Международный консорциум «Электронный университет»

Московский государственный университет экономики, статистики и информатики

Евразийский открытый институт

Белов В. С.

Информационно-аналитические системы

Основы проектирования и применения

Учебное пособие

Руководство по изучению курса

Практикум по курсу

Москва 2004

УДК 004.415
ББК 32.973.202
Б 435

Белов В.С. **ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИЕ СИСТЕМЫ.** ***Основы проектирования и применения****:* учебное пособие, руководство, практикум / Московский государственный университет экономики, статистики и информатики. – М., 2004. – 116 с.

ISBN 5-7764-0425-8

© Белов В.С., 2004
© Московский государственный университет экономики, статистики и информатики, 2004

ОГЛАВЛЕНИЕ

Введение **6**

УЧЕБНОЕ ПОСОБИЕ..... **7**

Глава 1. Базовые понятия информационно-аналитических систем
1.1. Роль и место анализа в процессе принятия решения. 8
1.2. Аспекты проблемы анализа и их реализация в программных продуктах 9

Глава 2. Информационное пространство как среда анализа
2.1. Понятие информационного пространства..... 12
2.2. Структура информационного пространства 12
2.3. Элементы структуры информационного пространства. Понятие показателя. 14
2.4. Пространственная интерпретация понятия показатель. 15
2.5. Содержание экономических показателей. 18
2.6. Системы показателей. 19

Глава 3. Технологии сбора и хранения данных – концепция информационных хранилищ
3.1. Технологии извлечения, преобразования и загрузки данных. 22
3.1.1. Понятие о гибкой архитектуре данных. 22
3.1.2. Сбор, повышение качества данных и приведение их в единую структуру. 25
3.2. Концепции организации хранения данных. 26
3.2.1. Концепция информационного хранилища. 27
3.2.2. Концепция централизованного хранилища данных..... 29
3.2.3. Концепция распределенного хранилища данных. 30
3.2.4. Концепция автономных витрин данных..... 31
3.2.5. Концепция единого интегрированного хранилища и многих витрин данных. 31
3.3. База метаданных информационного хранилища (репозиторий ИХ). **32**
3.3.1. Основные понятия о базе метаданных. 32
3.3.2. Классификация метаданных. 33
3.3.3. Модели метаданных по измерениям..... 35
3.3.3.1. Трехмерная система классификации и модель МД. 35
3.3.3.2. Размерностная модель МД информационного хранилища. 35
3.3.4. Содержание компонентов размерностной модели МД ИХ. 36
3.3.4.1. Метаданные, описывающие сущности. 36
3.3.4.2. Метаданные, относящиеся к размещению ресурсов. 37
3.3.4.3. Метаданные, описывающие временные аспекты ИХ. 36
3.3.4.4. Метаданные, относящиеся к пользователям и администраторам ИХ и ИАС. 37
3.3.4.5. Метаданные о движущих силах создания ИАС. 37
3.3.4.6. Действия, которые выполняются над данными. 37
3.4. Модели данных информационного хранилища. 38
3.4.1. Понятия модели данных информационного хранилища. 38

3.4.2. Элементы моделей данных информационного хранилища. 38
3.4.2.1. Содержание и назначение таблицы фактов. 38
3.4.2.2. Таблицы размерности (измерений), другие компоненты модели. 38
3.4.3. Схемы представления многомерных данных. 39

Глава 4. Признаки OLAP-систем, технологии оперативного и интеллектуального анализа данных
4.1. Подходы к выполнению анализа средствами информационных технологий (IT-анализа). 48
4.1.1. Классификация IT-анализа по режиму и темпу..... 48
4.1.2. Требования, предъявляемые к OLAP-системам. 49
4.1.3. Задачи и содержание оперативного (OLAP) анализа. 51
4.1.4. Типы многомерных OLAP-систем. 52
4.2. Интеллектуальный анализ данных Data mining. 54
4.2.1. Содержание понятия знания. Классификация видов знаний. 54
4.2.2. Задачи Data mining..... 56
4.2.3. Специфические методы и области применения data mining-а. 57

Глава 5. Содержание и методы анализа и прогнозирования бизнес-процессов (деятельности предприятий) как объект автоматизации
5.1. Содержание экономического анализа. 61
5.2. Классификация методов анализа. 61
5.3. Аннотация содержания методов анализа в экономической предметной области. 63
5.3.1. Методики проведения анализа в маркетинговой деятельности..... 63
5.3.2. Анализ обеспечения ресурсами..... 63
5.3.3. Анализ в области логистики. 64
5.3.4. Финансовый анализ. 64
5.3.5. Анализ инвестиций и инноваций. 65
5.3.6. Методы стратегического анализа..... 66
5.3.6.1. Анализ стратегической позиции предприятия. 66
5.3.6.2 Анализ ситуации по слабым сигналам и оценка рисков..... 66
5.3.6.3. Анализ отклонений. 66
5.3.6.4. Анализ полей бизнеса. 67
5.3.6.5. Бенчмаркинг. 67
5.3.7. Информационный обмен, связанный с аналитической работой..... 68

Глава 6. Основы создания и применения информационно-аналитических систем
6.1. Программные инструментальные средства ИАС. 71
6.1.1. Состав программных инструментальных средств ИАС. 71
6.1.2. Средства сбора и доработки данных. 71
6.1.3. Средства преобразования данных. 73
6.1.4. Средства оперативного (OLAP) анализа. 74
6.1.5. Средства интеллектуального анализа данных. 75
6.2. Управление и проектирование ИАС. 77
6.2.1. Управление информационно-аналитическими системами. 77
6.2.2. Задачи и средства администрирования ИАС. 78
6.2.3. Принципы проектирования информационных хранилищ ИАС. 79
6.2.4. Рынок инструментальных средств ИАС. 81

РУКОВОДСТВО ПО ИЗУЧЕНИЮ КУРСА 85

1. Сведения об авторе 86
2. Цели и задачи дисциплины, сфера профессионального применения. 86
3. Необходимый объем знаний для изучения дисциплины. 87
4. Перечень основных тем дисциплины. 87
   4.1. Тема 1. Базовые понятия информационно-аналитических систем. 87
   4.2. Тема 2. Информационное пространство и система экономических и других показателей как среда анализа. 88
   4.3. Тема 3. Технологии сбора и хранения данных – концепция информационных хранилищ. 90
   4.4. Тема 4. Признаки OLAP-систем, технологии оперативного и интеллектуального анализа данных. 92
   4.5. Тема 5. Содержание и методы анализ и прогнозирования бизнес-процессов (деятельности предприятий). 94
4.6. Тема 6. Основы создания и применения информационно-аналитических систем. 95
5. Итоговый контроль знаний по курсу. 97

ПРАКТИКУМ ПО КУРСУ 99

Введение 100

1. Краткое содержание Курса «Информационно-аналитические системы» 101
   1.1. Роль и место анализа в процессе принятия решений 101
   1.2. Задачи информационно-аналитических систем. 101
      1.2.1. Сбор данных, необходимых для принятия решений 101
      1.2.2. Информационные хранилища (ИХ), правила построения и функционирования. 101
   1.3. Выполнение анализа. 102
      1.3.1. Требования к оперативному анализу. 102
      1.3.2. Интеллектуальный анализ. 102
   1.4. Представление результатов анализа. 103
2. Создание и применение ИАС. 103
3. Задания по лабораторной работе. 105
   3.1. Задание 1. 105
   3.2. Задание 2. 105
   3.3. Задание 3. 105
4. Тематика лабораторных работ. 109

Литература 112

Введение

Настоящее издание «Информационно-аналитические системы. Основы проектирования и применения» является учебно-методическим комплексом по изучению дисциплин «Информационно-аналитические системы» для студентов, обучающихся по специальностям: «Менеджмент», «Маркетинг», «Антикризисное управление», «Мировая экономика», «Финансы и кредит»; «Проектирование информационно-аналитических систем» для студентов, обучающихся по специальности «Информационные системы в экономике».

Целью комплекса является информационная поддержка на основе современных информационных технологий процесса приобретения знаний и навыков по указанным дисциплинам.

В итоге освоения дисциплин на основе использования данного комплекса обучающиеся получат знания по проблемам создания и применения информационно-аналитических систем в интересах повышения эффективности принимаемых решений, приобретут навыки организации сбора, поддержания качества и обеспечения хранения данных, выполнения аналитических работ с помощью современных информационных технологий.

Издание состоит из разделов: учебное пособие, руководство по изучению курса, практикум.

Учебное пособие разделено на шесть глав.

Первая глава освещает роль и место разбора в процессе принятия решений, в ней рассматриваются аспекты проблемы анализа, базовые понятия информационно-аналитических систем, их виды, состав.

Вторая глава содержит основные сведения об информационном пространстве как среде анализа, описание его структуры в целом и её элементов. Раскрыто содержание понятия показатель, приведена его пространственная интерпретация, показана необходимость применения классификации и кодирования показателей в процессе подготовки данных к анализу. Рассмотрены системы экономических показателей.

В третьей главе рассматривается концепция информационных хранилищ. Раскрыто содержание понятия о гибкой архитектуре данных, описаны процессы сбора, повышения их качества и приведения в единую структуру; требования к структуре и правила построения информационных хранилищ на основе моделей многомерных данных. Приведён анализ концепций создания структур распределённых хранилищ.

Четвёртая глава включает сведения о признаках OLAP-систем, методологии анализа с использованием информационных технологий – ИТ(IT)-анализе. Рассмотрены его классификация по режиму и темпу, требования к OLAP-системам, содержание оперативного анализа. Дан обзор видов знаний, описаны задачи и методы интеллектуального анализа (Data mining).

В пятой главе, посвящённой рассмотрению процессов анализа в качестве объектов автоматизации, приведены классификация методов анализа в экономической предметной области, дан их обзор, описан информационный обмен, сопутствующий анализу.

Шестая глава включает анализ требований к инструментальным средствам создания и применения ИАС и обзор рынка этих средств, принципы и методики управления и проектирования ИАС,

«Руководство по изучению курса» включает содержание программу курса, методические рекомендации, ссылки на разделы учебного пособия и другие источники.

В «Практикуме» приведён краткий обзор основных понятий информационно-аналитических систем, даны установки на разработку модели ИАС и описана методика приобретения навыков аналитической работы, приведены требования к оформлению отчёта о практической работе.

Автор выражает признательность д.э.н. профессору Тельнову Ю.Ф. за ценные советы при разработке комплекса, компании «ТЕРН» за предоставление инструментального средства «Business Objects».

Учебное пособие

ГЛАВА 1.
Базовые понятия информационно-аналитических систем

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 подъема (2). 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 необходимы информация и знания, которые должны удовлетворять требованиям полноты, достоверности, своевременности (актуальности), полезности.

Основополагающую роль в подготовке принятия решений играет его обоснование по имеющейся у ЛПР информации. Ее, 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.

Объемы 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 серьезной, что появилось отдельное направление или вид информационных систем – *информационно-аналитические системы (ИАС)*, под которыми понимают комплекс аппаратных, программных средств, информационных ресурсов, методик, которые используются для обеспечения автоматизации аналитических работ в целях обоснования принятия управленческих решений и других возможных применений.

1.2. Аспекты проблемы анализа и их реализация в программных продуктах

Вся проблема аналитической подготовки принятия решений имеет следующие аспекты:

– извлечение из многих источников разнородных данных, представленных в различных форматах и приведение их к единому формату и единой структуре;

– организация хранения и предоставления пользователям необходимой для принятия решений информации;

– собственно анализ, в том числе оперативный и интеллектуальный, и подготовка плановой или регулярной оценки состояния управляемого объекта в документов или экранных форм;

– подготовка результатов оперативного и интеллектуального анализа для эффективного их восприятия потребителями и принятия на ее основе адекватных решений.

Аспект, касающийся сбора и хранения информации с сопутствующей доработкой, оформился в концепцию информационных хранилищ (Data Warehouse). Эта концепция состоит в том, что сведения о деятельности предприятия или иного объекта хозяйственной или иной деятельности накапливаются в течение длительного периода времени (годы) в информационном хранилище по определенным правилам. Они используются в различных временных режимах для анализа, как источник данных для разного рода отчетности и работы с партнерами (Reporting) и обоснования управленческих решений.

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

Интеллектуальный анализ информации – имеет также широко распространенное в русской специальной литературе англоязычное название Data mining. Предназначен для фундаментального исследования проблем в той или иной предметной области. Требования по времени менее жестки, но используются более сложные методики. Ставятся, как правило, задачи и получают результаты стратегического значения. При решении сложных задач в режиме Data mining 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. Соответственно 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 инструментальных средств создания и поддержки OLAP-систем, информационных хранилищ (DWH), СППР (DSS), интеллектуального анализа Data mining (DMg), который получил обобщенное название – Business intelligence (BI), которому пока не подобран русскоязычный термин.

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

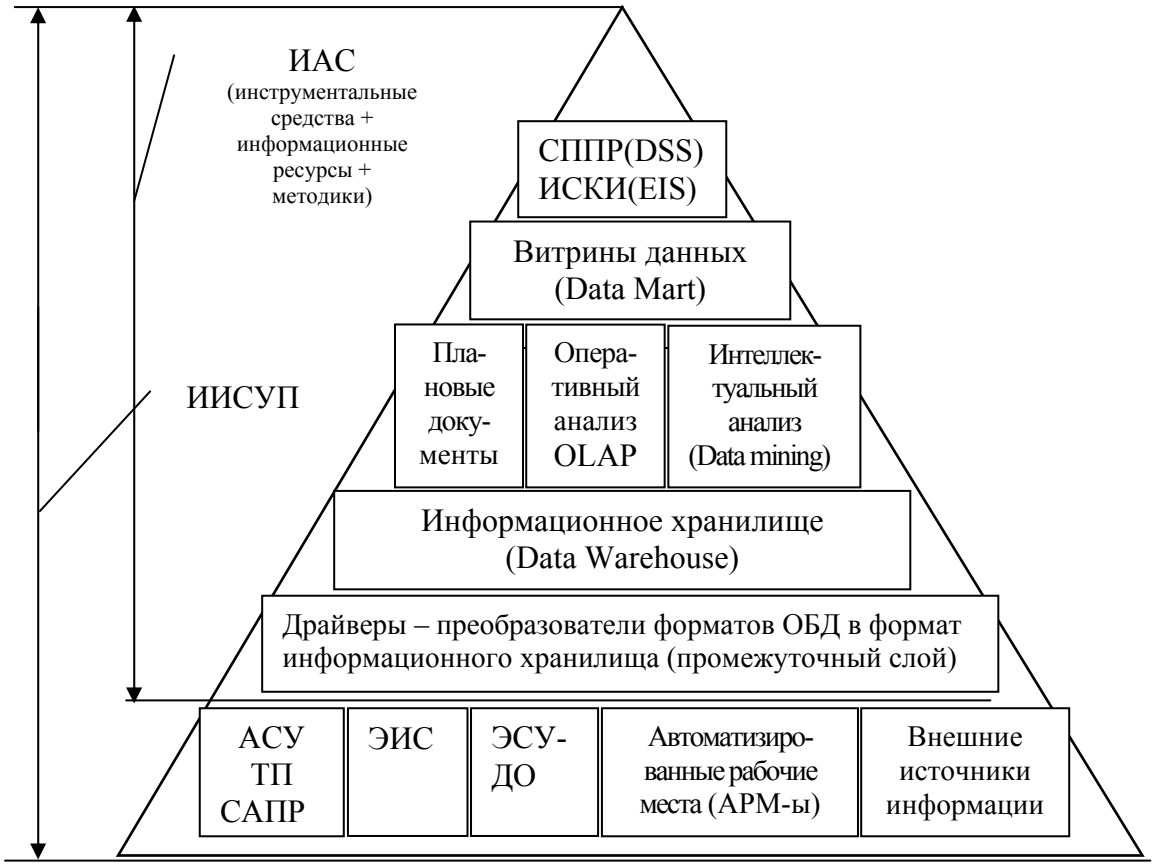

Рис. 1.1. Функциональный состав и место ИАС в обеспечении предприятия IT-технологиями

Функциональный состав и место ИАС в системе применяемых на предприятии информационных технологий отражен на рис. 1.1. Следует заметить, что ИАС играет объединяющую роль, консолидирует разрозненные IT-технологии в единую интегрированную информационную систему управления предприятием (корпорацией), как ее называют ИИСУП (26).

АСУ ТП – автоматизированные системы управления технологическими процессами.

САПР – системы автоматизированного проектирования.

ЭСУДО – электронные системы управления документооборотом.

ИИСУП – интегрированные системы управления предприятием.

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 разработанных в России пакетов программ такого назначения, способных по своим характеристикам успешно конкурировать с зарубежными, а по ряду параметров и превосходящие их. Главные преимущества российских продуктов по показателям цена/качество, отсутствие проблем локализации и др.

Вопросы для самопроверки

1. Какое значение имеет аналитическая работа для успеха предприятия?
2. Какие факторы влияют на деятельность предприятия?
3. Какие информационные ресурсы используются для подготовки принятия решений?
4. Что является предпосылкой для принятия правильных решений?
5. Какие объемы данных используются в процессе анализа?
6. Назовите требования к информации, которая используется для принятия решений.
7. Что такое информационно-аналитическая система?
8. Что вызвало появление и широкое распространение информационно-аналитических систем?
9. Назовите аспекты проблемы анализа в процессе подготовки принятия решений?
10. В чем заключаются аспекты сбора и хранения информации?
11. В чем состоит содержание аспектов анализа данных и предоставления результатов анализа пользователям?
12. Какие типы инструментальных средств для реализации информационно-аналитических систем вы можете назвать?
13. Какие информационные технологии и информационные системы на предприятии и из внешней среды являются источником данных для сосредоточения в информационном хранилище или непосредственно для анализа?
14. В каких видах информационных систем используются результаты анализа?

ГЛАВА 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 в среде – с одной стороны и, с другой стороны, производятся операции с элементами структуры этого пространства и сопряженных с ним информационных пространств.

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

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. (16).

Совокупность информационных объектов, информационно отображающих свойства системы и протекающие в ней процессы, называют информационным пространством. (4) Оно состоит из различных квантов или массивов данных в виде разного рода письменных (знаковых) и фиксированных на носителях информации кодограмм, буквенно-цифровых на естественном языке, устных и визуальных сообщений. Все виды сообщений предоставляются непосредственно потребителям информации или передаются по каналам связи, могут быть сохранены в различном виде с помощью современных технических средств и по мере необходимости воспроизводиться. Подавляющая часть сообщений, как правило, «загружается» в информационную систему и становится основой ее информационного обеспечения. Сочетание знаковых, звуковых (аудио) и образных видимых, в том числе с анимацией, сообщений (визуальных) называют мультимедийными сообщениями. Такого рода сообщения получили в последние годы широкое распространение в связи с высокой эффективностью восприятия и широким внедрением поддерживающих такие сообщения средств на практике и в состав информационных систем соответственно.

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

2.2. Структура информационного пространства

Характерным свойством информационного пространства является его структурированность. Это означает, выделены его элементы, установлены связи между ними, введены обозначения, элементы и связи упорядочены. Свойство структурированности в раз-

ных видах информационных пространств может быть выражено в разной степени. Высокий уровень обеспечивает возможность представления информации в виде документов и манипулирования данными с помощью программно-технических средств информационных систем.

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

- неструктурированное ИП (НИП);
- слабо структурированное (ССИП);
- структурированное (СИП);
- 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 строги 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 языке или по предельно упрощенным правилам.

2.3. Элементы структуры информационного пространства. Понятие показателя

В качестве элементов структуры информационного пространства выступают единицы информации. Это понятие, которое рассматривается в теории экономических информационных систем (ЭИС), выражает сущностное или смысловое наполнение элемента ИП. Под единицей информации (3) понимают «набор символов, которому придается определенный смысл». Рассматривается система единиц информации, которая имеет довольно сложную иерархическую структуру. Выделяют несколько уровней единиц информации в зависимости от смыслового (семантического) значения, его наполненности.

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

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

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

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

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

Реквизиты составляют две группы:

- реквизиты-признаки, выражающие качественные отличия показателя, его смысловое содержание, в частности экономическое;
- реквизиты-основания, содержащие количественные значения показателя.

Показатель теряет смысл без 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 (9): $P \Rightarrow \langle R, x \rangle$,

где:

P – показатель (может быть экономическим);

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

x – количественное значение показателя .

Идентификатор в свою очередь можно представить в виде двух частей:

$R \Rightarrow \langle S, Q \rangle$,где:

S – составленное из реквизитов наименование показателя, выявляющее его предметный смысл;

Q – дополнительные признаки показателя, составленные также из реквизитов и уточняющие его количественное значение.

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

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

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

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

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

У – признак стадии учета или, как сказано в (9) функции управления, то есть плановые, фактические, нормативные или какие либо другие значения показателя.

Представим эту структуру в виде соотношения:

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

таким образом $R \Rightarrow \langle S \langle E, C, B, У \rangle \rangle$

Наименование показателя может быть слитным (определенным одним реквизитом) или иметь свою структуру и в свою очередь состоять из реквизитов, таких как:

- Ф – формальная (вычисляемая) характеристика показателя, раскрывающая его структуру или алгоритм агрегации исходных детальных данных, например объем продаж, среднее, максимальное значение той или иной величины (подразумевается методика подсчета);
- П – обозначение отображаемого технологического или бизнес-процесса, например изготовление, реализация, перевозка и т.д.
- О – объект измерения, подсчета – виды товаров, оборудования, работники по категориям или общим числом.

Тогда $S \Rightarrow \langle Ф, П, О \rangle$

Таким образом общая структурная формула показателя примет вид:

$$P \Rightarrow < R, x >$$

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

$$S \Rightarrow \langle Ф, П, О \rangle,$$

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

$$R \Rightarrow \langle \langle Ф, П, О \rangle, \langle E, C, B, У \rangle \rangle,$$

$$P \Rightarrow \langle \langle Ф, П, О \rangle, \langle E, C, B, У \rangle, x \rangle$$

Данная структура, представленная в таблице 2.1, может отображать практически любой показатель.

Таблица 2.1

R							
S			Q				
Ф	П	О	Е	С	В	У	
P							
Ф	П	О	Е	С	В	У	x

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).

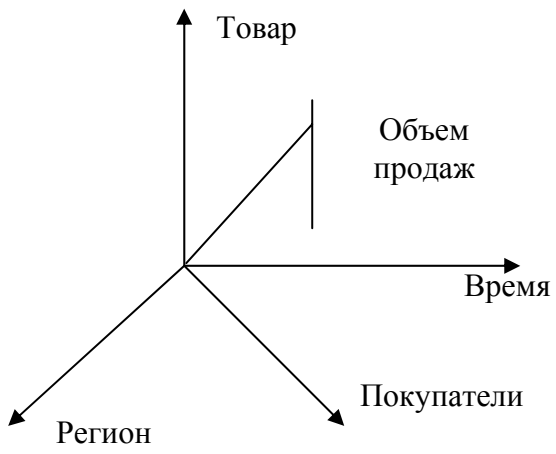

Рис. 2.1. Система координат многомерного информационного пространства показателя

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

Иллюстративные материалы специальной литературы и технической документации содержат упрощенные 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 приведен на рис. 2.1.

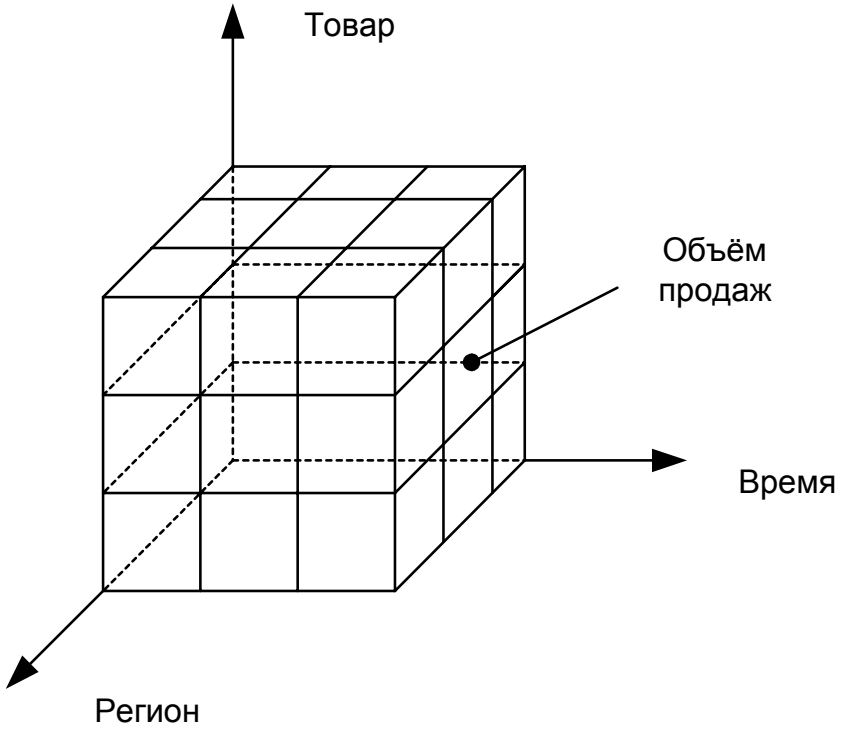

Рис. 2.2. Трехмерный гиперкуб информационного пространства показателя «Объем продаж»

Широко распространена также геометрическая интерпретация системы показателей в виде пирамиды, в основании которой лежат единичные признаки. Пирамида делится по высоте на «этажи», которые соответствуют иерархии признаков или системы показателей.

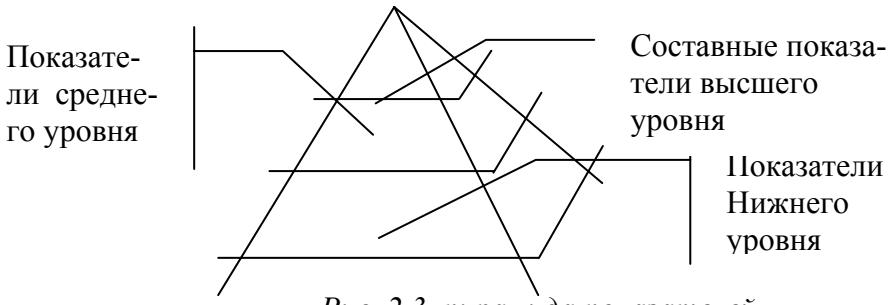

Рис. 2.3. пирамида показателей

Рис. 2.4. Диаграмма Ишикава

Соответствующие структуры представляют также в виде перевернутого дерева или «рыбьего скелета» – диаграммы Ишикава.

2.5. Содержание экономических 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 значениямиДля 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 ведется на предприятии, используются в оценках его состояния и динамики процессов, происходящих на предприятии, в информационных системах. Естественно, что при анализе их используют непосредственно и интерпретируют различным образом в целях извлечения знаний, формирования выводов и т.д.

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

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

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

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

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

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

Под классификацией понимают определение мест показателей в той или иной сфере деятельности, в частности экономической; среди объектов бизнеса, функций управления ими, их состояний, по единицам измерения, стабильности значения показателя, то есть упорядочение информации. Выражаясь формальным языком, «классификация преследует цель установления отношений между понятиями как отображениями объектов или групп объектов с общими свойствами, изучения структуры и упорядочения содержания данных». (9).

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

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

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

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

2.6. Системы 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 учета. Данные учета, 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 (15) – 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 определенным правилам все показатели нижних уровней.

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 учета, 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 $\frac{G_{IK}}{E}$ и скорости оборачиваемости капитала $\frac{E}{IK}$ и рассчитывается по формуле

$$ROI = \frac{G_{IK}}{IK} = \frac{G_{IK}}{E} \cdot \frac{E}{IK},$$

где:

G_{IK} – прибыль с капитала, рассчитываемая только для определенной, “работающей” на предприятии части имущества;

IK – сумма основных и оборотных средств.

У показателя ROI имеются недостатки (15), выражающиеся в том, что относительность составляющих показателя – реквизитов затрудняет выделение компонента, за счет 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 отмеченных 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 одноименной американской компанией. Pyramid Structure of Ratios разработана Британским институтом менеджмента для выполнения сравнительных аналитических работ по различным предприятиям.

Система ZWEI, разработанная в Германии, RL используют несколько обобщающих показателей, в том числе и абсолютных, к которым относятся Cach 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 ведется путем 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.

Вопросы для самопроверки

1. Дайте определение понятия информационного пространства.
2. В каких видах содержатся сведения в ИП и какие манипуляции совершаются над компонентами ИП?
3. Какое Вы знаете характерное свойство ИП и в чем оно состоит?
4. Какие единицы информации Вы знаете? В чем их содержание?
5. Дайте определение понятия показатель, исходя из формально-структурного подхода.
6. Каким образом строится пространственная интерпретация понятия показатель?
7. Какие виды геометрического представления показателей и их систем Вы знаете?
8. Что такое система оценок и в чем ее смысл?
9. Дайте определение показателя с точки зрения экономиста.
10. В чем содержание экономических показателей?
11. В чем заключается классификация показателей?
12. В чем состоит кодирование показателей?
13. Какие общегосударственные системы классификации и кодирования Вы знаете?
14. Как строятся системы показателей деятельности предприятия?
15. Что такое логико-дедуктивная система показателей?
16. Что такое эмпирико-индуктивная система показателей?

ГЛАВА 3.
Технологии сбора и хранения данных – концепция информационных хранилищ

3.1. Технологии извлечения, преобразования и загрузки данных

1. Понятие о гибкой архитектуре данных.
2. Сбор, повышение качества данных и приведение их в единую структуру.

3.1.1. Понятие о гибкой архитектуре данных

Объектом анализа являются данные, сконцентрированные в хранилище, а при необходимости и изымаемые непосредственно из первичных источников, которые должны быть структурированы в виде системы показателей исследуемой предметной области.

Прохождение данных из первичных источников как внутренних – предприятия или другого объекта управления, так и из внешних источников в информационное хранилище и далее к лицу(ам) или другим программам-приложениям, использующим их (данные) в процессе анализа и подготовки итоговых материалов различного назначения – в итоге для обоснования принятия решений, происходит в несколько этапов:

- Этап извлечения, преобразования и загрузки данных. На основе принятой системы показателей, характеризующих деятельность предприятия, подлежащих анализу и использованию в процессе принятия решений, и необходимых при разработке каких-либо документов, организуется сбор необходимых данных в хранилище и прорабатываются пути непосредственного извлечения в экстренных случаях необходимых детальных данных из первичных источников; этому этапу предшествует работа по созданию необходимой структуры перекачиваемых данных;
- Этап накопления, обеспечения готовности данных к использованию.
  По мере накопления в соответствующих зонах памяти выполняется периодическая загрузка данных из функциональных (транзакционных) подсистем интегрированной информационной системы (ИС) или автономных ИС, поддерживается необходимый уровень качества данных; в отдельных случаях допускается внеплановая загрузка по ситуации.
- Этап применения данных, содержащихся в хранилище и извлекаемых напрямую из первичных источников.

Для обеспечения процесса управления предприятием или другим объектом данные используются в трех основных режимах – создания плановых отчетных и других документов (Reporting), оперативного анализа в незапланированных ситуациях (OLAP-анализ), интеллектуального или углубленного анализа (Data mining). Накопленные в хранилище данные могут быть использованы в специальных программах, обеспечивающих развитие бизнеса.

В основе концепции информационных хранилищ заложена идея гибкой архитектуры данных. Это означает, что любому пользователю из числа доверенных лиц должна быть обеспечена возможность доступа к любому разрешенному для использования участку данных, которыми располагает предприятие (организация). Такой доступ осуществляется путем закладки в организационно-методологические основы построения системы сбора и хранения данных соответствующих возможностей, а также согласования принци-

пов построения и четкого взаимодействия аппаратного, программного комплексов и структуры накапливаемых и хранимых сведений.

Эта идея реализуется в основном в рамках свойств «открытых систем», определение которых POSIX 1003.0 принято Комитетом IEEE. В соответствии с этим определением открытая система есть «система, которая реализует открытые спецификации на интерфейсы, сервисы (услуги среды) и поддерживаемые форматы данных, достаточные для того, чтобы дать возможность должным образом разработанному прикладному программному обеспечению быть переносимым в широком диапазоне систем с минимальными изменениями, взаимодействовать с другими приложениями на локальных и удаленных системах, и взаимодействовать с пользователями в стиле, который облегчает переход пользователей от системы к системе».

Основные свойства открытых систем, очерченные этим определением, следующие:

- Расширяемость (extensibility) предполагает возможность включения новых или изменения некоторых прикладных функций ИС из числа уже реализованных, не изменяя при этом остальные функциональные подсистемы ИС.
- Масштабируемость (scalability) предусматривает применительно к прикладным программам и базам данных, реализуемым на разных прикладных платформах, возможность изменения их количественных характеристик (размерности решаемых задач, числа обслуживаемых пользователей и т.д.) путем настройки параметров, а не путем перепроектирования и программирования заново.
- Переносимость (portability) – это возможность перемещения ИС на другие аппаратно-программные платформы в случае их модернизации или замены с наименьшими затратами, сохраняя инвестиции, вложенные в разработку приложений, формирование массивов данных и обучение пользователей. Рассматривается переносимость приложений (application portability) и данных (data portability). Такая возможность обеспечивается соблюдением принятых стандартов обмена данными между приложениями и функциональной средой открытых систем. Определена «переносимость» пользователей (user portability), которая предусматривает возможность обеспечения стабильным дружественным пользовательским интерфейсом.
- Интероперабельность (interoperability) – свойство, обеспечивающее взаимодействие ИС с другими системами при обращении к информационным ресурсам (базам данных, базам знаний) этих систем или при решении определенных задач с использованием их вычислительных ресурсов, если собственные ресурсы недостаточны. Интероперабельность систем реализуется, прежде всего, форматами данных, принятыми в качестве стандартов электронного обмена данными (electronic data interchange -EDI) для разных прикладных областей. Интероперабельность систем при обращении запуске на исполнение к программам и данным, располагающихся в других системах, обеспечивается стандартами удаленного вызова процедур (remote procedure call – RPC).
- Способность к интеграции. При интеграции систем в целом (system integration) это свойство обеспечивает объединение нескольких ИС различного назначения в интегрированную многофункциональную ИС.

В случае интеграции баз данных (database integration) для прикладной программы или пользователя несколько баз данных представляются как одна логически единая база данных. При этом обеспечивается обращение пользователей к любой из этих баз независимо от ее места расположения в режиме коллективного доступа к данным, одновременная работа нескольких баз данных с каждой из прикладных программ ИС или пользователем.

При интеграции данных (data integration) обеспечивается возможность совместного использования запросом пользователя или программой одновременно нескольких файлов данных как единого целого. Рассматривается логическая интеграция, которая осуществляет объединение данных на логическом уровне, не затрагивая их физической организации, а также физическая, предполагающая слияние данных в единый информационный массив.

Высокая готовность (high availability) – свойство, означающее высокую отказоустойчивость (практически полную) системы (fault tolerance). В случае отказа какого-либо компонента обеспечивается гарантия автоматического восстановления работоспособности и сохранение целостности баз данных. Свойство готовности рассматривается и как мера способности системы принимать и успешно выполнять запросы и задания за доступный интервал времени.

На основе обзора свойств открытых систем можно сделать вывод об их определяющем значении при создании и эксплуатации информационно-аналитических систем. В соответствии с принципами открытых систем организуются и реализуются манипуляции с данными на всех рассмотренных выше этапах пути прохождения данных из источников к потребителям, а также подготовки на основе использования накапливаемых данных информации для тех или иных целей.

При организации работ по созданию, внедрению и применению ИАС важное значение имеет учет стандартов. Применение их облегчает организацию сбора, представление их в виде системы стандартных показателей в функциональных подсистемах ИАС, позволяет легко вписывать данные из первичных источников в структуру хранилища данных

Положения функциональной стандартизации в области информационных технологий определены стандартом ГОСТ Р ИСО/МЭК ТО 10000-99 «Информационная технология. Основы и таксономия международных функциональных стандартов».

Ориентация на стандарты при создании и ведении ИАС имеет два аспекта:

- использование их при создании архитектуры аппаратной и программной платформ ИАС (смотри приведенный выше стандарт);
- применение международных и российских стандартов в процессе классификации и кодирования систем показателей, что становится особенно актуальным при организации международного компьютерного обмена данными в процессе экономической или другой деятельности.

В настоящее время широко распространены системы электронного обмена данными (ЭОД), в англоязычной терминологии Electronic Data Interchange (EDI). Для создания системы компьютерного обмена данными, особенно в корпоративных распределенных структурах, тем более транснационального масштаба, в мировой практике широко применяются стандарты EDIFACT, которые были утверждены Международной организацией по стандартизации (International Organization for Standardization ISO) в 1988 году:

- ISO 7372-86 «Trade data interchange. Trade data elements directory. First edition. 1986-07-01» (Справочник элементов данных).
- ISO 9735-88 «EDI for administration, commerce and transport (EDIFACT). Syntax rules. 1988-07-15» (Синтаксические правила ЭДИФАКТ ООН).

В России появились два аналога приведенных выше стандартов:

- ГОСТ 6.20.2-91 «Элементы внешнеторговых данных (ISO 7372-86)
- ГОСТ 6.20.1.90 «Электронный обмен данными в управлении (ISO 9735-88), торговле и на транспорте (ЭДИФАКТ). Синтаксические правила.»

Введены эти стандарты 01.01.1992 года.

Стандарты ISO нашли дальнейшее развитие по отраслям деятельности, по технологиям обмена, в направлении применения Web –технологий.

3.1.2. Сбор, повышение качества данных и приведение их в единую структуру

Эти процессы, соответствующие этапу извлечения, преобразования и загрузки данных, в англоязычной терминологии называются Extraction, Transformation, Loading – ETL-процессы.

В процессе создания ИАС и ее центральной подсистемы – информационного хранилища как важная проблема выделяется обеспечение необходимого качества данных, в том числе достоверности, согласованности, соответствия установленным ограничениям и бизнес-правилам и т.д. Качество загружаемых и содержащихся в хранилище данных достигается системой организационных и программно-технических мероприятий. К ним относятся: "ручная" проверка разного рода ошибок, несоответствий, например неодинаковых названий полей с одним смыслом, и автоматическая с применением программных средств.

При сборе данных в информационное хранилище необходимо учитывать два основных аспекта: структурный и смысловой. Структурный аспект заключается в представлении данных из источников в тех или иных форматах программных сред, в которых были сформированы. Они должны приводиться к одному или группе форматов в системе сбора и хранения данных. Смысловой аспект состоит в содержательном наполнении знаковых структур данных. Даже при согласованных форматах данных могут быть разные толкования одинаково или близко по виду записанных данных и другие виды разночтений. Такие ситуации необходимо исключить еще на этапе формирования структуры ИХ.

ETL-процессы, которые реализуют требования по обеспечению качества, созданию необходимой структуры и поддержанию смысловых характеристик данных делятся на следующие стадии:

- Извлечение. На этой стадии производится перегрузка данных из источника, как правило, в промежуточную область хранилища. Под каждый источник в этой области создается своя таблица. Данные в источниках могут иметь различные форматы, в том числе текстовые неструктурированные, табличных процессоров, разного типа СУБД. Данные одного типа и структуры в первичных источниках сводят в одну таблицу, присваивая ей дополнительные поля.
- Структуризация. Ей подвергаются только неструктурированные данные. Они приводятся к пригодному для ввода в реляционные таблицы виду.
- Обработка. Изначально структурированные и подвергшиеся структуризации данные подвергаются обработке, которая заключается в очистке, фильтрации, согласовании данных.
- Пересылка и импорт данных. Современные СУБД обеспечивают возможности транспортировки данных внутри одного сервера, так и в распределенном режиме между серверами. Этот процесс требует тщательного квалифицированного администрирования. Необходимо обеспечивать защиту передачи данных по каналам связи. Может оказаться, что некоторые данные не могут быть вставлены в предназначенные им таблицы из-за ограничений или несогласованности типов данных. В таких случаях для них необходимо отводить отдельный участок памяти, где они сохраняются для дальнейшей оценки.

Отдельно и более подробно рассмотрим процессы повышения качества данных. Искажение данных может появляться на любом этапе, стадии, шаге ETL-процессов. Уп-

редить их возникновение полностью – задача практически невыполнимая, но необходимо принимать меры по снижению их количества и обнаружению. Рассматривают фатальные искажения, ошибки, к которым относятся: отсутствие данных в источнике, ошибка в подключении к источнику, проявляющаяся в отсутствии доступа к необходимым данным, проявление системного сбоя операционной системы.

Очистка данных состоит в исключении из общего потока тех данных, которые не отвечают заданным ограничениям на загрузку или бизнес-правилам. В этом случае данные из исходной таблицы разделяются на две части: отвечающие критериям качества, несоответствующие таковым и не попадающие на загрузку в хранилище.

Оценка качества данных производится:

- по критичности ошибок в данных (могут или не могут быть загружены) – ошибки в именах полей, типах данных;
- по правильности форматов и представлений данных;
- на соответствие данных ограничениям целостности;
- на уникальности внутренних и внешних ключей;
- по полноте данных и связей;

Приведенный перечень видов искажений, несоответствий в данных касается в основном структурного представления данных, технологических причин.

Смысловое содержание данных – знаковых структур очень важно при создании структуры хранилища и его поддержании, которое состоит в первоначальной загрузке и дальнейшем его заполнении и актуализации в процессе эксплуатации. В ходе ETL-процессов велика вероятность искажений, сбоев в работе по причине несогласованности в семантике. Такие ситуации называют «семантические разрывы» (46). В этой работе приводится четыре их вида:

- «Вавилонский» – это ситуация, когда одно и то же понятие или показатель обозначают разными именами.
- Кросспотоковый разрыв. В этом случае наименования разных по смысловому содержанию понятий одинаковые в знаковом представлении. Такое происходит при наличии в одной системе разных по смыслу данных, предоставляемых из разных источников. Например в одном источнике «Партия комплектов 50 штук» понимают комплекты узлов для компьютеров но этот комментарий опущен, в другом источнике это комплекты узлов для музыкальных центров. В связи со сложившейся привычкой комментарии также отсутствуют.
- Кроссязыковый разрыв происходит в тех случаях, когда передающее и принимающее звено не согласовали характеризующие признаки понятий – показателей. Например при оценке объемов поставок нефти передающий оценивает их в тоннах, а принимающий – в баррелях.
- Асинхронный разрыв связан с потерей или задержкой в передаче целой посылки в массиве или ее части, что нарушает целостность массива, его смысловое содержание или не дает возможности правильно понять и определить данные.

  В информационных хранилищах семантические разрывы в данных возможны на любом из переходов от этапа к этапу при прохождении данных от источников к конечным пользователям информацией и знаниями.

3.2. Концепции организации хранения данных

1. Концепция информационного хранилища.
2. Концепция централизованного хранилища данных

3. Концепция распределенного 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 ее объемы 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, интеллектуальный анализ, подготовка плановых и стандартных документов, которые используют информационные ресурсы, предоставляемые системами сбора и хранения данных.

Проблема сбора и хранения информационных ресурсов выделилась как занимающая особое место во всей системе управления предприятием (корпорацией) и оформилась в концепцию информационных хранилищ (ИХ) – англоязычный термин Data Warehouse (DW). Эта подсистема является центральной в интегрированной информационной системе, которой становится совокупность локальных информационных систем и информационных технологий, слитых воедино средствами ИАС.

3.2.1. Концепция информационного хранилища

Хранилища выполняют задачи накопления сведений о деятельности предприятия, партнеров, других информационных ресурсов из различных источников, в том числе 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) [11] называет такие источники данных операционными базами данных. Сбор перечисленных сведений сочетается, как правило, с доработкой исходных данных, которая заключается в проверке достоверности, устранении противоречивости, сортировке, систематизации в соответствии с заранее построенной на основе заданий пользователей единой структурой хранилища.

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 определенным предметным областям или целям, обеспечивающим подготовку и принятие соответствующих решений в соответствующей системе DSS или использования в каком-либо другом виде приложения, например для разработки планов развития предприятия, инвестиционных проектов, реинжиниринга и для любых работ, для выполнения которых необходимы исходные данные, накопленные в информационном хранилище. Необходимо подчеркнуть, что сосредотачиваемые в хранилище данные поддерживают аналитическую деятельность, а не функциональные или технологические процессы. Не непосредственную производственную деятельность, а анализ данных в историческом аспекте.

Интегрированность предусматривает сбор и доработку (предварительную обработку) информации по определенной предметной области из различных источников и превращение ее в организованный по заданным правилам, подчиненным определенной цели, массив в виде гиперкуба или системы поликубов информации. Пользователь легко использовать совместное использование данных из различных секций хранилища, относящихся к разным направлениям деятельности или функциям предприятия.

Неизменчивость состоит в том, что в хранилище данные из различных источников накапливаются, практически не подвергается каким-либо изменениям, только в случае крайней необходимости, в основном при обнаружении искажений. Они наращиваются по заданному, четко определенному графику. Пользователь имеет только право чтения информации. Манипуляции с содержимым хранилища доверяются только администраторам.

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.

Таким образом, идея хранилищ данных – это не просто единый подход к хранению необходимых данных, а создание единого многопрофильного в рамках одной концептуальной идеи информационного ресурса предприятия, направления исследований, корпоративной структуры и т.д.

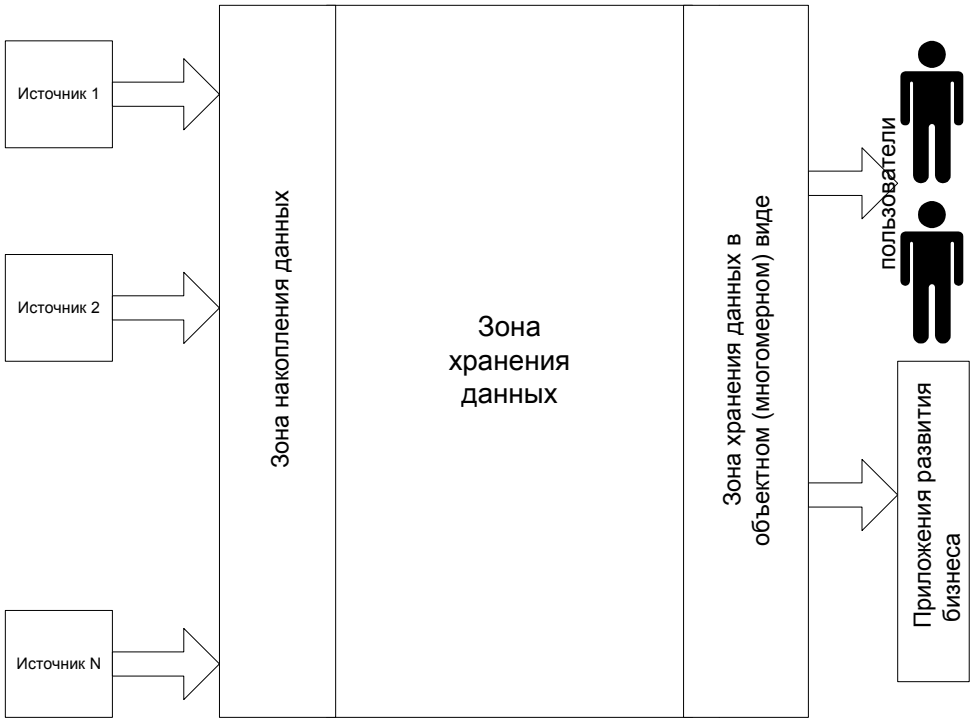

Рис. 3.1. Структура информационного хранилища

Данные в хранилище из источников накапливаются в течение определенного периода времени в зоне накопления. В течение этого времени производится работа по обеспечению необходимого качества данных по описанным выше правилам. В процессе перекачки из источников данные трансформируются в единый формат, проверяются их смыс-

ловая согласованность, наличие ошибок, проводятся мероприятия по повышению качества данных. При достижении необходимого уровня качества и определенного графиком работы момента времени данные переносятся в зону хранения.

В зоне хранения они могут быть представлены в виде реляционной или многомерной модели (объектном виде). В случае использования реляционной модели необходимо иметь в хранилище зону объектного представления данных для достижения уровня характеристик системы, соответствующего требованиям, предъявляемым к OLAP-системам. Содержание их будет рассмотрено в разделе 4.1.2.

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

3.2.2. Концепция централизованного хранилища данных

Такой подход означает, что при нескольких источниках информации – операционных базах данных создается 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 предъявлет более строгие требования к инструментальным средствам ИХ.

Рис. 3.2. 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 объемам 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.

3.2.3. Концепция распределенного 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 определенную предметную область или на регион в корпоративных структурах. Система локальных хранилищ действует в качестве распределенного 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 извлеченных 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 ее хранения – навязать единый порядок весьма затруднительно.

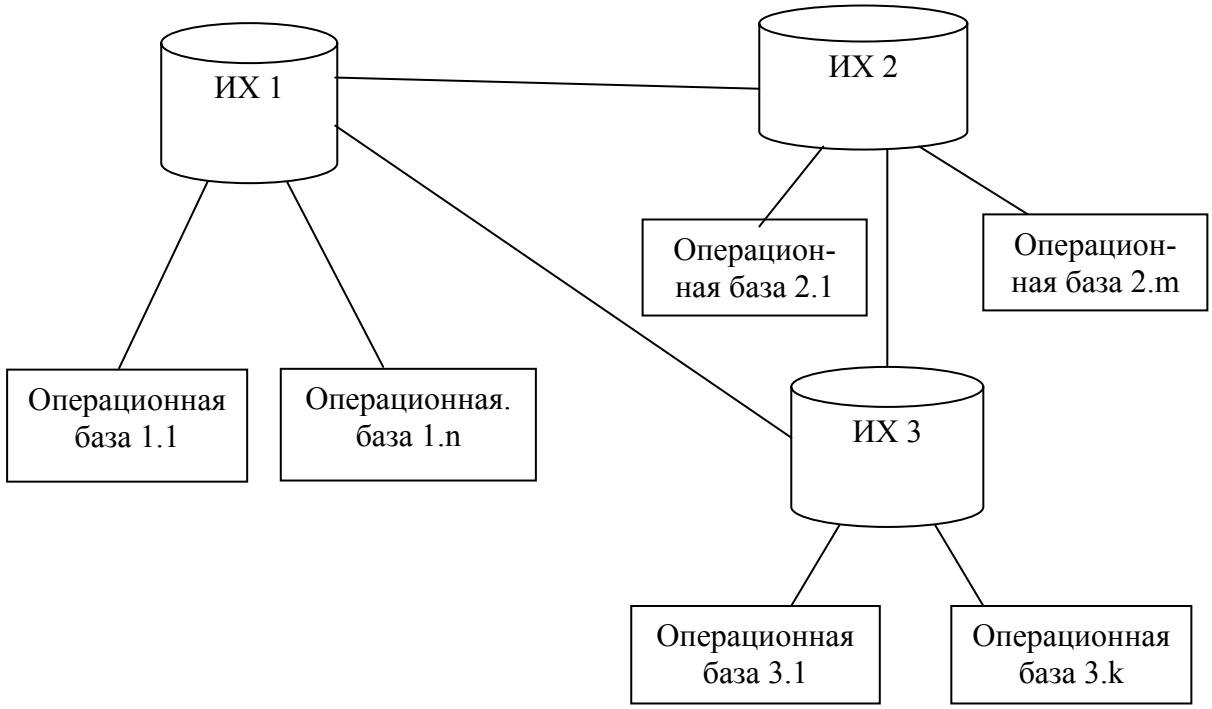

Рис. 3.3. Схема распределенного информационного хранилища

3.2.4. 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 распределенного ИХ в части придания функций предметной ориентированности некоторым локальным ИХ.

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

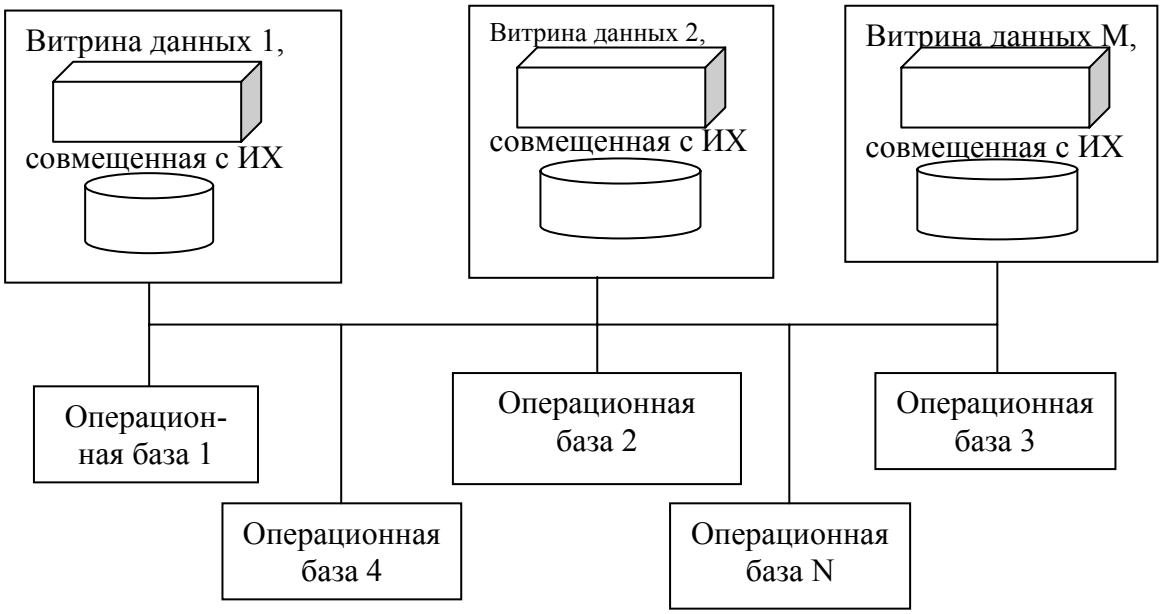

Рис. 3.4. Схема автономных витрин данных.

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

3.2.5. Концепция единого интегрированного хранилища и многих витрин данных

В 1994 году M. 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 достаточно полного удовлетворения необходимого множества запросов.

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

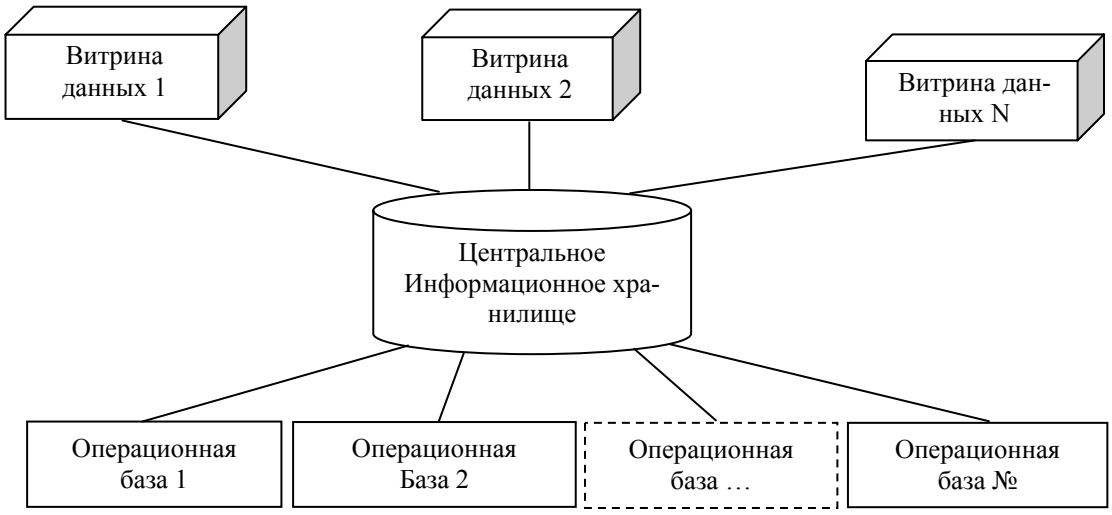

Рис. 3.5. Схема центрального информационного хранилища и многих витрин данных

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):

– уровень лиц, принимающих решения, который может быть совмещен 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.

3.3. База метаданных информационного хранилища (репозиторий ИХ)

1. Основные понятия о базе метаданных.
2. Классификация метаданных.
3. Модели метаданных по измерениям
4. Содержание компонентов размерностной модели МД ИХ.

3.3.1. Основные понятия о базе метаданных

Управление ИАС осуществляется посредством использования метаданных, 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 и 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, создается физическая структура хранения, обеспечивается вызов для реализации заданных в сценарии анализа функций.

Метаданные можно разделить на два типа – бизнес-метаданные и технические метаданные. Они позволяют ориентироваться в огромном объеме 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 метаданные 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:

- определенные 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, перечень и правила исключений;
- другие данные, относящиеся к техническому сопровождению.

3.3.2. Классификация метаданных

Под метаданными понимают также сведения в виде некоего справочника – репозитория, который помогает пользователям и лицам, обеспечивающим функционирование ИАС, ориентироваться в содержащихся в ИХ данных. Репозиторий является одним из центральных блоков инструментальной части ИАС.

Группа проектировщиков на основе рассмотрения требований пользователей, в которую входят их представители и работники ИТ-подразделения, составляет список подлежащих сосредоточению в ИХ показателей и их реквизитов. Эта работа выполняется на основе анализа документации, опросов, специальных исследований. Составляется перечень документов, аналитических сценариев с предполагаемыми запросами. Из полученных в результате обследования материалов выделяются сущности в разрезе достижения целей анализа. Их называют классы в терминологии объектно-ориентированного подхода, показатели в традиционной терминологии. Из всей совокупности материалов выделяют признаки (реквизиты, атрибуты), описывающие сущности (классы, показатели). На следующем этапе повторяющиеся по смыслу атрибуты представляют в виде одного согласованного по смыслу и синтаксису атрибута. В процессе слияния атрибутов в репозиторий появляются свои, присущие хранилищу атрибуты, их можно назвать системообразующими.

Метаданные должны содержать сведения о преобразовании исходных данных, истинные даты событий и временные характеристики отображаемых процессов, диапазоны для полей данных, сведения об источниках данных и их форматах и т. д.

В связи с обширностью сведений, содержащихся в МД, а также разнообразием вариантов реализации ИАС и ИХ возникла необходимость в более детальной классификации различных видов МД по нескольким признакам.

По динамике использования различают МД активные и пассивные. К активным относят те, которые контролируют функции или поведение приложений. Пассивные носят информационный характер и используются при просмотре содержимого ИХ.

По стадии применения различают проектные и рабочие метаданные

Проектные МД это совокупность атрибутов, содержащихся в исходных БД или файловых системах в составе их проектов, созданных с помощью CASE-средств. Воспринятые из этих источников МД и дополненные метаданными собственно хранилища должны уточняться в процессе обновления версий источников. Версии приложений для OLTP (функциональных систем) появляются реже, чем для систем поддержки принятия решений, поэтому проблема отслеживания проектных МД облегчается.

В шестимерной модели, которая далее будет рассмотрена, к проектным МД относятся измерения:

- МД для сущностей, описывающих содержимое ИХ,
- МД для расположения данных,
- МД для движущих сил развития ИХ.

Рабочие метаданные – это совокупность атрибутов, используемых в процессе эксплуатации системы (хранилища или ИАС в целом) Они делятся на активные и пассивные.

Активные рабочие МД имеют два источника: проектные метаданные и собственные компоненты метаданных ИХ. Этот тип МД используется для управления ИХ в частности и ИАС в целомна этапе эксплуатации системы. Активные рабочие метаданные используют для обеспечения безопасности и контроля за доступом к данным в ИХ, накопления статистики по обращениям к тем или иным разделам ИХ с целью совершенствования архитектуры ИАС. Информация, учитывающая время, может составлять значительный объем, к тому же она динамична. В размерностной (шестимерной) модели часть активных рабочих МД соответствует измерениям, описывающим операции над данными и людей, пользующихся хранилищем, а также временной аспект, то есть привязку данных и событий ко времени.

Пассивные рабочие МД привязаны к деловому аспекту, то есть целевому использованию ИХ. Пользователь, используя такие МД, приобретает возможность легче ориентироваться в огромных массивах данных, экономя время на доступ к ним, применять их для формирования запросов, для облегчения понимания смысла содержащихся в ИХ данных.

Пользователи должны знать источник содержащихся в ИХ данных, ответственных лиц, график загрузки. Эти сведения должны содержаться в репозитории МД. Помимо этого необходимо хранить данные о приложениях, языках программирования, датах переработки, о моделях данных, которые используются в CASE-средствах

3.3.3. Модели метаданных по измерениям

Под измерениями при рассмотрении совокупностей метаданных, понимают назначение и другие отличительные признаки входящих в эти совокупности групп МД. Различают (13) два основных вида классификации и соответствующих моделей: трехмерную и шестимерную.

3.3.3.1. Трехмерная система классификации и модель МД

Такая система подразумевает три типа метаданных:

– МД оперативных баз данных – источников данных для ИХ,

– МД об использовании данных, то есть об их назначении в интересах конечного пользователя,

– МД для управления информационным хранилищем.

Эта система классификации применяется для ИХ и ИАС малого масштаба с одной предметной областью или небольшим их количеством.

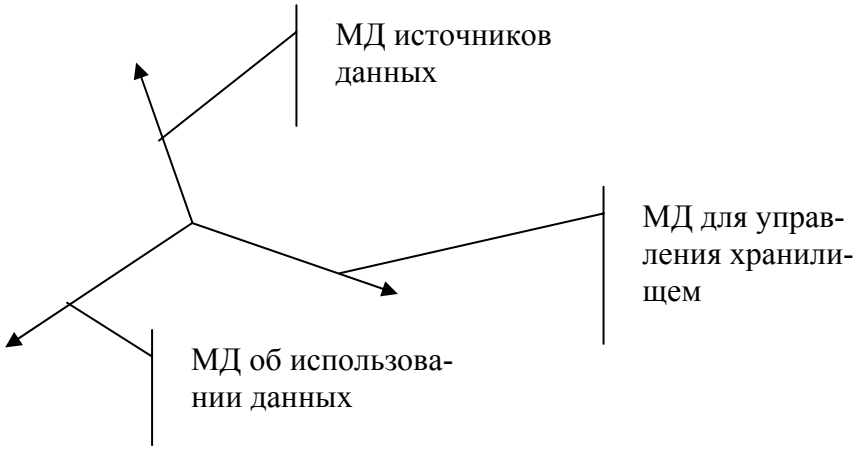

Рис. 3.6. Трехмерная модель метаданных.

3.3.3.2. Размерностная модель МД информационного хранилища

Эта модель обеспечивает более детальный и строгий учет данных, сосредоточенных в ИХ. Она отвечает на вопросы: что? где? когда? кто? почему? как?. Такая модель называется схемой Захмана.

Перечислим эти измерения:

1. Сущности, которые составляют содержимое ИХ – Что?
2. Размещение данных в операционных базах и непосредственно в ИХ – Где?

3. Моменты загрузки и агрегирования данных, время их происхождения – Когда?
4. Люди, которые используют и поддерживают хранилища – Кто?
5. Движущие силы создания и развития ИХ – Почему?
6. Действия, которые выполняются над данными – Как?

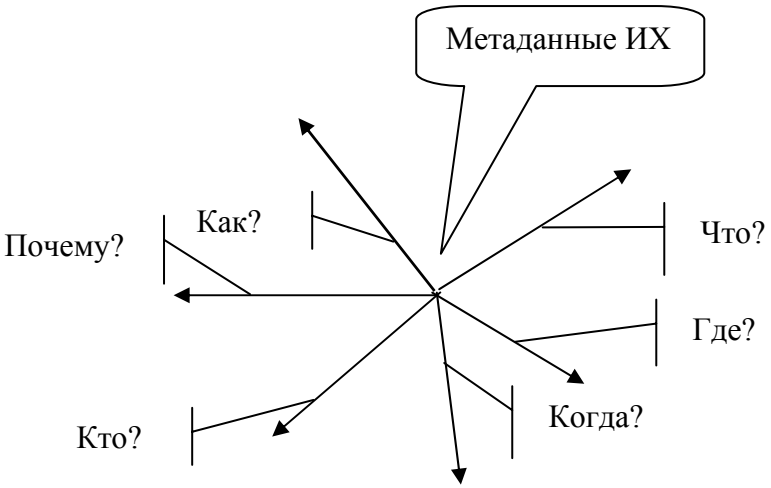

Рис. 3.7. Размерностная модель МД ИХ

3.3.4. Содержание компонентов размерностной модели МД ИХ

3.3.4.1. Метаданные, описывающие сущности

Это измерение является основой создания информационного хранилища. От правильности формирования содержимого хранилища зависит успех его использования. Должна быть документально определена предметная область или области и зафиксирована в метаданных. Если рассматривается несколько предметных областей, то в метаданных должны быть сведения о взаимодействии этих областей, а для одной предметной области о связях ее частей.

Сущности определяются на этапе моделирования. МД должны содержать имена полей, тип данных, домены включенных в ИХ данных, а также сведения о полях в источниках. Одна часть описаний полей может содержаться в источниках данных, другая – в хранилище, а также возможно их нахождение одновременно и в источниках и в хранилище.

Связи между сущностями определяются бизнес-правилами и они фиксируются в базе метаданных. При моделировании структуры метаданных ИХ целесообразно применение объектного подхода и соответствующих инструментальных средств.

3.3.4.2. Метаданные, относящиеся к размещению ресурсов

В базе метаданных должны быть сведения о размещении ресурсов ИХ и ИАС в целом. К таким ресурсам относятся: серверы, рабочие станции, сетевое оборудование, программные средства. Эти метаданные используются для управления конфигурацией, сопровождения системы, Они создаются путем загрузки файла конфигурации таких операционных систем как: UNIX, LINUX, Windows и других, а также путем выполнения инвентаризации технической и программной платформы.

3.3.4.3. Метаданные, описывающие временные аспекты ИХ

В хранилище отслеживаются:

– время происхождения событий в бизнес-сфере или другой предметной области;

– периоды времени, в течение которых происходят процессы;

– моменты появления сведений в источниках данных и в хранилище.

Имеются сущности из пользовательской сферы, которые акцентируют внимание на временных аспектах. К ним относятся: планы, графики работ, планируемая и фактическая их продолжительность.

В репозиторий вносятся МД, относящиеся к функционированию ИАС, в том числе: моменты обращения пользователей, длительность отклика на запросы, моменты начала и окончания загрузки данных в хранилище. По этим сведениям можно оценивать динамические характеристики ИАС.

3.3.4.4. Метаданные, относящиеся к пользователям и администраторам ИХ и ИАС

Функции, права доступа лиц, имеющих отношение к ИХ, фиксируются в соответствующих разделах базы МД. У названных лиц, независимо от названий их должностей в тех или иных организациях, имеются определенные права и обязанности, касающиеся работы в среде ИАС. Суть их работы определяется направлением деятельности: архитектор данных, специалист по информационным технологиям, руководитель проекта, администраторы баз данных первичных источников, бизнес-аналитик, менеджеры деловой сферы (лица, принимающие решения) и другие.

Все роли и лица их выполняющие должны быть зафиксированы в репозитории ИХ. Там же должна быть отражена организационная структура или ее часть, относящаяся к ИАС.

Защита данных в процессе создания ИАС должна осуществляться, но поэтапно, чтобы не слишком ограничивать действия исполнителей проекта. Полномасштабная корпоративная защита вводится по завершении проекта.

3.3.4.5. Метаданные о движущих силах создания ИАС

Под движущими силами авторы концепции размерностной модели метаданных ИХ понимают цели и задачи проекта ИАС, которые должны быть четко сформулированы и внесены в репозиторий. Необходимо дать описание выполнения задач бизнес-подразделений.

Целесообразно отобразить организационные и технические меры по обеспечению высокого качества данных. В состав МД вводят характеристики ИАС по возможностям смыслового содержания запросов, скорости ответа на них, которые определяют в ходе испытаний и опытной эксплуатации. Отражаются схемы и технологии взаимодействия с бизнес-пользователями.

3.3.4.6. Действия, которые выполняются над данными

В репозитории должны находиться МД о преобразованиях, которым должны подвергнуться перегружаемые из источников в хранилище данные. Должен быть разъяснен и унифицирован в хранилище их смысл. Необходимо отобразить те действия, которые совершаются над данными в хранилище заблаговременно с тем, чтобы не тратилось время на это при реализации запланированных запросов, а также возможности по совершению действий над данными при выполнении незапланированных запросов.

3.4. Модели данных информационного хранилища

1. Понятия модели данных информационного хранилища.
2. Элементы моделей данных информационного хранилища.
3. Схемы представления многомерных данных.

3.4.1. Понятия модели данных информационного хранилища

Многомерная модель данных представляет исследуемый объект в виде многомерной или объектно-ориентированной схемы данных, которая в геометрическом представлении представляет собой системы поликубов. Для зрительного восприятия используют совокупность фрагментарных трехмерных моделей. 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 трехмерных кубов с целью облегчения восприятия и квазиобъемного представления при формировании отчетных и аналитических документов и мультимедийных презентаций по материалам аналитических работ в системе поддержки принятия решений.

Многомерные данные могут быть отображены в моделях инструментами в виде СУБД на основе реляционных моделей данных, а также и специальными многомерными инструментальными средствами, называемыми объектными надстройками (32), многомерными и/или объектно-ориентированными СУБД.

3.4.2. Элементы моделей данных информационного хранилища

3.4.2.1. Содержание и назначение таблицы фактов

В многомерном пуле информации создается большая центральная таблица, называемая таблица факта (fact table). В ней помещаются все данные относительно интересующего пользователя обобщающего показателя, то есть объекта или события, которые интересуют пользователя. Таблицы фактов содержат числовые или качественные (содержательные) значения.

Рассматривают четыре вида фактов:

- транзакционные факты (transaction facts), отражающие происходящие в системе события, например финансовые и другие операции;
- «моментальные снимки» (Snapshot facts), фиксирующие состояния объекта в заданные моменты времени – наличие товаров на складах, состояния счетов в банке и так далее;
- элементы документов (Line-item facts), содержащие сведения о реквизитах документов – таких как количество отправленных, полученных товаров, ценах, дате и времени отправки;
- событие или состояние объекта (Event or state facts), отражающие акт совершения того или иного действия – отправку или прибытие товара, перечисление денежных сумм без описания их.

В таблицы фактов включаются, как правило, данные на уровне детализации, то есть самого нижнего уровня иерархии. Она имеет внутренний составной ключ из ключей окружающей ее меньших таблиц, которые содержат данные по признакам.

3.4.2.2. Таблицы размерности (измерений), другие компоненты модели

Таблицы, наполняющие факт-таблицы содержанием называют таблицы размерности или измерений (dimensional table).

Они содержат постоянные или редко и мало изменяемые данные и должны находиться в отношении «один ко многим» к таблице фактов. Таблицы размерности являются родительскими по отношению к таблице факта. Таблица факта является дочерней. В случае наличия в таблице измерений иерархии в ней должны быть поля, указывающие на «предков». Их называют еще консольные таблицы (outrigger table). Они присоединяются к таблицам размерности и детализируют отдельные атрибуты. Консольные таблицы являются родительскими по отношению к таблицам размерности.

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. Так, первичный ключ таблицы размерности является внешним ключом таблицы факта.

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

3.4.3. Схемы представления многомерных данных

Представление многомерных данных в рамках реляционных моделей может выполняться в виде трех вариантов схем:
– «звезда»;
– «снежинка»;
– «созвездие».
Линейное представление на плоскости отображено на рисунках 7.1–7.3

Схема «звезда» состоит только из таблиц размерности и таблицы факта. Она представлена на рис. 7.1 и 7.4.

Развитием схемы «звезда» является схема «снежинка» (snowflake schema). Ее 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. 7.2 и 7.5.

Многомерные модели данных на основе многомерных СУБД отличаются денормализацией, точнее отсутствием или неполнотой нормализации. Допускаются дублирование или избыточность данных, сбалансированные и несбалансированные (включающие неодинаковое количество уровней для разных ветвей) иерархии. Ячейки гиперкубов, формируемые такими средствами, имеют одинаковую размерность, что также приводит к избыточному расходу ресурсов системы.

Объектно-ориентированные средства разработки приложений имеют возможности создания OLAP-средств анализа в случаях необходимости разработки оригинальных приложений.

Факт-таблицы служат непосредственным материалом для осуществления анализа, в особенности в оперативном режиме, так как они имеют многомерную природу и обеспечивают быстрое извлечение из них нужной информации для обоснования решений или использования в приложениях.

OLAP-анализ может быть реализован в системе клиент-сервер. Клиентские средства реализуют небольшое число измерений. Возможны различные варианты распределения функций между серверной и клиентской частями ИАС в OLAP-анализе.

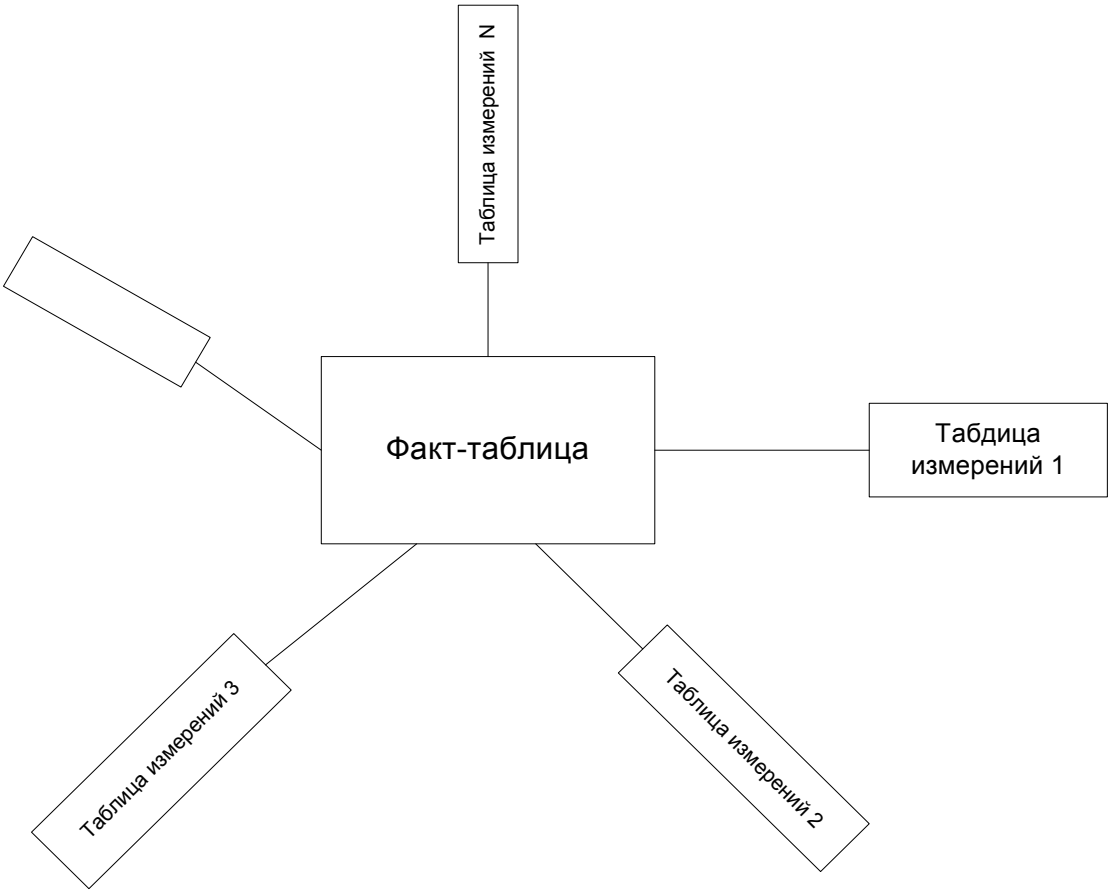

Рис. 3.8. Модель данных по схеме "звезда"

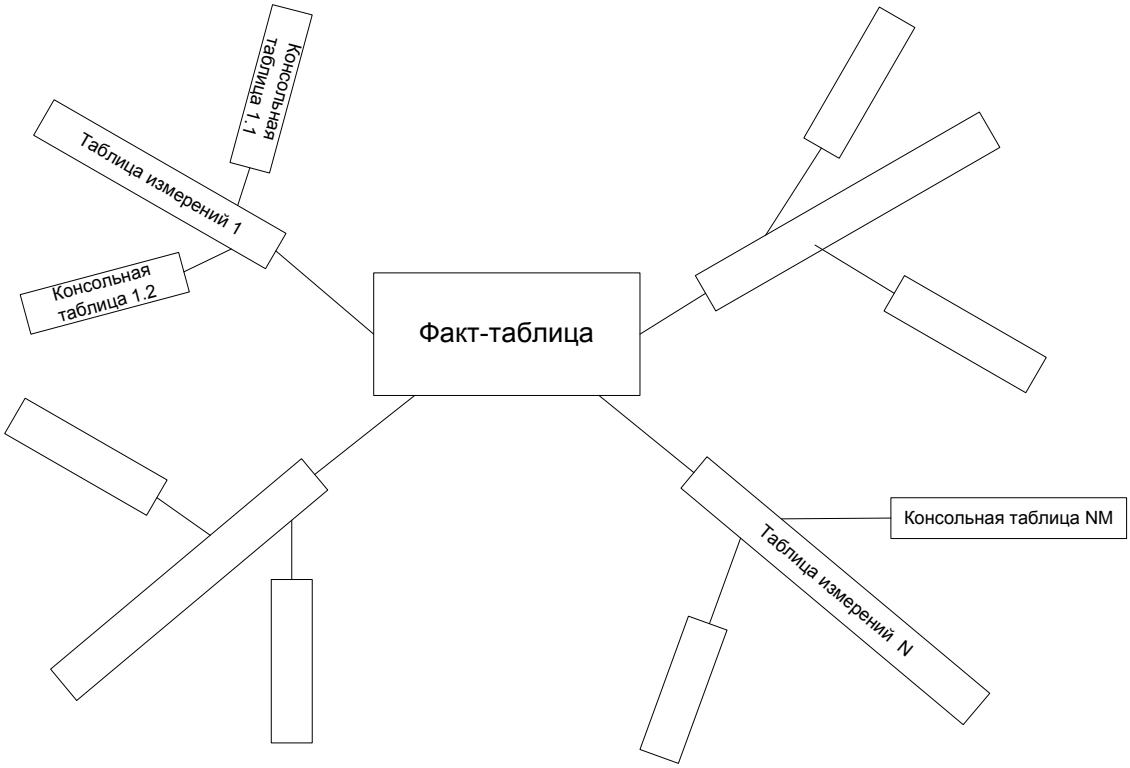

Рис. 3.9. Модель данных по схеме «снежинка»

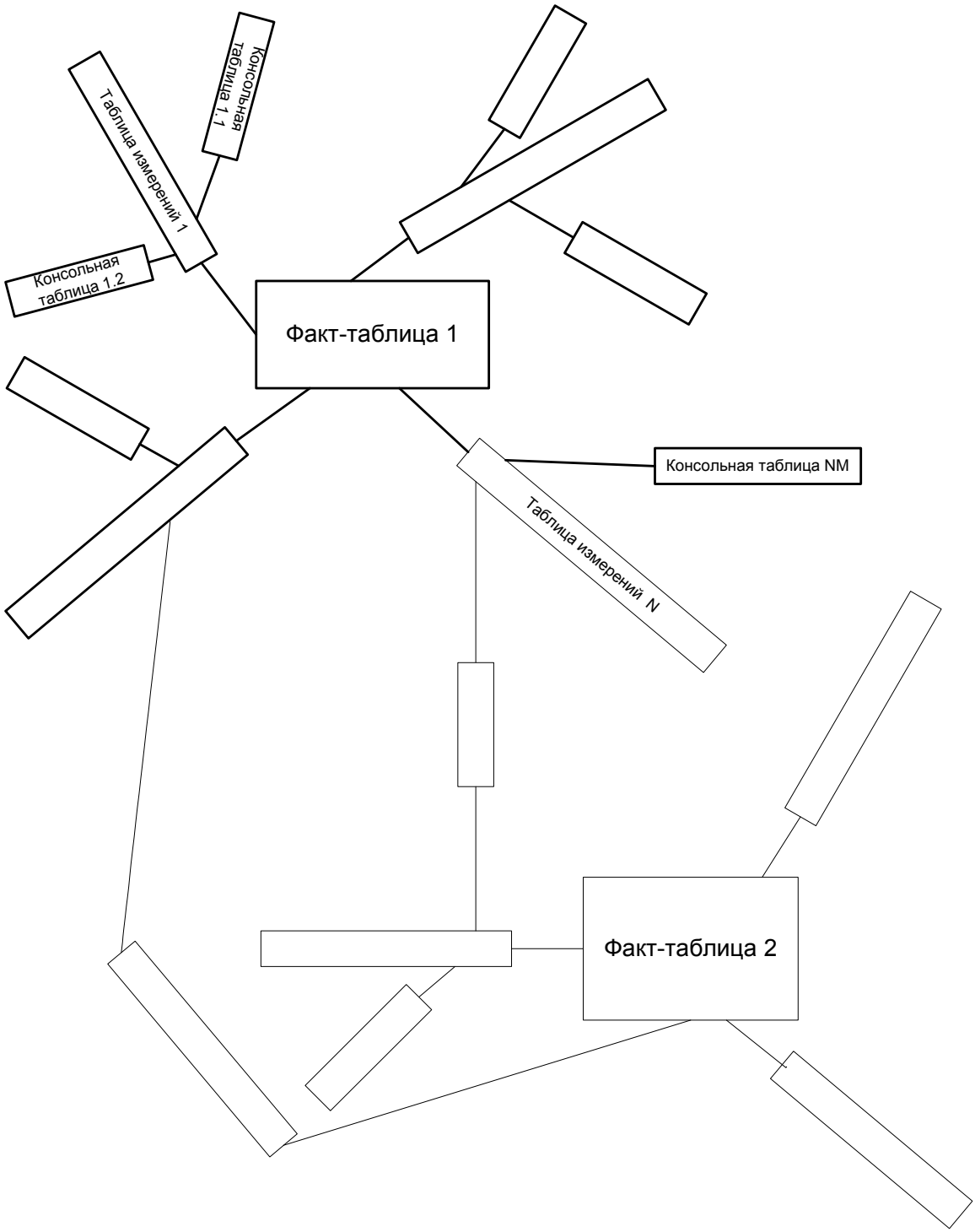

Рис. 3.10. Модель данных по схеме «созвездие»

Далее на рисунках 3.11 – 3.15 приведены примеры реализации названных моделей в среде СУБД MS SQL Server.

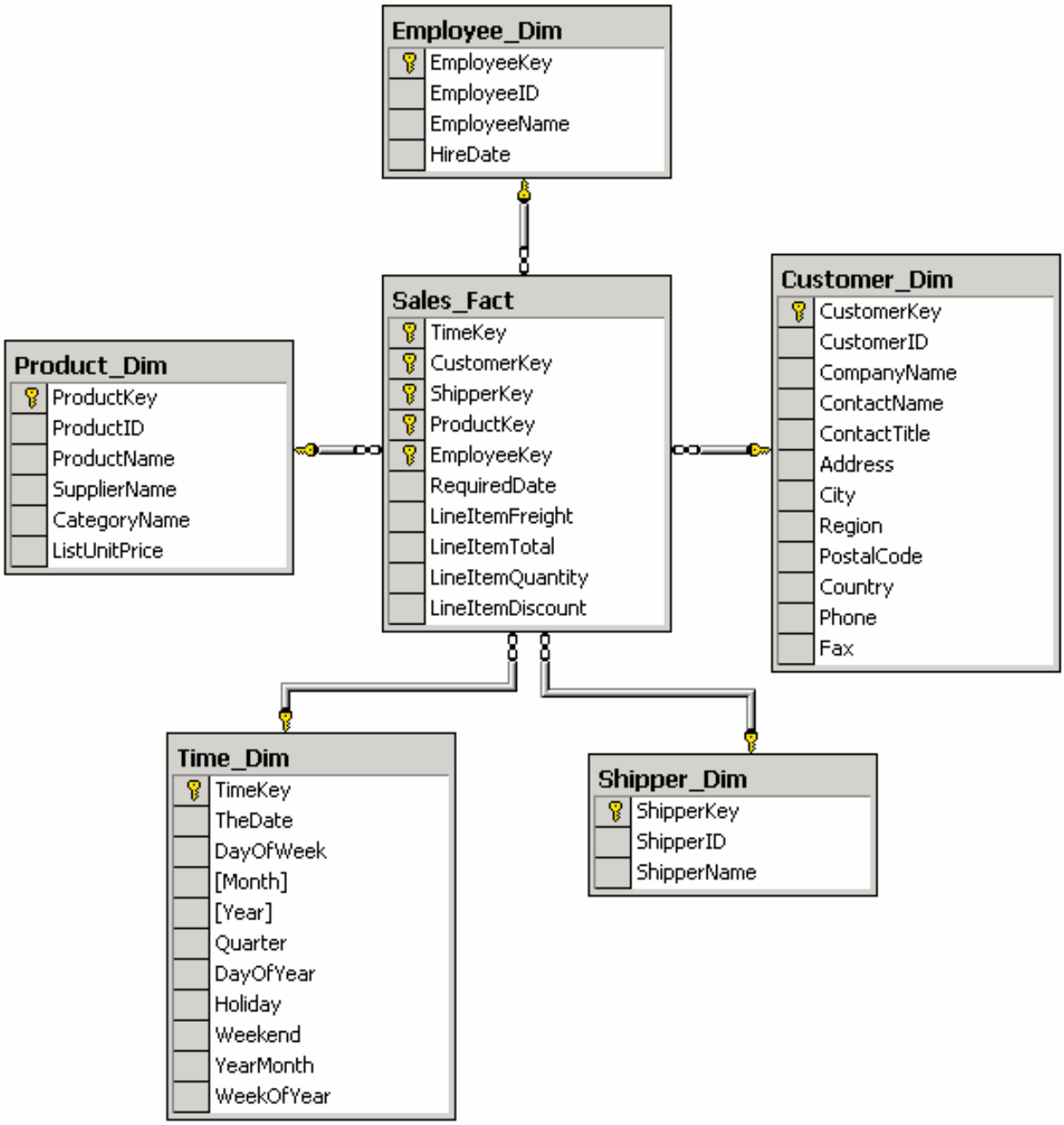

Рис. 3.11. Система таблиц по схеме «звезда»

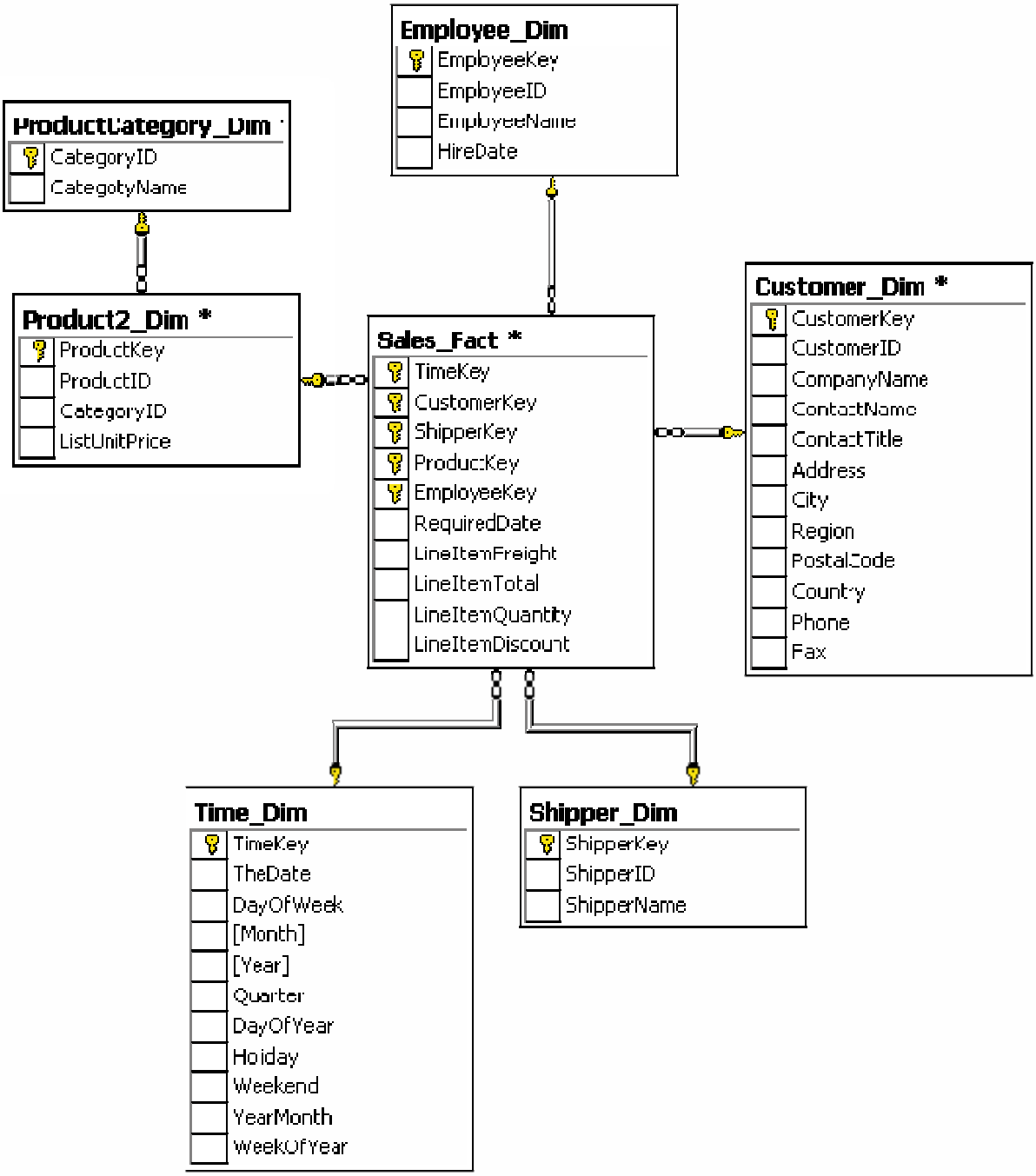

Рис. 3.12. Система таблиц по схеме «Снежинка»

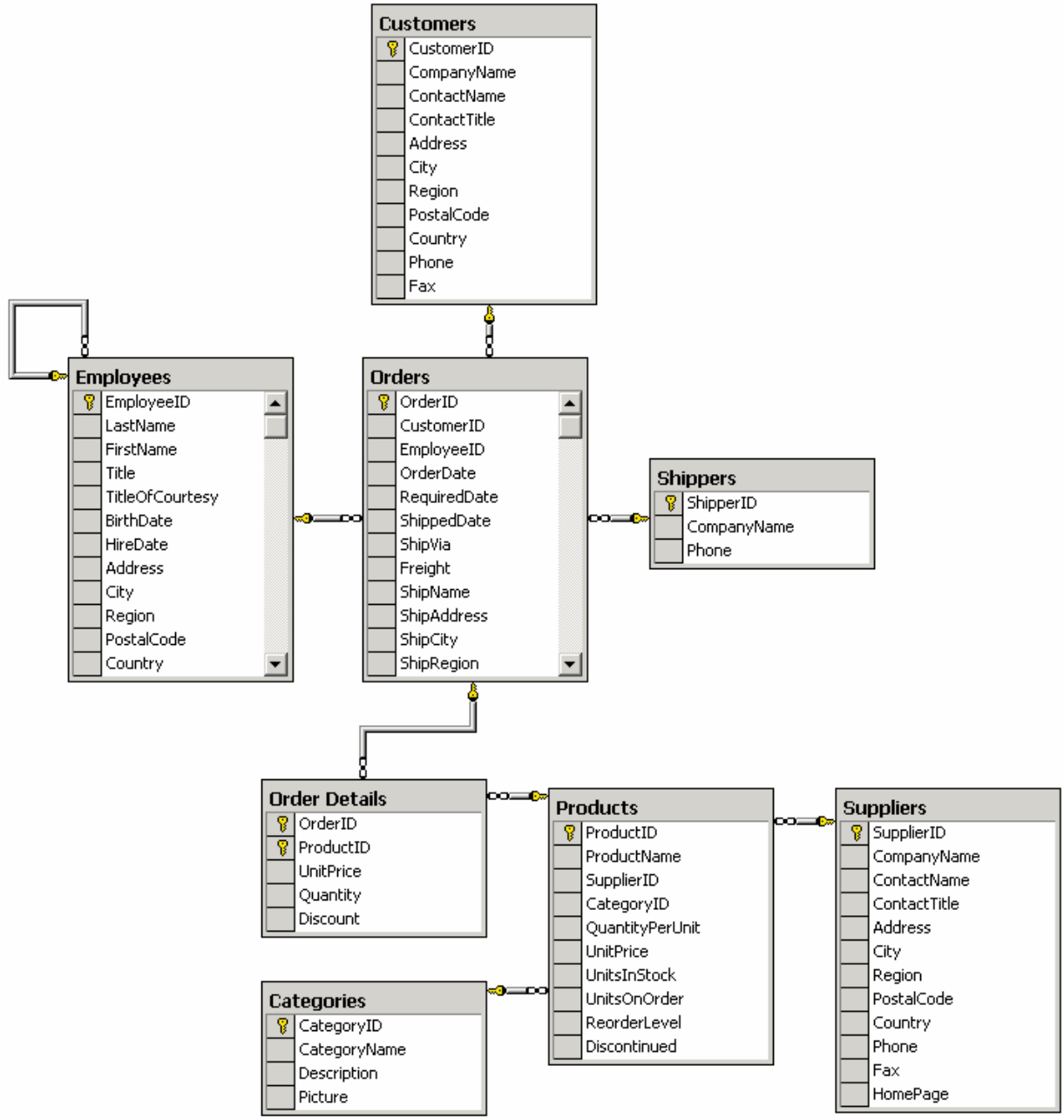

Рис. 3.13. Система таблиц по схеме «созвездие»

SQL Server Enterprise Manager - [3:Data in Table 'Sales_Fact' in 'Northwind_Mart' on 'MAINDESK']

Console Window Help

TimeKey	CustomerKey	ShipperKey	ProductKey	EmployeeKey	RequiredDate	LineItemFreight	LineItemTotal	LineItemQuantity	LineItemDiscount
3	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

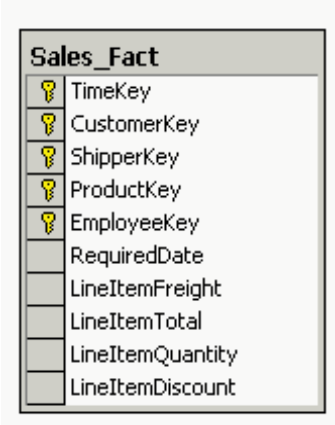

Рис. 3.14. Таблица фактов БД «Борей»

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

Console Window Help

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

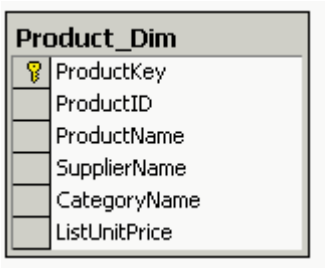

Рис. 3.15. Пример таблицы измерений БД «Борей»

Вопросы для самопроверки

1. Какие этапы прохождения данных от первичных источников к пользователям Вы знаете?
2. В чем состоит идея гибкой архитектуры данных?
3. Назовите свойства открытых систем.
4. Каково содержание свойств расширяемости и масштабируемости?
5. Каково содержание свойств переносимости и интероперабельности?
6. Каково содержание свойств способность к интеграции и высокой готовности?
7. Чем определяется значение стандартов при создании и применении ИАС?
8. Какие стандарты описания показателей Вы знаете?
9. В чем смысл стандартов электронного обмена данными?
10. Какие проблемы разрешаются ETL-процессами?
11. Назовите стадии ETL-процессов.
12. Какие аспекты характеризуют сбор данных?
13. Каковы требования к качеству данных, помещаемых в ИХ?
14. Какие методы повышения качества данных используются в процессе их продвижения в информационное хранилище?
15. Какие проблемы решаются при приведении данных к единому формату информационного хранилища?
16. В чем состоит смысл семантического разрыва?
17. Какие виды семантических разрывов Вы знаете?
18. Какие источники данных для информационного хранилища Вы можете назвать?
19. Назовите основные принципы построения информационных хранилищ – правила Инмона.
20. В чем заключаются требования предметной ориентированности?
21. Раскройте понятие интегрированности данных в ИХ?
22. В чем состоят требования неизменчивости?
23. В чем состоит поддержка хронологии в информационном хранилище?
24. Какова структура информационного хранилища?
25. Раскройте содержание концепции централизованного хранилища данных.
26. В чем состоит концепция распределенного хранилища данных?
27. Что такое автономных витрин данных?
28. В чем смысл концепции единого интегрированного хранилища и многих витрин данных?
29. Раскройте понятие «база метаданных – репозиторий ИХ
30. Каковы назначение и выполняемые функции метаданных?

31. Опишите принципы создания репозитория ИХ.
32. Каково назначение и содержание бизнес-метаданных?
33. Для чего нужны технические метаданные?
34. Чем различаются активные и пассивные метаданные?
35. Как делятся данные по стадиям применения?
36. В чем смысл разделения метаданных на активные рабочие и пассивные рабочие МД?
37. Какие компоненты входят в состав трехмерной классификации и модели метаданных?
38. На какие вопросы отвечает размерностная модель метаданных (модель Захмана) информационного хранилища?
39. Что содержат метаданные, описывающие сущности?
40. Каков состав метаданных, отвечающих на вопрос – где?
41. В чем смысл метаданных, отвечающих на вопрос – когда?
42. Какова идея совокупности метаданных, отвечающих на вопрос – кто?
43. Раскройте смысл совокупности метаданных отвечающих на вопрос – почему?
44. Какие метаданные описывают действия, выполняемые над данными?
45. Раскройте понятие модели данных информационного хранилища.
46. Назовите элементы модели данных информационного хранилища.
47. Что помещается в таблицы фактов?
48. Какие виды фактов Вы знаете?
49. Что представляют таблицы размерности?
50. Каково назначение консольных таблиц?
51. Перечислите виды многомерных схем данных и дайте им краткую характеристику.
52. В чем смысл схемы «звезда»?
53. В чем смысл схемы «снежинка»?
54. В чем смысл схемы «созвездие»?
55. Каковы особенности многомерных моделей данных?
56. Какие принципы положены в основу построения модели данных ИХ?

ГЛАВА 4.
Признаки OLAP-систем, технологии оперативного и интеллектуального анализа данных

4.1. Подходы к выполнению анализа средствами информационных технологий (IT-анализа)

1. Классификация IT-анализа по режиму и темпу.
2. Задачи и содержание оперативного (OLAP) анализа.
3. Типы многомерных OLAP-систем.

Информация, извлекаемая из информационных хранилищ и предоставляемая ее 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.

4.1.1. 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 (Exequtive Information system EIS);

– динамические – обеспечивают обработку нерегламентированных запросов и гибкую систему подготовки отчетов.

Статические ИАС при всей кажущейся простоте и соответственно привлекательности для ЛПР имеют ограниченные возможности по информационной поддержке принятия решений. Зачастую полученная в отчетах информация порождает вытекающие из ее 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 (23):

– -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 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, многомер-

ным 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).

4.1.2. Требования, предъявляемые к OLAP-системам

С 1993 года стал проявляться интерес к многомерному представлению данных – в этом году появилась программная статья Эдварда Кодда. В ней он сформулировал двенадцать основных требований к средствам реализации OLAP, дал критическую оценку реляционного подхода в связи с его малой пригодностью к реализации в задачах многомерного анализа данных с повышенными требованиями к времени отклика на аналитические запросы. Они состоят в следующем:

1. Многомерное представление данных.

Средства должны поддерживать многомерный на концептуальном уровне взгляд на данные.

2. Прозрачность.

Это требование заключается в том, что пользователь не должен знать о том, какие конкретные средства используются хранения и обработки данных, как они организованы и откуда они берутся.

3. Доступность.

Средства должны сами выбирать источник данных и связываться с ним для формирования ответа на данный запрос.

4. Согласованная производительность.

Производительность не должна зависеть от количества измерений в запросе.

5. Поддержка архитектуры «клиент-сервер»

Средства должны работать в архитектуре «клиент-сервер».

6. Равноправность всех измерений.

Ни одно из измерений не должно быть базовым, все они должны быть равноправными.

7. Динамическая обработка разреженных матриц.

Неопределенные значения должны храниться и обрабатываться наиболее эффективными способами.

8. Поддержка многопользовательского режима работы с данными.

Все многомерные операции должны единообразно и согласованно применяться к любому числу любых измерений.

9. Поддержка операций на основе различных измерений.

Все многомерные операции должны единообразно и согласованно применяться к любому числу любых измерений.

10. Простота манипулирования данными.

Средства должны иметь максимально удобный и естественный пользовательский интерфейс.

11. Развитые средства представления данных.

Средства должны поддерживать различные способы представления данных.

12. Неограниченное число измерений и уровней агрегации данных.

Не должно быть ограничений на число поддерживаемых измерений.

К 12 правилам впоследствии были присоединены еще 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 (39) 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 психофизиологичекими 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 обепечивать необходимый уровень защиты при множественном доступе для исключения взаимных помех, несанкционированного доступа. 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 ее из любых необходимых источников. Инструментальные средства оперируют с необходимыми объемами 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.п. 2.1-2.2. 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 ученых Ясина, Королева и др. еще в 70-х годах XX века. В трудах Кодда, Инмона легко узнаются основополагающие идеи этих и других ученых, которые были реализованы в большом числе проектов в разных предметных областях.

4.1.3. 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 трехмерного). Извлечение информации, как правило, сопровождается обработкой ее 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 объемы продаж, их динамику, привязку их к регионам, а также получать другие интересующие аналитика или ЛПР сведения.

Извлечение необходимой информации для построения отчетов производится путем использования ряда процедур.

К ним относятся:

– сечение или срез (slice and dice) – извлечение данных из факт-таблицы по каким-либо определенным значениям одного или нескольких измерений, например из гипер-куба (факт-таблицы), содержащей сведения об издержках в отчет (раздел отчета) помещают данные только по какому-либо одному или группе издержек;

– поворот, под которым понимают изменение координат, их порядка или добавление измерений; эта процедура обеспечивает замену в готовом отчете «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;

– свертка (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 ежесуточно в ИХ их можно передавать в агрегированном виде – еженедельно или ежемесячно, соответственно в агрегированные данные можно помещать в отчеты;

– развертка или раскрытие (roll up) – процедура, обратная свертке, 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 создается отчет, в котором сводятся данные об издержках и выручке по одним и тем же изделиям и временным периодам;
– проекция – конструирование отчетов, являющихся подмножествами из множества единичных реквизитов или атрибутов, содержащихся в операционных базах или в ИХ;
– 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.

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-таблиц-сущностей. 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. 1) представляет исследуемый объект в виде многомерного куба, чаще используют трехмерную 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 трехмерных кубов с целью облегчения восприятия и квазиобъемного представления при формировании отчетных и аналитических документов и мультимедийных презентаций по материалам аналитических работ в системе поддержки принятия решений.

4.1.4. Типы многомерных OLAP-систем.

Общие положения

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 изложены основные свойства многомерной и реляционной моделей OLAP-систем и различия между ними. Сущность смешанной OLAP-системы заключается в возможности использования многомерного и реляционного подхода в зависимости от ситуации: размерности информационных массивов, их структуры, частости обращений к тем или иным записям, вида запросов, наличия соответствующих инструментальных средств и т.д.

Рассмотрим подробнее сущность, достоинства и недостатки приведенных разновидностей OLAP-систем. При этом необходимо обратиться к рис. 3.1 «Структура инфор-

мационного хранилища». На нем приведены зоны хранилища, которые соответствуют этапам продвижения данных из источников к конечным пользователям. В различных вариантах OLAP-систем эти зоны будут заполняться данными, представленными в виде различных моделей.

Многомерные OLAP-системы

В многомерных СУБД данные организованы не в виде реляционных таблиц, а в виде упорядоченных многомерных массивов в виде 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. Структура хранилища в этом случае упрощается, так как отпадает необходимость в зоне хранения данных в многомерном или объектно-ориентированном виде. Снижаются огромные трудозатраты на создание моделей и систем преобразования данных из реляционной модели в объектную (см. рис. 3.1).

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, в ;

– за счет денормализации и предварительной агрегации многомерные массивы используют в 2,5-100 раз больше памяти, чем исходные данные (расход памяти при увеличении числа измерений растет по экспоненциальному закону);

– отсутствуют стандарты на интерфейс и средства манипулирования данными;

-имеются ограничения при загрузке данных.

Реляционные OLAP-системы

В настоящее время в массовых средствах, обеспечивающих аналитическую работу, преобладает использование инструментов на основе реляционного подхода. Структура хранилища остается в том виде, как представлено на рис. 3.1. Трудозатраты на создание зоны многомерных данных резко увеличиваются, так как практически отсутствуют в этой ситуации специализированные средства объективизации реляционной модели данных, содержащихся в информационном хранилище. Время отклика на запросы часто не может уложиться в рамки требований к OLAP-системам.

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, причем 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 учет 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, а также превзойти в экономии памяти.

Гибридные OLAP-системы

Представляют собой сочетание инструментов, реализующих реляционную и многомерную модель данных. Структура хранилища остается в основном такой же, как на рис. 3.1, однако зона многомерных данных создается специализированными средствами. Это позволяет резко снизить затраты ресурсов на создание и поддержание такой зоны, время отклика на запросы, в том числе незапланированные резко снижается, выполняются требования к OLAP-системам.

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, связанных с применением программных инструментальных средств в многомерном анализе.

Тем не менее встречаются обстоятельства, когда применение ROLAP– и HOLAP-систем становится невозможным из-за чрезвычайно жестких требований со стороны объектов управления или сответственно контролируемых процессов. Такие ситуации характерны для крупных промышленных, транспортных, энергетических комплексов, на финансовых рынках, при управлении объектами в критических ситуациях или их моделировании.

Для такого класса применения ИАС становится безальтернативным применение многомерных или объектно-ориентированных инструментальных средств и методов.

4.2. Интеллектуальный анализ данных Data mining

1. Содержание понятия знания. Классификация видов знаний.
2. Задачи Data mining.
3. Специфические методы data mining-а
4. Области применения методов Data Mining
5. Средства реализации и задачи проектированиякомплекса data mining

4.2.1. Содержание понятия знания. Классификация видов знаний

Для обоснования принятия решений необходимы знания. Их добывают из различных источников.

Понятие «знания» рассматривается с различных точек зрения. В соответствии с этим имеется много определений этого понятия. Энциклопедический словарь определяет знания как «проверенный практикой результат познания действительности, верное ее отражение в мышлении человека». Применительно к ситуации с использованием компьютерных информационных систем (ИС) и, в частности ИАС, можно добавить «и в компьютерной ИС». По определению Гавриловой Т.А. и Хорошевского В.Ф. [17] знания это «закономерности предметной области (принципы, связи, законы), полученные в результате

практической деятельности и профессионального опыта, позволяющие специалистам ставить и решать задачи в этой области».

На начальном этапе подготовки данных к использованию в аналитической сфере они же представляют знания как «хорошо структурированные данные или метаданные». Знания различаются по многим признакам. Соответственно в литературе [29,18] приводится классификация различных видов знаний.

1. Различают фактические и стратегические знания.

Фактические – это такие знания, которые позволяют специалисту предметной области решать конкретные задачи из бизнес-сферы или в каком-либо другом виде деятельности. К ним относятся факты, взаимосвязи, системы понятий, правила. Стратегические – позволяют определить поведение объектов в ближайшем или отдаленном будущем.

2. Факты и эвристики.

Факты – это хорошо известные и описанные обстоятельства. К ним относятся также экономические категории, известные и описанные закономерности и так далее.

Эвристики – знания, опыт, навыки специалистов в соответствующих предметных областях. Они являются объектом изучения и внедрения в информационные системы различного назначения.

3. Декларативные и процедурные знания.

Первые являются очевидными, например: выручка – сумма, полученная в результате продажи товаров. Товар – изделие, предназначенное для продажи.

Процедурные – по существу алгоритмы преобразования декларативных знаний, действий над ними.

4. Интенсиональные и экстенсиональные знания.

Первые являются знаниями о связях между объектами (их атрибутами) рассматриваемой предметной области. Вторые – свойства объектов, их состояния, значения свойств в пространстве и динамике.

5. Глубинные и поверхностные знания.

Глубинные знания содержат подробные сведения о структуре предметной области, законах поведения структуры в целом и отдельных ее элементов, достоверные и полные отражения взаимосвязей элементов структуры и т.д. Например: подробные сведения об устройстве компьютера или мобильного телефона, позволяющие производить проектирование их или ремонт.

Поверхностные знания касаются лишь внешних свойств и связей с рассматриваемым объектом(ами). Перечень необходимых сведений о пользовательских свойствах упомянутых или других изделий.

6. Жесткие и мягкие знания.

Жесткие знания отражают системы или объекты с четко выраженными свойствами, связями, поведением, которые легко описываются качественными и количественными признаками, например описываются логико-дедуктивной системой показателей.

Мягкие знания отображают соответственно системы и объекты с трудно поддающимися описанию или формализации свойствами и связями. Дают нечеткие, размытые решения и множественность рекомендаций.

Взаимосвязь между видами знаний отражена на рисунке 4.1.

Рис. 4.1. Характеристики знаний

4.2.2. Задачи Data mining

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

Предметом нашего изучения является Data mining.

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

В экономической предметной области применение методов поиска решений, условий неотрицательности и других свойств математических моделей путем дедуктивного

получения следствий, исходя из предварительно сформулированных предпосылок, относится к разделу экономической науки, называемому математическая экономика.

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

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

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

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

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

Основными задачами интеллектуального анализа [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.

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

Одновременно с этим многомерный статистический анализ твердо удерживает свои позиции в жесткой области знаний. Он делится на: факторный, дисперсионный, регрессионный, корреляционный, кластерный анализ (является также сферой интересов data mining-а). Эти методы позволяют решать многочисленные задачи в области экономики, менеджмента, юриспруденции, которые являются составной частью аналитической подготовки принятия решений.

4.2.3. Специфические методы и области применения data mining-а

Помимо перечисленных выше методов многомерного статистического анализа, ставших традиционными, все более широкое применение находят специфические методы

интеллектуального анализа, происходящие из смежных областей информационных технологий (IT-систем) и достижений различных областей науки.

К специфическим методам интеллектуального анализа относятся:
– методы нечеткой логики;
– системы рассуждений на основе аналогичных случаев;
– классификационные и регрессионные деревья решений;
– нейронные сети;
– генетические алгоритмы;
– байесовское обучение (ассоциации);
– кластеризация и классификация;
– эволюционное программирование;
– алгоритмы ограниченного перебора.

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

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

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

Нейронные сети – это упрощенная аналогия нервной системы живого организма. Разработаны модели нейронных сетей. Распространенной моделью является многослойный персептрон с обратным распространением ошибки. Нейроны работают в составе иерархической сети, в которой нейроны нижележащего слоя своими выходами соединены с входами нейронов вышележащего слоя. На нейроны нижнего слоя подаются значения входных параметров, которые являются сигналами, которые передаются в следующий слой. При этом они ослабляются или усиливаются в зависимости от числовых значений, которые придаются межнейронным связям, называемых весами. На выходе нейрона верхнего слоя вырабатывается сигнал, являющийся ответом сети на введенные значения входных параметров. Для получения необходимых значений весов сеть необходимо «тренировать» на примерах с известными значениями входных параметров и правильных ответов на них. Подбираются такие веса, которые обеспечивают наибольшую близость ответов нейронной сети к правильным.

Генетические алгоритмы представляют собой поисковый метод, используемый для нахождения наилучшего решения или совокупности решений. Он основан на идее естественного отбора. Начинается построение генетических алгоритмов с кодировки исходных логических закономерностей, называемых как и в биологии хромосомами. Набор таких кодов называют популяцией хромосом. Далее применяется функция пригодности, которая выделяет наиболее подходящие элементы для дальнейших операций. Это может быть отбор в какие-либо группы, но возможен и вариант применения скрещивания и му-

тации с целью получения «нового» поколения. Алгоритм работает над изменением старой популяции до тех пор, пока новая не будет отвечать заданным требованиям.

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

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

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

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

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

Области применения методов Data Mining

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

– выделение в данных групп, сходных по некоторым признакам записей;

– нахождение и аппроксимация зависимостей, связывающих анализируемые параметры или события;

– поиск наиболее значимых параметров в данной проблеме (задаче);

– выявление данных, характеризующих значительные или существенные отклонения от сложившихся ранее закономерностей (анализ отклонений);

– прогнозирование развития объектов, систем, процессов на основе хранящейся ретроспективной информации или с использованием принципов обучения на известных примерах и другие задачи.

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

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

Вопросы для самопроверки

1. В чем состоят задачи и содержание OLAP-анализа?
2. Какие функции выполняют процедуры сечение и поворот?
3. Какие функции выполняют процедуры свертка и развертка?
4. Что означают требования к OLAP-системе под рубрикой «быстрый»?
5. Что означают требования к OLAP-системе под рубрикой «анализ»?
6. Что означают требования к OLAP-системе под рубрикой «разделяемой»?
7. Что означают требования к OLAP-системе под рубрикой «многомерной»?
8. Что означают требования к OLAP-системе под рубрикой «информации»?
9. Какие типы OLAP-систем ВЫ знаете, в чем их различие?
10. В чем сущность MOLAP-системы?
11. В чем сущность ROLAP-системы?
12. В чем сущность HOLAP-системы?
13. В чем состоят задачи и содержание OLAP-анализа?
14. Какие функции выполняют процедуры сечение и поворот?
15. Какие функции выполняют процедуры свертка и развертка?
16. Какие функции выполняют процедуры проекция и построение трендов?
17. Для каких целей предназначена подсистема интеллектуального анализа данных?
18. Какие направления и методы поддерживают средства подсистемы интеллектуального анализа ИАС?
19. Раскройте содержание понятия «знания».
20. Каким образом классифицируются виды знаний?
21. В чем специфика методов интеллектуального анализа?
22. Какие методы интеллектуального анализа Вы знаете?
23. В чем состоит содержание методов нечеткой логики, системы рассуждений на основе аналогичных случаев?
24. В чем состоит содержание методов нейронных сетей и генетических алгоритмов?
25. В чем заключаются методы ассоциаций, кластеризации и классификации?
26. Раскройте сущность методов эволюционного программирования и алгоритмов ограниченного перебора.
27. Назовите процессы, явления, закономерности, при исследовании которых используются методы интеллектуального анализа.
28. В каких предметных областях и для каких целей используются методы интеллектуального анализа?
29. Какие средства интеллектуального анализа представлены на рынке, каковы их возможности?
30. Что рекомендуется для успешного комплексирования подсистемы интеллектуального анализа ИАС и достижения успеха при его выполнении?

ГЛАВА 5.
Содержание и методы анализа и прогнозирования бизнес-процессов (деятельности предприятий)

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

5.1. 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 непременным 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 определенные 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 назовем стратегическим анализом.

В процессе реализации планов должен осуществляться контроль и анализ хода их реализации, который назовем 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 (12) 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.

Важным обстоятельством является необходимость

5.2. 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, отраженную в документации и на различных носителях и содержащуюся в информационной системе – анализ фактов;
– на базе как прошлой, так и обращенной в будущее, то есть прогнозной информации – анализ событий и отклонений;
– анализ будущей информации – по существу оценка бюджетов и планов, их альтернатив.

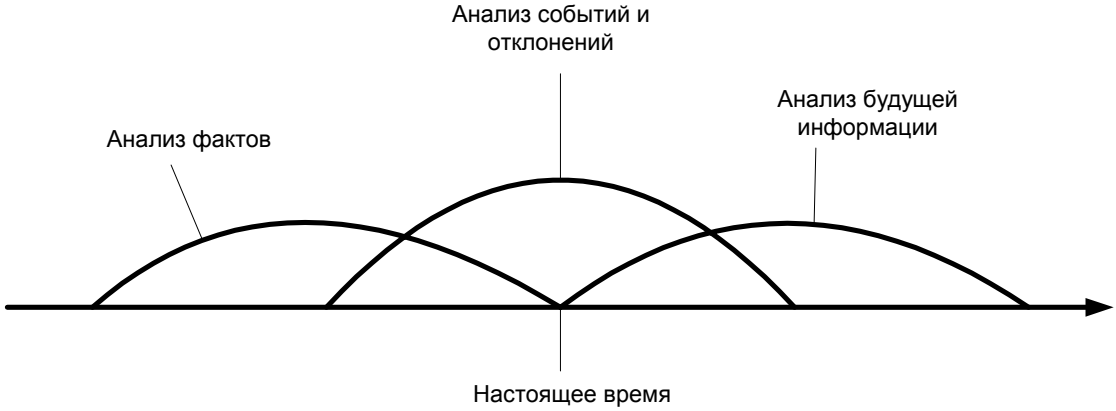

Рис. 5.1. Виды анализа в связи со временем

По масштабности решаемых или обслуживаемых задач анализ делится на:
– 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;
– производстенной или основной деятельности;
– в логистике;
– обеспечении ресурсами;
– финансовой;
– в сфере инвестиций и инноваций.

По методам различают анализ:
– 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, нечеткой 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 приведенная система признаков анализа не претендует на полноту и носит рекомендательный характер.

5.3. Аннотация содержания методов анализа в экономической предметной области

Рассмотрим содержание (краткий обзор) некоторых методов по предметным областям и методикам проведения

5.3.1. 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 учетом 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 расчеты. Определяют зависимость МП или выручки от расходов на рекламу, послепродуктовое обслуживание клиентов, торговые издержки, в том числе площади, других расходов на маркетинг.

5.3.2. 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.

ABC-анализ. Метод, позволяющий выделить наиболее значимые для предприятия группы товаров. Рассчитываются количество и стоимость потребляемых материалов. По итогам расчета 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.

5.3.3. 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 партнерами. 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 объема или по окончанию 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 приема-выдачи материалов и полуфабрикатов, склады, система транспортирования, сопутствующий информационный обмен, включая документооборот. Далее в соответствии с принятыми в логистической цепи методами учета 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.

5.3.4. 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 платежеспособность 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-счет Альтмана,* вычисляемый по балансу и отчету 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.

5.3.5. Анализ инвестиций и инноваций

Расчеты, 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 путем составления иерархии целей, их взвешивания, составления таблиц функций и определения альтернатив реализации функций. Производится расчет 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 неопределенности. Подробно описаны в (14), 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.

5.3.6. Методы стратегического анализа

5.3.6.1. 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 четырехклеточная матрица. Клетки ее 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. Наработаны типовые рекомендации, суть которых сводится к поддержке перспективных, ликвидации безнадежных 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 (14).

А*нализ цепочки создания стоимости и* к*онкурентный 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 партнеров 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, обусловливающие эти силы, оценивается их соотношение. По материалам анализа вырабатывается оптимальная стратегия. Конкретных рекомендаций методика не дает и ограничивается качественным анализом.

5.3.6.2. Анализ ситуации по слабым сигналам и оценка рисков

Методика анализа ситуации по слабым сигналам дает рекомендации по установке контрольных точек, определяет или устанавливает уровни нестабильности, осведомленности. 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.

5.3.6.3. 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 учетом 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* (14), которая представляет собой совокупность формул и схем расчета 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 ведется с использованием методов интеллектуального анализа (см. раздел 5.3). Анализ отклонения может быть обращен как в ретроспективу, так и в перспективу. Исследование ретроспективы ведется в интересах извлечения знаний и формирования на их основе выводов на перспективу.

5.3.6.4. 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.

5.3.6.5. 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 яв-

ляется 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.

5.3.7. Информационный обмен, связанный с аналитической работой

Аналитическая работа на предприятии осуществляется специальной группой. Она может быть автономной или включенной 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. 10.2.

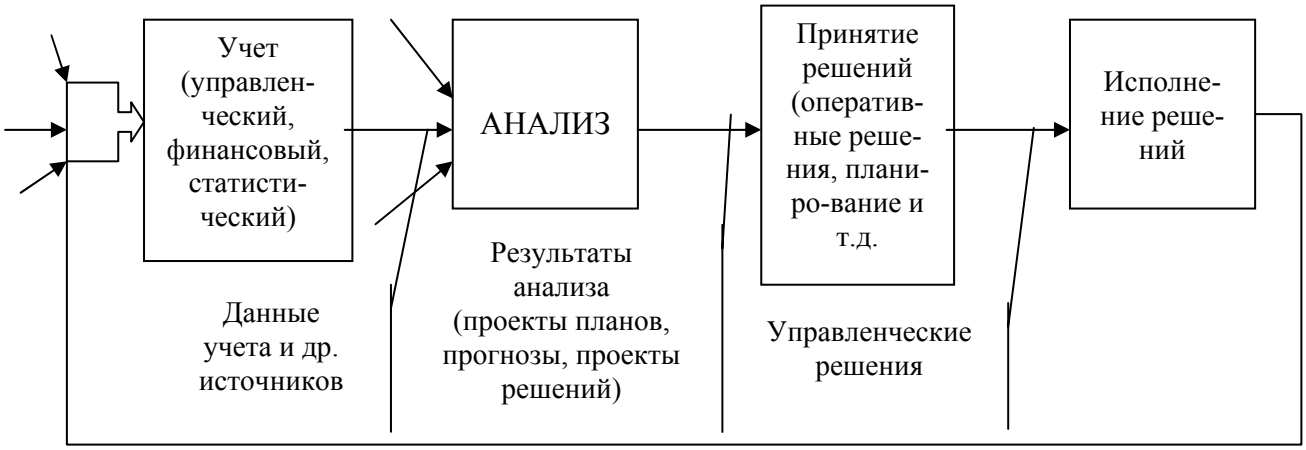

Рис. 5.2. Место анализа в цепочке принятия управленческих решений.

Источники информации для анализа делятся на внутренние и внешние.

К внутренним источникам относятся:

– бухгалтерский учет, включая аналитический и складской;
– статистический учет,
– управленческий учет;
– 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 учета бухгалтерский и статистический относятся к обязательным видам учета.

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 выдает информацию заинтересованным потребителям. Основной потребитель ее – лица, принимающие решения (ЛПР). На предприятии потребителями ее являются также службы управления предприятием. К ним относятся:
– бухгалтерская и финансовая служба;
– служба контроллинга или другие ее подразделения, если аналитики входят в ее 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 ее особой ценности и подчас конфиденциальности.

Форма представления информации

Информация для ЛПР и смежных служб может представляться на бумажных носителях в виде:
– аналитических записок;
– отчетов;
– 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 изготовляются с учетом рекомендаций инженерной психологии, обеспечивают потребителей сжатой, доходчивой информацией с использованием всех имеющихся на предприятии средств, в том числе и мультимедийных.

Вопросы для самопроверки

1. В чем состоит содержание экономического анализа?
2. Что является предметом текущего и стратегического анализа?
3. Перечислите признаки, по которым методики анализа можно разделить на группы.
4. В чем состоит содержание методов анализа в маркетинговой деятельности?
5. Какие методы анализа применяются в областях обеспечения ресурсами и логистики?
6. Каким образом поддерживаются методики анализа финансового состояния, инвестиций и инноваций средствами ИАС?

7. В чем сущность методик стратегического анализа?
8. Какие средства ИАС могут быть использованы при выполнении стратегического анализа?
9. Какие методики оценки текущего состояния предприятия Вы знаете?
10. В чем состоят методики анализа ситуации по слабым сигналам, оценки рисков и управления ими?
11. В чем состоит анализ отклонений?
12. Как организуется аналитическая работа на предприятии?
13. Из каких элементов состоит цепочка движения связанной с аналитической работой информации?
14. Перечислите источники, потребители информации, формы ее представления.

ГЛАВА 6.
Основы создания и применения информационно-аналитических систем

6.1. 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.

6.1.1. 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) усеченного или расширенного типа, в развитых ИАС в комплект входят специализированные языки различного уровня;
- средства графического и визуального конструирования отчетов, рассчитанные на конечного пользователя, как правило, дублируются языковыми средствами;
- средства удаленного доступа, обеспечения работы в распределенном 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. 11.1

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

6.1.2. Средства сбора и доработки данных

Сбор данных из различных источников сопряжен 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 определенной в ИХ структуре и форматам витрин данных.

Инструменты ИАС различного вида имеют в своем 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 и т.д. Мощные средства ИАС насчитывают до 50-ти типов форматов, с которыми может взаимодействовать система.

Рис. 6.1. 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 Da-

tabase 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 производится по определенному согласованному распорядку – сценарию, который может быть сформирован с помощью заимствованных средств, например блока преобразования данных DTS из состава MS SQL Server 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) или объектную надстройку, которая имеет различные названия в конкретных реализациях инструментов ИАС. Например: Юниверс (Univers) в системе Business Objects, выборка в системе «Контур Стандарт», гиперкуб – SAS, пул – SAP R-3, в некоторых источниках ее 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.

6.1.3. 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 ведется наблюдение в процессе эксплуатации за частостью 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.

6.1.4. Средства оперативного (OLAP) анализа

Задачи оперативного анализа изложены в разделе 4.1.3. Все ИАС, сертифицированные как OLAP-системы, имеют в своем составе соответствующие программные модули. 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.

Реализация их направлена на максимальное упрощение действий пользователя в процессе анализа. В процессе оперативного создания отчетов 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 отчетов. Представленная общая структура базы данных в виде схемы с условно отображенными атрибутами (реквизитами) и классами, являющаяся по существу заранее заготовленной витриной данных. Из нее простым перетаскиванием на отведенную часть экрана компонуется отчет. Отпадает необходимость написания SQL-запроса. В некоторых инструментах (Business Objects) в процессе графического конструирования автоматически формируется соответствующий SQL-запрос. Конструирование отчетов 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 отчетов или презентаций, которые состоят из следующих компонентов: страница, таблица, график. Для окончательного оформления с учетом 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 осуществлен и в Web-среду. 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 усеченные версии или специализированные фирменные языки. В различных инструментальных средствах применяются свои варианты инструментов:

– в продукте Business Objects используется стандартный SQL,
– в системе «Контур Стандарт» – объктно-ориентированный язык Python совместно с XML;
– в среде «Линтер-Невод» применяется свой язык LAB в совокупности с SQL;
– 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;
– в семействе продуктов Microsoft – языки VBA, SQL.

6.1.5. 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 крупных массивов данных (до миллионов записей) «Polyanalist» российской фирмы «Megaputer». Эта фирма пользуется мировой известностью, ее продукты используются более чем в 20 странах, занимали призовые места на международных конкурсах, их применяют такие известные фирмы как Chase Manhatten Bank, Dupont, Siemens и другие.

Все семейство продуктов обеспечивает наиболее полный набор методов Data mining и Text mining из известных комплексов инструментальных средств.

В том числе обеспечивается:

– извлечение знаний в больших массивах данных;

– автоматическое построение и тестирование формул, описывающих обнаруженные функциональные зависимости;

– составление классификационных правил по заданным примерам; формирование многомерных кластеров;

– предложение алгоритмов решений;
– текстовый анализ с возможностью связи понятий;
– визуализация результатов анализа и другие возможности, реализованные в шестнадцати модулях.

Набор средств реализован в шести конфигурациях, содержащих различные комбинации модулей:

Polyanalist 4.6 – однопользовательская версия:

Polyanalist 3.5 Lite, – студенческая версия

Polyanalist 3.5 Power – продукты для индивидуальных пользователей и малого бизнеса.

Polyanalist 3.5 Professional для MS Windows NT/2000/XP – мощная система Data Mining для профессионалов.

Polyanalist Knowledge server 4.6 – клиент-серверная сетевая 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 другим СУБД и OLAP-системам. В составе продукта имеется средство разработки собственных data mining-приложений и специализированных аналитических систем – Polyanalist COM-SDK, продукт, предназначенный для создания собственных приложений для Data mining, который имеет высокоуровневый API, библиотеку вызовов и по выбору все или отдельно каждый математический модуль.

Активно развивает свои продукты, ориентированные на реализацию методов Data mining российская фирма – Лаборатория BaseGroup (г. Рязань). Основная ориентация продуктов фирмы – обеспечение:

– подготовки массивов данных к анализу;
– прогнозирования событий и поведения процессов на основе применения большого набора нейросетевых методов;
– анализа ситуаций с помощью аппарата деревьев решений;
– применения методов ассоциативных правил, генетических алгоритмов, фильтрации.

Развитой системой в отношении возможностей, предоставляемых модулями и автономными подсистемами интеллектуального анализа, является инструментарий ИАС фирмы США «SAS Institute Inc». В комплект инструментальных средств, решающих задачи интеллектуального анализа, входят:

– SAS/ETS – в этом модуле реализуются методы анализа временных рядов, экономического системного моделирования и прогнозирования, финансового анализа и формирования отчетов; производится восстановление пропущенных значений методом интерполяции, изменение временной привязки временного ряда, выделение сезонной компоненты во временных рядах, построение трендов, выявление флюктуаций 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 2004 выпускается на рынок платформа SAS9 – Intelligence Platform, которая является интегрированной инструментальной средой, имеющей удобный интерфейс, разнообразные средства прогнозирования, аналитики и оптимизации, бизнес-приложения.

В то же время высокого уровня инструмент Oracle Express OLAP имеет в своем составе ограниченный состав средств интеллектуального анализа. Решение такого рода задач, выходящих за рамки возможностей этого инструмента, предлагается выполнять средствами MS Excel и другими специализированными инструментами Data mining, с которыми 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». Этой фирмой в 2004 году выпущена русская усовершенствованная версия этого инструментального средства. Широко распространенный пакет OLAP-анализа Business Objects имеет в своем 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.

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.

6.2. Управление и проектирование ИАС

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

6.2.1. Управление информационно-аналитическими системами

Управление информационно-аналитической системой является в определенной степени функцией администраторов комплекса информационных технологий, действующих на предприятии. Однако следует учесть, что ИАС обеспечивает необходимыми сведениями – знаниями непосредственно руководителей высшего уровня. Это определяет необходимость участия лиц, принимающих решения, в управлении созданием, развитием и применением комплекса средств ИАС, к которым относятся:

– информационные ресурсы;
– техническая платформа;
– системная платформа в составе операционных систем и сред;
– системы управления базами данных и специальные инструментальные средства создания и поддержки ИАС.

Эти лица ставят задачи на создание информационно-аналитических систем, определяют содержание и структуру сведений, подлежащих накоплению в хранилище, форму представления извлекаемых знаний, лично используют возможности ИАС в различных процессах и режимах (по определению ИАС). В работах по созданию и применению информационно-аналитических систем ЛПР опираются на администраторов, задачи которых будут рассмотрены ниже.

6.2.2. 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.

<u>Загрузка и обновление данных</u>

В ИХ состояние базы данных в отличие от OLTP меняется не постоянно, а по определенному 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.

Одним из важнейших этапов проектирования ИАС является создание общего представления о составе подлежащих анализу данных (содержащихся в ИХ и всех находящихся в сфере взаимодействия операционных баз данных). Такие сведения сосредотачиваются в базе метаданных (МД) – репозитории.

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 приемы перегрузки данных в рамках третьего подхода: замена всей таблицы из операционной БД, просмотр меток дат и времени в источниках информации, чтение операционных контрольных файлов, прием изменений на уровне СУБД, чтение регистрационных записей СУБД, сравнение состояний операционных БД «до» и «после».

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

Планирование

Операции загрузки и обновления должны быть четко спланированы, так как только такой подход обеспечит бесперебойность работы ИАС. В концепцию спроектированной ИАС или ИХ закладывается частность 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 определенное автоматическим планировщиком время. Программные системы могут использовать средства планирования смежных систем. И

Эксплуатация

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

6.2.3. Принципы проектирования информационных хранилищ ИАС

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 warehose 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.

На начальном этапе создания ИАС или соответствующего модуля в интегрированной ИС – фазе структуризации информационного пространства предприятия 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 партнерскими организациями;
- оценить информационный обмен, обусловленный внешними связями с учетом международных с этих позиций;
- рассмотреть используемую выбранными инструментальными средствами ИАС систему классификации и кодирования и интерфейсы с точки зрения оценки объема работ по увязке ее с имеющимися на предприятии наработками;
- разработать или модернизировать с учетом 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.

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 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 объемов 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 партнеров, 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».

Средства реализации и задачи проектирования комплекса data mining

Наиболее развитые инструментальные средства ИАС, такие как SAS, Oracle, особенно математические среды такие как Matlab, обеспечивают представление добытых в результате анализа данных в виде моделей различного рода. Специальные процедуры и языковые средства дают возможность снижения трудозатрат при построении моделей 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.

Задача проектировщика ИАС заключается в подборе, освоении в плане обеспечения соответствия параметров выбранных инструментальных средств задачам, стоящим перед предприятием (корпоративной структурой) в части интеллектуального анализа данных, необходимых для принятия стратегических решений.

Заметим, что комплекс этих средств должен обеспечить выполнение задач анализа, возникающих в трех названных разделах науки: математической экономики, эконометрики, специальных задач DM-g.

Условия успеха в интеллектуальном анализе данных:

- ясность в представлении цели анализа,
- подготовка существенных для проводимого исследования данных,
- правильный выбор методов и программных средств,
- квалифицированное и тщательное выполнение методов анализа.
- решение о применении результатов анализа.

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).

6.2.4. Рынок инструментальных средств ИАС

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

На мировом рынке отмечается бурный рост продаж инструментальных средств Business Intelligence при спаде спроса на функциональные (OLTP-системы) (49). Лидером рынка таких технологий стала компания Business Objects. Ее средства стали полнофункциональными с неограниченными возможностями по отношению к масштабам предприятия. Достигнуто это за счет развития собственного продукта и скупки компаний, владеющих смежными технологиями, необходимыми для достижения полной функциональности.

Как показано выше (п. 1, 2), все средства анализа и ИХ можно поделить на три больших класса:

1. Средства массового применения, к которым относятся Excel, Statistica.

2.Встроенные в □нформации□нные ИС инструментальные средства создания и поддержки ИХ и анализа.

3. Специализированные инструментальные средства создания ИХ и обеспечения различного вида анализа.

4. *Средства, относящиеся к первому классу*

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

– средства получения данных из операционных БД –ODBC и другие по желанию заказчика;
– развитые средства использования факт-таблиц (OLAP-кубов), разработанных в среде инструментов OLAP-средств, статистического и финансового анализа;
– широкий набор средств оформления отчетов, 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.

Упоминавшийся уже пакет Statistica широко распространен, 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 объемам выборок нет. Предназначен для профессиональной аналитической работы с неограниченными по объемам массивами данных в различных предметных областях.

5. *Встроенные средства OLAP и интеллектуального анализа*

Крупные ИИС, такие как SAP и другие системы подобного класса имеют в своем 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 могут служить разработки российских компаний «ПиБи» «OLAP 7.7», ориентированная на работу с широко распространенной системой «1С: Предприятие 7.7» и компании «1С-Рарус» – «1С-Рарус: Динамические отчеты», в которых продукт Контур интегрируется с системой 1С версий 7.7 и 8.

Необходимо при этом учитывать уровень пользователя. Можно приобрести дорогой инструмент с массой возможностей, но из-за недостаточного профессионализма нфозователей или из-за отсутствия постоянной необходимости в применении инструмента в полном объеме его возможностей он найдет 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.

6. *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 средства data mining представлены в ограниченном наборе. 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 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 продукты других фирм. Широко привлекаются компоненты для создания клиентских приложений.
– к мощным системам относятся также продукты фирм Informix, Sybase, IBM, Hiperion;
– к продуктам среднего класса можно отнести Seagete 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 версии 5.1 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. Следующая 6-я версия вывела продукт на нишу крупных пользователей.

Фирма реализовала комплексирование своих продуктов с пакетами программ закупленной фирмы Crystal Decisions, что резко расширяет границы его применения ввиду простоты применения, широкой функциональности и относительной ценовой доступности. С продуктами фирмы комплексируются такие известные ИИС как BAAN и другие, в том числе и российские.

Мощным игроком на рынке инструментов создания ИАС выступила российская фирма «Intersoft Lab», которая выпустила на рынок серию продуктов «Контур» в составе «Контур Стандарт», «Контур Корпорация». «Контур Дизайнер кубов», «Контур OLAP Browser», «Контур Contour Cube components», 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. Продукты «Контур Дизайнер кубов», «Контур OLAP Browser», «Контур Contour Cube components» являются инструментами создания объектных надстроек и OLAP-приложений.

Необходимо отметить на рынке российскую фирму «Релэкс» г. Воронеж. Эта фирма предлагает весьма совершенные оригинальные инструменты «Линтер» – СУБД класса MS SQL Server, Oracle, имеющий практически все достоинства названных систем и информационно-аналитическую систему «Невод», которая обеспечивает представление результатов интеллектуального анализа помимо традиционных представлений также 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 ценовая доступность продуктов.

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

Вопросы для самопроверки

1. Что входит в состав структуры программных средств ИАС?
2. Каково назначение средств сбора и доработки данных?
3. В чем состоят функции средств преобразования данных?
4. Какие задачи и какими методами выполняют средства оперативного анализа?
5. В чем особенности средств интеллектуального анализа?
6. Какие виды средств интеллектуального анализа Вы знаете, в чем их отличия?
7. Какими способами совмещаются программные средства реализации нформациионно-аналитических систем?
8. В чем сущность управления информационно-аналитической системой?
9. На какой структуре сведений о предметной области основано управление ИАС?
10. В чем содержание понятия «метаданные» и какое значение они имеют при создании и поддержке ИАС?
11. В чем заключается методика создания базы метаданных?
12. В чем заключаются задачи администрирования ИАС?
13. Раскройте содержание задач загрузки и обновления данных в ИХ.
14. Раскройте содержание планирования работы ИАС.
15. В чем содержание организации и осуществления эксплуатации ИАС?
16. В чем состоит основная задача проектирования ИАС?
17. Каковы условия превращения данных в знания?
18. Опишите содержание этапов проектирования ИАС.
19. Какие факторы необходимо учитывать при проектировании ИАС?

Руководство по изучению курса

1. Сведения об авторе

Кандидат технических наук
Профессор кафедры ПЭИС

Перечень научных и учебно-методических работ

№ п/п	*Наименование работы, ее вид*	*Выходные данные*	*Соавторы*
1.	Информационно-аналитические системы. Учебное пособие	МЭСИ, 2001 г.	
2.	Введение в информационно-аналитические системы	Сборник научных трудов «Информационно-аналитические системы». – М.; МЭСИ, 2002.	
	Инструментальные средства и информационное обеспечение реинжиниринга бизнес-процессов	Сборник научных трудов «Реинжиниринг бизес-процессов на основе информационных технологий. Системы управления знаниями». – М.; МЭСИ, 2002.	

2. Цели и задачи дисциплины, сфера профессионального применения

Целью курса «Информационно-аналитические системы» (ИАС)» является изучение студентами, обучающимися по специальностям «Менеджмент, Маркетинг, Антикризисное управление, Мировая экономика, Финансы и кредит», проблематики автоматизации анализа информационной подготовки принятия управленческих решений с использоваием современных информационных технологий на основе применения инструментальных средств широкого назначения и специализированных пакетов прикладных программ; освоение основ участия в разработке и сопровождении информационных хранилищ, технологий оперативного и интеллектуального анализа данных, отражающих деятельность предприятий в различных предметных областях.

Основной задачей курса является приобретение студентами прочных знаний и навыков, определяемых целью курса. Должно быть сформировано представление о содержании аналитической работы, необходимо получить знания технологии создания и сопровождения ИАС на основе использования современных инструментальных средств, приобрести навыки аналитической работы.

Сфера профессионального применения

Знания и навыки, приобретенные в ходе изучения курса ИАС, могут быть использованы при изучении других дисциплин в различных предметных областях: в менеджменте, маркетинге, дисциплинах экономического блока, гуманитарных – психологии, юриспруденции и т. д.

В практической деятельности они находят применение во всех направлениях профессиональной деятельности – в государственном управлении, экономике от малых предприятий до крупных корпоративных структур, на транспорте, в производстве и торговле, в гуманитарных направлениях деятельности и т.д.

3. Необходимый объем знаний для изучения дисциплины

Изучение теоретических основ дисциплины и овладение практическими навыками основывается на знаниях и практических навыках, приобретенных студентами при изучении дисциплин: «высшая математика», «линейная алгебра», «теория статистики», «базы данных», «стандарты электронного обмена данными», «микроэкономика», «макроэкономика», «основы бизнеса», «интеллектуальные информационные системы» и других дисциплин.

4. Перечень основных тем дисциплины

4.1. Тема 1. Базовые понятия информационно-аналитических систем

Цель изучения: Осветить содержание курса, подходы и методы, используемые в процессе анализа, при создании ИАС, в инструментальных средствах их создания и поддержки.

4.1.1 Подходы, используемые при автоматизации процессов экономического анализа. Изучаются основные понятия информационно-аналитических систем. Определяется роль и место анализа в процессе принятия решений. Рассматривается и обосновывается необходимость применения информационных технологий при выполнении аналитических работ с имеющимися в распоряжении ЛПР данными для выработки адекватных сложившимся ситуациям решений.

4.1.2 Аспекты проблемы анализа. Выделяются три основных аспекта: сбор и хранение данных, необходимых для аналитической подготовки принятия решений; собственно анализ, разделяющийся на оперативный и интеллектуальный; эффективное представление результатов анализа. Решаются проблемы автоматизации анализа посредством создания информационно-аналитических систем (ИАС). Дается определение системы.

4.1.3 Состав информационно-аналитической системы. Рассматривается состав ИАС, отвечающий полному набору функций в соответствии с приведенными выше аспектами. Отмечается консолидирующая роль средств ИАС в создании интегрированной информационной системы управления предприятием.

4.1.4 Типы инструментальных средств создания и поддержки ИАС. Рассматривается назначение инструментальных средств, приводится их классификация. Объектом изучения являются: аналитические инструментальные средства пакетов прикладных программ широкого применения MS Office – Excel, Access; Statistica, Mathcad,; специализированные средства SAS, Business Objects, Линтер-Невод, Контур-корпорация/контур-стандарт, Oracle Express; их структура и функции; соответствующие блоки интегрированных экономических информационных систем.

Изучив данную тему, студент должен

знать: место и значение анализа в системе управления предприятием, задачи информационно-аналитических систем, основные функции инструментальных средств ИАС, их классификацию и наиболее известные образцы;

уметь: определять класс инструментальных средств по составу функций, назначению, масштабности выполняемых задач.

При изучении темы 1 необходимо:

Читать учебное пособие [51] главу 1; учебно-справочное издание [50] главу 1 стр. 3–8; книгу [29] стр. 12–22; учебный курс [18] стр. 12–19.

Сосредоточить внимание на следующем:
Место и значение анализа в системе принятия решений по управлению предприятием, необходимость внедрения информационных технологий в аналитическую работу, проблемы аналитической подготовки принятия решений, функциональность и классификацию инструментальных средств ИАС.

Для самооценки знаний по теме 1 ответить на вопросы

1. Какое значение имеет аналитическая работа для успеха предприятия?
2. Какие факторы влияют на деятельность предприятия?
3. Какие информационные ресурсы используются для подготовки принятия решений?
4. Что является предпосылкой для принятия правильных решений?
5. Какие объемы данных используются в процессе анализа?
6. Назовите требования к информации, которая используется для принятия решений.
7. Что такое информационно-аналитическая система?
8. Что вызвало появление и широкое распространение информационно-аналитических систем?
9. Назовите аспекты проблемы анализа в процессе подготовки принятия решений?
10. В чем заключаются аспекты сбора и хранения информации?
11. В чем состоит содержание аспектов анализа данных и предоставления результатов анализа пользователям?
12. Какие типы инструментальных средств для реализации информационно-аналитических систем вы можете назвать?
13. Какие информационные технологии и информационные системы на предприятии и из внешней среды являются источником данных для сосредоточения в информационном хранилище или непосредственно для анализа?
14. В каких видах информационных систем используются результаты анализа?

4.2. Тема 2. Информационное пространство и система экономических и других показателей как среда анализа

Цель изучения: изучить понятие и структуру информационного пространства; понятия показателя, его пространственную интерпретацию; содержание экономических показателей, виды систем показателей.

4.2.1. Понятие информационного пространства (ИП). Показывается, что все действия, связанные с анализом и принятием решений, производятся в информационном пространстве. На основе понятия системы раскрывается сущность понятия информационного пространства, его содержание.

4.2.2. Структура информационного пространства и его элементы. Дается определение структуры ИП. Показывается, что свойство структурированности ИП дает возможность представлять сведения в виде документов и манипулировать ими. Рассматривается система единиц информации, даются их определения. Раскрывается содержание понятия показатель с двух точек зрения: структурно-формальной и экономической.

4.2.3. Пространственная интерпретация понятия показатель. Рассматривается представление совокупности признаков показателя в виде системы координат части инфор-

мационного пространства, отображающей соответствующий показатель. Представлены варианты геометрической интерпретации систем показателей.

4.2.4. Сущность и система экономических показателей. Дается понятие системы оценок как совокупности показателей с их критериальными значениями. Рассматривается понятие показателя с экономической точки зрения. Показано, что состояние и деятельность предприятия с необходимой полнотой возможно отобразить лишь системой показателей.

4.2.5. Содержание и структуризация систем экономических показателей. Рассматривается необходимость увязки микро- и макроэкономических показателей в экономической сфере, необходимость учета сложившейся структуры показателей в экономике страны и тесной увязки с ней структуры показателей предприятий. Раскрываются понятия классификации и кодирования систем показателей. Даются рекомендации по организации работ по структуризации ИП предприятия при создании и внедрении ИАС.

4.2.6. Виды систем показателей. Рассматриваются представление значений показателей в абсолютных и относительных величинах для оценки состояния предприятий и других объектов. Даются характеристики логико-дедуктивных и эмпирико-индуктивных систем показателей.

Изучив данную тему, студент должен

знать: содержание понятий: информационное пространство, система единиц информации, показатель и система показателей; сущность пространственной интерпретации показателей; сущность классификации и кодирования показателей; виды систем показателей.

уметь: применять полученные знания при внедрении и эксплуатации информационно-аналитических систем.

приобрести навыки классификации и кодирования показателей при создании структуры информационного хранилища.

При изучении темы 2 необходимо: Читать учебное пособие [51]главу 2; учебно-справочное издание [50] главу 1 стр. 3–8; книгу [3] стр. 12–22; учебник [1] стр. 12–19.

Сосредоточить внимание на следующем:

При создании и целевом использовании ИАС необходимо изучить требования рыночных инструментальных средств к структуризации информационного пространства предприятия с тем, что бы сократить сроки внедрения и избежать возможных ошибок.

Для самооценки по теме 2 ответить на вопросы

1. Дайте определение понятия информационного пространства.
2. В каких видах содержатся сведения в ИП и какие манипуляции совершаются над компонентами ИП?
3. Какое Вы знаете характерное свойство ИП и в чем оно состоит?
4. Какие единицы информации Вы знаете? В чем их содержание?
5. Дайте определение понятия показатель, исходя из формально-структурного подхода.
6. Каким образом строится пространственная интерпретация понятия показатель?
7. Какие виды геометрического представления показателей и их систем Вы знаете?

8. Что такое система оценок и в чем ее смысл?
9. Дайте определение показателя с точки зрения экономиста. В чем содержание экономических показателей?
10. В чем заключаются классификация и кодирование показателей?
11. Какие общегосударственные системы классификации и кодирования Вы знаете?
12. Что рекомендуется выполнить при внедрении на предприятии информационно-аналитической системы в части структуризации ИП?
13. Как строятся системы показателей деятельности предприятия?
14. Что такое логико-дедуктивная система показателей?
15. то такое эмпирико-индуктивная система показателей?

4.3. Тема 3. Технологии сбора и хранения данных – концепция информационных хранилищ

Цель изучения: Освоить технологии сбора данных из различных источников, основы построения структур предметных разделов и информационного хранилища в целом. Приобрести навыки построения рабочих хранилищ и витрин данных.

4.3.1. Понятие о гибкой архитектуре данных. Рассматриваются принципы гибкой архитектуры данных, обеспечивающей доступ с любой аппаратно-программной платформы к любому нужному источнику, а также принципы открытых систем, использование которых позволяет реализовать гибкую архитектуру данных. Раскрывается значение стандартов в создании системы сбора данных в информационное хранилище.

4.3.2. Повышение качества информации при сборе ее в информационное хранилище. Раскрываются проблемы обеспечения качества информации на предприятии и способы его повышения на пути движения данных из первичных источников к хранилищам и поддержания в процессе накопления и хранения данных в ИХ. Рассматриваются методы оценки данных – «ручные» и программные, а также оценки выгод, которые может получить и получает предприятие, обеспечивая должное качество данных.

4.3.3. Преобразование данных в единый формат и приведение их к единой структуре. Отмечается то, что приходится при пересылке данных из первичных источников решать проблему приведения их к единому формату из множества форматов, применяемых в истоках, а также согласования интерфейсов в смысловом значении данных.

4.3.4. Основные принципы построения информационных хранилищ. Рассматриваются правила Инмона – автора концепции ИХ, определяющие основные свойства ИХ. Раскрываются такие свойства как: предметная ориентированность, интегрированность, неизменчивость, поддержка хронологии.

4.3.5. Понятие о метаданных (МД), базе МД – репозитории, используемых в информационном хранилище. Приводится определение МД, рассматривается назначение, состав, выполняемые функции. Описывается классификация МД по различным признакам: по динамике использования, в зависимости от этапа применения, по детальности описания предметной области. Рассматриваются принципы создания репозитория ИХ.

4.3.6. Особенности построения модели данных информационного хранилища. Рассматривается принцип ориентации структуры ИХ на выполняемый конечным пользователем анализ содержащихся в нем данных. Схема данных, имеющихся в OLTP – транзакционных системах, должна обеспечивать их связь с данными, помещаемыми в хранилище. Обосновывается необходимость перехода от нормализованных данных в реляционных базах к схеме «звезда»-«снежинка»-«созвездие» для обеспечения представления данных в размерностном историческом аспекте. Раскрываются понятия «факт-таблица», «таблица измерения», OLAP-куб.

Изучив данную тему, студент должен

знать: принципы сбора данных из различных источников, понятия о базах метаданных (БМД), требования к информационным хранилищам, основы построения БМД и ИХ.

уметь: создавать логическую модель системы сбора данных, базы метаданных и информационного хранилища.

приобрести навыки формулирования потребностей бизнес-пользователей в составе, форматах и структуре исходных данных для выполнения аналитических работ, формирования состава базы метаданных.

При изучении темы 3 необходимо:

Читать учебное пособие [51] главу 3; учебно-справочное издание [50]главу 2 стр. 18–32; Книгу [29] главу 3 стр. 41–57; книгу [13] главу 8 стр. 241–269.

Сосредоточить внимание на следующем:

При сборе и обработке данных из различных источников и помещении в информационное хранилище необходимо использовать гибкую архитектуру данных, обеспечить должный уровень качества закладываемых в ИХ данных; в процессе преобразования исходных форматов в единый формат ИХ согласовать семантику данных первичных источников со смыслом атрибутов базы метаданных.

Для самооценки по теме 3 ответить на вопросы

1. Назовите принципы гибкой архитектуры данных.
2. Назовите принципы открытых систем.
3. Каковы требования к качеству данных, помещаемых в ИХ?
4. Какие методы повышения качества данных используются в процессе их продвижения в информационное хранилище?
5. Какие проблемы решаются при приведении данных к единому формату информационного хранилища?
6. Назовите основные принципы построения информационных хранилищ.
7. В чем заключаются требования предметной ориентированности и интегрированности данных в ИХ?
8. В чем заключаются требования неизменчивости и поддержки хронологии?
9. Раскройте понятие «база метаданных – репозиторий ИХ.
10. Каковы назначение, состав и выполняемые функции метаданных?
11. Опишите принципы создания репозитория ИХ.
12. Какие принципы положены в основу построения модели данных ИХ?
13. В чем заключается необходимость перехода от реляционной модели данных к схемам «звезда», «снежинка», «созвездие»?
14. Что такое «факт-таблица», «таблица измерения»?

4.4. Тема 4. Признаки OLAP-систем, технологии оперативного и интеллектуального анализа данных

Цель изучения: Усвоить требования, предъявляемые к OLAP-системам. Освоить технологии оперативного (OLAP) и интеллектуального (Data mining) анализа сосредоточенных в информационном хранилище данных.

4.4.1. Краткое содержание правил Кодда, которым должны соответствовать OLAP-системы, обобщение этих правил в требованиях теста FASMI. Раскрывается сущность требований: Быстрый, Анализ, Разделяемой, Многомерной, Информации.

4.4.2. Типы многомерных OLAP-систем. Проводится анализ трех вариантов получения данных в многомерном представлении, пригодном для OLAP-анализа. Рассматривается представление данных в виде многомерной модели: непосредственно специализированными многомерными средствами – многомерные MOLAP-системы; на основе рационального применения средств реляционных СУБД – реляционные ROLAP-системы; сочетанием реляционного и чисто многомерного подхода – гибридные HOLAP-системы.

4.4.3. Задачи и содержание OLAP-анализа. Рассматриваются способы извлечения необходимых данных в сочетании с обработкой их по несложным алгоритмам с применением многомерного статистического анализа. Раскрывается содержание специфических процедур OLAP-анализа: сечение или срез, поворот, свертка и развертка, проекция, построение трендов.

4.4.4. Назначение и состав выполняемых задач подсистемой интеллектуального анализа данных информационно-аналитической системы. Рассматривается проблема получения знаний из данных, сосредоточенных в ИХ и других источниках. Показано отличие области применения методов Data mining от подходов инженерии знаний. Показано, что предметом и средствами интеллектуального анализа в полнофункциональной ИАС могут быть методы, а также средства, поддерживающие в экономической предметной области эконометрику, математическую экономику и специфические средства Data mining.

4.4.5. Содержание понятия «знания». Классификация видов знаний. Дается краткая характеристика их видов. Рассматриваются фактические и стратегические знания, факты и эвристики, декларативные и процедурные, интенсиональные и экстенсиональные, глубинные и поверхностные, жесткие и мягкие знания.

4.4.6. Специфические задачи интеллектуального анализа. Перечисляются методики решения таких задач. Рассматриваются отличия этих задач от традиционного многомерного статистического анализа, являющегося основой арсенала методов эконометрики и математической экономики. Дается обзор методов интеллектуального анализа данных. Рассматривается содержание методов нечеткой логики, систем рассуждения на основе аналогичных случаев (CBR), генетических алгоритмов, эволюционного программирования, алгоритмов ограниченного перебора и других методов.

4.4.7. Области применения методов интеллектуального анализа. Рассматриваются процессы, явления и закономерности, при исследовании которых целесообразно применить те или иные виды интеллектуального анализа. Приводится перечень предметных областей в сфере экономики, управления и т.д.

Изучив данную тему, студент должен

знать: содержание требований к OLAP-системам, преимущества и недостатки трех их типов, процедуры OLAP-анализа; назначение и возможности подсистемы интеллекту-

ального анализа данных полнофункциональной информационно-аналитической системы, состав и выполняемые функции, используемые методики и области применения.
уметь: выбирать соответствующие потребностям предприятия типы OLAP-систем и инструментальных средств интеллектуального анализа, подбирать необходимые процедуры OLAP-анализа и методики интеллектуального анализа к задачам аналитической работы, выполнять задачи оперативного и интеллектуального анализа.
приобрести навыки применения средств оперативного и интеллектуального анализа при оценке данных и подготовке аналитических материалов для принятия решений.

При изучении темы 4 необходимо:

Читать учебное пособие [51] главу 4; учебно-справочное издание [50] главу 6 стр. 80–106; книгу [13] главу 11 стр. 325–334.

Сосредоточить внимание на следующем:

При подготовке подсистем оперативного и интеллектуального анализа ИАС оценить потребности предприятия (аналитика) в составе и содержании предполагаемых запросов, направлений исследований по методикам интеллектуального анализа; научиться выбирать соответствующий тип OLAP-системы и инструментальных средств Data mining, исходя из возможностей имеющейся или выбранной аппаратно-программной платформы.

Для самооценки по теме 4 ответить на вопросы

1. Что означают требования к OLAP-системе под рубрикой «быстрый, анализ, разделяемой, многомерной, информации»?
2. Какие типы OLAP-систем Вы знаете, в чем их различие?
3. В чем состоят задачи и содержание OLAP-анализа?
5. Какие функции выполняют процедуры сечение и поворот?
6. Какие функции выполняют процедуры свертка и развертка?
7. Какие функции выполняют процедуры проекция и построение трендов?
7. Для каких целей предназначена подсистема интеллектуального анализа данных?
8. Какие направления и методы поддерживают средства подсистемы интеллектуального анализа ИАС?
9. Раскройте содержание понятия «знания».
10. Каким образом классифицируются виды знаний?
11. Какие методы интеллектуального анализа Вы знаете, в чем их специфика?
12. В чем состоит содержание методов нечеткой логики, системы рассуждений на основе аналогичных случаев?
13. В чем состоит содержание методов нейронных сетей и генетических алгоритмов?
14. В чем заключаются методы ассоциаций, кластеризации и классификации?
15. Раскройте сущность методов эволюционного программирования и алгоритмов ограниченного перебора.
16. Назовите процессы, явления, закономерности, при исследовании которых используются методы интеллектуального анализа.
17. В каких предметных областях и для каких целей используются методы интеллектуального анализа?
18. Какие средства интеллектуального анализа представлены на рынке, каковы их возможности?
19. Что рекомендуется для успешного комплексирования подсистемы интеллектуального анализа ИАС и достижения успеха при его выполнении?

4.5. Тема 5. Содержание и методы анализа и прогнозирования бизнес-процессов (деятельности предприятий) как объект автоматизации

Цель изучения: изучить содержание экономического анализа, классификацию и содержание методов анализа в экономике, информационный обмен, связанный с аналитической работой.

4.5.1. Содержание экономического анализа. Рассматривается значение анализа в управлении предприятием и другими хозяйственными объектами, отмечается связь анализа с планированием деятельности, основанным на прогнозировании хода процессов. Раскрывается содержание стратегического и текущего анализа.

4.5.2. Классификация методов анализа. Рассматривается система признаков, по которым методы анализа делят на группы. Приводится состав групп по: целям, по временному фактору, масштабности решаемых задач, предметным областям экономики, применяемым специфическим методикам в сфере экономики, математическим методам. Отмечается, что применение информационных технологий, средств и методов информационно-аналитических систем повышает эффективность рассматриваемых методик анализа.

4.5.3. Методики проведения анализа в маркетинговой деятельности. Дается краткий обзор: анализа разрыва, маржинальной прибыли, портфолио-анализа, сравнительных расчетов с позиций поддержки этих методик средствами ИАС.

4.5.4. Анализ в области обеспечения ресурсами и логистики. В части обеспечения ресурсами рассматриваются ABC-анализ, определение верхних границ цен, анализ возможных прерываний бизнес-процесса. В логистике дается характеристика методов анализа цепочек логистических процессов и издержек, связанных с этими процессами.

4.5.5. Обеспечение средствами ИАС анализа финансового, инвестиций и инноваций. Рассматриваются проблемы реализации с помощью методик и средств оперативного и интеллектуального анализа традиционных подходов: баланса притока и оттока финансовых средств, финансовой паутины, финансово-стоимостного анализа проектов.

4.5.6. Обзор методов стратегического анализа в связи с использованием средств ИАС. Даются характеристики SWOT-анализа, матриц БКГ и Мак-Кинси, анализа цепочки создания стоимости и конкурентного анализа по Портеру, полей бизнеса и бенчмаркинга. Рассматриваются проблемы информационной поддержки этих методик в историческом аспекте.

4.5.7. Поддержка методик анализа текущего состояния предприятия средствами ИАС. Дается характеристика анализа ситуации по слабым сигналам, оценки рисков и управления ими, анализа отклонений. Рассматриваются виды отклонений и способы их обнаружения, оценки и выработки управлений.

4.5.8. Информационный обмен, связанный с аналитической работой. Рассматриваются варианты организации аналитических работ на предприятии, цепочка движения информации по этапам подготовки, принятия и реализации решений. Перечисляются внутренние и внешние источники информации, ее потребители, формы представления.

Изучив данную тему, студент должен

знать: содержание экономического анализа; основные группы методов анализа; назначение и возможности, предоставляемые этими методами, принципы использования

средств ИАС при выполнении анализа, систему информационного обмена, связанного с аналитической работой.

уметь: выбрать соответствующий стоящей перед аналитиком задаче метод анализа и обеспечить его необходимыми средствами из состава инструментов и информационных массивов, которыми располагает имеющаяся на предприятии ИАС.

приобрести навыки аналитической работы по выбранным методикам с применением средств ИАС.

При изучении темы 5 необходимо:

Читать учебное пособие [51] главу 5; книгу [18] глава 1 стр. 14–32; книгу [13] главу 12 стр. 357–378.

Сосредоточить внимание на следующем:

При организации аналитической работы необходимо тщательно изучить требования к результатам работы, выбрать адекватные им методику анализа и средства ИАС.

Для самооценки по теме 5 ответить на вопросы

1. В чем состоит содержание экономического анализа?
2. Что является предметом текущего и стратегического анализа?
3. Перечислите признаки, по которым методики анализа можно разделить на группы.
4. В чем состоит содержание методов анализа в маркетинговой деятельности?
5. Какие методы анализа применяются в областях обеспечения ресурсами и логистики?
6. Каким образом поддерживаются методики анализа финансового состояния, инвестиций и инноваций средствами ИАС?
7. В чем сущность методик стратегического анализа?
8. Какие средства ИАС могут быть использованы при выполнении стратегического анализа?
9. Какие методики оценки текущего состояния предприятия Вы знаете?
10. В чем состоят методики анализа ситуации по слабым сигналам, оценки рисков и управления ими?
11. В чем состоит анализ отклонений?
12. Как организуется аналитическая работа на предприятии?
13. Из каких элементов состоит цепочка движения связанной с аналитической работой информации?
14. Перечислите источники, потребители информации, формы ее представления.

4.6. Тема 6. Основы создания и применения информационно-аналитических систем

Цель изучения: Освоить методы создания информационно-аналитических систем; усвоить принципы создания базы метаданных как основы структуры ИАС, организации применения средств ИАС в целях обеспечения аналитической подготовки принятия решений.

4.5.9. Методы создания информационно-аналитических систем. Рассматривается последовательность и содержание этапов создания ИАС. Освещается роль и задачи бу-

дущих конечных пользователей системы – лиц, принимающих решения, и экспертов-аналитиков в ее создании.

4.5.10. Сущность управления информационно-аналитической системой. Рассматриваются задачи управления ИАС; показано, что основой проекта ИХ и ИАС в целом, а в дальнейшем и управления ИАС являются метаданные (МД).

4.5.11. Задачи и средства администрирования ИАС. Рассматривается перечень работ, выполняемых в рамках задач администрирования ИАС. Определяется место пользователей в решении задач администрирования. Раскрывается содержание задачи загрузки и обновления данных, функции планирования работы ИАС, организации и осуществления ее эксплуатации.

4.5.12. Основы проектирования ИАС. Рассматривается основная задача проектирования – обеспечение раскрытия знаний, содержащихся в данных, сосредоточенных в первичных источниках и информационном хранилище. Перечисляются условия превращения данных в информацию и знания как принципы проектирования ИАС. Раскрывается содержание этапов проектирования информационной системы применительно к созданию ИАС. Рассматриваются факторы, которые необходимо учесть при проектировании ИАС.

Изучив данную тему, студент должен

знать: задачи пользователя по созданию ИАС, назначение и состав базы метаданных (БМД); задачи и средства обеспечения применения (администрирования) ИАС.

уметь: выполнять задачи пользователя ИАС при выработке требований к характеристикам ИАС, в процессе проектировании базы метаданных, выборе и конфигурировании инструментальных средств создания и поддержки ИАС.

приобрести навыки анализа и моделирования предметной области с целью построения базы метаданных и структуры ИАС, наполнения ИХ и транспортировки данных из первичных источников в ИХ.

При изучении темы 6 необходимо:

Читать учебное пособие [51] главу 6, книгу [13] главу 2 стр. 53–95, главу 3 стр. 101–136, главу 4 стр. 139–163, главу 5 стр. 169–196, главу 6 стр. 201–221; учебно-справочное издание [23] главу 3 стр. 38–52, главу 9 стр. 107–125; учебно-справочное издание [50] главу 4 стр. 44–58, главу 5 стр. 60–78.

Сосредоточить внимание на следующем:

При создании ИАС, в том числе базы метаданных, структуры информационного хранилища максимально возможно удовлетворять потребности конечных пользователей по обоснованию решений; учитывать то, что необходимо тщательно строить модели данных, структуру запросов, стараясь сочетать требования к системе с располагаемыми ресурсами.

Для самооценки по теме 6 ответить на вопросы

1. В чем сущность управления информационно-аналитической системой?
2. На какой структуре сведений о предметной области основано управление ИАС?
3. В чем содержание понятия «метаданные» и какое значение они имеют при создании и поддержке ИАС?
4. В чем заключается методика создания базы метаданных?
5. В чем заключаются задачи администрирования ИАС?
6. Раскройте содержание задач загрузки и обновления данных в ИХ.

7. Раскройте содержание планирования работы ИАС.
8. В чем содержание организации и осуществления эксплуатации ИАС?
9. В чем состоит основная задача проектирования ИАС?
10. Каковы условия превращения данных в знания?
11. Опишите содержание этапов проектирования ИАС.
12. Какие факторы необходимо учитывать при проектировании ИАС?

5. Для проведения итогового контроля необходимо:

Изучить вопросы по дисциплине «Информационно-аналитические системы».

1. Основные задачи, которые выполняют ИАС.
2. Роль и место анализа в принятии решений.
3. Проблемы анализа в свете использования информационных технологий.
4. Содержание аспекта сбора и хранения данных.
5. Содержание аспекта анализа данных и предоставления результатов анализа пользователям.
6. Классификация средств выполнения анализа с помощью ИТ.
7. Состав информационных технологий и информационных систем на предприятии и из внешней среды – источников данных для сосредоточения в информационном хранилище или непосредственно для анализа.
8. Понятие и структура информационного пространства.
9. Элементы структуры информационного пространства.
10. Понятия показателя и реквизитов.
11. Пространственная интерпретация понятия показатель.
12. Содержание экономических показателей.
13. Виды систем экономических показателей.
14. Рекомендации по структуризации информационного пространства предприятия при создании ИАС.
15. Содержание экономического анализа.
16. Сущность системы оценок.
17. Принципы гибкой архитектуры данных и открытых систем, которыми руководствуются при создании ИАС.
18. Информационный обмен, связанный с аналитической работой.
19. Понятие информационного хранилища.
20. Принципы построения информационных хранилищ.
21. Требования к качеству данных и способы его обеспечения при загрузке в информационное хранилище.
22. Проблемы, разрешаемые при приведении данных к единой структуре информационного хранилища.
23. Концепции построения структур хранилищ данных.
24. Назначение, состав и выполняемые функции базы метаданных – репозитория ИХ.
25. Принципы создания репозитория ИХ.
26. Элементы моделей данных ИХ (факт-таблица, таблицы измерений, консольные таблицы).
27. Схемы представления – модели многомерных данных.
28. Признаки OLAP-систем.
29. Типы многомерных OLAP-систем.

30. Классификация ИТ-анализа по режиму и темпу.
31. Задачи и содержание оперативного (OLAP) анализа.
32. Содержание понятия «знания», классификация видов знаний.
33. Интеллектуальный анализ данных (Data mining), цели и решаемые задачи.
34. Состав и содержание специфических задач интеллектуального анализа.
35. Классификация методов анализа.
36. Содержание методов анализа в экономической предметной области.
37. Состав программных инструментальных средств ИАС.
38. Средства сбора и доработки данных
39. Средства оперативного OLAP– анализа.
40. Средства интеллектуального анализа данных.
41. Управление информационно-аналитическими системами.
42. Задачи и средства администрирования ИАС.
43. Технологии загрузки данных в информационное хранилище.
44. Содержание планирования работы ИАС.
45. Принципы и этапы проектирования ИАС.
46. Рынок инструментальных средств ИАС.

Выполнить типовое задание:

Темы лабораторных работ по созданию предметных секторов информационного хранилища и системы анализа сосредоточенных в нем данных изложены в практикуме по данной дисциплине.

Практикум по курсу

ВВЕДЕНИЕ В ПРАКТИКУМ

Назначение практикума по курсу «Информационно-аналитические системы» для студентов, обучающихся по специальностям «Менеджмент, Маркетинг, Антикризисное управление, Финансы и кредит», привить навыки выполнения анализа данных с помощью средств информационно-аналитических систем (ИАС) в соответствующих специальностям предметных областях, участия в создании ИАС.

В соответствии с этим занятия проводятся по двум направлениям:

1. Выполнение заданий по анализу массивов данных с использованием массовых и специализированных инструментальных средств методами оперативного и интеллектуального анализа.
2. Решение задач, которые стоят перед пользователями, в процессе создания и развития ИАС.

Работы могут выполняться студентами индивидуально или в составе группы из 2–3 человек самостоятельно или при консультировании преподавателем.

1. Краткое содержание курса «Информационно-аналитические системы»

1.1. Роль и место анализа в процессе принятия решений

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

Одной из первостепенных задач при подготовке и принятии решений является анализ имеющейся в распоряжении лиц, принимающих решения (ЛПР), данных. Потребности в быстром и высококачественном обеспечении принятия решений, а также высокий уровень развития информационных технологий привели к созданию отдельного направления информационных систем –

информационно-аналитических систем (ИАС), под которыми понимают комплекс аппаратных, программных средств, информационных ресурсов, методик, которые используются для обеспечения автоматизации аналитических работ в целях обоснования принятия управленческих решений и других возможных применений.

1.2. Задачи информационно-аналитических систем

Сбор, организация и обеспечение хранения необходимых для анализа данных в целях обоснования принятия решений; обеспечение анализа имеющихся в распоряжении данных, в том числе в режимах оперативного и интеллектуального анализа; подготовка результатов анализа, выполненного в соответствии с необходимостью в том или ином режиме, для эффективного восприятия потребителями.

1.2.1. Сбор данных, необходимых для принятия решений

С этой целью используются различные источники данных, в том числе:

– внутренние источники – специализированные информационные системы технологического, экономического, организационного назначения данного предприятия; коммерческие и иные базы данных; автоматизированные рабочие места.

– внешние источники – вышестоящие организации, в том числе государственные, партнеры по деятельности, мировые информационные ресурсы и другие.

В процессе сбора данных решают следующие задачи:

– прокладка путей от источника до центра сбора, хранения и использования данных – информационного хранилища;

– преобразование данных в соответствии с интерфейсами обмена между источниками и центром;

– приведение данных к единому формату, принятому в системе;

– доработка данных, состоящая в проверке достоверности, устранении противоречивости, сортировке, систематизации.

1.2.2. Информационные хранилища (ИХ), правила построения и функционирования (правила Инмона)

ИХ – это аппаратно-программная система, являющаяся модулем ИАС и предназначенная для длительного хранения данных, готовая их предоставить в любое время в целях проведения разного рода аналитических работ. ИХ строятся в соответствии с правилами Инмона, которые состоят в следующем:

– предметная ориентированность – компоновка пулов данных по предметным областям.
– интегрированность – сбор и доработка данных из различных источников и превращение их в организованный по заданным правилам массив.
– неизменчивость – данные не подвергаются частым обновлениям, а лишь наращиваются по заранее определенному графику, пользователь имеет право только чтения.
– поддержка хронологии – данные рассматриваются в историческом аспекте от прошлого к будущему, обязательна их привязка каким-либо способом ко времени.

1.3. Выполнение анализа

Выполнение анализа производится в двух основных режимах:
– оперативный анализ, широко распространено англоязычное название в виде аббревиатуры OLAP (On Line Analytical Processing).
– интеллектуальный анализ или Data mining.

Данные, содержащиеся в информационном хранилище, могут быть использованы для формирования типовых отчетных документов. Такой вариант использования данных называют Reporting.

1.3.1. Требования к оперативному анализу

Исходя из правил доктора Кодда, сформулирован перечень требований к информационным системам, называемый тест FASMI, которые обеспечивают режим оперативного анализа.

Название теста является аббревиатурой английских слов – формулировок требований к OLAP-системам:
Fast – быстрый, Analysis – анализ, Shared – разделяемой, Multidimensional – многомерной, Information – информации.

Смысл требований состоит в следующем:

Быстрый – это требование ко времени отклика на запросы к OLAP-системе, допускается в пределах 5 (пяти) секунд.

Анализ – должен выполняться без применения каких-либо языковых средств, обеспечивается выполнение непосредственно силами топ-менеджеров.

Разделяемый – обеспечение необходимого уровня защиты данных от несанкционированного доступа, от взаимных помех при многопользовательском доступе. Организуется комплексом организационно-технических мероприятий и программных средств.

Многомерный – средства OLAP-системы должны обеспечить работу с данными в многомерном представлении данных – естественном на уровне ненормализованной ER-модели с полной поддержкой иерархий независимо от того, какие типы баз данных используются в качестве источников.

Информация – должна обеспечиваться возможность получения данных из любых источников, операции производятся с любыми объемами и структурами данных.

1.3.2. Интеллектуальный анализ

Часто применяется англоязычный термин – Data mining. Для извлечения знаний из массивов данных, содержащихся в информационных хранилищах или получаемых напрямую из первоисточников, применяются методы и средства интеллектуального анализа.

Задачами интеллектуального анализа являются:
– выявление взаимозависимостей, причинно-следственных связей, ассоциаций и аналогий, определение значения фактора времени, локализация событий или явлений по месту.
– классификация событий и ситуаций, материальных и других объектов по совокупностям признаков, определение профилей различных факторов.
– прогнозирование событий, хода процессов.
– оценка эффективности деятельности, проектов.

1.4. Представление результатов анализа

Осуществляется с использованием методов инженерной, управленческой психологии с целью достижения максимального эффекта восприятия полученных знаний и/или воздействия на клиента, партнера и т.д.

Для этой цели используются все имеющиеся средства отображения, 3D-графика, анимации, в том числе мультимедиа.

Результаты анализа могут экспортироваться в специализированные средства поддержки принятия решений.

2. Создание и применение ИАС

Задачи, выполняемые пользователями системы – лицами, принимающими решения (ЛПР), аналитиками, экспертами на этапах ее создания и применения заключаются в выработке требований к содержанию и структуре данных в ИХ, контроле за наполнением хранилищ данных, разработке и реализации сценариев анализа.

Главным в формировании требований является определение состава данных, которые необходимо собирать в ходе деятельности предприятия из различных источников, приводить их к единому формату и структуре с целью удобного использования в процессе анализа. На следующем этапе работы пользователей по созданию ИАС определяется, что нужно делать с этими данными.

Задача специалистов по информационным технологиям заключается в реализации требований пользователей в виде конкретных физических моделей.

Состав данных представляется в виде ***базы метаданных*** (БМД) или репозитория. БМД представляет собой совокупность свойств объектов, подлежащих исследованию в ходе аналитических работ. Практически это совокупность атрибутов или реквизитов показателей, отображающих объекты. Она создается по результатам обследования предметной области, в которой создается ИАС или ее фрагмент, в том числе информационное хранилище или его секция.

В результате обследования пользователи должны получить состав документов или исходных материалов для анализа и последующего использования при принятии решений. Из таких источников извлекаются атрибуты, подлежащие переносу в информационное хранилище или в секцию его. Повторяющиеся атрибуты вносятся в базу МД один раз, но добавляются атрибуты, свидетельствующие об агрегации данных. На рис. 1 показана схема создания базы метаданных, касающихся конечных пользователей. К ним добавляются метаданные, относящиеся к компетенции администраторов информационных хранилищ и ИАС в целом, так называемые технические метаданные в отличие от бизнес-метаданных. Среди таблиц на рис. 1 могут быть и материалы, относящиеся к технической стороне реализации ИХ и ИАС.

Наименования столбцов или атрибутов в таблицах, представляющих материалы или отчеты, обозначают свойства описываемых объектов. В строках содержатся значения свойств экземпляров этих объектов.

Делается оценка количества экземпляров содержащихся в данной таблице объектов и размерности их атрибутов с целью определения объема требуемой памяти.

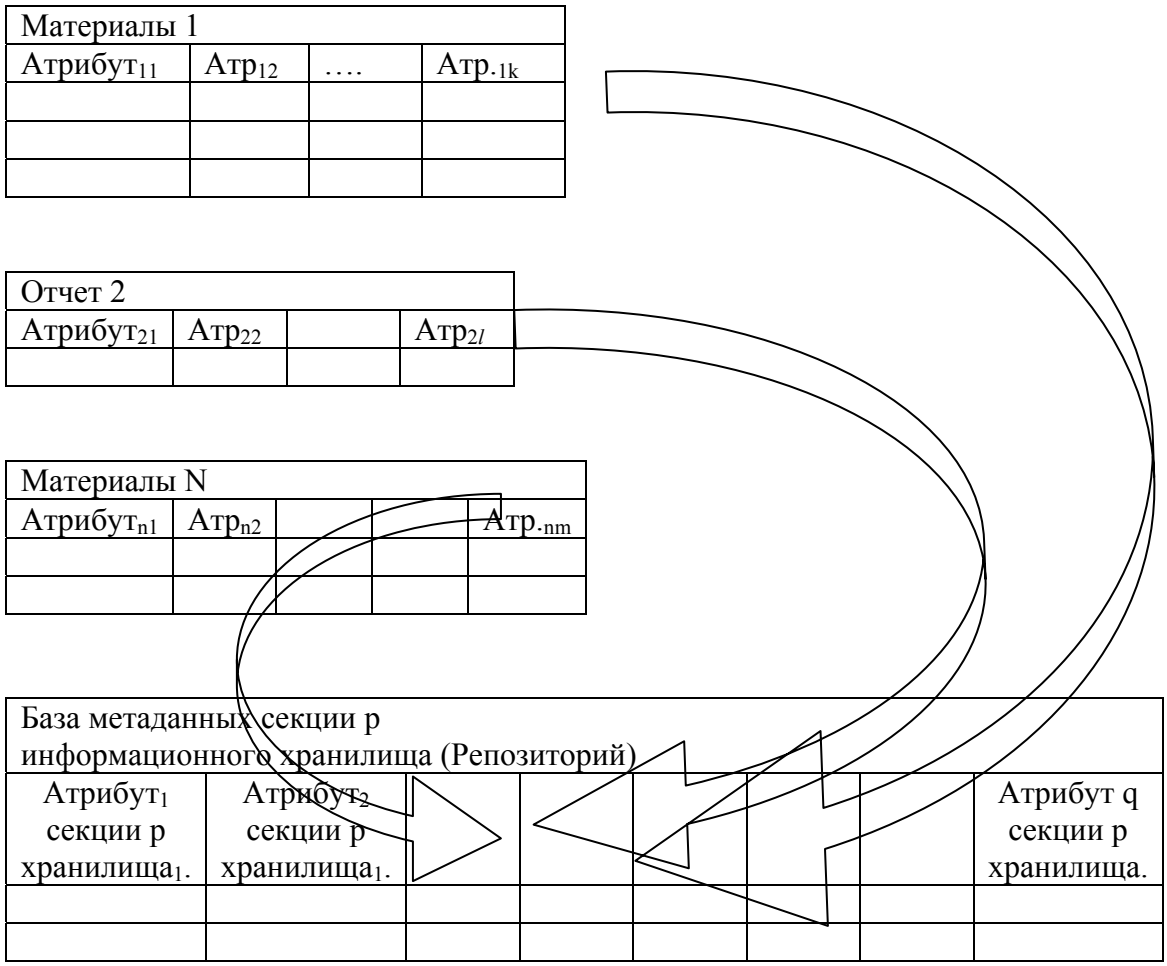

Рис. 1. Структура базы метаданных

Множество Q атрибутов секции P информационного хранилища определяется по формуле:

$$Q = \bigcup_{1}^{N} A_i \cup B \text{, где:}$$

A_i – множества атрибутов отчетов и материалов.

N – количество отчетов и материалов.

B – множество системообразующих атрибутов.

3. Задания по лабораторной работе

Для приобретения навыков аналитической работы с использованием специализированных и массовых инструментальных средств выполняются следующие задания.

Задание 1

Подобрать массивы данных по выбранной предметной области в соответствии с интересами по осваиваемой специальности, работе или другими обстоятельствами. При использовании Internet можно воспользоваться сайтами www//Рамблер.ru, www//Expert.ru, Яndex.ru и другими.

Экспортировать заинтересовавшие обучающегося данные в Excel или специализированное инструментальное средство в зависимости от наличия такового.

Транспорт в аналитический инструмент выполняется средствами Интернетэксплорер или опцией Excel (при транспорте в него): Данные – Внешние данные – Создать запрос.

Далее в зависимости от характера источника данных: Интернет; базы данных, поддерживаемые различными СУБД; электронные таблицы; текстовые файлы и т.д. выбираются соответствующие этим форматам данных опции.

После транспортировки в Excel форматы, вид и структура полученных данных приводятся к удобному для пользователя виду.

В случае, если данные получены в виде нескольких таблиц, их следует при необходимости свести в одну сводную таблицу с помощью опции Данные – Сводная таблица. Эта операция выполняется для того, чтобы можно было легко получать выборки, состоящие из атрибутов, находящихся в разных таблицах.

Задание 2.

На следующем этапе производится *оперативный анализ* подготовленных данных. При использовании инструмента Excel с помощью опций Данные –> Сортировка, Фильтр, Форма, Итоги, Группа и структура. Приведенные опции могут использоваться в последовательности, обеспечивающей выполнение поставленных пользователем задач анализа.

К ним относятся:

– выделение из подготовленной совокупности данных заинтересовавших пользователя объектов в виде группы атрибутов,
– сортировка значений атрибутов по определенным критериям,
– подведение итогов,
– группировка (агрегация) значений по заданным формулам и другие операции оперативного анализа.
– операции над атрибутами с целью получения новых свойств исследуемых объектов и т.д.

Выполнение *интеллектуального анализа* средствами Excel. В этом средстве имеется инструмент «Мастер функций», позволяющий выполнять математическую, статистическую, логическую обработку данных, финансовый анализ с оценкой проектов современными методами, прогнозирование хода процессов. Операции над сформированными данными выполняются с использованием аппарата справок. Сначала решаются находящиеся в справках примеры, затем выполняется анализ собственных примеров.

Задание 3.

При использовании специализированного инструмента Business Objects на начальном этапе приобретения навыков анализа используются данные примеров, содержащиеся

в самом инструменте, скомпонованные в виде факт-таблиц или, как называют на жаргоне фирмы, «юниверсов».

Запустить Business Objects, среди модулей Business Miner, Business Objects Designer, Business Objects (BO) вызвать BO; на стандартной панели инструментов нажать кнопку «Мастер Создания Нового Отчета». Будет вызвано окно «Мастер построения новых отчетов. Добро пожаловать» Рис. П1. Выбрать опцию «Создать стандартный отчет» – «Начать». Появится окно «Мастер построения новых отчетов. Выбор юниверса». Рис. П2.

Целесообразно выбрать юниверс Evaluation Kit, так как в его составе имеются и качественные и количественные атрибуты, здесь они называются объекты, которые понадобятся при выполнении упражнения.

Нажать на кнопку «Готово» – будет вызвана панель запроса в виде графического конструктора отчетов. Рис. П3.

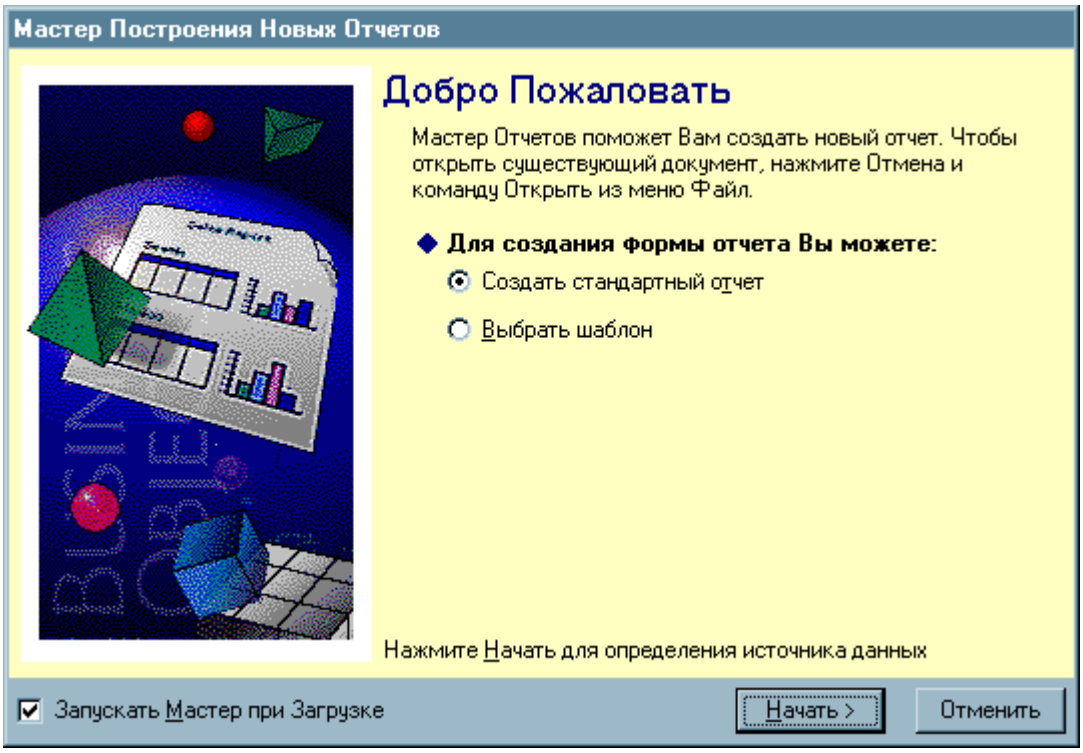

Рис. П1.

Запросы с исполнением в виде выборок с различными условиями легко реализуются перетаскиванием графических образов атрибутов – качественных в виде кубиков и количественных в виде шариков, называемых здесь объектами, в панель «объекты результата».

Для постановки условий выборок в окне «Условия» Панели Запросов щелкните на значении (или значениях), находящемся на правой стороне условия. Окно «Классы и Объекты» становится окном «Операнды». Затем нужно дважды щелкнуть на операнде «Показать Список Значений», после чего список значений появится в диалоговом окне «Список Значений».

Если нужно выбрать больше значений для условия, то удерживайте клавишу «Ctrl» нажатой и в окне «Список Значений» щелкните на каждом значении, затем нажмите «OK».

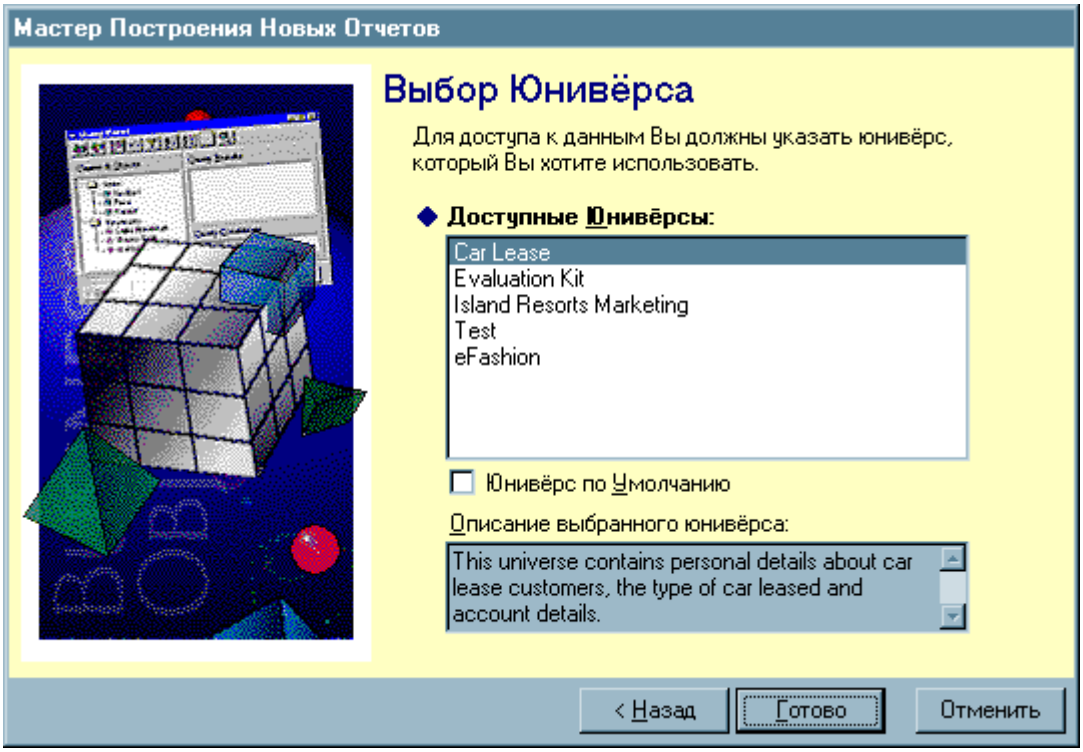

Рис. П2.

Созданные с помощью различных инструментов материалы могут быть помещены для дальнейшей аналитической обработки в среду Business Objects, для чего используется панель инструментов «Доступ к персональным данным» рис. П4.

Анализ перенесенных в среду Business Objects данных производится в главной панели, фрагмент которой приведен на рис. П5.

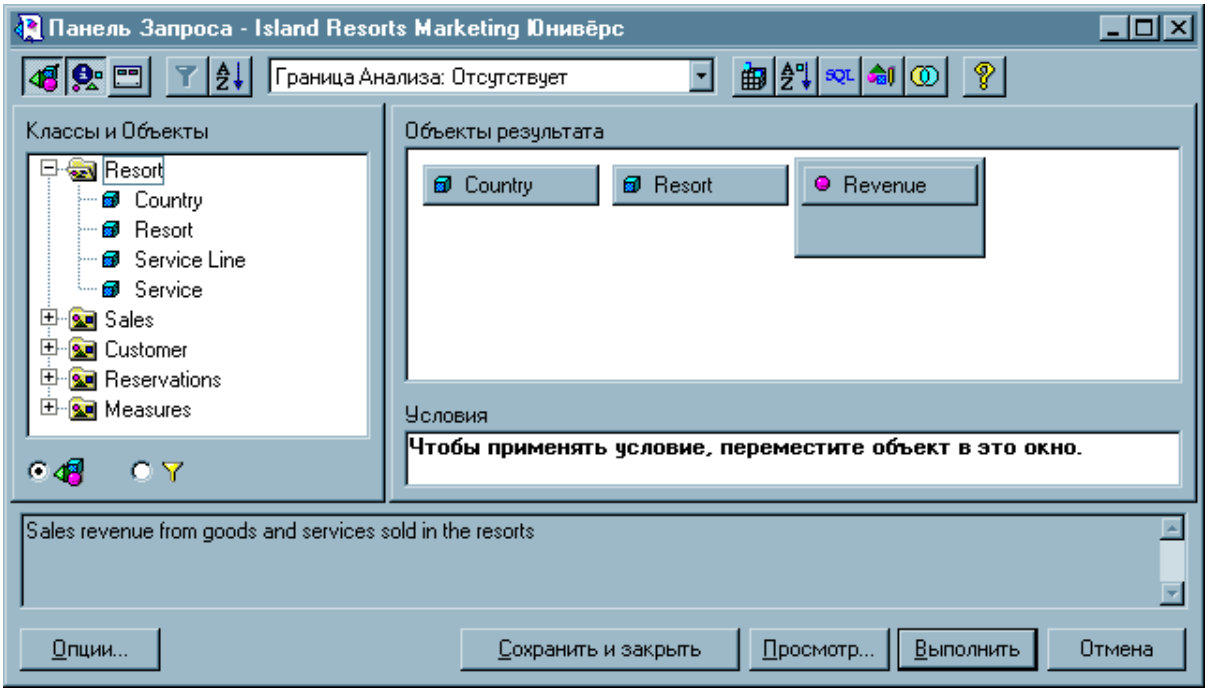

Рис. П3 .

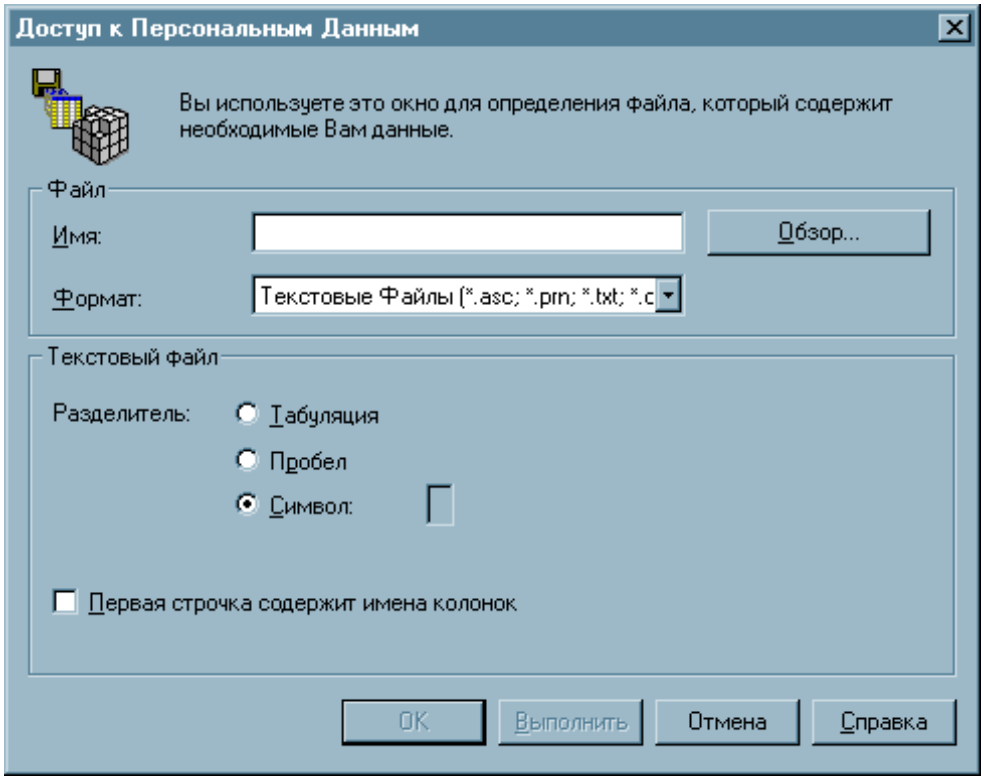

Рис. П4. Экранная форма обеспечения доступа к персональным данным

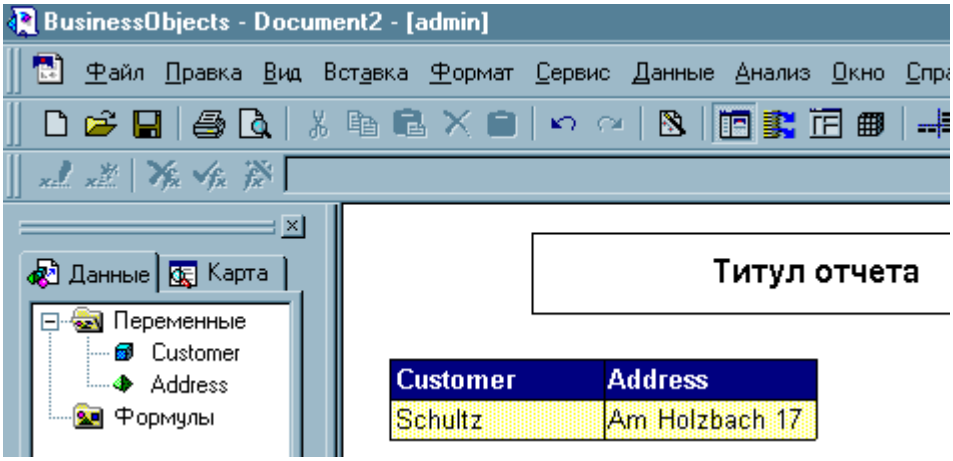

Рис. П5. Главная панель анализа Business Objects.

4. Тематика лабораторных работ

Темы работ выбираются в соответствии со специализацией и интересами обучающихся. В качестве ориентира на рис. 2 приведен примерный состав центров генерации данных в информационном пространстве предприятия в соответствии с выполняемыми функциями. Для специализаций менеджмент, антикризисное управление, финансы и кредит бригада подбирает из приведенного перечня несколько функций из одного направления деятельности: управления, основной деятельности (производство или торговля или другая сфера деятельности), логистики, обеспечения. Для специализаций маркетинг, мировая экономика подбираются темы материалов: обзор рынка товаров (услуг), исследования конкурентов по направлениям деятельности предприятия, привлекательность своих и конкурирующих изделий и т.д.

При очной форме обучения бригады подбирают из рекомендуемого перечня функций смежные так, чтобы в результате вся группа представила по возможности полный перечень функций предприятия. В результате должна получиться модель содержания и структуры данных информационного хранилища. Направление деятельности предприятия согласуется бригадами.

Технология создания базы метаданных описана в разделе 2 настоящего практикума. В соответствии с ней разрабатываются проекты документов пользовательского назначения. Это могут быть регулярно представляемые отчеты о деятельности подразделений или ответственных лиц, а также аналитические запросы руководителей предприятия по сценариями т.д. Затем атрибуты таких материалов должны войти в базу метаданных информационного хранилища. К ним дополняются системообразующие атрибуты, которые описывают состав источников данных, общие временные, региональные и другие характеристики.

Рис. Центры генерации данных в информационном пространстве предприятия

Следующим этапом работы является наполнение структуры модели ИХ учебными данными, которые могут быть почерпнуты из доступных источников, средств массовой информации, с предприятий (для работающих студентов) или смоделированы, исходя из предположений и знаний, полученных при изучении соответствующих курсов.

Требования к оформлению отчета по лабораторной работе

Отчет по лабораторной работе должен быть представлен в виде документа на бумажном носителе, выполнен по правилам оформления лабораторных работ. Он должен содержать:

1. Описание выбранной предметной области. Приводятся выполняемые задачи, объемы работ.

2. Макеты отчетных документов по выбранной группой предметной области в различных форматах, например xls, mdb, htm и/или других. На группу 2–3 человека – 5–6 документов. Для индивидуальной работы 2–3 документа. В документах должны быть макеты данных в объеме не менее десяти записей.

3. Фрагмент базы метаданных, относящийся к выбранной предметной области, содержащий атрибуты макетов, представленных по п. 2, а также макеты соответствующих данных в скоординированном виде. В совокупности фрагмент БМД и макетов записей составляют секцию информационного хранилища, относящуюся к исследуемой предметной области.

4. Макеты отчетных аналитических материалов. По 2–3 материала на группу из 2–3 человек. При индивидуальной работе – 1–2 материала. Аналитические отчеты должны сопровождаться краткими комментариями и диаграммами.

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., 2001.
3. Мишенин А.И. Теория экономических информационных систем. – М.: Финансы и статистика, 2000.
4. Королев 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. Бойченко А.В., Кондратьев В.К., Филинов Е.Н. Основы открытых информационных систем. Учебное пособие. – М., 2001.
8. Дубров А.М., Мхитарян В.С. Многомерные статистические методы и основы эконометрики. – М., 1998.
9. Ясин Е.Г. и др. Экономическая информация. – М. 1974.
10. Дейт К. Дж. Введение в системы баз данных. – М., СПБ, К., 2000.
11. .Саймон А. Р. Стратегические технологии баз данных. – М., 1999.
12. Хаан Д. ПиК Планирование и контроль: концепция контроллинга. – М., 1997.
13. Спирли, Эрик. Корпоративные хранилища данных. Планирование, разработка, реализация. Том 1.: Пер. с англ. – М.: Издательский дом «Вильямс», 2001.
14. Данилочкина Н.Г. и др. Контроллинг как инструмент управления предприятием. – М., 1999.
15. Карминский А.М. и др. Контроллинг в бизнесе. – М., 1998.
16. Макарова Н.В. и др. Информатика. Учебник. – М., 1998.
17. Гаврилова Т.А. Хорошевский В. Ф.: Базы знаний интеллектуальных систем. – СПб, Питер, 2001.
18. Дюк В., Самойленко А. Data mining:Учебный курс. – Питер, 2001.
19. Боровиков В.П. Программа Statistica для студентов и инженеров. – М., 2001.
20. Андрейчиков А.В., Андрейчикова О.Н. Анализ, синтез, планирование решений в экономике. – М., 2000.
21. Светлов Н.М. Руководитель. Российское программное обеспечение 2000/2001. – М., 2000.
22. Плис А.И., Сливина Н.А. Mathcad: математический практикум для экономистов и инженеров. – М., 1999.
23. Архипенков С.Я. Аналитические системы на базе Oracle Express OLAP. Проектирование, создание, сопровождение. – М., 2000.
24. Бритов П.А., Липчинский Е.А. Практика построения хранилищ данных: Система SAS. Ж. Системы управления базами данных. 04-05/1998/
25. Business Objects Руководство пользователя. Материалы фирмы Терн, 1996.
26. Калянов Г.Н. Консалтинг при автоматизации предприятий: научно-методическое издание. – М.: СИНТЕГ, 1997.
27. Кузьмичева Т.С. Контур Стандарт. Технология работы с системой. Руководство пользователя. ISL. – М., 1999.
28. Ивантер А. Деловой барометр показывает «Ясно» ж. «Эксперт» № 29, авг. 2001.
29. Корнев В.В., Гареев А.Ф., Васютин С.В., Райх В.В. Базы данных. Интеллектуальная обработка информации. – М., 2000.

30. Маклаков С.В. Bpwin Erwin. CASE-средства разработки информационных систем. – М., 2000.
31. Маклаков С.В. Создание информационных систем с AllFusion Modeling Suite. – М.:ДИАЛОГ-МИФИ, 2003.
32. Харрингтон Д. Проектирование объектно-ориентированных баз данных / Пер. с англ. – М.: ДМК Пресс, 2001.
33. Кирстен В., Ирингер М., Рериг Б., Шульте П. СУБД Cache: объектно-ориентированная разработка приложений. Учебный курс. – СПб.: Питер, 2001.
34. Замков О.О., Толстопятенко А.В., Черемных Ю.Н. Математические методы в экономике: Учебник / Под ред. д.э.н., проф. Сидоровича. – М.: Изд-во «Дело и сервис», 2001.
35. Вендров А.М. Проектирование программного обеспечения экономических информационных систем. Учебник. – М: Финансы и статистика. 2000.
36. Вендров А.М. Практикум по проектированию программного обеспечения экономических информационных систем: Учебное пособие. – М. Финансы и статистика, 2002.
37. Вебер А.В., Данилов А.Д., Шифрин С.И. Knoulidge-технологии в консалтинге и управлении предприятием / Под редакцией Финикова М.В. – СПб: Наука и техника, 2003.
38. Под ред. Максимовой В.Ф. Макроэкономика. – М., 2001.
39. Пендс Н. Перевод Абушаева Ш. Что относится к OLAP? OLAPReport.com Last updated on Jan 1999.
40. Щавелев Л.В. Способы аналитической обработки данных для поддержки принятия решений. – Ж. СУБД 04-05/1998/
41. Белов В.С. Информационно-аналитические системы. Учебное пособие МЭСИ. – М., 2001.
42. Шрек М. Перевод Intersoft Lab Технология Хранилищ данных: последние 10 лет, несомненно были впечатляющими http//www/iso/ru 08/01.
43. Елманова Н., Федоров А. Введение в OLAP: Основы OLAP. 28. Хранилища данных. КомпьютерПресс 5/2001.
44. Уринцов А.И. Основы электронного обмена данными. Учебное пособие. – М.: МЭСИ, 2004.
45. Островский Е.В. Порядок разработки ETL-процессов. Olap.ru. 2000.
46. Косов Д.В. Хранилища данных и семантические разрывы. Olap.ru. 2000.
47. Коломиец С.В. Проектирование процессов перегрузки данных. Olap.ru. 2000.
48. Федечкин С. Хранилище данных: вопросы и ответы. PCWeek, 31, 26.08.2003.
49. Строддер Д., Бурьесси Ж., Янг М. Лучшие ИТ-поставщики на 2004 год. Intelligent enterprise. Деловой ИТ-журнал. 26.01.2004.
50. Федоров А.Г., Елманова Н.З. Введение в OLAP-технологии Microsoft. Учебно-справочное издание. – М.: Диалог-МИФИ, 2002.
51. Белов В.С. Информационно-аналитические системы. Основы проектирования и применения. Учебное пособие. – М: МЭСИ, 2004.

Интернет-ресурсы

1. http://www.olap.ru – журнал СУБД
2. http://www.tern.ru – компания ТЕРН
3. http://www. iso.ru – компания Intersoftlab
4. http://www.sas.ru – SAS Institute
5. http://www.basegroup.ru – компания Basegroup РФ
6. http://www.banklist.ru – ЦБ РФ
7. http://www.megaputer.ru комания Megaputer РФ
8. http://www.expert.ru – журнал Эксперт РФ
9. http://www.relex.ru – компания РЕЛЭКС РФ