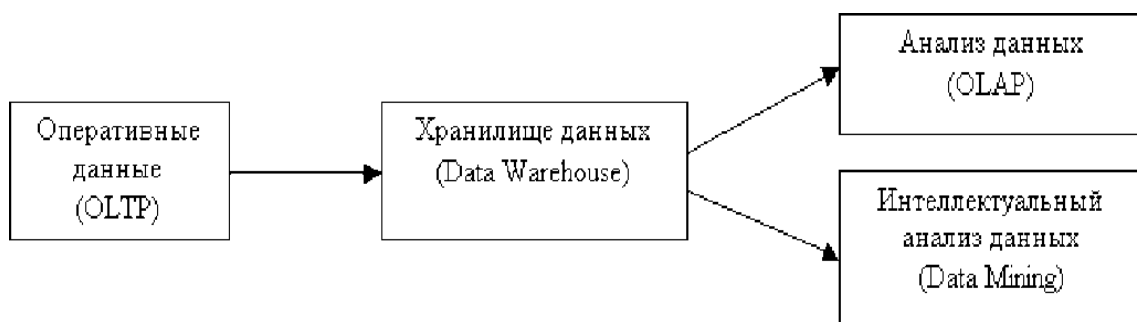


Рисунок 2 - Архитектурно-технологическая схема информационно-аналитической поддержки принятия решений

Сбор и хранение информации, а также решение информационно-поисковых задач средствами систем управления базами данных (СУБД), осуществляется в OLTP (Online Transaction Processing)-подсистемах, реализующих транзакционную обработку данных (группа последовательных операций, которая представляет собой логическую единицу работы с данными). Для предоставления необходимой для принятия решений информации обычно приходится собирать данные из нескольких транзакционных баз данных различной структуры и содержания. Основная проблема при этом состоит в несогласованности и противоречивости этих баз-источников, отсутствии единого логического взгляда на корпоративные данные. Поэтому для объединения в одной системе OLTP и СППР для реализации подсистемы хранения используется концепция хранилищ данных (ХД). В основе концепции ХД лежит идея разделения данных, используемых для оперативной обработки и для решения задач анализа, что позволяет оптимизировать структуры хранения. ХД интегрирует ранее разрозненные данные, содержащиеся в архивах, накапливаемых OLTP-системами из внешних источников, в единую базу данных, осуществляя предварительное согласование и агрегацию.

Подсистема анализа может быть построена на основе:

— подсистемы информационно-поискового анализа на базе реляцион-

ных СУБД и статических запросов с использованием языка SQL;

— подсистемы оперативного анализа. Для реализации таких подсистем применяется технология оперативной аналитической обработки данных OLAP, использующая концепцию многомерного представления данных;

— подсистемы интеллектуального анализа, реализующие методы и алгоритмы DataMining.

Архитектура СППР специалистами представляется по-разному. На рисунке 3 представлена схема, состоящая из следующих элементов: системы управления данными (thedatamanagementsystem, DBMS), системы управления моделями (themodelmanagementsystem, MBMS), машина знаний (theknowledgeengine, KE), интерфейс пользователя (theuserinterface) и пользователей - ЛПР (theuser(s)).

Для СППР отсутствует не только единое общепринятое определение, но и исчерпывающая классификация. Обобщая существующие взгляды, классификация СППР представляется следующим образом (см. таблицу 1).

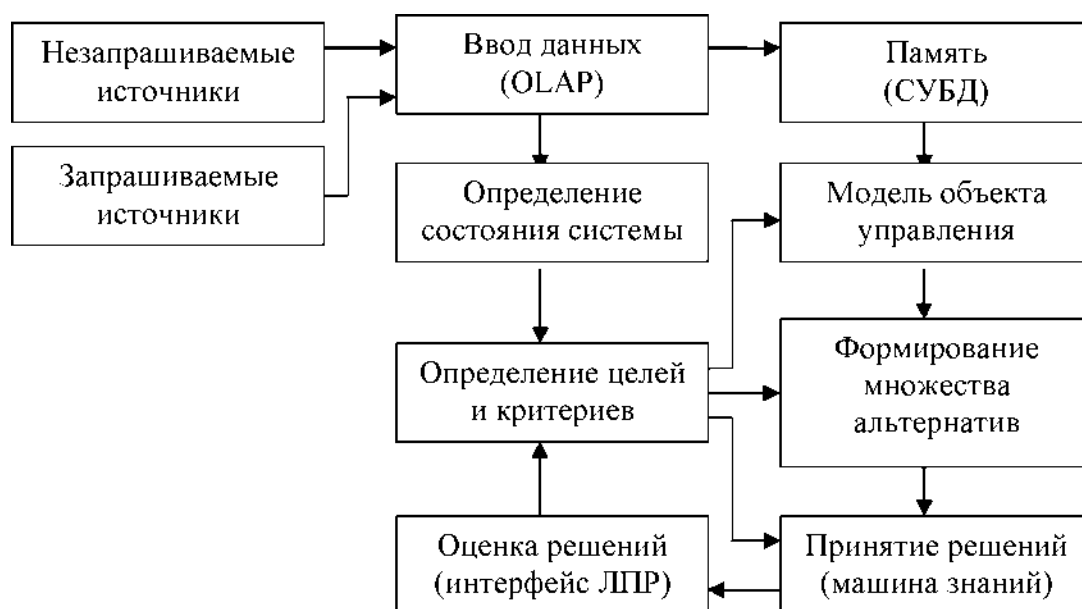


Рис.3. Обобщенная архитектура СППР

Таблица 1 - Классификация СППР

Признаки	Классы СППР
На концептуальном уровне	Управляемые сообщениями (Communication-DrivenDSS), управляемые данными (Data-DrivenDSS), управляемые документами (Document-DrivenDSS), управляемые моделями (Model-DrivenDSS).
На уровне пользователя	Пассивные, активные, интерактивные
На техническом уровне	СППР предприятия, настольные СППР
В зависимости от данных	Оперативные, стратегические

На концептуальном уровне СППР делятся на управляемые сообщениями (Communication-DrivenDSS), управляемые данными (Data-DrivenDSS), управляемые документами (Document-DrivenDSS), управляемые знаниями (Knowledge-DrivenDSS) и управляемые моделями (Model-DrivenDSS).

СППР, управляемые моделями, характеризуются доступом и манипуляциями с математическими моделями (статистическими, оптимизационными, имитационными). Некоторые OLAP-системы, позволяющие осуществлять сложный анализ данных, могут быть отнесены к гибридным СППР, которые обеспечивают моделирование, поиск и обработку данных.

Communication-DrivenDSS поддерживает группу пользователей, работающих над выполнением общей задачи.

Data-DrivenDSS и Data-orientedDSS в основном ориентируются на доступ и манипуляции с данными.

Document-DrivenDSS управляют, осуществляют поиск и манипулируют неструктурированной информацией, заданной в различных форматах.

Knowledge-DrivenDSS обеспечивают решение задач в виде фактов, правил, процедур.

На уровне пользователя СППР делятся на пассивные, активные и интерактивные системы. Пассивной СППР называется система, которая помогает

процессу принятия решения, но не может вынести предложение, какое решение принять. Активная СППР может сделать предложение, какое решение следует выбрать. Интерактивная СППР позволяет ЛПР изменять решения путем варьирования исходными данными или условиями решений.

На техническом уровне различается СППР предприятия и настольные СППР. СППР первого типа подключена к ХД и обслуживает многих менеджеров предприятия, а второй тип предназначен для отдельного пользователя.

В зависимости от данных, с которыми системы работают, СППР условно подразделяются на оперативные и стратегические. Оперативные СППР предназначены для немедленного реагирования на изменения текущей ситуации в управлении финансово-хозяйственными процессами компании. Стратегические СППР ориентированы на анализ значительных объемов разнородной информации, собираемых из различных источников. Важнейшей целью этих СППР является поиск наиболее рациональных вариантов развития бизнеса компании с учетом влияния различных факторов, таких как конъюнктура целевых для компании рынков, изменения финансовых рынков, изменения в законодательстве и др.

СППР первого типа получили название Информационных Систем Руководства - ИСР (ExecutiveInformationSystems, EIS). По сути, это конечные наборы отчетов, построенные на основании данных из транзакционной АИС предприятия, в идеале адекватно отражающей в режиме реального времени основные аспекты производственной и финансовой деятельности.

СППР второго типа предполагают достаточно глубокую проработку данных, специально преобразованных для использования в ходе процесса принятия решений. Неотъемлемым компонентом СППР этого уровня являются правила принятия решений, которые на основе агрегированных данных дают возможность менеджерам компании обосновывать свои решения, использовать факторы устойчивого роста бизнеса компании и снижать риски.. Технологии этого типа СППР строятся на принципах многомерного представления и анализа данных (OLAP).

В настоящее время при создании СППР получили развитие Web- технологии, которые для ряда компаний являются синонимами СППР.

На сегодняшний день можно выделить четыре наиболее популярных типа архитектур систем поддержки принятия решений:

- Функциональная СППР.
- Независимые витрины данных.
- Двухуровневое хранилище данных.
- Трехуровневое хранилище данных.

Функциональная СППР (Рис.4.) является наиболее простой с архитектурной точки зрения. Такие системы часто встречаются на практике, в особенности в организациях с невысоким уровнем аналитической культуры и недостаточно развитой информационной инфраструктурой.

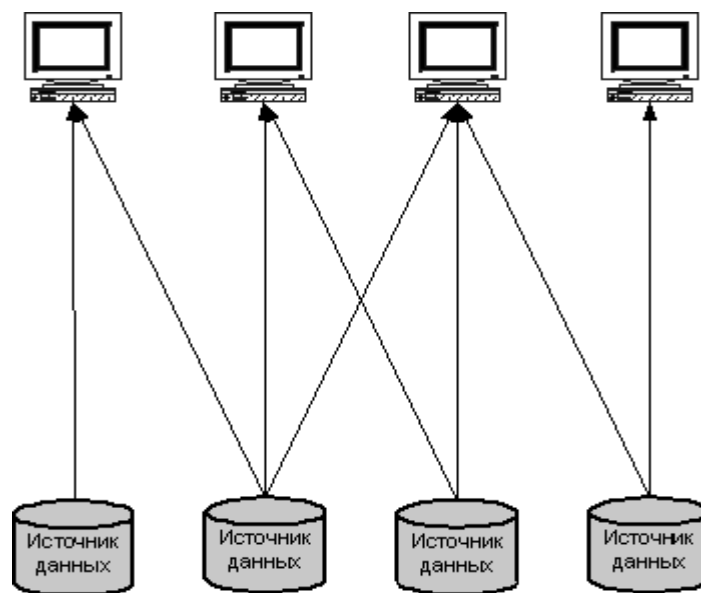


Рис.4. Функциональная СППР

Характерной чертой функциональной СППР является то, что анализ осуществляется с использованием данных из оперативных систем.

Преимущества:

- Быстрое внедрение за счет отсутствия этапа перегрузки данных в специализированную систему
- Минимальные затраты за счет использования одной платформы

Недостатки:

- Единственный источник данных, потенциально сужающий круг вопросов, на которые может ответить система
- Оперативные системы характеризуются очень низким качеством данных с точки зрения их роли в поддержке принятия стратегических решений. В силу отсутствия этапа очистки данных, данные функциональной СППР, как , обладают невысоким качеством
- Большая нагрузка на оперативную систему. Сложные запросы могут привести к остановке работы оперативной системы, что весьма нежелательно.

СППР с использованием независимых витрин данных(Рис.5.) встречаются в крупных организациях с большим количеством независимых подразделений, зачастую имеющих свои собственные отделы информационных технологий.

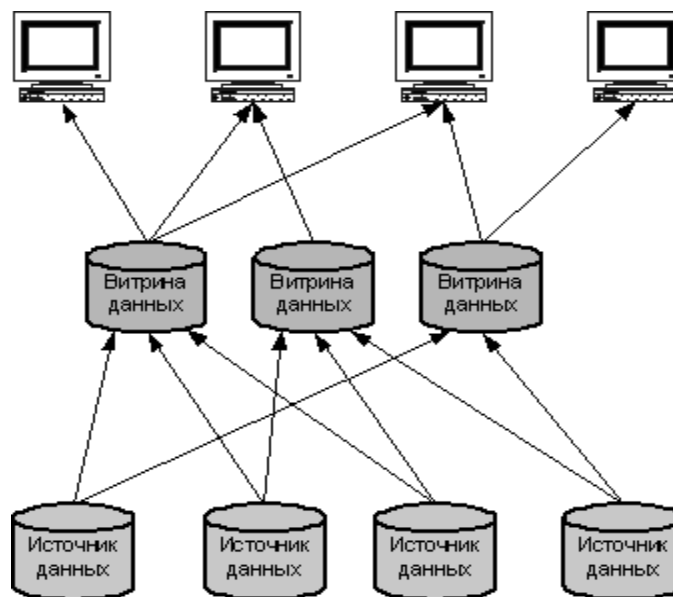


Рис.5. Независимые витрины данных

Преимущества:

- Витрины данных можно внедрять достаточно быстро
- Витрины проектируются для ответов на конкретный ряд вопросов
- Данные в витрине оптимизированы для использования определенными группами пользователей, что облегчает процедуры их наполнения, а также способствует повышению производительности.

Недостатки:

- Данные хранятся многократно в различных витринах данных. Это приводит к дублированию данных и, как следствие, к увеличению расходов на хранение и потенциальным проблемам, связанным с необходимостью поддержания непротиворечивости данных.
- Потенциально очень сложный процесс наполнения витрин данных при большом количестве источников данных.
- Данные не консолидируются на уровне предприятия, таким образом, отсутствует единая картина бизнеса.

СППР на основе двухуровневого хранилища данных(Рис.6.) строится централизованно для предоставления информации в рамках компании. Для поддержки такой архитектуры необходима выделенная команда профессионалов в области хранилищ данных.

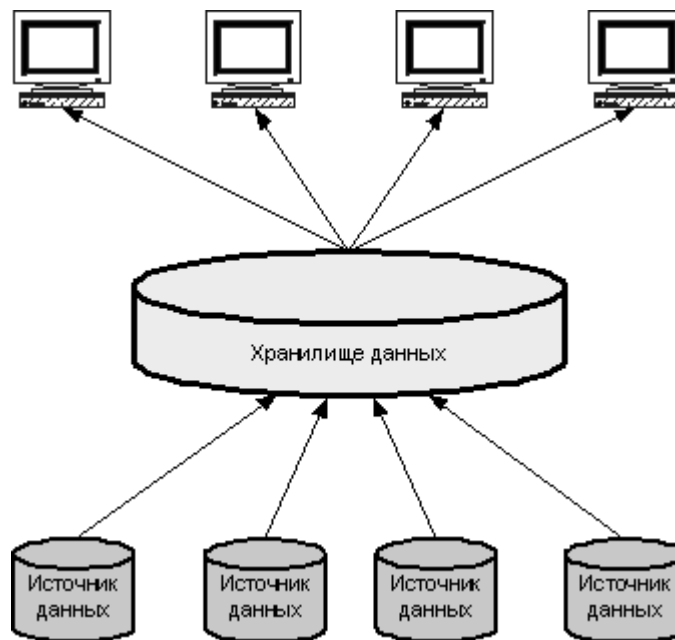


Рис.6. Двухуровневое хранилище данных

Это означает, что вся организация должна согласовать все определения и процессы преобразования данных.

Преимущества:

- Данные хранятся в единственном экземпляре.
- Минимальные затраты на хранение данных.

- Отсутствуют проблемы, связанные с синхронизацией нескольких копий данных.
- Данные консолидируются на уровне предприятия, что позволяет иметь единую картину бизнеса.

Недостатки:

- Данные не структурируются для поддержки потребностей отдельных пользователей или групп пользователей
- Возможны проблемы с производительностью системы
- Возможны трудности с разграничением прав пользователей на доступ к данным.

Архитектура СППР на основе трёхуровневого хранилища данных приведена на рис.7:

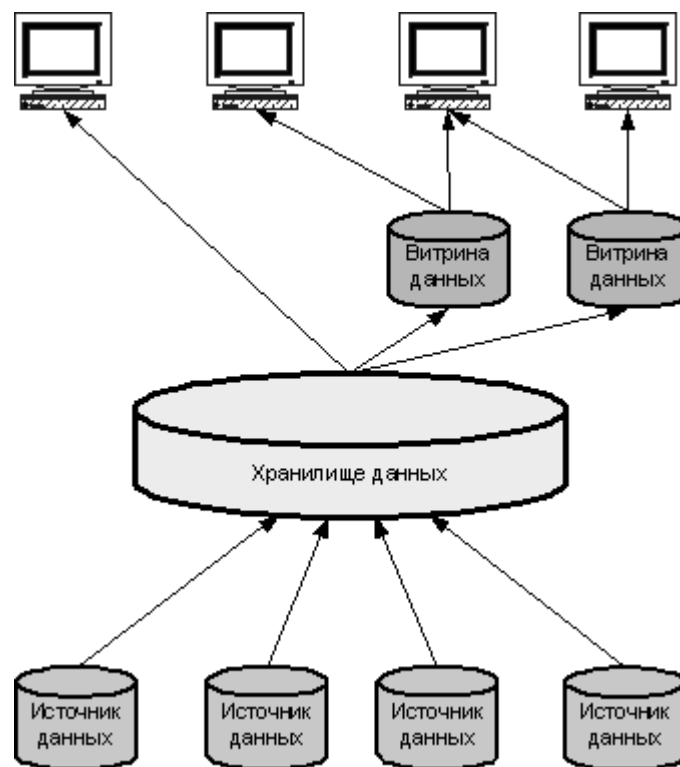


Рис.7. Трёхуровневое хранилище данных

Хранилище данных представляет собой единый централизованный источник корпоративной информации. Витрины данных представляют подмножества данных из хранилища, организованные для решения задач отдельных подразделений компании. Конечные пользователи имеют возможность досту-

па к детальным данным хранилища, в случае если данных в витрине недостаточно, а также для получения более полной картины состояния бизнеса.

Преимущества:

- Создание и наполнение витрин данных упрощено, поскольку наполнение происходит из единого стандартизованного надежного источника очищенных нормализованных данных.
- Витрины данных синхронизированы и совместимы с корпоративным представлением. Имеется корпоративная модель данных. Существует возможность сравнительно лёгкого расширения хранилища и добавления новых витрин данных.
- Гарантированная производительность.

Недостатки:

- Существует избыточность данных, ведущая к росту требований на хранение данных
- Требуется согласованность с принятой архитектурой многих областей с потенциально различными требованиями (например, скорость внедрения иногда конкурирует с требованиями следовать архитектурному подходу).

Нами рассмотрены основные варианты архитектур систем поддержки принятия решений, при этом выбор конкретного варианта зависит от комплекса решаемых задач на объекте автоматизации.

Система поддержки принятия решений (СППР) (англ. Decision Support System, DSS)— компьютерная автоматизированная система, целью которой является помощь людям, принимающим решение в сложных условиях для полного и объективного анализа предметной деятельности. СППР возникли в результате слияния управленческих информационных систем и систем управления базами данных.

Использование системы позволяет найти ответы на множество вопросов, возникающих у руководителей компании, например:

У генерального директора:

- На сколько процентов выполнен план по продажам, доходу, прибыли, расходам
- Какова доля рынка, принадлежащего компании
- Каковы тенденции развития сегмента рынка, на котором представлена компания
- Каковы ключевые показатели производительности компании в текущем периоде
- Каковы тенденции изменения ключевых показателей производительности компании со временем

У руководителя отдела по работе с партнерами

- Какие из партнеров приносят наибольший доход, прибыль
- Какие проекты, группы продуктов лучше всего продает данный партнер
- Каковы тенденции изменения продаж через партнеров

У руководителя финансового департамента

- Сколько каждый проект стоит моему предприятию
- Сколько стоит поддержка продаваемых проектов
- Какие проекты в этом году стоят больше, чем в прошлом
- Как расходы различных подразделений и компании в целом соотносятся с доходами

У руководителя департамента бюджетного планирования и контроля

- Насколько точно различные подразделения компании соблюдают установленный бюджет
- Каковы тенденции расходов по различным подразделениям, статьям бюджета

У руководителя департамента закупок

- Какие из моих поставщиков предлагают наилучшее соотношение цена/качество
- Какие из поставщиков доставляют товары быстрее остальных Медленнее остальных

- Как часто происходят задержки поставок от того или иного поставщика
- Каких поставщиков выбрать для поставок крупных/небольших партий продукта

У руководителя планового отдела (отдела стратегического планирования)

- Насколько предприятие выполняет план по продажам, доходам, прибыли
- Какие области бизнеса вносят положительный вклад, а какие - отрицательный
- Каков прогноз ключевых показателей производительности на следующий период (месяц, квартал, год)

У руководителя отдела сервисного обслуживания

- Каково среднее время выполнения заявки на обслуживание
- Каковы расходы на выполнение одной заявки
- Каково среднее время до первой поломки данной модели

У руководителя отдела кадров

- Какова производительность персонала, прошедшего определенное обучение перед теми, кто его не проходил
- Каковы тенденции ежегодного роста персонала компании в различных регионах, подразделениях
- Каково прогнозируемое количество персонала на следующий год
- Каковы прогнозы по поводу состава
- Какие сотрудники нуждаются в обучении
- Каким набором навыков должен обладать сотрудник, чтобы хорошо выполнять свои обязанности

У руководителя отдела анализа качества

- Какие проекты доставляются вовремя, какие - с запозданием
- Имеют ли определенные клиенты или проекты недопустимо долгий срок поставки
- Изменилось ли время доставки определенных продуктов со временем

- Насколько быстрее или медленнее стала поставка продуктов (услуг) в определенный сегмент рынка
- Каковы основные причины отказа от продукта (услуги).

Тема №6. Пример реализации СППР – «Монитор руководителя»

Одним из главных факторов, влияющих на успешную реализацию целей, поставленных перед государственными организациями в рамках различных инициатив по построению цифровой экономики, является возможность отслеживания всех стадий выполнения различных проектов и процессов, их промежуточных результатов, отклонений от планов. Полная прозрачность процессов помогает координировать действия различных министерств и ведомств, оперативно принимать более взвешенные и верные управленческие решения.

С принятием государственной программы "Цифровая экономика" стало очевидно, что вектор движения государства направленный на информатизацию государственных услуг придаст дополнительное ускорение таким направлениям ИТ-бизнеса как построение систем электронного документооборота, систем управления неструктурированным контентом в организациях, создание информационных порталов, построение систем поддержки принятия решений. Именно эти направления позволяют решать основные задачи, поставленные перед государственными организациями: повышение эффективности государственного управления и экономия затрат государства и граждан за счёт электронного взаимодействия.

Мониторинг и контроль проектов и бизнес-процессов в том или ином виде существует в любой государственной или коммерческой структуре. Важную роль при этом играет прозрачность и скорость этих процессов, оказывающих влияние на все ключевые функции организации и отражающиеся на показателях эффективности её деятельности. Сегодня руководителям федеральных и региональных органов государственного управления, крупных

предприятий, контролирующего множество процессов и исполнителей, приходится анализировать результаты деятельности организации и принимать важные управленческие решения в условиях непрерывного роста данных, постоянно меняющейся конъюнктуры, в режиме нехватки времени. Затраты ресурсов и времени на сбор и обработку информации из различных источников традиционными способами становятся критически большими, ко времени составления необходимого отчёта стратегическая ситуация меняется, информация устаревает и не находит применения.

Другая важная задача управления – фокусирование внимания на наиболее приоритетных, критичных областях и проблемах деятельности организации. В постоянно растущем потоке документов, отчётов, файлов, важные аналитические данные легко могут затеряться среди слабо актуализированной статистической информации.

Проблема некачественной информации, противоречивости данных, их неактуальности, недостаточности или, наоборот, излишней детализации, зачастую становится причиной неэффективного менеджмента, в котором не учитывается важная информация, а всё внимание уделяется вторичным деталям.

Что необходимо руководителю? Получать полную, но не избыточную, достоверную, объективную информацию, своевременно, в удобной форме, с откровенной индикацией кризисных показателей. Сегодня эти задачи успешно решаются при помощи передовых информационных технологий с применением различных систем поддержки принятия решений.

Однако существуют проблемы, которые могут помешать эффективному использованию подобной системы: низкая эргономичность интерфейса, инструментарий, требующий специальных знаний, неоднородность и многочисленность экранных форм, перегрузка избыточной функциональностью. Руководитель может просто перестать использовать систему, если посчитает, что это слишком сложно.

Совокупность различных факторов диктует определённые требования к системам класса поддержки принятия решения. Рассмотрим функционал системы поддержки принятия решений «Монитор руководителя», которая специально настроена под информационные нужды высшего менеджмента и предоставляет широкие возможности анализа данных в режиме реального времени. При этом данная система дополнительно предоставляет доступ к широкому набору информации о проектной деятельности, обладает простым интуитивно понятным интерфейсом, не требует специальных знаний для применения инструментов анализа и предоставляет информацию в удобной графической форме.

Приведем описание архитектуры решения.

На рабочем столе руководителя необходимо собрать всю нужную ему информацию по текущему состоянию проектов и поручений в простом и понятном виде. Наиболее сложная техническая проблема – получение данных из различных источников внутри и вне организации. В рамках проекта по построению системы поддержки принятия решения разрабатывается методология получения и контроля данных, шаблоны отчётов по проектам и механизмы загрузки этой информации в хранилище.

Информация по проектам и поручениям загружается в хранилище данных, которое пополняется с помощью автоматической загрузки планов и отчётов или вручную с помощью веб-форм. Затем информация экспортируется в витрину данных, которая может быть расположена на любом персональном компьютере и позволяет с помощью графических диаграмм или отчётов получать первому лицу актуальные данные о текущем состоянии всех проектов и поручений, с возможностью детализации вплоть до предоставления первичных данных.

Интерфейс системы обеспечивает представление текущей информации из различных источников по важнейшим направлениям деятельности в нужном виде и в нужное время.

Экраны "Монитора руководителя" могут содержать:

- информацию о стратегических целях и степени их достижения;
- информацию о выполнении бюджетов и расходовании средств, в том числе, в разрезе проектов, программ, целей, задач и ответственных;
- ключевые финансовые и статистические показатели деятельности организации, в том числе в сравнении с предыдущими периодами;
- информацию об отклонениях ключевых показателей деятельности от плановых значений;
- показатели реализации программ и проектов, информацию об их содержании, финансировании, плановых и фактических сроках реализации этапов, достижении целевых показателей;
- информацию об исполнении поручений;
- другую критичную для руководителя информацию.

Учитывая возросшую тенденцию к использованию свободного ПО в государственных структурах, в качестве технологической платформы для разработки аналитической системы «Монитор руководителя» используются инструменты с открытым кодом. Открытые решения позволяют снизить зависимость от вендора, сократить сроки разработки, а сами разработки вести в режиме "экстремального программирования", в условиях, когда функциональные требования к системе постоянно меняются и развиваются.

Среда разработки и исполнения - приложение AdobeFlex является свободно распространяемой высокоэффективной средой разработки для создания и обслуживания веб-приложений, совместимых со всеми наиболее распространенными обозревателями, платформами персональных компьютеров и версиями операционных систем. Многофункциональные интернет-приложения, созданные в среде Flex, могут выполняться как в браузере (с помощью программного проигрывателя AdobeFlashPlayer), так и автономно. Это обеспечивает их надежную работу во всех основных браузерах и операционных системах.

IBM ILOG Elixir - дополнительная библиотека для платформ AdobeFlex и Adobe AIR.

IBM ILOG Elixir включает графические компоненты Flex: трехмерные диаграммы, шкалы, карты, карты интенсивности, календари, объекты OLAP и сводные диаграммы, организационные схемы, древовидные схемы, лепестковые диаграммы, диаграммы ресурсов и задач Ганта.

Ext JS - фреймворк для построения веб-интерфейсов.

Ext JS - библиотека JavaScript для разработки веб-приложений и пользовательских интерфейсов, которая поддерживает технологию AJAX, анимацию, работу с DOM, реализацию таблиц, вкладок, обработку событий.

Веб-сервер Apache является кроссплатформенным программным обеспечением, поддерживающим операционные системы Linux, Mac OS, Microsoft Windows, и др. Основными достоинствами Apache считаются надёжность и гибкость конфигурации. Apache позволяет подключать внешние модули для предоставления данных, использовать СУБД для аутентификации пользователей, модифицировать сообщения об ошибках и т. д.

База данных MySQL используется для хранения информации. В зависимости от потребностей может быть использована любая другая СУБД.

Система может разрабатываться и функционировать практически на любых платформах. Доступ к ней возможен через Интернет, Интранет или с автономного персонального компьютера. Продуманная и преднастроенная структура объектов мониторинга делает возможным в короткий срок создать рабочий прототип системы. Модульная архитектура системы позволяет оперативно вносить изменения "на лету", быстро переходить на новые версии.

Благодаря тому, что представленная в системе «Монитор руководителя» информация является точной, релевантной, непротиворечивой и демонстрирует наиболее приоритетные для руководства показатели в оперативном режиме, система поддержки принятия решения позволяет увидеть целостную картину результатов деятельности организации и подчинённых в различных направлениях деятельности. Отчёты предоставляются в удобном для восприятия виде, с использованием различных видов деловой графики (диаграммы, географические карты, приборные доски, графики). Всё это позволяет снизить затраты на

обработку информации, необходимой для принятия управленческих решений, повысить их эффективность, оперативно реагировать на критические ситуации.

Одним из первых заказчиков, оценивших преимущества использования системы «Монитор руководителя» стало федеральное государственное учреждение "Аналитический центр при Правительстве Российской Федерации". Оперативный мониторинг, который осуществляется с помощью системы поддержки принятия решений, позволяет отслеживать ход реализации проектов Комиссии при Президенте Российской Федерации по модернизации и технологическому развитию экономики России первым лицам Аппарата Правительства и Аппарата Президента Российской Федерации.

Тема №7. Информационное пространство предприятия, показатели отчетности

Деятельность предприятия, другого объекта или системы отображается в информационном пространстве. Подготовка принятия решений и, соответственно, анализ происходят в этом пространстве как в среде - с одной стороны и, с другой стороны, производятся операции с элементами структуры этого пространства и сопряженных с ним информационных пространств.

Аналитик и другие лица, связанные с информационными процессами, используют в своей работе понятия, определяющие сущность, структуру, элементы информационного пространства, особенно при использовании современных информационных технологий.

Предприятие, корпорация, любой хозяйственный комплекс является системой. Систему можно представить с одной стороны объектом как единым целым, с другой стороны как совокупность (множество) связанных между собой и взаимодействующих составных частей - объектов, но меньшего масштаба. Информационное отображение физических объектов или процессов называют информационным объектом.

Совокупность информационных объектов, информационно отображающих свойства системы и протекающие в ней процессы, называют информационным пространством. Оно состоит из различных квантов или массивов данных в виде разного рода письменных (знаковых) и фиксированных на носителях информации кодограмм, буквенно-цифровых на естественном языке, устных и визуальных сообщений. Все виды сообщений предоставляются непосредственно потребителям информации или передаются по каналам связи, могут быть сохранены в различном виде с помощью современных технических средств и по мере необходимости воспроизводиться. Подавляющая часть сообщений, как , «загружается» в информационную систему и становится основой ее информационного обеспечения. Сочетание знаковых, звуковых (аудио) и образных видимых, в том числе с анимацией, сообщений

(визуальных) называют мультимедийными сообщениями. Такого рода сообщения получили в последние годы широкое распространение в связи с высокой эффективностью восприятия и широким внедрением поддерживающих такие сообщения средств на практике и в состав информационных систем соответственно.

Сообщения могут содержать оперативную информацию о технологических процессах, хозяйственных операциях; могут быть представлены в виде экономической, технической, организационно-распорядительной, отчетной документации и т.д.

Характерным свойством информационного пространства является его структурированность. Это означает, выделены его элементы, установлены связи между ними, введены обозначения, элементы и связи упорядочены. Свойство структурированности в разных видах информационных пространств может быть выражено в разной степени. Высокий уровень обеспечивает возможность представления информации в виде документов и манипулирования данными с помощью программно-технических средств информационных систем. Хотяшов Э.Н. и Королев М. А. различает пять степеней структурированности ИП:

- неструктурированное ИП (НИП);
- слабо структурированное (ССИП);
- структурированное (СИП);
- формализовано-структурированное (ФСИП);
- машинно-структурированное (МСИП).

Рассмотрим подробнее признаки степеней структурированности.

НИП - признаки структуризации крайне редки, примером служит человеческая речь, передача сообщений в животном мире от особи к особи.

ССИП - компоненты структуризации не имеют законченного вида, это естественный письменный язык, где признаками структуризации являются грамматические правила, которые зачастую неоднозначны, противоречивы, имеют исключения, недостаточно строги и т. д. СИП отличается преоблада-

нием структурированных компонентов, внедрено кодирование, информация документируется; это информация, подготовленная к «загрузке» в информационную систему.

ФСИП - имеются такие спецификации информационных объектов и их взаимосвязей, которые содержат алгоритмы получения любых значений элементов данных; обеспечиваются операции по управлению данными, возможны реорганизация и оптимизация структуры ЭИС, а также алгоритмов обработки информации.

МСИП - представлены в формализованном виде все информационные объекты и их взаимосвязи, процессы преобразования информации описаны на языках программирования, обеспечивается взаимодействие пользователя и ЭИС на естественном или близком к естественному языку или по предельно упрощенным правилам.

В качестве элементов структуры информационного пространства выступают единицы информации. Это понятие, которое рассматривается в теории экономических информационных систем (ЭИС), выражает сущностное или смысловое наполнение элемента ИП. Под единицей информации понимают «набор символов, которому придается определенный смысл». Рассматривается система единиц информации, которая имеет довольно сложную иерархическую структуру. Выделяют несколько уровней единиц информации в зависимости от смыслового (семантического) значения, его наполненности.

По возрастанию содержательности понятия определены следующие единицы информации: реквизит и составная единица информации (СЕИ), которая включает в себя такие единицы как показатель и база данных.

Элементарной единицей информации нижнего уровня является реквизит. Это информационное отображение свойства объекта, какого-либо процесса или явления. Сообщения состоят из определений свойств объектов, предметов, явлений, складывающихся некоторым образом из соответствующих реквизитов. Следует заметить, что синонимом понятия реквизит является атрибут, широко используемый в литературе по базам данных термин.

Отсюда составная единица информации собирается из набора соответствующих определению данного объекта реквизитов и представляет собой информационное отображение объекта или его части.

Разновидностью составной единицы информации является показатель. Это сложное понятие. Имеются его различные определения. Одни авторы подчеркивают сущностный смысл или характер, привязанный к предметной области, в частности экономической. Другие исходят из формально-структурного подхода, ориентированного на структуризацию содержащейся в показателе информации в целях приспособления его структуры для эффективного использования в информационной системе. Результаты такой структуризации используются и в информационно-аналитических системах.

Приведем в данном контексте определение формально-структурного подхода по Королеву М. А. в интерпретации Ясина Е.Г. «Показатель представляет высказывание с законченным смыслом, включающее как название переменной величины, так и ее конкретное количественное значение со всеми качественными признаками, необходимыми для идентификации последнего». Показатель образуется из набора реквизитов или терминов.

Реквизиты составляют две группы:

- реквизиты-признаки, выражающие качественные отличия показателя, его смысловое содержание, в частности экономическое;
- реквизиты-основания, содержащие количественные значения показателя. Показатель теряет смысл без реквизитов-признаков и реквизитов оснований. В совокупности они образуют высказывание (сообщение), имеющее законченный предметный смысл, что позволяет утверждать, что показатель является наименьшей составной единицей информации, которая достаточна для документообразования, передачи, хранения и восприятия сообщений.

При структуризации информационного пространства разрабатывается система показателей, анализируется их собственная структура. В ходе этой работы необходимо исследовать общие закономерности, выявить категории показателей - члены общей структурной формулы описания показателей.

В общем виде, структура показателя выглядит следующим образом:

$$P \rightarrow (R, x),$$

где: P - показатель (может быть экономическим);

R - набор реквизитов (терминов), идентифицирующих смысловое значение показателя; x - количественное значение показателя .

Идентификатор в свою очередь можно представить в виде двух частей:

$$R \rightarrow (S, Q),$$

где: S - составленное из реквизитов наименование показателя, выявляющее его предметный смысл; Q - дополнительные признаки показателя, составленные также из реквизитов и уточняющие его количественное значение.

Выделенные реквизиты могут быть в свою очередь составными, Для уточнения связей между ними строятся схемы, детализирующие объект до такой степени, что дальнейшая детализация невозможна или не имеет смысла. Реквизиты самого нижнего уровня называются единичными. Другие, расположенные на более верхних уровнях - множественными.

Дальнейший анализ проведем, начав с дополнительных признаков. Они могут состоять из:

E - единиц измерения, их может быть несколько в составе показателя;

C - субъектов, это могут быть наименования субъектов и объектов хозяйственной деятельности, регионов, места размещения предприятия и других объектов;

B - времени или реквизитов, определяющих временной аспект - моментов происхождения событий, периодов времени протекания хозяйственных или иных процессов, явлений;

U - признак стадии учета или, как говорят в функции управления, то есть плановые, фактические, нормативные или какие либо другие значения показателя. Представим эту структуру в виде соотношения:

$$Q \rightarrow (E, C, B, U), \text{ таким образом } R \rightarrow (S(E, C, B, U))$$

Наименование показателя может быть слитным (определенным одним реквизитом) или иметь свою структуру и в свою очередь состоять из рекви-

зитов, таких как:

- Φ - формальная (вычисляемая) характеристика показателя, раскрывающая его структуру или алгоритм агрегации исходных детальных данных, например, объем продаж, среднее, максимальное значение той или иной величины (подразумевается методика подсчета);

- Π - обозначение отображаемого технологического или бизнес-процесса, например изготовление, реализация, перевозка и т. д.

- O - объект измерения, подсчета - виды товаров, оборудования, работники по категориям или общим числом. Тогда $S \rightarrow \langle \Phi, \Pi, O \rangle$.

Таким образом, общая структурная формула показателя примет вид:

$$P \rightarrow (R, x)$$

$$R \rightarrow (S, Q),$$

$$S \rightarrow (\Phi, \Pi, O),$$

$$Q \rightarrow (E, C, B, Y),$$

$$R \rightarrow ((\Phi, \Pi, O), (E, C, B, Y)),$$

$$P \rightarrow ((\Phi, \Pi, O), (E, C, B, Y), x)$$

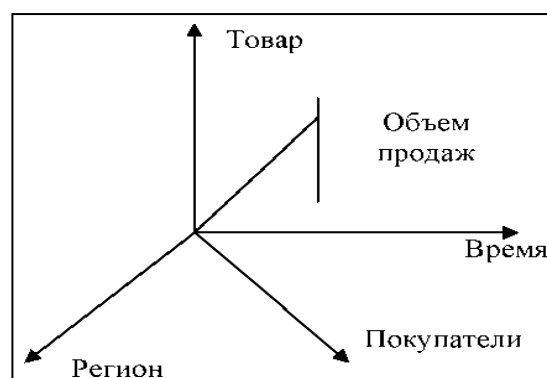
Данная структура, представленная в таблице 1, может отображать практически любой показатель.

Таблица 2. Структура показателя

P							
R							X
S			Q				
Φ	Π	O	E	C	B	Y	
P							
Φ	Π	O	E	C	B	Y	X

Представим пространство признаков показателя на системе координат.

Для образного восприятия сложных понятий удобна их пространственная или геометрическая интерпретация.



рическая интерпретация.

Рис. 8. Система координат многомерного информационного пространства показателя

Число измерений признаков этого пространства равно числу осей его системы координат, на которых откладывают отрезки, соответствующие реквизитам. Множественность реквизита выражается в длине отрезка. Тогда составляющие его единичные реквизиты укладываются на этом отрезке. Для трехмерного пространства это будет куб. Можно построить пространство системы взаимосвязанных показателей в виде набора соприкасающихся трехмерных кубов, Рис.9.

Иллюстративные материалы специальной литературы и технической документации содержат упрощенные геометрические модели рассматриваемых пространств. На осях откладываются реквизиты-признаки в виде отрезков равной длины, а их количественное наполнение содержится в реквизитах-основаниях, которые не отображаются геометрически.

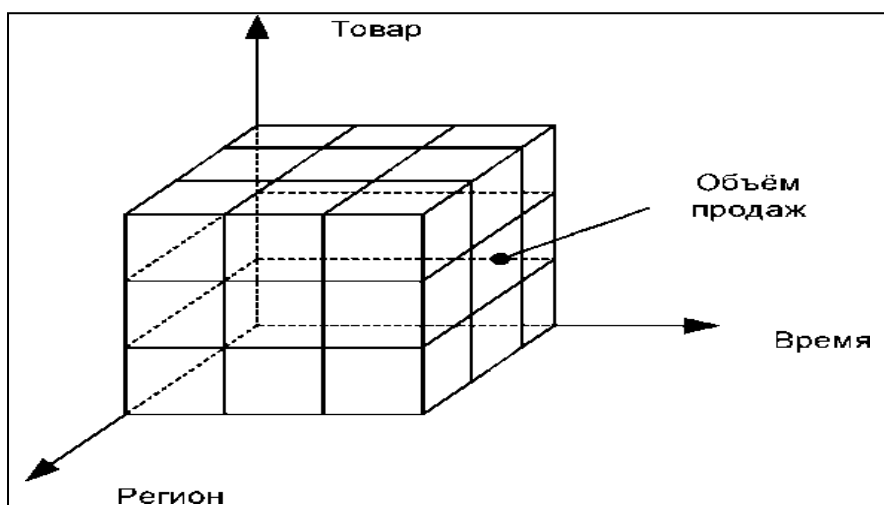


Рис.9. Трехмерный гиперкуб информационного пространства показателя

Пример такого пространства показателей приведен на рис.9. «Объем продаж». Широко распространена также геометрическая интерпретация системы показателей в виде пирамиды, в основании которой лежат единичные признаки. Пирамида делится по высоте на «этажи», которые соответствуют иерархии признаков или системы показателей.

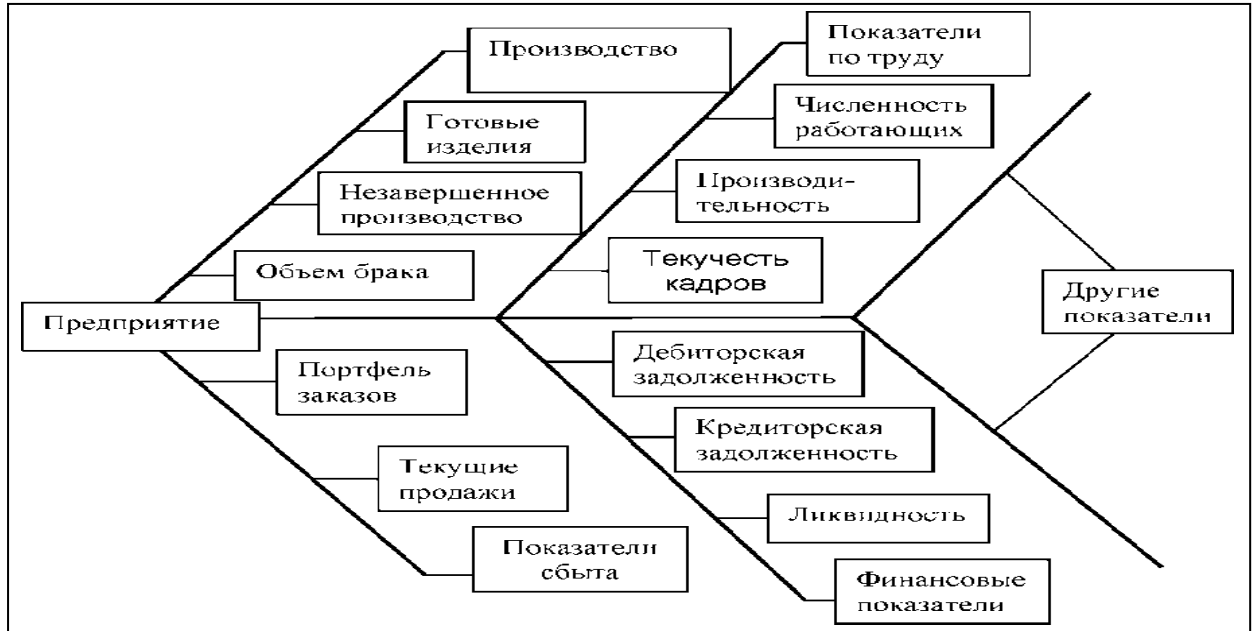


Рис.10. Диаграмма Ишикава

Соответствующие структуры представляют также в виде перевернутого дерева или «рыбьего скелета» - диаграммы Ишикава, Рис.10.

Для оценки ситуации на предприятии или другом объекте при подготовке и принятии решений по управлению объектом должна быть выработана или принята система оценок, которая ложится в основу аналитической работы с имеющейся и требуемой информацией. Систему оценок можно рассматривать как совокупность показателей деятельности объекта с их критериальными значениями.

Для экономической сферы деятельности это будут экономические показатели. Системы экономических показателей основываются на элементах структуры информационного пространства.

С экономической точки зрения под показателем деятельности предприятия понимают «конкретное проявление экономической категории в харак-

теристике объекта». Это наиболее строгое определение экономической сущности показателя. Отдельным показателем невозможно достаточно полно отобразить и оценить состояние экономического объекта, поэтому применяются системы показателей.

Система экономических показателей отображает финансово - хозяйственную деятельность как на уровне предприятия, то есть на микроуровне, так и на макроуровне. Два этих раздела показателей касаются как каждого предприятия, корпорации (объединения), так и государственных органов и негосударственных организаций регионального, государственного и межгосударственного масштаба (уровня). Показатели отражаются в документации, которая ведется на предприятии, используются в оценках его состояния и динамики процессов, происходящих на предприятии, в информационных системах. Естественно, что при анализе их используют непосредственно и интерпретируют различным образом в целях извлечения знаний, формирования выводов и т. д.

В связи с этим централизованно разработаны системы реквизитов соответствующих обязательных к использованию документов: государственные стандарты документооборота, формы документов, системы кодирования статистической, учетной, финансовой отчетности и другой документации. На предприятии разрабатывается внутренняя система показателей и соответствующих реквизитов.

Предприятие регулярно обменивается информацией с внешней средой в виде представления разного рода отчетности и других материалов, получения различных руководящих и установочных документов, информационных материалов.

Предприятие представляет в государственные органы ежеквартально и нарастающим итогом за полугодие, год:

- внешнюю финансовую отчетность в территориальные государственные органы - налоговую инспекцию и финансовое управление;
- статистическую отчетность о различных сторонах деятельности

предприятия в территориальные органы Госкомстата РФ;

- отчетность о финансово-хозяйственной деятельности в государственных фонды (предстоит их реформирование в плане объединения и упрощения отчетности).

Выдается информация в виде отчетности и других документов в вышестоящие органы. Госпредприятия и организации представляют ее в государственные и муниципальные органы управления, негосударственные - в корпоративные органы. Имеется документооборот и другого рода: информационные потоки с банковскими и другими финансовыми учреждениями, например страховыми, взаимодействующими и партнерскими организациями и т. д.

В государственных органах проведены структуризация, формализация, классификация и кодирование многих видов документов на основе разработки соответствующей системы реквизитов и показателей, которая постоянно видоизменяется и далека от совершенства, а самое главное от общепринятых в мировой практике международных стандартов, систем отчетности, классификации, кодирования. Особенно страдает от такого положения дел финансовая сфера управленческой деятельности.

Под классификацией понимают определение мест показателей в той или иной сфере деятельности, в частности экономической; среди объектов бизнеса, функций управления ими, их состояний, по единицам измерения, стабильности значения показателя, то есть упорядочение информации. Выражаясь формальным языком, «классификация преследует цель установления отношений между понятиями как отображениями объектов или групп объектов с общими свойствами, изучения структуры и упорядочения содержания данных».

Кодирование заключается в заинтересованном обозначении элементов данных, которое имело бы необходимую длину и удобство представления при наименьших затратах на перекодирование при обмене данными, их обработке, контроле, хранении.

Задачи классификации и кодирования тесно увязаны. Примерами общегосударственных систем классификации и кодирования служат ОКП (Общесистемный классификатор промышленной и сельскохозяйственной продукции), ОКОНХ (Общесистемный классификатор отраслей народного хозяйства), ОКПО (общесистемный классификатор предприятий и организаций) и другие, однако упомянутые системы плохо увязываются с системами национальных счетов, принятыми в международной практике.

На микроуровне стоит подобная задача меньшего масштаба, но не менее трудная. При внедрении информационно-аналитической системы как автономной, которая увязывается с локальными базами данных или информационными системами, так и встроенной в интегрированную ЭИС возникает проблема структуризации информационного пространства предприятия, которая выражается в первую очередь в классификации и кодировании.

Как , на предприятии имеются фрагменты автоматизации: то ли локальные информационные системы типа бухгалтерских, электронного документооборота, различные базы данных в сфере логистики, производственного учета и т. д. Все они имеют какую-либо структуризацию, классификацию и кодирование данных. Зачастую это носит частный характер и не стыкуется или плохо согласуется с другими массивами данных.

Тема №8. Интеграция данных в рамках СППР из различных источников

Формирование разнообразной отчетности — важнейшая функция информационной системы предприятия. В зависимости от потребностей пользователей выделяют различные виды отчетов и способы хранения и извлечения необходимых данных.

Формируемую на предприятии отчетность принято разделять на оперативную и аналитическую. Задача оперативной отчетности — отражение текущего состояния предприятия и его различных подразделений. Такие отчеты требуются с высокой частотой и, как правило, являются довольно узкоспециализированными, т. е. охватывают какое-то одно направление деятельности.

Оперативные отчеты можно получать непосредственно из оперативных (OLTP) систем. Так как используемые для отчетов данные хранятся в разрозненных источниках, такой способ хранения является децентрализованным (рис. 11).

OLTP-базы оптимизированы для максимально быстрой обработки большого числа транзакций и хранят данные за текущий период времени. Это позволяет им быстро обрабатывать текущие задачи. Однако в качестве источника данных для аналитических отчетов децентрализованные источники данных имеют ряд существенных недостатков.

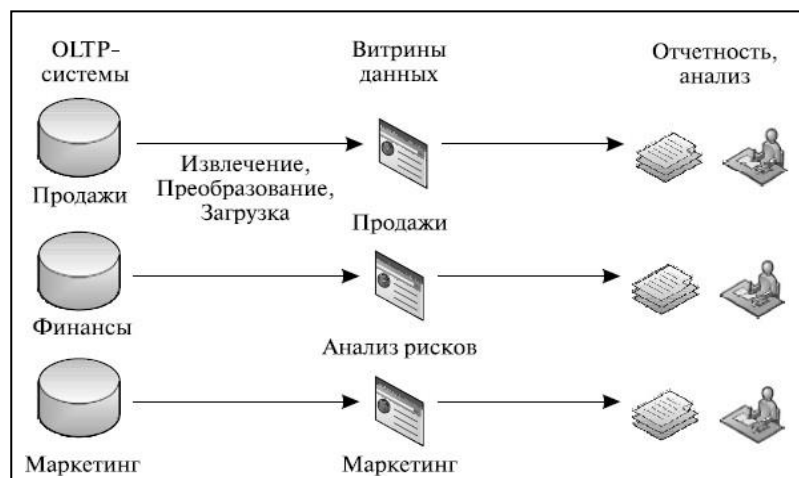


Рис. 11. Децентрализованные источники данных

В первую очередь оперативные системы должны обеспечивать приемлемую производительность при выполнении своей прямой задачи — обработки операций. Генерация отчетов ложится на них дополнительной нагрузкой и замедляет работу.

Более того, если отчет охватывает зоны ответственности разных OLTP-систем или требует подключения внешних данных, то часто требуется какое-то программирование, экспорт в промежуточные форматы и дополнительные расчеты. Такие операции обычно осуществляются вручную, а это порождает сразу две проблемы: на формирование отчета требуется много времени, и при этом высока вероятность возникновения ошибок.

На корректность получаемой информации негативно влияет также и то, что порядок и формат ввода оперативных данных в OLTP-системы зачастую не отвечают требованиям их аналитической и статистической обработки. Данные могут повторяться или быть сохранены в виде, затрудняющем их машинную обработку.

Оперативные системы накладывают существенные ограничения на хранение данных за большой период времени, так как попросту не предназначены для этого.

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

Наполняющие витрину данные проходят более или менее сложные шаги извлечения, преобразования и загрузки (Extract, Transform, Load, ETL). Такой подход позволяет выбрать из оперативных систем-источников только требуемые данные, выверить и очистить их, а также посчитать и сохранить промежуточные итоги. Благодаря этому витрины могут обеспечить высокую производительность при построении отчетов и низкие затраты на сопровождение данных.

Недостатком витрин является их ограниченность рамками конкретной задачи, например, деятельностью одного департамента или периодом времени для анализа.

Решить все эти проблемы хранения и анализа данных в масштабе всего предприятия призвана технология хранилищ данных (Data Warehouse).

Корпоративное хранилище данных (КХД)— это специальным образом организованный массив данных организации, обрабатываемый и хранящийся в едином аппаратно-программном комплексе. КХД обеспечивает быстрый доступ к оперативной и исторической информации и предоставляет широкие возможности для построения отчетов и анализа хранимых данных.

Корпоративное хранилище данных представляет собой единый централизованный источник информации в масштабе предприятия (рис. 12).

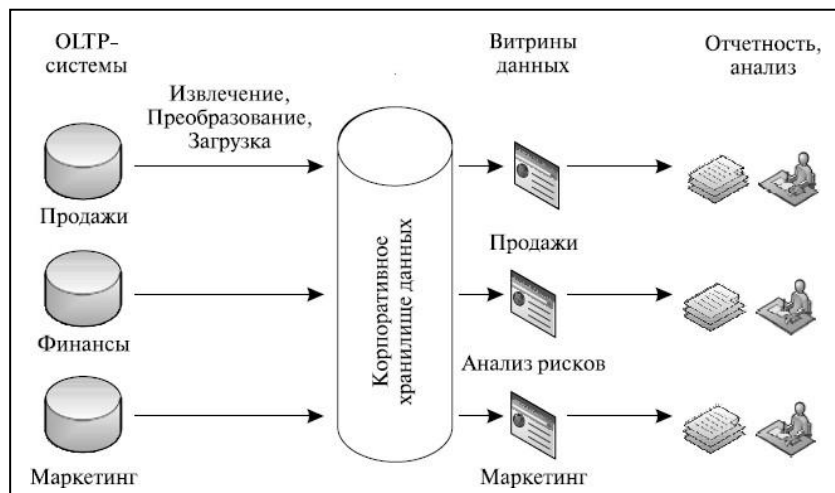


Рис. 12. Централизованный источник данных на основе ХД

Хранилище данных обладает рядом важных особенностей:

- Предметная ориентированность.
- Интегрированность.
- Привязка ко времени.
- Неизменяемость.

Хранилище может содержать данные не только из OLTP-систем, но и из других разнообразных источников, таких как внешние базы данных, базы

унаследованных (существовавших на предприятии раньше и впоследствии замененных) информационных систем или отдельные файлы (например, текстовые файлы или документы форматов Microsoft Office).

Распространенным подходом является построение витрин данных, представляющих собой определенный срез информации в хранилище. При использовании этого подхода доступ к данным существенно упрощается, так как каждая категория пользователей работает с данными, которые представлены наиболее удобным для них образом.

Рассмотрим технологии интеграции данных. Существование множества разрозненных источников данных и отсутствие механизмов их объединения и согласования значительно усложняют использование информации. Интеграция данных позволяет представить корпоративные данные из разрозненных источников в виде цельной картины, обеспечивающей полноту и удобство доступа к ним. Интеграция данных может быть описана с помощью модели, которая включает приложения, продукты, технологии и методы (см. рис. 13). Приложения — это конечные программные средства, созданные поставщиками информационных систем в соответствии с требованиями клиента. Приложения используют один или несколько продуктов интеграции данных. Продукты — это готовые к использованию решения, поддерживающие одну или несколько технологий интеграции данных. Технологии представляют собой реализацию одного или нескольких методов интеграции данных.

Метод — это способ интеграции данных, не привязанный к какому-либо конкретному механизму доступа к данным.

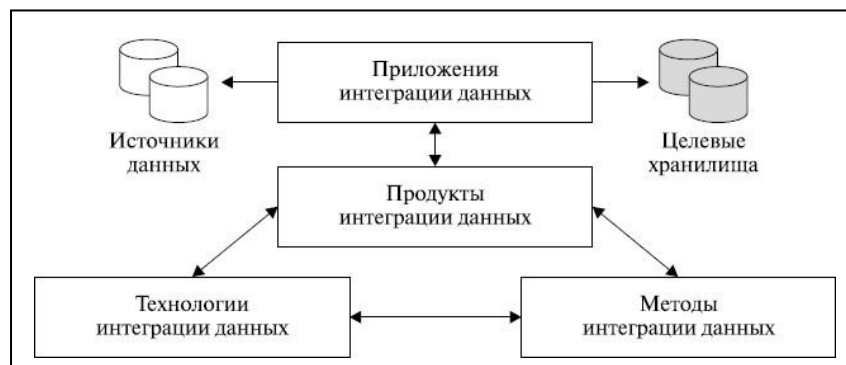


Рис. 13. Модель интеграции данных

Выделяют три основных метода интеграции данных: консолидация, федерализация и распространение (рис. 14).

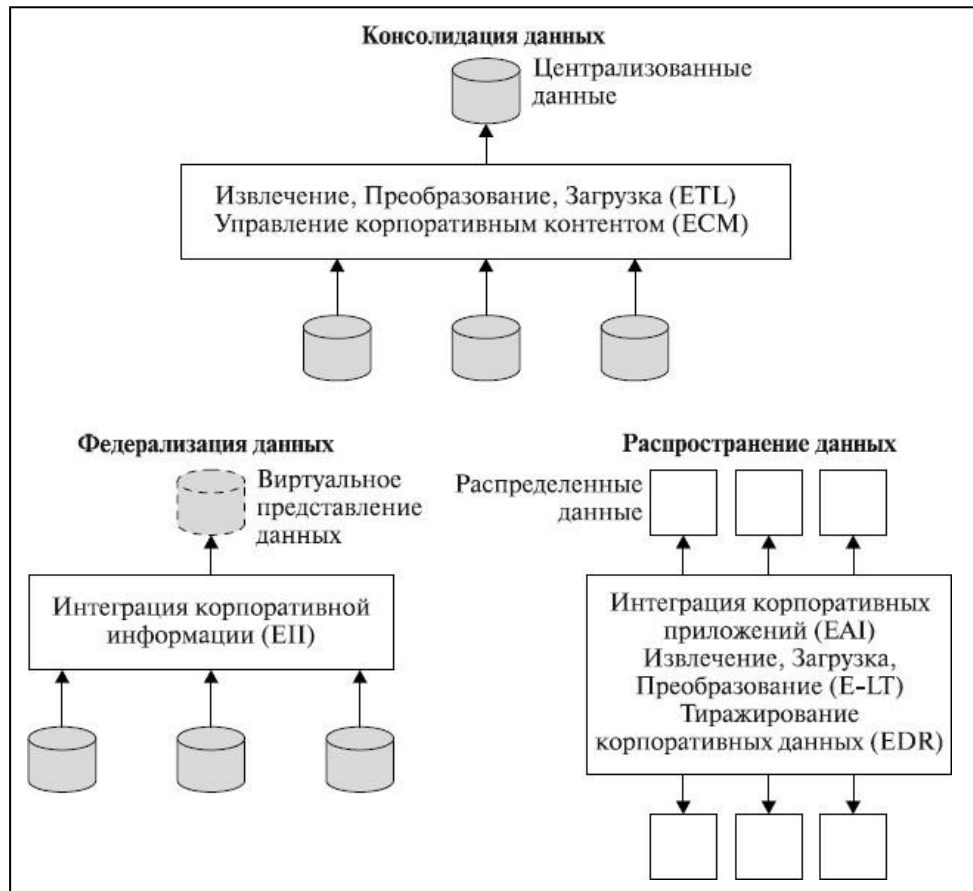


Рис. 14. Методы интеграции данных

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

«Извлечение» противопоставляется «проталкиванию» («передача», англ. push): при «проталкивании» когда в источнике данных произошло какое-то изменение, то он сам передает обновленные данные в целевое хранилище. При этом становится возможным проследить всю историю изменения данных во времени.

«Проталкивание» данных применяется для обновления данных в режиме, приближенном к реальному времени. Приложения, обеспечивающие такой

режим, называются приложениями оперативной интеграции данных. Такие приложения фиксируют изменения данных в исходной системе и передают эти изменения в целевое хранилище.

Соответственно, такой подход требует, чтобы приложения консолидации могли определить произошедшие изменения. Для этого используются различные методы захвата изменившихся данных (Changed Data Capture, CDC).

Извлечение и передача данных могут работать взаимосвязано. Например, приложения оперативной интеграции передают произошедшие изменения в некоторое временное хранилище, откуда накопленные изменения извлекаются пакетным приложением и загружаются уже в постоянное хранилище. При этом следует обратить внимание на то, что оперативная интеграция запускается в момент изменения данных, т. е. происходит «по событию» (event-driven). А пакетная интеграция запускается «по требованию» (ondemand) (рис. 15):

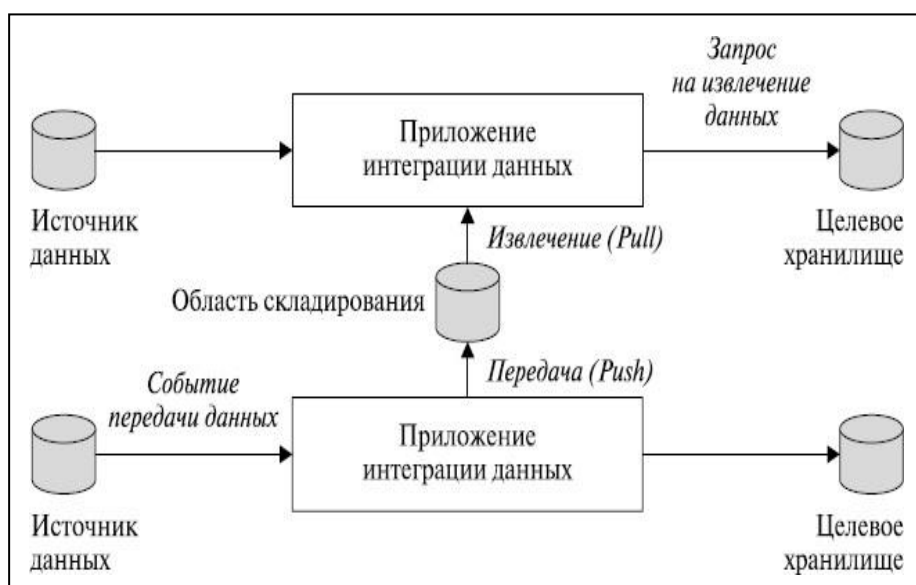


Рис. 15. Иллюстрация работы приложений интеграции данных

Бизнес-приложения, которые обрабатывают консолидированное хранилище, могут генерировать запросы к данным, создавать на их основе требуемые отчеты и проводить анализ этих данных. Как , результаты обработки консолидированных данных не передаются обратно в первичные системы из-за

ряда проблем, связанных с подобной синхронизацией. Однако существуют такие приложения, которые могут производить обновления консолидированного хранилища данных и передавать эти изменения обратно в первичные системы. Например, такой подход может использоваться для создания еженедельной модели ценообразования в торговой организации. Модель может оптимизироваться и обновляться в течение недели в центральном хранилище, а затем снова быть загружена в первичную систему в начале следующей недели. Схожим образом работают, например, системы прогнозирования и планирования бюджета.

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

Консолидация данных — это основной подход, который используется для построения и поддержки оперативных складов данных и корпоративных хранилищ данных. Выделяют две технологии консолидации: «Извлечение, Преобразование, Загрузка» (Extract, Transform, Load, ETL) и «Управление контентом предприятия» (Enterprise Content Management, ECM).

Областью применения ECM являются неструктурированные источники данных, такие как пользовательские документы и веб-страницы.

Задачей приложений ECM является построение такого «слоя» доступа к данным, который будет обеспечивать поиск и управление метаданными, шаблонами и различными версиями данных из этих источников. Технология ETL наилучшим образом подходит для извлечения и обработки значительных объемов данных из структурированных источников.

ETL - технология включает в себя этапы извлечения данных (extract) из различных источников, их преобразования (transform) и загрузки в целевое хранилище (load) (рис. 16).

Данные обычно извлекаются из структурированных источников, таких как оперативные базы данных (OLTP), структурированные файлы (например, в формате XML) или любые другие. Источником данных для приложений ETL

также может послужить результат работы приложений ЕСМ, когда проблема извлечения данных из неструктурированных документов уже решена.

Этап преобразования включает в себя реструктурирование, выверку, очистку и агрегирование данных. На этом этапе данные приводятся к единому стандарту представления, очищаются от ошибок, дубликатов и противоречивых значений. При необходимости данные также могут быть дополнены какими-то дополнительными значениями из других источников («обогащены»).

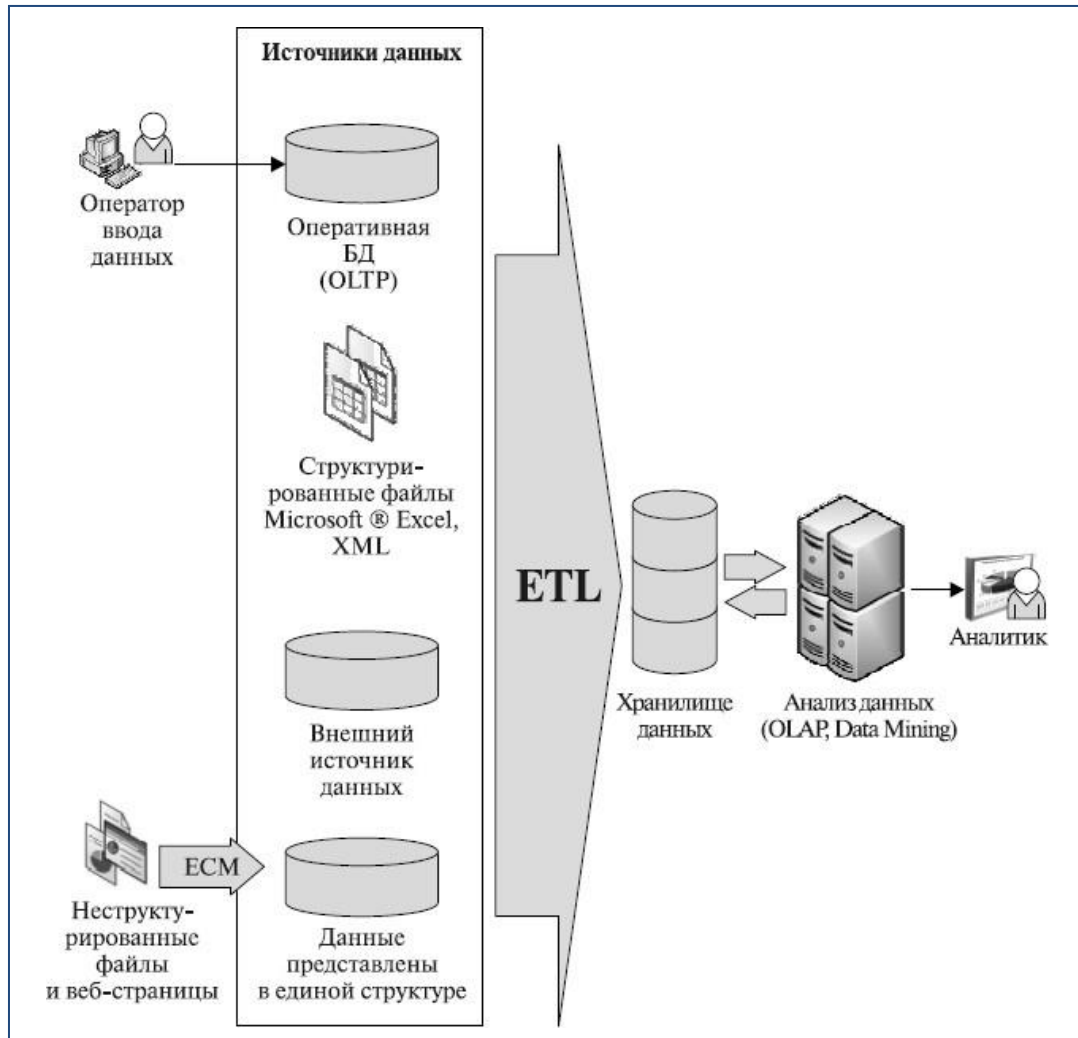


Рис.16. Технология ETL

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

Федерализация данных. Метод федерализации представляет требуемые в момент запроса данные из различных источников в виде единого виртуального представления (view). В отличие от консолидации данные не перемещаются из источников в хранилище. Когда бизнес-приложение генерирует запрос, используя виртуальное представление, то процессор федерализации данных извлекает требуемые данные из соответствующих распределенных источников, представляет их таким образом, чтобы они отвечали требованиям запроса, и отправляет результаты бизнес-приложению, от которого пришел запрос. Это происходит незаметно для бизнес-приложения, оно работает с виртуальным представлением как с целостным источником данных (рис.17).

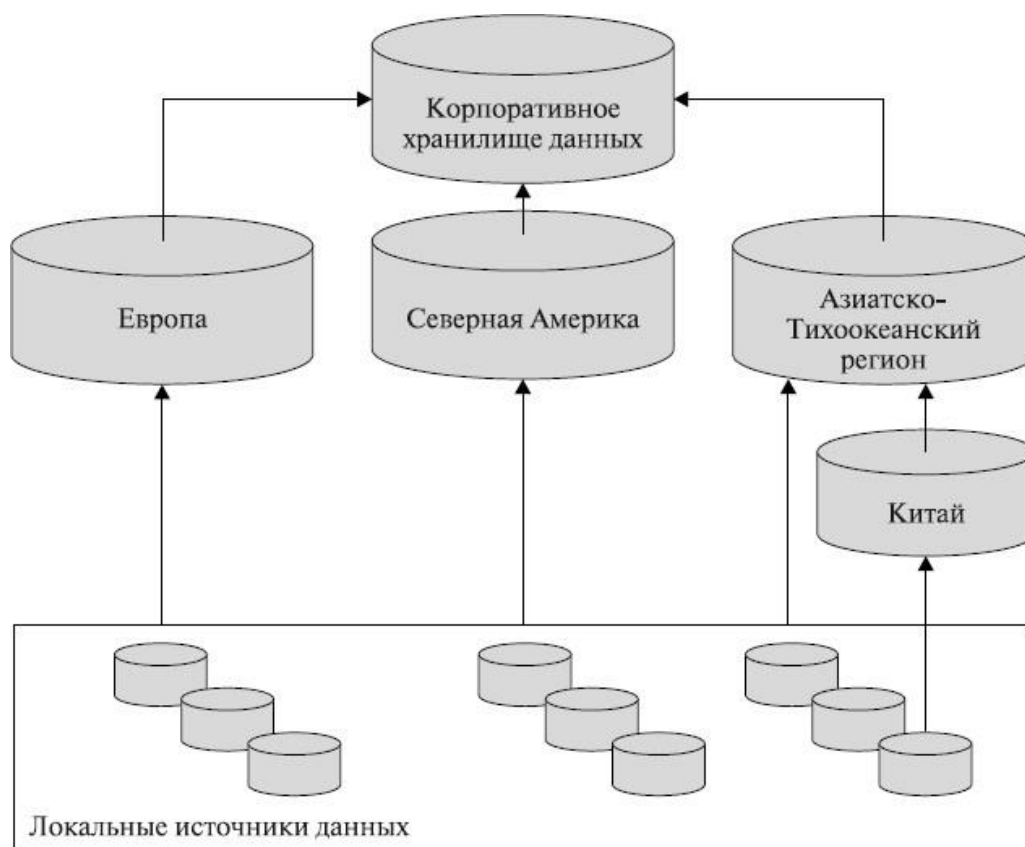


Рис. 17. Федерализованное представление по географическому признаку

Технологией, которая поддерживает метод федерализации, является «Интеграция корпоративной информации» (Enterprise Information Integration, EII), в рамках которой воплощается уровень абстракции между пользовательскими запроса-

ми и разнородными источниками данных. Пользовательские приложения генерируют запросы к этому уровню как к единому источнику данных.

В упрощенном представлении при использовании ЕИ запрос от пользователя поступает к серверу интеграции (рис. 18). Затем этот запрос разбивается на подзапросы, каждый из которых адресуется соответствующему источнику данных. Этими источниками могут быть корпоративные хранилища данных, базы различных приложений, структурированные документы и веб-страницы, операционные хранилища данных и даже неструктурированные документы, доступ к которым можно получить посредством ЕСМ. Сервер интеграции приводит каждый полученный от источника ответ к нужному виду в соответствии с требованиями исходного запроса, объединяет все ответы в едином представлении и передает обратно приложению, от которого поступил запрос.



Рис.18. Интеграция корпоративной информации

В основе метода распространения лежит копирование данных из системы-источника в одну или несколько систем-получателей. Такое копирование осуществляется посредством передачи изменений, произошедших в одной системе, в другие связанные с ней системы. Ключевое понятие метода распространения — «передают» («проталкивают»): если при консолидации и федерализации

зации данные извлекаются из исходных систем, то при распространении сами первичные системы передают произошедшие изменения системам-получателям. Благодаря такому подходу реализуется главное преимущество метода — обеспечение актуальности данных в режиме реального времени или близком к нему. Так как передача данных инициируется при наступлении определенного события, то метод распространения является событийным (eventdriven).

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

Среди технологий, поддерживающих распространение данных, выделяют интеграцию корпоративных приложений (Enterprise Application Integration, EAI), технологию «Извлечение, Загрузка, Преобразование» (Extract, Load, Transform, E-LT) и тиражирование корпоративных данных (Enterprise Data Replication, EDR). Тиражирование корпоративных данных в основном реализуется в приложениях резервного копирования, архивирования и равномерного распределения нагрузки между различными системами.

Технологии «Извлечение, Загрузка, Преобразование» (E-LT) перекликаются с ETL - технологией, реализующей метод консолидации. Действительно, буквы, образующие названия этих технологий, обозначают одни и те же операции. Однако различие заключается в последовательности шагов. В технологии E-LT преобразование данных происходит не до, а после их загрузки в целевое хранилище. При этом данные преобразуются не сразу после загрузки, а по пользовательскому запросу или определенному расписанию. Благодаря этому достигается высокая скорость обмена данными, но значительно урезаются возможности построения единого хранилища данных.

Интеграция корпоративных приложений (Enterprise Application Integration) позволяет различным приложениям «общаться» между собой посредством определенных стандартных интерфейсов. Приложения могут обмениваться между собой сообщениями, транзакциями и дан-

ными. При этом обмен происходит «прозрачно» для всех участников, без необходимости предусматривать месторасположение и особенности форматов каждой из вовлеченных обмен систем.

В рамках данного пункта технология EAI рассматривается с позиции методов передачи данных. Действительно, обеспечение единого пространства взаимодействия между различными приложениями невозможно без передачи данных между ними. Однако интеграция корпоративных приложений представляет собой гораздо более обширное понятие, нежели просто передача данных. EAI — это технология, обеспечивающая взаимодействие корпоративных приложений в режиме реального времени различными способами. В качестве метода интеграции данных здесь используется метод распространения.

В технологии EAI объем данных, которыми обмениваются приложения, обычно невелик. Передача данных практически всегда происходит в рамках одной единой бизнес-транзакции. Наглядным примером, иллюстрирующим такое взаимодействие, может быть бронирование путешествий: запрос на бронирование распределяется по отдельным, но взаимосвязанным системам бронирования авиабилетов, номера в отеле, трансферов и аренды автомобиля.

Рынок средств интеграции приложений Интеграция приложений — это сложная технология, которая позволяет различным корпоративным приложениям обмениваться сообщениями, транзакциями и данными.

Потребность в интеграции приложений очень велика. Как одна ERP-система не может обеспечить покрытие всех потребностей предприятия, поэтому возникает необходимость обеспечить совместную работу программных систем и СУБД разных производителей. Помимо этого, многие проекты ERP могут внедряться поэтапно, а для взаимодействия отдельных модулей также требуется единое информационное пространство. Зачастую возникает задача объединить системы разных подразделений или разных предприятий при их слиянии и поглощении.

Наконец, на большинстве предприятий работает одна или несколько унаследованных корпоративных систем предыдущих версий или собственной разработки, которые требуется связать с новой системой.

Интеграция корпоративных приложений призвана решить подобные проблемы. Комплексные системы интеграции для современных предприятий носят название интеграционных платформ. Ключевые понятия современной интеграции приложений. Ключевые понятия, на которые опираются интеграционные платформы на текущем этапе развития, это сервис-ориентированная архитектура (Service-oriented Architecture, SOA), сервисная шина предприятия (Enterprise Service Bus, ESB) и сервер приложений (Application Server). Следует обратить внимание на то, что в современной литературе и обзорах различных экспертов толкование и взаимосвязь этих понятий могут различаться. Однако можно выделить наиболее универсальную трактовку.

В основе сервис-ориентированной архитектуры лежит идея совокупности программных компонентов — сервисов, имеющих стандартные интерфейсы для использования этих компонентов и доступа к ним посредством сетевых протоколов. Интерфейсы в SOA независимы от платформ развертывания сервисов и технологий их реализации. В качестве сервиса в SOA может выступать как целое приложение, так и отдельные его функциональные модули.

Сервисная шина предприятия представляет собой инфраструктуру для реализации концепции SOA. Под ESB, как , подразумевается воплощение транспортного уровня, на котором происходит взаимодействие программных компонентов между собой. С точки зрения практической реализации ESB является программным обеспечением промежуточного уровня (middleware), универсальной шиной, к которой подключаются различные сервисы посредством унифицированных интерфейсов (адаптеров).

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

Основное предназначение EAS —обеспечивать разработку, хранение и запуск программных модулей, созданных с помощью поддерживаемых сервером моделей разработки(программных библиотек). Модели разработки — это каркас, при помощи которого программисты реализуют приложения, используя один или несколько языков программирования. Основные современные модели разработки — это JavaEnterpriseEdition (Java EE) и Microsoft.Net Framework.При выполнении приложений сервер распределяет между ними системные ресурсы (такие как память и процессорное время) и обеспечивает поддержку взаимодействия приложений между собой, СУБД и внешними системами посредством сетевых протоколов.

Функциональность большинства предлагаемых на рынке серверовприложений может быть значительно расширена за счет использованиядополнительных компонентов, набор которых различается в зависимости от производителя. Поэтому современный сервер приложений редковстречается «голым», а зачастую представляет собой комплекс программных средств, «обернутых» вокруг основного ядра, отвечающего за выполнение приложений.

Преобладающими архитектурами в настоящее время являются .Net,разрабатываемая компанией Microsoft, и JavaEnterpriseEdition (бывшаяJ2EE), которую развивают несколько производителей. Хотя эти технологиидоминируютнарынкесерверов приложений,имион не исчерпываетсяи представлен целым рядом менее заметных платформ, такими как, например, SpringFramework, RubyonRails, ApexCode, PlainOldJavaObject(POJO).

Следует отметить, что технология Java обладает некоторыми весомыми преимуществ перед .Net:

- независимость от аппаратной платформы предоставляет большуюгибкость в выборе оборудования, позволяет запускать одно и то жеприложение на разных платформах и переносить их между платформами;

- единые технологические стандарты дают заказчикам возможностьвыбирать из множества средств разработки, операционных среди приложений промежуточного слоя те, которые наиболее отвечают их требованиям.

Платформа от Microsoft реализует схожую функциональность, что и технологии Java, но в рамках систем Windows. Основные поставщики интеграционных платформ. Ключевые позиции на рынке корпоративных серверов приложений занимают компании IBM, Oracle, Microsoft и RedHat.

Рассмотрим основные особенности ведущих продуктов.

IBM2. Интеграционная платформа от компании IBM носит название WebSphere. Основой WebSphere является сервер приложений WebSphere Application Server (WAS), построенный на Java EE. Это масштабная платформа, которая включает в себя десятки компоненты средств мониторинга и администрирования, функциональность которых может быть существенно расширена за счет дополнительных пакетов (FeaturePacks). Это одна из сильных сторон предложения от IBM, так как она предлагает образцовые продукты, охватывающие практически все требования рынка. Помимо этого, в пользу IBM относят полную поддержку Java EE и многолетнее лидерство на рынке EAS, что обеспечивает широкую поддержку продуктов IBM сторонними производителями.

Oracle. Линейка продуктов корпорации Oracle представлена двумя семействами: перспективная Oracle WebLogic Server и Oracle Application Server. Семейство WLS включает в себя различные версии сервера приложений и пакет WebLogic Suite, расширяющий функциональные возможности сервера. Помимо этого имеются дополнительные инструменты администрирования, мониторинга и разработки. Оба семейства поддерживают Java.

RedHat (JBoss). JBoss EAS — это основанный на Java EE сервер приложений, который имеет свободно распространяемую версию с открытым исходным кодом. Эта версия не покрывается поддержкой со стороны производителя в отличие от платной корпоративной версии. Платформу интеграции JBoss можно дополнительно расширить средствами построения ESB, инструментами управления бизнес-процессами и другими модулями. Компания RedHat также предлагает среду разработки JBoss Developer Studio, платформу построения веб-порталов Enterprise Portal Platform, веб-сервер JBoss Enterprise Web Server и сервер

веб-приложений ApacheTomcat — все доступно как в виде свободно распространяемых версий без поддержки производителя, так и с платной поддержкой. RedHat прочно удерживает лидерство в сегменте продуктов с открытым исходным кодом. Большое количество дополнительных open-source модулей, в том числе сторонних производителей, позволяет ему нарастить функциональность, подчас не уступающую гигантам с закрытым кодом. Кроме того, EAS JBoss имеет отличную репутацию за свои технические качества.

Microsoft. Технология .Net Framework в совокупности с Internet Information Services (сокр. IIS, проприетарный набор серверов для нескольких служб Интернета от компании Microsoft) обеспечивают полное покрытие функционала серверов приложений, несмотря на то, что продукт с названием «сервер приложений» в линейке компании нет. Функцию сервисной шины предприятия выполняет Microsoft BizTalk Server, который включает в себя средства управления бизнес-процессами, интеграции приложений (EAI) и адаптеры ко всем основным протоколам и системам.

Преимуществом продуктов от Microsoft считается их тесная интеграция с другими продуктами компании, прежде всего СУБД Microsoft SQL Server и пакетом Microsoft Office.

SAP. Развитие SAP на рынке EAS основано на двух ключевых технологиях. Java EE сервер приложений SAP NetWeaver Application Server является основой инфраструктуры приложений, в состав которой входят, в частности, средства построения порталов SAP NetWeaver Portal, инструменты управления и интеграции бизнес-процессов SAP NetWeaver BPM и SAP NetWeaver Process Integration. Разработка бизнес-приложений ведется на проприетарном языке ABAP (Advanced Business Application Programming), который клиенты SAP используют для адаптации бизнес-приложений под свои нужды. Слабыми сторонами платформы SAP относительно лидеров рынка EAS считаются сравнительно небольшое количество внедрений и невысокая поддержка сторонних приложений.

Тема №9. «Оперативная аналитическая обработка данных в СППР»

Оперативная аналитическая обработка данных выражается термином OLAP (On-Line Analytical Process) и очень часто понимается как интерактивная аналитическая обработка данных. Возможность работы в интерактивном режиме при анализе данных выгодно отличает OLAP-системы от любых других систем подготовки отчетности, в том числе регламентированной. OLAP может работать с любыми OLTP-базами, но наилучшего эффекта можно достичь с использованием хранилищ данных (Data Warehouse). Основной задачей хранилища является предоставление данных для анализа в простой и понятной форме.

Для эффективной работы аналитику требуется централизация всех данных и соответствующее структурирование информации, а также удобные инструменты для просмотра и визуализации информации. Традиционные отчеты, построенные на основе единого хранилища, лишены гибкости. С ними нельзя работать в интерактивном режиме, т. е. их нельзя свернуть, развернуть, покрутить, чтобы получить необходимое представление данных. OLAP организует данные в виде многомерных кубов (cubes). Получаемая при этом модель данных является естественной и интуитивно понятной. В качестве осей многомерной системы координат выступают основные атрибуты анализируемого бизнес-процесса.

Аналитик при анализе показателей деятельности оперирует многомерными представлениями, и к примеру, рассматривает продажи в зависимости от потребительского спроса на конкретный товар, от региона продаж, от конкретного интервала времени. Исходя из этого, объем продаж можно представить в виде трехмерного куба (рис. 19.), грани которого изображают массивы данных по товарам, регионам и времени, а внутри куба находятся значения объема продаж.

Такое представление данных обеспечивает эффективный механизм для анализа и создания запросов, в которых, выбирая конкретный временной

период, товар и регион, можно получить соответствующий показатель объема продаж.

Многомерный анализ определяется как одновременный анализ по нескольким измерениям и по каждому измерению производится консолидация данных. Любое направление консолидации включает серию последовательных уровней обобщения, где каждый вышестоящий уровень соответствует большей степени агрегации данных по соответствующему измерению.

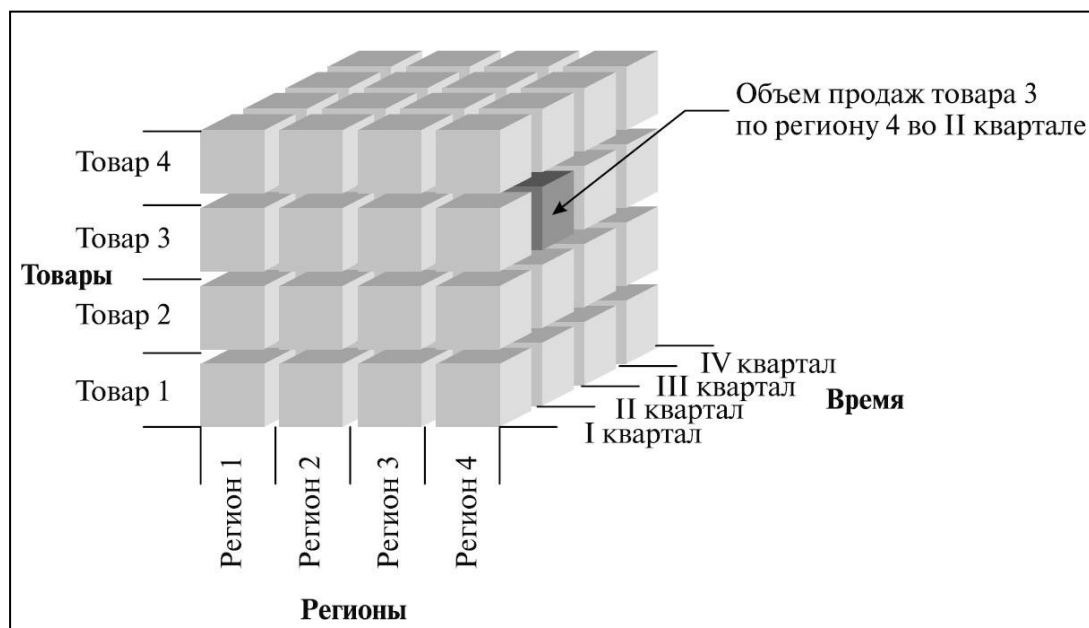


Рис.19. Вид информационного куба «Объем продаж»

Например, измерение Регион может определяться следующим направлением консолидации, состоящим из уровней обобщения «страна — регион — город — магазин — менеджер». Измерение Время может включать направление консолидации—«год—квартал—месяц—день».

BOLAP - такие многоуровневые объединения называются иерархиями, см. рис.20, где приведен пример географической иерархии. Средства оперативной аналитической обработки данных позволяют перейти на нужный уровень иерархии, но сам информационный куб для анализа не пригоден. Представить или изобразить адекватно трехмерный куб пользователь еще в состоянии, а вот пятнадцатимерный просто невозможно. Поэтому при анализе из многомерного куба извлекают обычные двумерные таблицы.

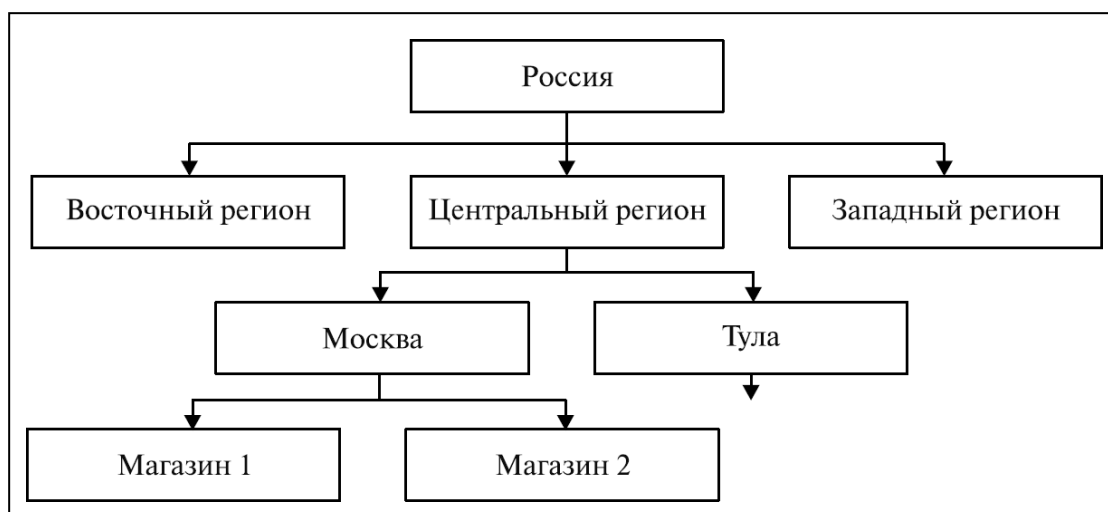


Рис.20.Пример иерархии

Рассмотрим кратко методы извлечения информации из кубов данных. Для извлечения информации из кубов данных используются различные операции манипулирования:

Операция «Вращение». Изменение порядка представления (визуализации) измерений называется Вращением (Rotate). Эта операция обеспечивает возможность визуализации данных в форме, наиболее комфортной для их восприятия. Например, если аналитик первоначально вывел отчет, в котором Регионы были перечислены по оси X, а Товары по оси Y, он может решить, что такое представление неинформативно, и поменять местами координаты (выполнить Вращение на 90°).

Отношения и Иерархические Отношения. В нашем примере значения показателей определяются только тремя измерениями. На самом деле их может быть гораздо больше, и между их значениями обычно существует множество различных Отношений (Relation). В свою очередь, множество Отношений может иметь иерархическую структуру — Иерархические отношения. И часто более удобно не объявлять новые измерения и затем устанавливать между ними множество Отношений, а использовать механизм Иерархических отношений. В этом случае все потенциально возможные значения из различных измерений объединяются в одно множество. Например, мы можем добавить к множеству значений измерения Менеджер («Петров», «Сидоров», «Иванов»),

«Смирнов») значения измерения Подразделение («Филиал 1», «Филиал 2», «Филиал 3») и измерения Регион («Восток», «Запад») и затем определить между этими значениями Отношение Иерархии.

Операция Агрегации. Сточки зрения пользователя измерения Страна, Регион, Город, Магазин являются точно такими же измерениями, как Менеджер (рис. 21.). Но каждое из них соответствует новому, более высокому уровню агрегации значений показателя Объем продаж. В процессе анализа пользователь переходит от детализированных данных к агрегированным, т. е. производит операцию Агрегации (DrillUp).



Рис. 21. Операция подъема (агрегация)

Операция Детализации. Переход от более агрегированных к более детализированным данным называется операцией Детализации (DrillDown). Например, начав анализ на уровне Региона, пользователь может захотеть получить более точную информацию о работе конкретного Магазина или Менеджера (рис. 22.).

Формирование «Среза». Пользователя редко интересуют все потенциально возможные комбинации значений измерений. Более того, он практически никогда не работает одновременно сразу со всем кубом данных. Подмножество куба, получившееся в результате фиксации значения одно-

го или более измерений, называется Срезом (Slice), а сама операция называется «разрезанием» куба.



Рис.22. Операция спуска (детализация)

Аналитик как бы берет и «разрезает» измерения куба по интересующим его меткам. Этим способом аналитик получает двумерный срез куба (отчет) и с ним работает. Структура отчета представлена на рис. 23.

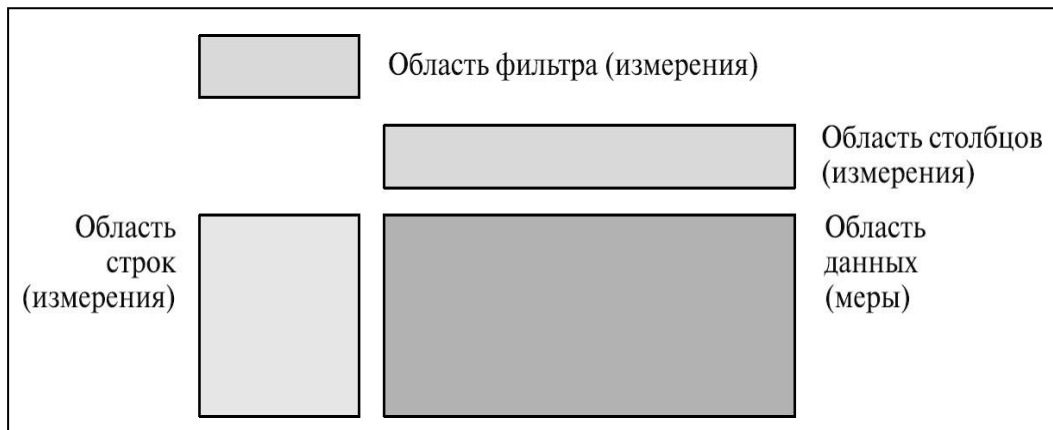


Рис.23. Структура аналитического отчета

OLAP-отчет — управляемая, динамическая OLAP-таблица, которая сопровождается синхронной диаграммой (графиком). Измерения отображаются в названиях строк и столбцов таблицы, соответствующие им факты и итоги (аг-

регированные факты) — в ячейках таблицы. Колонки и строки являются инструментами управления таблицей.

Пользователь может перемещать их, фильтровать, сортировать, детализировать/обобщать и выполнять другие OLAP-операции. При этом таблица автоматически вычисляет новые итоги (агрегаты). Управляя OLAP-таблицей, можно из одного набора данных сформировать множество отчетов. В OLAP-таблице можно условно выделить несколько рабочих областей (рис. 24.).

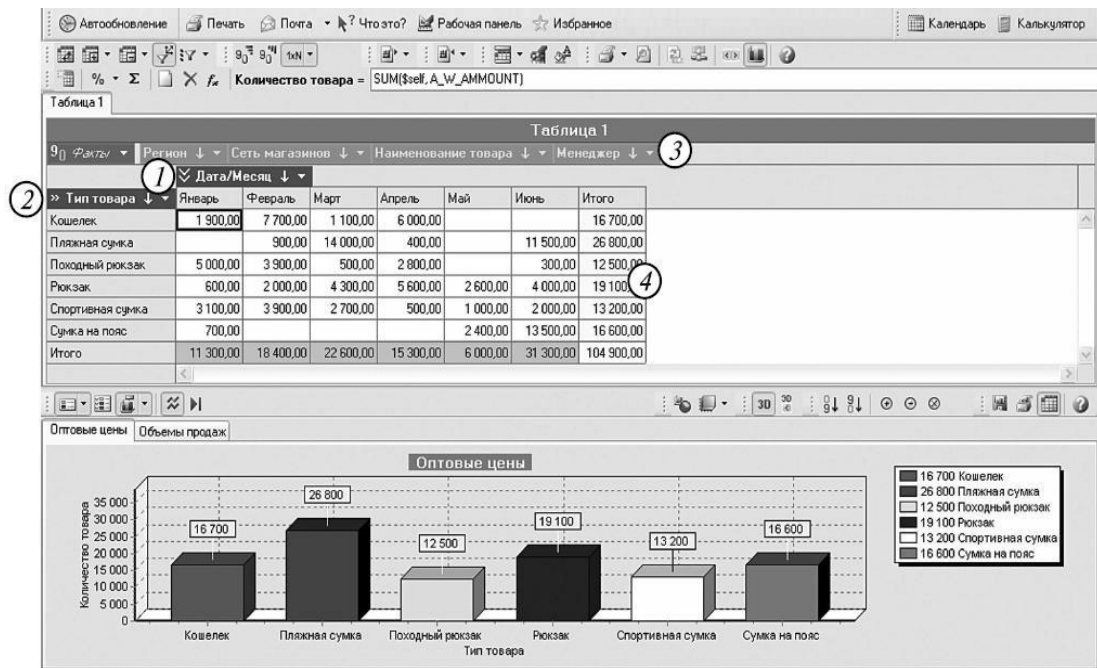


Рис.24. OLAP-отчет, 1—область колонок; 2—область строк; 3—область фильтров; 4—OLAP-таблица

Один из базовых принципов OLAP — способность интуитивно манипулировать данными—требует быстрого извлечения информации.

В качестве примера типа данных, который допустимо рассчитать заранее, можно привести сводные данные — например, показатели продаж по месяцам, кварталам или годам, для которых исходными данными являются ежедневные показатели (табл. 3). При получении команд пользователя OLAP-отчет перестраивается за доли секунды, позволяя из одной отчетной формы

создать множество других. При этом используются заранее рассчитанные агрегатные значения (табл. 4).

OLAP-отчет отличается рядом принципиальных особенностей:

- агрегированный отчет, позволяющий углубиться в детали;
- отчет, позволяющий легко изменять аналитические разрезы;
- отчет, позволяющий фильтровать данные по сочетаниям их категорий;
- экранный отчет с возможностью вывода на принтер;
- отчет, настраиваемый пользователем без программирования;
- отчет, имеющий простую, регулярную структуру;
- отчет, предоставляющий пользователю интерактивный способ работы с данными.

Таблица 3. OLAP – отчет

Регион	Товар	Квартал	Сумма
Москва	Товар 1	Квартал 1	25 000 000
		Квартал 2	5 000 000
		Итого	30 000 000
	Товар 2	Квартал 1	5 000 000
		Квартал 2	5 000 000
		Итого	10 000 000
	Итого		
Санкт-Петербург	Товар 1	Квартал 1	5 000 000
		Квартал 2	1 000 000
		Итого	6 000 000
	Товар 2	Квартал 1	2 000 000
		Квартал 2	1 000 000
		Итого	3 000 000
	Итого		
Всего			49 000 000

Таблица 4. OLAP – отчет по регионам

Регион	Сумма
Москва	40 000 000
Санкт-Петербург	9 000 000
Итого	49 000 000

Оперативность при аналитической обработке необходима для поддержания необходимого ритма работы специалиста в процессе анализа и принятия решений. Она реализуется за счет интерактивности, непосредственного доступа к данным, возможности сформулировать запрос к данным для получения динамических отчетов. Оперативность и гибкость создания динамических отчетов требуют возможности выбора, задания или изменения:

- состава показателей и связанных с ними разрезов и реквизитов;
- вида таблицы, уровня детальности (агрегирования) разрезов и конкретных значений реквизитов;
- условий отбора значений фактов и реквизитов, формул расчета;
- форматирования табличного представления и графического оформления результатов.

Имея инструментарий, обладающий подобными комбинаторными свойствами, аналитик может проводить интерактивное исследование данных, создавая нужные отчеты и представления данных.

Ему предоставлена возможность менять структуру отчета, переставляя строки и столбцы таблицы, переключаться с табличного представления на графическое, убирать ненужные сведения и детализировать нужные, переходить с детальных данных на итоги.

Средства OLAP как раз и являются такими инструментами, обеспечивающими оперативность и гибкость в получении различных отчетов, существенно расширяя возможности участников процесса подготовки и принятия решений.

Аббревиатура OLAP была впервые введена Э. Ф. Коддом, известным ученым в области реляционных БД, создателем широко распространенной реляционной модели данных.

Кодд отмечал основной недостаток реляционной модели данных как невозможность «объединять, просматривать и анализировать данные с точки зрения множественности измерений», т. е. самым понятным для корпоративных аналитиков способом. Исходя из этого, он определил общие требования к

системам OLAP, расширяющим функциональность реляционных СУБД и включающим многомерный анализ как одну из своих характеристик.

Рассмотрим требования Кодда к средствам оперативной аналитической обработки, которые вошли в концепцию комплексного многомерного анализа данных, накопленных в хранилище, в виде 12 основных правил:

1. Многомерное концептуальное представление данных (Multi-Dimensional Conceptual View).
2. Прозрачность (Transparency).
3. Доступность (Accessibility).
4. Устойчивая производительность (Consistent Reporting Performance).
5. Клиент-серверная архитектура (Client-Server Architecture).
6. Равноправие измерений (Generic Dimensionality).
7. Динамическая обработка разреженных матриц (Dynamic Sparse Matrix Handling).
8. Поддержка многопользовательского режима (Multi-User Support).
9. Неограниченная поддержка кроссмерных операций (Unrestricted Cross-dimensional Operations).
10. Интуитивное манипулирование данными (Intuitive Data Manipulation).
11. Гибкий механизм генерации отчетов (Flexible Reporting).
12. Неограниченное количество измерений и уровней агрегации (Unlimited Dimensions and Aggregation Levels).

Nigel Pendse и Richard Creeth, считая, что для большинства людей слишком обременительно помнить все правила Кодда, переработали их в так называемый тест FASMI (Fast Analysis of Shared Multidimensional Information). В тесте OLAP-определение резюмировано только пятью ключевыми словами: Быстрый Анализ Разделяемой Многомерной Информации.

Это определение по достоинству было оценено и до сих пор используется для характеристики OLAP-средств.

Тест FASMI расшифровывается следующим образом:

Fast — означает, что система должна обеспечивать выдачу большинства ответов пользователям в пределах нескольких секунд.

Analysis—означает, что система может справляться с любым логическим и статистическим анализом, характерным для данного приложения, и обеспечивает его сохранение в виде, доступном для конечного пользователя.

Shared — означает, что система осуществляет все требования защиты конфиденциальности, если множественный доступ для записи необходим, обеспечивает блокировку модификаций на соответствующем уровне. Система должна предоставлять широкие возможности разграничения доступа к данным и одновременной работы многих пользователей.

Multidimensional — система должна обеспечить многомерное концептуальное представление данных, включая полную поддержку иерархий.

Information — возможность обращаться к любой нужной информации независимо от ее объема и места хранения.

В настоящее время OLAP-технологии становятся обычным рабочим инструментом менеджеров всех уровней предприятия.

Рассмотрим сферы применения OLAP-технологий:

Продажи. Закупки. Цены. Маркетинг. Склад. Движение денежных средств. Бюджет. Бухгалтерские счета. Финансовая отчетность. Посещаемость сайта. Результаты выборов. Результаты социологических опросов. Объемы производства. Потребление электроэнергии. Потребление расходных материалов. Использование помещений. Заработная плата. Текучесть кадров в регионе. Текучесть кадров на предприятии. Метеорология. Уровень жизни населения. Пассажирские перевозки. Грузовые перевозки. Простои транспорта (вагонов, самолетов, пароходов, грузовиков). Автомобильный трафик. Заболеваемость персонала (учащихся, трудящихся). Многомерный хит-парад. Выбор недвижимости (офисов, складов, квартир). Урожайность агро-культур. Использование автотранспорта.

Тема №10. «Хранилища данных, архитектура, классификация, примеры проектных решений»

Одной из главных целей разработки хранилищ данных (ХД) является информационное обеспечение для систем поддержки принятия решений по всем или основным видам деятельности организации. Вид деятельности организации или направление бизнеса совместно со спектром соответствующих ему бизнес-процессов определяют предметную область ХД.

ХД — это сложная компьютерная система. Под архитектурой ХД понимают совокупность программно-аппаратных компонент, совокупность технологических и организационных решений, предпринимаемых для создания, разработки и функционирования ХД, т.е. выбор аппаратного и программного обеспечения, выбор способов взаимодействия программно-аппаратных компонент, выбор способа решения проектной задачи по разработке и созданию ХД. Как правило, в архитектуру ХД включают следующие компоненты:

- средства извлечения данных из различных БД OLTP-систем, унаследованных систем и других внешних источников данных;
- средства трансформации и очистки данных;
- программное обеспечение БД;
- средства для соединения источников данных с хранилищем и клиентов с сервером.

Кроме этого, необходимы специальные программные средства проектирования хранилища, средства работы с репозитарием метаданных и OLAP-средства.

Характер и масштаб решаемых задач анализа данных организации оказывает решающее значение на выбор архитектуры ХД и методы его проектирования. С одной стороны, ХД создается для решения конкретных, строго определенных задач анализа и воспроизводства новых данных, с другой — ХД должно обеспечивать корпоративную отчетность в рамках всей организации. Таким образом, определяющим моментом в построении ХД являются задачи

обработки, анализа данных и генерации отчетов.

Характер и масштаб решаемых задач анализа данных определяет и подходы к выбору архитектуры и проектированию ХД.

Выбор архитектуры ХД относится к сфере компетенции руководителя ИТ-проекта по созданию системы складирования данных. На такой выбор влияют несколько различных факторов:

- инфраструктура организации,
- производственная и информационная среда организации,
- управление и контроль,
- масштабы проекта,
- возможности аппаратно-технологического обеспечения,
- готовность персонала и имеющиеся ресурсы.

Правильный выбор архитектуры ХД обычно определяет успех конкретного проекта по созданию системы складирования данных. Существует несколько факторов, влияющих на принятие решений о выборе способа реализации: время, отведенное на проект, возврат инвестиций, скорость ввода ХД в эксплуатацию, потребности пользователей, потенциальные угрозы по переделке, требования к ресурсам, необходимым в определенный момент времени, выбранная архитектура ХД, совокупная стоимость владения ХД.

На рис. 25 приведена типовая обобщенная концептуальная схема для архитектуры ХД. В конкретных решениях некоторые компоненты могут отсутствовать.

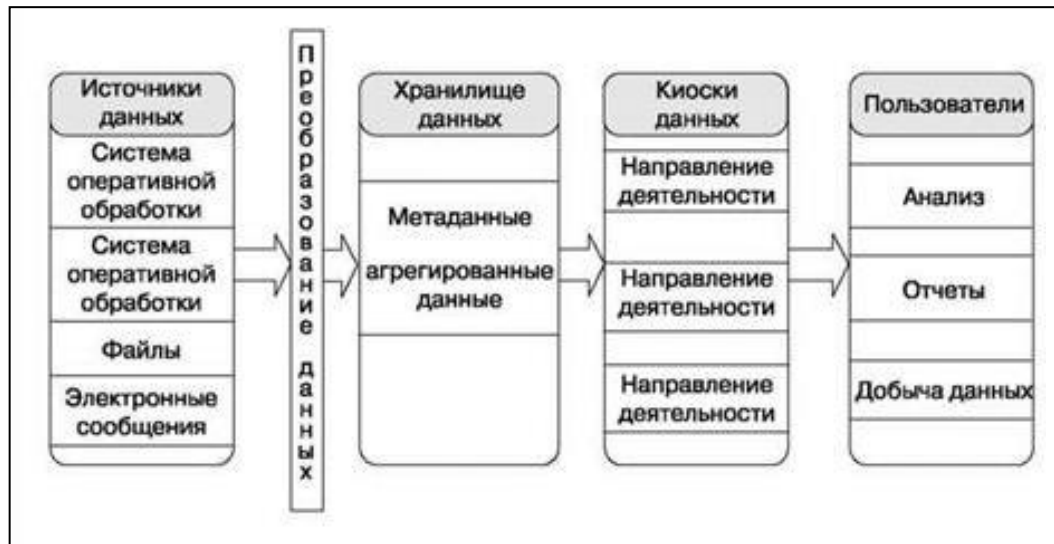


Рис. 25. Типовая обобщенная концептуальная схема для архитектуры ХД

Рассмотрим компоненты типовой архитектуры хранилища данных:

- Программное обеспечение промежуточного слоя для обеспечения доступа к сети и доступа к данным. Сюда относят сетевые и коммуникационные протоколы, драйверы, системы обмена сообщениями и т.д.
- Базы данных систем оперативной обработки данных (OLTP) и данные внешних источников. Для OLTP-систем характерна целевая направленность на эффективную обработку структур данных в рамках относительно небольшого числа четко определенных типовых транзакций.
- Предварительная обработка и загрузка данных. Предварительная обработка, связанная с фильтрацией, очисткой и преобразованием данных из OLTP-систем и внешних источников, обычно выполняется в некотором промежуточном файле, который называется иногда загрузочной секцией. После обработки данные загружаются в ХД.
- Хранилище данных. Представляет собой ядро системы складирования данных. Это могут быть один или несколько серверов БД для поддержки ХД.
- Метаданные. Метаданные представляют собой репозиторий, который играет роль справочника о данных. Он включает терминологию предметной области, сведения об источниках данных, описание источников исходных

данных, сведения об алгоритмах обработки исходных данных и т.д.

- Уровень доступа к данным включает в себя программное обеспечение, которое обеспечивает взаимодействие конечных пользователей с данным ХД. Универсальным средством общения служат SQL и его расширения.
- Уровень информационного доступа обеспечивает непосредственное общение пользователя с ХД. В качестве таких средств могут выступать стандартные пакеты MSOffice, LotusNotes или специальные программные продукты.
- Уровень администрирования. Компоненты этого уровня отслеживают выполнение процедур обновления ХД, включающих процедуры подкачки данных, обновления индексов, суммирования и агрегации данных, репликацию данных в распределенной вычислительной среде, авторизацию пользователя и разграничение доступа.

Типовыми архитектурами для систем складирования данных принято считать следующие:

- системы с глобальным ХД;
- системы с независимыми киосками данных;
- системы с интегрированными киосками данных;
- системы, разработанные на основе комбинации из вышеперечисленных архитектур.

Глобальное хранилище данных (GlobalDataWarehouse), или хранилище данных масштаба организации, — это такое ХД, в котором будут поддерживаться все данные организации или большая их часть. Это наиболее полное интегрированное ХД с высокой степенью интенсивности доступа к консолидированным данным и использованием его всеми подразделениями организации или руководством организации в рамках основных направлений деятельности организации. Таким образом, такое ХД проектируется и конструируется на основе потребностей аналитической информационной поддержки организации в целом.

Глобальное ХД необязательно должно быть реализовано физически

как централизованное. Термин «глобальное» используется для отражения масштаба использования и доступа к данным в рамках всей организации. Глобальное ХД может быть физически как централизованным, так и распределенным.

Централизованное глобальное ХД характерно для организаций, расположенных территориально в одном здании. Оно поддерживается отделом информационных систем организации. Распределенное глобальное ХД также может быть использовано в рамках организации в целом. Оно физически распределяется по подразделениям организации и также поддерживается отделом информационных систем.

Поддержка ХД отделом информационных систем вовсе не означает, что именно эта служба управляет ХД. Например, отдельные части распределенного ХД могут управляться в рамках подразделений или направлений бизнеса.

Управление ХД определяет:

- какие данные должны поступать в ХД;
- когда данные должны поступать в ХД;
- когда данные должны обновляться;
- кому разрешен доступ к данным в ХД.

Таким образом, для глобального ХД существуют два основных архитектурных решения, рис. 26.

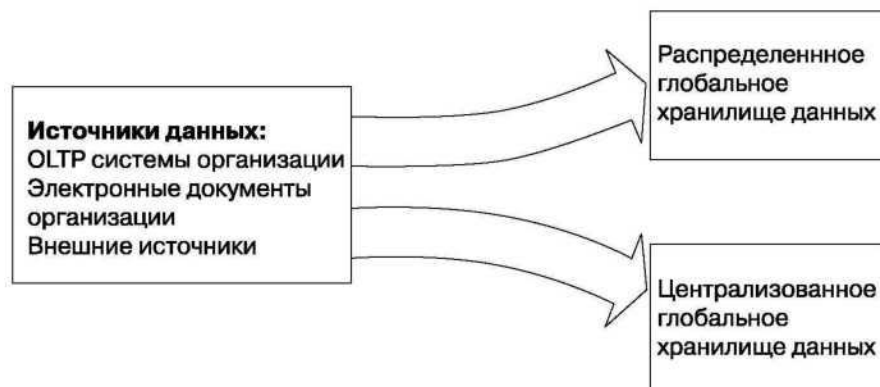


Рис.26. Основные архитектурные решения глобального ХД

Данные для ХД обычно извлекаются из OLTP-систем организации,

электронных документов организации и внешних источников данных. После фильтрации, очистки и преобразования они помещаются в ХД. Затем пользователи получают доступ к этим данным в соответствии с правилами управления доступом к данным, принятыми в организации.

Преимуществом глобального ХД является предоставление конечным пользователям доступа к информации в масштабах предприятия, недостатком — высокие затраты на реализацию, в том числе временные затраты на создание ХД.

Независимые киоски данных включают в себя автономные или независимые киоски данных (Stand-alone Data Marts), которые управляются рабочими группами, отделами или направлениями бизнеса и разрабатываются исключительно для реализации аналитических потребностей последних. Данные могут извлекаться из OLTP-систем, в частности, при помощи информационных служб организации. Информационные службы могут поддерживать вычислительную среду для киосков данных, но не управляют информацией в них. Данные в киоски могут поступать и из глобального ХД.

Киоски данных могут быть взаимозависимы или взаимосвязаны. Такая архитектура ХД включает в себя совокупность киосков данных, которые управляются рабочими группами, отделами или направлениями бизнеса, но разрабатываются в рамках единой для организации схемы удовлетворения информационных и аналитических потребностей. Для взаимосвязанных киосков данных типична распределенная архитектура реализации. Несмотря на то, что отдельные киоски данных реализуются в рамках рабочих групп, подразделений и направлений бизнеса, они могут быть интегрированы для представления данных в рамках организации в целом. Фактически, на наиболее высоком уровне интеграции, они могут стать глобальным ХД. В такой архитектуре пользователи одних подразделений могут получать доступ к данным других подразделений в рамках своих полномочий.

Требования интеграции данных в рамках архитектуры взаимосвязанных киосков данных делают реализацию ХД более сложной по сравнению с неза-

висимыми киосками данных. Например, необходимо решить вопрос, кто будет управлять данными, и кто будет поддерживать вычислительную среду. Важным становится вопрос о том, что делать с данными, которые являются общими для нескольких киосков данных, а также как разработать схему разграничения доступа пользователей к киоскам данных в рамках всей организации.

Главным достоинством создания ХД такой архитектуры является более глобальное представление данных. Взаимосвязанные киоски данных могут управляться в рамках того подразделения, в котором они создаются. Реализация такой архитектуры не выдвигает высоких требований к программно-аппаратному обеспечению, и стоимость ее может быть невысокой. Однако время реализации будет больше по сравнению с независимыми киосками данных. Возрастают также сложность и стоимость процедур проектирования.

В заключение следует отметить, что развитие программно-вычислительных средств позволяет создавать так называемые виртуальные ХД, которые работают над OLTP-системами, ХД с многоуровневой архитектурой и так называемые встроенные ХД, которые встраиваются в существующую систему обработки данных организации.

Рассмотрим методологические подходы в организации работ по созданию хранилища данных. К ХД применимы следующие основные подходы:

- «сверхувниз» (Top down design);
- «снизуверх» (Bottom down design);
- «изсередины» (Middle of design).

На выбор подхода к реализации ХД оказывают влияние следующие факторы:

- состояние текущей информационной инфраструктуры организации;
- имеющиеся в наличии ресурсы и требования по возврату инвестиций;
- потребности организации в интегрированном представлении дан-

ных о своей деятельности;

- скорость реализации.

Выбор методологического подхода к реализации ХД влияет на объем и тщательность проектирования.

Подход «сверху вниз» требует детального планирования и проектирования ХД в рамках ИТ-проекта до начала выполнения проекта. Это связано с тем, что необходимо привлекать всех потенциальных пользователей ХД для выяснения их информационных потребностей в аналитической обработке данных, принимать решения об источниках данных, безопасности, структурах данных, качестве данных, стандартах данных. Все эти работы должны быть документированы и согласованы. При этом подходе модель ХД должна быть разработана до начала реализации.

Обычно такой подход практикуют при создании глобального ХД. Если киоски данных включаются в конфигурацию, то они могут быть построены позже. Достоинством такого подхода является получение более согласованных определений данных и бизнес-правил организации в самом начале работы над созданием ХД.

Для этого подхода характерны большие затраты времени, что откладывает начало реализации и задерживает возврат инвестиций. Подход «сверху вниз» хорошо применять в организациях с четко организованной информационно-вычислительной структурой, когда программно-аппаратная платформа определена и существуют слаженно работающие источники данных.

При использовании подхода «снизу вверх» начинают с планирования и проектирования киосков данных подразделений без предварительной разработки глобальной информационно-вычислительной инфраструктуры организации. Такой подход является более приемлемым во многих случаях, поскольку он быстрее приводит к конечным результатам. У него есть и недостатки: данные могут дублироваться и быть несогласованными в разных киосках данных. Чтобы избежать этого, необходимо тщательное планирование и проектирование.

Подходы «снизу вверх» и «сверху вниз» могут комбинироваться в зависимости от поставленных перед руководителем проекта по созданию ХД целей. Подход «проектирование из середины» представляет собой комбинацию вышеперечисленных подходов, которые применяются как бы по спирали. Сначала создается ядро системы (подход «сверху вниз»), а затем оно поэтапно наращивается за счет добавления новой или дополнительной функциональности (подход «снизу вверх»). Таким образом, на каждом витке спирали может быть использован каждый из двух указанных выше подходов.

Существуют и другие комбинации. Выбор подхода к реализации ХД наряду с выбором архитектуры ХД определяет тактические решения в проектировании и управлении проектом создания системы складирования данных.

Приведем краткий обзор решений основных производителей программного обеспечения для разработки ХД. При изложении материала используем следующую схему:

- название проекта компании и его цель;
- архитектурные решения;
- СУБД и используемая модель данных;
- возможности языка обработки данных;
- степень охвата жизненного цикла (анализ — проектирование — реализация — поддержка);
- возможные конкурентные преимущества.

IBM. Решения компании IBM представлены подробно на сайте <https://www.ibm.com/ru>.

Oracle. Решения, предлагаемые компанией, предоставлены на сайте <https://www.oracle.com/index.html>.

NCR. Решение этой компании в области складирования данных ориентировано на организации, у которых имеются потребности в системах DSS (система поддержки и принятия решений) и системах OLAP. Предлагаемые продукты прописаны на сайте <http://www.ncr.ru>. NCR предлагает решения по автоматизации гипермаркетов, супермаркетов, универмагов, бутиков, ресто-

ранов, бензоколонок и других предприятий торговли.

SASInstitute. Компания считает себя поставщиком решений для организации ХД. Компания предлагает методологию RapidDataWarehousing для быстрого создания и наполнения ХД., см. <https://www.sas.com/ru>.

Конкурентным преимуществом компании является наличие у нее длинной линейки программных продуктов для статистического и сравнительного анализа данных, который интегрирован в ее методологию построения и использования ХД.

Sybase. Стратегии компании приведены на официальном сайте <https://www.sybase.ru/>. Конкурентным преимуществом компании является наличие набора программных продуктов для обеспечения полного жизненного цикла разработки ХД.

Microsoft. Конкурентным преимуществом компании является наличие у нее набора программных продуктов для обеспечения разработки и поддержки ХД, см. сайт <https://www.microsoft.com/ru>.

Тема №11. «Области применения хранилищ данных»

Приведем краткий обзор сфер применения технологий хранилищ данных.

Концепция хранилищ данных находит применение во многих сферах бизнеса, науки и управления. Типовые решения для бизнеса можно разделить на следующие основные группы.

1. Разработка основы для создания аналитических подсистем сопровождения бизнеса.
2. Разработка ХД как составной части виртуального предприятия.
3. Разработка ХД для цифровых (электронных) библиотек и мультимедиа.

Имеется тенденция расширения проникновения концепции в те сферы бизнеса, где необходимо выполнять, с одной стороны, сравнительный анализ, искать зависимости в данных, выявлять тренды в рядах динамики, а с другой — использовать системы складирования данных в связке с системами операционной обработки.

Рассмотрим несколько примеров применения технологии складирования данных в области создания аналитических подсистем информационного сопровождения бизнеса.

Компонентами типовой архитектуры ХД являются:

- программное обеспечение промежуточного слоя. Основное назначение этих компонент состоит в обеспечении доступа к сети и доступа к данным;
- БД OLTP систем и данные внешних источников;
- предварительная обработка и загрузка данных;
- ХД, реализованное средствами СУБД;
- метаданные, которые играют роль справочника о данных;
- уровень доступа к данным — программное обеспечение, которое обеспечивает взаимодействие конечных пользователей с данными ХД;
- уровень информационного доступа, который обеспечивает непосредственное общение пользователя с ХД;

- уровень администрирования.

Основные сферы применения технологии складирования данных приведены в табл. 5. Имеется тенденция расширения проникновения концепции в те сферы бизнеса, где необходимо выполнять, с одной стороны, сравнительный анализ, искать зависимости в данных, выявлять тренды в рядах динамики, а с другой — использовать системы складирования данных в связке с системами операционной обработки.

Таблица 5. Области применения концепции складирования данных

Сфера деятельности	Комментарий
Сегментация рынка	CRM
Планирование продаж, прогнозирование и управление	CRM, SCM
Опека клиентов	CRM
Схемы лояльности	
Проектирование и разработка продуктов	MRP/ERP
Интеграция цепочки поставок	SCM, ERP/MRP, SCP, SCE, DRP, JIT
Инновации и новые возможности	
Новые возможности приложений с использованием Интернет/Интранет	eBusiness, TMP
Приложения, основанные на агентах программного обеспечения	
Приложения для извлечения знаний и киберорганизация	EIF, виртуальное предприятие
Распространение DW из области стратегического планирования в область бизнес операций	VDW
Приложения для вертикальных секторов индустрии	CRM, TMP
Готовые DW (off-the-shelf)	
Автоматизация принятия решений	DSS. EIS
Новые категории оперативных приложений, ориентированные на клиента	OLAP
Сбор и анализ экспериментальных данных в химии, физике, биологии	EDW
Хранение мультимедийной информации в DW	DL

Сокращения:

- CRM (Customer Relationship Management) — управление взаимоотношениями с клиентами;
- SCM (Supply Chain Management) — управление цепочкой поставок;

- SCP (SupplyChainPlaning — планирование управления цепочкой поставок;
- SCE (SupplyChainExecuting) — реализация управления цепочкой поставок;
- DRP (DistributionResourcePlaning) — планирование потребностей распределения;
- JIT (Just-in-Time) — точновсрок;
- MRP (Manufacturing Resource Planing) — планирование материальных затрат;
- VDW (VirtualDataWarehouse) — виртуальные хранилища данных;
- DL (Digital Library) — цифровые библиотеки;
- ERP (EnterpriseResourcePlaning) — системы планирования масштаба предприятия;
- TMP (TradingPartnerManagement) — управление деловыми партнерами;
- EIF (Enterprise Information Factory) — виртуальное предприятие.

Приведем примеры применения технологий складирования данных в области создания аналитических подсистем информационного сопровождения бизнеса.

Аналитические CRM-системы. Оперативные системы CRM содержат следующие компоненты: центры обработки мобильных сообщений, данные по обслуживанию клиентов, данные из отдела продаж, данные о продажах через интернет-магазины, данные ERP систем, данные из ИСП (EIS) и других внешних источников.

Эти системы выступают источниками данных для аналитических CRJV1. Типовая структура аналитического ХД CRM-системы приведена на рис. 27.

Внедрение такого решения позволяет оптимизировать цепочки работы с клиентами, провести персонализацию обслуживания клиентов, повысить доходы от продаж, а также позволяют разрабатывать стратегии расширения

рынка за счет привлечения клиентов на основе индивидуального подхода.

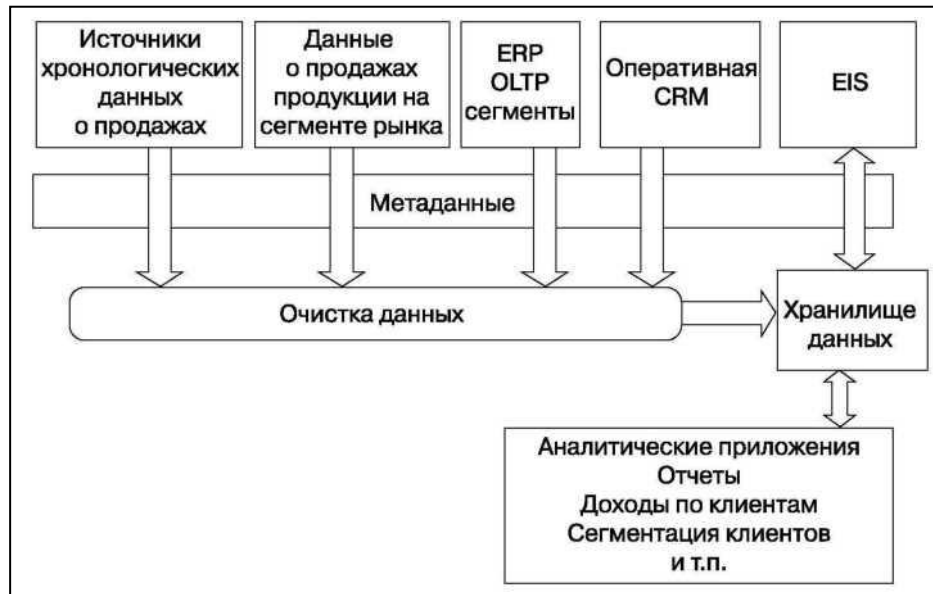


Рис. 27. Архитектура аналитической CRM-системы

Наиболее известное работающее решение в области аналитических CRM в телекоммуникациях имеет компания SAS Institute (US WEST Communications).

Аналитические SRM (SupplyRelationshipManagement) системы занимаются управлением взаимоотношениями с поставщиками. Пример архитектуры для ХД аналитических SRM систем приведен на рис. 28.

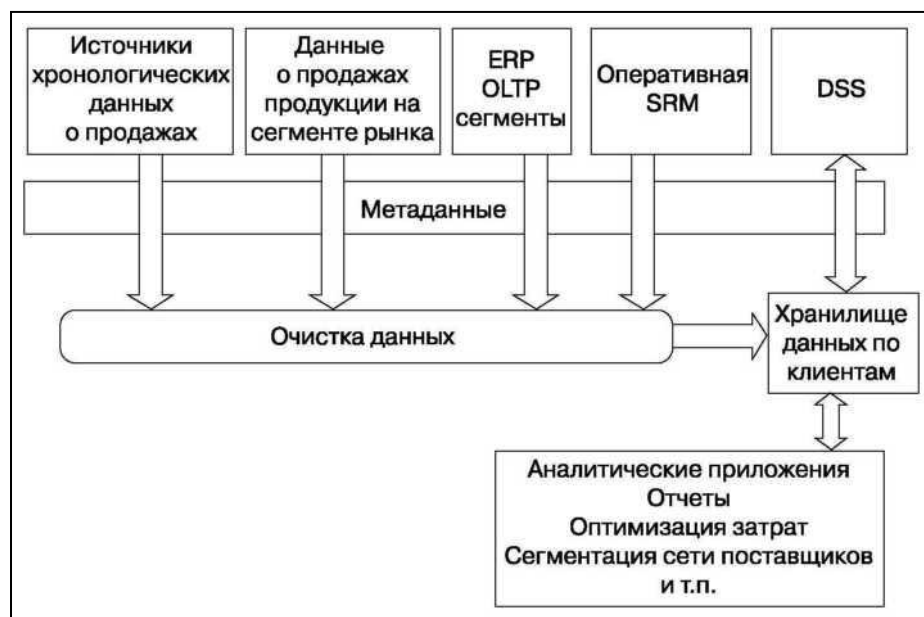


Рис.28. Архитектура аналитической SRM-системы

Наиболее известное решение в области создания аналитических SRM-

систем разработано компанией SASInstitute.

Аналитические SCM-системы, не встроенные в ERP-системы, представляют собой информационные системы для решения задач анализа и оптимизации в управлении жизненным циклом продукции. Пример типовой архитектуры для ХД аналитической SCM-системы приведен на рис. 29.



Рис. 29. Архитектура аналитической SCM-системы

По уровню использования SCM-решений телекоммуникации занимают второе место в мире (после нефти и газа).

Перечень наиболее удачных решений в области оперативных SCM-систем приведен в табл. 6.

Таблица 6. Решения в области оперативных SCM

Компания	Программные продукты
IBM	WebSphere (for e-Business), интеграция с ERP
SAP	BusinessInformationWarehouse, SAPAdvancedPlaner&OptimizerLogisticsExecutionSystem
BAAN	IBAAN с совокупностью модулей в архитектуре ПО BAAN, в том числе и использованием хранилища данных

Виртуальные предприятия. Одной из перспективных областей применения систем складирования данных является разработка ХД как составной части виртуального предприятия. В этом случае ХД рассматривается как часть интегрированной информационной структуры организации, которая имеет типовую архитектуру, показанную на рис. 30.



Рис. 30. Место хранилища данных в виртуальном предприятии

Мультимедийные хранилища данных. Перспективным становится разработка ХД для цифровых (электронных) библиотек и мультимедиа. Современные СУБД имеют ряд встроенных возможностей для хранения и выборки мультимедийных данных. Однако большинство решений по созданию мультимедийных хранилищ данных.

тимедийных баз данных реализуется на реляционных СУБД, обладающих возможностью работы с BLOB-данными и имеющими поддержку очень больших БД. Типичными представителями таких СУБД являются СУБД Oracle (имеет специальные средства выборки визуальной информации — VIR и интернет-систему обработки файлов iFS), DB2 и Informix (IBM).

Примерами мультимедийных ХД являются разрабатываемые во всем мире электронные хранилища музейных данных (образы картин и других экспонатов).

Обсудим особенности типового решения создания мультимедийных ХД. Следует отметить следующие свойства медиаданных:

- неструктурированная форма с точки зрения теории реляционных баз данных;
- размер элемента медиаданных очень большой;
- данные не имеют фиксированного максимального размера;
- внутренний формат для представления таких данных не может быть выражен простым типом данных реляционных СУБД;
- поиск данных затруднен или просто невозможен стандартными средствами СУБД.

В этом отношении хранилище мультимедийных данных имеет типовую архитектуру, в которой медиаданные быстро извлекаются и визуализируются.

Корпоративные информационные фабрики. В кругу бизнес-пользователей ИТ- технологий обсуждается предложенная Биллом Инмоном концепция так называемой корпоративной информационной фабрики (Corporate Information Factory, CIF), как одной из основополагающих вычислительных архитектур для производства информационных продуктов предприятия. Для любого предприятия реализацию такой концепции можно рассматривать как важную перспективную задачу, решение которой не только позволит повысить качество управления взаимоотношениями с внешними организациями (налоговыми и финансовыми государственными структурами) и партнерами, но и значительно увеличить производительность его подразделе-

ний, поставляющих информацию, необходимую для принятия стратегических решений.

Корпоративная информационная фабрика — это логическая архитектура программно-аппаратного решения по производству, складированию, управлению и доставке данных для поддержки принятия стратегических и тактических решений в масштабе организации. Концепция CIF подразумевала системно организованное взаимодействие репозитариев оперативных данных (OperationalDataStore), центрального ХД, витрин данных и системы интеллектуального анализа данных (DataMining) за счет создания технологических цепочек переработки и доставки данных.

В абстрактной форме процесс производства информации в CIF был представлен в аналогии с производством некоторого продукта. В соответствии с этим были выделены основные стадии производства информации (новых данных): получение исходных данных (сырья), их преобразование (производство отдельных деталей), складирование данных, создание информационных продуктов (из деталей готовой продукции) и доставка данных их потребителям (распределение конечной продукции).

Основная идея, положенная в основу концепции CIF, состоит в выделении элементов информационной архитектуры на основе их функционального назначения и регламентирования технологических процедур обработки данных.

Ключевым моментом правильно спроектированной CIF являются метаданные. Задача этого слоя — описать в рамках единой терминологической базы (метаданные бизнес-пользователя) всю совокупность объектов управления средой CIF (метаданные администрирования). Только подход позволяет из потока входной информации получить однородное описание среды и предметной области, что дает возможность обращаться к измерениям, кубам, отчетам и бизнес-объектам на основе произвольных выборок. Таким образом, обеспечивается высокое качество циркулирующей в CIF информации.

В основе CIF лежит модель функционального разделения процессов

производства новых данных (информационных продуктов) и доставки информационных продуктов их потребителям, а также управления этими процессами.

Производители информационного продукта собирают данные из доступных источников (чаще всего из оперативных систем ввода и обработки данных), преобразуют и интегрируют их, размещая в системе складирования данных в унифицированном регламентированном формате. Потребители информационных продуктов извлекают необходимые тематические выборки из системы складирования данных (через специализированные предварительно настроенные интерфейсы — витрины данных) и затем используют их в процессе принятия решений.

Логическая структура СІФ включает в себя несколько типовых архитектурных элементов (табл. 7).

Таблица 7. Типовые архитектурные элементы логической структуры СІФ

Элемент	Характеристика
Системы, доставшиеся «по наследству» (LegacySystems)	Поддерживают бизнес-функции, которые были созданы в организации ранее. В таких системах обычно компоненты, обеспечивающие формирование отчетов и ввод и передачу данных, реализуются в рамках единого программного блока, что затрудняет решение задач по интеграции и преобразованию данных в соответствии с новыми требованиями бизнеса
Приложения оперативного управления организацией (OLTP)	Обеспечивают быструю обработку данных в рамках бизнес-направлений деятельности организации. Как правило, такие системы приобретаются у компании-разработчика, которая осуществляет их техническую поддержку?
Оперативные склады данных (Operational Data Store — ODS)	Этот элемент наделяется свойствами как оперативных, так и аналитических систем. Основное его назначение — обеспечить осуществление анализа информации практически сразу после ее обновления в оперативных системах

Компоненты преобразования данных (ETL-tools, Staging Area, Near-line Storage)	Служат для перегрузки данных из одних программных компонентов в другие (с промежуточной очисткой и согласованием данных, получаемых из различных источников)
Корпоративное хранилище данных (Enterprise Data Warehouse)	Здесь накапливается детальная информация, необходимая для выполнения анализа. Данные перегружаются в корпоративное хранилище из оперативных элементов — унаследованных систем, автоматизированных банковских систем или оперативных складов данных. Как правило, обновление информации в EDW происходит с большой задержкой. Для разрешения этой проблемы используются ODS-элементы
Витрины данных (Data Marts)	Предназначены для хранения аналитической информации уровня подразделения или направления бизнеса
Приложения принятия решений (DSS) и приложения анализа данных (DM)	DSS, примером функционала которых могут быть системы анализа клиентской базы банка, обеспечивают поддержку принятия решений. Разнообразный статистический анализ выполняется в DM
Инфраструктура сетевых коммуникаций	Обеспечивает публикацию данных в сети Интранет/ Интернет), а также обработку результатов ввода информации пользователями.

На предприятии производственные и финансовые потоки тесно взаимосвязаны с потоками информационными, которые отражают их динамические показатели и текущее состояние. Кроме того, такие информационные потоки являются источником данных для анализа при определении трендов изменений и их количественных характеристик.

Описанная выше в общих чертах схема превращения данных в информационные продукты и составляет суть концепции CIF на любом предпри-

ятии (рис. 31):

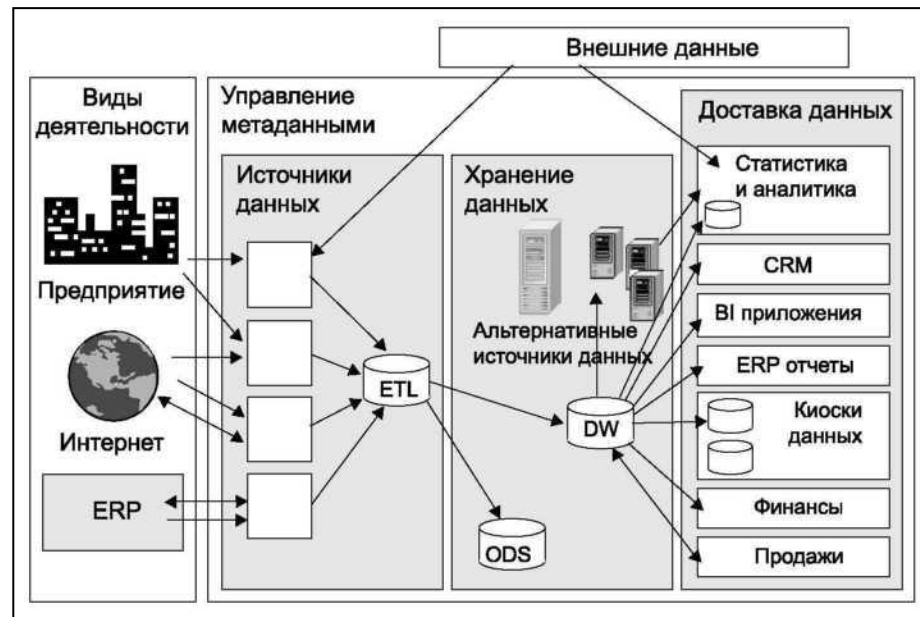


Рис.31. «Корпоративная информационная фабрика»

Складирование данных — это технология, с помощью которой можно оперативно собрать данные и на их основе решать разнообразные задачи по финансовому планированию, бюджетированию, риск-менеджменту, анализу взаимоотношений с партнерами, маркетинговому анализу и т.д.

Главное преимущество отлаженной архитектуры CIF заключается в адаптации вычислительной среды как под четко определенные информационные потоки небольшого предприятия, так и под сложные схемы консолидации, которые характерны для предприятий с развитой филиальной структурой и входящих в состав холдингов и отраслевых объединений предприятий.

Рассмотрим, как CIF функционирует на предприятии.

Первоначальное наполнение корпоративного ХД и постоянное поддержание его в актуальном состоянии — это отнюдь не тривиальные задачи. Особые требования здесь предъявляются к качеству информации, кроме того, высока степень риска — ошибочные решения на основе неверных исходных посылок могут обернуться серьезными потерями.

На предприятиях основными источниками данных являются ERP-системы. Они представляют собой семейство оперативных приложений, обеспечивающих обработку производственных и финансовых данных, включая выполнение бухгалтерских проводок, логистических операций, генерацию текущей оперативной отчетности. Модули ERP ориентированы на информационные продукты, которые они сопровождают или поддерживают. Разумеется, ERP не предназначены для обработки информации в историческом аспекте и не имеют развитого инструментария для агрегации и систематизации данных предприятия. Из-за строгой предметной направленности у подсистем ERP, как правило, слабо развиты взаимосвязи на уровне данных: обычно у них информационный обмен осуществляется небольшими объемами.

Таким образом, на первом шаге построения CIF-системы источники данных накапливают информацию в масштабе предприятия в «сыром» виде: она не подготовлена для анализа и компиляции аналитической отчетности.

Организация процесса интеграции является еще одним фактором успеха в создании CIF: информация извлекается из разнородной вычислительной среды ERP, преобразуется с целью повышения ее качества и складировается. Все это делается для того, чтобы системы поддержки и принятия решений могли в дальнейшем ее активно использовать.

Для наполнения корпоративного ХД в нем обычно предусматриваются инструментальные средства:

- для извлечения и доставки из различных оперативных БД и внешних источников;
- для очистки, преобразования и интеграции;
- для загрузки;
- для актуализации.

ХД — это предметно-ориентированная, интегрированная, неизменяемая и поддерживающая хронологию коллекция данных, используемая для поддержки принятия решений. С позиций CIF хранилище является отправной точкой при преобразовании данных в информационные продукты (аналитиче-

ские отчеты и пр.). Оно всегда предоставляет своим потребителям проверенные и согласованные данные по всей организации в целом, независимо от источника их происхождения.

Процесс управления данными предусматривает комплекс процедур, отвечающих за прохождение информации в СІF. Он включает в себя архивацию и восстановление данных, секционирование, управление перемещением данных в системе, агрегацию и т.д.

В конечном информация должна попасть к потребителю в заданном виде для принятия взвешенных управленческих решений. Логично на выходе СІF применять:

- средства для многомерного представления данных и манипулирования ими;
- средства для формирования отчетов;
- систему информационных запросов.

Хранилища данных с архитектурой шины данных. В данной архитектуре ХД с архитектурой шины данных, предложенной Ральфом Кимболлом, первичные данные преобразуются в необходимые структуры на стадии подготовки данных. При этом обязательно принимаются во внимание требования к скорости обработки информации и качеству данных. ХД с архитектурой шины данных изначально ориентированы на использование многомерной модели данных.

В отличие от корпоративной информационной фабрики, в ХД с архитектурой шины данных чаще используются связанные киоски данных, которые разрабатываются для обслуживания бизнес-процессов (бизнес - показателей или бизнес-событий), а не направлений бизнеса. Например, данные о заказах, которые должны быть доступны для общекорпоративного использования, вносятся в ХД только один раз, в отличие от СІF, в котором их пришлось бы трижды копировать в витрины данных отделов маркетинга, продаж и финансов. После того, как в ХД появляется информация об основных бизнес-процессах, консолидированные киоски данных могут выдавать их перекрест-

ные характеристики.

Матрица шины данных корпоративного ХД с архитектурой шины выявляет и усиливает связи между показателями бизнес-процессов (фактами) и описательными атрибутами (измерениями). ХД с архитектурой шины данных состоит из набора взаимосвязанных киосков данных, которые созданы для обслуживания бизнес-процессов организации (см. рис. 32.).

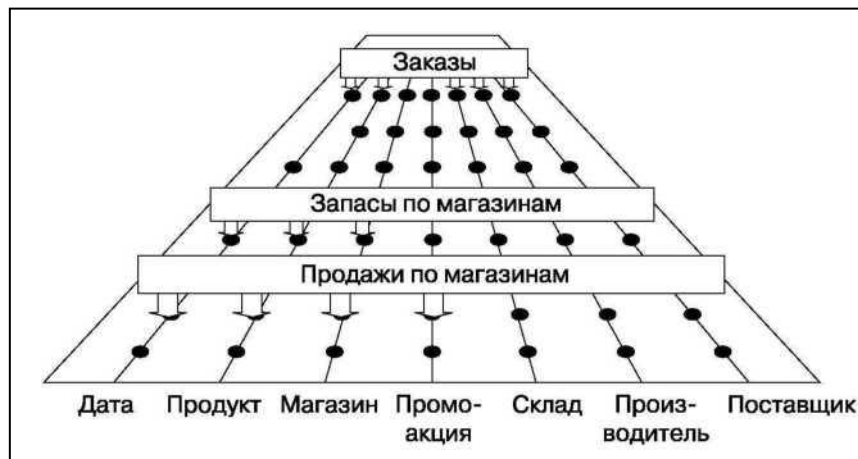


Рис. 32. Хранилище данных с архитектурой шины данных

Суммируя все вышесказанное, можно отметить типичные характеристики ХД с архитектурой шины данных.

1. Использование многомерной модели организации данных с архитектурой «звезда» (starscheme).
2. Использование двухуровневой архитектуры, которая включает стадию подготовки данных, недоступную для конечных пользователей, и собственно ХД с архитектурой шины. В состав последнего входят несколько киосков атомарных данных, несколько киосков агрегированных данных и персональный киоск данных, но оно не содержит одного физически целостного или централизованного ХД.
3. ХД не является единым физическим репозитарием (в отличие от СІF). Это «виртуальное» ХД, представляющее коллекцию витрин данных, каждая из которых имеет архитектуру типа «звезда».

Отметим, что и корпоративная информационная фабрика, и ХД с архи-

тектурой шины данных имеют своей целью создание корпоративного ХД.

Обе эти архитектуры отличаются в основном способами представления данных.

Отметим, что в последнее время возрастает практический интерес к использованию ХД при формировании информационной инфраструктуры организаций.

Преимущества, которые получает организация от внедрения хранилищ данных.

- *Взгляд на данные организации, как на единое целое.* Это ответы на такие вопросы: сколько продуктов реально производится? Что влияет на изменение спроса? Какие товары или услуги приносят наибольший доход? А также возможность учитывать особенности и предпочтения клиентов.
- *Возврат инвестиций, вложенный в создание хранилища данных.*
- *Возрастает надежность данных для принятия решений.* Данные, загружаемые в хранилище данных, подвергаются очистке — согласуются, проверяются, уточняются.
- *Геопространственный анализ данных.* Анализ такой информации имеет решающее значение в принятии решений по всем вопросам, связанным с географией бизнеса.
- *Исследование трендов и колебаний в бизнес - данных.* Позволяет достаточно надежно прогнозировать развитие бизнес-процессов организации во времени.

Тема №12. «Реализация СППР в банковской сфере»

Системой поддержки принятия решения назовем совокупность объектов и субъектов, которые позволяют лицам, принимающим решения; организовывать процесс поиска оптимального управленческого решения.

В данном случае в качестве объектов выступают компьютерные системы с установленным специализированным программным обеспечением, системы сбора, хранения, обработки и передачи информации, телекоммуникационные системы и т. д. (назовем это многообразие объектов аналитическим программным комплексом), а в качестве субъектов - консультанты, аналитики, эксперты, коллегиальные органы при руководстве, заместители, помощники и т. д.

Аналитический программный комплекс – универсальное инструментальное средство сбора, консолидации, обработки и анализа больших объемов информации, способное на основе имеющихся данных, например, о клиентах банка, филиалах, совершенных банковских операциях, последовательности внешних событий и т.п.:

- анализировать текущее положение, как головного банка, так и его филиалов, а также клиентов банка;
- оценивать состояние банков-корреспондентов;
- прогнозировать поведение различных показателей;
- автоматически отслеживать происходящие и надвигающиеся критические события;
- анализировать взаимосвязь событий и процессов, происходящих одновременно, а также событий, смещенных во времени;
- на основе выявленных взаимосвязей прогнозировать поведение одних показателей в зависимости от значений других.

АПК состоит из 4 подсистем: подсистема ввода данных (Input), подсистема хранения аналитических данных (Data Warehouse), подсистема анализа и отчетов (Analysis and Reporting), подсистема администрирования (Supervisor).

Основное программное обеспечение:

- Центральное хранилище аналитических данных – SybaseAdaptiveServer IQ;

- Подсистема анализа и отчетности – BusinessObjects и WebIntelligence.

В число функций подсистем АПК входят импорт и структурирование данных, работа с хранилищем данных, ведение системы показателей и справочной системы, графическое представление информации, анализ данных, прогнозирование временных рядов и т.д.

Приведем краткое описание функций основных модулей Business Objects, используемых в СППР:

- **Designer** - рабочее место разработчика моделей данных.
- **InfoView** - место конечного пользователя, позволяющее просматривать и обновлять подготовленные отчеты.
- **Reporter** - встраиваемая в InfoView опция, дающая возможность пользователю формировать запрос к базе в терминах своей предметной области.
- **Explorer** - встраиваемая опция, необходимая для пользователей, которые строят отчеты и занимаются анализом данных. InfoView + Reporter + Explorer могут быть основой для разработки приложений OLAP.
- **BusinessQuery** - работа с семантической моделью ВО из Excel.
- **BusinessMiner** - рабочее место аналитика данных и пользователей, которым необходимо быстро принять решение с учетом накопленной на предприятии статистики (навигация по дереву решений - метод "что-если").
- **Supervisor** - рабочее место администратора системы.
- **BroadcastAgent** - сервер документов, позволяющий организовывать выполнение заданий по расписанию и по событиям в источниках информации.
- **WebIntelligence** – обеспечивает доступ к системе BusinessObjects из Internet/Intranet.

Рассмотрим технологические принципы работы системы.

1) В основу работы системы положена технология хранилищ данных (datawarehouse) – информация собирается в единый банк данных, системати-

зируется и приводится к единому формату. После этого она доступна для просмотра и анализа.

2) Данные на центральный сервер АПК можно вводить вручную или принимать из различных источников в разных форматах (OLTP АБС, Excel, Oracle, MS-SQL, текстовые файлы и т.п.). С процессом импорта данных связан также процесс их интерпретации, то есть привязка всех классифицированных значений к справочникам системы.

3) С помощью универсальных средств построения запросов центрального сервера данные при необходимости группируются и превращаются в метаданные.

4) В DESIGNER администратор объединяет и преобразует метаданные из центрального хранилища, а также из альтернативных баз данных в каталог бизнес терминов (universe). Задаёт исходные иерархии измерений и исходные форматы отображения объектов в отчетах.

5) Процессы анализа, прогнозирования и представления данных проводятся в BusinessMiner + BusinessObjects, или WebIntelligence. Доступен If-Else-анализ, анализ взаимосвязей (Data Mining), и т.д. Через Business Miner возможен анализ бизнес-данных в Microsoft Excel. Возможен дистанционный анализ и просмотр отчетов через сеть Internet посредством Web-browser без установки дополнительного оборудования (функции WebIntelligence). Возможен экспресс-анализ данных и построение отчетов в BusinessObjects.

Комплекс предусматривает использование программ внешнего анализа и прогнозирования, такие как Статистика, NCSS и т.п., которые цепляются к центральному хранилищу через ODBC. Широкие возможности графического интерфейса делает возможной перенастройку в режиме реального времени визуальных свойств объектов, присутствующих на экране – графиков, диаграмм, списков и т.д., а также оперативную настройку “горячих” клавиш. Все графики и диаграммы в системе снабжены средствами “погружения в данные”, которые позволяют немедленно выяснить, каким образом были получены те или иные результаты, вплоть до доступа к первичной информации.

Домен документов является местом централизованного хранения и дистрибуции разработанных отчетов между пользователями DSS. Через этот домен пользователь может отправить документ на обработку заданий (обновление отчета, печать, экспорт отчета на Web, рассылка по списку и выполнение скриптов) по расписанию или наступлению predeterminedенного события с помощью BROADCAST SERVER.

В АПК реализована удобная справочная система. Возможна настройка отображения каждого существующего справочника, а также создание новых справочников. Предусмотрены широкие функции администрирования. При работе с системой производится регистрация пользователей, предусмотрена защита от несанкционированного доступа. Кроме того, возможно задание ограничений прав конкретного пользователя.

Рассмотрим функционал отдельных подсистем. Подсистема ввода данных (Input) состоит из источников информации и средств упорядочивания входных данных согласно установленному стандарту.

В качестве исходных данных АПК использует информацию из учетных банковских систем, из витрин данных и всевозможную внешнюю информацию (результаты биржевых торгов, публичные отчеты других банков и т.п.).

Входные данные 1. АПК закачивает данные непосредственно из хранилища, если существует возможность присоединения к базам, или через промежуточные таблицы. Частота обновления информации зависит от необходимости иметь актуальные данные

Входные данные 2. Витрины данных – базы данных, которые используются для локальных задач, которые требуют меньших объемов информации, большой мобильности данных и специфических приемов обработки информации.

Входные данные 3. Платная информация таких агентств как Reuters, или свободная информация в сети Internet являются неотъемлемой частью входных данных аналитической системы.

Входные данные 4. В АПК предусмотрен ввод данных вручную. Эксперты вводят оценочные коэффициенты, администраторы - описания аргументов.

Данные для хранения и анализа могут загружаться в АПК из таблиц различных форматов (Dbase, Excel, Oracle, MS-SQL, InterBase и т.п.) или вводиться вручную. При этом они могут быть преобразованы в один из поддерживаемых хранилищем видов данных (одномерные, многомерные или неструктурированные данные). Данные могут рассчитываться в ходе первичной обработки.

Настройка системы в соответствии с тем или иным видом источника осуществляется в зависимости от решаемой задачи. При этом используются специальные шаблоны загрузки. Загрузка данных может осуществляться системой автоматически в соответствии с заданным предварительно расписанием или по событию.

Поступившие в систему данные можно анализировать, рассчитывать по ним коэффициенты, проверять, удовлетворяют ли эти коэффициенты заданным условиям (т.е. осуществлять мониторинг) и представлять их в виде графика (по отдельности или группами).

Подсистема хранения аналитических данных (Data Warehouse) состоит из Центрального хранилища, витрин данных, модуля создания метаданных, инструмент описания семантического слоя.

Вся поступающая в АПК информация содержится в удобном для системы виде в Центральном хранилище данных.

К функциям хранилища данных АПК можно отнести:

- преобразование разнородных данных из различных источников к единому формату и ввод данных,
- проверка логической корректности и полноты вводимой информации,
- контроль за выполнением регламента поступления и полноты вводимой информации,

- хранение текущих и исторических данных,
- обеспечение удобного доступа к хранимой информации.

При этом поддерживаются три вида данных:

- **одномерная информация** – измерения и их атрибуты, например, атрибуты валют, балансовых счетов или филиалов;
- **многомерная информация** – примитивы в контекстах, например, данные об остатках и оборотах, разложенные по филиалам, балансовым счетам, валюте и времени;
- **неструктурированная информация** – документы, например, проводки и кредитные договора.

Развитая сеть запросов позволяет рассчитывать на основе информации из хранилища данных различные показатели - метаданные.

Возможна ситуация, когда необходимо, чтобы система автоматически (без участия пользователя) реагировала на различные события или напоминала пользователю об их наступлении. Для этого в центральном хранилище предусмотрена программная компонента, именуемая *системным агентом*. Системный агент позволяет, например, автоматически вести мониторинг, рассчитывать показатели, а также напоминать о необходимости ручного ввода информации.

Подсистема анализа и отчетов (AnalysisandReporting) имеет широкие возможности анализа, прогнозирования и представления данных предоставляет АПК. Благодаря семантическому слою, пользователи обращаются к данным на языке бизнес-терминов.

При построении различных отчетов АПК обладает широкими возможностями масштабирования, “разворачивания” формул (“drill-down”), построения производных графиков, а также специальными аналитическими функциями.

Все это возможно за счет обширного инструментария.

Первым инструментом является комплекс продуктов Reporter+Explorer+Miner. С их помощью осуществляется:

- **Прогноз.** На основании имеющихся данных о значениях показателей система может прогнозировать их поведение в будущем. Результат прогноза отображается на графике.
- **Анализ “If-Else”.** Используя полученный прогноз, пользователь может проводить анализ “If-Else”. В процессе его система находит закономерности, учитывающие взаимосвязь между графиками показателей.
- **Анализ взаимосвязей (Data Mining).** На основании значений показателей, рассчитанных для прошедших периодов времени, система может находить и исследовать зависимость поведения графиков показателей друг от друга не только в один и тот же момент времени, но и со сдвигом во времени.
- **Мониторинг.** Существует множество методик, позволяющих анализировать состояние дел в той или иной области и сигнализировать о неблагоприятной ситуации (или, наоборот, о возможности совершения выгодных операций).

С помощью инструментов мониторинга, реализованных в АПК, в системе автоматически ведутся регулярный расчет и проверка значений большого числа показателей. В случае нарушения заданных ограничений система информирует об этом банковского аналитика.

В состав АПК входит специальный набор программных средств и инструментов, которые позволяют анализировать данные о произвольном наборе объектов и формировать на основании их показателей различные рейтинги.

Рейтинг-анализ позволяет оценивать как текущее состояние совокупности объектов, так и их состояние в прошлом. При этом производится сравнение полученного результата с состоянием других аналогичных совокупностей или с заданными предварительно средними характерными значениями.

Реализован широкий спектр возможностей просмотра различных диаграмм и составления рейтинг-отчетов.

Вторым инструментом анализа и представления данных является WebIntelligence, который обладает мощными средствами построения html-отчетов через Web-browser. Пользователь может строить и просматривать от-

четы, задавая произвольные запросы к базам данных опять же в терминах своего бизнеса. WebIntelligence использует те же universe, что и другие средства Business Objects.

Благодаря возможностям WebIntelligence, аналитик может проанализировать ситуацию по отчету, который он сможет создать и вне месторасположения банка. Причем пользователь не ограничен в этом случае html-отчетами. Любой отчет можно экспортировать в Excel.

Третий инструмент аналитика в АПК – это BusinessQuery, посредством которого реализована возможность проводить анализ в MS Excel, пользуясь все теми же данными и терминами семантического слоя. О простоте и богатстве пакета анализа MS Excel говорить не приходится – комментарии излишни.

Четвертый инструмент – возможность анализа данных в отдельных аналитических пакетах, таких как Статистика, NCSS и т.п. В этом случае эти пакеты используют стандартные средства доступа к центральному хранилищу. Рассчитанные данные из этих пакетов могут пополнять центральную базу.

Пятый инструмент – экспресс-отчеты. Это возможность ВО строить отчеты минуя семантический слой. Это может быть полезно для быстрого выхватывания неструктурированной информации опытными пользователями, знающими структуру центрального хранилища.

Отдельно хочется выделить то, что аналитические отчеты хранятся в одном месте. По желанию можно загрузить любой отчет из Домена документов. Для удобства пользователей выделена система справочников.

Механизм работы со справочниками позволяет пользователю просматривать информацию по лицевым и балансовым счетам, филиалам, валютам, клиентам.

При просмотре пользователь может производить сортировку, поиск нужного элемента, переходить между справочниками по ссылкам и осуществлять фильтрацию (например, просматривать только клиентские лицевые счета,

только счета крупных клиентов либо только лицевые счета, открытые на данном балансовом счете).

Пользователю предоставляется возможность создавать нестандартные справочники, отбирая информацию для просмотра, а также настраивать перекрестные ссылки на другие справочники.

Изменения конфигурации справочника сохраняются в памяти системы для одного и того же пользователя до следующего сеанса работы с ним. Другой пользователь, соответственно, может настроить конфигурацию для себя.

Благодаря BroadcastAgent, в комплексе автоматически выполняются различные события по обмену, обновлению, удалению документов.

Инструменты администрирования разделены на 2 части – средства администрирования хранилища данных (Sybase) и средства (BO и Wi) по управлению системой анализа и отчетности, представленные в Supervisor. Первые позволяют обслуживать хранилище данных и настраивать физические формы доступа к объектам базы, вторые - регламентировать работу пользователей и продуктов BO и Wi.

Администратор отвечает за следующие действия в системе:

- определяет характеристики отдельных атрибутов объектов данных;
- создает структуры для физического хранения отдельных объектов данных;
- создает слой метаданных;
- создает семантический слой;
- задает правило логической непротиворечивости для данного объекта данных;
- определяет права отдельных пользователей по манипулированию отдельными атрибутами объектов данных;
- задает способы, источники и временной регламент поступления данных;
- ввод данных из источников данных;
- расчет агрегатов данных;

- импортирование рассчитанных агрегатов данных во внешние структуры хранения;
- сброс неактуальных данных в архив;
- определяет права отдельных пользователей по манипулированию отдельными программными средствами;
- описывает структуры общедоступных документов;
- обмен документами между пользователями системы;
- сброс неактуальных документов в архив.

Построение и внедрение систем поддержки принятия решений – это качественный скачок в развитии организации. Однако для совершения подобного скачка необходимо вложить некоторую критическую массу денег и времени, после чего возможна отдача. Барьером для внедрения многих готовых комплексов становится стоимость. Не готовы еще руководители выкладывать кругленькую сумму денег за расплывчатые возможности и обещания. Ведь заказчику предоставляется в основном инструментарий, включающий средства разработки и набор готовых методов. Часто для настройки под конкретную систему ведения бизнеса нужно потратить столько же усилий, сколько бы хватило на построение своей собственной системы, удовлетворяющей современному развитию технологий. Опыт показывает, что срок введения готовой системы составляет от шести месяцев до одного года.

Разрабатывать систему лучше самому, используя передовые разработки в области программного обеспечения – это наиболее дешевый вариант системы. Причем принципиальным будет необходимость сразу формировать информационное поле, ориентированное на аналитическую обработку – как по структуре, так и по содержанию. Развитие системы будет представлять введение новых модулей, отвечающие за отдельные задачи и неразрывно связанные с другими компонентами системы. В этом случае срок внедрения системы явно увеличится в два раза, однако система будет включать большее разнообразие в средствах аналитической обработки.

Разработка и внедрение системы поддержки принятия решений, подобной описанной выше проходит согласно следующему плану.

Разработка

- 1. Определить потребности и описать основных пользователей системы.
- 2. Определить желаемые результаты функционирования системы.
- 3. Определиться с источниками данных.
- 4. Разработать информационную модель системы.
- 5. Выбрать способ хранения данных и тип хранилища.
- 6. Выбрать систему визуализации и анализа данных.
- 7. Разработать технологическую модель системы.
- 8. Подготовить график внедрения системы, ответственных исполнителей.

Внедрение и настройка системы

- Покупка и установка программного обеспечения.
- Разграничение пользовательских прав доступа к системе.
- Подготовка структуры базы для заполнения данными, и установление взаимосвязей.

- Наполнение хранилища данными.
- Пробная реализация одной из подзадач СППР.
- Демонстрация и оценка реализованных возможностей системы.
- Корректировка дальнейших планов по развитию системы.
- Реализация оставшихся подзадач.

Сопровождение и дальнейшее развитие системы

- 1. Текущее администрирование системы
- 2. Обучение пользователей
- 3. Дальнейшая работа по модернизации системы.

Примерная архитектура СППР в банковской сфере приведена на Рис.33.

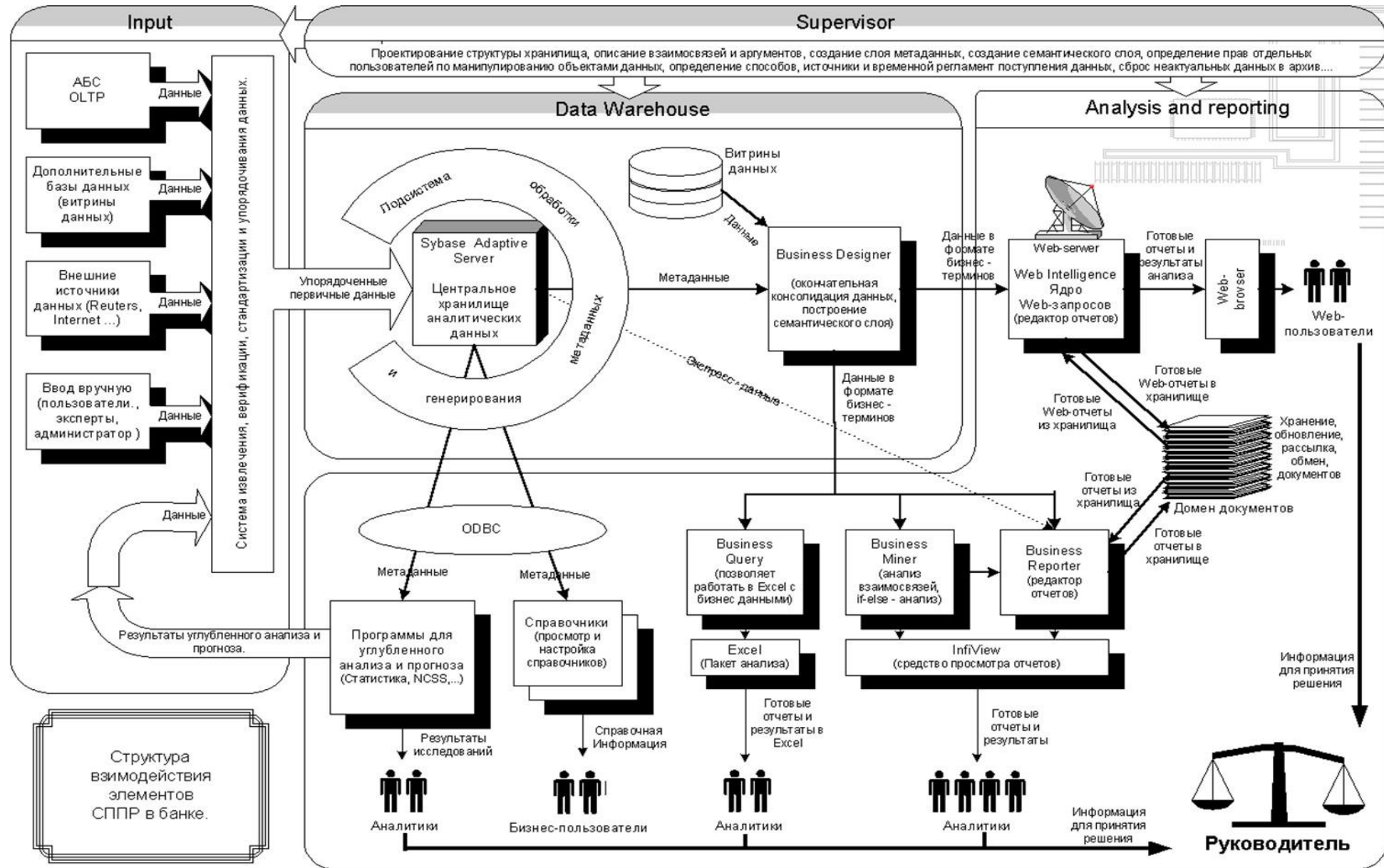


Рис. 33. Архитектура СППР для банковской сферы

Тема №13. Интеллектуальный анализ данных – технология Data Mining

- Задачи Data Mining.
- Специфические методы Data Mining-a
- Области применения методов Data Mining
- Средства реализации и задачи проектирования комплекса Data Mining

Задачи Data Mining

Следует различать два различных процесса получения знаний. Первый – это «извлечение» их из живого источника – эксперта, специалиста с целью их идентификации и возможной формализации, помещения в базу знаний и построения на этой основе экспертных систем, а также в других целях. Такой процесс относят к инженерии знаний. Другой – это «добыча» скрытых от пользователя знаний из данных, помещенных в различного рода компьютерные информационные системы, в том числе базы данных различного назначения, информационные хранилища. Процесс второго рода называют Data Mining – используют русский перевод «интеллектуальный анализ».

Предметом нашего изучения является Data Mining.

Для обработки накопленных в различных источниках и местах сбора и хранения данных и выполнения интеллектуального анализа используются все достижения математической науки и информационных технологий. В первую очередь используются методы линейной алгебры, классического математического анализа, дискретной математики, многомерного статистического анализа.

В экономической предметной области применение методов поиска решений, условий неотрицательности и других свойств математических моделей путем дедуктивного получения следствий, исходя из предварительно

сформулированных предпосылок, относится к разделу экономической науки, называемому математическая экономика.

Анализ количественных закономерностей и взаимозависимостей в экономике, который выполняется статистическими методами, относится к эконометрике.

Традиционная математическая статистика долгое время являлась основной методологией анализа данных в экономической и других предметных областях. Однако базовая концепция усреднения по выборке часто приводит к операциям над фиктивными величинами. В экономике средние значения ряда показателей по различным предприятиям иногда создают искаженное представление об отсталости или, наоборот, о незаурядных успехах ряда предприятий, отраслей или регионов – сглаживают их.

По этой причине появился ряд методик, которые относят к специфическим для Data Mining-а. Эти методики позволяют избежать таких ситуаций. В таблице приведены примеры постановок задач для методик, основанных на математической статистике и специфических методах, см. Табл. 8.:

Таблица №8. Специфика технологий =DataMiningиOLAP

OLAP	Data Mining
Каковы средние показатели рентабельности предприятий в регионе?	Какова характерная совокупность значений показателей финансово-хозяйственной деятельности предприятий в регионе?
Каковы средние размеры счетов клиентов банка – физических лиц?	Каков типичный портрет клиента – физического лица, отказывающегося от услуг банка?
Какова средняя величина ежедневных покупок по украденной или фальшивой кредитной карточке?	Существуют ли стереотипные схемы покупок для случаев мошенничества с кредитными карточками?

Выше показано, что работа по интеллектуальной обработке данных происходит в сфере закономерностей.

Основными задачами интеллектуального анализа являются:

- выявление взаимозависимостей, причинно-следственных связей, ассоциаций и аналогий, определение значений факторов времени, локализация событий или явлений по месту;
- классификация событий и ситуаций, определение профилей различных факторов;
- прогнозирование хода процессов, событий.

Главной задачей здесь является выявление закономерностей в исследуемых процессах, взаимосвязей и взаимовлияния различных факторов, поиск крупных «непривычных» отклонений, прогноз хода различных процессов в области мягких и глубинных знаний.

Одновременно с этим многомерный статистический анализ твердо удерживает свои позиции в жесткой области знаний. Он делится на: факторный, дисперсионный, регрессионный, корреляционный, кластерный анализ (является также сферой интересов Data Mining-a).

Эти методы позволяют решать многочисленные задачи в области экономики, менеджмента, юриспруденции, которые являются составной частью аналитической подготовки принятия решений.

Специфические методы и области применения Data Mining-a

Помимо перечисленных выше методов многомерного статистического анализа, ставших традиционными, все более широкое применение находят специфические методы интеллектуального анализа, происходящие из смежных областей информационных технологий (IT-систем) и достижений различных областей науки.

К специфическим методам интеллектуального анализа относятся:

- методы нечеткой логики;
- системы рассуждений на основе аналогичных случаев;
- классификационные и регрессионные деревья решений;
- нейронные сети;
- генетические алгоритмы;

- байесовское обучение (ассоциации);
- кластеризация и классификация;
- эволюционное программирование;
- алгоритмы ограниченного перебора.

Методы нечеткой логики используются для описания плохо формализуемых объектов из состава «мягких» знаний. Над ними также совершаются мягкие вычисления. Используется понятие «лингвистическая переменная», значения которой определяются через нечеткие множества, а они представляются базовым набором значений или базовой числовой шкалой.

Системы рассуждений на основе аналогичных случаев CaseBasedReasoning (CBR) основаны на том, что принятие решения осуществляется по прецеденту, наиболее подходящему к данной ситуации с учетом определенных корректив. Иногда решение принимается на основе учета всех примеров, находящихся в хранилище данных.

Деревья решений основаны на иерархической древовидной структуре классифицирующих правил. Решения об отнесении того или иного объекта или ситуации к соответствующему классу принимаются по ответам на вопросы, стоящие в узлах дерева. Положительный ответ означает переход к правому узлу следующего уровня, отрицательный – к левому узлу. Процесс разделения продолжается до полного ответа на все поставленные вопросы.

Нейронные сети – это упрощенная аналогия нервной системы живого организма. Разработаны модели нейронных сетей. Распространенной моделью является многослойный персептрон с обратным распространением ошибки. Нейроны работают в составе иерархической сети, в которой нейроны нижележащего слоя своими выходами соединены с входами нейронов вышележащего слоя. На нейроны нижнего слоя подаются значения входных параметров, которые являются сигналами, которые передаются в следующий слой. При этом они ослабляются или усиливаются в зависимости от числовых значений, которые придаются межнейронным связям, называемых весами. На выходе нейрона верхнего слоя вырабатывается сигнал, являющийся

ответом сети на введенные значения входных параметров. Для получения необходимых значений весов сеть необходимо «тренировать» на примерах с известными значениями входных параметров и правильных ответов на них. Подбираются такие веса, которые обеспечивают наибольшую близость ответов нейронной сети к правильным.

Генетические алгоритмы представляют собой поисковый метод, используемый для нахождения наилучшего решения или совокупности решений. Он основан на идее естественного отбора. Начинается построение генетических алгоритмов с кодировки исходных логических закономерностей, называемых как и в биологии хромосомами. Набор таких кодов называют популяцией хромосом. Далее применяется функция пригодности, которая выделяет наиболее подходящие элементы для дальнейших операций. Это может быть отбор в какие-либо группы, но возможен и вариант применения скрещивания и мутации с целью получения «нового» поколения. Алгоритм работает над изменением старой популяции до тех пор, пока новая не будет отвечать заданным требованиям.

Байесовское обучение или ассоциации применяются в тех случаях, когда сложилась ситуация увязки между собой некоторых событий. Например, заселение новостроек сопровождается приобретением мебели и других предметов домашнего обихода. Необходимо выявить количественные характеристики этой связи.

Кластеризация и классификация. Слово кластеризация происходит от английского cluster – пучок, сгусток. Кластеризация предусматривает разделение совокупности схожих объектов на группы – кластеры по наибольшей близости их признаков. Проблема состоит в том, что оценка производится не по одному какому либо признаку, а одновременно по их совокупности. Разработаны алгоритмы кластеризации, которые пересчитывают значения признаков в некоторую величину, характеризующую «расстояние» между объектами рассматриваемой совокупности и объединяют близкие объекты в кластеры. Классификация отличается тем, что выявляются признаки, объеди-

няющие объекты, которые уже состоят в группах. Этими методами занимается также и эконометрика.

Эволюционное программирование. В этой методике предположения о виде аппроксимирующей функции строятся в виде программ на внутреннем языке программирования. Процесс построения программ выглядит как эволюция в среде программ. После нахождения в этой среде подходящей программы система начинает вносить в нее необходимые корректировки. Эта методика реализована российской системой Polyanalyst. Специальный модуль этой системы переводит найденные зависимости на доступный язык формул, таблиц.

Алгоритмы ограниченного перебора. Они вычисляют частоты комбинаций простых логических событий в группах данных. На основании оценки полученных частот делается заключение о полезности комбинаций для обнаружения ассоциаций в данных, прогнозирования и других целей.

Эти методы стали весьма широко и эффективно применяться в связи с бурным развитием в последнее десятилетие XX века самих методик и соответствующих инструментальных средств. Они находят применение в тех ситуациях, когда обычные методы анализа трудно или невозможно применить из-за отсутствия сведений о характере или закономерностях исследуемых процессов, взаимозависимостях явлений, фактов, поведении объектов и систем из различных предметных областей, в том числе в социальной и экономической.

Области применения методов Data Mining

С помощью этих методов при отсутствии априорной информации об объектах и их поведении и значительной ее неполноте решаются следующие задачи:

- выделение в данных групп, сходных по некоторым признакам записей;
- нахождение и аппроксимация зависимостей, связывающих анализируемые параметры или события;

- поиск наиболее значимых параметров в данной проблеме (задаче);
- выявление данных, характеризующих значительные или существенные отклонения от сложившихся ранее закономерностей (анализ отклонений);
- прогнозирование развития объектов, систем, процессов на основе хранящейся ретроспективной информации или с использованием принципов обучения на известных примерах и другие задачи.

Решение перечисленных задач может осуществляться каким-либо из перечисленных выше методов или комплексно для получения наиболее адекватного решения.

Средствами ИАС обеспечивается также оценка полученных результатов анализа и моделирования, в том числе оценка точности и устойчивости результатов, верификация моделей на тестовых наборах данных.

Тема №13. «Адаптация корпоративного портала для принятий решений (на примере 1С:Битрикс24)»

Битрикс24 – это корпоративный портал, который пытается охватить практически все, поскольку в нем заявлены функциональные возможности социальных сетей, проектов, задач, управления персоналом, и многое другое, в том числе CRM.

CRM Битрикс24 является частью общего программного продукта, поэтому время от времени будем касаться некоторых других частей общей системы Битрикс24, которые не относятся к CRM, но необходимы для понимания ее возможностей и особенностей.

Тарифы – это первое, с чем сталкивается каждый пользователь при выборе CRM системы. И первое, с чем нужно определиться, какой из вариантов Битрикс24 подойдет:

- работа «в облаке»;
- коробочное решение.

Работа «в облаке» — это покупка SaaS-решения, при этом вы оплачиваете доступ к системе, а вся работа ведется на серверах компании Битрикс24.

Коробочное решение – это вариант покупки программного продукта Stand-Alone, т.е. покупается программное обеспечение для установки на собственный сервер.

Для начала важно понимать, что CRM устанавливается как один из компонентов общей корпоративной системы Битрикс24. А потому, когда мы заходим в рабочую область пользователя (в «облаках» или на своем сервере после установки), то сразу попадаем в CRM-окружение, а видим общие инструменты портала.

Окружение, которое отображается на стартовой странице, состоит из таких компонентов:

- Мой диск
- Сообщения

- Календарь
- Задачи
- Лента и т.д.

Таким образом, после входа в систему пользователь видит перечень функций, которые напрямую не связаны с CRM. Да, они также нужны при определенных условиях, но для пользователя CRM не являются основными инструментами.

В таком подходе есть свои плюсы и минусы.

Плюсы:

- Вы получаете намного больше, чем планировали.

В принципе, на это и направлены все маркетинговые усилия Битрикс24. Пользователь понимает, что получил намного больше разнообразных инструментов, чем при покупке отдельно CRM.

Минусы:

- Наличие ненужных в работе инструментов усложняют навигацию и работу пользователя в системе.
- Дополнительные функции требуют определенных ресурсов, но при этом они не используются.

Битрикс24 предлагает программное обеспечение, интерфейс, возможности которого направлены не только и не столько на продажи, сколько на другие функции, которые нужны и очень полезны.

Рассмотрим Битрикс24 в целом. Если большинство CRM-систем являются небольшими и простыми в настройке SAAS-решениями, то Битрикс24 – это целый портал. И как для пользователя, так и для администратора, здесь имеется большое число функций. Например, в Битрикс24 администратор получает доступ и настраивает такие опции, как документооборот, фотогалерея, проактивная защита и т.д. В результате процесс внедрения заметно усложняется.

С одной стороны, широкие возможности – это мощно, интересно. С другой – это сложно и для пользователя, и для администратора.

Рассмотрим структуру CRM Битрикс24.

Последовательность действий в продажах обычно выглядит следующим образом:

- Регистрируется лид, т.е. в систему приходит запрос о желании совершить покупку. Этот запрос, а также контактные данные потенциального клиента регистрируются, и менеджеру отправляется уведомление (в системе и/или по почте).
- Формируется потенциальная сделка, на основе которой можно сформировать счет, можно отправить коммерческое предложение.
- После получения согласия клиента в момент продажи или заключения договора, на основе лида создается контакт.
- Для этого контакта оформляется продажа, проводятся отгрузка и оплата.
- Сделка закрывается.

Каким образом это реализовано в CRM Битрикс?

Здесь есть, так называемая, рабочая область CRM.

И в ней присутствуют следующие возможности:

- Лиды. В данном случае лид – это потенциальный клиент, т.е. в лиде отражаются контакты человека, который еще не является вашим покупателем, но уже проявил определенный интерес к вашей компании и оставил вам свои контактные данные. Лиды собирают с лендингпейдж, с рекламных объявлений, из интернет-магазинов, по телефону.
- Контакты и Компании. Сформированный лид, который был принят в работу, можно автоматически конвертировать в контакт и/или в Компанию.
- Потенциальная сделка. Также создается в системе на основе лида или вручную с использованием конкретного контакта или компании.

Создать любую из этих форм – лида, контакта, компании – не сложно. Но здесь также есть свои особенности – это дополнительные поля, которые можно создать для фиксации адреса, дополнительного телефона, любых других данных.

При создании дополнительного поля кроме типа данных (строка, число, дата и пр.), нужно выбрать такие качества, которые не имеют никакого отношения к CRM:

- Привязка к разделам информационных блоков.
- Привязка к элементам информационных блоков.

Что это такое, пользователи обычно не понимают, и для работы в CRM эти привязки, на самом деле, не нужны. Но выбрать эти параметры необходимо, так как мы пользуемся не просто CRM-системой, а большим порталом Битрикс24 .

Важно помнить: все, что касается Битрикс24 , но при этом не требуется для CRM Битрикс, остается видимым. Т.е. интерфейс относится не просто к CRM, но к большому portalу, а потому он всегда перегружен.

Сама по себе форма создания «Лида», а также последующая его конвертация проста и понятна интуитивно, но место при этом используется не экономно, а интерфейс выглядит несбалансированным.

Внешне формы Лида, Контакта и Компании – очень просты и понятны интуитивно. Более того, они выглядят идентичными. Разница – в названии и в сути:

- Лид – это контакты человека (организации), который проявил интерес к вашей продукции.
- Контакт – это человек, с которым уже работает отдел продаж. Ведет переговоры, решает какие-то организационные вопросы и т.д.
- Компания – это, по сути, тот же контакт, но на уровне организации (юридического лица). Т.е. контакты – это представители компании. И если контакт – это всегда человек, то компания – это организация, которую этот контакт (или даже несколько контактов) представляет.

Таким образом, переговоры, общение, обсуждение всех вопросов будет проводиться с тем или иным контактом, а заключение договора, выставление счета, оформление продажи и отгрузки — проводится с компанией. Эту разницу нужно знать и понимать.

Потенциальная сделка: особенности оформления. Сделка в этой CRM системе удобна и понятна пользователям. При ее создании можно помимо каких-то произвольных полей, добавить прямо в этот документ еще и товары. Такая возможность предусмотрена далеко не в каждой CRM системе.

В Битрикс24 можно в процессе общения с компанией создать Сделку и зафиксировать в ней, какие товары были предложены и по какой цене. Важно не путать Сделку и Счет. Счета выписываются отдельно, причем, их можно сформировать на основе Сделки, а не только создать вручную. В сделке просто фиксируются все договоренности предварительно: товары, объемы, цены.

Предложение. Помимо сделок, счетов, лидов, контактов и компаний, в CRM Битрикс есть еще один интересный документ – это Предложение. Этот документ является некой формой коммерческого предложения, где также можно перечислить все товары и цены. В принципе, предложение во многом дублирует Счета и Сделки. Есть некоторые отличия (преимущественно в «шапке» и печатной форме документа), но в остальном этот документ выглядит точно так же, как Сделка и Счет.

Отчеты. Отчеты – это важная часть в любой CRM-системе. Успешные продажи во многом зависят от своевременного контроля работы менеджеров, которая обычно осуществляется на основе отчетов.

В CRM Битрикс отчеты, с моей точки зрения, оказались одним из самых слабых мест. Перечень готовых отчетов очень мал, большинство из них – малоинформативные. Также существует конструктор отчетов, который предназначен для создания собственных вариантов отчетности, но он также очень ограничен в плане возможностей. Из-за ограниченного перечня настроек новые отчеты будут во многом дублировать стандартные, а возможность внедрения серьезной аналитики в CRM Битрикс не предусмотрена.

Воронка продаж – это отчет, который даже в меню системы вынесен отдельно, так как он наиболее востребован в текущей работе любого отдела продаж.

Воронка продаж – это некая диаграмма, которая демонстрирует динамику и процент сделок по статусу: от тех, которые находятся в финальной стадии (продажа) до тех, которые еще находятся в самой первой начальной стадии (получен лид, первые звонки).

Оформлен отчет также аккуратно, информативно, хотя гибкости настройки и этому отчету также недостает, настойки и здесь реализованы слабо. Впрочем, этот отчет стандартен, и что-то особенное в нем придумать сложно.

Товарная база. В CRM Битрикс имеется собственная база товаров и услуг, которые впоследствии могут использоваться в Сделках, Счетах и Предложениях. Реализация товарной базы удобная, во многом похожа на реализацию Лидов, Контактных и Клиентов, т.е. также есть обязательные и настраиваемые поля, что удобно для работы с разными типами товаров и услуг.

Это хорошее решение. Все же один раз заполнить перечень товаров и потом им пользоваться – это действительно быстро и привычно для большинства компаний. Например, счет покупателю обычно выписывают на основе документа Заказ в 1С, который обязательно резервирует товар на складе для клиента. А если выписать этот же счет в CRM, товар не будет зарезервирован, и у вас постоянно будут возникать накладки с товарными остатками (вы будете получать оплату за товарные позиции, которые отсутствуют на складе).

Мои дела. В разделе Мои дела находится история всех взаимодействий пользователя. Здесь находятся все входящие звонки, задания от руководства, текущие задачи по работе с клиентами и т.д.

Лента и Рабочий стол. Эти элементы относятся к общему порталу Битрикс. Рабочий стол – это аналог Рабочего стола компьютера, где также присутствуют различные рабочие значки. Лента – это перечень всех событий, которые связаны с действиями и задачами клиента.

Создание бизнес-процессов. Создание и разработка бизнес-процессов в любой CRM-системе — задача всегда сложная. Важно продумать цепочку

действий, создать последовательность задач. Например, после получения лида менеджер должен получать задачу “позвонить клиенту”, а после звонка — зафиксировать договоренности в CRM и т.д.

В CRM Битрикс создать такую цепочку очень сложно, в первую очередь, с технической точки зрения. Конструктор, при помощи которого можно создать тот или иной бизнес-процесс, имеется. Но он очень сложен, перегружен различными возможностями и настроен, скорее, на работу в общей системе Битрикс24, а не в CRM Битрикс.

Телефония в Битрикс24 и – это отдельное понятие, она даже в тарифной сетке указывается отдельно. Но с компонентом CRM Битрикс-телефония работает без проблем. Здесь также есть свои особенности:

- Вся телефония работает только через браузер. Т.е. для того, чтобы зафиксировать входящий или исходящий звонок, понадобится подключиться к браузеру через гарнитуру. Даже если у вас есть какие-то SIP-трубки или работает мини-АТС, вы не сможете пользоваться всеми возможностями одновременно. Либо преимущества вашего варианта телефонии, либо работа с CRM.

- Все исходящие звонки, независимо от того, куда они совершаются, даже внутри офиса при наличии мини-АТС, – в CRM-Битрикс платные.

Выводы. Система Битрикс24 предназначена, в первую очередь, для крупного бизнеса. Это большой многофункциональный портал, замкнутый сам на себя. При установке CRM Битрикс выбирается не CRM система, а огромный портал с огромным числом возможностей, который в числе прочих включает CRM-систему. Некоторые специалисты рекомендуют CRM Битрикс24 для крупного бизнеса по причине сочетания большого перечня возможностей со сложностью настройки и обслуживания.

Интеграция «1С-Битрикс» с другим ПО. Компания «1С-Битрикс» последовательно реализует решения по интеграции своих продуктов. Сегодня реализована возможность интеграции CRM Битрикс24 с «1С: Управление торговлей», что позволяет синхронизировать работу со счетами и каталогом то-

варов. До появления возможности интеграции часть действий приходилось дублировать или выгружать из одной программы в другую.

Особенность интеграции от «1С-Битрикс» – наличие штатных механизмов двунаправленного обмена данными. Обмен выполняется с использованием открытого Стандарта обмена коммерческой информацией CommerceML. Для обмена информацией в электронной коммерции необходим общий язык, с помощью которого компании могли бы обмениваться структурированными данными между своими разнотипными компьютерами.

Язык Internet первого поколения, HTML, не подходит для этой цели – он описывает форматирование информации, но не ее смысл. И вот появился XML – ExtensibleMarkupLanguage (расширяемый язык разметки). Как и HTML, он содержит текст, размеченный тегами. Но теги в XML описывают уже и смысл и структуру информации, позволяя напрямую обрабатывать ее программными средствами. Например, Международный совет по прессе и телекоммуникациям недавно утвердил NewsML как основную систему разметки новостной информации, также был создан MathML для математических документов и др.

Для интеграции приобретать ничего не потребуется - нужно выполнить упрощённую настройку продуктов «1С» и «1С-Битрикс». Интеграция производится с каталогом товаров и функционалом выставления счетов из CRM.

Интеграция Корпоративного портала с «1С» предоставляет:

- Актуальные данные в CRM: обновляемый каталог товаров, действующий прайс-лист со всеми внесёнными изменениями, точная информация об остатках на складе.

- Актуальные данные на портале: о структуре компании, о сотрудниках, кадровых перестановках.

- Исполнение ожиданий клиента, потому что информация по всем интересующим его вопросам будет достоверна на любом этапе взаимодействия с клиентом, и счета не потребуют коррекции из-за не вовремя полученной информации, например, о скидках. При этом в курсе состояния сделки в рамках

важной для них информации и бухгалтерия, и менеджер по работе с клиентами, и склад. А руководитель в любой момент получает доступ к информации, имеющейся в «1С».

