

**Московский международный институт эконометрики,
информатики, финансов и права**

БЕЛОВ В.С.

**«Информационно-аналитические
системы»**

Москва 2003

УДК	004
ББК	32.973.202
Б	20

Белов В.С. Учебное пособие «Информационно-аналитические системы». / Московский международный институт эконометрики, информатики, финансов и права. М., 2003 г. – 70.

Рекомендовано Учебно-методическим объединением по образованию в области прикладной информатики (по областям) в качестве учебного пособия для студентов высших учебных заведений, обучающихся по специальности 351400 «Прикладная информатика (по областям)» и другим междисциплинарным специальностям.

© Белов Вячеслав Сергеевич, 2003

© Московский международный институт эконометрики, информатики, финансов и права, 2003

СОДЕРЖАНИЕ

1. Введение в информационно-аналитические системы.....	5
1.1 Роль и место анализа в процессе принятия решения	5
1.2 Аспекты проблемы анализа и их реализация в программных продуктах	7
2. Информационное пространство как среда анализа	9
2.1 Понятие информационного пространства	9
2.2 Структура информационного пространства	10
2.3 Элементы структуры информационного пространства. Понятие показателя	11
2.4 Пространственная интерпретация понятия показатель	13
3. Сущность и система экономических показателей	16
3.1 Содержание экономических показателей	16
3.2 Виды систем показателей	18
4. Анализ и прогнозирование бизнес-процессов (деятельности) предприятий как объект автоматизации	21
4.1 Содержание экономического анализа	21
4.2 Классификация методов анализа	22
4.3 Содержание методов анализа в экономической предметной области	23
4.3.1 Методики проведения анализа в маркетинговой деятельности	23
4.3.2 Анализ обеспечения ресурсами	24
4.3.3 Анализ в области логистики	25
4.3.4 Финансовый анализ	25
4.3.5 Анализ инвестиций и инноваций	26
4.3.6 Методы стратегического анализа	27
4.3.7 Информационный обмен, связанный с аналитической работой	30
5. Информационное хранилище и технологии анализа данных - основа информационно-аналитической системы	33
5.1 Признаки OLAP-систем	33
5.2 Схемы представления многомерных данных	35
5.3 Типы многомерных OLAP-систем	41
5.3.1 Многомерные OLAP-системы	41
5.3.2 Реляционные OLAP-системы	42

5.3.3 Гибридные OLAP-системы	42
6. Концепции организации хранения данных	43
6.1 Понятие информационного хранилища	43
6.2 Концепция централизованного хранилища данных	44
6.3 Концепция распределённого хранилища данных	45
6.4 Концепция автономных витрин данных	46
6.5 Концепция единого интегрированного хранилища и многих витрин данных	48
7. Подходы к выполнению анализа средствами информационных технологий (it-анализа)	49
7.1 Классификация IT-анализа по режиму и темпу	49
7.2 Задачи и содержание оперативного (OLAP) анализа	50
7.3 Интеллектуальный анализ данных Data mining	52
8. Программные инструментальные средства ИАС	54
8.1 Состав программных инструментальных средств ИАС	54
8.2 Средства сбора и доработки данных	55
8.3 Средства преобразования данных	57
8.4 Средства оперативного (OLAP) анализа	58
8.5 Средства интеллектуального анализа данных	59
9. Управление и проектирование ИАС	61
9.1 Управление информационно-аналитическими системами	61
9.2 Задачи и средства администрирования ИАС	62
9.3 Принципы проектирования информационно-аналитических систем	64
9.4 Рынок инструментальных средств ИАС	65
10. Литература	69

1. Введение в информационно-аналитические системы

1. Роль и место анализа в процессе принятия решения
2. Аспекты проблемы анализа и их реализация в программных продуктах

1.1 Роль и место анализа в процессе принятия решения

Современный этап развития рыночных отношений в российской экономике (начало XXI века) характеризуется началом экономического подъёма (23). Период времени быстрых, в значительной мере интуитивных, импровизационных, а зачастую и силовых решений меняется на зону продуманных, просчитанных выводов и решений - оперативных, инвестиционных.

Необходимо также принимать во внимание открытость экономики России и связанной с ней конкуренции с высокоразвитыми экономическими субъектами. В регионах мира со сложившейся развитой рыночной экономикой достижение заметного повышения прибыли (от долей процента) связано со сложной аналитической работой с использованием новейших достижений науки: математики всех направлений, информационных технологий (IT), которые питают и подкрепляют экономические науки, менеджмент, маркетинг, социологию, юриспруденцию и т.д. Начинают приобретать определяющее значение знания о протекающих хозяйственных процессах.

На успех ведения дела влияют как объективные, так и субъективные факторы. К объективным факторам можно отнести:

- закономерности протекания хозяйственных процессов,
- правовую среду,
- неписаные правила и традиции ведения дел,
- экономическую конъюнктуру и т.д.

Большое значение имеет субъективный фактор, под которым будем понимать влияние на ход бизнес-процессов работников предприятия и в особенности лиц, принимающих решения (ЛПР).

Для выработки и принятия соответствующих складывающейся обстановке решений необходима информация, которая должна удовлетворять требованиям полноты, достоверности, своевременности (актуальности), полезности.

Основополагающую роль в подготовке принятия решений играет его обоснование по имеющейся у ЛПР информации. Её, как правило, получают из различных внутренних и внешних источников. В интересах выработки адекватного решения используются внутренние информационные ресурсы, которые складываются из отражения деятельности (функционирования) объекта в документах, других видах и способах сбора, обработки, хранения информации, а также внешние по отношению к объекту информационные ресурсы, например (если это предприятие) - корпорации, отрасли, региона, а также глобальные - из средств массовой информации, специальной литературы, всемирной информационной сети Internet и т.д.

Таким образом, границы информационного пространства как отображения деятельности предприятия и его взаимодействия с внешней средой, в рамках которого принимаются решения, выходят далеко за пределы предприятия.

Одной из первостепенных задач при подготовке и принятии решений является анализ имеющейся в распоряжении ЛПР информации, который является фундаментом обоснования решения.

Объёмы информации, необходимой и используемой при принятии решений, достигают десятков и сотен мегабайт, а в крупных корпоративных и общегосударственных системах и терабайт (10^{12} байт). Информация характеризуется многоплановостью, сложностью отображаемых объектов и систем, а также связей между объектами, явлениями и процессами, скрытостью закономерностей.

Эти обстоятельства вынуждают использовать имеющиеся в настоящее время весьма развитые программно-технические средства. Широкое и эффективное применение этих средств стало одним из факторов выживаемости и успеха предприятия в условиях острой конкурентной борьбы. Получили широкое распространение автоматизированные информационные системы, которые в последние годы чаще называют информационные системы, подразумевая, что без автоматизации их просто невозможно представить.

Проблема анализа исходной информации для принятия решений оказалась настолько серьёзной, что появилось отдельное направление или вид информационных систем - *информационно-аналитические системы (ИАС)*.

1.2 Аспекты проблемы анализа и их реализация в программных продуктах

Вся проблема аналитической подготовки принятия решений имеет три аспекта:

- сбор и хранение необходимой для принятия решений информации;
- собственно анализ, в том числе оперативный и интеллектуальный;
- подготовка результатов оперативного и интеллектуального анализа для эффективного их восприятия потребителями и принятия на её основе адекватных решений.

Аспект, касающийся сбора и хранения информации с сопутствующей доработкой, оформился в концепцию информационных хранилищ (Data Warehouse).

В связи с большим объёмом и сложностью аспект проблемы собственно анализа имеет два направления - оперативный анализ данных (информации), широко распространена англоязычная аббревиатура названия - On-Line Analytical Processing - OLAP. Основной задачей оперативного или OLAP-анализа является быстрое (в пределах секунд) извлечение необходимой аналитики или ЛПП для обоснования или принятия решения информации.

Интеллектуальный анализ информации – имеет также широко распространённое в русской специальной литературе англоязычное название Data mining. Предназначен для фундаментального исследования проблем в той или иной предметной области. Требования по времени менее жёстки, но используются более сложные методики. Ставятся, как правило, задачи и получают результаты стратегического значения.

Жёстких границ между OLAP и интеллектуальным анализом нет, но при решении сложных задач приходится использовать весьма мощные специальные программные средства или, как говорят, инструменты.

Аспекты проблемы анализа и необходимые для их разрешения функции нашли выражение в соответствующих программных продуктах. Средства автоматизации анализа представлены в различных видах. Имеются комплексные информационно-аналитические системы, выполняющие в той или иной степени функции в соответствии с рассмотренными аспектами. Представлены на рынке программных продуктов и целевые программные системы, выполняющие в увеличенном объеме, расширенном составе и повышенной сложности какие-либо функции, например оперативного или интеллектуального анализа. ИАС информационно подпитывают системы поддержки

принятия решений (СППР), в литературе также применяют аббревиатуру DSS (Decisin Support Sistem).

В целом сложился рынок OLAP-систем, информационных хранилищ (DWH), СППР (DSS), интеллектуального анализа Data mining (DMg), который получил обобщённое название - Business intelligence, которому пока не подобран русскоязычный термин.

Как правило, все инструментальные средства, предназначенные для автоматизации аналитических работ, приспособлены для обработки многомерных массивов информации, имеют возможность импорта/экспорта данных в другие операционные среды, развитые средства визуального двумерного (2D) и трёхмерного (3D) представления информации.

Модули, предназначенные для выполнения функций OLAP-анализа, входят также и в состав интегрированных информационных систем (ИИС) (системы, выполняющие весь комплекс автоматизации работ в информационном пространстве экономического или какого-либо другого объекта). Наиболее развитые ИИС выполняют функции и оперативного и интеллектуального анализа.

Необходимо отметить следующее обстоятельство. Целевые программные продукты и ИИС весьма дороги и пока малодоступны для массового российского потребителя. Выходом из этого положения является использование редко применяемых на практике возможностей массовых программных инструментальных средств Excel, Mathcad, Stadia, Statistica и др.

До 2000 года господствующее положение на этом рынке занимали программные продукты иностранных фирм. В настоящее время положение меняется – появилось несколько пакетов программ такого назначения, по своим характеристикам способных успешно конкурировать с зарубежными, а по ряду параметров и превосходящие их. Главные преимущества российских продуктов по показателям цена/качество, отсутствие проблем локализации и др.

2. Информационное пространство как среда анализа

1. Понятие информационного пространства.
2. Структура информационного пространства.
3. Элементы структуры информационного пространства. Понятие показателя.
4. Пространственная интерпретация понятия показатель.

Деятельность предприятия, другого объекта или системы отображается в информационном пространстве. Подготовка принятия решений и, соответственно, анализ происходят в этом пространстве как среде – с одной стороны и производятся операции с элементами структуры этого пространства и сопряжённых 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.

2.1 Понятие информационного пространства

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. (14).

Совокупность информационных объектов, информационно отображающих свойства системы и протекающие в ней процессы, называют информационным пространством. (4) Оно состоит из различных квантов или массивов информации в виде разного рода письменных (знаковых) и фиксированных на носителях информации кодограмм, буквенно-цифровых на естественном языке, устных и визуальных сообщений. Все виды сообщений передаются непосредственно потребителям информации или по каналам связи, могут быть сохранены в различном виде с помощью современных технических средств и по мере необходимости могут воспроизводиться.

Подавляющая часть сообщений, как правило, “загружается” в информационную систему и становится основой её информационного обеспечения. Сочетание знаковых, звуковых (аудио) и образных видимых, в том числе с анимацией, сообщений (визуальных) называют мультимедийными сообщениями, что получило в последние годы широкое распространение в связи с высокой эффективностью

восприятия и широким внедрением таких средств на практике и в состав информационных систем соответственно.

Сообщения могут содержать оперативную информацию о технологических процессах, хозяйственных операциях; экономическую, техническую, организационно-распорядительную, отчётную документацию и т.д.

2.2 Структура информационного пространства

Характерным свойством информационного пространства, как показано выше, является его структурированность. Это свойство обеспечивает возможность представления информации в виде документов и манипулирования данными с помощью программно-технических средств информационных систем.

Хотяшов Э.Н. и Королёв М.А. различает пять степеней структурированности ИП:

- неструктурированное ИП (НИП);
- слабо структурированное (ССИП);
- структурированное (СИП);
- формализованно-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 строгие и т.д.

СИП – отличается преобладанием 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 упрощённым правилам.

2.3 Элементы структуры информационного пространства.

Понятие показателя

В теории ЭИС рассматривается понятие единицы информации (З), под которой понимают “набор символов, которому придаётся определённый смысл”. Выделяют несколько уровней единиц информации в зависимости от смыслового (семантического) значения, его наполненности.

По возрастанию содержательности понятия определены следующие единицы информации: реквизит, составная единица информации (СЕИ), показатель, база данных.

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 подчёркивают 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.

Приведём в данном контексте определение формально-структурного подхода по Королёву М.А. в интерпретации (изложении) Ясина Е.Г. “Показатель представляет высказывание с законченным смыслом, включающее как название переменной величины, так и её 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 какого-либо из названных реквизитов. 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 (8):

$$P \Rightarrow \langle I, X \rangle, \text{ где:}$$

P – показатель (может быть экономическим);

I – набор реквизитов (терминов), идентифицирующих смысловое значение показателя;

X – количественное значение показателя.

Идентификатор в свою очередь можно представить в виде двух частей:

$$I \Rightarrow \langle S, Q \rangle, \text{ где:}$$

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. Реквизиты самого нижнего уровня называются единичными. Другие, расположенные на более верхних уровнях - множественными.

Дальнейший анализ проведём, начав с дополнительных признаков. Они могут состоять из единиц измерения E, уровня показателя Y, времени B, субъекта C. Представим эту структуру в виде соотношения:

$$Q \Rightarrow \langle E, C, B, Y \rangle,$$

таким образом $I \Rightarrow \langle S \langle E, C, B, Y \rangle \rangle$

Наименование показателя может быть слитным (определённым одним реквизитом) или иметь свою структуру и в свою очередь состоять из реквизитов, таких как формальная (вычисляемая) характеристика

показателя - Φ ; обозначение отображаемого технологического или бизнес-процесса - Π ; объект измерения, подсчёта - O .

Тогда $S \Rightarrow \langle \Phi, \Pi, O \rangle$

Таким образом общая структурная формула показателя примет вид:

$$C \Rightarrow \langle S, Q \rangle,$$

$$S \Rightarrow \langle \Phi, \Pi, O \rangle,$$

$$Q \Rightarrow \langle E, C, B, Y \rangle,$$

$$C \Rightarrow \langle \langle \Phi, \Pi, O \rangle, \langle E, C, B, Y \rangle \rangle,$$

$$m \Rightarrow \langle \langle \Phi, \Pi, O \rangle, \langle E, C, B, Y \rangle, x \rangle$$

Данная структура может отображать документ вида

Таблица 2.1

И						
S			Q			
Φ	Π	O	E	C	B	Y

2.4 Пространственная интерпретация понятия показатель

Для образного восприятия сложных понятий удобна их пространственная или геометрическая интерпретация. Представим пространство признаков показателя на системе координат. Число измерений признаков этого пространства равно числу осей его системы координат, на которых откладывают отрезки, соответствующие реквизитам (терминам).

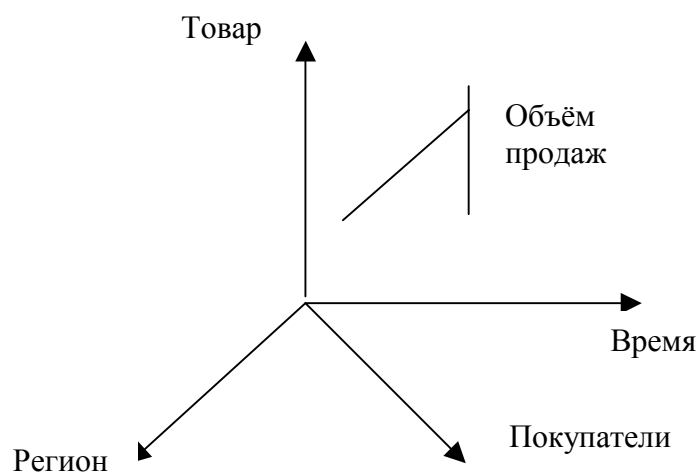


Рис.2.1 Система координат многомерного информационного пространства показателя.

Множественность реквизита выражается в длине отрезка. Тогда составляющие его единичные реквизиты укладываются на этом отрезке. Для трёхмерного пространства это будет куб. В свою очередь можно построить пространство системы взаимоувязанных показателей в виде набора соприкасающихся трёхмерных кубов.

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

Пример такого пространства показателей приведён на рис. 2.2.

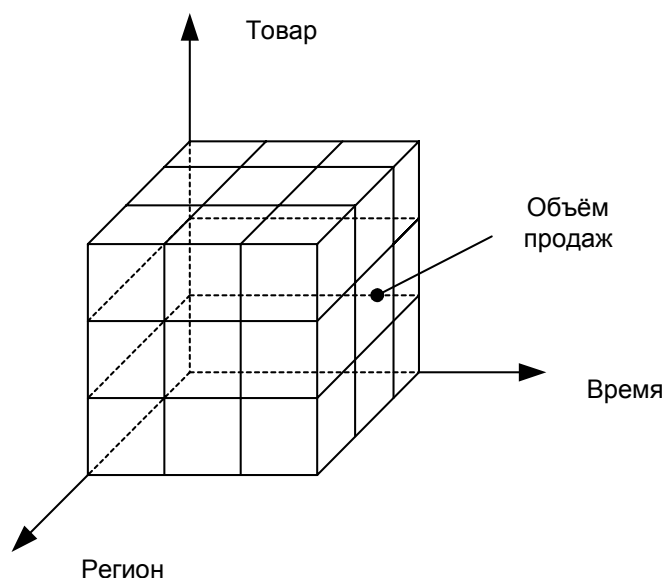


Рис.2.2 Трёхмерный гиперкуб информационного пространства показателя “Объём продаж”

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.

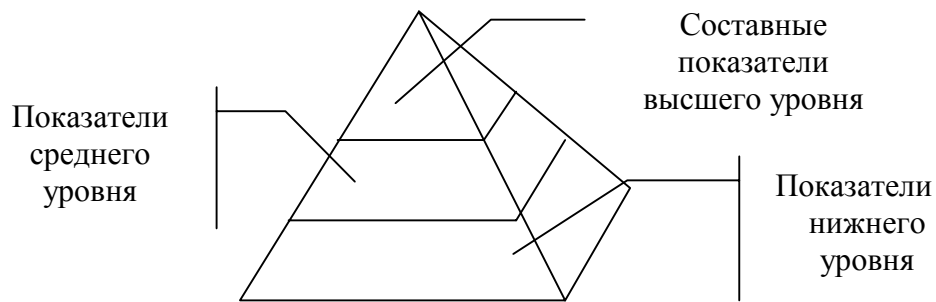


Рис.2.3. Пирамида показателей

Соответствующие структуры представляют также в виде перевёрнутого дерева или “рыбьего скелета” - диаграммы Ишикава.



Рис. 2.4. Диаграмма Ишикава

3. Сущность и система экономических показателей

1. Содержание экономических показателей.
2. Виды систем показателей.

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.

3.1 Содержание экономических показателей

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 характеристике объекта” (5). Это наиболее строгое, на наш взгляд, определение экономической сущности показателя из встречающихся в литературе.

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 ведётся 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, учётной, финансовой отчётности и другой документации. На предприятии разрабатывается внутренняя система показателей и соответствующих реквизитов.

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

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

- внешнюю финансовую отчётность в территориальные государственные органы – налоговую инспекцию и финансовое управление;
- статистическую отчётность о различных сторонах деятельности предприятия в территориальные органы Госкомстата РФ;
- отчётность о финансово-хозяйственной деятельности в государственные фонды (предстоит их реформирование в плане объединения и упрощения отчётности).

Выдаётся информация в виде отчётности и других документов в вышестоящие органы. Госпредприятия и организации представляют её 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 партнёрскими 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 отчётности, 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”. (8).

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

При внедрении ИАС или интегрированной ЭИС необходимо выполнить следующие работы в части структуризации ИП предприятия (корпорации):

- провести анализ состояния, характера и уровня выполненных на предприятии (корпорации) в этой части работ;
- согласовать основные положения структуры ИП с вышестоящими органами (если они имеются), взаимодействующими и партнёрскими организациями;
- оценить с этих позиций информационный обмен, обусловленный внешними связями с учётом международных;
- рассмотреть используемую выбранными инструментальными средствами ИАС систему классификации и кодирования и интерфейсы с точки зрения оценки объёма работ по увязке её с имеющимися на предприятии наработками;
- разработать или модернизировать с учётом имеющегося на предприятии (в корпорации) опыта и наработок систему классификации, кодирования, реквизитов, показателей, ориентируясь на требования, которые предъявляют сложившиеся условия и возможности, предоставляемые внедряемыми инструментальными средствами.

3.2 Виды систем показателей

Количественные показатели отражают в абсолютных величинах происходящие на предприятии процессы в монетарном (денежном) или штучном выражении. Широко используются и относительные величины. Они служат основой финансового и управленческого учёта. Данные учёта, прошедшие аналитическую обработку различного уровня в зависимости от стоящих задач и возможностей, в свою очередь,

являются фундаментом управленческих решений. Как показано выше, одного или нескольких показателей мало для создания достаточно полной картины (представления) происходящих на предприятии процессов и его состояния. На практике (в основном зарубежной) сложилось несколько систем показателей, основанных на экономическом характере объективно существующих между показателями связей. По видам связей различают две группы систем показателей (13) - логико-дедуктивные и эмпирико-индуктивные.

Логико-дедуктивная система показателей строится в виде пирамиды, в основе которой лежат частные показатели, находящиеся в смысловой, подчас сложной взаимосвязи между собой и показателями, находящимися на более высоких “этажах”. На вершине пирамиды находится обобщающий показатель, как бы впитывающий в себя по определённым правилам все показатели нижних уровней (См. рис. 2.3.).

Эмпирико-индуктивные системы составлены с помощью использования статистического отбора показателей, наиболее существенных и значимых с точки зрения подготовки принятия решений.

Логико-дедуктивные системы показателей.

Эти системы имеют широкую область применения, наиболее широко распространены и положены в основу инструментальных средств ИАС и интегрированных ЭИС, используются для целей учёта, анализа, планирования и контроля. В качестве обобщающего показателя используется чаще всего - рентабельность инвестированного капитала (Return On Investment ROI), определяемую с помощью показателей оборота (выручки) E и рассчитанных на этой основе рентабельности оборота капитала $\frac{G_{\text{ИК}}}{E}$ и скорости оборачиваемости капитала $\frac{E}{\text{ИК}}$ и рассчитывается по формуле

$$ROI = \frac{G_{\text{ИК}}}{\text{ИК}} = \frac{G_{\text{ИК}}}{E} \cdot \frac{E}{\text{ИК}}, \text{ где:}$$

$G_{\text{ИК}}$ - прибыль с капитала, рассчитываемая только для определённой, “работающей” на предприятии части имущества;

ИК - сумма основных и оборотных средств.

У показателя ROI имеются недостатки (13), выражающиеся в том, что относительность составляющих показателя - реквизитов затрудняет выделение компонента, за счёт которого произошло изменение показателя; при оптимизации локальные и глобальный оптимум могут войти в противоречия, некоторые весьма важные для оценки ситуации характеристики объекта не охватываются этим показателем.

Однако и при наличии этих замечаний на основе этого показателя построен ряд логико-дедуктивных систем показателей. К ним относятся: Du Pont, Pyramid Structure of Ratios, система показателей аналитических

блоков интегрированных ЭИС SAP R3 и BAAN. Система Du Pont создана и используется одноимённой американской компанией. Pyramid Structure of Ratios разработана Британским институтом менеджмента для выполнения сравнительных аналитических работ по различным предприятиям.

Системы ZWEI, RL используют несколько обобщающих показателей, в том числе: рентабельность, ликвидность и абсолютные, к которым относятся Cash flow, выручка, портфель заказов и др.

Эмпирико-индуктивные системы показателей

Эмпирико-индуктивные системы строятся на основе анализа методами математической статистики результатов деятельности крупных выборок предприятий - около 100. Из большого числа экономических показателей выбираются наиболее существенные для соответствующих целей. В основном такие системы применяются для ранней диагностики финансового состояния предприятий. К таким системам относятся Beaver, Weibel. Эти системы используют для оценки шесть-семь наиболее значимых показателей. Оценка ведётся путём сравнения показателей обследуемых предприятий с “эталонными”.

В настоящее время активно ведутся работы по созданию подобных отечественных систем. В частности широко представлены на российском рынке программные продукты российских фирм “ИНЭК”, Pro Invest Consulting и других фирм.

4. Анализ и прогнозирование бизнес-процессов (деятельности) предприятий как объект автоматизации

1. Содержание экономического анализа.
2. Классификация методов анализа.
3. Содержание методов анализа в экономической предметной области.

4.1 Содержание экономического анализа

В управлении предприятием важными компонентами являются анализ и планирование его деятельности. При этом процесс анализа сочетается с прогнозированием хода различных процессов. Они могут быть контролируемыми, т.е. управляемыми субъектом, или неконтролируемыми, на которые он не в состоянии оказать достаточно заметно влияющее воздействие.

При анализе имеющейся информации наблюдаемые процессы можно разделить по этому признаку (влияем или не влияем на данный процесс) с дальнейшим выявлением существенных факторов, степени их влияния на рассматриваемый процесс и т.д. Такого рода анализ, сочетающийся с прогнозированием, является неременным условием эффективного планирования деятельности предприятия, обоснования принимаемых управленческих решений.

Наблюдаемые или исследуемые процессы - это, в первую очередь, протекающие на предприятии (в корпорации) бизнес-процессы. Под ними подразумевают совокупность работ по выполнению какой-либо задачи предприятия. На основе анализа хода этих процессов, внешних условий, внутреннего состояния предприятия, в том числе финансового, делаются определённые выводы, вырабатываются или корректируются цели предприятия. В соответствии с выработанными целевыми установками ставятся задачи, осуществляется планирование мероприятий и деятельности предприятия в целом. При этом учитываются принятые критерии оценки, вырабатываются управленческие решения по реализации планов. Эту часть анализа назовём стратегическим анализом.

В процессе реализации планов должен осуществляться контроль и анализ хода их реализации, который назовём текущим анализом. Результаты его являются материалом для выработки решений по корректировке, с одной стороны уже состоявшихся управленческих решений, а с другой - по доработке самих планов или даже целевых установок в случае значительных отклонений полученных показателей от запланированных, отсутствия ресурсов или в связи с какими-либо другими обстоятельствами.

В целом содержание анализа (11) состоит в систематизации, оценке полученных параметров в соответствии с принятой системой показателей, изучении и оценке факторов, влияющих на деятельность предприятия, выявлении его сильных и слабых сторон, определении возможностей и рисков.

4.2 Классификация методов анализа

Существует большое количество методов анализа, которые делятся на группы по различным признакам.

Рассмотрим систему признаков, характеризующих методы анализа. Их можно сгруппировать:

По целям - это:

- оценка состояния и результатов деятельности предприятия;
- постоянный контроль рациональности ведения хозяйственной деятельности, выявление резервов для обеспечения выполнения поставленных задач;
- прогнозирование хода внутренних процессов на предприятии и внешних факторов, влияющих на его деятельность.

По временному фактору анализ разделяют на:

- использующий прошлую информацию, отражённую в документации и на различных носителях и содержащуюся в информационной системе - анализ фактов;
- на базе как прошлой, так и обращённой в будущее, то есть прогнозной информации - анализ событий и отклонений;
- анализ будущей информации - по существу оценка бюджетов и планов, их альтернатив.

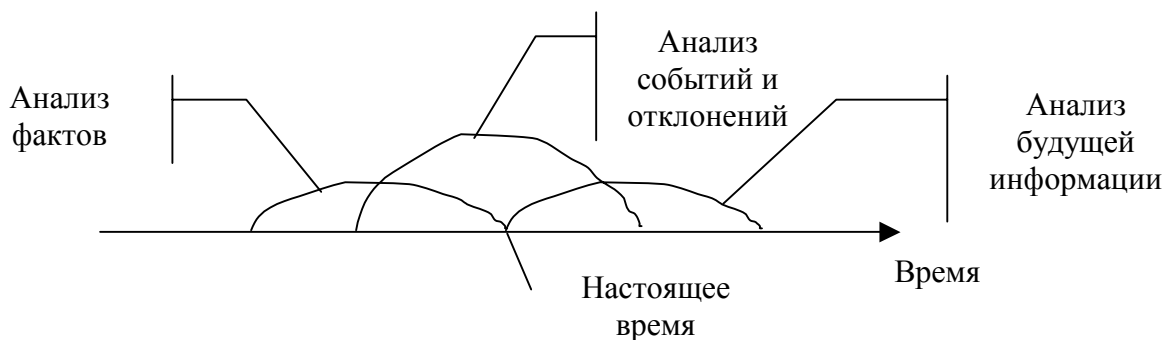


Рис. 4.1. Деление анализа по временному фактору.

По масштабности решаемых или обслуживаемых задач:

- стратегический, сюда можно отнести оценку эффективности целей, долгосрочные прогнозы, исторические оценки процессов и явлений и т.д.;
- оперативный - это оценка текущего состояния, выявление узких мест и отклонений;
- система раннего предупреждения.

По предметным областям, в рассматриваемом случае – экономики различают анализ:

- в маркетинге;
- производственной или основной деятельности;
- в логистике;
- обеспечении ресурсами;
- финансовой;
- в сфере инвестиций и инноваций.

По методам различают:

- сравнительный по подразделениям, предприятиям, регионам, временным периодам и т.д.;
- анализ отклонений;
- функционально-стоимостный;
- анализ цепочки создания стоимости и конкурентный анализ по Портеру;
- анализ полей бизнеса (Profit Impact of Market Strategies - PIMS);
- бенчмаркинг (Beanchmarking);
- интеллектуальный анализ (Data mining).

В процессе анализа используются различные математические методы, в том числе:

- математической статистики;
- многомерного статистического анализа,
- эконометрики;
- алгебры – линейная, логики, предикатов, нечёткой логики;
- численные методы анализа.

Необходимо заметить, что какой-либо конкретный аналитический процесс или аналитическая работа могут характеризоваться одновременно несколькими из перечисленных выше признаков. Понятно, что приведённая система признаков анализа не претендует на полноту и носит рекомендательный характер.

4.3 Содержание методов анализа в экономической предметной области

Рассмотрим содержание (краткий обзор) некоторых методов по предметным областям и методикам проведения. Знание методов анализа необходимо для сценариев OLAP – и интеллектуального анализа. Это облегчает реализацию рассматриваемых ниже методик (ряд из них без использования этих средств будут мало эффективными).

4.3.1 Методики проведения анализа в маркетинговой деятельности

Анализ разрыва (Gap analysis) является средством долгосрочного (стратегического) планирования. Основой метода является сравнение стратегий оптимального и возможного развития. Составляется матрица

оценок стратегий по принятым показателям, по ним строятся профили стратегий. Используются экспертные оценки по балльной качественной системе. .

Портфолио-анализ - подбор такого портфеля инвестиций с учётом рисков, то-есть сочетания возможных потерь и доходности, который обеспечил бы наименьшие потери с максимально возможными доходами.

Анализ маржинальной прибыли (МП). Исследуется реакция величины МП на маркетинговые мероприятия. Объектами анализа являются продукты, регионы, заказы, группы клиентов и т.д. Производится выявление причин убытков или резкого повышения прибыли, их локализация и вырабатываются предложения по ликвидации "узких" мест или распространению передового опыта. Величину МП распределяют по различным объектам исследования: продуктам, группам продуктов, продуктовым сегментам рынка, предприятиям, корпорации в целом.

Сравнительные расчёты. Определяют зависимость МП или выручки от расходов: на рекламу, послепродуктовое обслуживание клиентов, торговые издержки, в том числе площади; других расходов на маркетинг.

4.3.2 Анализ обеспечения ресурсами

Общие подходы заключаются в исследовании рынков закупки товаров и анализе поставщиков, разделении материалов и комплектующих по номенклатуре, качеству, количеству, ценам у различных поставщиков. Выделяются факторы, влияющие на процесс снабжения, оценивается "совокупное предложение на рынке", а затем выделяются нужные или подходящие поставщики.

АВС-анализ. Метод, позволяющий выделить наиболее значимые для предприятия группы товаров. Рассчитываются количество и стоимость потребляемых материалов. По итогам расчёта формируются три группы товаров А, В и С. Товары А обладают наибольшей кумулятивной стоимостью КС (количество единиц товара×цена ед. товара). Товары группы С имеют наименьшую кумулятивную стоимость. При этом соблюдается условие: КС группы А=50% всей КС; Совместная КС групп А и В = 90% всей КС. Такое разделение позволяет сосредоточить внимание на направлениях, где ожидается наибольшая польза. Этот метод применяется не только в анализе обеспечения ресурсами.

Анализ возможных прерываний бизнес-процесса. Из-за непоступления исходных материалов или необеспеченности другими ресурсами или услугами может быть нарушен производственный или другой процесс. Остановы могут быть частичными или полными. Результатами их являются потери, затраты, упущенная выгода.

Оцениваются факторы, связанные с убытками, готовность поставщиков, подбираются альтернативы.

Определение верхних границ цен. Под ними понимают максимальную цену, которую готов заплатить покупатель, в том числе и при закупке материалов. Этот уровень зависит от потребности и значимости товара или материала для обеспечиваемого бизнес-процесса. Цена альтернативного товара-заменителя служит ориентиром верхней границы цены.

4.3.3 Анализ в области логистики

Логистика это наука и практика управления продвижением (перемещением и хранением) товара от производителя к потребителю.

Анализ цепочек логистических процессов позволяет обеспечить руководство предприятия информацией по предметам логистики и выработать соответствующие решения, осуществить согласование и оптимизацию материальных и сопутствующих им информационных потоков с другими процессами, протекающими на предприятии, и с партнёрами. Определяются потребности в материалах, транспорте, складских площадях и т.д., исходя из планов заказов и производства. Выбирается методика управления складскими запасами по ритму или срокам поставок. Определяется потребность в поставке по мере достижения минимально допустимого объёма или по окончании рассчитанного периода времени. Например вычислением средней скорости потребления по прошлым периодам или по производственной мощности, темпам производственного процесса.

Анализ издержек логистических процессов

Выявляются места возникновения издержек, к ним относятся объекты приёма-выдачи материалов и полуфабрикатов, склады, система транспортирования, сопутствующий информационный обмен, включая документооборот. Далее в соответствии с принятыми в логистической цепи методами учёта определяются составляющие издержек. В процессе анализа оцениваются прогнозные и реальные издержки. Проводится контроль экономичности по принятой системе показателей, оценивается степень готовности поставок и продвижения, информационное обеспечение и другие показатели.

4.3.4 Финансовый анализ

Задачи и содержание финансового анализа подробно рассматриваются в соответствующих курсах, широко освещаются в литературе. Стержнем насчитывающей несколько десятков показателей

системы оценки финансового состояния предприятия являются показатели ликвидности и рентабельности, отражающие платёжеспособность и прибыльность предприятия.

Методами анализа улавливаются неблагоприятные или критические ситуации, принятые по результатам анализа меры обеспечивают приемлемые или оптимальные значения и соотношения показателей, подтверждающие выправление положения. Необходимо отметить, что анализ в финансовой сфере тесно увязан с планированием. Каждые плановые предложения или решения должны тщательно оцениваться на предмет реализуемости планов, недопущения недостатка или избытка средств, достижения необходимой и достаточной эффективности их использования.

Методическими инструментами финансового анализа, которые заложены в программные информационно-аналитические средства, являются:

Анализ потоков платежей (Cash flow analysis) КФ -анализ - баланс притока и оттока финансовых средств. На его основе определяются показатели: маржинальной прибыли; безубыточности, в том числе точка безубыточности, точка закрытия предприятия, кромка безопасности; эффект операционного рычага; коэффициент выручки. Важное значение среди других имеет показатель работающего капитала, который должен быть положительным

Для прогнозирования критического состояния используется *показатель Z-счёт Альтмана*, вычисляемый по балансу и отчёту о прибылях и убытках.

Финансовая "паутина". Для поддержки принятия решений важное значение имеют графические методы представления состояния объекта, в данном случае финансового состояния. Одним из таких инструментов является этот метод, который является одним из вариантов многомерного графического представления данных. В ИАС входит как один из способов наглядного отображения финансового состояния.

4.3.5 Анализ инвестиций и инноваций

Расчёты, связанные с инвестициями и проектированием представляют собой самостоятельное направление финансово-экономического планирования, тесно связанного с инженерно-технологическими исследованиями и решениями.

Задачи анализа в этой области экономической деятельности заключаются в сравнительных оценках альтернатив, мониторинге реализации инвестиционных и инновационных проектов по принятой системе показателей. Помимо ставших традиционными финансовых оценок по ряду специальных показателей используется также *функционально-стоимостный анализ*. В основе его лежат субъективные оценки проектов путём составления иерархии целей, их взвешивания,

составления таблиц функций и определения альтернатив реализации функций. Производится расчёт полезности и формирование последовательности в матрице ценности целей. Функционально-стоимостный анализ завершается анализом чувствительности полученных данных к изменению весовых коэффициентов целей, оценкой и выдачей результата.

Используются различные методы оценки инвестиционных и инновационных проектов в условиях неопределённости. Подробно описаны в (12), к ним относятся:

- анализ ставки дисконтирования с поправкой на риск;
- метод достоверных эквивалентов с вариантами использования в качестве их математического ожидания денежных потоков и состояния предпочтения;
- методики принятия решений без использования численных значений вероятностей, основанные на построении и анализе матрицы стратегий и состояний природы для инвестиционного проекта методами максимакса, максимина, минимакса и компромиссного - Гурвица;
- опционный, использующий подходы, принятые при оценке ценных бумаг.

4.3.6 Методы стратегического анализа

4.3.6.1 Анализ стратегической позиции предприятия

Для оценки стратегической позиции предприятия используются несколько методик.

SWOT анализ - аббревиатура английских слов strengths, weaknesses, opportunities, threats сильные, слабые имеются в виду стороны предприятия, возможности, опасности. На основе анализа внутренней и внешней среды, выявления ключевых факторов успеха, социальных аспектов строится четырёхклеточная матрица. Клетки её заполняются соответствующими данными. Полученные данные позволяют сформировать стратегию предприятия, которая закладывается в планы, исполняется, результаты подвергаются очередному этапу анализа.

Матрица БКГ (Бостонской консультативной группы). Схожий подход. Результаты аналитической работы представляются таким же образом. Определяются позиции предприятия на рынке по сравнению с ведущей фирмой в данном сегменте рынка, все направления деятельности разбиваются на четыре группы. В их отношении вырабатываются соответствующие стратегии. Нарботаны типовые

рекомендации, суть которых сводится к поддержке перспективных, ликвидации безнадёжных направлений деятельности.

Матрица Мак-Кинси является развитием матрицы БКГ. Эта методика предусматривает использование формализованных показателей привлекательности рынка и конкурентного статуса. В исходных данных используются экспертные оценки прогнозные показатели. Подробно метод рассмотрен в (12).

Анализ цепочки создания стоимости и конкурентный анализ по Портеру. Им предложено представить совокупность выполняемых предприятием функций в виде цепочек процессов создания стоимости. В начале и конце цепочек деятельность предприятия интегрируется (согласуется) с деятельностью партнёров по бизнесу.

Конкурентный анализ проводится на "поле сил", действующих на предприятие. Автор выделил пять основных, среди которых: влияние покупателей, влияние поставщиков; возможность появления новых конкурентов, существование товаров-заменителей, действия конкурентов внутри отрасли. Исследуются факторы, обуславливающие эти силы, оценивается их соотношение. По материалам анализа вырабатывается оптимальная стратегия. Конкретных рекомендаций методика не даёт и ограничивается качественным анализом.

4.3.6.2 Анализ ситуации по слабым сигналам и оценка рисков.

Методика анализа ситуации по слабым сигналам даёт рекомендации по установке контрольных точек, определяет или устанавливает уровни нестабильности, осведомлённости. Предусматриваются варианты реакции на сигналы.

Оценка рисков и управление ими. Риск рассматривается как возможность потерь в виде убытков, упущенной выгоды или как степень нестабильности, непредсказуемых исходов.

Проводится качественный и количественный анализ рисков. При качественном анализе выявляются факторы, зоны опасности, виды рисков.

Количественный анализ использует методы аналогий, Монте-Карло, экспертные, анализа чувствительности (что..., если...), сценариев.

4.3.6.3 Анализ отклонений.

В комплексе аналитических работ на предприятии анализ отклонений играет весьма существенную роль. После разработки

системы целей, выбора стратегий и рассчитанных на их основе планов и бюджетов в процессе их реализации необходим контроль. В идеале он должен сопровождать каждый процесс и быть непрерывным. На практике он реализуется выборочно для наиболее значимых и существенных процессов с допустимой периодичностью. Выводы о степени реализации планов и бюджетов делают посредством анализа отклонений числовых и/или качественных показателей в принятой на предприятии системе.

Различают *абсолютные и относительные* показатели. В экономической и других предметных областях имеется проблема знака отклонения. Иногда снижение значения показателя означает "хорошо" и наоборот. Это обстоятельство необходимо учитывать. Например рост прибыли и убытков.

Селективные отклонения предусматривают сравнения во временном аспекте. Отрезок времени текущего года или другого периода сравнивается с таким же - предыдущего.

Кумулятивное отклонение получаем при сравнении значений показателей, полученных нарастающим итогом.

Рассматриваются отклонения *план-факт, факт-факт* - сравнение с прошлым фактом в сопоставимом отрезке времени, *план-желаемый результат*, когда сравнивается плановый показатель с желательным с учётом изменившихся условий.

Оценку отклонений производят по допустимым пределам и по влиянию на прибыль или другой обобщающий показатель, например ROI.

В процессе анализа выявляются места и причины отклонений. Для оценки величин отклонений может быть использована *методика цепных подстановок* (12), которая представляет собой совокупность формул и схем расчёта на основе цепочек создания стоимости, позволяющая в итоге вычислить отклонения по обобщающему показателю на основе имеющихся исходных данных. Для реализации этого метода необходимо реализовать на предприятии систему классификации и кодирования показателей, которая была рассмотрена выше. В интегрированных ЭИС имеются модули, выполняющие подобные задачи на основе принятой в конкретном программном инструментальном средстве системы классификации и кодирования.

Исследование причинно-следственных связей и других интересующих ЛПР и аналитиков явлений ведётся с использованием методов интеллектуального анализа (см. раздел 5.3). Анализ отклонения

может быть обращён как в ретроспективу, так и в перспективу. Исследование ретроспективы ведётся в интересах извлечения знаний и формирования на их основе выводов на перспективу.

4.3.6.3 Анализ полей бизнеса

Это исследование воздействия рыночных стратегий на прибыль для данного предприятия или для отдельных полей бизнеса, или видов деятельности на базе информации о более чем 2000 предприятий, содержащейся в базах данных специализированных фирм. Учитывается взаимовлияние специфической внешней среды данного вида бизнеса и внутренней ситуации на предприятии. В качестве обобщающих показателей используются ROI и денежные потоки - Cash-balance.

4.3.6.4 Бенчмаркинг

Одним из условий выживаемости предприятия, что особенно актуально для нынешних российских условий, является достижение мирового уровня рыночной привлекательности продукции или услуг. Здесь имеется в виду совокупная оценка свойств продукции, связанных с ней услуг, а также процессов на самом предприятии. Целью анализа является выявление лучшего в отрасли или на данном поле бизнеса продукта или предприятия, выявление и оценка уровня собственного отставания или опережения. Сравниваются также производственные, управленческие и иные функции. На основании анализа вырабатываются меры по устранению отставания или закреплению успехов.

4.3.7 Информационный обмен, связанный с аналитической работой

Аналитическая работа на предприятии осуществляется специальной группой. Она может быть автономной или включённой в какое-либо подразделение. В последнее время создаются подразделения контроллинга, в чьи функции в качестве основной включается эта деятельность. В отдельных, особо сложных ситуациях пользуются услугами консультантов. На малых предприятиях эта работа может быть возложена на одного из заместителей руководителя или эксперта.

Для уяснения функций ИАС необходимо изучить информационный обмен, связанный с аналитической работой. В общей постановке анализ основан на переработке информации, которую аналитики должны где-то получить, и выдаче информации заинтересованным лицам или организационным единицам. Место

аналитического процесса в цепочке других, связанных с управлением процессов показано на рис. 4.1.

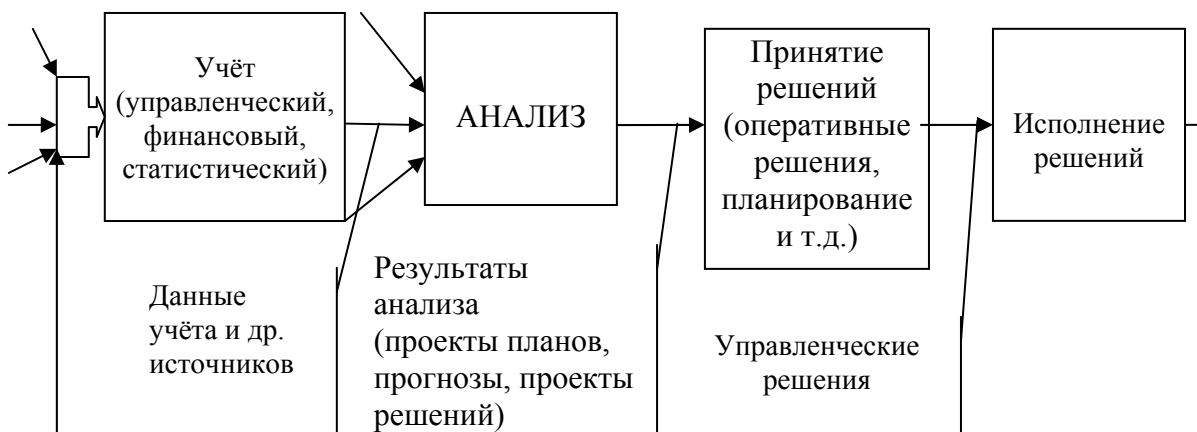


Рис. 4.1. Место анализа в цепочке принятия управленческих решений.

Источники информации для анализа делятся на внутренние и внешние.

К внутренним источникам относятся:

- бухгалтерский учёт, включая аналитический и складской;
- статистический учёт,
- управленческий учёт;
- деловая переписка;
- материалы различных исследований и обследований, выполненных на предприятии
- текущая документация, в том числе материалы ревизий и аудиторских проверок и т.д.;
- зафиксированные данные опросов;
- устная информация;
- информация из баз данных, эксплуатирующихся на предприятии ЭИС и автономных автоматизированных рабочих мест (АРМ).

Из перечисленных видов учёта бухгалтерский и статистический относятся к обязательным видам учёта.

К внешним источникам информации относятся:

- установочная информация из государственных органов и вышестоящих организаций (для зависимых предприятий) это правовые и руководящие документы, инструкции и т.д., определяющая условия функционирования;
- информация из специализированных информационных организаций и их информационных хранилищ, к ним относятся различные фонды, финансовые и биржевые и т.д.;
- библиотечные фонды и информационные хранилища;

- средства массовой и специализированной информации;
- глобальные информационные ресурсы, например сеть Internet и другие;
- данные деловой разведки и прочие возможные источники информации.

С другой стороны служба анализа выдаёт информацию заинтересованным потребителям. Основной потребитель её - лица, принимающие решения (ЛПР). На предприятии потребителями её являются также службы управления предприятием. К ним относятся:

- бухгалтерская и финансовая служба;
- служба контроллинга или другие её подразделения, если аналитики входят в её состав;
- маркетинговое подразделение;
- служба логистики;
- технологические и производственные;
- информационная и PR;
- другие заинтересованные структуры и лица.

На предприятии должен быть установлен порядок доступа к такой информации по причине её особой ценности и подчас конфиденциальности.

Форма представления информации

Информация для ЛПР и смежных служб может представляться на бумажных носителях в виде:

- аналитических записок;
- отчётов;
- предложений, справок и т.д.

Виды и формы документов должны соответствовать российским и международным стандартам документооборота. Это не означает, что исключаются какие-либо иные формы.

Аналитические материалы и другие данные для обоснования и принятия решений представляются и на других видах носителей. Формируются в ИАС и ИЭИС в виде тех же документов в качестве электронной копии или специально сформированных информационных массивов, которые изготавливаются с учётом рекомендаций инженерной психологии, обеспечивают потребителей сжатой, доходчивой информацией с использованием всех имеющихся на предприятии средств, в том числе и мультимедийных.

5. Информационное хранилище и технологии анализа данных - основа информационно-аналитической системы

1. Признаки OLAP-систем.
2. Схемы представления многомерных данных.
3. Типы многомерных OLAP-систем.

Программные инструментальные средства, обеспечивающие автоматизацию аналитических работ в целях поддержки принятия решений, в литературе получили два распространенных названия: OLAP - системы и информационные хранилища. Использование того или иного зависит от приверженности автора к продукту той или иной фирмы. Под этими названиями подразумевают и полный набор средств и частные подсистемы.

5.1 Признаки OLAP-систем

Доктор Кодд сформулировал в 1993 году 12 правил, определяющих требования к программным системам, претендующим на название OLAP. К 12 правилам впоследствии были присоединены ещё шесть.

В них содержатся некоторые противоречия, не все авторы безусловно их принимают, к тому же имеется некоторая расплывчатость определений.

В конце 90-х годов получил распространение свод требований (27) к информационно-аналитическим системам в виде "теста FASMI" - аббревиатуры английских слов, определяющих требования к OLAP-системам:

Fast Analysis Shared Multidimensional Information - русский перевод Быстрый Анализ Разделяемой Многомерной Информации.

Раскроем содержание перечисленных свойств, которыми должна обладать ИАС.

Fast Быстрый - это свойство выражается во временных требованиях к ответам системы на запросы пользователей. Ответ должен быть получен обычно за время в пределах секунды. Более сложные запросы допускается обрабатывать в течение 5-ти секунд и лишь отдельные запросы допускаются с 20-секундной реакцией. Такие требования связаны с психофизиологическими показателями аналитиков и ЛПР, обусловлены достижением наиболее значимых результатов анализа при выполнении этих требований. Специальные исследования показали, что при времени ответа более 30-ти секунд наступает раздражение и возможна реакция в виде перезапуска системы.

Analysis Анализ - возможности системы выполнять аналитические работы различного характера в предметной области пользователя собственными средствами, не прибегая к программированию. Для

описания специфических для данного пользователя аналитических процессов могут применяться встроенные средства в виде языков высокого уровня, электронных таблиц со встроенными функциями, графических конструкторов, визуальных средств с применением кнопочных и рамочных технологий.

Shared Разделяемый - система должна обеспечивать необходимый уровень защиты при множественном доступе для исключения взаимных помех, несанкционированного доступа, ведь ценность результатов анализа гораздо выше исходной информации.

Multidimensional Многомерный - определяющее требование. Средства OLAP-системы должны обеспечить работу с данными в многомерном представлении на концептуальном уровне с полной поддержкой иерархий. Требование считается выполненным независимо от того, какой тип базы данных используется, не устанавливаются рамки количества измерений.

Information Информация - должна обеспечиваться возможность получения её из любых необходимых источников. Инструментальные средства оперируют с необходимыми объёмами и структурами данных.

Более подробно рассмотрим свойство многомерности, так как оно является наиболее характерным отличительным от других систем свойством, в частности OLTP.

Как показано в п.2. информационное пространство, отображающее функционирование объекта, многомерно. Естественно стремление аналитика и ЛПР к тому, чтобы иметь дело с моделью данных в наиболее естественном виде. Это обстоятельство привело к тому, что с помощью современных программно-технических средств, имеющих широкие возможности интерпретации данных, были созданы соответствующие многомерные модели. Теоретические основы были заложены в трудах крупных российских учёных Ясина, Королёва и др. ещё в 70-х годах XX века. В трудах Кодда, Инмона легко узнаются основополагающие идеи этих и других учёных, которые были реализованы в большом числе проектов в разных предметных областях.

В последнее десятилетие XX века основной моделью данных, использованной в многочисленных инструментальных средствах создания и поддержки баз данных – СУБД, была реляционная модель. Данные в ней представлены в виде множества связанных ключевыми полями двумерных таблиц – отношений. Для устранения дублирования, противоречивости, уменьшения трудозатрат на ведение баз данных применяется формальный аппарат нормализации отношений. Однако применение его связано с дополнительными затратами времени на формирование ответов на запросы к базам данных, хотя и экономятся ресурсы памяти.

Многомерная модель данных (см п. 2) представляет исследуемый объект в виде многомерного куба, чаще используют трёхмерную модель. По осям или граням куба откладываются измерения или реквизиты-

признаки. Реквизиты-основания являются наполнением ячеек куба. Многомерный куб или как иногда называют пул данных может быть представлен комбинацией трёхмерных кубов с целью облегчения восприятия и квазиобъёмного представления при формировании отчётных и аналитических документов и мультимедийных презентаций по материалам аналитических работ в системе поддержки принятия решений.

5.2 Схемы представления многомерных данных

Многомерные данные могут быть отображены инструментами в виде СУБД на основе реляционных моделей данных, а также и специальными многомерными инструментальными средствами.

Представление многомерных данных в рамках реляционных моделей может выполняться в виде трёх вариантов схем:

- "звезда";
- "снежинка";
- "созвездие".

Линейное представление на плоскости отображено на рис 5.1.

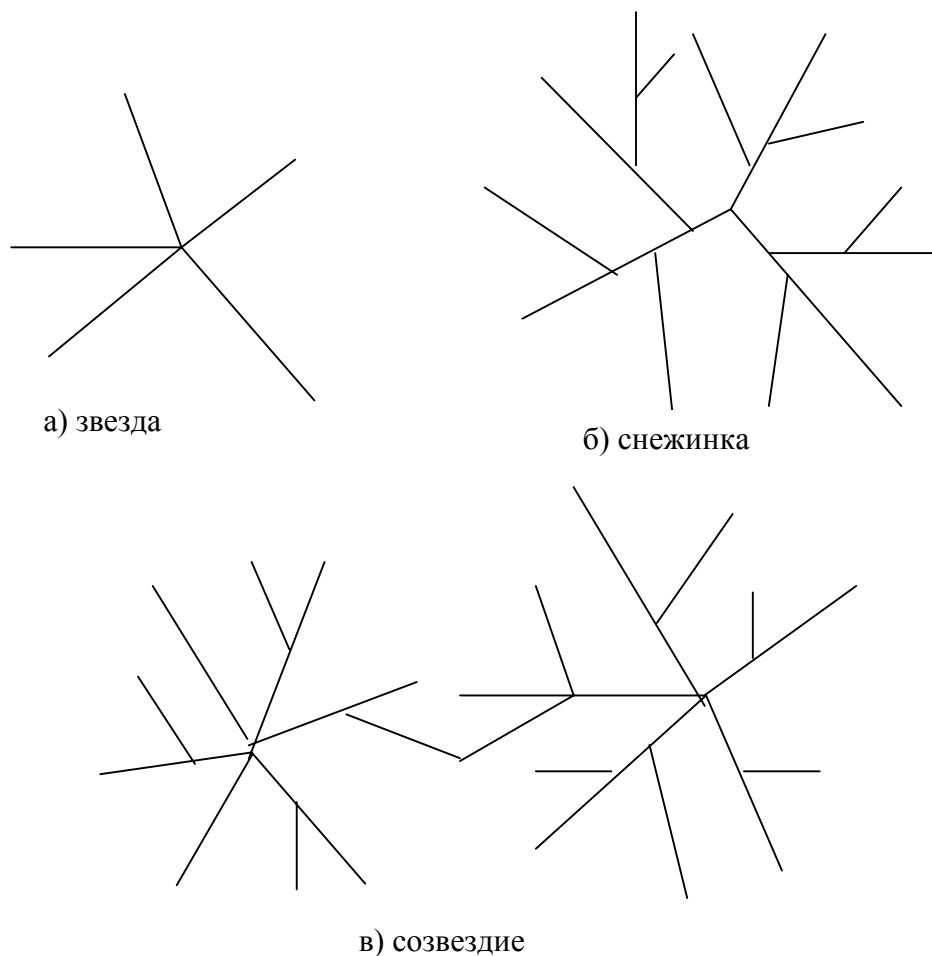


Рис. 5.1. Линейное представление схем многомерных данных

Названные схемы являются системами таблиц реляционной модели.

На рис. 5.2. представлены схема базы данных Northwind, входящей в комплект поставки СУБД MS SQL Server и MS Access, а также варианты схем построенных на их основе кубов данных.

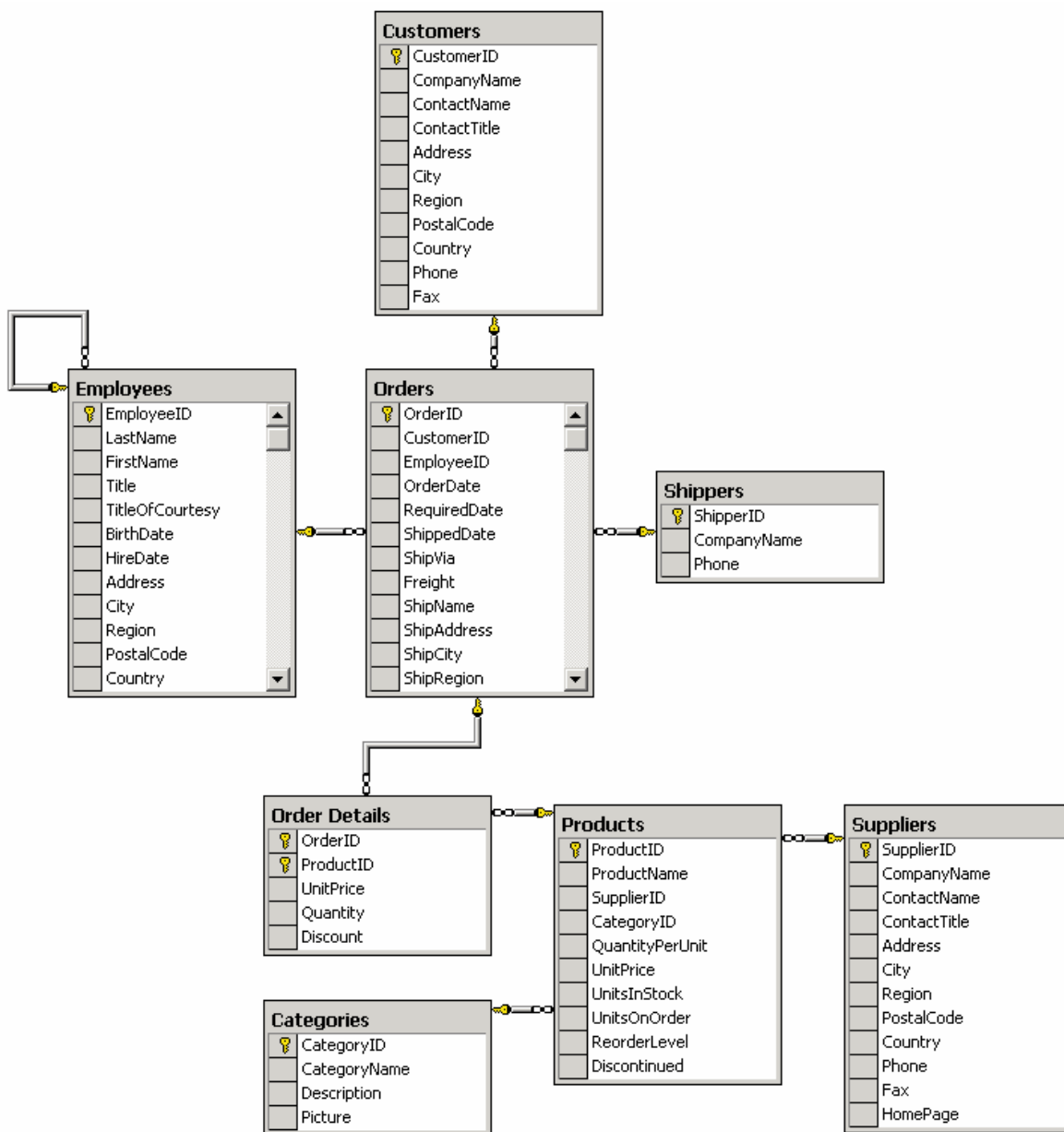


Рис.5.2. Схема базы данных Northwind

В многомерном пуле информации создаётся большая центральная таблица, называемая таблица факта (fact table). В ней помещаются все данные относительно интересующего пользователя обобщающего показателя. Её окружают меньшие таблицы, содержащие данные по признакам, называемые таблицы размерности или иногда их называют измерений (dimensional table).

Таблицы размерности являются родительскими по отношению к таблице факта. Таблица факта является дочерней. Могут быть также

консольные таблицы (outrigger table). Они присоединяются к таблицам размерности и детализируют отдельные атрибуты. Консольные таблицы являются родительскими по отношению к таблицам размерности.

Таблицы фактов содержат числовые или качественные (содержательные) значения.

TimeKey	CustomerKey	ShipperKey	ProductKey	EmployeeKey	RequiredDate	LineItemFreight	LineItemTotal	LineItemQuantity	LineItemDiscount
5	85	4	11	5	01.08.1996	14.3904	168	12	0
5	85	4	42	5	01.08.1996	11.992	98	10	0
5	85	4	72	5	01.08.1996	5.996	174	5	0
1	79	1	14	6	16.08.1996	2.1321	167.4	9	0
1	79	1	51	6	16.08.1996	9.476	1696	40	0
3	34	2	41	4	05.08.1996	10.971	77	10	0
3	34	2	51	4	05.08.1996	38.3985	1484	35	222.6
3	34	2	65	4	05.08.1996	16.4565	252	15	37.8
4	84	1	22	3	05.08.1996	6.0492	100.8	6	5.04
4	84	1	57	3	05.08.1996	15.123	234	15	11.7
4	84	1	65	3	05.08.1996	20.164	336	20	0
2	76	2	20	4	06.08.1996	19.54	2592	40	129.6
2	76	2	33	4	06.08.1996	12.2125	50	25	2.5
2	76	2	60	4	06.08.1996	19.54	1088	40	0
5	34	2	31	3	24.07.1996	11.404	200	20	0

Sales_Fact	
Primary Key	TimeKey
Primary Key	CustomerKey
Primary Key	ShipperKey
Primary Key	ProductKey
Primary Key	EmployeeKey
	RequiredDate
	LineItemFreight
	LineItemTotal
	LineItemQuantity
	LineItemDiscount

Рис.5.3 Таблица факта.

SQL Server Enterprise Manager - [4:Data in Table 'Product_Dim' in 'Northwind_Mart' on 'MAINDESK']

Console Window Help

SQL

	ProductKey	ProductID	ProductName	SupplierName	CategoryName	ListUnitPrice
▶	1	1	Chai	Exotic Liquids	Beverages	18
	2	2	Chang	Exotic Liquids	Beverages	19
	3	3	Aniseed Syrup	Exotic Liquids	Condiments	10
	4	4	Chef Anton's Cajun Seasoning	New Orleans Cajun Delights	Condiments	22
	5	5	Chef Anton's Gumbo Mix	New Orleans Cajun Delights	Condiments	21.35
	6	6	Grandma's Boysenberry Spread	Grandma Kelly's Homestead	Condiments	25
	7	7	Uncle Bob's Organic Dried Pears	Grandma Kelly's Homestead	Produce	30
	8	8	Northwoods Cranberry Sauce	Grandma Kelly's Homestead	Condiments	40
	9	9	Mishi Kobe Niku	Tokyo Traders	Meat/Poultry	97
	10	10	Ikura	Tokyo Traders	Seafood	31
	11	11	Queso Cabrales	Cooperativa de Quesos 'Las Cabras'	Dairy Products	21
	12	12	Queso Manchego La Pastora	Cooperativa de Quesos 'Las Cabras'	Dairy Products	38
	13	13	Konbu	Mayumi's	Seafood	6
	14	14	Tofu	Mayumi's	Produce	23.25
	15	15	Genen Shouyu	Mayumi's	Condiments	15.5
	16	16	Pavlova	Pavlova, Ltd.	Confections	17.45
	17	17	Alice Mutton	Pavlova, Ltd.	Meat/Poultry	39
	18	18	Carnarvon Tigers	Pavlova, Ltd.	Seafood	62.5
	19	19	Teatime Chocolate Biscuits	Specialty Biscuits, Ltd.	Confections	9.2

Product_Dim

🔑	ProductKey
	ProductID
	ProductName
	SupplierName
	CategoryName
	ListUnitPrice

Рис. 5.4 Таблица измерений.

При разработке базы данных по схеме "звезда" или по другой многомерной схеме необходимо глубоко и тщательно проанализировать предметную область; поместить в центральную таблицу факта все характеризующие исследуемый объект данные, предварительно разработав систему признаков.

Консольные и таблицы размерности, а также таблица факта соединяются идентифицирующими связями. Первичные ключи родительских таблиц являются внешними ключами дочерних. Например первичный ключ таблицы размерности является внешним ключом таблицы факта.

Схема "звезда" состоит только из таблиц размерности и таблицы факта (рис. 5.1. а)

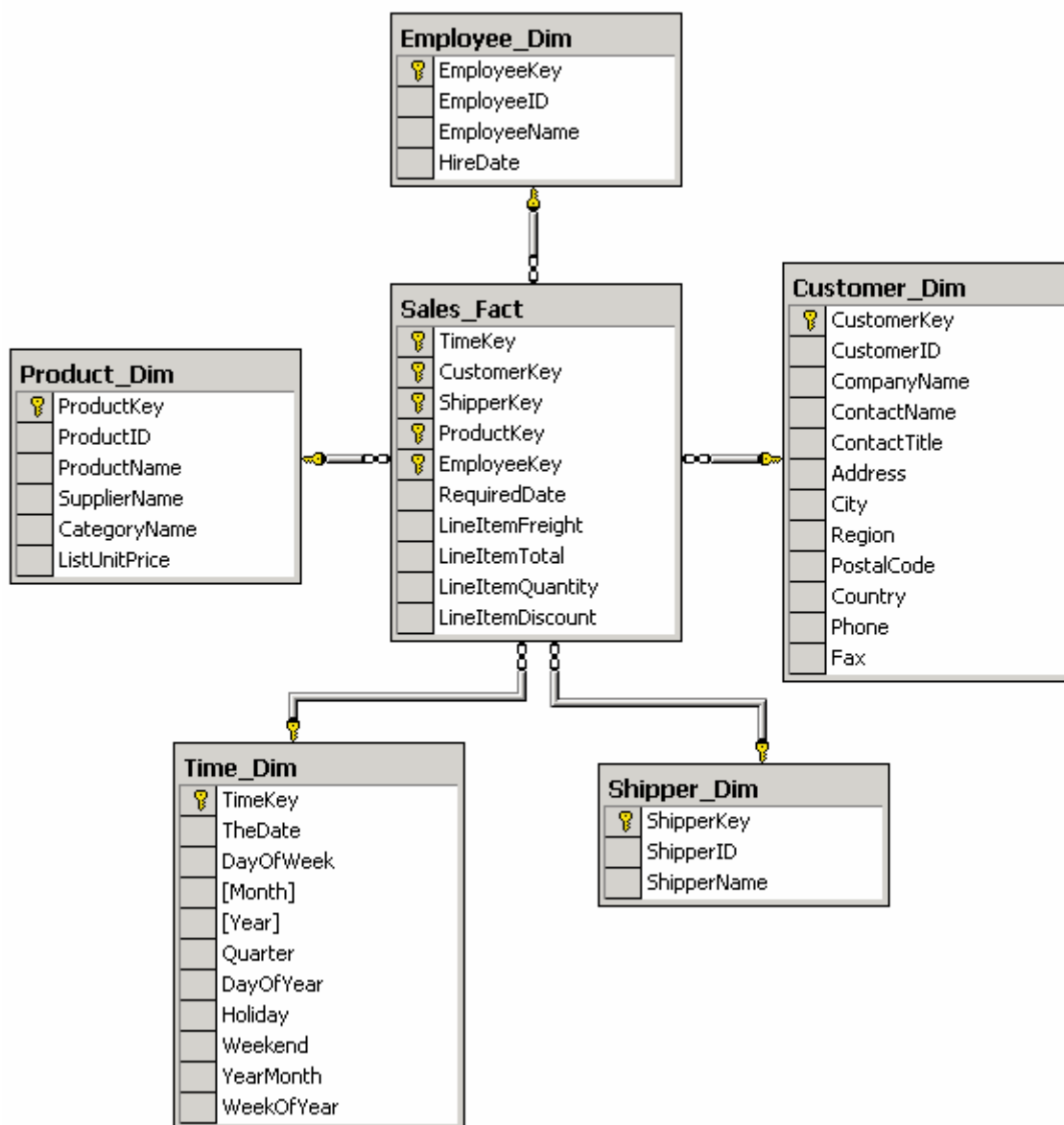


Рис.5.5. Система таблиц по схеме “звезда”

Развитием схемы "звезда" является схема "снежинка" (snowflake schema). Её отличает от первой схемы большое количество консольных таблиц, они имеются практически на каждой таблице размерности и могут иметь несколько уровней иерархии, как показано на рис. 5.1.б.

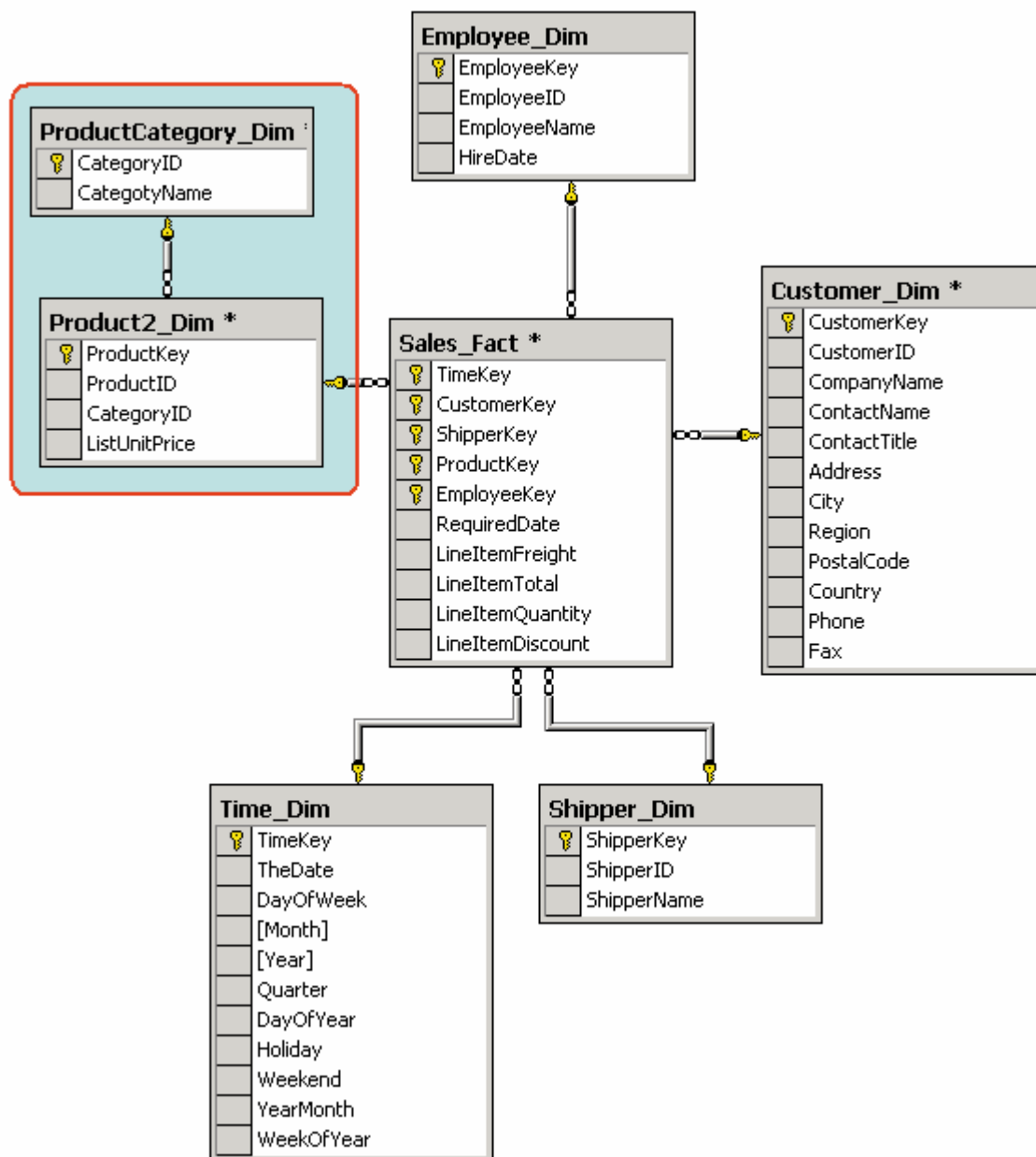


Рис. 5.6 Система таблиц по схеме "снежинка"

Схема "созвездие" (fact constellation schema) получается из нескольких таблиц фактов (Рис.5.1.6). В этом варианте многомерной модели через консольные или таблицы размерности сообщаются несколько таблиц фактов, отображающих несколько объектов с общими атрибутами.

В схемах "снежинка" и "созвездие" применение консольных таблиц приводит к дополнительным затратам времени на реализацию запроса. При проектировании этот фактор должен учитываться. При создании многомерных моделей на основе реляционной базы данных рекомендуют "создавать длинные и узкие таблицы фактов и сравнительно небольшие и широкие таблицы размерности (измерений).

Схемы реализации многомерных баз на основе реляционной СУБД в табличном виде приведены на рис. 5.3. – 5.6.

Многомерные модели данных на основе многомерных СУБД отличаются денормализацией, точнее отсутствием или неполнотой нормализации. Допускаются дублирование или избыточность данных. Ячейки гиперкубов, формируемые такими средствами, имеют одинаковую размерность, что также приводит к избыточному расходу ресурсов системы.

5.3 Типы многомерных OLAP-систем

В рамках OLAP-технологий на основе того, что многомерное представление данных может быть организовано как средствами реляционных СУБД, так многомерных специализированных средств, различают три типа многомерных OLAP-систем:

- многомерный (Multidimensional) OLAP- MOLAP
- реляционный (Relation) OLAP - ROLAP
- смешанный или гибридный (Hibrid) OLAP - HOLAP

Выше по существу изложены существо и различия между многомерной и реляционной моделью OLAP-систем. Сущность смешанной OLAP-системы заключается в возможности использования многомерного и реляционного подхода в зависимости от ситуации: размерности информационных массивов, их структуры, частоты обращений к тем или иным записям, вида запросов и т.д.

Рассмотрим подробнее достоинства и недостатки приведённых разновидностей OLAP-систем.

5.3.1 Многомерные OLAP-системы

В многомерных СУБД данные организованы не в виде реляционных таблиц, а упорядоченных многомерных массивов или гиперкубов, когда все хранимые данные должны иметь одинаковую размерность, что означает необходимость образовывать максимально полный базис измерений. Данные могут быть организованы в виде поликубов, в этом варианте значения каждого показателя хранятся с собственным набором измерений, обработка данных производится собственным инструментом системы.

Достоинствами MOLAP являются:

- более быстрое, чем при ROLAP получение ответов на запросы - затрачиваемое время на один-два порядка меньше;
- из-за ограничений SQL затрудняется реализация многих встроенных функций.

К ограничениям MOLAP относятся:

- сравнительно небольшие размеры баз данных - предел десятки Гигабайт;
- за счёт денормализации и предварительной агрегации многомерные массивы используют в 2,5-100 раз больше памяти, чем исходные данные;
- отсутствуют стандарты на интерфейс и средства манипулирования данными;
- имеются ограничения при загрузке данных.

5.3.2 Реляционные OLAP-системы

В настоящее время в массовых средствах, обеспечивающих аналитическую работу, преобладает использование инструментов на основе реляционного подхода.

Достоинствами ROLAP-систем являются:

- возможность оперативного анализа непосредственно содержащихся в хранилище данных, так как большинство исходных баз данных - реляционного типа;
- при переменной размерности задачи выигрывают ROLAP, так как не требуется физическая реорганизация базы данных;
- ROLAP-системы могут использовать менее мощные клиентские станции и серверы, причём на серверы ложится основная нагрузка по обработке сложных SQL-запросов;
- уровень защиты информации и разграничения прав доступа в реляционных СУБД несравненно выше, чем в многомерных.

Недостатком ROLAP-систем является меньшая производительность, необходимость тщательной проработки схем базы данных, специальная настройка индексов, анализ статистики запросов и учёт выводов анализа при доработках схем баз данных, что приводит к значительным дополнительным трудозатратам.

Выполнение же этих условий позволяет при использовании ROLAP-систем добиться схожих с MOLAP-системами показателей в отношении времени доступа и даже превзойти в экономии памяти.

5.3.3 Гибридные OLAP-системы

Представляют собой сочетание инструментов, реализующих реляционную и многомерную модель данных.

При таком подходе используются достоинства первых двух подходов и компенсируются их недостатки. В наиболее развитых программных продуктах такого назначения реализован именно этот принцип.

Использование гибридной архитектуры в OLAP-системах - это наиболее приемлемый путь решения проблем в применении программных инструментальных средств в многомерном анализе.

6. Концепции организации хранения данных

1. Понятие информационного хранилища.
2. Концепция централизованного хранилища данных
3. Концепция распределённого хранилища данных.
4. Концепция автономных витрин данных.
5. Концепция единого интегрированного хранилища и многих витрин данных.

Подготовка принятия решений требует сосредоточения значительного, а подчас колоссального количества информации (выше приведены возможные её объёмы в зависимости от масштабов и сложности решений) на месте его подготовки. Естественно стремление приблизить места хранения и использования информации. Проблемы подготовки принятия решений разрешаются с использованием инструментальных систем поддержания принятия решения Decision Support Sistem (DSS). В них большое место стали занимать OLAP-технологии. Проблема сбора и хранения информации выделилась как занимающая особое место во всей системе управления предприятием (корпорацией) и оформилась в концепцию информационных хранилищ (ИХ) - англоязычный термин Data Warehouse (DW).

6.1 Понятие информационного хранилища

ИХ выполняют задачи сбора информации из баз данных, отображающих отдельные бизнес-процессы, автоматизированных рабочих мест, информационных систем и других источников информации, в том числе из глобальных информационных сетей, как например Internet.

Simon (Саймон) (10) называет такие источники данных операционными базами данных. Сбор информации сочетается, как правило, с доработкой исходных данных, которая заключается в проверке достоверности, устранении противоречивости, сортировке, систематизации, построении заданной единой структуры хранилища и т.д.

Инмон (Inmon) - автор концепции информационных хранилищ выделяет следующие характерные для них свойства:

- предметную ориентированность;
- интегрированность;
- неизменчивость;
- поддержка хронологии.

Свойство предметной ориентированности означает компоновку пулов информации по определённым предметным областям или целям, обеспечивающим подготовку и принятие необходимых решений в соответствующей системе DSS.

Интегрированность предусматривает сбор и доработку (предварительную обработку) информации по определённой предметной области из различных источников и превращение её в организованный по заданным правилам, подчинённый определённой цели, массив в виде гиперкуба или системы поликубов информации.

Неизменчивость состоит в том, что информация не подвергается частым обновлениям, а только в случае крайней необходимости. В основном наращивается по заданному, чётко определённой графике. Пользователь имеет только право чтения информации.

Поддержка хронологии заключается в обязательности привязки данных ко времени. Информация, содержащаяся в хранилище, рассматривается в историческом аспекте. Используют следующие типы привязки ко времени:

- к моменту совершения события или факта;
- к моменту фиксации его информационными средствами;
- комбинированные методы, сочетающие оба подхода.

Реализация концепции ИХ может быть осуществлена несколькими способами - имеются несколько вариантов концепций ИХ.

6.2 Концепция централизованного хранилища данных

Такой подход означает, что при нескольких источниках информации - операционных базах данных создаётся единое централизованное хранилище. В первичных источниках информация хранится в "сыром" - недоработанном виде, то есть в структуре информационного пространства данного источника информации или операционной БД. Вся поступающая в ИХ информация должна быть преобразована в принятую в данном ИХ структуру. Передача данных из операционных БД в ИХ, которая сопровождается доработкой, может быть организована по заданному временному графику и правилам доработки с соблюдением принципов Инмона. Допускаются неожиданные запросы "на лету", что предъявляет более строгие требования к инструментальным средствам ИХ.



Рис.6.1 Схема централизованного хранения данных

При реализации такой концепции возникает потребность в мощном компьютере. В зависимости от масштабов предметной области это будет или персональный компьютер с предельно высокими характеристиками, особенно в части требований к объёмам памяти или майнфрейм и даже суперкомпьютер. Необходимо наличие развитых средств телекоммуникаций, обеспечивающих информационный обмен "операционные БД - ИХ". Это требование относится к любому варианту концепции ИХ.

6.3 Концепция распределённого хранилища данных

Возможен и имеет место противоположный подход к хранению данных на основе распределения функций ИХ по местам их возникновения или группировки нескольких операционных БД вокруг локального или регионального информационного хранилища. Эти хранилища могут быть ориентированы на определённую предметную область или на регион в корпоративных структурах. Система локальных хранилищ действует в качестве распределённого хранилища. Не исключается и наличие центрального хранилища, но в такой структуре требования к его размерности значительно облегчаются.

Такой подход предусматривает трансляцию каждого запроса к каждому источнику (базе данных), обработку, увязывание, согласование, компоновку извлечённых данных "на лету" и предоставление их пользователю.

Такой подход при экономии ресурсов на создание крупного централизованного хранилища имеет ряд недостатков, к которым можно отнести:

-в связи с нормализованностью данных в операционных базах и длительностью доступа из "центра" общее время отклика такой системы выходит за рамки допустимого;

-должны быть обеспечены постоянство нахождения в сети и открытость всех источников информации, так как отсутствие какого либо из них может сорвать весь процесс анализа;

-возможна противоречивость и несогласованность ответов из различных источников из-за различных форматов представления, разницы в темпах обновления, правил привязки ко времени, изменения смысловой нагрузки данных и т. д.;

-практическая невозможность комплексного исторического обзора содержащейся в разнородных источниках информации из-за различного порядка её хранения - навязать единый порядок весьма затруднительно.

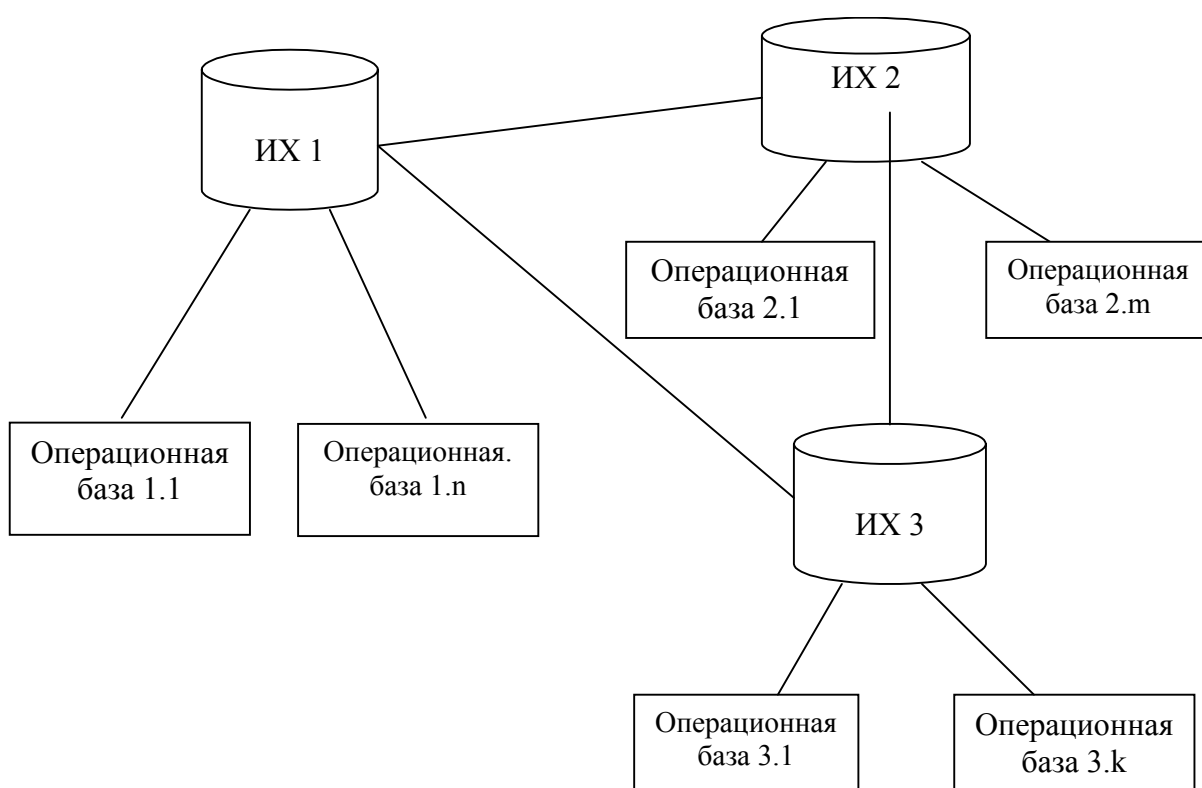


Рис. 6.2 Схема распределённого информационного хранилища.

6.4 Концепция автономных витрин данных

Одним из вариантов организации централизованного хранения и представления информации является концепция витрин данных (Data Mart). Она предложена Forrester Research в 1991 году. При таком подходе информация, относящаяся к крупной предметной области - например информационному пространству крупной корпоративной системы, имеющей несколько достаточно самостоятельных направлений деятельности, группируется по этим направлениям в специально организованных базах данных, которые называют витринами данных.

Этот подход является развитием концепции распределённого ИХ в части придания функций предметной ориентированности некоторым локальным ИХ.

Такой подход позволяет обойтись сравнительно менее ресурсоёмкими аппаратными и программными средствами,

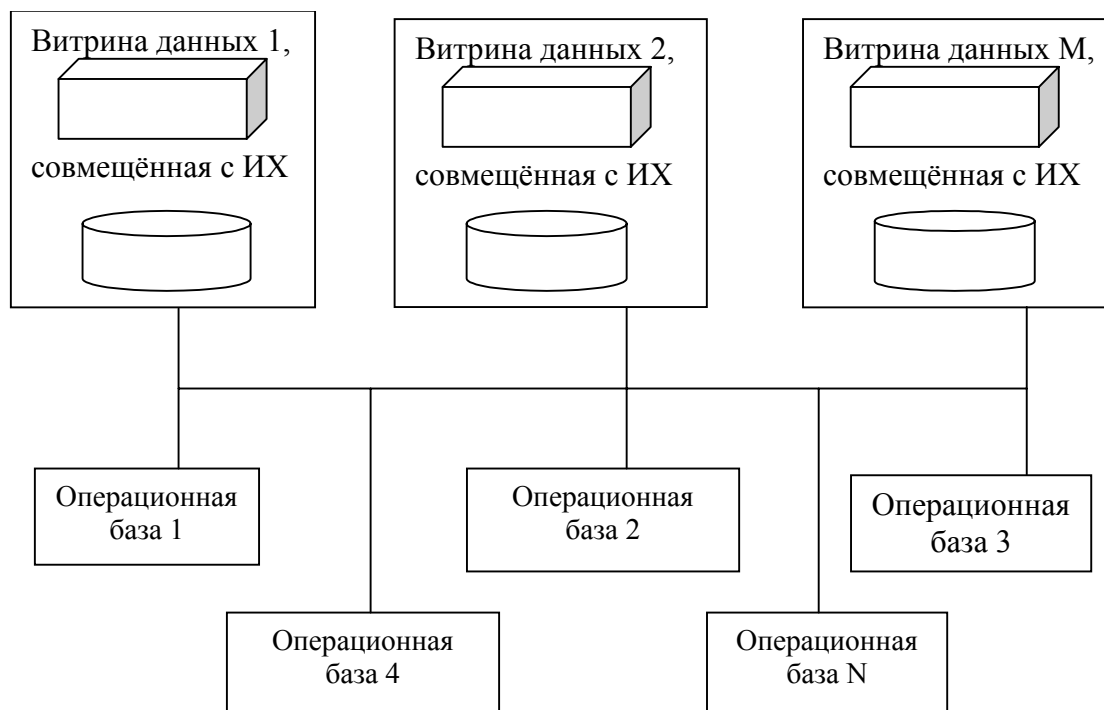


Рис.6.3 Схема автономных витрин данных.

обеспечивает повышение адаптируемости системы к изменяющимся условиям, расширяет доступность для внедрения. Пользователь предприятия или другого подразделения корпорации получает своё ИХ, обслуживающее местные потребности.

6.5 Концепция единого интегрированного хранилища и многих витрин данных

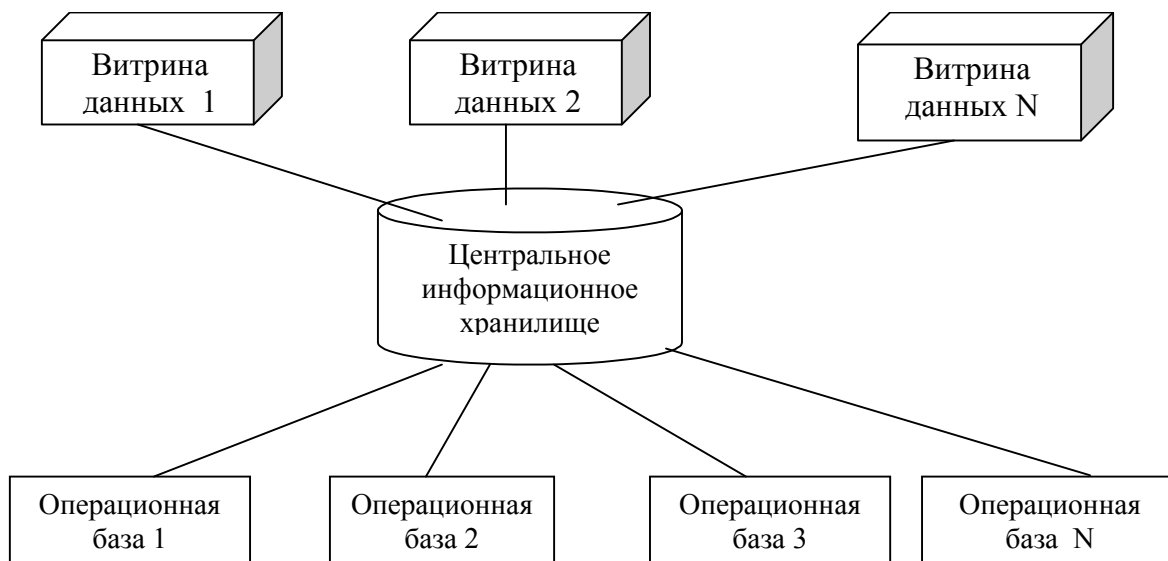


Рис. 6.4. Схема центрального информационного хранилища и многих витрин данных

В 1994 году М. Demarset предложил объединить две концепции: единого интегрированного хранилища и связанных с ним и получающих из него информацию витрин данных. В таком варианте имеется крупное информационное хранилище агрегированной и подработанной информации, которое может удовлетворить потенциальные запросы по отдельным направлениям деятельности.

Здесь очевидны преимущества: данные заранее агрегируются, обеспечивается единая хронология, согласованы различные форматы, устраняются противоречивость и неоднозначность данных - информация приобретает необходимую кондицию для быстрого и достаточного полного удовлетворения необходимого множества запросов.

Недостатком является необходимость применения высокопроизводительных аппаратных средств и специализированных многомерных или гибридных программных инструментальных средств.

В таком варианте ИАС приобретает иерархическую многоуровневую структуру, содержащую следующие уровни:

- общекорпоративное централизованное хранилище данных;
- витрины данных по направлениям деятельности;
- локальные или региональные базы и хранилища данных;
- операционные базы данных, автоматизированные рабочие места пользователей автономных программ и АЭИС.

Пунктам концентрации информации соответствуют иерархические уровни использования при подготовке, принятии и реализации решений

данных, которые являются появляющейся в результате функционирования предприятия (корпорации) информации:

- уровень лиц, принимающих решения, который может быть совмещён с уровнем витрин данных;
- уровень рабочих мест аналитиков и других заинтересованных пользователей.

Рассмотренные концепции охватывают лишь те стороны функционирования ИАС, которые относятся к организации хранения данных. Они не определяют требования и подходы к выполнению анализа, способы представления данных в ИХ - реляционный или многомерный.

7. Подходы к выполнению анализа средствами информационных технологий (it-анализа)

1. Типы многомерных OLAP-систем.
2. Задачи и содержание оперативного (OLAP) анализа.
3. Интеллектуальный анализ данных Data mining.

Информация, извлекаемая из информационных хранилищ и предоставляемая её конечным потребителям независимо от архитектуры ИХ, способов представления в базах данных, должна отвечать предъявляемым требованиям по форме представления, содержанию, своевременности, достоверности, воспринимаемости и т.д. Применяемые методы анализа должны обеспечивать необходимое содержание и достоверность предоставляемой пользователям информации.

7.1 Классификация IT-анализа по режиму и темпу

Различают два вида информационно-аналитических систем по режиму и темпу анализа:

-статические - имеют заранее разработанный сценарий обработки данных при весьма ограниченных возможностях вариаций запросов - так называемые информационные системы руководителя (Executive Information system EIS);

-динамические - обеспечивают обработку нерегламентированных запросов и гибкую систему подготовки отчётов.

Статические ИАС при всей кажущейся простоте и соответственно привлекательности для ЛПР имеют ограниченные возможности по информационной поддержке принятия решений. Зачастую полученная в отчётах информация порождает вытекающие из её содержания вопросы, на которые в допустимое время ответ не может быть получен.

Динамические ИАС предназначены для обработки незапланированных заранее, неожиданных (ad hoc) запросов.

Пользователи системы работают с такой ИАС в интерактивном режиме. Обрабатывается серия непредусмотренных заранее запросов, которые возникают в ходе полготовки и принятия решения. Заранее может быть подготовлена некоторая цепочка действий или сценарий, который может корректироваться.

Поддержка принятия управленческих решений осуществляется в следующих режимах или базовых сферах (28):

- сфера детализированных данных;
- сфера агрегированных показателей;
- сфера закономерностей.

В сфере детализированных данных подсистемы ИАС или автономные ИС нацелены на поиск данных. Эту задачу отлично выполняют реляционные СУБД. В качестве языка манипулирования данными, ставшего стандартным, используется, как правило, SQL. Для поиска детализированной информации используются информационно-поисковые системы, которые могут работать как с операционными, локальными или региональными базами и хранилищами данных, так и совместно с центральным ИХ.

Сфера агрегированных показателей отличается агрегацией данных, оперативной аналитической обработкой, многомерным представлением в виде гиперкубов, многомерным анализом. В этой сфере используются специальные многомерные СУБД. Допустимо использование реляционных представлений данных. При правильном применении реляционных СУБД, как показано выше, показатели эффективности ИАС сопоставимы со специализированными многомерными. Агрегированные массивы при реляционном подходе представлены в виде описанных выше схем: “звезды” и других. Агрегация может производиться также “на лету” при обработке запроса.

Анализ детализированных данных и агрегированных показателей относится к оперативному или OLAP-анализу.

Сфера закономерностей связана или основана на интеллектуальной обработке данных. Главной задачей здесь является выявление закономерностей в исследуемых процессах, взаимосвязей и взаимовлияния различных факторов, поиск крупных “непривычных” отклонений, прогноз хода различных существенных процессов. Эта сфера относится к интеллектуальному анализу (Data mining).

7.2 Задачи и содержание оперативного (OLAP) анализа

Оперативный анализ – это функция ИАС, обеспечивающая быстрый, в соответствии с правилами FASMI, доступ к любой необходимой информации, содержащейся в ИХ или, точнее в факт-таблице, представляемой также в виде многомерного куба (на практике

трёхмерного). Извлечение информации, как правило, сопровождается обработкой её по несложным алгоритмам, как то:

производится суммаризация, определение процентов от заданных величин, получение относительных показателей, вычисление величин с заданными коэффициентами и другие действия над данными на разных уровнях детализации. Анализ производится с данными, представленными в виде электронных таблиц, над которыми предоставляется возможность оперативно производить различные более сложные вычисления.

Примерами такого рода целей OLAP-анализа могут быть. Определение суммарных издержек на производство всей совокупности изделий предприятия в течение заданного периода, начиная с большого периода времени (года). Последующими этапами анализа могут быть – получение данных по этому показателю по каждому изделию за более короткие промежутки времени (полугодие, квартал, месяц) и т.д. Затем можно выявить наиболее затратные процессы, места их возникновения. Список задач можно продолжить. В сбытовой сфере, к примеру, можно изучать объёмы продаж, их динамику, привязку их к регионам, а также получать другие интересующие аналитика или ЛПР сведения.

Извлечение необходимой информации для построения отчётов производится путём использования ряда процедур.

К ним относятся:

- сечение или срез (slice and dice) - извлечение данных из факт-таблицы по каким-либо определённым значениям одного или нескольких измерений, например из гиперкуба (факт-таблицы), содержащей сведения об издержках, в отчёт (раздел отчёта) помещают данные только по какому-либо одному виду или группе издержек;
- поворот, под которым понимают изменение координат, их порядка или добавление измерений; эта процедура обеспечивает замену в готовом отчёте “Издержки”, к примеру, аргумента - время на регионы или центры затрат; если рассматривалась взаимозависимость “возраст - семейное положение” то можно в качестве аргумента брать любое из этих измерений и менять их местами;
- свёртка (drill up) - агрегируются данные по заданным признакам и алгоритмам; можно группировать необходимые данные, содержащиеся в ИХ в детальном виде, так при занесении сведений в операционную БД ежедневно в ИХ их можно передавать в агрегированном виде – еженедельно или ежемесячно, соответственно агрегированные данные можно помещать в отчёты;
- развёртка или раскрытие (roll up) – процедура, обратная свёртке, данные детализируются, например группы товаров представляются по конкретным товарам, более крупные временные периоды разбиваются на мелкие и т.д.
- создание кросс-таблиц - то-есть совмещение данных из разных таблиц по заданным признакам; например создаётся отчёт, в котором

- сводятся данные об издержках и выручке по одним и тем же изделиям и временным периодам;
- проекция - конструирование отчётов, являющихся подмножествами из множества единичных реквизитов или атрибутов, содержащихся в операционных базах или в ИХ; (См. также п.2.3.)
 - построение трендов – зависимость числовых или качественных значений показателя от тех или иных параметров, времени, технологии и т.д.

Инструменты OLAP –систем обеспечивают возможность сортировки и выборки данных по заданным условиям. Могут задаваться различные качественные и количественные условия.

7.3 Интеллектуальный анализ данных Data mining

Последующим этапом анализа является интеллектуальный анализ данных. В англоязычных источниках получивший название Data mining, иногда переводят как – добыча знаний.

Главным предназначением технологий интеллектуального анализа данных является извлечение и представление знаний из накопленной в базах данных, информационных хранилищах и других источниках информации.

При наличии большого количества определений понятия “Знания” здесь мы подразумеваем под ними зависимости и взаимосвязи, скрытые закономерности. Извлечение и представление знаний или познание скрытых связей и закономерностей в совокупностях данных различного объёма представляет собой проблему и для любых объёмов данных, когда связи между явлениями, процессами, фактами выражены неявно и неизвестны закономерности, согласно которым протекают процессы, происходят те или иные явления, события и факты.

Основными задачами интеллектуального анализа (2) 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.

Для 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 свою очередь многомерный статистический анализ делится на: факторный, дисперсионный, регрессионный, корреляционный, кластерный анализ. Эти методы позволяют решать многочисленные

задачи в области экономики, менеджмента, юриспруденции, которые являются составной частью аналитической подготовки принятия решений.

Помимо перечисленных выше методов, ставших традиционными, всё более широкое применение находят специфические методы интеллектуального анализа, происходящие из смежных областей информационных технологий (ИТ-систем) и получившие в них дальнейшее развитие, – интеллектуальные информационные системы. К ним относятся методы искусственного интеллекта и систем подготовки принятия решений (пересекаются с информационно-аналитическими системами).

К специфическим методам интеллектуального анализа относятся:

- методы нечёткой логики;
- классификационные и регрессионные деревья решений;
- нейронные сети;
- генетические алгоритмы;
- байесовское обучение и кластеризация.

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, о поведении объектов и систем из различных предметных областей, в том числе в социальной и экономической.

С помощью этих методов при отсутствии априорной информации об объектах и их поведении и значительной её 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 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.

Наиболее развитые ИАС, такие как SAS, Oracle обеспечивают представление добытых в результате анализа данных в виде моделей различного рода. Специальные процедуры и языковые средства дают возможность построения моделей автоматически на основе анализа имеющихся данных об исследуемых объектах.

Развитые средства data mining имеют графические средства анализа в виде графических конструкторов моделей, развитых средств отображения результатов, в том числе 3-D.

8. Программные инструментальные средства ИАС

1. Состав программных инструментальных средств ИАС.
2. Средства сбора и доработки данных.
3. Средства преобразования данных.
4. Средства оперативного (OLAP) анализа.
5. Средства интеллектуального анализа данных.

8.1 Состав программных инструментальных средств ИАС

В предыдущих разделах были рассмотрены основные функции информационно-аналитических систем, в том числе: информационного хранилища ИХ (DW), оперативного анализа (OLAP), интеллектуального анализа (DMg), представления пользователю результатов анализа и подготовки принятия решений (DMt). Для реализации функций применяется набор программных инструментальных средств. Полный набор этих средств включает ряд крупных модулей.

Крупные функциональные модули могут быть органической частью ИАС или самостоятельным программным продуктом. Во втором варианте они входят в ИАС в качестве комплектующего элемента. В свою очередь эти модули состоят из ряда блоков. Перечислим основные блоки:

- средства импорта, перекачки данных из операционных баз и других источников информации, взаимодействующие с различными операционными системами и СУБД ;
- средства преобразования данных, осуществляющие проверку на правильность, преобразование структур, агрегирование;
- набор или комплекс программ, которые выполняют операционные функции оперативного (OLAP) анализа; основу их составляет язык запросов Structured Query Language (SQL) усечённого или расширенного типа, в развитых ИАС в комплект входят специализированные языки различного уровня;

- средства графического и визуального конструирования отчётов, рассчитанные на конечного пользователя, как правило, дублируются языковыми средствами;
- средства удалённого доступа, обеспечения работы в распределённом и режиме "клиент-сервер", коллективного доступа и работы в глобальных сетях;
- средства администрирования ИАС;
- средства интеллектуального анализа данных, обычно называемые "Miner";
- приложения, разработанные встроенными в комплекс программ средствами.
- средства моделирования объектов и процессов.

Структура программных средств ИАС изображена на рис. 8.1.

Рассмотрим подробнее назначение и функции программных модулей, входящих в состав ИАС. Многие из них имеют самостоятельное значение и могут быть использованы для различных целей.

8.2 Средства сбора и доработки данных

Сбор данных из различных источников сопряжён с тем, что информация в них формируется в различных форматах, имеет разнообразную структуру. В информационном хранилище и в системах поддержки принятия решений или в витринах данных информация должна быть приведена к определённой в ИХ структуре и форматам витрин данных.

ИАС различного вида имеют в своём составе средства сбора данных из источников, созданных на различных аппаратных платформах: Intel, RISC, AS/400 и работающих во многих вариантах операционных сред, в том числе: в различных версиях Windows, Unix, AIX, Linux, OS-2 и т.д. Обеспечивается совместная работа со многими СУБД, как то: Access, Paradox, Fox Pro, MS SQL-Server, Oracle, DB-2, Informix, «Линтер» и т.д. Мощные ИАС насчитывают до 50-ти типов форматов, с которыми может взаимодействовать система.



Рис.8.1. Структура программных средств ИАС

Доступ к данным может быть реализован (организован) в двух вариантах.

В первом варианте производится непосредственное обращение из ИАС в базы данных. Этот процесс осуществляется с помощью драйверов из библиотек BDE (Borland Database Engine) и ODBC (Open Database Connectivity). Доступ к необходимым данным возможен после создания псевдонима данных – описателя, содержащего сведения, однозначно определяющие необходимую базу данных. Состав их зависит от типа СУБД и режима работы. При обращении к SQL-серверу достаточно указать наименование сервера и имя непосредственно базы данных. В случае работы с файловыми системами необходимо указать полный путь к искомому файлу. В системе ODBC следует задавать источник данных ODBS DSN. Этот способ выборки данных позволяет реализовать режим непосредственного обмена on line, когда каждое изменение в источнике мгновенно отражается в OLAP-системе. Однако реализовать этот режим не всегда возможно в силу того, что источник может быть выключен или закрыт для доступа, из-за неудовлетворительного состояния каналов связи и по другим причинам.

Предусматривается в виду изложенного и другой режим. Перегрузка информации из операционных баз данных в этом режиме производится по определённому согласованному распорядку – сценарию, который может быть сформирован с помощью заимствованных средств, например блока преобразования данных DTS

из состава MS SQL Server 7.0 или собственных средств разработчика OLAP- системы.

Операция перегрузки может производиться:

- непосредственно из баз данных с их форматами,
- из текстовых файлов, предварительно сформированных из таблиц операционных баз данных.

Все источники информации должны быть зарегистрированы в ИАС. Собранные из первичных источников данные компонуются в новую многомерную базу данных (МБД), которая имеет различные названия в конкретных реализациях ИАС. Например: Юнивёрс (Univers) в системе Business Objects, выборка в системе “Контур Стандарт”, гиперкуб - SAS, пул – SAP R-3, в некоторых источниках её называют факт-таблица (см. выше) и т.д. При построении МБД используются таблицы-справочники для доступа к информации, относящейся к различным классификаторам. Связи между объектами из нескольких источников данных определяются по ключевым полям.

Средства OLAP-систем предоставляют широкие возможности применения фильтров в процессе компоновки МБД.

8.3 Средства преобразования данных

Ввиду чрезвычайно высокой значимости информации, используемой для анализа, которая повышается после выполнения аналитических работ, обязательна процедура предварительной очистки данных. Производится обеспечение и проверка достоверности различными аппаратно-программными средствами и многими способами, в том числе:

-обратная проверка, контрольное суммирование, помехозащитное кодирование, семантический контроль и т. д.

Помимо этого производится логическая обработка данных в виде отбраковки нехарактерных для данного набора значений разного рода показателей. В этом случае используются средства статистической обработки данных и другие методы.

Необходимость преобразования связана с тем, что физическое представление данных в ИХ, как правило, сильно отличается от представлений в источниках. Ставится также задача эффективной реализации запросов и выполнения требования по времени отклика системы. Имеется потребность в унификации форматов представления данных.

Для этих целей используются языки обработки реляционных и многомерных данных, а также специальные процедуры. Языки могут быть заимствованными или фирменными.

При сборе (доставке), упаковке данных в ИХ возникает естественная потребность в агрегировании данных. По материалам исследований 90% пользователей используют в работе сильно агрегированную

информацию. Необходимость в использовании информации детального уровня возникает весьма редко. Для обеспечения эффективного использования ИХ ведётся наблюдение в процессе эксплуатации за частотой и характером запросов к хранимым данным. Выявляются закономерности, определяются наиболее частые типы запросов. На основании наблюдений заранее готовится агрегированная информация по соответствующим показателям. Такие меры резко снижают время отклика, соответственно повышается эффективность ИАС. Для реализации такого подхода применяются специальные процедуры.

8.4 Средства оперативного (OLAP) анализа

Задачи оперативного анализа изложены в разделе 7.2. Все ИАС, сертифицированные как OLAP-системы, имеют в своём составе соответствующие программные модули. Реализация их направлена на максимальное упрощение действий пользователя в процессе анализа. В процессе оперативного создания отчётов DSS - EIS затруднительно по времени использование языковых средств. По этой причине широко используются упрощающие такую работу и сокращающие трудозатраты средства графического конструирования. Основной принцип действия - сборка из элементов, представленных в графическом виде структур отчётов. Представленная общая структура базы данных в виде схемы с условно отображёнными атрибутами (реквизитами) и классами, являющаяся по существу заранее заготовленной витриной данных. Из неё простым перетаскиванием на отведённую часть экрана komponуется отчёт. Отпадает необходимость написания SQL-запроса. В некоторых инструментах (Business Objects) в процессе графического конструирования автоматически формируется соответствующий SQL-запрос. Конструирование отчётов может быть организовано и на основе электронных таблиц. Для реализации такого подхода созданы специальные программные модули, реализующие систему диалоговых меню, шаблоны, графические конструкторы. Сценарий анализа формируется с использованием технологии drag&drop (перенести и оставить), а также выбором из предложенных в диалоговом окне альтернатив, нажатием курсором соответствующих кнопок, определяющих те или иные режимы, настройкой иерархического дерева атрибутов и т.д. С помощью перечисленных методов и соответствующих средств выбирается нужная база данных, соответствующие поля и записи, определяются подлежащие анализу показатели, задаются режимы фильтрации, взаиморасположение функций и параметров (повороты, срезы и другие процедуры OLAP-анализа).

Результаты анализа представляются в виде напечатанных отчётов или презентаций, которые состоят из следующих компонентов: страница, таблица, график. Для окончательного оформления с учётом

эстетических и психофизиологических требований используют собственные средства ИАС или полученные результаты анализа в составе перечисленных компонентов переносят в другие среды, в которых могут быть добавлены различные дополнительные компоненты в виде рисунков, кино-, фото-, аудио-, видеоматериалов. Экспорт данных может быть осуществлён и в Web-среде. Возможен и обратный вариант: все перечисленные дополнительные компоненты импортируются в качестве внедряемых OLE-объектов.

При создании сложных сценариев OLAP-анализа использование мнемонических методов приводит к повышенным трудозатратам и может даже усложнить процесс в связи с их громоздкостью. В таких случаях используются программные методы. В качестве языков программирования применяют стандартный язык запросов SQL, расширенные или усечённые версии или специализированные фирменные языки. В продукте Business Objects используется стандартный SQL, в системе “Контур Стандарт” проблемно-ориентированный язык Python, в системе Oracle Express OLAP работы в таких случаях ведутся в профессиональной инструментальной среде для визуальной объектно-ориентированной разработки приложений Express Objects, в составе которой имеется язык Express Basic. Упомянутые средства программного метода подготовки OLAP-анализа позволяют выполнять анализ любой степени сложности с минимальными трудозатратами, но требуют достаточно высокой квалификации участников этого процесса.

8.5 Средства интеллектуального анализа данных

Это наиболее сложная, интеллектуально насыщенная часть информационно-аналитических систем, что служит причиной наличия соответствующего модуля лишь в составе наиболее развитых систем. К тому же состав задач, выполняемых модулями такого назначения, зависит от уровня развития системы.

Интеллектуальный анализ чаще реализуется автономными программными системами в связи со сложностью выполняемых задач.

В то же время OLAP-системы частично выполняют самые отработанные и легко реализуемые функции интеллектуального анализа.

К специализированным пакетам интеллектуального анализа относится высокоразвитый набор инструментальных средств ”для интеллектуального анализа баз данных “Polyanalyst” российской фирмы “Megaputer”. Всё семейство продуктов обеспечивает извлечение знаний в больших массивах данных, автоматически строит и тестирует формулы, описывающие обнаруженные функциональные зависимости; составляет классификационные правила по заданным примерам,

формирует многомерные кластеры, предлагает алгоритмы решений. Набор средств реализован в четырёх версиях:

Polyanalyst Lite, Polyanalyst Power – продукты для индивидуальных пользователей и малого бизнеса.

Polyanalyst Professional для MS Windows NT – мощная система Data Mining для профессионалов.

Polyanalyst Knowledge server – клиент-серверная версия продукта, объединена в Knowledge server. Предназначен для работы на высокопроизводительных платформах, включая SMP-машины, обеспечивает доступ к SQL-СУБД: Oracle, DB-2, Informix, MS SQL – Server и другими СУБД и OLAP-системами. В составе продукта имеется средство разработки собственных data mining-приложений и специализированных аналитических систем – Polyanalyst Development Toolkit, который имеет высокоуровневый API, библиотеку вызовов и по выбору все или отдельно каждый математический модуль.

Наиболее развитой системой в отношении возможностей, предоставляемых модулями и автономными подсистемами интеллектуального анализа, является ИАС фирмы США “SAS Institute Inc”. В комплект инструментальных средств, решающих разнообразные задачи интеллектуального анализа, входят:

- SAS/ETS – в этом модуле реализуются методы анализа временных рядов, экономического системного моделирования и прогнозирования, финансового анализа и формирования отчётов; производится восстановление пропущенных значений методом интерполяции, изменение временной привязки временного ряда, выделение сезонной компоненты во временных рядах, построение трендов, выявление флуктуаций и нерегулярной части;

- SAS/STAT – использует статистические методы регрессионного, дисперсионного анализа, нелинейного моделирования, анализа категориальных данных, многомерного, в том числе факторного анализа, кластерного и непараметрического анализа;

- SAS/INSIGHT – динамическое средство для исследования и анализа данных; использует методы статистического исследования одномерных и многомерных данных;

- SAS/IML – интерактивный матричный язык программирования, оперирующий с матрицами данных, которые могут быть числовыми и символьными;

- SAS/OR – инструмент моделирования, анализа, решения задач исследования операций, управления проектами;

- модули решения задач нейросетевыми методами.

В то же время высокого уровня инструмент Oracle Express OLAP не имеет в своём составе средств интеллектуального анализа. Решение этих задач предполагается выполнять средствами MS Excel, с которым имеется дружественный интерфейс. Необходимо заметить, что многие

ИАС возлагают на MS Excel задачи интеллектуального анализа, которые выполняются программными модулями “Мастера функций” этого популярного инструмента.

Среди массовых инструментов необходимо выделить пакет “Stadia-6.2” российской фирмы InCo, реализующий все методики статистического анализа, прогнозирование, сравнение возможных альтернатив выбора.

Пакет “Statistica” – с его помощью возможно решение всех задач статистического анализа, а также выявления закономерностей и прогнозирования нейросетевыми и другими современными методами.

Нейросетевое дополнение к пакету “Statistica” разработано российской фирмой “Статсофт”.

Широко распространённый пакет OLAP-анализа Business Objects имеет в своём составе модуль интеллектуального анализа “Miner”, выполняющий ряд задач этого класса, в том числе деревья решений, кластерного анализа. Однако основные функции интеллектуального анализа предусмотрено выполнять средствами MS Excel. В дальнейшем предусмотрена интеграция с пакетом Statistica.

Вышедший на рынок в 1999 году российский продукт “Контур Стандарт” из задач интеллектуального анализа выполняет лишь функции построения и анализа временных рядов (трендов) и кластерного анализа.

9. Управление и проектирование ИАС

1. Управление информационно-аналитическими системами.
2. Задачи и средства администрирования ИАС.
3. Принципы проектирования информационно-аналитических систем.
4. Рынок инструментальных средств ИАС.

9.1 Управление информационно-аналитическими системами

Управление ИАС осуществляется посредством использования метаданных (МД), то есть данных о том, что представляют собой и как хранятся данные и по существу содержат информацию о предметной области. Благодаря их использованию предоставляется информация, необходимая для автоматизации процессов сбора, преобразования и загрузки данных в ИХ для использования всей ИАС.

Проектирование ИХ, разработка сценариев OLAP и интеллектуального анализа начинается с формирования массивов метаданных, которые реализуются соответствующими программными модулями. Этими модулями описываются атрибуты метаданных, проверяется их корректность, создаётся физическая структура хранения,

обеспечивается вызов в репозиторий для реализации заданных в сценарии анализа функций.

Метаданные можно разделить на два типа – бизнес-метаданные и технические метадаанные, они позволяют ориентироваться в огромном объёме информации

Бизнес-метаданные служат в основном интересам пользователя. К ним относятся определения данных, величины атрибутов и областей, своевременность данных, бизнес-правила, соотношения данных, их охват.

Технические метадаанные используются службой администратора. К ним относятся сведения о времени обновления и преобразовании данных, права доступа и т.д.

В связи со сложностью потоков данных из операционных БД технические МД необходимы для работы многих программ.

Более подробно, в состав бизнес-метаданных входят:

- определённые объекты;
- иерархии;
- столбцы углубления;
- столбцы анализа;
- столбец фактических значений в прогнозных или бюджетных документах;
- столбцы бюджетных значений в прогнозных или бюджетных документах;
- временные измерения;
- значения, определяющие успех предприятия;
- категории и классификации данных;
- сведения о зависимых и независимых переменных;
- вид анализа и ограничений и т.д.

В состав технических метаданных входят:

- структура и семантика данных;
- алгоритмы агрегирования;
- сведения о разграничениях доступа и защите информации;
- периодичность загрузки и очистки;
- права собственности на информацию;
- поисковые таблицы и перечень и правила исключений;
- другие данные, относящиеся к техническому сопровождению

9.2 Задачи и средства администрирования ИАС

К задачам администрирования ИХ относятся:

- загрузка и обновление данных;
- планирование работы с ИХ;
- эксплуатация ИХ;
- создание (участие в создании) моделей предметной области, гиперкубов и интерфейсов пользователей;
- контроль за их наполнением и т.д.

Загрузка и обновление данных.

В ИХ состояние базы данных в отличие от OLTP меняется не постоянно, а по определённому графику перемещением из операционных БД статичных "кадров". Такой подход обеспечивает стабильность данных, их устойчивость от сиюминутных изменений.

Рассматриваются три стратегии загрузки:

- загрузка предварительно заархивированных данных;
- загрузка данных, содержащихся в существующих приложениях;
- пошаговое внесение изменений из среды OLTP, начиная с момента последней загрузки.

Применение первого способа обеспечивает быструю перегрузку, но необходимо использовать специальные программы для преобразования, при этом затруднительно масштабирование.

Загрузка из приложений во многом сходна с первым способом, при этом используется много различных технологий и файловых систем, необходимы значительные ресурсы. Выходом из положения является подработка данных внутри приложения и выдача их в согласованном формате.

Третий способ предусматривает загрузку только тех изменений, которые появились с момента последнего "Общения" операционной базы с ИХ. Применяются следующие приёмы перегрузки данных в рамках третьего подхода: замена всей таблицы из операционной БД, просмотр меток дат и времени в источниках информации, чтение операционных контрольных файлов, приём изменений на уровне СУБД, чтение регистрационных записей СУБД, сравнение состояний операционных БД "до" и "после".

Постепенно третий подход, то есть сбор изменённых данных становится единственным.

Планирование

Операции загрузки и обновления должны быть чётко спланированы, так как только такой подход обеспечит бесперебойность работы ИАС. В концепцию спроектированной ИАС или ИХ закладывается частота обновления данных после первоначальной загрузки. Периодичность таких акций тщательно планируется, для этого применяются мощные и гибкие инструменты. Эти средства могут работать в автоматическом режиме, то-есть сами запускаться в определённое автоматическим планировщиком время. Программные системы могут использовать средства планирования смежных систем.

Эксплуатация

Производится контроль состояния системы с помощью встроенных программных средств, ведётся анализ и отчётность о работе системы, содержащейся в ней информации. Производится настройка систем руководителя, поддержки принятия решений. Сгенерированные отчёты систематизируются, хранятся в специальном каталоге приложений, что экономит время и средства на создание новых отчётов. Эта работа в продвинутых ИАС максимально автоматизирована.

9.3 Принципы проектирования информационно-аналитических систем

Эта проблема актуальна не только для проектировщиков программных продуктов, администраторов ИАС, но и для пользователей, в особенности наиболее представительной по численности группы - аналитиков. В статье Computerworld 03.02 1995 года отмечалось, что "Data warehouse нельзя купить, его надо строить самим". В этом строительстве должны участвовать все - от руководителей до аналитиков, членов специальной группы проектировщиков ИХ-ИАС. Программные средства, которые можно купить, являются лишь инструментом для создания и поддержки ИХ-ИАС.

Одной из важнейших задач ИАС является преодоление кризиса оперативного анализа или как его называют в англоязычной литературе Data in Jail (DIJ) - "данные в тюрьме". Смысл в том, что при обилии исходных данных без квалифицированного использования их пользователь не в состоянии извлечь из них информацию и приобрести знания о процессах, происходящих в Вашем бизнесе или другой предметной области. Здесь следует напомнить, что данные и информация не одно и то же.

Для превращения данных в информацию необходимо соблюдение ряда условий:

1. Необходимо, чтобы принятое по выбору инструмента решение покрывало бизнес-потребности предприятия. Комплект программных продуктов должен обеспечивать выполнение полного спектра задач ИАС, а также обеспечивать техническую поддержку, обучение за приемлемую цену или возможности самообучения благодаря наличию добротной бумажной или электронной документации, другие сервисные услуги и возможности. Помимо инструментально-технологической основы комплект должен содержать готовые приложения, которые могли бы быть использованы немедленно, а также служить ориентиром для дальнейших разработок. Сюда могут быть включены аналитические задачи, поддерживающие поиск возможностей роста, обеспечивающие финансовую эффективность и т.д.

2. Должна обеспечиваться *интегрированность* - решение должно хорошо сочетаться с существующей платформой или средой. Необходимо реализовать гармоничное взаимодействие между всеми модулями системы на основе действующих стандартов в индустрии программных продуктов.

3. Необходимо выполнение условия *неограниченности* – то есть выбранная конфигурация средств должна быть адаптируема к изменениям, и расширению на увеличение числа пользователей и объёмов данных.

4. Свойство *гарантированности* - решение должно быть проверенным в смысле получения тех свойств, которые мы собирались обнаружить в данном продукте - получении бизнес-преимуществ и качества технологии. Поставщик продукта должен иметь прочные позиции на рынке, устойчивое финансовое состояние, хорошую клиентскую базу, большое число партнёров, пользующихся его технологиями.

При проектировании ИАС, под которым будем подразумевать выбор инструментальных средств, структурирование управленческих решений в рамках структурирования информационного пространства предприятия, подготовка витрин данных, исследование потребностей ЛПР в составе и других характеристиках информации, необходимой для принятия решений и т.д.

Основная задача администраторов ИАС - создание систем факт-таблиц с таблицами размерности и консольными таблицами. При этом должны соблюдаться требования потребителей информации. Должен быть обеспечен диалог аналитика и соблюдение требований к динамике процесса. Необходимо учитывать при проектировании “витрин данных” и других элементов ИАС психофизиологические требования и нормы к процессам анализа и обеспечения принятия решений.

Вопросы денормализации реляционных баз и нормализации многомерных баз также являются одними из основных при создании ИАС. Следует учитывать, что инструменты создания гиперкубов также допускают связи между таблицами “многие ко многим”.

Остальные принципы ложатся в рамки традиционного канонического и индустриального проектирования информационных систем. Они достаточно подробно рассмотрены в учебнике “Проектирование экономических информационных систем” в разделе “Проектирование систем оперативного анализа данных” (1).

9.4 Рынок инструментальных средств ИАС

На российском рынке инструментальных средств ИАС представлены в основном зарубежные продукты американских фирм. В последние годы стали появляться российские аналитические системы, но они носят как правило целевой характер, в основном в области

финансового анализа. Как показано выше (п.1), все средства анализа и ИХ можно поделить на три больших класса:

1. Средства массового применения, к которым относятся Excel, Statistica.
2. Встроенные в интегрированные ЭИС инструментальные средства создания и поддержки ИХ и анализа.
3. Специализированные инструментальные средства создания ИХ и обеспечения различного вида анализа.

Средства, относящиеся к первому классу.

Программный инструмент MS Excel, в особенности версия Excel 2000, имеет практически полный набор признаков, отвечающих требованиям, предъявляемым к ИАС:

- средства получения данных из операционных БД -ODBC и другие по желанию заказчика;
- развитые средства OLAP, статистического и финансового анализа;
- широкий набор средств оформления отчётов, базирующийся на средства MS Office, которыми широко пользуются другие пакеты.

Существенным ограничением является размерность рядов чисел, которая не может превышать 30 (тридцати). Тем самым сужаются рамки применения этого средства индивидуальным и малым бизнесом или малоразмерными массивами данных.

Пакет Statistica широко распространён, имеет широкие возможности по осуществлению интеллектуального анализа, интегрируется с другими средствами через инструмент OLE. Имеются намерения у фирмы-разработчика доведения продукта до требований Codd,a. Ограничений по объёмам выборок нет. Предназначен для профессиональной аналитической работы с неограниченными по объёмам массивами данных в различных предметных областях.

Встроенные средства OLAP и интеллектуального анализа

Крупные ИИС, такие как SAP R-3 и другие системы подобного класса имеют в своём составе полный набор средств создания и поддержки ИХ, OLAP и Data mining. Менее развитые системы, в особенности российские включают в свой состав инструментальные средства других фирм или имеют собственные разработки ограниченных масштабов и возможностей. Примером может служить разработка российской компании "ПиБи" "OLAP 7.7", ориентированная на работу с широко распространённой системой "1С: Предприятие 7.7".

Необходимо при этом учитывать уровень пользователя. Можно приобрести дорогой инструмент с массой возможностей, но из-за недостаточного профессионализма пользователей или из-за отсутствия постоянной необходимости в применении инструмента в полном объёме его возможностей он найдёт лишь ограниченное применение. В таком случае целесообразно пойти по пути привлечения консультационных фирм, которые выполняли бы такие работы по мере необходимости.

Примером встраивания заимствованных пакетов могут служить "Галактика", использующая инструменты Oracle, Никос-софт с продуктом NS-2000 и встроенным инструментарием анализа и ИХ канадской фирмы COGNOS.

Многие российские фирмы - производители ЭИС на интуитивном уровне встраивают в свои системы средства OLAP и Data mining, не осознавая, что придают своим продуктам такие свойства.

Специализированные инструментальные средства ИАС

Наиболее известными специализированными инструментами создания и поддержки ИАС являются продукты фирм:

- SAS Institute - комплекс программ, обеспечивающих проведение всех работ по созданию и поддержке ИХ, проведению всех видов анализа, имеет инструменты моделирования, имеется и собственная объектная СУБД;
- Oracle – наиболее полный набор программных средств, включая СУБД, CASE-средства и инструменты имитационного моделирования, но отсутствуют средства data mining. Программный комплекс ориентирован на мощные платформы в виде суперкомпьютеров, майнфреймов. Для персональных компьютеров имеются адаптации, но с ограниченными возможностями, которые подчас трудно на практике выделить.
- набор специализированных программных продуктов для создания и поддержки корпоративных ИХ и систем поддержки принятия решений компании *Microsoft*. Продукты Microsoft SQL Server 7.0 с подсистемами Microsoft Data Transformation services и Microsoft Decision Support services (DSS). Названные продукты обеспечивают

создание и поддержку ИХ, а также выполнение OLAP-анализа. Для осуществления Data mining привлекаются упомянутые уже продукты фирмы Cognos.. Широко привлекаются компоненты для создания клиентских приложений.

- к мощным системам относятся также продукты фирм Informix, Sybase, IBM, Hiperion;
- к продуктам среднего класса можно отнести Seagate Software, Act, Arbor Software, Эти пакеты выполняют в основном функции создания и поддержки ИХ, OLAP-анализа;
- продуктом массового применения, обеспечивающим выполнение названных для продуктов среднего класса функций является комплекс программ фирмы Business objects, включающий модули BO, BM, BQ. Главное достоинство его - возможность работы на платформах персональных компьютеров, в локальных сетях уровня Windows. Этот продукт наиболее приемлем для средних предприятий.

Фирма наметила комплексирование с пакетом Statistica, что резко расширяет границы его применения ввиду простоты применения и относительной ценовой доступности. С продуктами фирмы комплексированы такие известные ИИС как BAAN и другие, в том числе и российские.

Мощным конкурентом выступила российская фирма "Intersoft Lab", которая выпустила на рынок серию продуктов "Контур" в составе "Контур Стандарт", "Контур Профи", "Контур Корпорация". Названные продукты удовлетворяют всем требованиям к OLAP-системам и информационным хранилищам, имеют некоторые преимущества в части скорости доступа, удобства интерфейса. Различия между продуктами в масштабах обслуживаемых объектов и систем: малый, средний объекты, корпоративная система.

Необходимо отметить на рынке российскую фирму "Релэкс" г. Воронеж. Эта фирма предлагает весьма совершенные оригинальные инструменты "Линтер" - СУБД класса SQL Server 7.0 - Oracle, имеющий практически все достоинства названных систем и информационно-аналитическую систему "Невод", которая обеспечивает представление результатов интеллектуального анализа помимо традиционных представлений в виде графических структур выявленных связей и ассоциаций. Производится непрерывное развитие названных продуктов. Отличительной чертой является самая высокая степень защиты информации из всех средств, представленных на рынке и дешевизна продуктов.

В целом российские фирмы в настоящее время предлагают полный набор весьма совершенных продуктов, выполняющих функции инструментальных средств создания и ведения информационно-аналитических систем.

10. Литература

1. Смирнова Г.Н., Сорокин А.А., Тельнов Ю.Ф. Проектирование экономических информационных систем. Учебник. М. 2001.
2. Тельнов Ю.Ф. Интеллектуальные информационные системы в экономике. М. 1999.
3. Мишенин А.И. Теория экономических информационных систем. М. Финансы и статистика, 2000.
4. Королёв М.А., Мишенин М.И., Хотяшов Э.Н. Теория экономических информационных систем. М. 1984.
5. Жеребин В.М., Информационное обеспечение АСУ М. 1975.
6. Козлов В.А. Открытые информационные системы. М. 1999.
7. Дубров А.М., Мхитарян В.С. Многомерные статистические методы и основы эконометрики. М. 1998.
8. Ясин Е.Г. и др. Экономическая информация. М. 1974.
9. Дейт К. Дж. Введение в системы баз данных. М. СПб, К. 2000.
10. Саймон А. Р. Стратегические технологии баз данных. М. 1999.
11. Хан Д. ПиК Планирование и контроль: концепция контроллинга. М. 1997.
12. Данилочкина Н.Г. и др. Контроллинг как инструмент управления предприятием. М. 1999.
13. Карминский А.М. и др. Контроллинг в бизнесе. М.1998.
- 14.Макарова Н.В. и др. Информатика. Учебник. М.1998.
- 15.Боровиков В.П. Программа Statistica для студентов и инженеров. М.2001.
16. Андрейчиков А.В., Андрейчикова О.Н. Анализ, синтез, планирование решений в экономике. М. 2000.
- 17.Светлов Н.М. (руководитель) Российское программное обеспечение 2000/2001 М. 2000.
- 18.Плис А.И., Сливина Н.А. Mathcad: математический практикум для экономистов и инженеров. М. 1999.

- 19.Архипенков С.Я. Аналитические системы на базе Oracle Express OLAP. Проектирование, создание, сопровождение. М. 2000.
- 20.Бритов П.А., Липчинский Е.А. Практика построения хранилищ данных: Система SAS. Ж. Системы управления базами данных. 04-05/1998/
- 21.Business Objects Руководство пользователя. Материалы фирмы Терн 1996.
- 22.Кузьмичёва Т.С. Контур Стандарт. Технология работы с системой. Руководство пользователя. ISL М. 1999.
- 23.Ивантер А. Деловой барометр показывает “Ясно” ж. “Эксперт” №29 авг. 2001.
- 24.Корнев В.В., Гареев А.Ф., Васютин С.В., Райх В.В. Базы данных. Интеллектуальная обработка информации. М. 2000.
- 25.Маклаков С.В. BPwin ERwin. CASE-средства разработки информационных систем. М. 2000.
- 26.Под ред. Максимовой В.Ф. Макроэкономика. М. 2001.
- 27.Пендс Н. Перевод Абушаева Ш. Что относится к OLAP? OLAPReport.com Last updated on Jan 1999.
- 28.Щавелев Л.В. Способы аналитической обработки данных для поддержки принятия решений. Ж. СУБД 04-05/1998/
- 29.Белов В.С. Материалы лекций по курсу “Информационно-аналитические системы” Рукопись. М.2000.
- 30.Шрек М. Перевод Intersoft Lab Технология Хранилищ данных: последние 10 лет, несомненно были впечатляющими <http://www.iso.ru> 08/01.
- 31.Елманова Н., Фёдоров А. Введение в OLAP: Основы OLAP. Хранилища данных. КомпьютерПресс 5/2001.