

АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ ОБРАЗОВАТЕЛЬНАЯ
ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
ЦЕНТРОСОЮЗА РОССИЙСКОЙ ФЕДЕРАЦИИ
«РОССИЙСКИЙ УНИВЕРСИТЕТ КООПЕРАЦИИ»
КАЗАНСКИЙ КООПЕРАТИВНЫЙ ИНСТИТУТ (ФИЛИАЛ)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

ХРАНИЛИЩА ДАННЫХ

Направление подготовки: 38.03.05 Бизнес-информатика

Направленность (профиль): «Электронный бизнес»

Форма обучения: очная, заочная

Квалификация (степень) выпускника: бакалавр

Срок обучения: очная форма - 4 года, заочная форма – 4 года 6 мес.

Вид учебной работы	Трудоемкость, часы (з.е.)	
	Очная форма	Заочная форма
1. Контактная работа обучающихся с преподавателем:	36(1)	10(0,28)
Аудиторные занятия, часов всего, в том числе:	36(1)	8(0,22)
• лекции	18(0,5)	4(0,11)
• лабораторные	18(0,5)	4(0,11)
Промежуточная аттестация (контактная работа)	-	2(0,06)
2. Самостоятельная работа студентов, всего	72(2)	96(2,67)
• др. формы самостоятельной работы	72(2)	96(2,67)
3. Промежуточная аттестация: зачет с оценкой	-	2(0,06)
Итого	108(3)	108(3)

Хрущева Л.Б. Хранилища данных: Рабочая программа дисциплины (модуля). – Казань: Казанский кооперативный институт (филиал) Российского университета кооперации, 2018. – 85 с.

Рабочая программа по дисциплине (модулю) «Хранилища данных» по направлению подготовки 38.03.05 Бизнес - информатика, направленность «Электронный бизнес» составлена Хрущевой Л.Б., старшим преподавателем кафедры естественных дисциплин, сервиса и туризма Казанского кооперативного института (филиала) Российского университета кооперации в соответствии с требованиями Федерального государственного образовательного стандарта высшего образования по направлению подготовки «Бизнес - информатика», утвержденного приказом Министерства образования и науки Российской Федерации от «11» августа 2016 г. № 1002, и учебными планами по направлению подготовки 38.03.05 Бизнес-информатика, направленность (профиль) «Электронный бизнес» (год начала подготовки -2018).

Рабочая программа:

обсуждена и рекомендована к утверждению решением кафедры естественных дисциплин, сервиса и туризма Казанского кооперативного института (филиала) Российского университета кооперации от «10» мая 2018 г., протокол № 3.

одобрена Научно-методическим советом Казанского кооперативного института (филиала) от 23.05.2018, протокол № 5.

утверждена Ученым советом Российского университета кооперации от 30.05.2018, протокол № 7.

© АНОО ВО ЦС РФ
«Российский университет
кооперации» Казанский
кооперативный институт
(филиал), 2018
© Хрущева Л.Б., 2018

СОДЕРЖАНИЕ

1 Цели, задачи освоения дисциплины (модуля)	4
2. Место дисциплины (модуля) в структуре основной образовательной программы	4
3. Перечень планируемых результатов обучения по дисциплине (модулю)	4
4. Объем дисциплины (модуля) и виды учебной работы	5
5. Содержание дисциплины (модуля)	6
5.1. Содержание разделов, тем дисциплины (модуля)	6
5.2. Междисциплинарные связи с обеспечиваемыми (последующими) дисциплинами (модулями)	7
5.3. Разделы, темы дисциплины (модуля) и виды занятий	7
6. Лабораторный практикум	7
7. Практические занятия (семинары)	8
8. Примерная тематика курсовых проектов (работ)	8
9. Самостоятельная работа студента	8
10. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)	9
11. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)	10
12. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля)	10
13. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)	10
14. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)	11
15. Методические указания для обучающихся по освоению дисциплины (модуля)	11
16. Методические рекомендации по организации изучения дисциплины для преподавателей, образовательные технологии	12
ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)	14
1. Паспорт фонда оценочных средств	15
1.1. Компетенции, формируемые в процессе изучения дисциплины	15
1.2. Сведения об иных дисциплинах (преподаваемых, в том числе, на других кафедрах) участвующих в формировании данных компетенций	15
1.3. Этапы формирования и программа оценивания контролируемой компетенции	15
1.4. Показатели и критерии оценивания компетенций, шкала оценивания	17
2. Типовые контрольные задания для оценки результатов обучения по дисциплине и иные материалы для подготовки к промежуточной аттестации	22
2.1. Материалы для подготовки к промежуточной аттестации	22
2.2. Комплект билетов для проведения промежуточной аттестации	29
Комплект тестовых заданий для проведения зачета с оценкой по дисциплине	30
2.3. Критерии оценки для проведения зачета с оценкой по дисциплине	32
2.4. Методические материалы, определяющие процедуру оценивания по дисциплине	33
ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ И ТЕКУЩЕЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)	34
1. Материалы для текущего контроля	35
КОМПЛЕКТ ЗАДАНИЙ ДЛЯ КОНТРОЛЬНОЙ РАБОТЫ	35
ТЕМЫ ДОКЛАДОВ	40
ЛАБОРАТОРНАЯ РАБОТА	41
2. Материалы для проведения текущей аттестации	69
КОМПЛЕКТ ЗАДАНИЙ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ АТТЕСТАЦИИ №1	69

1 Цели, задачи освоения дисциплины (модуля)

Цель освоения дисциплины - формирование профессиональных компетенций, связанных с использованием теоретических знаний и практических умений и навыков в области управления хранилищами данных, достаточных для дальнейшего продолжения образования и самообразования в области разработки и эксплуатации систем деловой осведомленности.

Задачи освоения дисциплины:

- изучение основных методов и принципов проектирования хранилищ данных с применением современных информационных средств и технологий;
- изучение основных методов и принципов организации хранилищ данных с применением современных информационных средств и технологий;
- изучение основных методов и принципов оперирования хранилищами данных с применением современных информационных средств и технологий;

2. Место дисциплины (модуля) в структуре основной образовательной программы

Дисциплина относится к вариативной части дисциплин блока Б1 «Дисциплины (модули)»

Для изучения учебной дисциплины необходимы следующие знания, умения и владения навыками, формируемые предшествующими дисциплинами:

Теоретические основы информатики (ОПК-3);

Программирование (ОПК-3; ПК-18).

3. Перечень планируемых результатов обучения по дисциплине (модулю)

Изучение учебной дисциплины направлено на формирование у обучающихся следующей профессиональной компетенции:

ПК-13 - умение проектировать и внедрять компоненты ИТ-инфраструктуры предприятия, обеспечивающие достижение стратегических целей и поддержку бизнес-процессов.

В результате изучения дисциплины обучающийся должен:

Формируемые компетенции (код компетенции)	Планируемые результаты обучения по дисциплине (модулю), характеризующие этапы формирования компетенций	Наименование оценочного средства
ПК-13	Знать теоретические основы информатики	Доклад
	Знать рынки программно-информационных продуктов и услуг	
	Знать методы проектирования, внедрения и организации эксплуатации корпоративных ИС и ИКТ	
	Знать виды контента информационных ресурсов предприятия и Интернет-ресурсов, процессы управления жизненным циклом цифрового контента, процессы создания и использования информационных сервисов	
	Уметь систематизировать и обобщать информацию, организовывать и проводить исследования в области экономики, управления и ИКТ	Лабораторная работа
	Уметь управлять процессами жизненного цикла контента предприятия и Интернет-ресурсов, управлять процессами создания и использования информационных сервисов (контент-сервисов)	
	Уметь проектировать, внедрять и организовывать эксплуатацию ИС и ИКТ	
	Уметь проводить OLAP анализ для различных проблемных областей; формулировать запросы к многомерным базам данных с помощью языка MDX, строить модели DataMining;	Контрольная работа
	Владеть методами поиска, хранения и обработки информации	
	Владеть методами управления процессами жизненного цикла контента предприятия и Интернет-ресурсов;	
	Владеть методами проектирования, внедрения и организации эксплуатации ИС и ИКТ	
	Владеть методами управления и процессами создания и использования информационных сервисов (контент-сервисов)	

4. Объем дисциплины (модуля) и виды учебной работы

очная форма обучения

Вид учебной деятельности	Часов	
	Всего	По семестрам
1. Контактная работа обучающихся с преподавателем:		6
Аудиторные занятия всего, в том числе:	36	36
Лекции	18	18
Лабораторные работы	18	18
Промежуточная аттестация (контактная работа)	-	-
2. Самостоятельная работа студента всего, в том числе:	72	72
Другие виды самостоятельной работы	72	72
Вид промежуточной аттестации – зачет с оценкой	-	-
ИТОГО:	часов	108
Общая трудоемкость	зач. ед.	3

заочная форма обучения

Вид учебной деятельности	Часов	
	Всего	По курсам
1. Контактная работа обучающихся с преподавателем:	10	10
Аудиторные занятия всего, в том числе:	8	8
Лекции	4	4
Лабораторные работы	4	4
Промежуточная аттестация (контактная работа)	2	2
2. Самостоятельная работа студента всего, в том числе:	96	96
Другие виды самостоятельной работы	96	96
Вид промежуточной аттестации – зачет с оценкой	2	2
ИТОГО:	часов	108
Общая трудоемкость	зач. ед.	3

5. Содержание дисциплины (модуля)**5.1. Содержание разделов, тем дисциплины (модуля)****Тема 1. Введение в хранилища данных.**

1. Цели и задачи дисциплины, ее роль и место в общей системе подготовки специалиста. Основные понятия. История вопроса.

2. Понятие OLAP. Различия между транзакционными и аналитическими системами. Области применения хранилищ данных.

Тема 2. Общие свойства и структура хранилищ данных.

1. Общие свойства хранилищ данных.

2. Данные хранилища.

3. Компоненты хранилища.

Тема 3. Многомерные кубы.

1. Основные понятия кубов. Иерархии измерений. Структура ХД. Примеры кубов.

2. Способы хранения агрегатных данных.

Тема 4. Методология построения хранилищ данных.

1. Подходы к стратегии построения. Компонентная архитектура. Техническая архитектура.

2. Модели разработки. Этапы спиральной модели применительно к разработке хранилищ данных.

Тема 5. Методы реализации хранилищ данных.

1. Две группы аналитических платформ. Обзор рынка BI.

2. Продукция Microsoft. Продукция Sybase. Продукция Oracle.

Тема 6. Технология Data Mining.

1. Виды Общие понятия. История вопроса. Приложения. Data Mining в

MSSAS. инвестиций компании активами компании.

2. Технология (процесс) добычи знаний. Решаемые задачи.
3. Математические основы (РАД).

5.2. Междисциплинарные связи с обеспечиваемыми (последующими) дисциплинами (модулями)

Дисциплина «Хранилища данных» формирует ПК-13 компетенцию, необходимую в дальнейшем для формирования компетенций ПК-14; ПК-16

5.3. Разделы, темы дисциплины (модуля) и виды занятий

очная форма обучения

№ п/п	Наименование темы дисциплины (модуля)	Виды занятий, включая самостоятельную работу студентов (в часах)			
		Лекции	Лабораторные занятия	Самостоятельная работа	Всего
1.	Введение в хранилища данных.	2	-	8	10
2.	Общие свойства и структура хранилищ данных.	4	-	10	14
3.	Многомерные кубы.	4	4	12	20
4.	Методология построения хранилищ данных.	6	4	14	24
5.	Методы реализации хранилищ данных.	-	4	14	18
6.	Технология Data Mining.	2	6	14	22
	Итого	18	18	72	108

заочная форма обучения

№ п/п	Наименование темы дисциплины (модуля)	Виды занятий, включая самостоятельную работу студентов (в часах)			
		Лекции	Лабораторные занятия	Самостоятельная работа	Всего
1.	Введение в хранилища данных	1	-	16	17
2.	Общие свойства и структура хранилищ данных	1	-	16	17
3.	Многомерные кубы	1	1	16	18
4.	Методология построения хранилищ данных	1	1	16	18
5.	Методы реализации хранилищ данных.	-	1	16	17
6.	Технология Data Mining.	-	1	16	17
	Итого	4	4	96	104

6. Лабораторный практикум

Лабораторные занятия проводятся с целью формирования компетенций обучающихся, закрепления на практике полученных на лекциях и в процессе самостоятельного изучения обучающимися специальной литературы

очная форма обучения

№ п/п	Наименование темы дисциплины (модуля)	Тематика лабораторных занятий	Трудоемкость (час.)
1.	Многомерные кубы.	Примеры кубов. Способы хранения агрегатных данных.	4
2.	Методология построения хранилищ данных.	Спиральная модель процесса разработки хранилища данных.	4
3.	Методы реализации хранилищ данных.	Обзор рынка BI.	4
4.	Технология Data Mining.	Технологические, статистические и кибернетические методы Data Mining.	6
	Итого		18

заочная форма обучения

№ п/п	Наименование темы дисциплины (модуля)	Тематика практических занятий (семинаров)	Трудоемкость (час.)
1.	Многомерные кубы.	Примеры кубов. Способы хранения агрегатных данных.	1
2.	Методология построения хранилищ данных.	Спиральная модель процесса разработки хранилища данных.	1
3.	Методы реализации хранилищ данных.	Обзор рынка BI.	1
4.	Технология Data Mining.	Технологические, статистические и кибернетические методы Data Mining.	1
	Итого		4

7. Практические занятия (семинары)

Практические занятия не предусмотрены учебными планами.

8. Примерная тематика курсовых проектов (работ)

Курсовые проекты (работы) не предусмотрены учебными планами.

9. Самостоятельная работа студента

№ п/п	Наименование темы дисциплины (модуля)	Виды самостоятельной работы студента	Оценочное средство
1.	Введение в хранилища данных Рекомендации: Обратить внимание на отличие БД от Хранилищ данных	чтение литературы;	Доклад
2.	Общие свойства и структура хранилищ данных Рекомендации: Обратить внимание на подробное изучение вопроса о структуре хранилища данных	конспектирование текста с составлением схем и таблиц	Доклад
3.	Многомерные кубы Рекомендации: Обратить внимание на способы хранения агрегатных данных	работа с конспектом лекций; подготовка сообщений к выступлению на	Доклад Лабораторная работа

№ п/п	Наименование темы дисциплины (модуля)	Виды самостоятельной работы студента	Оценочное средство
		практическом занятии;	
4.	Методология построения хранилищ данных Рекомендации: Обратить внимание на модели разработки	ответы на контрольные вопросы;	Устный опрос Лабораторная работа
5.	Методы реализации хранилищ данных Рекомендации: Обратить внимание на обзор рынка BI.	решение задач по образцу	Лабораторная работа
6.	Технология Data Mining Рекомендации: Обратить внимание на технологию (процесс) добычи знаний	решение задач по образцу	Лабораторная работа

10. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

Самостоятельная работа студентов предполагает тщательное освоение учебной и научной литературы по изучаемой дисциплине.

При изучении основной рекомендуемой литературы студентам необходимо обратить внимание на выделение основных понятий, их определения, научно-технические основы, узловые положения, представленные в изучаемом тексте.

При самостоятельной работе студентов с дополнительной литературой необходимо выделить аспект изучаемой темы (что в данном материале относится непосредственно к изучаемой теме и основным вопросам).

Дополнительную литературу целесообразно прорабатывать после основной, которая формирует базис для последующего более глубокого изучения темы. Дополнительную литературу следует изучать комплексно, рассматривая разные стороны изучаемого вопроса. Обязательным элементом самостоятельной работы студентов с литературой является ведение необходимых записей: конспекта, выписки, тезисов, планов.

Для самостоятельной работы по дисциплине используется следующее учебно-методическое обеспечение:

а) основная литература

1. Агальцов В.П. Базы данных. В 2-х кн. Книга 2. Распределенные и удаленные базы данных : учебник / В.П. Агальцов. — М. : ИД «ФОРУМ» : ИНФРА-М, 2017. — 271 с. - Режим доступа: <http://znanium.com/catalog/product/652917>

2. Основы проектирования баз данных: Учебное пособие / Голицына О.Л., Партыка Т.Л., Попов И.И., - 2-е изд. - М.: Форум, НИЦ ИНФРА-М, 2016. - 416 с. - Режим доступа: <http://znanium.com/catalog/product/552969>

б) дополнительная литература

1. Култыгин, О. П. Администрирование баз данных. СУБД MS SQL Server: учеб. пособие / О. П. Култыгин. - М.: МФПА, 2012. - 232 с. - Режим доступа: <http://znanium.com/catalog/product/451114>

11. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

а) нормативные документы

1. Гражданский кодекс Российской Федерации (часть первая) от 30.11.1994 № 51–ФЗ (ред. от 26.07.2006) // Собрание законодательства РФ. - 1994. - № 32. – Ст. 3301. - Консультант Плюс. [Электрон. ресурс]. - Электрон. дан. - [М., 2016.]

б) основная литература

1. Агальцов В.П. Базы данных. В 2-х кн. Книга 2. Распределенные и удаленные базы данных : учебник / В.П. Агальцов. — М. : ИД «ФОРУМ» : ИНФРА-М, 2017. — 271 с. - Режим доступа: <http://znanium.com/catalog/product/652917>

2. Основы проектирования баз данных: Учебное пособие / Голицына О.Л., Партыка Т.Л., Попов И.И., - 2-е изд. - М.: Форум, НИЦ ИНФРА-М, 2016. - 416 с. - Режим доступа: <http://znanium.com/catalog/product/552969>

в) дополнительная литература

1. Култыгин, О. П. Администрирование баз данных. СУБД MS SQL Server: учеб. пособие / О. П. Култыгин. - М.: МФПА, 2012. - 232 с. - Режим доступа: <http://znanium.com/catalog/product/451114>

12. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля)

- <https://www.book.ru/> - ЭБС Book.ru
- <http://www.iprbookshop.ru> - ЭБС IPRbooks
- <https://ibooks.ru/> -ЭБС Айбукс.ru/ibooks.ru
- <https://rucont.ru/> - ЭБС «Национальный цифровой ресурс «Руконт»
- <http://znanium.com/> - ЭБС Znanium.com
- <https://dlib.eastview.com/>- База данных East View

13. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)

1. Desktop School ALNG LicSAPk MVL.
 - a. Office ProPlus All Lng Lic/SA Pack MVL Partners in Learning (лицензия на пакет Office Professional Plus)
 - b. Windows 8
2. Система тестирования INDIGO.
3. Adobe Acrobat – свободно-распространяемое ПО

4. Интернет-браузеры Google Chrome, Firefox – свободно-распространяемое ПО

5. Консультант + версия проф.- справочная правовая система

Каждый обучающийся в течение всего обучения обеспечивается индивидуальным неограниченным доступом электронно-библиотечной системе и электронной информационно-образовательной среде.

14. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Образовательный процесс обеспечивается специальными помещениями, которые представляют собой аудитории для проведения занятий лекционного типа, занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, помещения для самостоятельной работы студентов и помещения для хранения и профилактического обслуживания учебного оборудования.

Специальные помещения соответствуют действующим противопожарным правилам и нормам, укомплектованы специализированной мебелью.

Аудитории лекционного типа, оснащенные проекционным оборудованием и техническими средствами обучения, обеспечивающими представление учебной информации большой аудитории, демонстрационным оборудованием.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой, обеспечивающей доступ к сети Интернет и электронной информационно-образовательной среде университета.

15. Методические указания для обучающихся по освоению дисциплины (модуля)

Дисциплина «Хранилища данных» состоит из 6 тем и изучается на лекциях, лабораторных занятиях и при самостоятельной работе обучающихся. Обучающийся для полного освоения материала должен не пропускать занятия и активно участвовать в учебном процессе. Кроме того, обучающиеся должны ознакомиться с программой дисциплины и списком основной и дополнительной рекомендуемой литературы.

Основной теоретический материал дается на лекционных занятиях. Лекции включают все темы и основные вопросы теории и практики в области управления хранилищами данных. Для закрепления теоретического материала, формирования профессиональных компетенций и практических навыков в области разработки и эксплуатации систем деловой осведомленности со студентами бакалавриата проводятся лабораторные занятия. В ходе лабораторных занятий разбираются основные и дополнительные теоретические вопросы, решаются практические задачи,

проводятся тестирования по результатам изучения тем.

На изучение каждой темы выделено в соответствии с рабочей программой дисциплины количество часов лабораторных занятий, которые проводятся в соответствии с вопросами, рекомендованными к изучению по определенным темам. Обучающиеся должны регулярно готовиться к лабораторным занятиям. При подготовке к занятиям следует руководствоваться конспектом лекций и рекомендованной литературой.

Для эффективного освоения материала дисциплины учебным планом предусмотрена самостоятельная работа, которая должна выполняться в обязательном порядке. Выполнение самостоятельной работы по темам дисциплины, позволяет регулярно проводить самооценку качества усвоения материалов дисциплины и выявлять аспекты, требующие более детального изучения. Задания для самостоятельной работы предложены по каждой из изучаемых тем и должны готовиться индивидуально и к указанному сроку. По необходимости студент бакалавриата может обращаться за консультацией к преподавателю. Выполнение заданий контролируется и оценивается преподавателем.

В случае посещения обучающегося лекций и практических занятий, изучения рекомендованной основной и дополнительной учебной литературы, а также своевременного и самостоятельного выполнения заданий, подготовка к зачету с оценкой по дисциплине сводится к дальнейшей систематизации полученных знаний, умений и навыков.

16. Методические рекомендации по организации изучения дисциплины для преподавателей, образовательные технологии

Оценочные средства для контроля успеваемости и результатов освоения дисциплины (модуля):

- а) для текущей успеваемости: доклад, лабораторная работа, контрольная работа;
- б) для самоконтроля обучающихся: тесты;
- в) для промежуточной аттестации: вопросы для зачета с оценкой, лабораторные работы

При реализации различных видов учебной работы по дисциплине «Хранилище данных» используются следующие образовательные технологии:

- 1) лекции с использованием методов проблемного изложения материала;
- 2) лабораторные работы;

№	Занятие в интерактивной форме	Количество часов по очной форме		Количество часов по заочной форме	
		Лекция	Лаборат.	Лекция	Лаборат.
1	Введение в хранилища данных Вид: Лекции с использованием методов	-	-	1	-

№	Занятие в интерактивной форме	Количество часов по очной форме		Количество часов по заочной форме	
		Лекция	Лаборат.	Лекция	Лаборат.
	проблемного изложения материала; Лабораторные занятия: решение и защита лабораторных работ;				
2	Общие свойства и структура хранилищ данных Вид: Лекции с использованием методов проблемного изложения материала; Лабораторные занятия: решение и защита лабораторных работ;	1	-	1	-
3	Многомерные кубы Вид: Лекции с использованием методов проблемного изложения материала; Лабораторные занятия: решение и защита лабораторных работ;	1	2	1	1
4.	Методология построения хранилищ данных Вид: Лекции с использованием методов проблемного изложения материала; Лабораторные занятия: решение и защита лабораторных работ;	1	2	1	1
5.	Методы реализации хранилищ данных. Вид: Лекции с использованием методов проблемного изложения материала; Лабораторные занятия: решение и защита лабораторных работ;	-	2	-	1
6.	Технология Data Mining. Вид: Лекции с использованием методов проблемного изложения материала; Лабораторные занятия: решение и защита лабораторных работ;	1	2	-	1
	Всего	4	8	4	4

**АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ ОБРАЗОВАТЕЛЬНАЯ
ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
ЦЕНТРОСОЮЗА РОССИЙСКОЙ ФЕДЕРАЦИИ
«РОССИЙСКИЙ УНИВЕРСИТЕТ КООПЕРАЦИИ»
КАЗАНСКИЙ КООПЕРАТИВНЫЙ ИНСТИТУТ (ФИЛИАЛ)**

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ
ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО
ДИСЦИПЛИНЕ (МОДУЛЮ)**

ХРАНИЛИЩА ДАННЫХ

Направление подготовки: 38.03.05 Бизнес - информатика
Направленность: «Электронный бизнес»

1. Паспорт фонда оценочных средств

1. 1. Компетенции, формируемые в процессе изучения дисциплины

Индекс	Формулировка компетенции
ПК-13	умение проектировать и внедрять компоненты ИТ-инфраструктуры предприятия, обеспечивающие достижение стратегических целей и поддержку бизнес-процессов

1.2. Сведения об иных дисциплинах (преподаваемых, в том числе, на других кафедрах) участвующих в формировании данных компетенций

1.2.1 Компетенция ПК-13 формируется в процессе изучения дисциплин (прохождения практик):

Разработка ИТ- сервисов предприятия

Производственная практика. Практика по получению профессиональных умений и опыта профессиональной деятельности

1.3. Этапы формирования и программа оценивания контролируемой компетенции

№	Код контролируемой компетенции	Контролируемые темы дисциплины	Наименование оценочного средства
1	ПК-13	Тема 1. Введение в хранилища данных Тема 2. Общие свойства и структура хранилищ данных Тема 3. Многомерные кубы Тема 4. Методология построения хранилищ данных Тема 5. Методы реализации хранилищ данных Тема 6. Технология Data Mining	Доклад Лабораторная работа Контрольная работа

Процедура оценивания

1. Процедура оценивания результатов освоения программы учебной дисциплины включает в себя оценку уровня сформированности компетенций студента при осуществлении текущего контроля и проведении промежуточной аттестации.

2. Уровень сформированности компетенции (ПК-13) определяется по качеству выполненной студентом работы и отражается в следующих формулировках: высокий, хороший, достаточный, недостаточный.

3. При выполнении студентами заданий текущего контроля и промежуточной аттестации оценивается уровень обученности «знать», «уметь», «владеть» в соответствии с запланированными результатами обучения и содержанием рабочей программы дисциплины:

– профессиональные знания студента могут проверяться при ответе на теоретические вопросы, выполнении тестовых заданий, практических работ,
– степень владения профессиональными умениями – при решении ситуационных задач, выполнении практических работ и других заданий.

4. Результаты выполнения заданий фиксируются в баллах. Общее количество баллов (макс. - 15 б.) складывается из:

– 5 баллов (33,3% от общей оценки) за выполнение практических заданий на выявление уровня обученности «уметь»,

– 5 баллов (33,3% от общей оценки) за выполнение практических заданий на выявление уровня обученности «владеть»,

– 3 балла (20% оценки) за ответы на теоретические вопросы,

– 2 балла (13,3% оценки) за ответы на дополнительные вопросы.

5. По итогам текущего контроля и промежуточной аттестации в соответствии с показателями и критериями оценивания компетенций определяется уровень сформированности компетенций студента и выставляется оценка по шкале оценивания.

1.4. Показатели и критерии оценивания компетенций, шкала оценивания

Компетенции	Показатели оценивания	Критерии оценивания компетенций				Итого:
		Высокий (верно и в полном объеме) 5 б.	Средний (с незначительными замечаниями) 4 б.	Низкий (на базовом уровне, с ошибками) 3 б.	Недостаточный (содержит большое количество ошибок/ответ не дан) – 2 б.	
<i>Теоретические показатели</i>						
<i>ПК-13</i>	Знает теоретические основы информатики <hr/> <i>Доклад</i>	Верно, и в полном объеме знает теоретические основы информатики	С незначительными замечаниями знает теоретические основы информатики	На базовом уровне, с ошибками знает теоретические основы информатики	Не знает теоретические основы информатики	20
	Знает рынки программно-информационных продуктов и услуг <hr/> <i>Доклад</i>	Верно, и в полном объеме знает рынки программно-информационных продуктов и услуг	С незначительными замечаниями знает рынки программно-информационных продуктов и услуг	На базовом уровне, с ошибками знает рынки программно-информационных продуктов и услуг	Не знает рынки программно-информационных продуктов и услуг	
	Знает методы проектирования, внедрения и организации эксплуатации корпоративных ИС и ИКТ <hr/> <i>Доклад</i>	Верно, и в полном объеме знает методы проектирования, внедрения и организации эксплуатации корпоративных ИС и ИКТ	С незначительными замечаниями знает методы проектирования, внедрения и организации эксплуатации корпоративных ИС и ИКТ	На базовом уровне, с ошибками знает методы проектирования, внедрения и организации эксплуатации корпоративных ИС и ИКТ	Не знает методы проектирования, внедрения и организации эксплуатации корпоративных ИС и ИКТ	
	Знает виды контента информационных ресурсов предприятия и Интернет-ресурсов, процессы управления жизненным циклом цифрового контента, процессы создания и	Верно, и в полном объеме знает виды контента информационных ресурсов предприятия и Интернет-	С незначительными замечаниями знает виды контента информационных ресурсов предприятия и Интернет-ресурсов,	На базовом уровне, с ошибками знает виды контента информационных ресурсов предприятия и Интернет-ресурсов, процессы управления жизненным циклом	Не знает виды контента информационных ресурсов предприятия и Интернет-ресурсов, процессы	

Компетенции	Показатели оценивания	Критерии оценивания компетенций				Итого:
		Высокий (верно и в полном объеме) 5 б.	Средний (с незначительными замечаниями) 4 б.	Низкий (на базовом уровне, с ошибками) 3 б.	Недостаточный (содержит большое количество ошибок/ответ не дан) – 2 б.	
	использования информационных сервисов <i>Доклад</i>	ресурсов, процессы управления жизненным циклом цифрового контента, процессы создания и использования информационных сервисов	процессы управления жизненным циклом цифрового контента, процессы создания и использования информационных сервисов	цифрового контента, процессы создания и использования информационных сервисов	управления жизненным циклом цифрового контента, процессы создания и использования информационных сервисов	
<i>Практические показатели</i>						
<i>ПК-13</i>	Умеет систематизировать и обобщать информацию, организовывать и проводить исследования в области экономики, управления и ИКТ <i>Лабораторная работа</i>	Верно, и в полном объеме может систематизировать и обобщать информацию, организовывать и проводить исследования в области экономики, управления и ИКТ	С незначительными замечаниями может систематизировать и обобщать информацию, организовывать и проводить исследования в области экономики, управления и ИКТ	На базовом уровне, с ошибками может систематизировать и обобщать информацию, организовывать и проводить исследования в области экономики, управления и ИКТ	Не может систематизировать и обобщать информацию, организовывать и проводить исследования в области экономики, управления и ИКТ	20
	Умеет управлять процессами жизненного цикла контента предприятия и Интернет- ресурсов, управлять процессами создания и использования информационных сервисов (контент-сервисов)	Верно, и в полном объеме может управлять процессами жизненного цикла контента предприятия и Интернет-	С незначительными замечаниями может управлять процессами жизненного цикла контента предприятия и Интернет-ресурсов,	На базовом уровне, с ошибками может управлять процессами жизненного цикла контента предприятия и Интернет- ресурсов, управлять процессами создания и использования	Не может управлять процессами жизненного цикла контента предприятия и Интернет- ресурсов,	

Компетенции	Показатели оценивания	Критерии оценивания компетенций				Итого:
		Высокий (верно и в полном объеме) 5 б.	Средний (с незначительными замечаниями) 4 б.	Низкий (на базовом уровне, с ошибками) 3 б.	Недостаточный (содержит большое количество ошибок/ответ не дан) – 2 б.	
	<i>Лабораторная работа</i>	ресурсов, управлять процессами создания и использования информационных сервисов (контент-сервисов)	управлять процессами создания и использования информационных сервисов (контент-сервисов)	информационных сервисов (контент-сервисов)	управлять процессами создания и использования информационных сервисов (контент-сервисов)	
	Умеет проектировать, внедрять и организовывать эксплуатацию ИС и ИКТ <i>Лабораторная работа</i>	Верно, и в полном объеме может проектировать, внедрять и организовывать эксплуатацию ИС и ИКТ	С незначительными замечаниями может проектировать, внедрять и организовывать эксплуатацию ИС и ИКТ	На базовом уровне, с ошибками может проектировать, внедрять и организовывать эксплуатацию ИС и ИКТ	Не может проектировать, внедрять и организовывать эксплуатацию ИС и ИКТ	
	Умеет проводить OLAP анализ для различных проблемных областей; формулировать запросы к многомерным базам данных с помощью языка MDX, строить модели DataMining; <i>Лабораторная работа</i>	Верно, и в полном объеме может проводить OLAP анализ для различных проблемных областей; формулировать запросы к многомерным базам данных с помощью языка MDX, строить модели DataMining;	С незначительными замечаниями может проводить OLAP анализ для различных проблемных областей; формулировать запросы к многомерным базам данных с помощью языка MDX, строить модели DataMining;	На базовом уровне, с ошибками может проводить OLAP анализ для различных проблемных областей; формулировать запросы к многомерным базам данных с помощью языка MDX, строить модели DataMining;	Не может проводить OLAP анализ для различных проблемных областей; формулировать запросы к многомерным базам данных с помощью языка MDX, строить модели DataMining;	

Компетенции	Показатели оценивания	Критерии оценивания компетенций				Итого:
		Высокий (верно и в полном объеме) 5 б.	Средний (с незначительными замечаниями) 4 б.	Низкий (на базовом уровне, с ошибками) 3 б.	Недостаточный (содержит большое количество ошибок/ответ не дан) – 2 б.	
<i>Владеет</i>						
<i>ПК-13</i>	Владеет методами поиска, хранения и обработки информации <hr/> <i>Контрольная работа</i>	Верно, и в полном объеме владеет методами поиска, хранения и обработки информации	С незначительными замечаниями владеет методами поиска, хранения и обработки информации	На базовом уровне, с ошибками владеет методами поиска, хранения и обработки информации	Не владеет методами поиска, хранения и обработки информации	20
	Владеет методами управления процессами жизненного цикла контента предприятия и Интернет-ресурсов; <hr/> <i>Контрольная работа</i>	Верно, и в полном объеме владеет методами управления процессами жизненного цикла контента предприятия и Интернет-ресурсов;	С незначительными замечаниями владеет методами управления процессами жизненного цикла контента предприятия и Интернет-ресурсов;	На базовом уровне, с ошибками владеет методами управления процессами жизненного цикла контента предприятия и Интернет-ресурсов;	Не владеет методами управления процессами жизненного цикла контента предприятия и Интернет-ресурсов;	
	Владеет методами проектирования, внедрения и организации эксплуатации ИС и ИКТ <hr/> <i>Контрольная работа</i>	Верно, и в полном объеме владеет методами проектирования, внедрения и организации эксплуатации ИС и ИКТ	С незначительными замечаниями владеет методами проектирования, внедрения и организации эксплуатации ИС и ИКТ	На базовом уровне, с ошибками владеет методами проектирования, внедрения и организации эксплуатации ИС и ИКТ	Не владеет методами проектирования, внедрения и организации эксплуатации ИС и ИКТ	
	Владеет методами управления и процессами создания и использования информационных сервисов (контент-сервисов)	Верно, и в полном объеме владеет методами управления и процессами создания и	С незначительными замечаниями владеет методами управления и процессами создания и	На базовом уровне, с ошибками владеет методами управления и процессами создания и информационных сервисов	Не владеет методами управления и процессами создания и использования	

Компетенции	Показатели оценивания	Критерии оценивания компетенций				Итого:
		Высокий (верно и в полном объеме) 5 б.	Средний (с незначительными замечаниями) 4 б.	Низкий (на базовом уровне, с ошибками) 3 б.	Недостаточный (содержит большое количество ошибок/ответ не дан) – 2 б.	
	<i>Контрольная работа</i>	использования информационных сервисов (контент- сервисов)	использования информационных сервисов (контент- сервисов)	(контент-сервисов)	информационных сервисов (контент- сервисов)	
<i>ВСЕГО:</i>						<i>60</i>

Шкала оценивания:

Оценка	Баллы	Уровень сформированности компетенции
отлично	52-60	высокий
хорошо	43-51	хороший
удовлетворительно	30-42	достаточный
неудовлетворительно	29 и менее	недостаточный

2. Типовые контрольные задания для оценки результатов обучения по дисциплине и иные материалы для подготовки к промежуточной аттестации

2.1. Материалы для подготовки к промежуточной аттестации

Вопросы к зачету с оценкой:

1. Понятие OLAP.
2. Различия между транзакционными и аналитическими системами.
3. Области применения хранилищ данных.
4. Раскройте суть технологии OLAP?
5. Укажите основные различия между OLAP и OLTP.
6. Чем обусловлено развитие хранилищ данных?
7. Какие основные свойства хранилищ данных, которые отличают их от транзакционных систем?
8. Перечислите наиболее распространенные области применения хранилищ данных.
9. Перечислите основные аспекты, в которых проявляется интегрированность данных хранилища.
10. В чем проявляется зависимость данных хранилища от времени?
11. Общие свойства хранилищ данных.
12. Данные хранилища.
13. Компоненты хранилища.
14. Откуда появилась необходимость в хранилищах данных?
15. Каковы отличия характеристик данных в оперативных и аналитических системах?
16. Перечислите основные требования к данным хранилищ данных.
17. Как соотносятся между собой концепции хранилищ данных и анализа данных?
18. В чем сложности анализа данных из оперативных баз данных?
19. Какие общие требования к данным предъявляются при анализе данных?
20. Назовите основные компоненты хранилищ данных.
21. Какие типовые архитектуры хранилищ данных вам известны?
22. Чем обусловлен выбор модели хранилища данных?
23. В чем особенность звездообразных схем структуры хранилищ данных?
24. Структура хранилищ данных.
25. Способы хранения агрегатных данных.
26. Дайте определение куба, измерения и меры. Приведите примеры.
27. Что такое "разрезание куба"?
28. Опишите роль меток, иерархий и уровней в структуре куба.
29. Чем отличается гиперкуб от поликуба?
30. Какие требования к приложениям для многомерного анализа

предъявляются по тесту FASMI?

31. Что такое многомерный куб? Что хранится в ячейке куба?
32. Что такое меры и измерения куба?
33. Виды иерархий. Приведите примеры.
34. Способы хранения OLAP-кубов.
35. Технология доступа к кубам из клиентских приложений.
36. Какие программы — клиенты OLAP Вы знаете?
37. Подходы к стратегии построения хранилищ данных.
38. Этапы спиральной модели применительно к разработке хранилищ данных.
39. Почему неоднородность программной среды является важным фактором, учитываемым при проектировании хранилищ данных?
40. Почему распределенность данных является важным фактором, учитываемым при проектировании хранилищ данных?
41. Почему использование метаданных конечными пользователями является важным фактором, учитываемым при проектировании хранилищ данных?
42. Почему защита данных является важным фактором, учитываемым при проектировании хранилищ данных?
43. Почему большие объемы хранимых данных являются важным фактором, учитываемым при проектировании хранилищ данных?
44. Какие подходы к стратегии построения корпоративного хранилища данных вы знаете?
45. Перечислите основные стадии спиральной модели разработки хранилищ данных.
46. Какие виды работ проводятся на этапе системно-аналитического обследования?
47. Перечислите пять классов данных информационного обеспечения систем поддержки принятия решений.
48. На каком этапе разработки хранилища данных определяются состав, содержание и источники потоков данных, которые будут поступать из источников в хранилище?
49. Информация каких типов включается в общую структуру позитария хранилища?
50. Приведите примеры общего и специального программного обеспечения систем поддержки принятия решений.
51. Какие требования предъявляются к серверному и клиентскому программному обеспечению?
52. Какие работы проводятся на этапе внедрения хранилищ данных?
53. Архитектура аналитических платформ.
54. Обзор рынка BI.
55. Какие продукты фирмы Microsoft используются для реализации хранилищ данных?
56. Какие продукты фирмы Oracle используются для реализации хранилищ данных?

57. Сферы применения Data Mining.
58. Процесс Data Mining. Начальные этапы.
59. Процесс Data Mining. Очистка данных.
60. Процесс Data Mining. Построение и использование моделей.
61. В чем отличия Data Mining от других методов анализа данных?
62. Из каких последовательных стадий состоит процесс Data Mining?
63. Назовите области применения Data Mining.
64. На какие две группы подразделяются методы Data Mining по принципу работы с исходными обучающими данными?
65. На каком этапе пересекается работа специалиста предметной области и специалиста по добыче данных?
66. На каком этапе пересекается работа специалиста предметной области и администратора баз данных?

Типовые контрольные задания:

Вариант 1

Свойства хранилища данных
Данные хранилища
Источники данных

Вариант 2

Витрины данных
Метаданные
Компоненты хранилища

Вариант 3

Выбор метода реализации хранилищ данных
Проектирование хранилища данных
OLAP

Вариант 4

MOLAP
HOLAP
Таблица фактов

Вариант 5

Таблицы измерений
OLAP на клиенте и на сервере

Образцы тестовых заданий для проведения текущего контроля и промежуточной аттестации по итогам освоения дисциплины, а также для контроля самостоятельной работы:

1. Какими свойствами должны обладать средства хранения данных:
интегрированные, неизменчивые, поддерживающие хронологию.
 - а) предметно-ориентированные, интегрированные, неизменчивые, поддерживающие хронологию.
 - б) предметно-ориентированные, неизменчивые, поддерживающие хронологию.
 - в) изменчивые, поддерживающие хронологию.

2. На основе чего реализуется концептуальная многомерная модель данных:
на основе представления данных в виде многомерного пространства, размерность которого определяется количеством измерений.
 - а) на основе представления данных в виде многомерного пространства, размерность которого определяется количеством граней куба.
 - б) на основе представления данных в виде бесконечного пространства.
 - в) на основе представления данных в виде пространства, ограниченного многомерным кубом.

3. Размерность многомерного пространства данных для анализа математически определяется:
 - а) сложением размеров всех измерений в модели данных;
 - б) количеством атрибутов в реляционной таблице фактов;
 - в) количеством таблиц содержащих измерения;
 - г) перемножением размеров всех измерений в модели данных.

4. Размер или кардинальность измерения определяется:
 - а) количеством атрибутов и свойств в измерении;
 - б) количеством значений ключа в таблице измерения;
 - в) количеством элементов в измерении;
 - г) количеством записей в таблице измерений;

5. Основное назначение систем интеллектуального анализа (Data Mining):
 - а) обнаружение в сырых данных скрытых знаний;
 - б) проведение статистического анализа;
 - в) решения задач математического программирования;
 - г) поиск агрегированных данных;

6. Правильная последовательность в Business Intelligence:
 - а) данные-информация-знания-принятие решения
 - б) информация-данные-знания-принятие решения

- в) данные-знания-информация-принятие решения
- г) принятие решения-информация-данные-знания

7. В платформе для бизнес-анализа должны быть реализованы:

- а) 10 ключевых возможностей
- б) 12 ключевых возможностей
- в) 15 ключевых возможностей
- г) 20 ключевых возможностей

8. Ключевые возможности систем BI сгруппированы:

- а) по двум основным категориям
- б) по трем основным категориям
- в) по четырем основным категориям
- г) по пяти основным категориям

9. «BI-инфраструктура» относится к категории:

- а) представление информации
- б) анализ данных
- в) возможность интеграции
- г) является основной категорией

10. Перечислите правильную последовательность этапов Knowledge Discovery in Databases –процесса обнаружения знаний в базах данных:

- а) трансформация, интерпретация результатов, выборка, очистка, построение моделей.
- б) построение моделей, выборка, очистка, трансформация, интерпретация результатов.
- в) построение моделей, выборка, очистка, трансформация, интерпретация результатов,
- г) выборка, очистка, трансформация, построение моделей, интерпретация результатов.

11. OLAP-системы это:

- а) информационные системы оперативной транзакционной обработки данных.
- б) информационные системы оперативного анализа данных.
- в) информационные системы автоматической обработки данных.
- г) информационные системы алгоритмической обработки данных.

12. В процессе погружения в информационное хранилище данные ...

- а) очищаются от ненужной для анализа информации;
- б) агрегируются;
- в) преобразуются из разных типов данных предметных приложений в единую структуру хранения;
- г) индексируются;

д) синхронизируются.

13. ETL (Extraction, Transformation, Loading) -процессы сбора, преобразования и загрузки данных обеспечивают:

- а) подготовку результатов анализа;
- б) создание массива данных в информационном хранилище.

14. Найдите ошибочный критерий. В процессе продвижения данных в информационное хранилище используются следующие критерии оценки качества данных по структурному представлению:

- а) по критичности ошибок в данных — ошибки в именах полей, типах данных;
- б) по правильности форматов и представлений данных;
- в) на соответствие ограничениям целостности;
- г) на кроссязыковый разрыв;
- д) уникальности внутренних и внешних ключей;
- е) по полноте данных и связей.

15. Найдите ошибочный ответ. Нарушения смыслового содержания данных (семантические разрывы) бывают следующих видов:

- а) «вавилонский»;
- б) кросспотоковый разрыв;
- в) разрыв кодограмм;
- г) кроссязыковый разрыв;
- д) асинхронный разрыв.

16. Основными составляющими хранилища данных являются:

- а) данные о деятельности предприятия за определенный период
- б) таблица фактов и таблица измерений
- в) таблица фактов

17. Агрегированные данные – это

- а) данные, объединенные на основании запроса пользователя
- б) данные, собранные некоторым агрегатом
- в) данные, предназначенные нескольким пользователям

Литература для подготовки к зачету с оценкой

а) нормативные документы

1. Гражданский кодекс Российской Федерации (часть первая) от 30.11.1994 № 51–ФЗ (ред. от 26.07.2006) // Собрание законодательства РФ. - 1994. - № 32. – Ст. 3301. - Консультант Плюс. [Электрон. ресурс]. - Электрон. дан. - [М., 2016.]

б) основная литература

1. Агальцов В.П. Базы данных. В 2-х кн. Книга 2. Распределенные и удаленные базы данных : учебник / В.П. Агальцов. — М. : ИД «ФОРУМ» : ИНФРА-М, 2017. — 271 с. - Режим доступа: <http://znanium.com/catalog/product/652917>

2. Основы проектирования баз данных: Учебное пособие / Голицына О.Л., Партыка Т.Л., Попов И.И., - 2-е изд. - М.: Форум, НИЦ ИНФРА-М, 2016. - 416 с. - Режим доступа: <http://znanium.com/catalog/product/552969>

в) дополнительная литература

1. Култыгин, О. П. Администрирование баз данных. СУБД MS SQL Server: учеб. пособие / О. П. Култыгин. - М.: МФПА, 2012. - 232 с. - Режим доступа: <http://znanium.com/catalog/product/451114>

Промежуточная аттестация

2.2. Комплект билетов для проведения промежуточной аттестации

**АВТНОМНАЯ НЕКОММЕРЧЕСКАЯ ОБРАЗОВАТЕЛЬНАЯ
ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
ЦЕНТРОСОЮЗА РОССИЙСКОЙ ФЕДЕРАЦИИ
«РОССИЙСКИЙ УНИВЕРСИТЕТ КООПЕРАЦИИ»
КАЗАНСКИЙ КООПЕРАТИВНЫЙ ИНСТИТУТ (ФИЛИАЛ)**

Направление подготовки: 38.03.05 Бизнес – информатика

Направленность: «Электронный бизнес»

Дисциплина: «Хранилища данных»

БИЛЕТ № 1

1. Отличие хранилища данных от баз данных
2. Назначение языка программирования SQL

БИЛЕТ № 2

1. Два метода создания хранилища данных.
2. Каковы области использования языка программирования SQL

БИЛЕТ № 3

1. Назначение Power Builder (PB) и его основные компоненты
2. Среда хранилища данных. Перечислите основные компоненты среды.

БИЛЕТ № 4

1. На чем основан EAR - метод моделирования данных
2. Что представляет собой база знаний, чем она отличается от хранилища данных?

БИЛЕТ № 5

1. Перечислите основные функции хранилища данных
2. Архитектура хранилищ данных

БИЛЕТ № 6

1. Концепция хранилищ данных
2. Многомерная модель данных и ее использование в OLAP-технологии.

Промежуточная аттестация
Комплект тестовых заданий для проведения зачета с оценкой по дисциплине

АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ ОБРАЗОВАТЕЛЬНАЯ
ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
ЦЕНТРОСОЮЗА РОССИЙСКОЙ ФЕДЕРАЦИИ
«РОССИЙСКИЙ УНИВЕРСИТЕТ КООПЕРАЦИИ»
КАЗАНСКИЙ КООПЕРАТИВНЫЙ ИНСТИТУТ (ФИЛИАЛ)

Направление подготовки: 38.03.05 Бизнес – информатика

Направленность: «Электронный бизнес»

Дисциплина: «Хранилища данных»

Тестовые задания для проверки уровня обученности **ЗНАТЬ:**

1. Мастер интеллектуального анализа предназначен для работы
 - а) с моделями интеллектуального анализа данных;
 - б) с таблицами источника данных для интеллектуального анализа;
 - в) со структурами и моделями интеллектуального анализ
 - г) со структурами интеллектуального анализа данных.

2. С какой целью создаются хранилища данных:
 - а) для хранения одном месте любых данных.
 - б) для интеграции разрозненных данных.
 - в) для агрегации ранее разъединенных детализированных данных.
 - г) для интеграции в одном месте, согласования и, возможно, агрегации ранее разъединенных детализированных данных.

3. Роль унифицированной многомерной модели заключается:
 - а) в создании концептуальной модели хранилища данных;
 - б) в определении функциональной зависимости между данными;
 - в) в определении реляционных отношений между сущностями;
 - г) в создании моста между пользователем и источниками данных.

4. Какова основная цель создания и использования хранилищ данных:
 - а) анализ данных для принятия управленческих решений;
 - б) надежное хранение, накопленных данных;
 - в) резервное копирование данных

5. Какие из вариантов ответов являются характерными требованиями к хранению данных для принятия решений в хранилищах данных?
 - а) данные ориентированы на приложения;
 - б) данные управляются транзакциями;
 - в) данные обобщены либо очищены.

6. CASE-технология

- а) технология, использования графических пакетов для моделирования
- б) совокупность компьютерных методов и средств для проектирования систем и моделирования бизнес-процессов

7. Централизованное хранилище данных - это:

- а) способ реализации концепции хранилища данных;
- б) синоним понятия «информационное хранилище»;
- в) сервер хранилища данных

Тестовые задания для проверки уровня обученности **УМЕТЬ, ВЛАДЕТЬ:**

1. Информационные хранилища созданы для удобства ...

- а) руководителей всех уровней для принятия решений;
- б) стратегического планирования;
- в) реорганизации бизнеса;
- г) предметных приложений;
- д) редактирования данных.

2. Многомерный анализ – это:

- а) одновременный анализ по нескольким измерениям;
- б) одновременный анализ по нескольким параметрам;
- в) одновременный анализ по нескольким данным.

3. Формирование подмножества многомерного массива данных, соответствующего единственному значению одного или нескольких элементов измерений, не входящих в это подмножество, называется

- а) гиперкубом;
- б) срезом гиперкуба;
- в) базой данных;
- г) витриной данных.

4. Технология оперативной аналитической обработки данных, использующая методы и средства для сбора, хранения и анализа многомерных данных в целях поддержки процессов принятия решений, называется

- а) OLTP;
- б) OLAP;
- в) Data Mining.

5. Языком запросов к реляционным базам данных является ...

- а) SQL
- б) Pascal
- в) C#

г) SSH

6. Что входит в состав хранилища данных:

- а) организационная структура, технические средства, базы или совокупности баз данных и программное обеспечение.
- б) базы или совокупности баз данных и программное обеспечение.
- в) источники данных и программное обеспечение.
- г) организационная структура и программное обеспечение

7. Что является основными составляющими структуры хранилищ данных:

- а) таблица исходной информации и таблица запросов.
- б) таблица базы данных и запросы.
- в) таблица фактов и таблица измерений.
- г) таблица запросов и таблица данных.

8. На основе чего реализуется концептуальная многомерная модель данных:

- а) на основе представления данных в виде многомерного пространства, размерность которого определяется количеством измерений.
- б) на основе представления данных в виде многомерного пространства, размерность которого определяется количеством граней куба.
- в) на основе представления данных в виде бесконечного пространства.
- г) на основе представления данных в виде пространства, ограниченного многомерным кубом.

9. Для чего предназначен язык MDX (Multidimensional Expressions)

- а) язык для создания приложений пользователя
- б) язык для более простого и эффективного доступа к многомерным структурам данных (для формирования запросов и описания алгоритмов получения вычисляемых значений)

10. Что означает аббревиатура OLAP?

- а) оперативный анализ данных;
- б) данные, используемые в производственной деятельности;
- в) организационная структура предприятия;
- г) объединение баз;
- д) централизованная единая база данных организации

2.3. Критерии оценки для проведения зачета с оценкой по дисциплине

После завершения тестирования на зачете с оценкой на мониторе компьютера высвечивается результат – процент правильных ответов.

Результат переводится в баллы и суммируется с текущими семестровыми баллами.

Максимальная сумма (100 баллов), набираемая студентом по дисциплине, предусматривающей в качестве формы промежуточной аттестации зачет с оценкой, включают две составляющие.

Первая составляющая – оценка регулярности и своевременности качества выполнения студентом учебной работы по изучению дисциплины в течение семестра (сумма не более 60 баллов).

Вторая составляющая – оценка знаний студента на зачете с оценкой (не более 40 баллов).

Перевод полученных итоговых баллов в оценки осуществляется по следующей шкале:

- с 86 до 100 баллов – «отлично»;
- с 71 до 85 баллов – «хорошо»;
- с 50 до 70 баллов – «удовлетворительно»

Если студент при тестировании отвечает правильно менее, чем на 50 %, то автоматически выставляется оценка «неудовлетворительно» (без суммирования текущих рейтинговых баллов), а студенту назначается переэкзаменовка в дополнительную сессию.

2.4. Методические материалы, определяющие процедуру оценивания по дисциплине

Общая процедура оценивания определена Положением о фондах оценочных средств.

1. Процедура оценивания результатов освоения программы дисциплины включает в себя оценку уровня сформированности общекультурных и профессиональных компетенций студента, уровней обученности: «знать», «уметь», «владеть».

2. При сдаче зачета с оценкой:

– профессиональные знания студента могут проверяться при ответе на теоретические вопросы, при выполнении тестовых заданий, практических работ;

– степень владения профессиональными умениями, уровень сформированности компетенций (элементов компетенций) – при решении ситуационных задач, выполнении практических работ и других заданий.

3. Результаты промежуточной аттестации фиксируются в баллах. Общее количество баллов складывается из следующего:

- до 60% от общей оценки за выполнение практических заданий,
- до 30% оценки за ответы на теоретические вопросы,
- до 10% оценки за ответы на дополнительные вопросы

**АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ ОБРАЗОВАТЕЛЬНАЯ
ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
ЦЕНТРОСОЮЗА РОССИЙСКОЙ ФЕДЕРАЦИИ
«РОССИЙСКИЙ УНИВЕРСИТЕТ КООПЕРАЦИИ»
КАЗАНСКИЙ КООПЕРАТИВНЫЙ ИНСТИТУТ (ФИЛИАЛ)**

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ
ТЕКУЩЕГО КОНТРОЛЯ И ТЕКУЩЕЙ АТТЕСТАЦИИ ПО
ДИСЦИПЛИНЕ (МОДУЛЮ)**

ХРАНИЛИЩА ДАННЫХ

Направление подготовки: 38.03.05 Бизнес - информатика
Направленность: «Электронный бизнес»

1. Материалы для текущего контроля

АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ ОБРАЗОВАТЕЛЬНАЯ
ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
ЦЕНТРОСОЮЗА РОССИЙСКОЙ ФЕДЕРАЦИИ
«РОССИЙСКИЙ УНИВЕРСИТЕТ КООПЕРАЦИИ»
КАЗАНСКИЙ КООПЕРАТИВНЫЙ ИНСТИТУТ (ФИЛИАЛ)

КОМПЛЕКТ ЗАДАНИЙ ДЛЯ КОНТРОЛЬНОЙ РАБОТЫ

по дисциплине «Хранилища данных»

Контрольные работы по дисциплине «Хранилища данных» состоят из тестовых вопросов.

I. ТЕСТОВЫЕ ВОПРОСЫ ДЛЯ ПРОВЕДЕНИЯ КОНТРОЛЬНЫХ РАБОТ:

1. Первая промышленная СУБД система IMS фирмы IBM была введена в эксплуатацию в:

- а) 1960 г.;
- б) 1963 г.;
- в) 1965 г.;
- г) 1968 г.

2. Эпоха персональных компьютеров _____ этап в развитии СУБД:

- а) первый;
- б) второй;
- в) третий;

3. СУБД - это:

- а) интегрированная совокупность взаимосвязанных данных;
- б) специальные пакеты программ, обеспечивающие ввод, поиск, хранение и корректировку данных в базах данных;
- в) совокупность программ на языке VisualBasic;
- г) инструмент, позволяющий производить сложные математические расчеты, с использованием электронных таблиц.

4. Сетевая модель данных использует:

- а) дерево данных;
- б) произвольный граф (или множество взаимосвязанных деревьев);
- в) данные в виде связанных таблиц;

5. Дерево является основной структурной единицей:

- а) иерархической модели данных;
- б) сетевой модели данных;
- в) реляционной модели данных;
- г) объектной модели данных.

6. Данные в виде связанных таблиц представлены:

- а) сетевой модели данных;
- б) иерархической модели данных;
- в) реляционной модели данных;

7. Данные в реляционной модели представлены как факты:

- а) сущностях и связях;
- б) сущностях;
- в) связях;

8. Объект, предназначенный для облегчения ввода данных:

- а) форма;
- б) таблица;
- в) запрос;
- г) отчет.

9. Совокупность структур данных и операций по их обработке это:

- а) технология обработки данных;
- б) способ доступа к данным;
- в) уровень, элемент, связь;
- г) модель данных.

10. Реляционная база данных представляет собой:

- а) сегмент данных;
- б) массив данных;
- в) последовательный набор данных;
- г) совокупность связанных между собой таблиц.

11. Совокупность связанных данных, правила организации которых основаны на общих принципах описания, хранения и манипулирования данными называется:

- а) пакетом прикладных программ;
- б) документом;
- в) базой данных;
- г) шаблоном.

12. Совокупность средств, методов и персонала для организации хранения информации в компьютерной среде называется:

- а) базой данных;

- б) базой знаний;
- в) банком данных;

13. Специальным образом организованное хранение информационных ресурсов в виде интегрированной совокупности файлов, обеспечивающее удобное взаимодействие между ними и быстрый доступ к данным называется:

- а) базой данных;
- б) экспертной базой;
- в) базой знаний;

14. Модель, при которой может находиться несколько элементов, связанных не только вертикально, но и горизонтально называется:

- а) иерархической;
- б) сетевой;
- в) реляционной;

15. Способ хранения и поиск информации в виде карточек, применяемый в библиотеках, является характерным примером _____ модели данных:

- а) иерархической;
- б) сетевой;
- в) дескрипторной;

16. Модель данных, напоминающая по своему строению дерево называется:

- а) иерархической;
- б) сетевой;
- в) реляционной;

17. Основопологающим элементом построения базы данных является:

- а) структура данных;
- б) модель данных;
- в) иерархия данных;

18. Наиболее широко применяемой сегодня моделью построения баз данных является:

- а) иерархическая;
- б) сетевая;
- в) реляционная;

19. База данных, состоящая из одной таблицы называется:

- а) реляционной базой данных;
- б) развернутой базой данных;
- в) базой данных с плоским файлом;

20. Какая база данных состоит из нескольких, возможно пересекающихся или даже дублирующих друг друга частей, хранимых в различных ЭВМ вычислительной сети:

- а) централизованная;
- б) распределенная;
- в) реляционная;
- г) композитная;

21. Поле базы данных это:

- а) элементарная единица логической организации данных;
- б) совокупность логически связанных единиц данных;
- в) совокупность экземпляров записей одной структуры;

22. Необходимым условием создания базы данных является:

- а) структурирование данных;
- б) унификация данных;
- в) оба ответа верны;

23. Организация файлов в памяти компьютера относится к:

- а) немашинному информационному обеспечению;
- б) внутримашинному информационному обеспечению;
- в) традиционной организации файлов в памяти компьютера;

24. Отчет в Access можно создавать на основе:

- а) только формы;
- б) таблицы или запроса;
- в) только запроса;
- г) только таблицы.

25. Формы нужны:

- а) для хранения информации;
- б) для извлечения информации из таблиц;
- в) для заполнения и редактирования таблиц;

26. Какого запроса в MS Access не бывает:

- а) простого;
- б) перекрестного;
- в) периферийного;

27. Таблицы в реляционной базе данных увязываются между собой с помощью:

- а) схемы таблиц;
- б) схемы данных;
- в) структуры данных;
- г) структуры таблиц;

28. Какого типа данных в MS Access для поля не существует:

- а) числовой;
- б) денежный;
- в) текстовой;
- г) процентный;

29. Объектом базы данных, представляющим собой бланк, подлежащий заполнению, или маску, накладываемую на набор данных является:

- а) отчет;
- б) запрос;
- в) таблица;
- г) форма;

30. Объектом базы данных Microsoft Access, позволяющий выбирать из БД только необходимую информацию - это:

- а) отчет;
- б) запрос;
- в) таблица;
- г) форма;

31. Ввод записи в таблицу осуществляется в режимах:

- а) таблицы;
- б) формы;
- в) отчета;
- г) запроса.

Критерии оценки:

- оценка «отлично» выставляется студенту, если задание выполнено верно и в полном объеме;

- оценка «хорошо» выставляется студенту, если задание выполнено с незначительными замечаниями;

- оценка «удовлетворительно» выставляется, если задание выполнено на базовом уровне, но с ошибками;

- оценка «неудовлетворительно» выставляется, если содержится большое количество ошибок, задание не выполнено.

АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ ОБРАЗОВАТЕЛЬНАЯ
ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
ЦЕНТРОСОЮЗА РОССИЙСКОЙ ФЕДЕРАЦИИ
«РОССИЙСКИЙ УНИВЕРСИТЕТ КООПЕРАЦИИ»
КАЗАНСКИЙ КООПЕРАТИВНЫЙ ИНСТИТУТ (ФИЛИАЛ)

ТЕМЫ ДОКЛАДОВ

по дисциплине «Хранилища данных»

1. Организация словаря-справочника данных (каталога) (на примере конкретной СУБД).
2. Управление доступом к данным (на примере конкретной СУБД).
Права доступа, роли, группы пользователей.
3. Полнотекстовое индексирование (на примере конкретной СУБД).
4. Особенности объектно-ориентированных СУБД.
5. Транзакционные параллельные СУБД.
6. Хранилища данных: назначение, принципы организации.
7. GRID-технологии.
8. GRID-технология Oracle.
9. Big Data.
10. Многомерные базы данных.
11. Поточковые базы данных.
12. Нереляционные базы данных. MongoDB, Redis.
13. SciDB.
14. TinyDB для беспроводных сенсорных сетей.
15. Облачные технологии (на примере конкретной СУБД).

Требования к оформлению

Доклад (реферат) выполняется студентом самостоятельно в отдельной папке с титульным листом на стандартных листах формата А4, шрифт Times New Roman, кегель 14, интервал – 1,5, поля: верхнее, нижнее – 2,0; слева – 3,0; справа – 1,0. Форматирование по ширине. Отступ первой строки - 1,25.

Доклад 10-12 минут, презентация в формате PowerPoint (ppt)

Критерии оценки:

- оценка «зачтено» выставляется студенту, если доклад носит исследовательский характер, имеет грамотно изложенный анализ теории по выбранному вопросу, проведен анализ точек зрения различных авторов или литературных источников, логично и последовательно изложен материал, сделаны соответствующие выводы.

- оценка «не зачтено» выставляется студенту, если доклад не отвечает основным требованиям, имеет поверхностный анализ и недостаточный уровень самостоятельности студента, материал изложен непоследовательно

ЛАБОРАТОРНАЯ РАБОТА

по дисциплине «Хранилища данных»

1. ПРОЕКТИРОВАНИЕ, РЕАЛИЗАЦИЯ И НАПОЛНЕНИЕ БАЗЫ ДАННЫХ, ЯВЛЯЮЩЕЙ ИСТОЧНИКОМ ДАННЫХ ДЛЯ ХРАНИЛИЩА

1.1. Основные понятия

Сущность – это некоторая абстракция реально существующего объекта (класса объектов), процесса или явления, о котором необходимо хранить информацию в БД.

Атрибут – это поименованная характеристика сущности, которая принимает значения из некоторого множества значений. В модели атрибут выступает в качестве средства, с помощью которого моделируются свойства сущностей.

Связи выступают в модели в качестве средства, с помощью которого представляются отношения между сущностями, имеющими место в предметной области. При анализе связей между сущностями могут встречаться бинарные (между двумя сущностями), тернарные (между тремя сущностями) и, в общем случае n-арные связи.

Потенциальный ключ – подмножество атрибутов, обладающее свойствами уникальности и минимальности. В отношении может существовать несколько потенциальных ключей.

Первичный ключ – это любой из потенциальных ключей, выбираемый для использования в связях между отношениями. Если выбран первичный ключ – все остальные потенциальные ключи называют альтернативными ключами.

Внешний ключ – столбец или комбинация столбцов, значения которых необходимы для сопоставления с первичным ключом в другой таблице. Используется для объединения родственных таблиц.

1.2. Анализ предметной области

Рассмотрим пример. Необходимо спроектировать базу данных супермаркета. В этой базе данных должна храниться информация о товарах, поставщиках, заказах, доставках, клиентах, сотрудниках супермаркета. В ходе анализа предметной области необходимо учесть её специфику. В

частности, будущая база данных должна учитывать следующие факты:

- в одном заказе может быть несколько товаров;
- один товар может фигурировать в нескольких заказах;
- отношения между сотрудниками супермаркета могут быть представлены в виде иерархии, то есть один сотрудник может быть и начальником, и подчиненным одновременно.

С учетом изложенных фактов можно приступить к созданию концептуальной модели.

1.3. Концептуальное моделирование

Модель базы данных, отражающая взгляд пользователей на предметную область с точки зрения решаемых задач, называется *концептуальной моделью данных*. Она складывается под воздействием ряда концептуальных требований, определяющихся множеством задач, для решения которых проектируется конкретная БД. Концептуальные требования определяют набор объектов, описываемых в БД, набор их атрибутов, хранящихся в БД, а также связи между объектами.

При создании концептуальной модели сначала определяются сущности и их атрибуты, назначаются первичные ключи. После создания сущностей определяются связи между ними.

Для создания моделей баз данных используются специальные программные продукты, названный CASE-средствами. Одним из таких продуктов является программа Power Designer. Он разработан фирмой Sybase и предназначен для создания различных моделей данных, отчетов и репозитариев.

Для построения концептуальной модели после запуска программы Power Designer, нужно в меню File выбрать пункт New. В появившемся окне в поле Model type нужно выбрать Conceptual model и нажать кнопку ОК.

Чтобы создать сущности, нужно на панели Palette выбрать значок сущности, а затем щелкнуть мышкой на рабочем поле. Чтобы задать имя и сущности и ее атрибуту, нужно дважды щелкнуть по ней мышкой. В появившемся окне на закладке General в соответствующем поле вводится имя сущности. Атрибуты, их типы, обязательность их заполнения в БД и наличие ключа задаются на закладке Attributes (в столбцах Name, Data type, Mandatory и Primary Key соответственно).

Чтобы создать связи между сущностями, нужно на панели Palette выбрать значок связи и протянуть связь от одной сущности к другой. Чтобы изменить сущность, нужно дважды по ней щелкнуть. В появившемся окне на закладке General определяются общие свойства. Закладка Details позволяет определить направленность связи (One – One, One – Many, и т.д.), мощность связи (0,1, 1,n и т.д.) и зависимость одной сущности от другой.

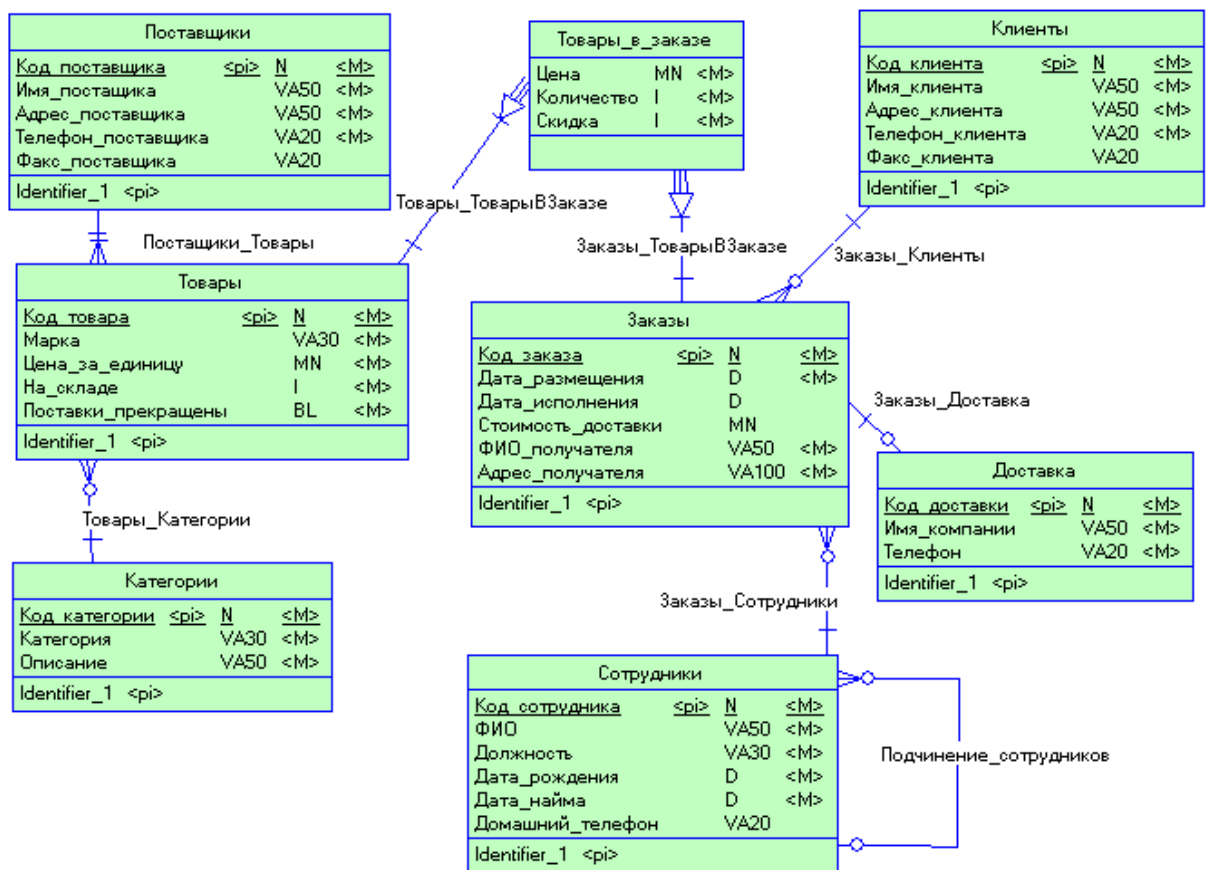


Рис. 1.1. Концептуальная модель

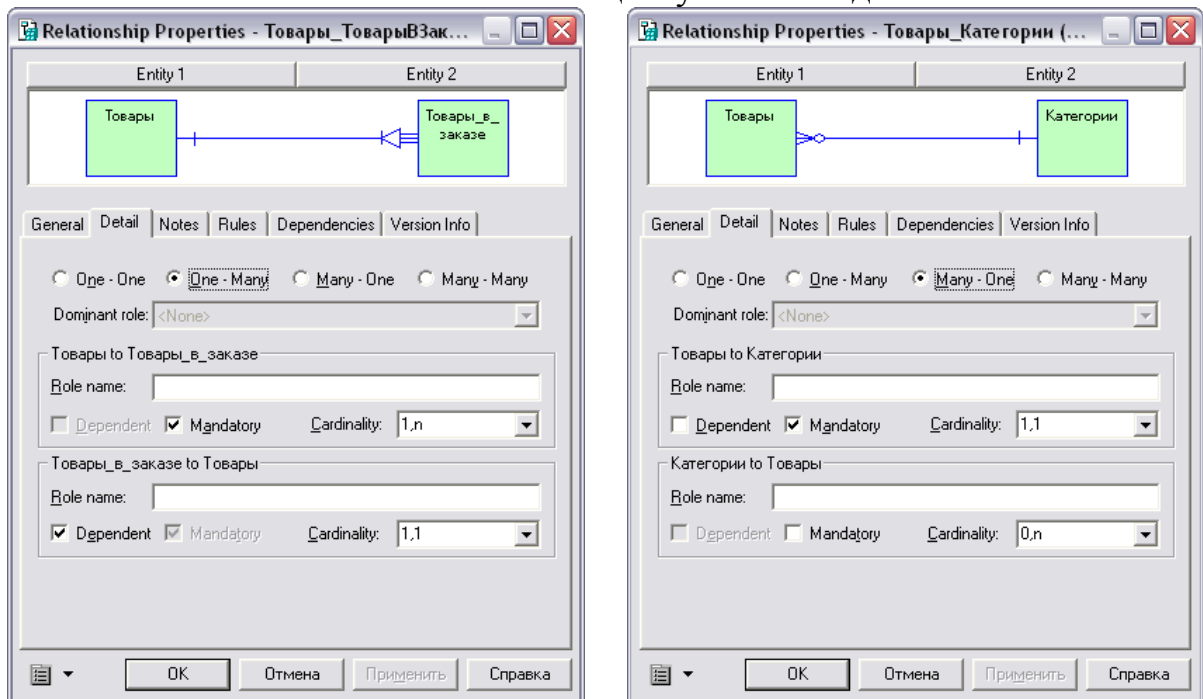


Рис. 1.2. Свойства связей

Свойства связей Товары_ТоварыВЗаказе и Товары_Категории представлены на рис. 1.2.

Если одна сущность зависима от другой, то при формировании физической модели в состав зависимой сущности в качестве ключевых атрибутов войдут ключи из связанной сущности.

Пример концептуальной модели, построенной в программном продукте

Power Designer, представлен на рис. 1.1.

1.4. Физическое моделирование

Отображение внутренней модели данных на внешний носитель с учетом особенностей файловой системы и системы ввода-вывода ЭВМ, называется физической БД или физической моделью данных.

При создании физической модели предоставляется возможность настройки модели: выбор сервера, имя, записываются ли изменения в физической модели поверх концептуальной и пр.

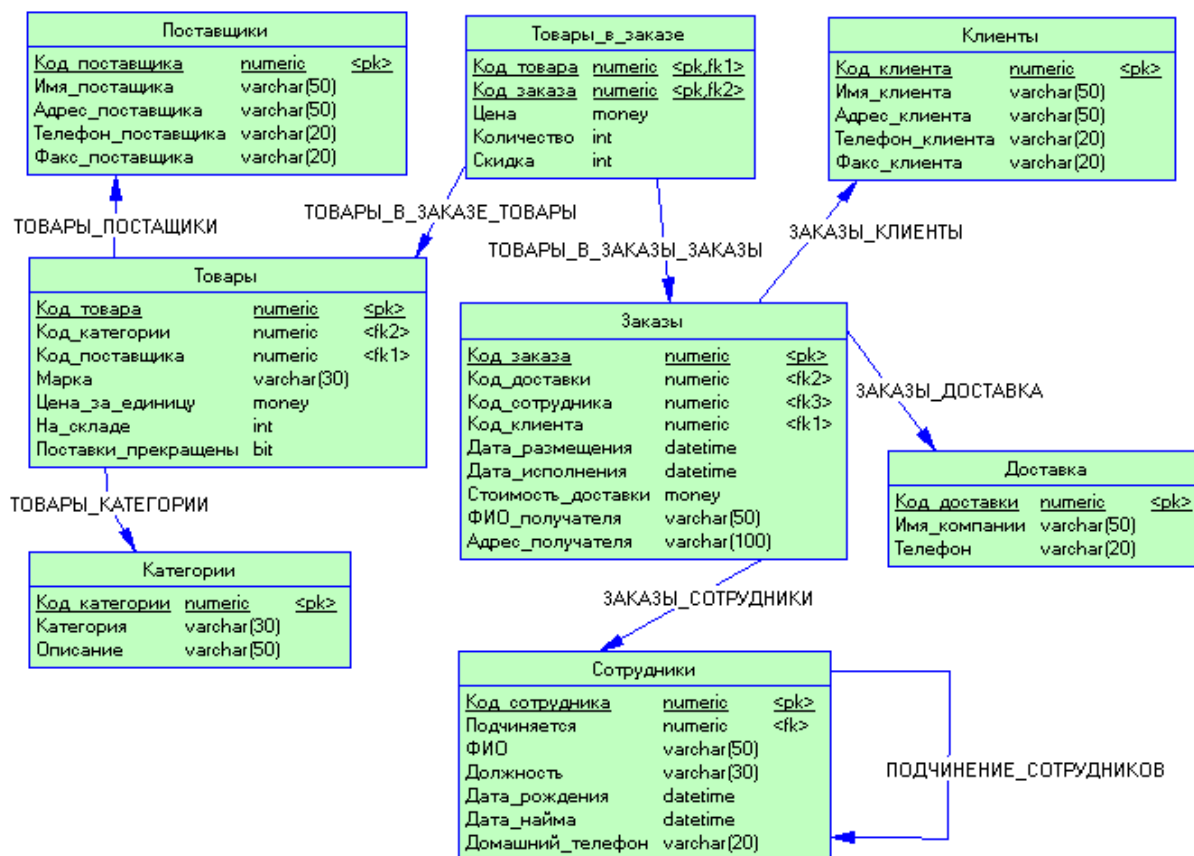


Рис. 1.3. Физическая модель

Отношения многие ко многим реализуется через создание специальных сущностей.

Чтобы создать физическую модель, нужно, прежде всего, открыть концептуальную модель. Затем в меню Tools нужно выбрать пункт Generate Physical Data Model. После этого физическая модель будет сгенерирована программой автоматически.

Пример физической модели приведен на рис. 1.3.

1.5. Создание базы данных

В соответствии с созданной физической моделью создается база данных. В программном продукте Power Designer предусмотрено средство генерации базы данных по существующей физической модели данных.

Чтобы сгенерировать базу данных, нужно в меню Database выбрать

пункт Generate Database. В открывшемся диалоговом окне можно настроить параметры генерации базы данных. После этого нужно запустить программу Microsoft SQLServer Enterprise Manager и создать новую базу данных. Для этого нужно в дереве консоли открыть папку Databases, щелкнуть в любом месте окна правой кнопкой мыши и выбрать пункт контекстного меню New Database. После этого нужно запустить программу Microsoft SQLServer Query Analyzer, выбрать в меню File пункт Open и в появившемся окне выбрать файл, созданный программой Power Designer в процессе генерации базы данных. Затем нужно на панели управления выбрать базу данных, созданную в программе Enterprise Manager и нажать кнопку Execute Query, расположенную на панели управления. Программа Query Analyzer создаст в базе данных таблицы, прописанные в файле, сгенерированном программой Power Designer.

1.6. Заполнение базы данных

На этом этапе каждая из созданных таблиц заполняется информацией.

Работа по заполнению базы данных информацией проводится в программе Microsoft SQLServer Enterprise Manager. Чтобы открыть базу данных, нужно выбрать ее в дереве консоли, расположенной в левой части рабочей области. Чтобы открыть таблицу в режиме заполнения, необходимо нажать правой кнопкой мыши по ее названию и в появившемся контекстном меню выбрать Open Table → Return all rows. В открывшемся окне вводятся значения атрибутов для каждой строки.

1.7. Контрольные вопросы:

1. Что такое сущность, атрибут, связи?
2. Какие типы ключей вы знаете?
3. Перечислите основные этапы проектирования реляционных баз данных.

2. ПРОЕКТИРОВАНИЕ СТРУКТУРЫ ХРАНИЛИЩА ДАННЫХ

2.1. Определение и типовые архитектуры хранилищ данных

Определение понятия «хранилище данных» первым дал Уильям Инмон в своей монографии – это «предметно-ориентированная, интегрированная, содержащая исторические данные, неразрушаемая совокупность данных, предназначенная для поддержки принятия управленческих решений».

Данные из различных источников помещаются в хранилище, а их описания — в репозиторий метаданных. Конечный пользователь, используя различные инструменты (средства визуализации, построения отчетов, статистической обработки и т.д.) и содержимое репозитория анализирует данные в хранилище. Результатом является информация в виде готовых отчетов, найденных скрытых закономерностей, каких-либо прогнозов. Так как средства работы конечного пользователя с хранилищем данных могут быть самыми разнообразными, то теоретически их выбор не должен влиять

на структуру хранилища и функции его поддержания в актуальном состоянии. Физическая реализация данной концептуальной схемы может быть самой разнообразной.

Двухуровневая архитектура хранилища данных подразумевает построение витрин данных (data mart) без создания центрального хранилища, при этом информация поступает из регистрирующих систем и ограничена конкретной предметной областью. При построении витрин используются основные принципы построения хранилищ данных, поэтому их можно считать хранилищами данных в миниатюре. Плюсы: простота и малая стоимость реализации; высокая производительность за счет физического разделения регистрирующих и аналитических систем, выделения загрузки и трансформации данных в отдельный процесс, оптимизированной под анализ структурой хранения данных; поддержка истории; возможность добавления метаданных.

Построение полноценного корпоративного хранилища данных обычно выполняется в трехуровневой архитектуре. На первом уровне расположены разнообразные источники данных — внутренние регистрирующие системы, справочные системы, внешние источники (данные информационных агентств, макроэкономические показатели). Второй уровень содержит центральное хранилище, куда стекается информация от всех источников с первого уровня, и, возможно, оперативный склад данных, который не содержит исторических данных и выполняет две основные функции. Во-первых, он является источником аналитической информации для оперативного управления и, во-вторых, здесь подготавливаются данные для последующей загрузки в центральное хранилище. Под подготовкой данных понимают их преобразование и проведение определенных проверок. Наличие оперативного склада данных просто необходимо при различном регламенте поступления информации из источников. Третий уровень представляет собой набор предметно-ориентированных витрин данных, источником информации для которых является центральное хранилище данных. Именно с витринами данных и работает большинство конечных пользователей.

2.2. Виды моделей хранилища данных

Хранилища строятся на основе многомерной модели данных, подразумевающей выделение отдельных измерений (время, география, клиент, счет) и фактов (объем продаж, доход, количество товара) с их анализом по выбранным измерениям. Многомерная модель данных физически может быть реализована как в многомерных, так и в реляционных СУБД. В последнем случае она выполняется по схеме «звезда» или «снежинка». Каждая таблица фактов содержит детальные данные и внешние ключи на таблицы измерений.

Теория построения многомерной модели данных и ее воплощение в реляционной структуре известна, однако информации по проблеме представлению иерархий очень мало. В качестве примера измерения, широко применяющегося при анализе деятельности предприятия и имеющего

иерархическую структуру, можно привести справочник статей затрат (рис. 2.1).

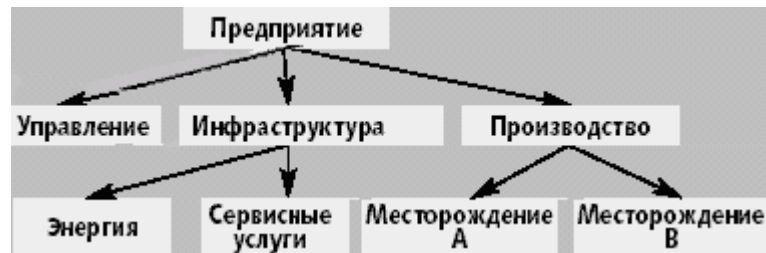


Рис. 2.1. Модель иерархического справочника

Классически проблема представления иерархий решается с помощью рекурсивной связи, что позволяет помещать в одной таблице дерево любой глубины и размерности. В нашем случае рассматриваемые данные представлены в виде таблицы 2.1.

Таблица 2.1. Представление иерархий с помощью рекурсивной связи

ID	NAME	PARENT ID
1	Предприятие	
2	Управление	1
3	Инфраструктура	1
4	Производство	1
5	Энергия	3
6	Сервисные услуги	3
7	Месторождение А	4
8	Месторождение Б	4

Однако в простоте этого решения скрывается и основной его недостаток: стандарт SQL не поддерживает рекурсивные указатели, поэтому для представления деревьев в хранилище данных используют другие методы.

Метод, предложенный Джо Селко, основан на теории множеств — все узлы дерева проходятся в прямом порядке и для каждого узла заполняются два значения (рис. 2.3).

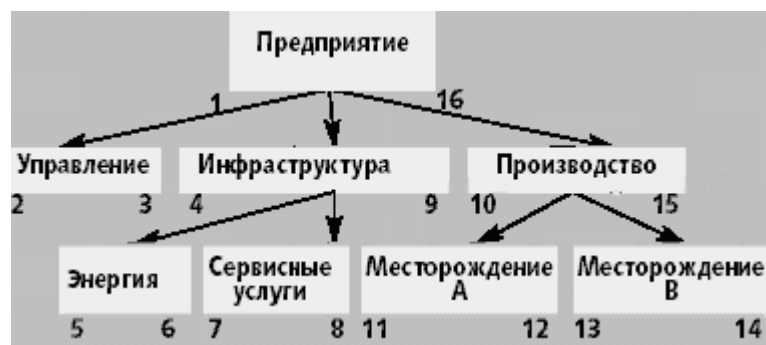


Рис. 2.3. Нумерация левой и правой границ узлов дерева

Сначала заполняется левая граница и лишь затем правая — при движении обратно от потомков к родителям. При такой нумерации узлов каждый родитель содержит потомков, левая и правая граница которых лежит в интервале между левой и правой границей родителя. Аналогично все родители потомка имеют левую границу, которая меньше левой границы потомка и правую, большую правой границы потомка. Следовательно, сумму затрат для конкретного места возникновения затрат и всех его составляющих можно получить одним запросом. Например, для получения затрат по инфраструктуре можно выполнить следующий SQL-запрос:

```
select sum(fact_table.cost)
from fact_table, dimension_table D1, dimension_table D2
where fact_table.dimension_id = D2.id
and D2.left >= D1.left
and D2.right <= D1.right
and D1.name = «Инфраструктура»
```

Для простоты работы с таким справочником кроме полей left, right стоит добавить еще два поля: «Level» — уровень узла в дереве, «Is_leaf» — флаг, показывающий является ли узел листом в дереве или нет. Таким образом, мы получаем таблицу «dimension_table» (таблица 2.2), которая позволяет хранить дерево любой глубины вложенности и размерности и выбирать потомков и родителей с помощью одного запроса.

Таблица 2.2. Представление иерархий с помощью левой и правой границ

ID	NAME	LEFT	RIGHT	LEVEL	IS_LEAF
1	Предприятие	1	16	1	N
2	Управление	2	3	2	Y
3	Инфраструктура	4	9	2	N
4	Производство	10	15	2	N
5	Энергия	5	6	3	Y
6	Сервисные услуги	7	8	3	Y
7	Месторождение А	11	12	3	Y
8	Месторождение Б	13	14	3	Y

Другой способ, описанный Ральфом Кимбаллом, основан на введении вспомогательной таблицы («helper-table»), через которую осуществляется связь таблицы фактов с таблицей измерения. Эта вспомогательная таблица отражает иерархическую структуру измерения и подчиняется следующему закону: вспомогательная таблица содержит весь набор пар «родитель-потомок», причем потомок может не быть непосредственным потомком родителя. Структура такой таблицы и ее содержимое показано в таблице 2.3.

Таблица 2.3. Структура и содержание вспомогательной таблицы

PARENT ID	CHILD ID	DISTANCE	IS LEAF
1	1	0	N
1	2	1	N
1	3	1	N
1	4	1	N
1	5	2	N
1	6	2	N
1	7	2	N
1	8	2	N
2	2	0	Y
3	3	0	N
3	5	1	N
3	6	1	N
4	4	0	N
4	7	1	N
4	8	1	N
5	5	0	Y
6	6	0	Y
7	7	0	Y
8	8	0	Y

Связывая таблицу фактов (рис. 2.4) с идентификатором потомка во вспомогательной таблице, а таблицу измерений с идентификатором родителя, мы можем вычислять сумму затрат для каждого места возникновения затрат и всех его составляющих одним запросом, как и в предыдущем случае. При этом, добавляя ограничения на поля «Distance» и «Is Leaf», мы можем легко считать затраты для любого уровня в иерархии.

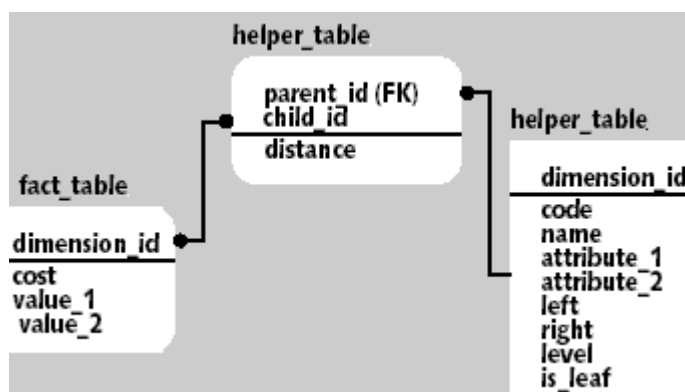


Рис. 2.4. Модель иерархического справочника с вспомогательной таблицей

Например, для того, чтобы посчитать сумму затрат, возникающих в местах, находящихся по иерархии на один уровень ниже «Инфраструктуры», необходимо выполнить следующий SQL-запрос:

```

select sum(fact_table.cost)
from fact_table, dimension_table, helper_table
where fact_table.dimension_id = helper_table.child_id
and dimension_table.dimension_id = helper_table.parent_id
and dimension_table.name = «Инфраструктура»
and helper_table.distance = 1
    
```

Проблема проектирования иерархических справочников еще более

усложняется когда измерение может иметь несколько альтернативных иерархий и становится совсем трудноразрешимой при необходимости поддерживать историю изменения таблицы измерения.

Важный момент, с которым часто приходится сталкиваться разработчику хранилища данных, связан с агрегатными значениями. Этот класс задач условно можно разделить на два подкласса. Первый рассматривает задачи создания и поддержания агрегатов по имеющимся детальным данным и широко освещен в литературе. Второй связан с тем, что источники данных для хранилища предоставляют собой не детальные значения, а уже некоторый набор агрегированных данных. Такая ситуация типична при создании хранилищ для управляющих компаний и государственных контролирующих органов, собирающих множество отчетных форм.

Крайним случаем такого подхода является модель, которую условно можно назвать «показатель-значение». Суть ее состоит в том, что собирается большой набор показателей, характеризующих финансово-хозяйственную деятельность предприятия. Эти показатели могут быть как связанными между собой функционально, так и нет, могут отражать одни и те же величины, но с разной степенью детализации и т.д. При попытке представить такие данные в виде многомерной модели разработчик сталкивается со значительными проблемами и очень часто идет по пути создания не хранилища данных, а хранилища форм. Типичное хранилище форм строится на основе трех измерений — экономические показатели, время, отчетные формы; таблицы фактов — значения экономических показателей и вспомогательных таблиц, описывающих, как показатели и их значения расположены в отчетных формах. При анализе таких данных аналитик будет испытывать значительные трудности, связанные главным образом с тем, что показатели различных форм нельзя сравнивать между собой. Единственное, что ему остается — это отслеживание изменений показателей одной формы во времени.

2.3. Многомерное моделирование и звездообразные схемы

Выше мы вскользь упомянули о том, что хранилища данных можно моделировать методами многомерного моделирования с использованием звездообразной схемы. Повторим, что традиционные методы моделирования данных и проектирования баз данных, как правило, не позволяют получить структуры данных, оптимизированные для массовых запросов. Поэтому придется отбросить некоторые традиционные идеи о проектировании и выучить ряд новых приемов. В этом разделе частично затрагиваются вопросы анализа, потому что это важно для понимания того, как определяются эти модели.

Как и большинство понятий в компьютерном мире, звездообразная схема имеет несколько синонимов: *многомерная схема*, *куб данных* и *соединение по схеме "звезда"*. Причина, по которой эту схему называют звездообразной, заключается в том, что ее графическое представление

напоминает звезду.

Звездообразная модель состоит из центральной *таблицы фактов (fact table)*, которая окружена несколькими *таблицами измерений (dimension table)*. Физически таблица фактов часто представляет собой несколько секционированных таблиц. Термин "секционирование" очень широко используется специалистами по базам данных и в общем случае означает формирование подмножеств данных.

Отношения между таблицей фактов и таблицами измерений должны быть простыми, чтобы существовал только один возможный путь соединения любых двух таблиц и чтобы смысл этого соединения был очевиден и хорошо понятен. Преимущество простоты отношений состоит в том, что это помогает повысить производительность.

Все, что аналитик должен сделать при многомерном моделировании, — это выявить факты и их измерения. Факты обычно представляют собой основные виды бизнес-деятельности организации и факторы, влияющие на данный бизнес или его сектор. Что можно сказать о таблицах измерений? В широком смысле слова — это элементы, которые могут оказывать определенное влияние или порождать различные тенденции в развитии фактов. Исходя из этого, измерения можно разбить на следующие категории: *люди, места, вещи и время* (рис. 2.5).

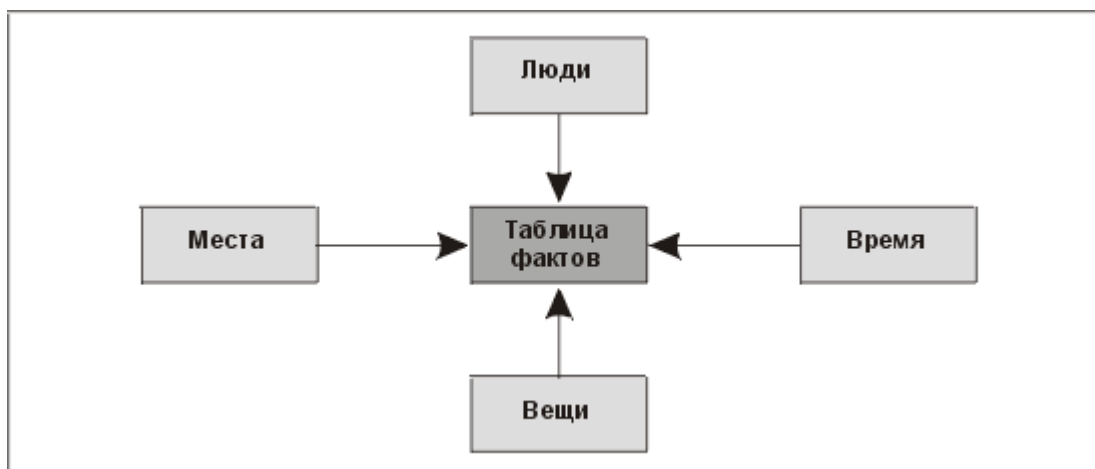


Рис. 2.5. Классификация измерений: люди, места, вещи и время

Рассматривайте таблицу фактов в свете реляционной теории: у нее есть внешний ключ к каждой соответствующей таблице измерения. Она объединяет измерения в нечто, представляющее собой значимое событие, и содержит качественную и количественную информацию об этом событии. В сценарии продаж таблица фактов может содержать данные о том, сколько товара реализовано и на какую сумму, а также внешние ключи к таблицам измерений, характеризующих операцию продажи (какой товар, когда он был продан, кто продал его и какой способ платежа был использован).

Временное измерение в звездообразной схеме обладает рядом интересных свойств. Факт, скорее всего, будет связан не с одним моментом времени, а с каким-то временным промежутком. Следовательно, временное

измерение обычно является обобщающим. При выборе единицы изменения времени приходится выбирать между хранением огромного числа фактов в хранилище (которые в большинстве запросов необходимо будет обобщать) и отсутствием данных о сезонных тенденциях вследствие того, что при обобщении выбран слишком большой период времени. Временное измерение также часто выступает в качестве основы для секционирования таблицы фактов. Во многих случаях секционирование необходимо для уменьшения объема таблицы фактов и облегчения управления ею. Время же используется потому, что при этом, как правило, получают приблизительно равные по объему кванты, и оно часто является наиболее логичным для секционирования измерением. Кроме того, время — наиболее вероятная основа для усечения таблицы фактов (если когда-нибудь удастся заставить предприятие согласиться на удаление данных из его хранилища).

Выбор степени детализации временного измерения (неделя, день, месяц и т.д.) может иметь исключительную важность. Если, например, выбрать месяц, то средство анализа, возможно, сможет обобщить данные в еще большей степени и показать результаты по годам. Затем с помощью различных методов детализации можно представить данные по кварталам нужного года, а потом вновь по месяцам. Очевидно, что в этом случае нельзя выполнить более подробную временную детализацию, поскольку для этого нет данных. Выбор степени детализации по времени имеет большое значение, потому что позволяет уравновесить два фактора — необходимость в более детальной разбивке и объем хранилища данных. Таким образом, при проектировании хранилища данных необходимо уделять должное внимание временным измерениям и в каждом конкретном случае принимать решение о необходимой для обобщения степени детализации.

2.4. Пример «Система мониторинга подземных вод»

Рассмотрим процесс проектирования хранилища данных на примере системы мониторинга подземных вод.

Первым шагом проектирования структуры хранилища данных является выявление измерений и фактов. Количество таблиц фактов и измерений для этих фактов может варьироваться в зависимости от размера хранилища и его использования. В нашем примере это четыре измерения («Скважины», «Количество воды», «Цели отбора» и время.) и один факт (сам отбор воды в определенное время с целью мониторинга состояния подземных вод).

Очень важным моментом является выбор степени детализации временных измерений, она может составлять годы, месяцы, недели, дни и так далее. Эта характеристика выбирается непосредственно исходя из специфики проектируемого хранилища данных и его будущего использования. В данном примере целесообразно обобщать данные по неделям.

На рис. 2.6. представлена звездообразная схема для системы мониторинга подземных вод.

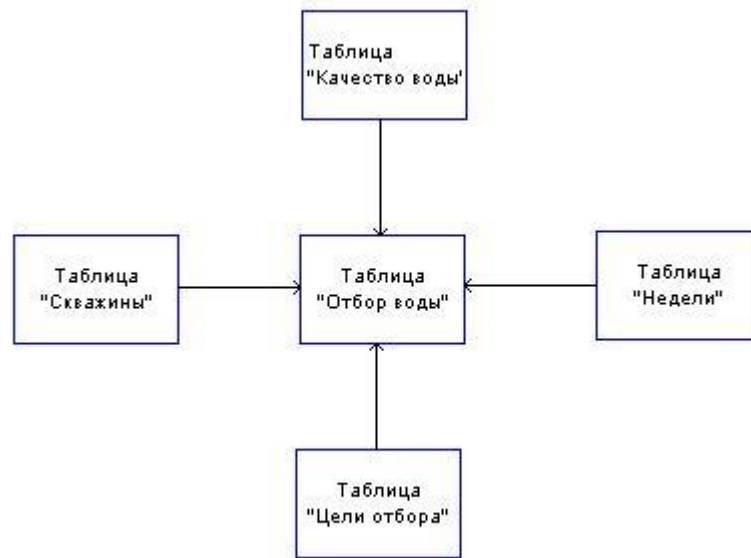


Рис. 2.6. Звездообразная схема системы мониторинга подземных вод

Далее следует определить таблицы измерений и фактов, т.е. наполнение их непосредственно атрибутами. На рис. 2.7 показаны определения таблиц, которые используются для реализации звездообразной схемы, изображенной на рис. 2.6.

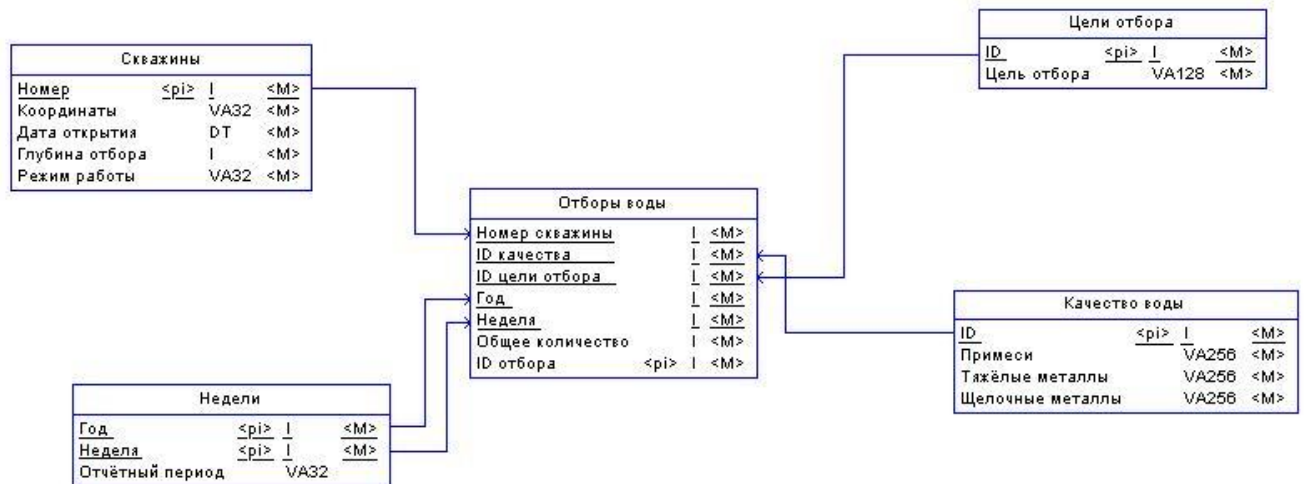


Рис. 2.7. Определения таблиц для реализации звездообразной схемы сбыта продукции

После определения таблиц необходимо проверить таблицы измерений на предмет множественного соединения между ними. В данном случае таблицы «Скважины» единственным образом связана с таблицей «Цели отбора» через таблицу фактов. Тоже самое справедливо и для остальных таблиц, что удовлетворяет основному требованию к таблицам измерений и

фактов.

Для исключения возможности использования лишних измерений необходимо проверить, все ли из выбранных измерений влияют на таблицу фактов. Так например, «Качество воды» влияет на таблицу фактов, т.к. при обобщении сведений по неделе может оказаться, что при неизменных остальных измерениях «Качество воды» будет иным, чем в предыдущий промежуток, тем самым факт изменился. Если окажется так, что данное измерение лишнее, его следует удалить.

2.5. Контрольные вопросы

1. Какие типовые архитектуры хранилищ данных вам известны?
2. Чем обусловлен выбор модели хранилища данных?
3. В чем особенность звездообразных схем структуры хранилищ данных?

2.6. Задание

На основании БД, созданной в ходе предыдущей лабораторной работы, спроектируйте структуру хранилища данных. Для этого:

1. Выделите измерения и факты хранилища данных.
2. Определите структуры измерений и фактов хранилища данных.
3. Проверьте таблицы на предмет множественности связей между ними.
4. Проверьте, все ли измерения влияют на таблицу фактов.

3. РЕАЛИЗАЦИЯ ХРАНИЛИЩА ДАННЫХ ПОД УПРАВЛЕНИЕМ MICROSOFT SQL SERVER 2000 ANALYSIS SERVICES

3.1. Измерения(Dimension)

В службах Microsoft SQL Server 2000 Analysis Services (SSAS) измерения являются основными компонентами куба. В измерениях данные привязаны к некоторой предметной области, например заказчики, магазины или служащие. В службах Analysis Services измерения содержат атрибуты, которые соответствуют столбцам в таблицах измерения. Эти атрибуты выглядят в виде иерархий атрибутов и могут быть организованы в многоуровневые иерархические структуры. Эти иерархии применяются для организации мер, которые содержатся в кубе.

Все измерения основаны на таблицах или представлениях в представлении источника данных. Измерения существуют независимо от куба, могут использоваться в нескольких кубах и могут использоваться многократно в одном кубе. Измерение, существующее независимо от куба, называется измерением базы данных, а измерение, используемое в кубе, называется измерением куба.

Структура измерения в основном определяется структурой таблицы

или таблиц базового измерения. Простейшая структура называется схемой "звезда", в которой каждое измерение основано на одной таблице измерения, которая непосредственно связана с таблицей фактов связью первичного и внешнего ключей.

Простейшим методом определения измерений базы данных и куба является использование мастера кубов для создания их одновременно с определением куба. Мастер кубов создаст измерения на основе таблиц измерений, найденных им или указанных пользователем из представления источника данных, используемого для куба. После этого мастер создает измерения базы данных и добавляет их к новому кубу, создавая измерения куба.

При создании куба также можно добавить любые измерения, которые уже существуют в базе данных. Эти измерения могли быть ранее определены для другого куба или мастером измерений. После создания измерения базы данных его можно отредактировать в конструкторе измерений.

3.2. Мера(Measure)

Как правило, мера представляет собой столбец, содержащий количественные, статистически вычисляемые (обычно числовые) данные, соответствующие столбцу в таблице фактов. Выражение меры также может использоваться для определения значения меры на основе столбца в таблице фактов, измененной многомерным выражением. Выражение меры позволяет выполнить взвешивание значения меры, например, для преобразования валют при взвешивании меры продаж по валютному курсу.

Для определения мер можно использовать столбцы атрибутов таблиц измерений, но такие меры являются, как правило, полуаддитивными или неаддитивными в зависимости от режима статистического вычисления

Кроме того, можно определить вычисляемую меру с помощью многомерных выражений, чтобы получить вычисляемое значение меры на основе других мер в кубе. Такие меры, называемые вычисляемыми, добавляют кубу в службах Analysis Services гибкие функциональные возможности анализа. В кубе меры сгруппированы по таблице фактов в группу мер. Группы мер используют для связи измерений с мерами. Группы мер также используют для меры с отдельным счетчиком в качестве режима статистического вычисления, что позволяет выполнять оптимальное статистическое вычисления.

Пересечение отдельной меры и отдельного элемента всех измерений, включенных в группу мер, представлено в службах Analysis Services в виде отдельной ячейки куба.

Меры и группы мер имеют свойства, позволяющие определять и контролировать, как меры и группы мер функционируют и отображаются для пользователей.

Свойства группы мер

Свойства группы мер определяют характеристики всей группы мер и

устанавливают характеристики по умолчанию для определенных свойств мер в группе мер.

Свойство	Определение
AggregationPrefix	Общий префикс, используемый для имен агрегатов.
DataAggregation	Определяет, могут ли службы Analysis Services осуществлять статистическое вычисление сохраняемых данных или кэшированных данных для группы мер. По умолчанию статистическое вычисление сохраняемых данных и кэшированных данных разрешено.
ErrorConfiguration	Настраиваемые параметры обработки ошибок для обработки дублирующихся ключей, неизвестных ключей, пределов ошибок, действий при обнаружении ошибки, файла журнала ошибок и ключа NULL.
IgnoreUnrelatedDimensions	Определяет, будут ли несвязанные измерения принудительно перемещаться на их верхний уровень, когда элементы измерений, не связанных с группой мер, включаются в запрос. Значение по умолчанию равно True.
ProactiveCaching	Настраиваемые параметры обработки ошибок для обработки дублирующихся ключей, неизвестных ключей, пределов ошибок, действий при обнаружении ошибки, файла журнала ошибок и ключа NULL
ProcessingMode	Указывает, должны ли статистическое вычисление и индексирование выполняться во время или после обработки (параметры - обычная или отложенная).
ProcessingPriority	Определяет приоритет обработки куба во время фоновых операций, например отложенных статистических вычислений или индексирования. Значение по умолчанию равно 0.
StorageLocation	Место хранения файловой системы для группы показателей. Если местоположение не указано, то оно наследуется из куба, содержащего эту группу мер.
StorageMode	Режим хранения для группы мер, значения MOLAP, ROLAP или HOLAP.
Type	Указывает тип группы мер.

Свойства мер

Меры наследуют определенные свойства у группы мер, элементами которых они являются, если только эти свойства не переопределены на уровне мер. Свойства мер определяют статистическое вычисление мер, их тип данных, папку для отображения мер, строку форматирования, любые выражения мер, их понятные имена, базовые исходные столбцы и видимость для пользователей.

Свойство	Определение
AggregateFunction	Определяет статистическое вычисление мер. Дополнительные сведения приведены ниже.
DataType	Тип данных столбца базовой таблицы фактов, с которым связана мера.
Description	Описание меры, которая может открываться в клиентских приложениях.
DisplayFolder	Папка, в которой группа мер будет отображаться, когда пользователи подключаются к кубу, позволяющая распределять меры по категориям для повышения удобства просмотра пользователями куба, содержащего множество мер.
FormatString	Определяет формат отображения. Более подробные сведения приведены ниже.
ID	Уникальный идентификатор (ID) меры.
MeasureExpression	Содержит многомерное выражение, определяющее меру.
Name	Имя меры, понятное для пользователя.
Source	Столбец в представлении источника данных, к которому привязана мера.
Visible	Определяет, отображается мера или нет

Куб (cube)

В службах Microsoft SQL Server 2000 Analysis Services куб разрабатывается на основе таблиц и представлений, моделируемых в представлении источника данных. Куб представляет собой набор мер, которые являются фактами, и измерений, которые являются интересующими областями, такими как, например время, продукт и заказчик. Куб дополняется вычислениями, ключевыми индикаторами производительности, действиями, секциями, перспективами и преобразованиями. Куб, в сущности, является синонимом унифицированной многомерной модели. Меры куба основаны на столбцах из одной или нескольких таблиц фактов, а элементы измерений куба основаны на столбцах из одной или нескольких таблиц измерений. Куб также может быть разработан без базового реляционного источника данных. В этом случае может быть сформирована базовая реляционная структура для поддержки куба. Факты в кубе статистически вычисляются на основе иерархий измерений.

Свойства кубов

Кубы имеют ряд свойств, изменяя параметры которых можно влиять на поведение всего куба, и несколько неизменяемых свойств. Данные свойства перечислены в следующей таблице:

Свойство	Определение
AggregationPrefix	Общий префикс, используемый для имен агрегатов.
Collation	Идентификатор языка и флаг сравнения, отделенный подчеркиванием, например Latin1_General_C1_AS.
DefaultMeasure	Многомерное выражение, определяющее меру по умолчанию для куба.
Description	Описание куба, которое может открываться в клиентских приложениях.
ErrorConfiguration	Настраиваемые параметры обработки ошибок для обработки дублирующихся ключей, неизвестных ключей, пределов ошибок, действий при обнаружении ошибки, файла журнала ошибок и ключа NULL.
EstimatedRows	Количество предполагаемых строк в кубе.
ID	Уникальный идентификатор (ID) куба.
Language	Идентификатор языка куба.
Name	Имя куба, понятное для пользователя.
ProactiveCaching	Настройки упреждающего кэша для куба.
ProcessingMode	Указывает, должны ли статистическое вычисление и индексирование выполняться во время или после обработки (параметры - обычная или отложенная).
ProcessingPriority	Определяет приоритет обработки куба во время фоновых операций, например отложенных статистических вычислений или индексирования. Значение по умолчанию равно 0.
ScriptCacheProcessingMode	Указывает на необходимость создания кэша сценария во время или после обработки (параметры - обычный или отложенный).
ScriptErrorHandlingMode	Определяет обработку ошибок, параметры - IgnoreNone или IgnoreAll.
Source	Представление источника данных, используемое для куба.
StorageLocation	Место хранения файловой системы для куба. Если не задано иное, то место хранения наследуется из

	базы данных, содержащей объект куба.
StorageMode	Режим хранения для куба, значения - MOLAP, ROLAP или HOLAP.
Visible	Определяет, отображается куб или нет.

3.4. Реализация

Analysis Manager – программа, работающая под управлением Microsoft® Management Console (MMC). Она представляет собой утилиту, входящую в состав Analyses Services и предназначенную главным образом для администраторов баз данных OLAP.

Как запускать Analysis Manager

Щелкнуть кнопку Start, пункт Программы, Microsoft SQL Сервер, Analyses Services и затем щелкните Analysis Manager.

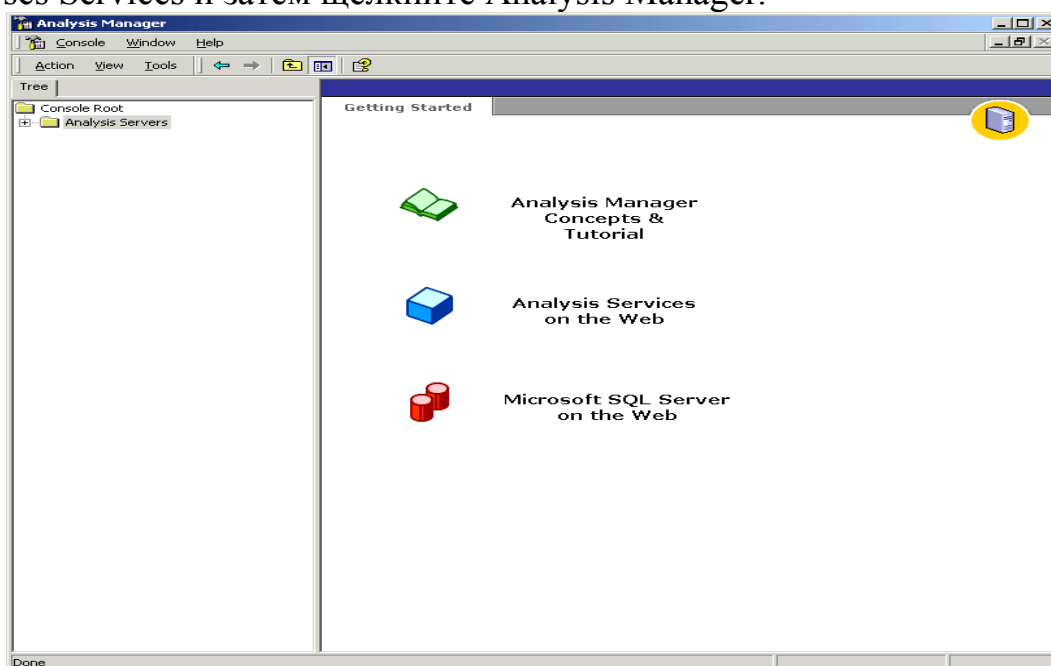


Рис.3.1. Окно программы Analysis Manager

Как настраивать структуру базы данных

1. В Analysis Manager в дереве раскройте Серверы Анализа.
2. Щелкнуть название вашего сервера. Подключение с сервером анализа будет установлено.
3. Щелкнуть правой кнопкой мыши на названии вашего сервера, и затем щелкните New Database.
4. В диалоговом окне Database, в блоке имени Базы данных, вводите Tutorial, и затем щелкните ОК.
5. В области окна дерева Analysis Manager, раскройте сервер, и затем расширите базу данных Tutorial, которую Вы только что создавали.

Ваша новая база данных Tutorial содержит следующие элементы:

- Источники Данных
- Кубы

- Общедоступные Измерения
- Добывающие Модели
- Роли Базы данных

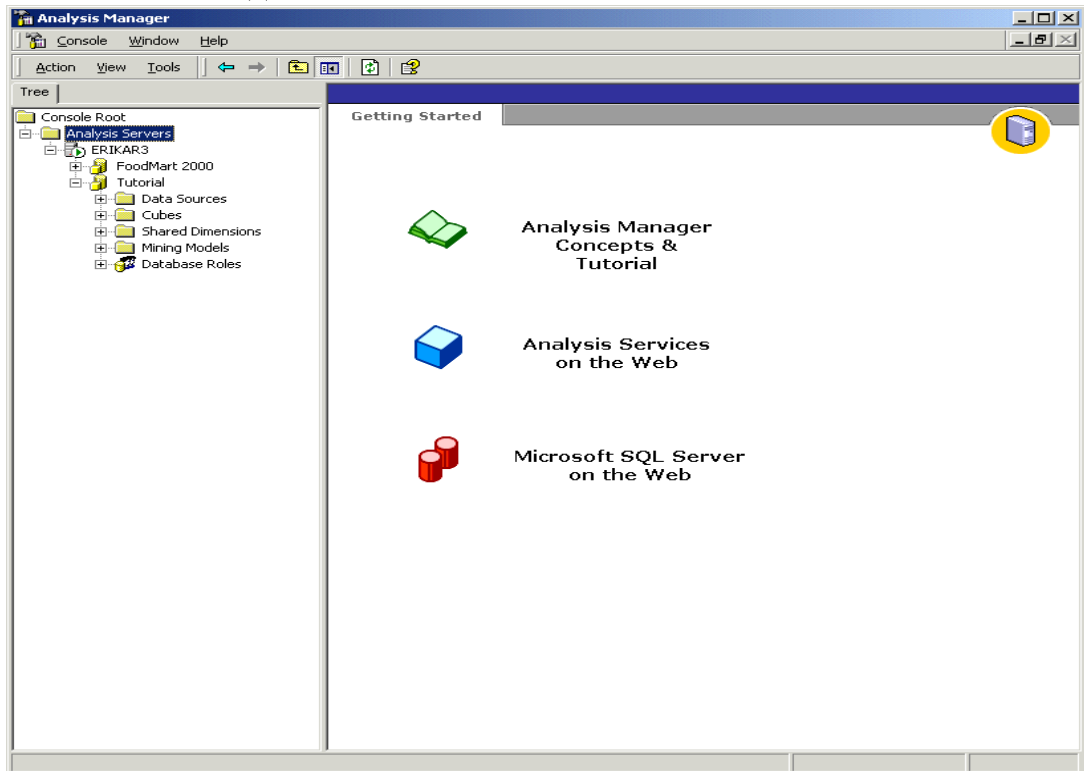


Рис.3.2. База данных Tutorial

Затем установите подключение к типовым данным в источнике данных Tutorial. Этот пример будет использоваться для всех упражнений в этой обучающей программе.

Установка источника данных в Менеджере Анализа подключает вашу базу данных к системному имени источника данных (DSN), которое устанавливается в ODBC Администраторе Источника Данных.

Как настраивать источник данных

1. В Analysis Manager в области дерева щелкните правой кнопкой мыши на папке Источников Данных под базой данных Tutorial, и затем щелкните New Data Source.

2. В диалоговом окне Data Link Properties, щелкните вкладку Provider, и затем щелкните Microsoft OLE DB Provider for ODBC Drivers.

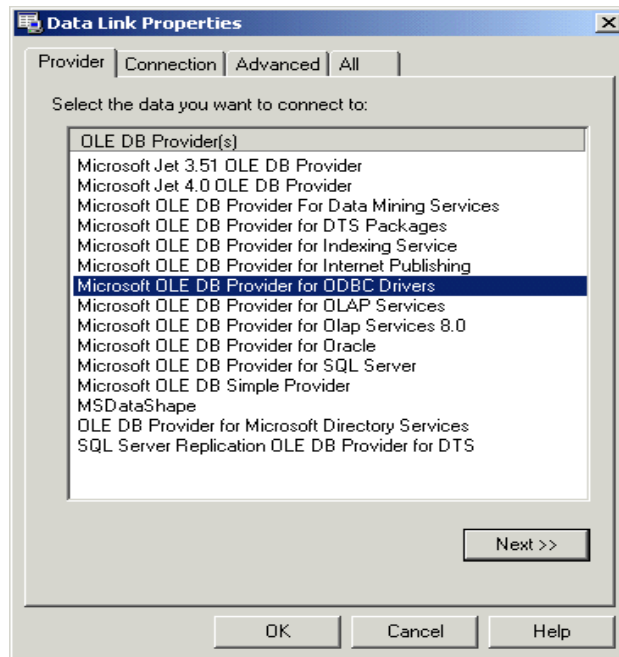
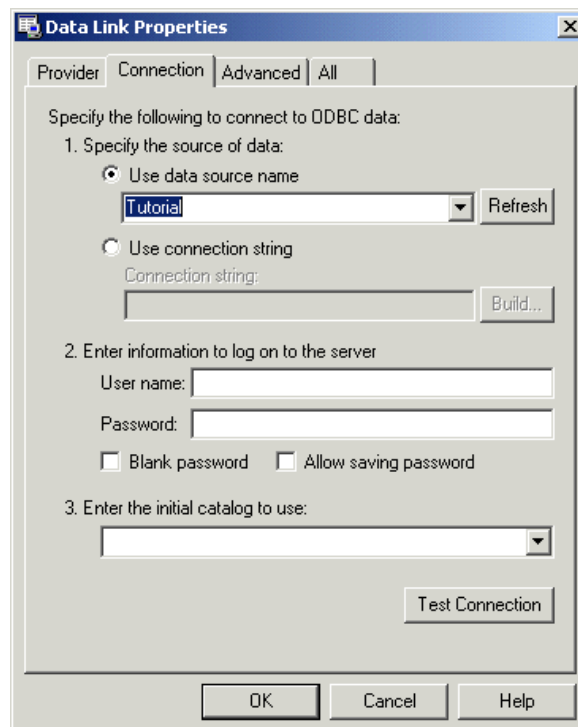


Рис.3.3. Выбор провайдера

3. Щелкнуть вкладку Connection, и затем в списке имен источников данных щелкните Tutorial.



4. Щелкнуть Test Connection, чтобы убедиться, что все работает.

5. Щелкнуть ОК, чтобы закрыть диалоговое окно Data Link Properties.

Построение куба

Куб - многомерная структура данных. Кубы определены набором измерений и мер.

Как открыть Мастер Куба

В Менеджере Анализа в области окна дерева, под базой данных Tutorial, щелкнете правой кнопкой мыши на папке кубов, выберите новый куб и затем щелкнете Wizard.

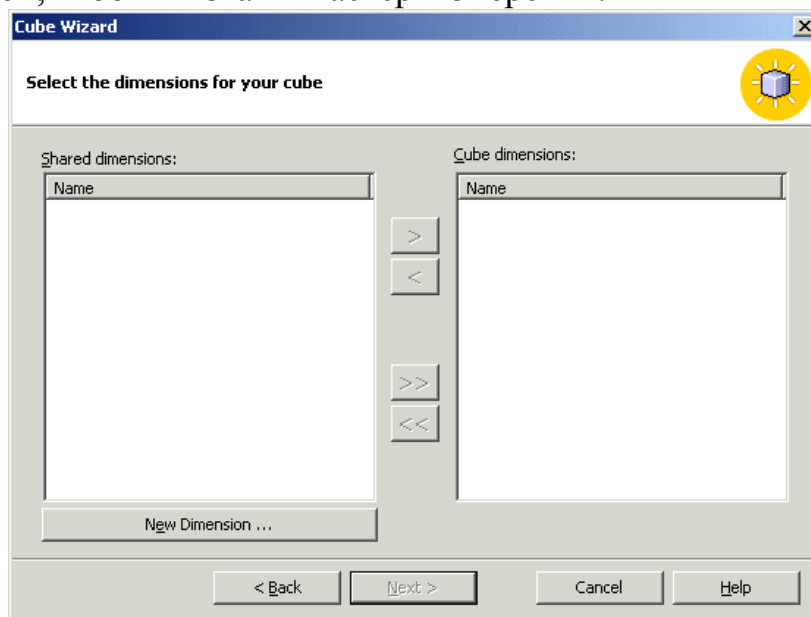
Как добавлять меры к кубу

Меры - количественные значения в базе данных, которую Вы хотите анализировать. Обычно-используемые меры коммерческие, стоимость, и данные бюджета. Меры проанализированы против различных категорий измерения куба.

1. На шаге Welcome Мастера Куба, щелкните Next.
2. В выборе таблицы факта на шаге источника данных, раскройте источник данных Tutorial, и затем щелкните sales_fact_1998.
3. Вы можете рассматривать данные в таблице sales_fact_1998, щелкая данные Browse.
4. Чтобы определить меры для вашего куба, под таблицей Факта, дважды щелкните на столбце store_sales. Повторите эту процедуру для столбцов store_cost и unit_sales, и затем щелкните Next.

Как построить измерение времени

1. На шаге Select the dimensions for your cube мастера Cube, щелкните New Dimension, чтобы вызвать Мастер Измерения.



2. На шаге Welcome щелкните Next.
3. Выберите Star Schema для одиночной таблицы измерения и затем щелкните Next.
4. На шаге выбора таблицы измерения, щелкните time_by_day. Вы можете просматривать данные, содержащиеся в time_by_day таблице, щелкая Browse Data.

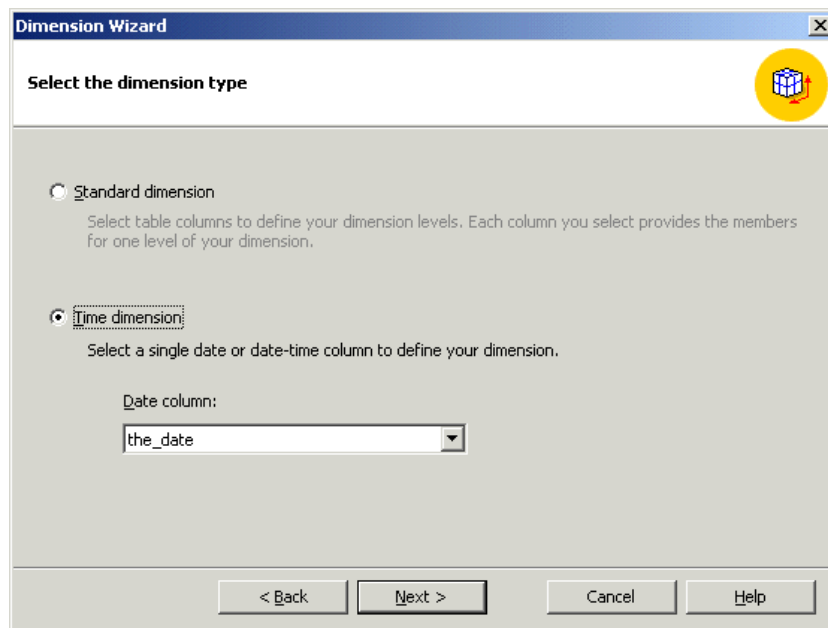


Рис.4. Окно выбора типа измерения

5. На шаге выбора типа измерения, выберите Time Dimension, и затем щелкните Next.

6. Затем определите уровни для вашего измерения. На следующем шаге щелкните Select time level, выберите Year, Quarter, Month, и затем щелкните Next.

7. На шаге дополнительных параметров выбора, щелкните Next.

8. На последнем шаге мастера введите имя вашего нового измерения.

Обратите внимание: Вы можете определить, будет ли это измерение разделено или возможно только частное использование ресурсов этого измерения, при помощи переключателя, который расположен в нижнем левом углу экрана.

9. Щелкните Finish, чтобы вернуться к Мастеру Куба.

10. В Мастере Куба теперь отражается измерение времени в списке измерений Куба.

Как построить измерение product

1. Щелкнуть New dimension снова. На шаге Welcome Мастера Измерения щелкните Next.

2. Выберите Snowflake Schema и затем щелкните Next.

3. На шаге выбора таблиц измерения дважды щелкните product и product_class, чтобы добавить их к выбранным таблицам. Щелкните Next.

4. Эти две таблицы, которые вы выбрали на предыдущем и текущем шаге соединяются между собой. Щелкните Next.

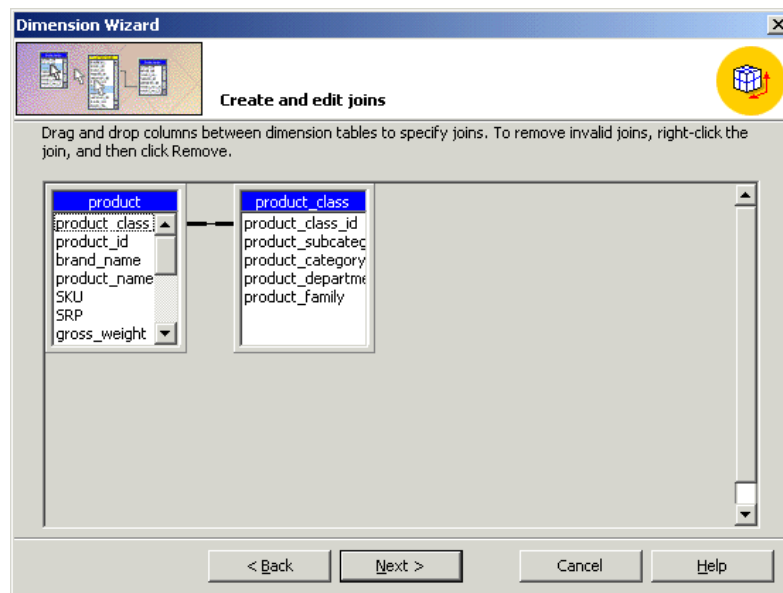


Рис.3.5. Окно создания и редактирования соединения

5. Чтобы определить уровни для измерения, под столбцами Available, дважды щелкните product_category, product_subcategory, и brand_name столбцы, в том же порядке. После того, как вы дважды щелкаете каждый столбец, его имя появляется под уровнями измерений. Щелкните Next после того, как выбрали все три столбца.

6. На шаге определения ключей столбцов щелкните Next.

7. На шаге дополнительных параметров выбора щелкните Next.

8. На последнем шаге мастера введите product в блок названия измерения. Щелкните Finish.

9. Вы должны увидеть измерение product в списке измерений куба.

Как построить измерение Customer

1. Щелкнуть New dimension.

2. На шаге Welcome Мастера Измерения щелкните Next.

3. Выберите Star Schema и затем щелкните Next.

4. На шаге выбора таблицы измерения, щелкните Customer, и затем щелкните Next.

5. На шаге выбора типа измерения щелкните Next.

6. Чтобы определить уровни для вашего измерения, под столбцами Available дважды щелкните Country, State_Province, City, и lname столбцы, в том же порядке. После того, как вы дважды щелкаете каждый столбец, его имя появляется под уровнями измерений. После того, как вы выбрали все четыре столбца, щелкните Next.

7. На шаге определения ключей столбцов щелкните Next.

8. На шаге дополнительных параметров выбора щелкните Next.

9. На последнем шаге мастера введите Customer в блок названия измерения. Щелкните Finish.

10. Вы должны увидеть измерение Customer в списке измерений куба.

Как построить измерение store

1. Щелкнуть New dimension.
2. На шаге Welcome Мастера Измерения щелкните Next.
3. Выберите Star Schema и затем щелкните Next.
4. На шаге выбора таблицы измерения, щелкните store, и затем щелкните Next.
5. На шаге выбора типа измерения щелкните Next.
6. Чтобы определить уровни для вашего измерения, под столбцами Available дважды щелкните store_country, store_state, store_city, и store_name столбцы, в том же порядке. После того, как вы дважды щелкаете каждый столбец, его имя появляется под уровнями измерений. После того, как вы выбрали все четыре столбца, щелкните Next.
7. На шаге определения ключей столбцов щелкните Next.
8. На шаге дополнительных параметров выбора щелкните Next.
9. На последнем шаге мастера введите store в блок названия измерения. Щелкните Finish.
10. Вы должны увидеть измерение store в списке измерений куба.

Как закончить построение куба

1. В мастере куба щелкните Next.
2. Щелкнуть Yes в окне сообщения.

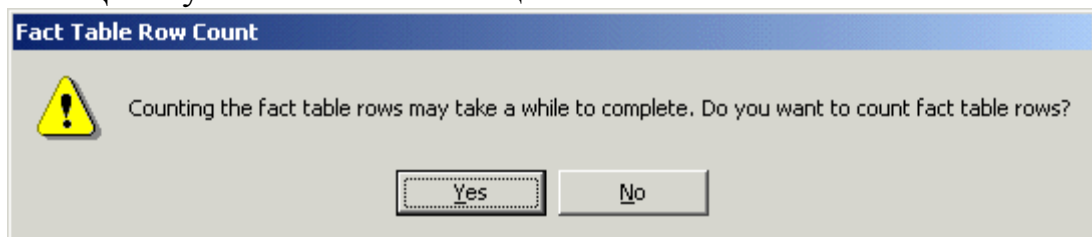


Рис.6. Окно Fact Table Row Count

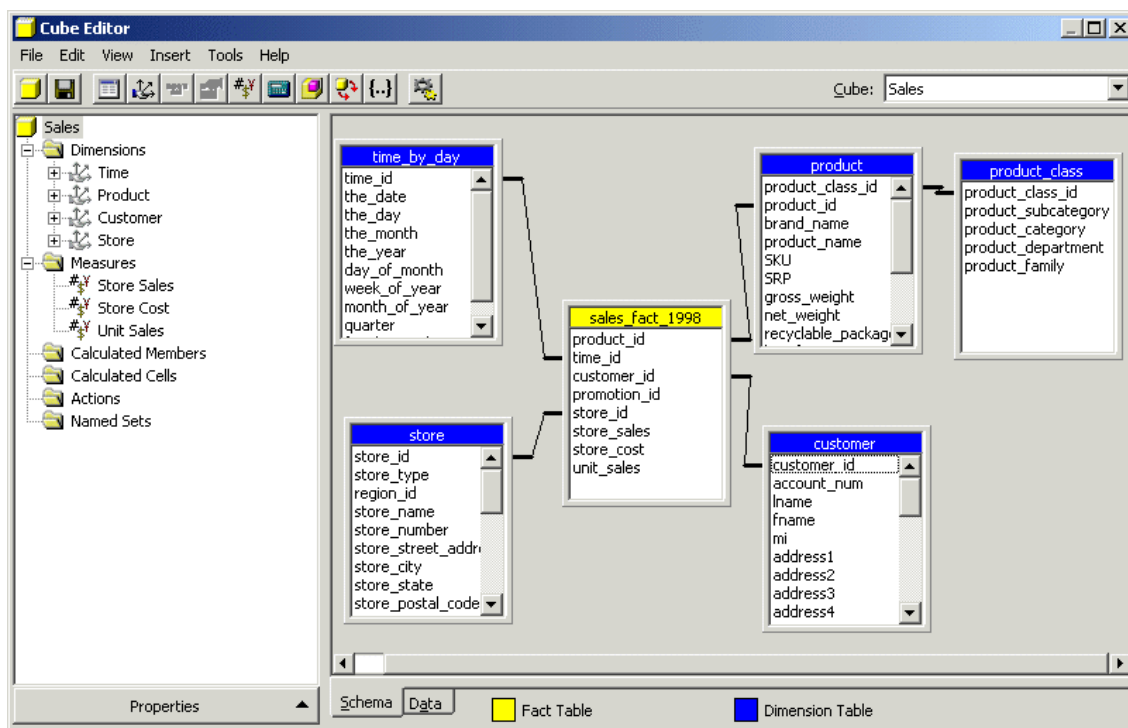


Рис.3.7. Окно редактирования куба

3. На последнем шаге мастера куба назовите ваш куб Sales и затем щелкните Finish.

4. Мастер закрывается и затем запускается Редактор Куба, который содержит новый куб. Нажимая по синим или желтым областям заголовка, упорядочьте таблицы так, чтобы они соответствовали следующей иллюстрации. Обратите внимание: вы не должны закрывать Редактор Куба, т.к. будете редактировать куб в следующем разделе обучающей программы.

Редактирование куба

Вы можете использовать два метода для доступа к редактору куба:

- в менеджере анализа в области дерева щелкните правой кнопкой мыши на существующем кубе и затем щелкните Edit;
- можно создать новый куб, используя редактора куба непосредственно. Этот метод не рекомендуется, если вы не опытный пользователь.

Если вы продолжаете от предыдущего раздела, вы должны уже быть в редакторе куба.

В области окна схемы редактора куба, вы можете видеть таблицу фактов (с желтой областью заголовка) и соединенными таблицами измерений (синие области заголовка). В области окна дерева редактора куба возможен предварительный просмотр структуры куба в иерархическом дереве. Вы можете редактировать свойства куба, щелкая кнопку Properties внизу левой области окна.

Хранение проекта и обработка куба

Вы можете проектировать опции хранения для данных в вашем кубе.

Прежде, чем вы можете использовать или просматривать данные в ваших кубах, вы должны обработать их.

Как проектировать хранение, используя мастер проекта хранения

1. В область дерева менеджера анализа раскройте папку кубов, щелкните правой кнопкой мыши на Commerce Cube и затем щелкните Design Storage.

2. На шаге Welcome щелкните Next.

3. Выберите MOLAP в качестве вашего типа хранения данных и затем щелкните Next.

4. Под опциями набора нажмите Performance gain reaches. В блоке введите 40%. Это настраивает Analyses Services на повышение производительности 40 %, независимо от того, сколько дискового места это требует. Администраторы могут использовать эту способность настройки для сбалансирования потребности в производительности запросов и дискового пространства, требуемого, чтобы хранить данные агрегатов.

5. Щелкните Start.

6. Вы можете наблюдать выполнение в правой части мастера. По окончании работы щелкните Next.

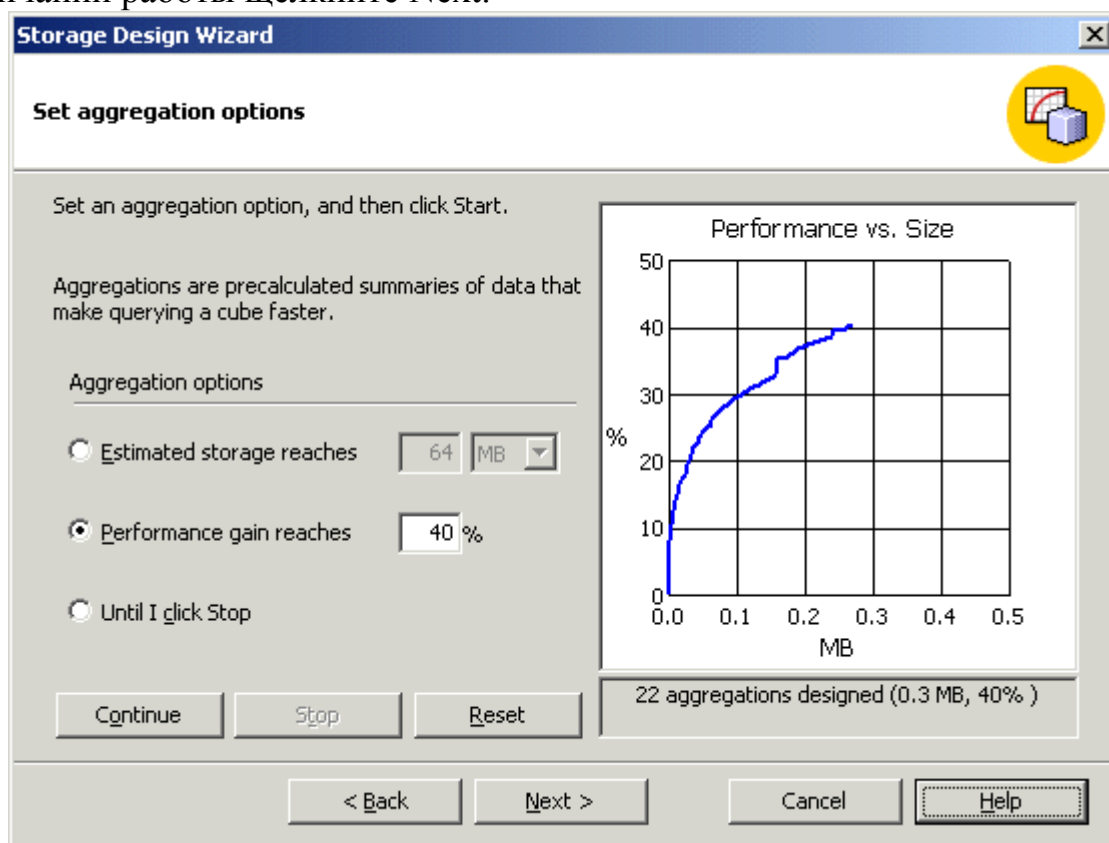


Рис.3.9.Окно мастера проекта хранения

7. На следующем шаге выберите Process и затем щелкните Finish.

Обратите внимание: обработка агрегатов может занимать некоторое время.

8. В появившемся окне вы можете наблюдать ваш куб, в то время как происходит обработка. Когда обработка закончена, выдается сообщение,

подтверждая, что обработка была закончена успешно.

9. Необходимо щелкнуть Close для возвращения к менеджеру анализа. Теперь вы готовы просмотреть данные в кубе.

3.5. Контрольные вопросы

1. В чем разница между измерением и мерой?
2. Каковы особенности реализации многомерных кубов данных в Microsoft SQL Server 2000 Analysis Services?
3. Опишите основные этапы реализации хранилищ данных в Microsoft SQL Server 2000 Analysis Services

Критерии оценки:

- оценка «отлично» выставляется студенту, если он показал высокий уровень сформированности компетенции, верно и в полном объеме показал глубокие исчерпывающие знания всего программного материала по дисциплине, понимание сущности проблемы; дал логически последовательные, содержательные, полные правильные и конкретные ответы на все вопросы.

- оценка «хорошо» выставляется студенту, если он показал средний уровень сформированности компетенции, с незначительными замечаниями показал твердые и достаточно полные знания всего программного материала по дисциплине, правильное понимание сущности проблемы. Дал последовательные, правильные, конкретные ответы на поставленные вопросы при свободном устранении замечаний по отдельным вопросам.

- оценка «удовлетворительно» выставляется студенту, если он показал низкий уровень сформированности компетенции, на базовом уровне с ошибками показал знание и понимание сущности проблемы; дал правильные и конкретные, без грубых ошибок ответы на поставленные вопросы.

- оценка «неудовлетворительно» выставляется студенту, если он показал недостаточный уровень сформированности компетенции, дал неправильный ответ хотя бы на один из основных вопросов, сделал большое количество ошибок в ответе, не понимает сущности излагаемых вопросов; дает неуверенные и неточные ответы на дополнительные вопросы.

2. Материалы для проведения текущей аттестации

Текущая аттестация 1

**АВТНОМНАЯ НЕКОММЕРЧЕСКАЯ ОБРАЗОВАТЕЛЬНАЯ
ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
ЦЕНТРОСОЮЗА РОССИЙСКОЙ ФЕДЕРАЦИИ
«РОССИЙСКИЙ УНИВЕРСИТЕТ КООПЕРАЦИИ»
КАЗАНСКИЙ КООПЕРАТИВНЫЙ ИНСТИТУТ (ФИЛИАЛ)**

КОМПЛЕКТ ЗАДАНИЙ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ АТТЕСТАЦИИ №1 (в форме лабораторной работы)

по дисциплине «Хранилища данных»

Лабораторная работа «Создание хранилища данных в аналитической»
платформе Deductor

1. Общие сведения о хранилищах данных

Современные информационные системы предприятий и организаций являются не только средством хранения и передачи информации, но и источником полезных знаний, которые могут быть использованы для повышения эффективности принятия управленческих решений. Такие знания могут быть представлены в виде зависимостей, закономерностей и структур в данных, интерпретация и осмысление которых специалистом позволяют ему делать выводы и суждения о процессах и явлениях в социальной и экономической сфере.

Обычно, с точки зрения ценности обнаруженных знаний, представляют наибольший интерес именно закономерности и зависимости скрытые в больших массивах данных, таких, как клиентские базы торговых сетей и телекоммуникационных компаний, истории продаж за длительный период времени, базы данных систем бронирования авиа и железнодорожных билетов и т.д. Знания, лежащие на поверхности являются тривиальными, общеизвестными и не позволяют добиться конкурентных преимуществ. Поэтому предприятия заинтересованы в организации процессов добычи из данных именно глубинных, «новых, нетривиальных, интерпретируемых и практически полезных знаний, которые могут быть использованы при принятии управленческих решений».

Современным направлением информационных технологий поиска знаний в массивах данных, содержащихся в информационных системах компаний, является Data Minig (раскопка, разработка данных) или интеллектуальный анализ данных (ИАД). В рамках ИАД используются различные модели и алгоритмы машинного обучения и прикладной

статистик для восстановления скрытых зависимостей и закономерностей. Очевидно, что подход к поиску знаний «от данных» предъявляет к последним ряд требований, нарушение которых может привести к получению недостоверных знаний и принятию ошибочных управленческих решений. Основными из них являются хронологичность, целостность, полнота, непротиворечивость, актуальность, отсутствие дубликатов и аномалий. Кроме этого, большую роль играет скорость доступа к данным для применения к ним различных методов аналитической обработки.

Обычные СУБД не гарантируют выполнения перечисленных требований и поэтому их использование для ИАД, может потребовать сложных процедур предобработки. Именно поэтому для поддержки процесса ИАД и поиска знаний получила развитие новая концепция консолидации данных – хранилища данных (Data Warehouse).

Хранилище данных (ХД) представляет собой предметно-ориентированный, интегрированный, неизменяемый и поддерживающий хронологию набор данных, организованный для целей поддержки принятия решений. ХД являются основным источником данных оперативно-аналитических подсистем (OLAP-систем) современных систем поддержки принятия решений (СППР) (рис. 1).

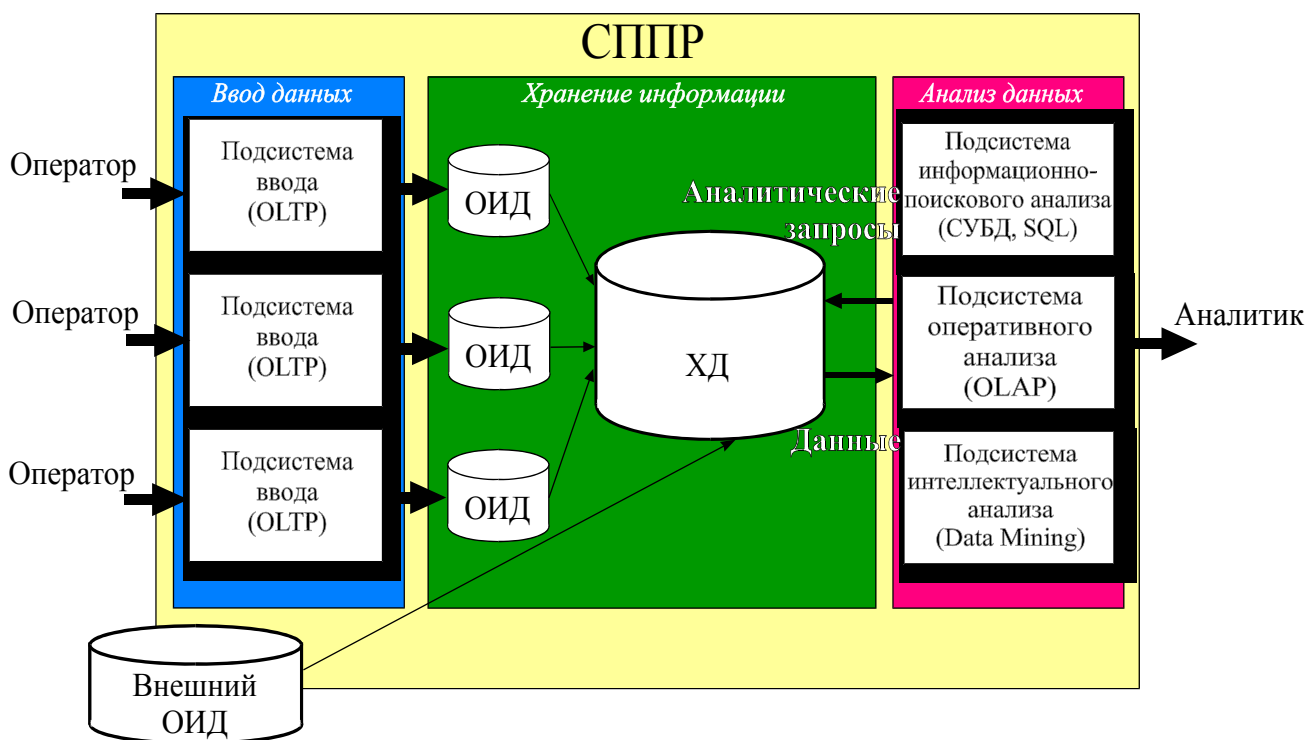


Рис. 1

Предметная ориентированность ХД означает, что данные в нем сгруппированы тематически, по определенным проблемным областям.

Интегрированность означает, что ХД должно поддерживать хранение данных различной природы и типов, отражающих различные аспекты деятельности предприятия, а не отдельные бизнес-функции.

Неизменяемость – внутри ХД данные не изменяются, могут только

пополняться новыми. Если нужны какие-либо изменения «перегружается» все хранилище целиком.

Хронологичность – поддерживается хронология данных внутри ХД

Таким образом, предоставляется возможность эффективнее и быстрее извлекать необходимую информацию для проверки выдвинутых гипотез.

Логическая структура данных в ХД использует реляционную модель. Используются две модели – «звезда» и «снежинка». В основе обеих моделей лежат следующие понятия:

- **измерение** – последовательность значений определенного признака. Например, измерение «Клиент» будет содержать последовательность наименований клиентов, измерение «Товар» - перечень товаров и т.д.;

- **атрибут** – свойство измерения, которое может быть полезно для его анализа. Например, для товара это может быть цвет, вес, форма и т.д.

- **факт** – значение, определяющее количественное выражения события или явления. Факты всегда связаны с определенным измерением. Например, для измерения «Товар» могут быть указаны следующие факты – цена, количество и сумма, для измерения «Клиент» - возраст, доход, стаж работы и т.д.

- **процесс** – совокупность измерений и фактов, связанных с определенным объектом, процессом или явлением. Процессов в хранилище может быть несколько и каждый из них образует отдельной «снежинкой» или «звездой».

«Звезда» включает в себя единственную таблицу фактов и множество таблиц измерений, соединенных с ней. Каждая таблица измерений должна находиться в отношении «один ко многим» с таблицей фактов.

Модель «снежинка» отличается от модели «звезда» тем, что содержит хотя бы одну таблицу измерений, ссылающуюся на таблицу, которая находится по отношению к ней в связи «один ко многим». Иными словами, измерения в «снежинке», могут иметь иерархически подчиненные измерения. Например, измерение «Регион» может быть связано с измерением «Город», т.е. для каждого региона будет содержаться список расположенных в нем городов.

Таблица фактов является основной таблицей процесса. Как правило, она содержит сведения об объектах или о событиях, совокупность которых будет в дальнейшем анализироваться. Обычно такая таблица содержит уникальный составной ключ, объединяющий первичные ключи таблиц измерений. Помимо этого таблица фактов содержит одно или несколько числовых полей, на основании которых в процессе выполнения аналитических запросов вычисляются агрегатные данные.

Агрегатными (агрегированными) называются значение, вычисленные на основе совокупности некоторого набора исходных значений. Например, суммы продаж за неделю (месяц, год) есть результат агрегирования ежедневных продаж с помощью функции суммирования. Агрегирование позволяет управлять детальностью представления явлений в анализируемых данных, что позволяет снизить количество обрабатываемых значений,

избавиться от случайных флуктуаций. Агрегатные данные могут формироваться и содержаться в ХД, что при большом количестве автоматически вычисляемых агрегатов может привести к информационному взрыву хранилища. Другой вариант – вычислять агрегаты «на лету», в процессе выполнения запроса к ХД, что может увеличить время ожидания.

Пример формирования представленных схем на основе реляционной БД содержащей сведения о производстве и сбыте изделий (рис. 2).



Рис. 2 - БД производство и сбыт изделий



Рис. 3 а) схема «снежинка», б). Схема «звезда»

Заполнение таблицы фактов происходит посредством обработки запросов к БД.

Классическое представление данных, разработанное основателем теории ХД – Биллом Инмоном, описывается следующими структурами (рис. 4):

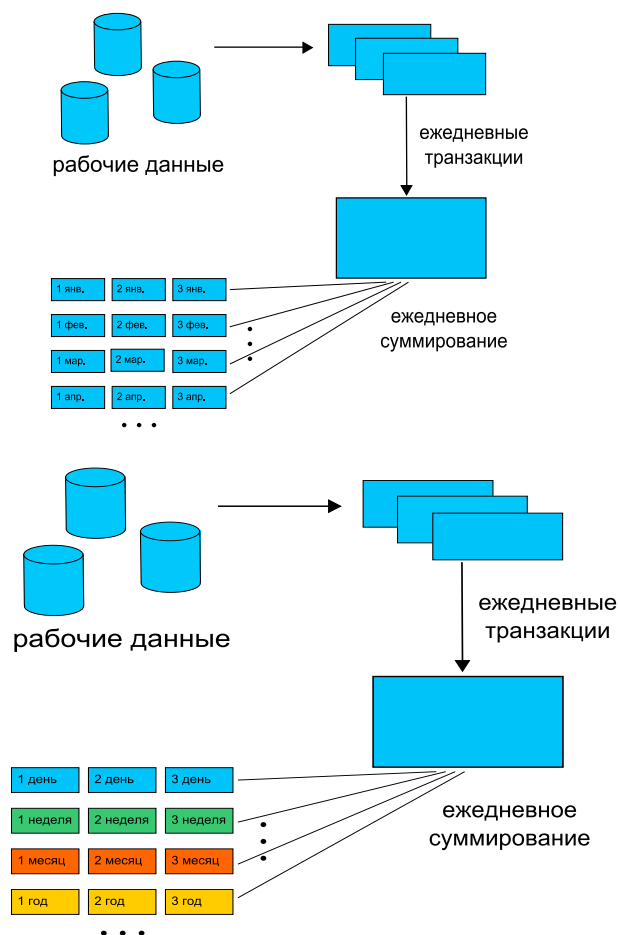


Рис. 4. Простейшие структуры данных в ХД: а) коммутативная, б) круговое суммирование

Коммутативная схема отражает ежедневные транзакции, которые перемещаются из операционной среды. После этого они «собираются» в запись ХД в соответствии с заданным регламентом (ежедневно, ежечасно и т.д.). При круговом суммировании данные также передаются из операционной но из различных структур. Для первых семи дней недели транзакции суммируются в семь ежедневных слотов. На восьмой день семидневные слоты складываются и размещаются в слот первой недели и т. д. В конце месяца недельные слоты складываются и размещаются в первый слот месяца. Затем недельный слот обнуляется и т.д.

Все загружаемые в ХД данные обязательно должны быть определены как измерение, атрибут либо факт (рис. 5)

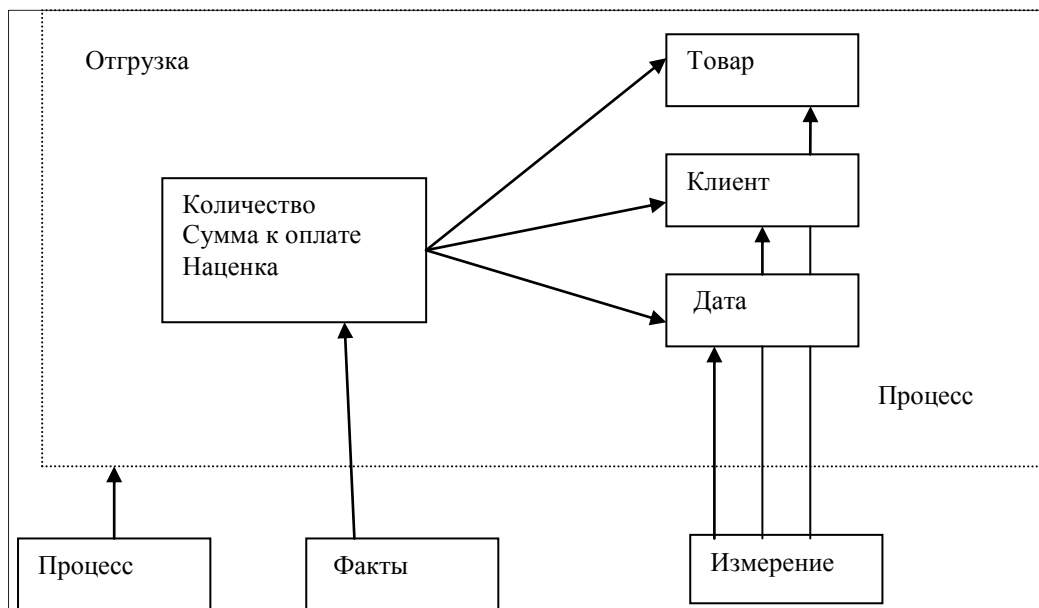


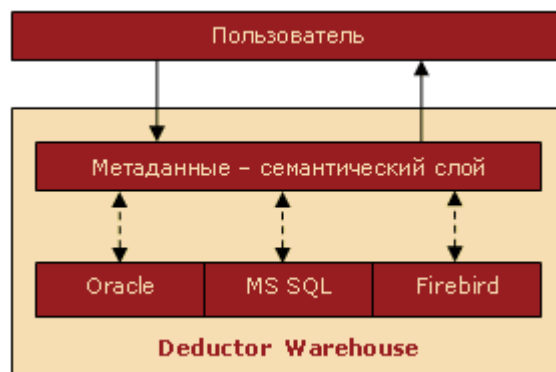
Рис. 5 Проектирование структуры хранилища

Кроме собственно данных, описывающих исследуемые процессы и явления, важнейшим элементом ХД являются *метаданные*. Метаданные (данные о данных) – это служебные данные, описывающие структуру хранилища, содержащие информацию о принадлежности данных к тому или иному типу (измерение, атрибут или факт).

С помощью метаданных формируется семантический слой, который обеспечивает визуальное представление хранилища и визуальные средства манипуляции с данными и метаданными.

2. Хранилище данных Deductor Warehouse

Deductor Warehouse – многомерное кросс-платформенное ХД, консолидирующее всю необходимую для анализа предметной области информацию. Использование единого хранилища позволяет обеспечить удобный доступ, высокую скорость обработки, непротиворечивость информации, централизованное хранение и автоматическую поддержку всего процесса анализа данных.



3. Разработка и создание ХД аптечной сети на основе Deductor Warehouse

3.1. Спроектировать структуру хранилища. Все данные представлены в 4-х таблицах, структуры и фрагменты приведены ниже (табл.1-табл.4)

Таблица 1. Группы товаров (фрагмент)

Код группы	Наименование группы
33	Иммуномодуляторы
34	Иммунодепрессанты
38	Секретолитики и стимуляторы моторной функции дыхательных путей
39	Противокашлевые средства
41	Стимуляторы дыхания
46	Психостимуляторы и ноотропы
48	Общетонизирующие средства и адаптогены
50	Местные анестетики
...	...

Таблица 2. Товары (фрагмент)

Код товара	Наименование товара	Код группы
35	Адреналина гидрохлорида раствор 0,1% р-р д/ин. 0,1 % амп. 1 мл [с нож.амп.] уп.контурн.яч. 5 Московс	159
354	Андрокур табл. 50 мг фл. 20 кор. 1 Schering-Plough Labo N.V.	178
477	Аспаркам табл. уп.контурн.б/яч. 10 Медисорб	108
485	Аспаркам табл. уп.контурн.б/яч. 10 Усолье-Сибирский ХФК	108
487	Аспаркам табл. уп.контурн.яч. 10 Ай Си Эн Лексредства	108
504	Аспирин табл. 500 мг бл. 10 кор. 2 Bayer С.С.	171
506	Аспирин табл. 100 мг бл. 10 кор. 2 Bayer С.С.	171
598	Ацетилсалициловая кислота табл. 0,5 г уп.контурн.б/яч. 10 Татхимфармпрепараты	171
682	АЦЦ 100 табл.шип. 100 мг туба 20 кор. 1 Hexal AG	38
686	АЦЦ 200 табл.шип. 200 мг туба 20 кор. 1 Hexal AG	38
690	АЦЦ лонг табл.шип. 600 мг туба 10 кор. 1 Hexal AG	38

Таблица 3. Отделы

Код отдела	Наименование отдела
1	Аптека №1
2	Аптека №3
3	Аптека №2

Таблица 4. Продажи (фрагмент)

Дата продажи	Товар.Код	Отдел.Код	Час покупки	Кол-во	Сумма
01.12.2008	477	1	11	1	4,45
01.12.2008	477	1	17	2	8,89
01.12.2008	485	1	11	1	4,56
01.12.2008	1813	1	15	1	54,17
01.12.2008	2206	1	14	1	11,58
01.12.2008	2290	1	16	1	179,38
01.12.2008	3381	1	15	1	98,08
01.12.2008	4138	1	15	1	78,74

01.12.2008	6100	1	10	2	116,73
01.12.2008	10208	1	10	1	6,6

Определим, какие данные являются измерениями, какие атрибутами, а какие фактами и что представляет собой процесс.

В табл.1 *Код группы* является измерением, *Наименование группы* - его атрибутом.

В табл.2 *Код товара* является измерением, *Наименование товара* - его атрибутом, а *Код группы* – ссылкой на одноименное измерение.

В табл.3 *Код отдела* является измерением, *Наименование отдела* - его атрибутом.

В табл.4 *Дата*, *Код отдела*, *Код товара* и *Час покупки* являются измерениями, *Количество* и *Сумма* – факты. То есть табл.4 представляет собой описание процесса продаж в трех аптеках.

3.2. Создание нового ХД

1) Запустить программу Deductor Studio Academic. Для создания нового пустого хранилища или подключения к существующему переходим на вкладку Подключения меню Вид, щелкаем правой кнопкой мыши и запускаем Мастер подключений (рис. 6).

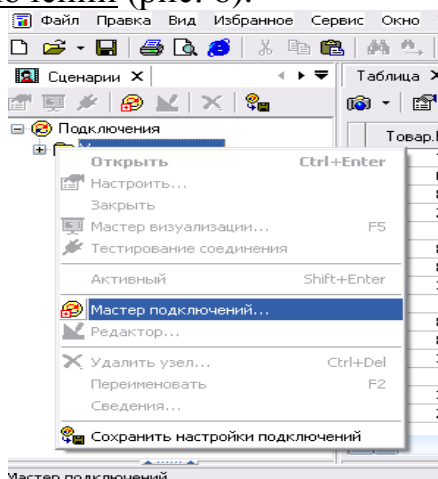


Рис. 6. Создание (подключение) хранилища данных

На первом шаге Мастера подключений следует выбрать тип источника (приемника) – Deductor Warehouse (рис. 7).

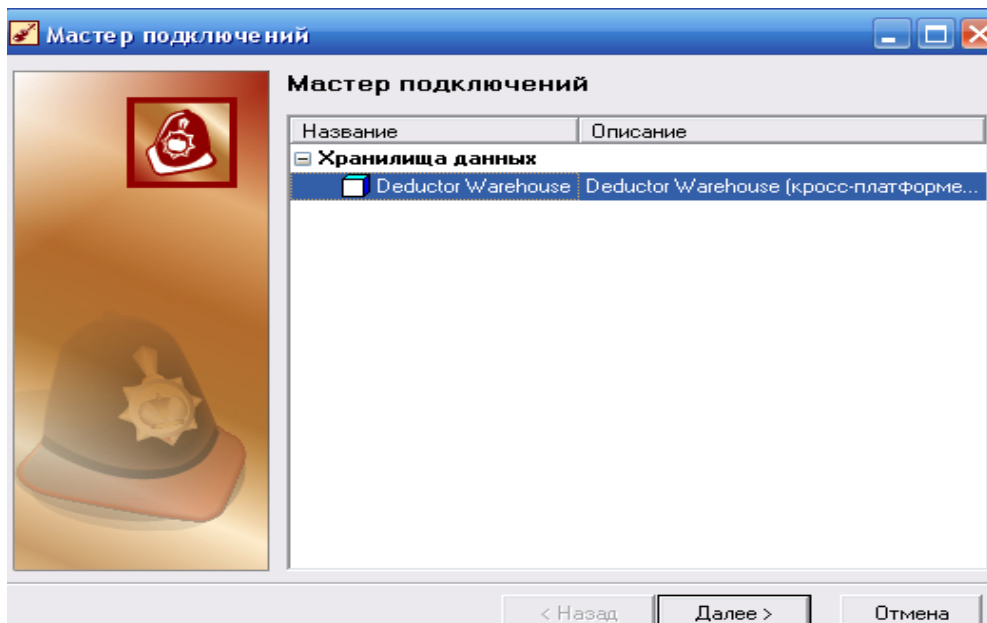


Рис. 7. Окно выбора типа подключения

На следующем шаге из списка базы данных выбираем Firebird. Задаем параметры базы данных, в которой будет создана физическая и логическая структура хранилища данных (рис. 8).

- база данных – *D:\farma.gdb* (или любой другой путь);
- логин *sysdba*, пароль *masterkey*;
- устанавливаем флажок **Сохранять пароль**.

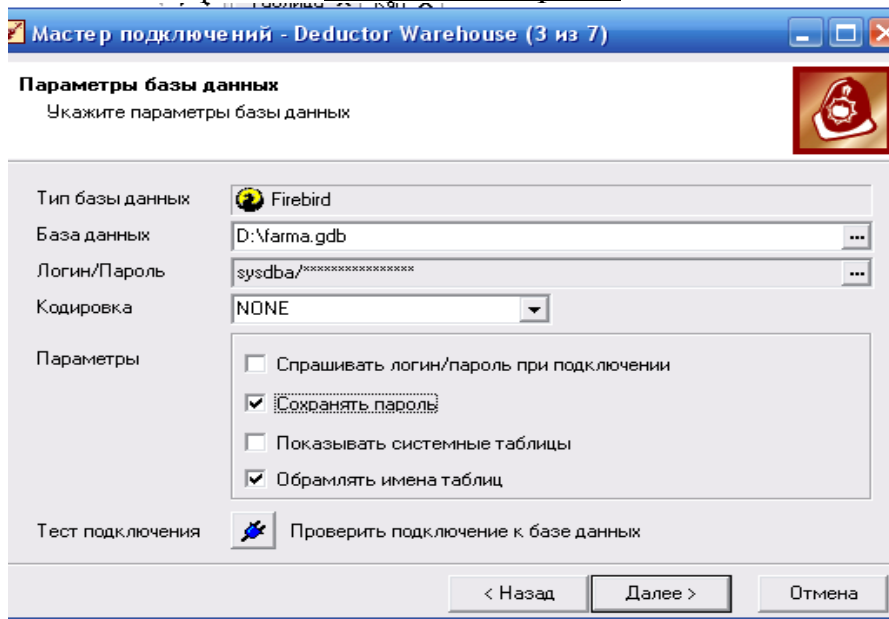



Рис. 8. Установка параметров базы данных

На следующей вкладке выбираем версию для работы с ХД Deductor Warehouse 6. Нажимаем кнопку Создать файл базы данных с необходимой структурой метаданных , и по указанному ранее пути будет создан файл *farma.gdb*. Это и есть пустое ХД.

Осталось выбрать визуализатор для подключения (здесь это Сведения и Метаданные) и задать имя, метку и описание нового хранилища (рис. 9).

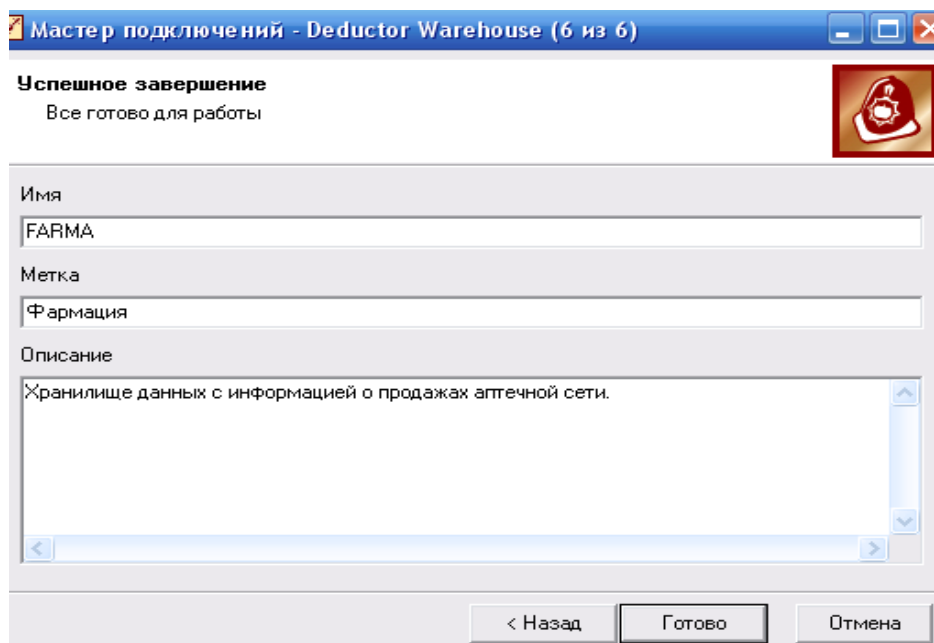


Рис. 9 Настройка семантики имен для узла подключения

Имя хранилища может быть введено только латинскими буквами. После нажатия кнопки Готово на дереве узлов подключений появится метка хранилища (рис.10)

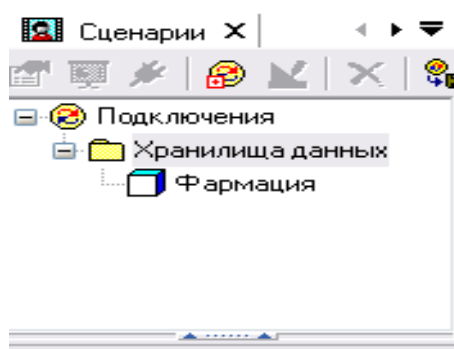







Рис. 10. Хранилище данных «Фармация»

Для проверки доступа к новому хранилищу данных воспользуемся кнопкой Тестирование соединения . Если спустя некоторое время появится сообщение «Тестирование соединения прошло успешно», то хранилище готово к работе. Сохраняем настройки подключений с помощью кнопки .

Если соединение по какой-либо причине установить не удалось, то будет выдано сообщение об ошибке. В этом случае нужно проверить параметры подключения хранилища данных и при необходимости внести в них изменения (кнопка «Настроить подключения» ). Таким образом создано пустое хранилище, в котором нет ни одного объекта (процесса, измерения, факта).

3.2. Перенос структуры данных в ХД. Для формирования структуры данных предназначен редактор метаданных, который вызывается нажатием кнопки «Открыть редактор метаданных»  на вкладке Подключения.

Для перехода в режим внесения изменений в структуру хранилища нажимаем кнопку «Разрешить редактировать» . Появится диалоговое окно

с предупреждением о том, что эта операция небезопасна. Вызвать контекстное меню для пункта *Измерения* и выбрать команду «Добавить». Создать первое измерение *Код группы* со следующими параметрами:

- имя – GR_ID;
- метка – Группа.Код;
- тип данных – целый.


Проделав аналогичные действия для создания всех остальных измерений, взяв параметры из табл.5, сохраним изменение структуры ХД с помощью кнопки Принять изменения .

Таблица 5. Параметры измерений

Измерение	Имя	Метка	Тип данных
Код группы	GR_ID	Группа.Код	Целый
Код товара	TV_ID	Товар.Код	Целый
Код отдела	PART_ID	Отдел.Код	Целый
Дата	S_DATE	Дата	Дата/Время
Час покупки	S_HOUR	Час	Целый

В результате структура метаданных хранилища будет содержать пять измерений (рис.11).

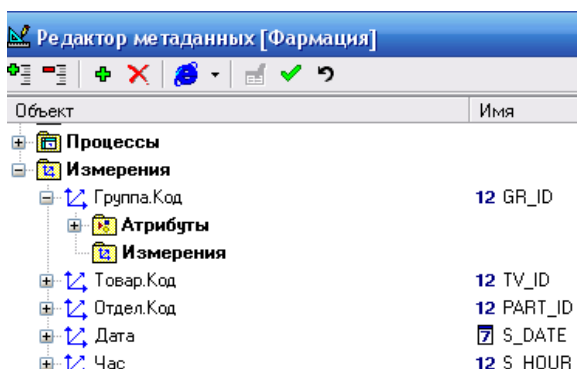



Рис. 11. Структура метаданных хранилища

К каждому измерению, кроме *Дата* и *Час*, добавим текстовый атрибут. Для измерения *Группа.Код* это будет *Группа.Наименование*, для измерения *Товар.Код* – *Товар.Наименование*, для измерения *Отдел.Код* – *Отдел.Наименование*.

Каждое измерение может ссылаться на другое измерение, реализуя тем самым иерархию измерений. В нашем случае измерение *Товар.Код* ссылается на *Группа.Код*. Эту ссылку и установим путем простого добавления (ссылка на измерение отображается значком ) , а имя ссылки зададим *GR_ID_1*. Результат работы иллюстрирует рис. 12.

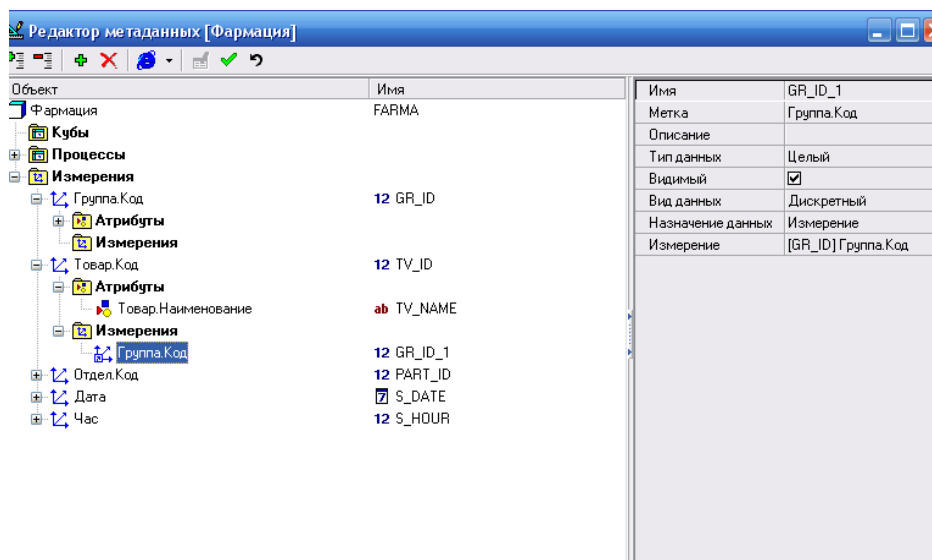



Рис. 12. Формирование ссылки на измерение

После того как все измерения и ссылки на измерения созданы, приступаем к формированию процесса («снежинки»). Назовем его *Продажи* и добавим в него ссылки на четыре существующих измерения: *Дата*, *Отдел.Код*, *Товар.Код*, *Час* (кнопка ). Кроме них, в процессе участвуют два факта: *Количество* и *Сумма*, причем первый - целочисленный, а второй - вещественный (рис. 13).

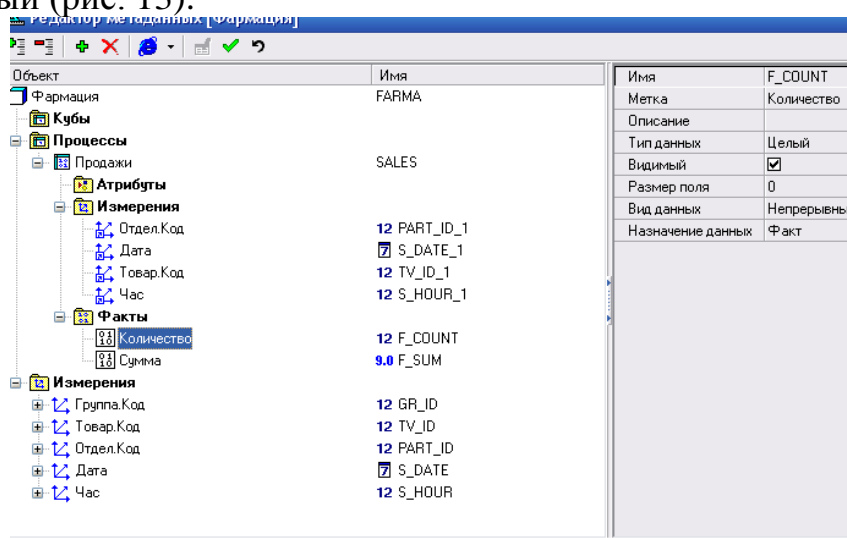


Рис. 13. Создание метаданных процесса

На этом проектирование структуры и метаданных ХД закончено.

3.2. Заполнение хранилища данных

После создания структуры ХД оно представляет собой пустой файл с настроенным семантическим слоем. В таком виде ХД готово к загрузке в него данных из внешних структурированных источников. Для этого необходимо создать соответствующий сценарий в Deductor Studio. Он должен выполнять следующие функции:

- импорт данных в Deductor Studio из источника данных (БД, учетной системы или файлов пользователя);
- загрузку данных в измерения и процессы хранилища Deductor

Warehouse.

Исходные данные для ХД содержатся в четыре текстовых файлах: *Группы товаров.txt*, *Товары.txt*, *Отделы.txt*, *Продажи.txt*. Поэтому сценарий загрузки должен быть настроен на использование этих файлов в качестве источников данных. Сценарий содержит следующие шаги:

- **импорт всех четырех текстовых файлов в Deductor**. Для этого переходим на вкладку Сценарии и из контекстного меню или нажатием клавиши F6 вызываем Мастер импорта, выбираем тип источника – текстовый файл и настраиваем параметры импорта. Последовательность создания узлов импорта должна быть такой, чтобы первыми следовали узлы импорта из файлов с таблицами измерений, и только в конце – таблица процесса *Продажи.txt*.

- **последовательно выполнить загрузку данных в измерения**. Поясним последовательность действия на примере первого измерения *Группа.Код*. Выделим в сценарии узел, соответствующий файлу из которого будет производиться загрузка, и вызовем Мастер экспорта и в списке типов получателей выбираем Deductor Warehouse (рис.14, 15)

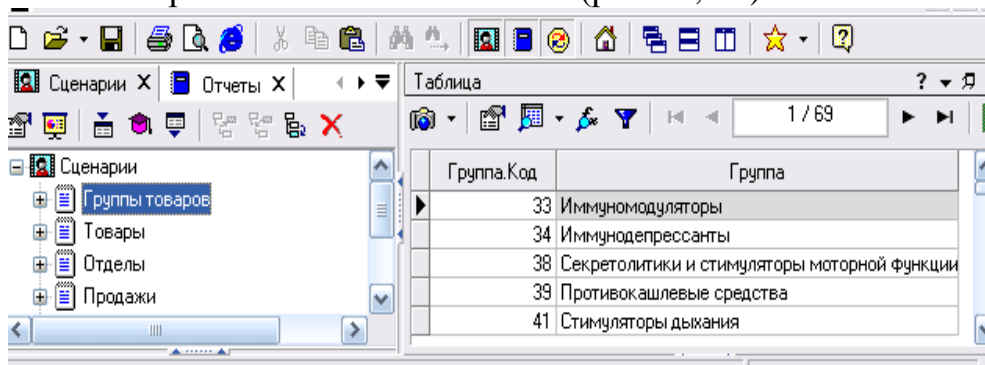


Рис.14. Сценарий в Deductor

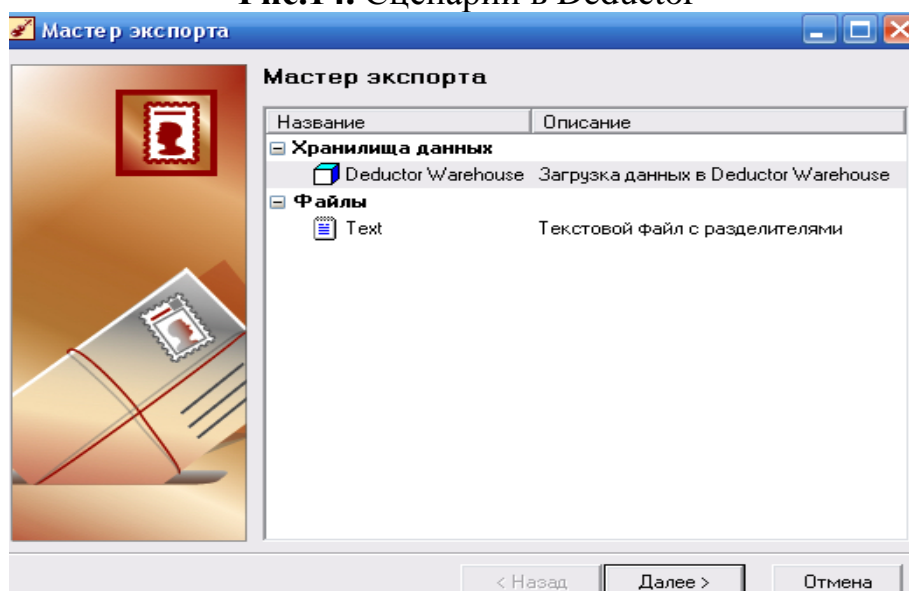


Рис.15. Экспорт в хранилище данных

На следующем шаге из списка доступных хранилищ выберем нужное под названием «Фармацея». На следующем шаге требуется указать, в какое именно измерение будет загружаться информация. Это *Группа.Код* (рис.16).

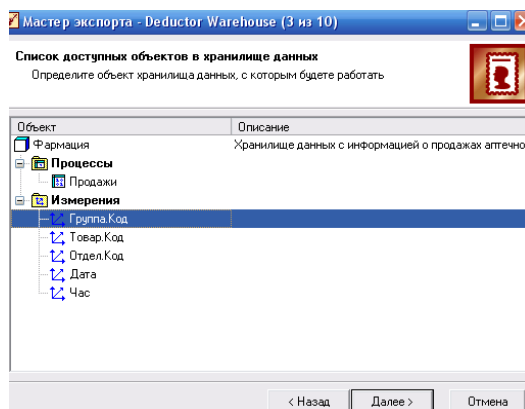


Рис. 16. Выбор измерения для загрузки данных.

Замечание. При установке соответствия, необходимо, чтобы тип данных в поле источника данных из которого будет производиться загрузка, и типа данных, объявленный для измерения, совпадали. Выбрать тип данных для поля в источнике можно в Мастере импорта.

Установить соответствие элементов объекта в хранилище данных с полями входного источника данных (то есть таблицы *Группы товаров.txt*). В случае, когда имена полей и (или) метки в семантическом слое хранилища данных совпадают, делать ничего не нужно (рис.17).

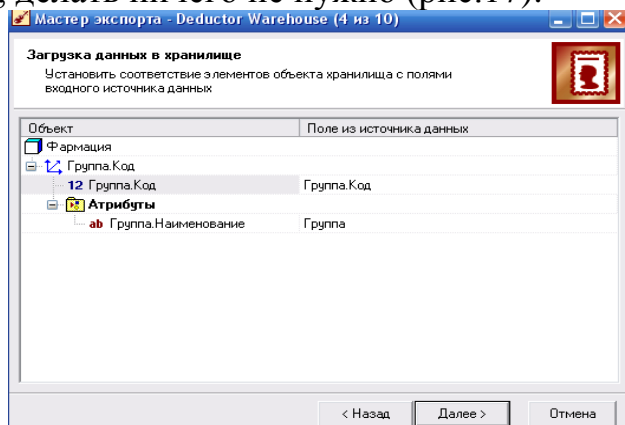


Рис.17. Настройка соответствия полей и загружаемых

Нажатие кнопки Пуск на следующем шаге загрузит в измерение данные. При этом старые данные, если они были, обновятся. Прделав аналогичные действия еще для двух измерений – *Отдел.Код*, *Товар.Код*, получим следующий сценарий (рис.18).

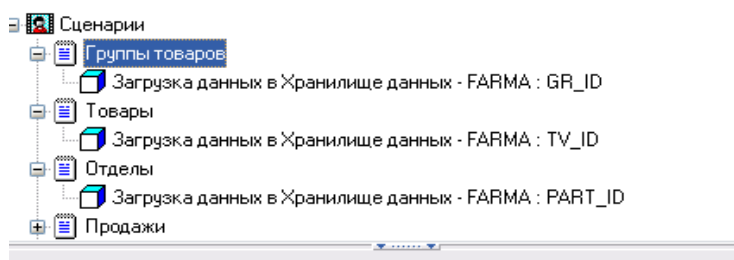


Рис. 18. Сценарий загрузки данных в ХД

Загрузка измерений на этом заканчивается, несмотря на то, что остались еще два измерения *Дата* и *Час*. Но они не имеют атрибутов и не

участвуют в иерархии, поэтому их значения можно загрузить на этапе экспорта в процесс.

Загрузим данные в процесс *Продажи*. В отличие от загрузки измерений, в Мастере экспорта появляются два специфических шага.

На одном из них нужно задать параметры контроля непротиворечивости данных в хранилище – указать измерения, по которым следует удалять данные из хранилища (рис.19).

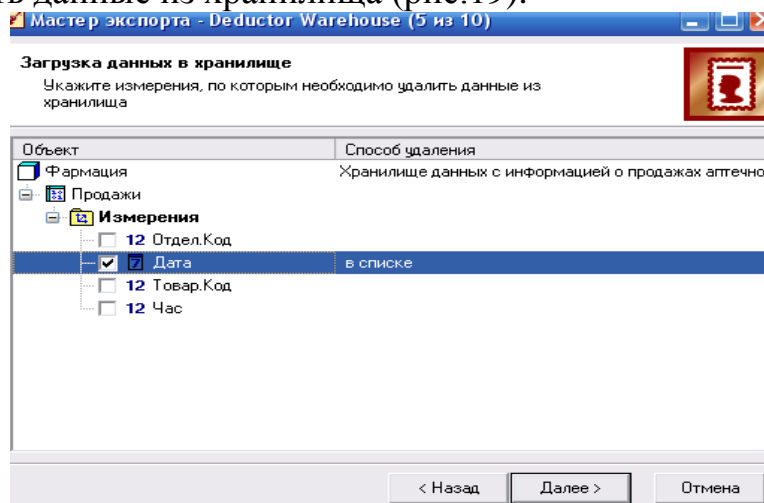


Рис.19. Параметры для контроля непротиворечивости информации
На последней странице Мастера экспорта оставить настройки по умолчанию (рис.20).

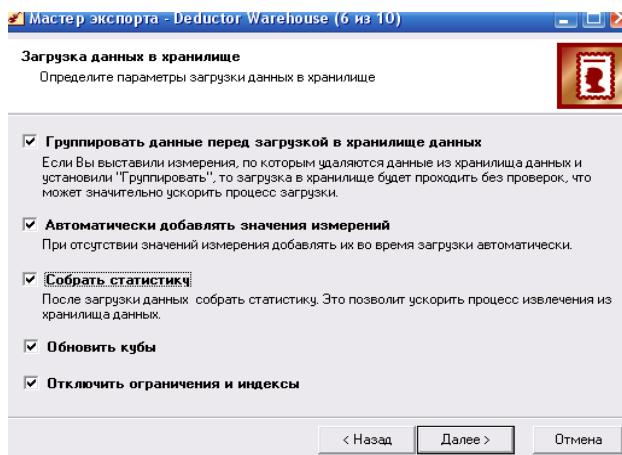


Рис. 20. Дополнительные параметры загрузки в процесс
В Мастере экспорта можно задать вариант агрегации фактов (рис.21).

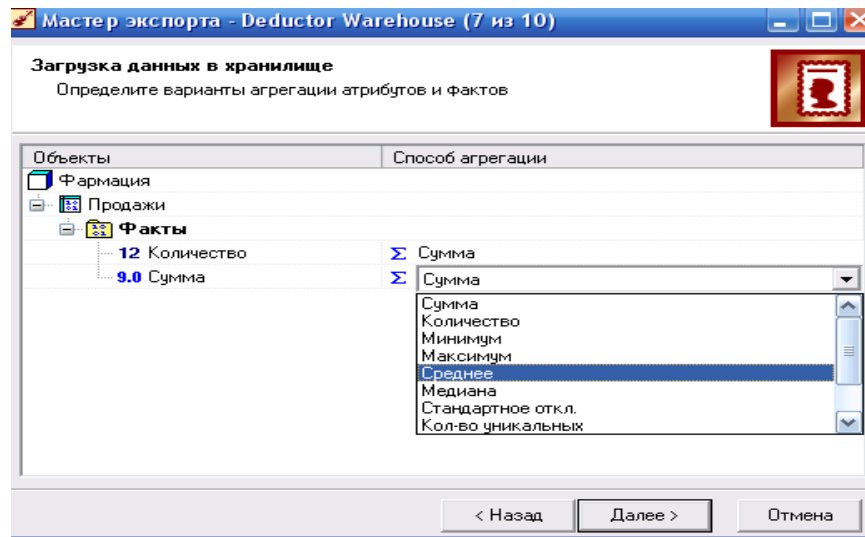


Рис. 21. Дополнительные параметры загрузки в процесс
Окончательный вариант загрузки приведен на рис.22.

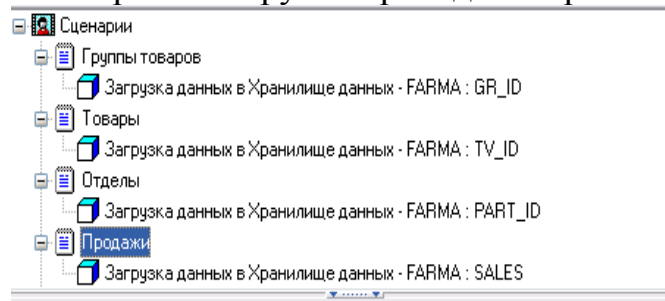


Рис.18. Окончательный сценарий загрузки

Таким образом, в результате описанных действий будет:

- создано и наполнено хранилище данных;
- написан сценарий загрузки информации из источников в ХД;
- продуман контроль непротиворечивости данных в ХД.

Контрольные вопросы

1. Что такое хранилище данных, какие требования к нему предъявляются? В чем отличие ХД от обычной базы данных.
2. Каковы причины появления и развития концепции ХД.
3. Дайте понятие метаданных и семантического слоя в ХД.
4. Перечислите основные элементы логической структуры ХД, поясните их смысл, приведите примеры.
5. Дайте понятие измерения как элемента структуры данных ХД. Приведите примеры измерений. В чем отличие измерений от фактов.
6. Дайте понятие процесса в ХД. Приведите пример.
7. Логическая структура данных на основе схемы «звезда».
8. Логическая структура данных на основе схемы «снежинка».
9. Что такое агрегирование (агрегация) данных? Приведите пример.
10. Опишите последовательность действий при загрузке данных в ХД Deductor Warehouse.

Критерии оценки:

- оценка «отлично» выставляется студенту, если он показал высокий

уровень сформированности компетенции, верно и в полном объеме показал глубокие исчерпывающие знания всего программного материала по дисциплине, понимание сущности проблемы; дал логически последовательные, содержательные, полные правильные и конкретные ответы на все вопросы.

- оценка «хорошо» выставляется студенту, если он показал средний уровень сформированности компетенции, с незначительными замечаниями показал твердые и достаточно полные знания всего программного материала по дисциплине, правильное понимание сущности проблемы. Дал последовательные, правильные, конкретные ответы на поставленные вопросы при свободном устранении замечаний по отдельным вопросам.

- оценка «удовлетворительно» выставляется студенту, если он показал низкий уровень сформированности компетенции, на базовом уровне с ошибками показал знание и понимание сущности проблемы; дал правильные и конкретные, без грубых ошибок ответы на поставленные вопросы.

- оценка «неудовлетворительно» выставляется студенту, если он показал недостаточный уровень сформированности компетенции, дал неправильный ответ хотя бы на один из основных вопросов, сделал большое количество ошибок в ответе, не понимает сущности излагаемых вопросов; дает неуверенные и неточные ответы на дополнительные вопросы.