

Документ подписан простой электронной подписью  
Информация о владельце: МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФИО: Силин Яков Петрович  
Должность: Ректор  
Дата подписания: 08.06.2026 14:30:04  
Уникальный программный ключ:  
24f866be2aca16484036a8cbb3c509a9531e6034

Одобрена  
на заседании кафедры

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФГБОУ ВО «Уральский государственный экономический университет»

02.12.2025 г.  
протокол № 3  
Зав. кафедрой Назаров Д.М.

Утверждена  
Советом по учебно-методическим  
вопросам и качеству образования

16 декабря 2025 г.  
протокол № 4  
Председатель Карх Д.А.



### РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Наименование дисциплины	Интеллектуальные системы бизнес-анализа
Направление подготовки	38.03.05 Бизнес-информатика
Профиль	Цифровой бизнес
Форма обучения	очно-заочная
Год набора	2026

Разработана:  
Доцент, к.ф.-м.н  
Тюлюкин В.А.

Екатеринбург  
2025 г.

## СОДЕРЖАНИЕ

<b>ВВЕДЕНИЕ</b>	<b>3</b>
<b>1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ</b>	<b>3</b>
<b>2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП</b>	<b>3</b>
<b>3. ОБЪЕМ ДИСЦИПЛИНЫ</b>	<b>3</b>
<b>4. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ОПОП</b>	<b>3</b>
<b>5. ТЕМАТИЧЕСКИЙ ПЛАН</b>	<b>10</b>
<b>6. ФОРМЫ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ШКАЛЫ ОЦЕНИВАНИЯ</b>	<b>11</b>
<b>7. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ</b>	<b>13</b>
<b>8. ОСОБЕННОСТИ ОРГАНИЗАЦИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ ДЛЯ ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ</b>	<b>15</b>
<b>9. ПЕРЕЧЕНЬ ОСНОВНОЙ И ДОПОЛНИТЕЛЬНОЙ УЧЕБНОЙ ЛИТЕРАТУРЫ, НЕОБХОДИМОЙ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ</b>	<b>15</b>
<b>10. ПЕРЕЧЕНЬ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, ВКЛЮЧАЯ ПЕРЕЧЕНЬ ЛИЦЕНЗИОННОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНФОРМАЦИОННЫХ СПРАВОЧНЫХ СИСТЕМ, ОНЛАЙН КУРСОВ, ИСПОЛЬЗУЕМЫХ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ</b>	<b>16</b>
<b>11. ОПИСАНИЕ МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ</b>	<b>17</b>

## ВВЕДЕНИЕ

Рабочая программа дисциплины является частью основной профессиональной образовательной программы высшего образования - программы бакалавриата, разработанной в соответствии с ФГОС ВО

ФГОС ВО	Федеральный государственный образовательный стандарт высшего образования - бакалавриат по направлению подготовки 38.03.05 Бизнес-информатика (приказ Минобрнауки России от 29.07.2020 г. № 838)
---------	---

### 1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель освоения дисциплины – формирование навыков создания интеллектуальных систем бизнес-анализа средствами Data Mining.

### 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина относится к части, формируемой участниками образовательных отношений.

### 3. ОБЪЕМ ДИСЦИПЛИНЫ

Промежуточная аттестация	Часов					3.е.
	Всего за семестр	Контактная работа (по уч.зан.)			Самостоятельная работа в том числе подготовка контрольных и курсовых	
		Всего	Лекции	Лабораторные		
Семестр 7						
Зачет	144	16	8	8	124	4
Семестр 8						
Экзамен, Контрольная работа	180	12	0	12	159	5
	324	28	8	20	283	9

### 4. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ОПОП

В результате освоения ОПОП у выпускника должны быть сформированы компетенции, установленные в соответствии ФГОС ВО.

Шифр и наименование компетенции	Индикаторы достижения компетенций
аналитический	

<p>ПК-1 Выявление, формирование и согласование требований к результатам аналитических работ с применением технологий больших данных</p>	<p>ИД-1.ПК-1 Знать:</p> <p>Регламенты организации по оформлению требований к результатам аналитических исследований с использованием технологий больших данных</p> <p>Технологии межличностной и групповой коммуникации в деловом взаимодействии, основы конфликтологии</p> <p>Технологии подготовки и проведения презентаций</p> <p>Предметная область анализа больших данных в соответствии с требованиями заказчика</p> <p>Возможности имеющейся у исполнителя методологической и технологической инфраструктуры анализа больших данных</p> <p>Современный опыт использования анализа больших данных</p> <p>Теоретическая и прикладная информатика</p> <p>Теоретические и прикладные основы анализа данных</p> <p>Основы бизнес-интеллекта, типы систем бизнес-интеллекта</p> <p>Теория принятия решений</p> <p>Математическое моделирование</p> <p>Типы анализа больших данных, виды аналитики</p> <p>Теория вероятностей и математическая статистика</p> <p>Современные методы и инструментальные средства анализа больших данных</p> <p>Стандарты проведения анализа данных</p> <p>Методы оценки временных и стоимостных характеристик технологий больших данных</p> <p>Источники информации, в том числе информации, необходимой для обеспечения деятельности в предметной области заказчика исследования</p> <p>Современная технологическая инфраструктура высокопроизводительных и распределенных вычислений</p> <p>Методы интерпретации и визуализации больших данных</p> <p>Правила деловой переписки</p>
	<p>ИД-2.ПК-1 Уметь:</p> <p>Проводить презентации при консультировании заказчика, согласовании и утверждении требований к результатам аналитических работ с использованием технологий больших данных</p> <p>Подготавливать документы, регламентирующие требования к результатам аналитического исследования с использованием технологий больших данных в соответствии с существующими регламентами организации</p> <p>Использовать имеющуюся у исполнителя методологическую и технологическую инфраструктуру анализа больших данных для выполнения аналитических работ</p> <p>Проводить сравнительный анализ методов и инструментальных средств анализа больших данных</p> <p>Проводить анализ больших данных в соответствии с утвержденными требованиями к результатам аналитического исследования</p>

<p>ПК-1 Выявление, формирование и согласование требований к результатам аналитических работ с применением технологий больших данных</p>	<p>ИД-3.ПК-1 Уметь:</p> <p>Проводить презентации при консультировании заказчика, согласовании и утверждении требований к результатам аналитических работ с использованием технологий больших данных</p> <p>Подготавливать документы, регламентирующие требования к результатам аналитического исследования с использованием технологий больших данных в соответствии с существующими регламентами организации</p> <p>Использовать имеющуюся у исполнителя методологическую и технологическую инфраструктуру анализа больших данных для выполнения аналитических работ</p> <p>Проводить сравнительный анализ методов и инструментальных средств анализа больших данных</p> <p>Проводить анализ больших данных в соответствии с утвержденными требованиями к результатам аналитического исследования</p> <hr/> <p>ИД-4.ПК-1 Иметь практический опыт:</p> <p>Выявление требований заказчика к результатам анализа, определение возможностей применения анализа больших данных в предметной области и конкретных задачах заказчика</p> <p>Консультирование заказчика по возможностям имеющейся методологической и технологической инфраструктуры анализа больших данных и результатам применения технологий больших данных к аналогичным задачам</p> <p>Согласование с заказчиком и утверждение требований к результатам аналитического исследования</p>
<p>ПК-2 Планирование и организация аналитических работ с использованием технологий больших данных</p>	<p>ИД-1.ПК-2 Знать:</p> <p>Возможности имеющейся у исполнителя методологической и технологической инфраструктуры анализа больших данных</p> <p>Возможности использования свободно распространяемого программного обеспечения для анализа больших данных</p> <p>Предметная область анализа больших данных в соответствии с требованиями заказчика</p> <p>Основы планирования аналитических работ</p> <p>Стандарты проведения анализа данных</p> <p>Методы и инструментальные средства управления аналитическими проектами по исследованию больших данных</p> <p>Содержание и последовательность выполнения этапов аналитического проекта по исследованию больших данных</p> <p>Содержание этапов жизненного цикла больших данных</p> <p>Типы анализа больших данных, виды аналитики</p> <p>Теоретические и прикладные основы анализа больших данных</p> <p>Современные методы и инструментальные средства анализа больших данных</p> <p>Теория вероятностей и математическая статистика</p> <p>Источники информации, в том числе информации, необходимой для обеспечения деятельности в предметной области заказчика исследования</p> <p>Методы интерпретации и визуализации анализа больших данных</p> <p>Технологии межличностной и групповой коммуникации в деловом взаимодействии, основы конфликтологии</p> <p>Технологии подготовки и проведения презентаций</p> <p>Правила деловой переписки</p>

<p>ПК-2 Планирование и организация аналитических работ с использованием технологий больших данных</p>	<p>ИД-2.ПК-2 Уметь:</p> <p>Проводить переговоры при определении содержания аналитических работ с использованием технологий больших данных</p> <p>Представлять содержание и результаты работ по анализу больших данных</p> <p>Вести протоколы мероприятий по анализу больших данных</p> <p>Планировать аналитические работы с использованием технологий больших данных</p> <p>Проводить аналитические работы с использованием технологий больших данных</p> <p>Проводить анализ больших данных</p> <p>Осуществлять интеграцию и преобразование данных в ходе работ по анализу больших данных</p>
	<p>ИД-3.ПК-2 Иметь практический опыт:</p> <p>Разработка, обсуждение и утверждение содержания аналитических работ с использованием технологий больших данных</p> <p>Определение состава группы для проведения анализа больших данных</p> <p>Определение необходимых ресурсов для проведения аналитических работ</p> <p>Разработка, обсуждение и утверждение плана аналитических работ</p> <p>Распределение ролей и состава аналитических работ между участниками группы для анализа больших данных</p>

<p>ПК-3 Подготовка данных для проведения аналитических работ по исследованию больших данных</p>	<p>ИД-1.ПК-3 Знать:</p> <p>Возможности имеющейся у исполнителя методологической и технологической инфраструктуры анализа больших данных</p> <p>Предметная область анализа</p> <p>Теоретические и прикладные основы анализа больших данных</p> <p>Современные методы и инструментальные средства анализа больших данных</p> <p>Современный опыт использования анализа больших данных</p> <p>Типы больших данных: метаданные, полуструктурированные, структурированные, неструктурированные</p> <p>Виды источников данных: созданные человеком, созданные машинами</p> <p>Источники информации, в том числе информации, необходимой для обеспечения деятельности в предметной области заказчика исследования</p> <p>Методы извлечения информации и знаний из гетерогенных, мультиструктурированных, неструктурированных источников, в том числе при потоковой обработке</p> <p>Российские и международные стандарты информационной безопасности</p> <p>Современная технологическая инфраструктура высокопроизводительных и распределенных вычислений</p> <p>Режимы получения и обработки данных, поддержка режима реального времени</p> <p>Технологии хранения и обработки больших данных в организации: базы данных, хранилища данных, распределенная и параллельная обработка данных, вычисления в оперативной памяти</p> <p>Облачные технологии, облачные сервисы</p> <p>Методы оценки временных и стоимостных характеристик технологий больших данных</p> <p>Технологии межличностной и групповой коммуникации в деловом взаимодействии, основы конфликтологии</p> <p>Правила деловой переписки</p>
	<p>ИД-2.ПК-3 Уметь:</p> <p>Определять требования к поставщикам данных из гетерогенных источников</p> <p>Осуществлять взаимодействие с внутренними и внешними поставщиками данных из гетерогенных источников</p> <p>Разрабатывать и оценивать модели больших данных</p> <p>Использовать инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников, в том числе в режиме реального времени</p> <p>Производить очистку данных для проведения аналитических работ</p> <p>Проводить интеграцию и преобразование больших объемов данных</p> <p>Оценивать соответствие наборов данных задачам анализа больших данных</p> <p>Оценивать стоимость данных для проведения аналитических работ</p>

<p>ПК-3 Подготовка данных для проведения аналитических работ по исследованию больших данных</p>	<p>ИД-3.ПК-3 Иметь практический опыт: Определение источников больших данных для анализа, идентификация внешних и внутренних источников данных для проведения аналитических работ Получение и фильтрация больших объемов данных из гетерогенных источников Извлечение, проверка и очистка больших объемов данных из гетерогенных источников Агрегация и разработка представления больших объемов данных из гетерогенных источников Оценка соответствия набора данных предметной области и задачам аналитических работ</p>
---	--

<p>ПК-4 Проведение аналитического исследования с применением технологий больших данных в соответствии с требованиями заказчика</p>	<p>ИД-1.ПК-4 Знать:  Содержание и последовательность выполнения этапов аналитического проекта  Основы управления аналитическими работами  Основы управления малыми аналитическими группами  Возможности имеющейся у исполнителя методологической и технологической инфраструктуры анализа больших данных  Предметная область анализа  Теория принятия решений  Математическое моделирование  Теоретические и прикладные основы анализа больших данных  Современный опыт использования анализа больших данных  Технологии анализа данных: статистический анализ, семантический анализ, анализ изображений, машинное обучение, методы сравнения средних, частотный анализ, анализ соответствий, кластерный анализ, дискриминантный анализ, факторный анализ, деревья классификации, многомерное шкалирование, моделирование структурными уравнениями, методы анализа выживаемости, временные ряды, планирование экспериментов, карты контроля качества  Нейронные сети: полносвязные, свёрточные и рекуррентные нейронные сети, методы обучения нейронных сетей, нейросетевые методы понижения размерности  Статистические модели  Статистический анализ: метод многовариантного тестирования, корреляционный анализ, регрессионный анализ  Статистические методы: параметрические, непараметрические, управляемые, неуправляемые, полуправляемые, кластеризация  Семантический анализ: обработка естественного языка, сентиментный анализ, анализ текста  Алгоритмы машинного обучения: обучение с учителем, обучение без учителя, полуправляемое обучение, обучение с подкреплением  Машинное обучение: классификация, кластеризация, обнаружение выбросов, фильтрация  Методы и модели классификации: логистическая регрессия, деревья решений, предредукция, постредукция, модели, основанные на правилах, вероятностные классификаторы, усиление энтропии информации  Фильтрация шумовых выбросов, виды шумовых выбросов: глобальный, контекстуальный, коллективный  Анализ изображений, анализ сетей, анализ пространственных данных, анализ временных рядов  Методы идентификации шаблонов  Методы оценки моделей: оценка качества построенной модели по тестовой выборке и анализ обобщающих способностей алгоритма  Распределенный анализ данных  Анализ данных в реальном времени  Правила деловой переписки  Методы разработки отчетной аналитической документации</p>
--	---

ПК-4 Проведение аналитического исследования с применением технологий больших данных в соответствии с требованиями заказчика	<p>ИД-2.ПК-4 Уметь:</p> <p>Планировать аналитические работы с использованием технологий больших данных</p> <p>Проводить аналитические работы с использованием технологий больших данных, как индивидуально, так и, осуществляя руководство малыми аналитическими группами</p> <p>Использовать имеющуюся у исполнителя методологическую и технологическую инфраструктуру анализа больших данных для выполнения аналитических работ</p> <p>Проводить сравнительный анализ методов и инструментальных средств анализа больших данных</p> <p>Разрабатывать и оценивать модели больших данных</p> <p>Программировать на языках высокого уровня, ориентированных на работу с большими данными: для статистической обработки данных и работы с графикой, для работы с разрозненными фрагментами данных в больших массивах, для работы с базами структурированных и неструктурированных данных</p> <p>Адаптировать и развертывать модели в предметной среде</p> <p>Решать задачи классификации, кластеризации, регрессии, прогнозирования, снижения размерности и ранжирования данных</p> <p>Решать проблемы переобучения и недообучения алгоритма</p> <p>Формировать предложения по использованию результатов анализа</p> <p>Оформлять результаты аналитического исследования для представления заказчику</p> <p>Разъяснять заказчику результаты аналитической работы</p> <p>Осуществлять поиск информации о новых и перспективных методах анализа больших данных, выполнять сравнительный анализ методов</p>
	<p>ИД-3.ПК-4 Иметь практический опыт:</p> <p>Выбор методов и инструментальных средств анализа больших данных для проведения аналитических работ</p> <p>Разработка, проверка, оценка используемых моделей больших данных</p> <p>Адаптация и развертывание моделей больших данных в предметной среде</p> <p>Выбор средств представления результатов аналитики больших данных</p> <p>Подготовка отчета по результатам аналитических работ с использованием технологий больших данных</p> <p>Консультирование заказчика по результатам аналитических работ с применением технологий больших данных</p> <p>Мониторинг эффективности работы аналитики больших данных</p> <p>Формирование предложений по использованию результатов анализа больших данных: рассылка, создание приложений, оптимизация процессов</p> <p>Формирование предложений по развитию существующей методологической и технологической инфраструктуры анализа больших данных на основе выполненных работ</p>

## 5. ТЕМАТИЧЕСКИЙ ПЛАН

Тема	Часов						
	Наименование темы	Всего часов	Контактная работа (по уч.зан.)			Самост. работа	Контроль самостоятельной работы
			Лекции	Лабораторные	Практические занятия		
Семестр 7		24					

Тема 1.	Введение в интеллектуальные системы бизнес-анализа (ПК-1, ПК-2)	24	4			20	
Семестр 7		116					
Тема 2.	Извлечение знаний из структурированных данных. (ПК-3)	116	4	8		104	
Семестр 8		54					
Тема 3.	Технологии KDD и Data Mining (ПК-4)	54		4		50	
Семестр 8		117					
Тема 4.	Самообучающиеся системы на основе нейронных сетей. Подготовка данных для машинного обучения (ПК-4)	117		8		109	

## 6. ФОРМЫ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ШКАЛЫ ОЦЕНИВАНИЯ

Раздел/Тема	Вид оценочного средства	Описание оценочного средства	Критерии оценивания
Текущий контроль (Приложение 4)			
Введение в интеллектуальные системы бизнес-анализа. Извлечение знаний из структурированных данных.	Контрольная работа №1 (Приложение 4)	Контрольная работа состоит из практического задания.	100
Технологии KDD и Data Mining	Контрольная работа №2 (Приложение 4)	Контрольная работа состоит из практического задания	100
Самообучающиеся системы на основе нейронных сетей. Подготовка данных для машинного обучения	Контрольная работа №3 (Приложение 4)	Контрольная работа состоит из практического задания	100
Извлечение знаний из структурированных данных.	Контрольная работа №4 (Приложение 4)	Контрольная работа состоит из практического задания	100
Промежуточная аттестация(Приложение 5)			
8 семестр (Эк)	Экзаменационные билеты	20 билетов Каждый билет состоит из 2 теоретических вопросов и задачи	менее 30 - 2 31<...<60 - 3 61<...<85 - 4 86<...<100 - 5

7 семестр (За)	Тест	Тест состоит из 20 вопросов	менее 30 - 2 31<...<60 - 3 61<...<85 - 4 86<...<100 -5
----------------	------	-----------------------------	---

### ОПИСАНИЕ ШКАЛ ОЦЕНИВАНИЯ

Показатель оценки освоения ОПОП формируется на основе объединения текущего контроля и промежуточной аттестации обучающегося.

Показатель рейтинга по каждой дисциплине выражается в процентах, который показывает уровень подготовки студента.

Текущий контроль.Используется 100-балльная система оценивания. Оценка работы студента в течении семестра осуществляется преподавателем в соответствии с разработанной им системой оценки учебных достижений в процессе обучения по данной дисциплине.

В рабочих программах дисциплин и практик закреплены виды текущего контроля, планируемые результаты контрольных мероприятий и критерии оценки учебный достижений.

В течение семестра преподавателем проводится не менее 3-х контрольных мероприятий, по оценке деятельности студента. Если посещения занятий по дисциплине включены в рейтинг, то данный показатель составляет не более 20% от максимального количества баллов по дисциплине.

Промежуточная аттестация. Используется 5-балльная система оценивания. Оценка работы студента по окончанию дисциплины (части дисциплины) осуществляется преподавателем в соответствии с разработанной им системой оценки достижений студента в процессе обучения по данной дисциплине. Промежуточная аттестация также проводится по окончанию формирования компетенций.

Порядок перевода рейтинга, предусмотренных системой оценивания, по дисциплине, в пятибалльную систему.

Высокий уровень – 100% - 70% - отлично, хорошо.

Средний уровень – 69% - 50% - удовлетворительно.

Показатель оценки	По 5-балльной системе	Характеристика показателя
100% - 85%	отлично	обладают теоретическими знаниями в полном объеме, понимают, самостоятельно умеют применять, исследовать, идентифицировать, анализировать, систематизировать, распределять по категориям, рассчитать показатели, классифицировать, разрабатывать модели, алгоритмизировать, управлять, организовать, планировать процессы исследования, осуществлять оценку результатов на высоком уровне
84% - 70%	хорошо	обладают теоретическими знаниями в полном объеме, понимают, самостоятельно умеют применять, исследовать, идентифицировать, анализировать, систематизировать, распределять по категориям, рассчитать показатели, классифицировать, разрабатывать модели, алгоритмизировать, управлять, организовать, планировать процессы исследования, осуществлять оценку результатов.  Могут быть допущены недочеты, исправленные студентом самостоятельно в процессе работы (ответа и т.д.)
69% - 50%	удовлетворительно	обладают общими теоретическими знаниями, умеют применять, исследовать, идентифицировать, анализировать, систематизировать, распределять по категориям, рассчитать показатели, классифицировать, разрабатывать модели, алгоритмизировать, управлять, организовать, планировать процессы исследования, осуществлять оценку результатов на среднем уровне. Допускаются ошибки, которые студент затрудняется исправить самостоятельно.
49 % и менее	неудовлетворительно	обладают не полным объемом общих теоретическими знаниями, не умеют самостоятельно применять, исследовать, идентифицировать, анализировать, систематизировать, распределять по категориям, рассчитать показатели, классифицировать, разрабатывать модели, алгоритмизировать, управлять, организовать, планировать процессы исследования, осуществлять оценку результатов. Не сформированы умения и навыки для решения профессиональных задач
100% - 50%	зачтено	характеристика показателя соответствует «отлично», «хорошо», «удовлетворительно»
49 % и менее	не зачтено	характеристика показателя соответствует «неудовлетворительно»

## 7. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

### 7.1. Содержание лекций

Тема 1. Введение в интеллектуальные системы бизнес-анализа (ПК-1, ПК-2)  
Основные понятия и определения интеллектуальных систем бизнес-анализа. Приобретение и преобразование знания в процессе обучения и адаптации. Информационная модель внешней среды.

Тема 2. Извлечение знаний из структурированных данных. (ПК-3)  
Построение моделей. Методика извлечения знаний. KDD (Knowledge Discovery in Databases): выборка данных, очистка данных, трансформация данных, Data Mining, интерпретация.

### 7.2 Содержание практических занятий и лабораторных работ

Тема 3. Технологии KDD и Data Mining (ПК-4)

Data Mining задачи: классификация, регрессия

Тема 4. Самообучающиеся системы на основе нейронных сетей. Подготовка данных для машинного обучения (ПК-4)

Обучение нейронных сетей. Ошибка обучения. Ошибка обобщения. Инициализация начальных весов. Выбор порога ошибки. Выбор числа итераций. Режимы обучения (обучение по шагам, обучение по эпохам). Обучающая выборка. Переобучение. Размер обучающей выборки.

### 7.3. Содержание самостоятельной работы

Тема 2. Извлечение знаний из структурированных данных. (ПК-3)

Настройка набора данных, скользящее окно, приведение типов, выделение временных интервалов, квантование, слияние с узлом, замена данных, кросс-таблица, свертка столбцов, сортировка, группировка, конечные классы.

Тема 3. Технологии KDD и Data Mining (ПК-4)

Data Mining алгоритмы: самообучающиеся

Тема 4. Самообучающиеся системы на основе нейронных сетей. Подготовка данных для машинного обучения (ПК-4)

Формализация данных. Методы сбора данных. Стоимость сбора данных. Информативность данных. Очистка данных и обогащение. ETL-процесс. Извлечение данных в ETL. Преобразование данных в ETL.

7.3.1. Примерные вопросы для самостоятельной подготовки к зачету/экзамену  
Приложение 1

7.3.2. Практические задания по дисциплине для самостоятельной подготовки к зачету/экзамену  
Приложение 2

7.3.3. Перечень курсовых работ  
Материалы не предусмотрены

7.4. Электронное портфолио обучающегося  
Размещается контрольная работа

7.5. Методические рекомендации по выполнению контрольной работы  
приложение 6

7.6 Методические рекомендации по выполнению курсовой работы  
Материалы не размещаются

## **8. ОСОБЕННОСТИ ОРГАНИЗАЦИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ ДЛЯ ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ**

### ***По заявлению студента***

В целях доступности освоения программы для лиц с ограниченными возможностями здоровья при необходимости кафедра обеспечивает следующие условия:

- особый порядок освоения дисциплины, с учетом состояния их здоровья;
- электронные образовательные ресурсы по дисциплине в формах, адаптированных к ограничениям их здоровья;
- изучение дисциплины по индивидуальному учебному плану (вне зависимости от формы обучения);
- электронное обучение и дистанционные образовательные технологии, которые предусматривают возможности приема-передачи информации в доступных для них формах.
- доступ (удаленный доступ), к современным профессиональным базам данных и информационным справочным системам, состав которых определен РПД.

## **9. ПЕРЕЧЕНЬ ОСНОВНОЙ И ДОПОЛНИТЕЛЬНОЙ УЧЕБНОЙ ЛИТЕРАТУРЫ, НЕОБХОДИМОЙ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ**

**Сайт библиотеки УрГЭУ**  
<http://lib.usue.ru/>

### **Основная литература:**

2. Назаров Д. М., Бегичева С. В., Ковтун Д. Б., Назаров А. Д. Data Science и интеллектуальный анализ данных: производственно-практическое издание. - Москва: Ай Пи Ар Медиа, 2023. - 305

## Дополнительная литература:

### **10. ПЕРЕЧЕНЬ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, ВКЛЮЧАЯ ПЕРЕЧЕНЬ ЛИЦЕНЗИОННОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНФОРМАЦИОННЫХ СПРАВОЧНЫХ СИСТЕМ, ОНЛАЙН КУРСОВ, ИСПОЛЬЗУЕМЫХ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ**

#### **Перечень лицензионного программного обеспечения:**

Microsoft Windows 10 .Договор № 52/223-ПО/2020 от 13.04.2020, Акт № Tr000523459 от 14.10.2020. Срок действия лицензии -Без ограничения срока.

Microsoft Office 2016.Договор № 52/223-ПО/2020 от 13.04.2020, Акт № Tr000523459 от 14.10.2020 Срок действия лицензии -Без ограничения срока.

Astra Linux Common Edition. Договор №0417-ПО/2019 от 08.05.2019, Акт №Sk000343 от 24.05.2019 и Контракт № 35-У/2018 от 13.06.2018, Акт № УТ213 от 17.12.2018. Срок действия лицензии - без ограничения срока.

МойОфис стандартный. Соглашение № СК-281 от 7 июня 2017. Дата заключения - 07.06.2017. Срок действия лицензии - без ограничения срока.

Deductor Academic. Лицензия Deductor Academic. Срок действия лицензии - без ограничения срока.

R Studio (среда для языка программирования R).Лицензия GNU Affero General Public License v3.Срок действия лицензии - без ограничения срока.

Язык программирования R.Лицензия GNU GPL 2.Срок действия лицензии - без ограничения срока.

#### **Перечень информационных справочных систем, ресурсов информационно-телекоммуникационной сети «Интернет»:**

Справочно-правовая система Гарант. Договор № 58419 от 22 декабря 2015. Срок действия лицензии -без ограничения срока

Справочно-правовая система Консультант +. Договор № 143/223-У/2025 от 02.12.2025 Срок действия лицензии до 31.12.2026

## **11. ОПИСАНИЕ МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ**

Реализация учебной дисциплины осуществляется с использованием материально-технической базы УрГЭУ, обеспечивающей проведение всех видов учебных занятий и научно-исследовательской и самостоятельной работы обучающихся:

Специальные помещения представляют собой учебные аудитории для проведения всех видов занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду УрГЭУ.

Все помещения укомплектованы специализированной мебелью и оснащены мультимедийным оборудованием спецоборудованием (информационно-телекоммуникационным, иным компьютерным), доступом к информационно-поисковым, справочно-правовым системам, электронным библиотечным системам, базам данных действующего законодательства, иным информационным ресурсам служащими для представления учебной информации большой аудитории.

Для проведения занятий лекционного типа презентации и другие учебно-наглядные пособия, обеспечивающие тематические иллюстрации.

**Приложение 1**  
**к рабочей программе**  
**Интеллектуальные системы**  
**бизнес-анализа**

Федеральное государственное бюджетное образовательное учреждение высшего образования  
**УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ**

УТВЕРЖДЕНЫ  
на заседании кафедры бизнес-  
информатики

**Примерные вопросы для самостоятельной подготовки к зачету/экзамену**

**К зачету**

1. Приведите примеры применения различных методов анализа сезонности продаж.
2. Какие факторы необходимо учитывать при анализе закупок?
3. Для каких целей проводится мониторинг условий договоров и поставок товара.
4. Основные критерии оценки поставщиков продукции.
5. Как разрабатываются рейтинговые оценки поставщиков и товаров?
6. Как разрабатываются системы управления входящей логистики (заявки, доставка)?
7. Для каких целей разрабатываются и применяются рейтинги запасов?
8. Какие методы применяются для анализа товарного ассортимента?
9. Поясните основные критерии выбора поставщика
10. В чем заключается эффективность визуального представления данных?

**К экзамену**

1. Учет фактора сезонности в коммерческой деятельности.
2. Основные факторы, определяющие глубину и широту ассортимента торгового предприятия.
3. Мониторинг договоров поставок с учетом скидок, бонусов и штрафов.
4. Формирование базового и экспериментального ассортимента.
5. Контроль дефицита и неликвидов в ассортименте запасов.
6. Мониторинг оборачиваемости запасов.
7. Какими методами оценивается качество данных при прогнозировании спроса?
8. Для чего применяются методы кластеризации при анализе продаж?
9. Какие требования предъявляются к данным при выполнении сравнительного анализа в экспертной деятельности?
10. Как оценивается достоверность классификации в экспертизе товаров?
11. В каких информационных системах создаются динамические модели оценки качества продукции?

**Практические задания по дисциплине для самостоятельной  
подготовки к зачету/экзамену**

**Открытые вопросы (ПК-1, ПК-2, ПК-3, ПК-4)**

1. Назовите основные этапы разработки интеллектуальных систем бизнес-анализа;
2. В чем состоит подготовка исходных данных к машинному обучению;
3. Создание Хранилища данных и организация доступа. Назовите структуры хранилищ данных.
4. Для чего проводится оценка качества данных.
5. Какую цель преследует редактирование выбросов и аномальных значений;
6. Что понимаете под поиском и редактирование дубликатов и противоречий;
7. Что дает спектральная обработка ряда;
8. Квантование данных. Назначение квантования. Параметры квантования, определение границ интервалов и задание меток. Приведите пример.
9. Подготовка временного ряда к прогнозированию. Скользящее окно. Критерии выбора погружения в ряд. Приведите пример скользящего окна
10. Применение скриптов в сценариях обработки данных. Приведите пример
11. Выявление закономерностей между связанными событиями. Ассоциативные правила. Шаблоны покупок. Достоверность и Лифт ассоциативных правил;
12. Самоорганизующиеся карты (Self Organizing Maps - SOM). Кластеризации многомерных векторов – алгоритм проецирования с сохранением топологического подобия. Инициализация начальных весов. Область применения карт Кохонена.
13. Линейная регрессия. Коэффициенты регрессии. Область применения;
14. Логистическая регрессия. Прогнозирование бинарной переменной. Коэффициенты регрессии. ROC-анализ. Lift-анализ. Оценка качества модели;
15. Нейронная сеть. Требования к обучающей и тестовой выборке. Задание топологии (структуры) нейронной сети. Скрытые слои и активационная функция. Определение числа связей и переобучение сети. Область применения нейронных сетей.

**Закрытые вопросы (ПК-1, ПК-2, ПК-3, ПК-4)**

Что означает аббревиатура BI?

- a. Business Insurance
- b. Business Intelligence
- c. Business Innovation
- d. Business Investment

Какой тип данных необходим для построения отчетов в BI системах?

- a. Картинки
- b. Текстовые данные
- c. Числовые данные
- d. Звуковые данные

Какая функция BI систем позволяет пользователю производить анализ данных самостоятельно?

- a. Интеграция данных
- b. Визуализация данных
- c. Самообучение системы
- d. Интерактивные отчеты

Какое ПО можно использовать для создания отчетов в BI системах?

- a. Excel
- b. Word
- c. PowerPoint
- d. Outlook

Что такое ETL в BI системах?

- a. Экстракция, трансформация и загрузка данных
- b. Экспорт, трансляция и локализация данных
- c. Эксперименты, тестирование и логирование данных
- d. Экономический анализ, тренды и линейная регрессия данных

Какая из перечисленных функций относится к аналитике данных в BI системах?

- a. Импорт данных
- b. Экспорт данных
- c. Визуализация данных
- d. Форматирование отчетов

Какой тип графиков позволяет отобразить связь между двумя переменными в BI системах?

- a. Линейный
- b. Круговой
- c. Столбчатый
- d. Диаграмма рассеяния

Какую роль играют дашборды в BI системах?

- a. Они позволяют отображать данные в реальном времени
- b. Они помогают анализировать данные на основе определенных критериев
- c. Они позволяют обновлять исходные данные в системе
- d. Они предоставляют возможность редактировать отчеты пользователю

Какой тип анализа данных позволяет определить наиболее эффективную стратегию продаж?

- a. SWOT-анализ
- b. Анализ дерева принятия решений
- c. Анализ мнений экспертов
- d. Анализ продуктовых матриц

Что означает аббревиатура BI?

- A. Business Insider
- B. Business Intelligence
- C. Business Interconnect
- D. Business Investigation

Какие данные включает в себя система BI?

- A. Только структурированные данные
- B. Только неструктурированные данные
- C. Структурированные и неструктурированные данные

D. Данные только из внутренних источников

Какую функцию выполняют инструменты визуализации данных в BI системах?

- A. Анализируют данные
- B. Хранят данные
- C. Позволяют быстро и удобно просматривать данные
- D. Извлекают данные из внешних источников

Какие процессы включает в себя цикл BI?

- A. Извлечение, преобразование, загрузка
- B. Планирование, анализ, отчетность
- C. Визуализация, прогнозирование, принятие решений
- D. Исследование, разработка, тестирование

Какая компонента BI системы обеспечивает анализ и обработку данных?

- A. Интерфейс пользователя
- B. База данных
- C. ETL-система
- D. Аналитический движок

Что означает аббревиатура ETL?

- A. Extract, Transform, Load
- B. Extract, Test, Launch
- C. Extract, Transform, Link
- D. Extract, Train, Learn

Какие типы аналитики включает в себя BI система?

- A. Описательная, диагностическая, предиктивная, прескриптивная
- B. Функциональная, структурная, объектно-ориентированная, логическая
- C. Вероятностная, кластерная, регрессионная, классификационная
- D. Линейная, целочисленная, векторная, матричная

Какой процесс BI цикла отвечает за извлечение данных из различных источников?

- A. Преобразование
- B. Загрузка
- C. Извлечение
- D. Анализ

Что такое BI система?

- a. Система бизнес-интеллекта
- b. Система бизнес-информации
- c. Система бизнес-индексации
- d. Система бизнес-инвентаризации

Какой из этих компонентов BI системы отвечает за анализ данных?

- a. ETL
- b. OLAP
- c. Reporting
- d. Data mining

Что означает термин ETL в BI системах?

- a. Extract, Transform, Load

- b. Explore, Transform, Learn
- c. Extract, Transfer, Load
- d. Explore, Transfer, Learn

Какой тип анализа данных позволяет выявлять скрытые зависимости между различными переменными?

- a. Deskриптивный анализ
- b. Кластерный анализ
- c. Анализ временных рядов
- d. Анализ данных на основе ассоциаций

Какой из этих типов отчетности позволяет анализировать данные на основе разных измерений?

- a. Табличный отчет
- b. Сводная таблица
- c. Графический отчет
- d. Динамический отчет

Какие инструменты BI системы позволяют создавать интерактивные дашборды?

- a. OLAP
- b. Data mining
- c. Reporting
- d. Visualization

Что такое ключевые показатели эффективности (KPI) в BI системах?

- a. Набор параметров, которые отображают текущее состояние бизнеса
- b. Метрики, которые помогают оценить эффективность бизнеса
- c. Инструменты, которые помогают анализировать данные
- d. Методы, которые позволяют прогнозировать развитие бизнеса

Какой тип анализа данных используется для прогнозирования будущих результатов?

- a. Deskриптивный анализ
- b. Диагностический анализ
- c. Прогностический анализ
- d. Поиск аномалий

Какой из этих компонентов BI системы отвечает за хранение данных?

- a. ETL
- b. OLAP
- c. Data warehouse
- d. Data mining

Какой тип графика чаще всего используется для отображения изменения данных со временем?

- a. Столбчатая диаграмма
- b. Круговая диаграмма
- c. Линейный график
- d. Гистограмма

Что такое BI система?

- a) Система бизнес-аналитики
- b) Система управления проектами

- c) Система для разработки программного обеспечения
- d) Система для автоматизации производства

Какие из перечисленных инструментов используются в BI системах?

- a) ETL
- b) CRM
- c) ERP
- d) OLAP

Какую роль играют панели управления в BI системах?

- a) Позволяют анализировать данные в реальном времени
- b) Строят графики и диаграммы на основе данных
- c) Отображают важные показатели в виде карточек
- d) Обеспечивают хранение и обработку данных

Какая функция отвечает за связь между исходными данными и BI системой?

- a) ETL
- b) OLAP
- c) BI-сервер
- d) CRM

Какие типы отчетов можно создавать в BI системах?

- a) Отчеты о прибыли и убытках
- b) Отчеты о посещаемости веб-сайта
- c) Отчеты о количестве продаж
- d) Все вышеперечисленные

Что такое OLAP-куб?

- a) Способ хранения и анализа данных
- b) Система автоматизации производства
- c) Система управления ресурсами предприятия
- d) Система машинного обучения

Какой инструмент используется для визуализации данных в BI системах?

- a) Tableau
- b) SAP
- c) Oracle
- d) Microsoft Dynamics

Что такое ETL процесс?

- a) Процесс выгрузки данных из BI системы
- b) Процесс загрузки данных в BI систему
- c) Процесс преобразования данных перед загрузкой в BI систему
- d) Процесс просмотра отчетов в BI системе

Какие типы данных можно анализировать в BI системах?

- a) Данные о продажах
- b) Данные о клиентах
- c) Данные о производственных мощностях
- d) Все вышеперечисленные

Какое преимущество дает использование BI систем?

- a) Более точное прогнозирование будущих результатов
- b) Увеличение скорости производства
- c) Снижение затрат на персонал
- d) Улучшение качества продукции

УТВЕРЖДЕНЫ  
на заседании кафедры бизнес-  
информатики

**МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ВЫПОЛНЕНИЮ КОНТРОЛЬНОЙ РАБОТЫ  
ДЛЯ СТУДЕНТОВ ОЧНО-ЗАОЧНОЙ ФОРМЫ ОБУЧЕНИЯ**

**по дисциплине  
Интеллектуальные системы бизнес-анализа**

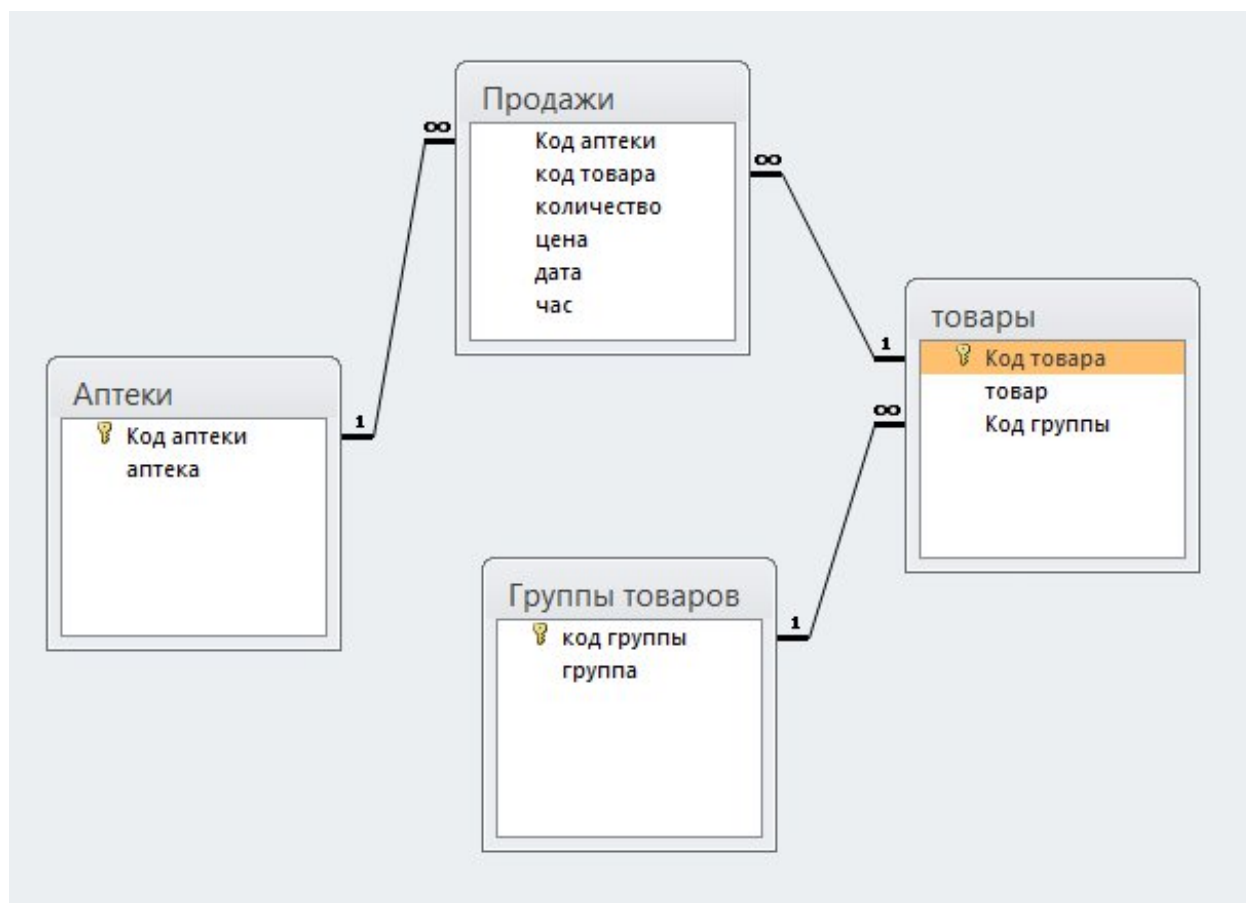
## Контрольная работа

**Постановка задачи.** Компания, владеющая небольшой аптечной сетью, занимается розничной продажей лекарственных препаратов. Руководство компании приняло решение о внедрении системы аналитической OLAP-отчетности, в которой его интересует информация о динамике продаж, загруженности торговых точек, самых продаваемых товарах в различных разрезах.

Так как существующая учетная система испытывает нагрузки (компания постоянно расширяет свою сеть), было решено создать единый консолидированный источник — хранилище данных, которое послужит базой для OLAP-отчетности.

**Исходные данные.** Представлены в четырех файлах: Группы товаров.txt, Товары.txt, Аптеки.txt, Продажи.txt.

Все данные взяты из четырех таблиц БД Access, имеющей следующую структуру данных:



Покажем последовательность решения задачи в аналитической платформе Deductor.

### *Deductor Warehouse*

Хранилище данных Deductor Warehouse — это специально организованная база данных, ориентированная на решение задач анализа данных и поддержки принятия решений, обеспечивающая максимально быстрый и удобный доступ к информации.

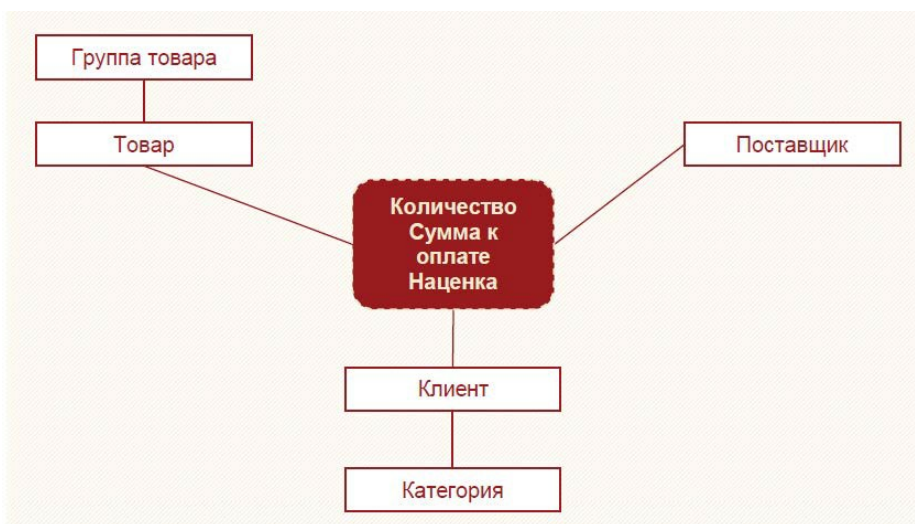
ХД Deductor Warehouse соответствует модели ROLAP (схема «снежинка») и может быть развернуто на одной из следующих СУБД: Firebird, SQL Server, Oracle (в версии Academic — только на Firebird).

Хранилище данных Deductor Warehouse включает в себя потоки данных, поступающие из различных источников, и специальный семантический слой, содержащий так называемые метаданные (данные о данных). Семантический слой и сами данные хранятся в одной СУБД.



Запрос к хранилищу осуществляется непосредственно сквозь семантический слой, который через внутреннюю систему команд (скрытую от пользователя и аналитика) подбирает запрашиваемую информацию из многообразия хранимых данных. Работу семантического слоя можно сравнить с деятельностью библиотекаря, который по просьбе читателя достает с разрозненных полок книги и раскрывает их на нужных страницах.

Все данные в Deductor Warehouse хранятся в структурах типа «снежинка», где в центре расположены таблицы фактов, а «лучами» являются измерения, причем каждое измерение может ссылаться на другое измерение. Именно эта схема чаще всего встречается в реляционных хранилищах данных.



В Deductor Warehouse имеются следующие типы объектов.

**Измерение** — последовательность значений одного из анализируемых параметров. Например, для параметра *Время* это последовательность календарных дней, для параметра *Регион* — список городов. Каждое значение измерения может быть представлено координатой в многомерном пространстве процесса, например, *Товар*, *Клиент*, *Дата*.

**Атрибут** — свойство измерения (то есть точки в пространстве). Атрибут как бы скрыт внутри другого измерения и помогает пользователю полнее описать исследуемое измерение. Атрибутами измерения *Товар* могут выступать *Цвет*, *Вес*, *Габариты*.

**Факт** — значение, соответствующее измерению. Факты — это данные, отражающие сущность события. Как правило, фактами являются численные значения, например, сумма и количество отгруженного товара, скидка.

**Ссылка на измерение** — установленная связь между двумя и более измерениями. Дело в том, что некоторые бизнес-понятия (соответствующие измерениям в хранилище данных) могут образовывать иерархии, например, Товары могут включать Продукты питания и Лекарственные препараты, которые, в свою очередь, подразделяются на группы продуктов и лекарств и т. д. В этом случае первое измерение содержит ссылку на второе, второе — на третье и т. д.

**Процесс** — совокупность измерений, фактов и атрибутов. По сути, процесс и есть «куб», «снежинка». Процесс описывает определенное действие, например продажи товара, отгрузки, поступления денежных средств и пр.

**Атрибут процесса** — свойство процесса. Атрибут процесса, в отличие от измерения, не определяет координату в многомерном пространстве. Это справочное значение, относящееся к процессу, например, № накладной, Валюта документа и т. д. Значение атрибута процесса, в отличие от измерения, не всегда может быть определено.

В Deductor Warehouse может одновременно храниться множество процессов («звезд» или «снежинок»), имеющих общие измерения, например, измерение Товар, фигурирующее в процессах Поступления и Отгрузка.

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



Информация о принадлежности данных к тому или иному типу (измерение, ссылка на измерение, атрибут или факт) содержится в семантическом слое хранилища. Обратим внимание на то, что:

- таблицы измерений содержат только справочную информацию (коды, наименования и т. п.) и ссылки на другие измерения при необходимости;

- таблица процесса содержит только факты и коды измерений (без их атрибутов).

## 1. Проектирование хранилища «Фармация»

Первая подзадача — спроектировать структуру хранилища нашей аптечной сети. Все данные представлены в четырех файлах\таблицах.

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

В таблице «Группы товаров»: *Код группы* является измерением, а *Группа* — его атрибутом.

В таблице «Товары»: *Код товара* является измерением, *Товар* — его атрибутом, а *Код группы* ссылкой на одноименное измерение.

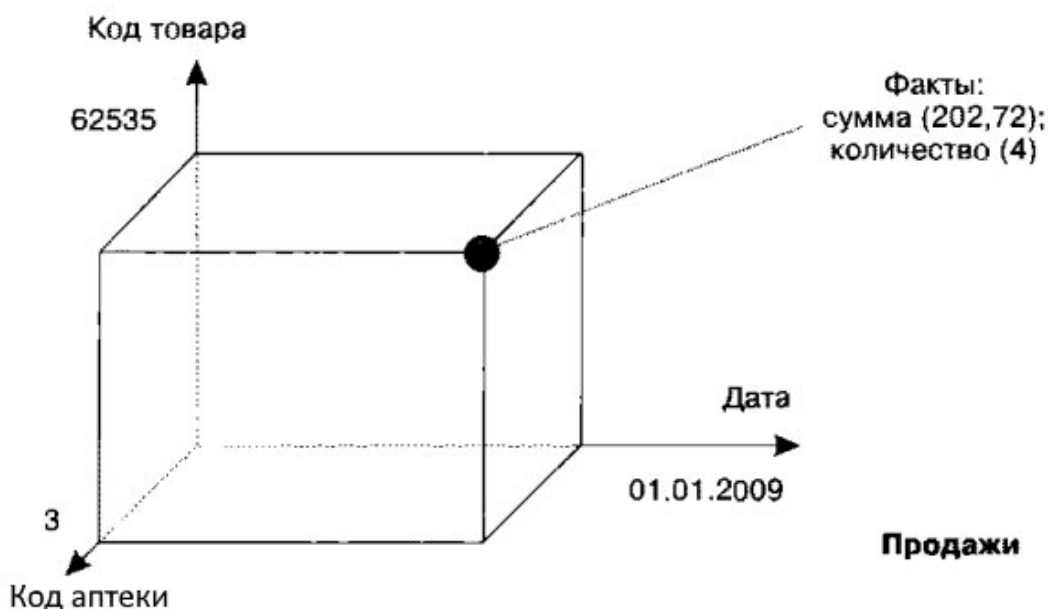
В таблице «Аптеки»: *Код аптеки* — измерение, *Аптека* — его атрибут.

В таблице «Продажи»: *Код аптеки*, *Код товара*, *Дата*, *Час* — измерения, *Количество* и *Цена* — факты.

Таблица «Продажи» представляет собой описание процесса продаж в трех аптеках.

При такой структуре ХД мы предполагаем, что уникальность точки в пространстве определяется совокупностью измерений *Дата* + *Товар* + *Код отдела* + *Час* покупки. То есть если в одной и той же аптеке в один и тот же день и час будет совершено несколько покупок, скажем, препарата «анальгин», то в хранилище данных будет отражена только одна запись.

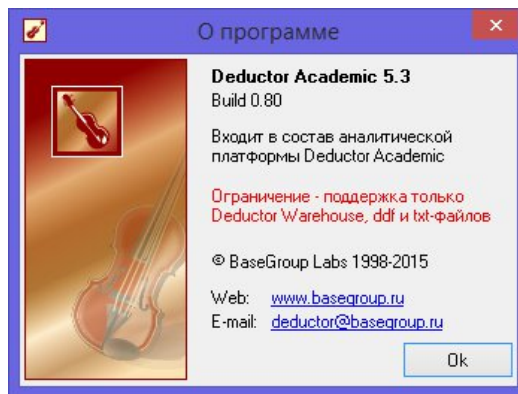
Взаимоотношение измерений, атрибутов и фактов внутри процесса продаж в трех аптеках показано на рисунке



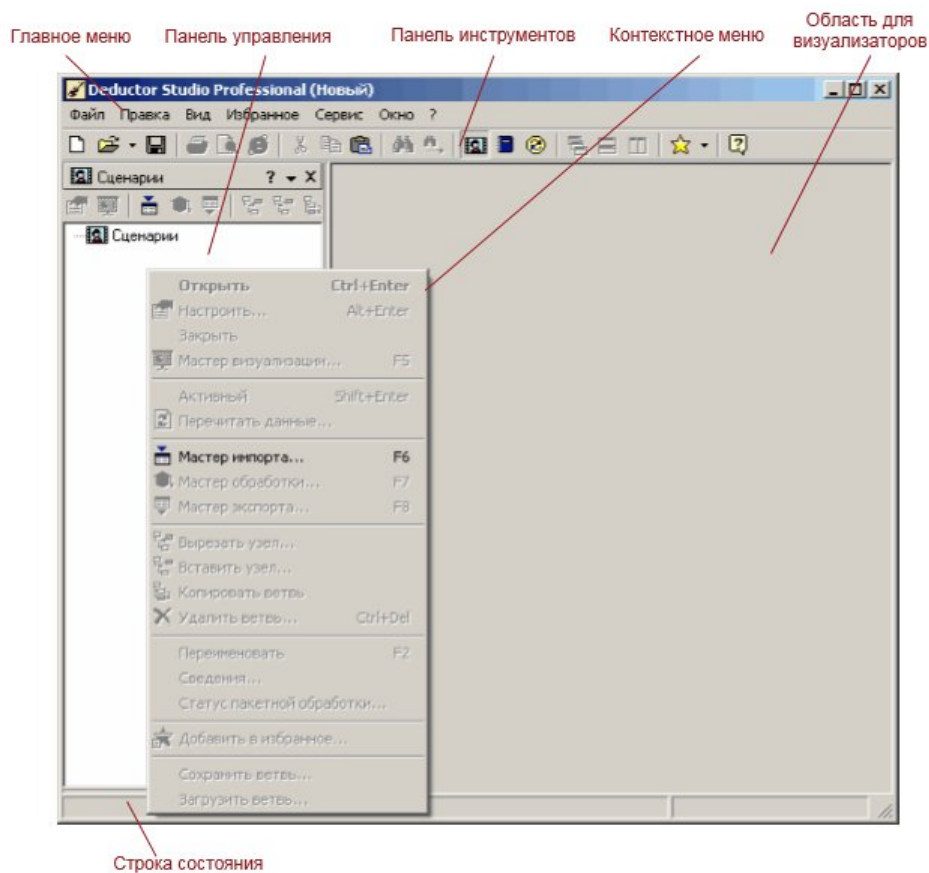
### **Создание Хранилища Данных**

Запустите программу *Deductor Academic* и исследуйте ее информационную среду

Окно запуска программы содержит информацию о версии программы, ограничениях, фирме **BaseGroup Labs**, выпускающей программу.



После нажатия кнопки ОК открывается стартовая страница рабочее окно программы Deductor Studio Academic 5.3.



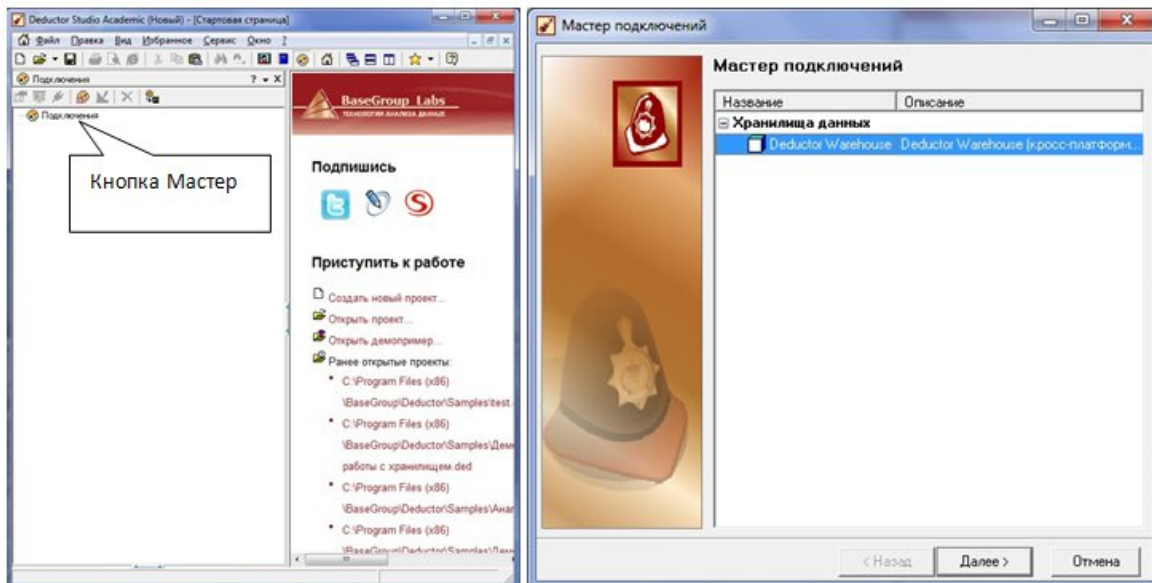
Строка меню содержит следующие элементы:

- Меню **Файл** предназначено для работы с файлами проекта;
- Меню **Правка** позволяет редактировать данные;
- Меню **Вид** включает команды подготовки сценариев (моделей обработки данных), отчетов и выполнение подключений к базам данных или текстовым документам;
- Меню **Сервис** состоит из команд настройки рабочей среды.

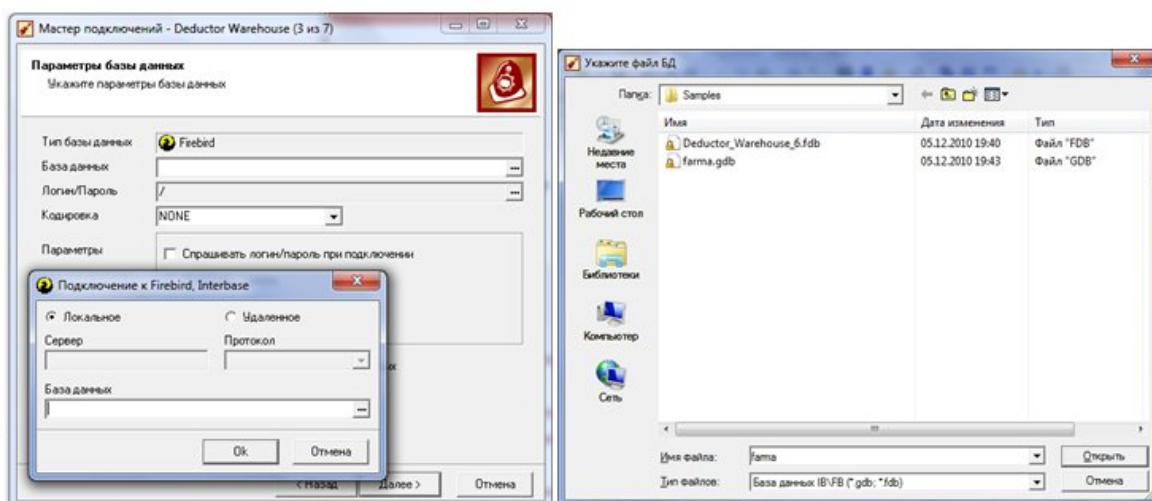
Deductor является аналитической платформой. Реализованные в Deductor технологии позволяют на базе единой архитектуры пройти все этапы построения аналитической системы от создания хранилища данных до автоматического подбора моделей и визуализации полученных результатов.

Создание нового хранилища в Deductor включает три этапа: проектирование хранилища, создание пустого хранилища (создание структуры хранилища) и заполнение хранилища данными.

Для создания нового пустого хранилища данных (ХД) в меню *Вид* выберите вкладку *Подключения*. На этой вкладке запустите *Мастер подключений*. На первом шаге мастера выберите тип источника (приемника) – *Deductor Warehouse*.



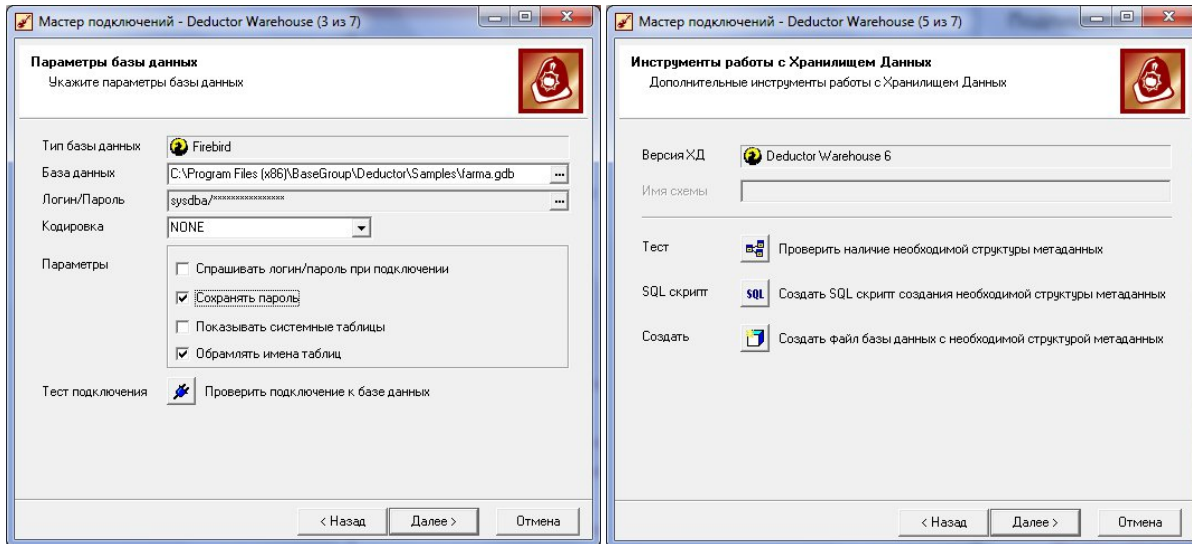
На следующем шаге задайте параметры базы данных, в которой будет создана физическая и логическая структура хранилища данных. В поле *База данных* нажмите кнопку с тремя точками и в открывшемся диалоговом окне *Подключение...* В поле *База данных* этого окна нажмите свою кнопку с тремя точками. Откроется диалоговое окно *Укажите файл БД*. В поле *Имя файла* этого окна запишите имя нового хранилища (например: *farma*) и нажмите кнопку *Открыть*.



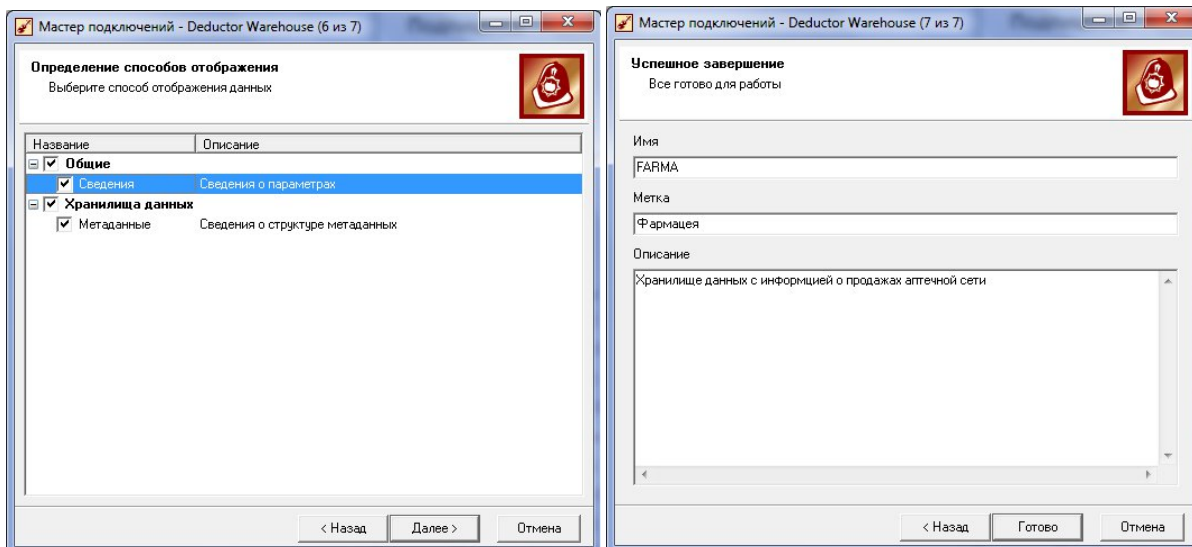
В окне мастера *Параметры базы данных* появится путь к базе данных. Введите логин – *sysdba* и пароль – *masterkey*. Установите флажок *Сохранить пароль* и нажмите кнопку *Далее*. (Пароль можно не вводить).

В окне **Инструменты работы с Хранилищем Данных** нажмите кнопку *Создать файл данных с необходимой структурой метаданных*. Будет создан файл, содержащий пустое хранилище данных.

В нашем примере файл хранилища данных имеет имя: *farma.gdb*. *Этот файл необходимо сдавать преподавателю на проверку*.



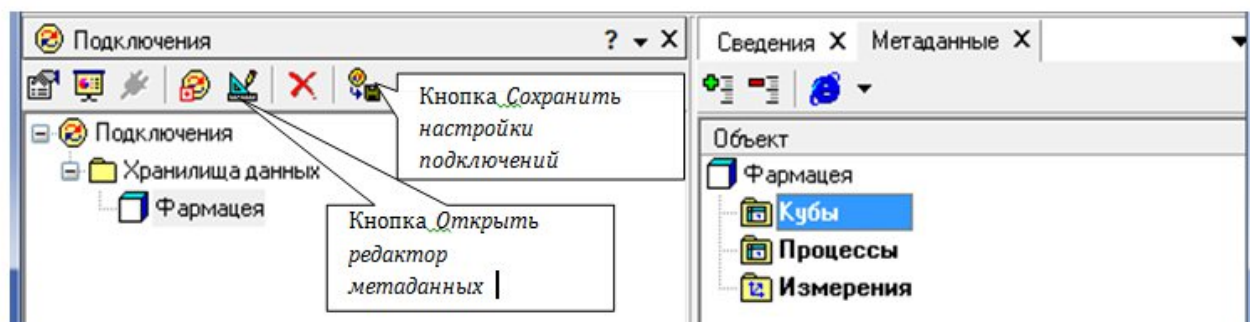
В следующих окнах нужно выбрать визуализатор для подключения (*Сведения и Метаданные*), задать имя, метку и описание нового хранилища. Имя хранилища может быть задано только латинскими буквами.



После нажатия кнопки *Готово* на дереве узлов подключений появится метка хранилища. Нажмите кнопку *Сохранить настройки подключений*.

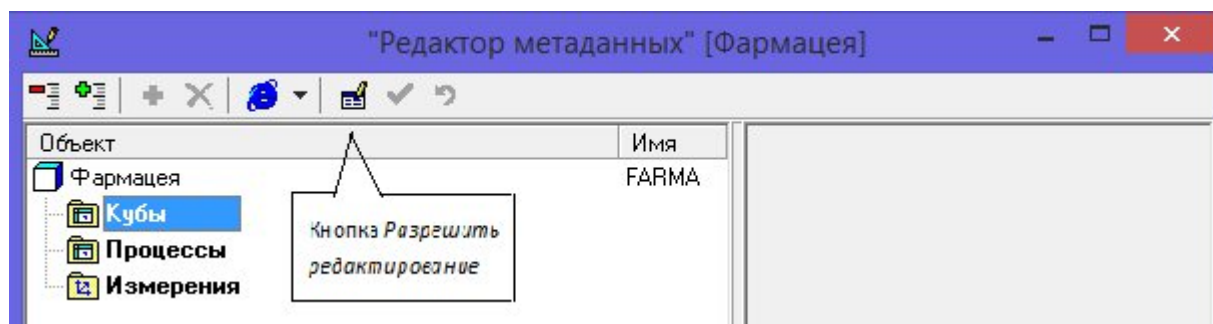
Таким образом, создано пустое хранилище, в котором пока нет ни одного объекта (процесса, измерения, факта).

Теперь нужно отобразить спроектированную базу данных аптечной сети в хранилище. Для этого предназначен редактор метаданных.



Откройте диалоговое окно *Редактор метаданных*, нажав одноименную кнопку на вкладке *Подключение*.

Для перехода в режим изменения структуры хранилища нажмите кнопку *Разрешить редактирование*. Появится диалоговое окно с предупреждением о том, что это небезопасная операция.



Выберите узел *Измерения*, щелкнув по нему правой кнопкой мыши. В контекстном меню выберите пункт *Добавить* и создайте первое измерение *Код группы* со следующими параметрами:

- Имя – *GR\_ID*;
- Метка – *Группа.Код*;
- Тип данных – *целый*.

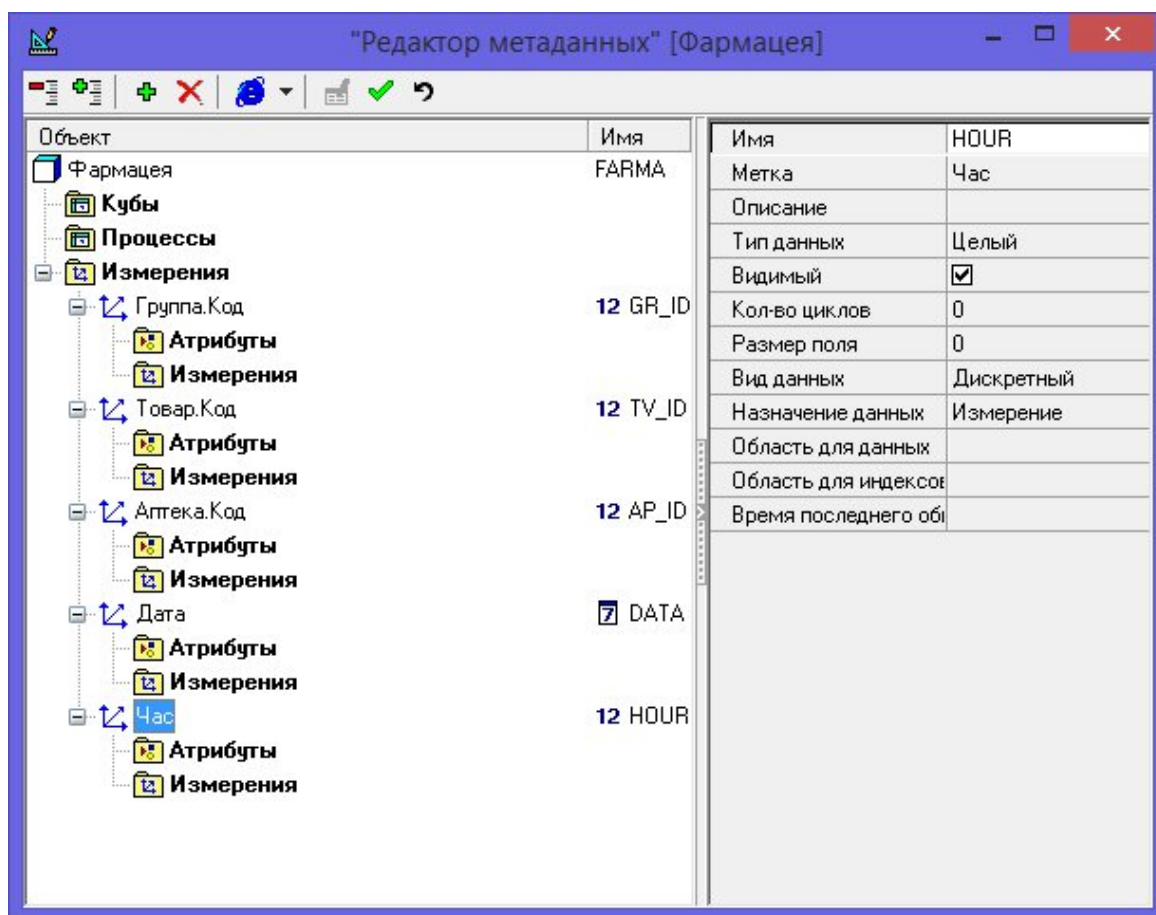
Метка – это семантическое название объекта хранилища данных, которое увидит пользователь, работающий с хранилищем данных.

Проделайте аналогичные действия для создания всех остальных измерений, взяв параметры из таблицы. Зафиксируйте изменения структуры ХД, нажав кнопку *Принять изменения*.

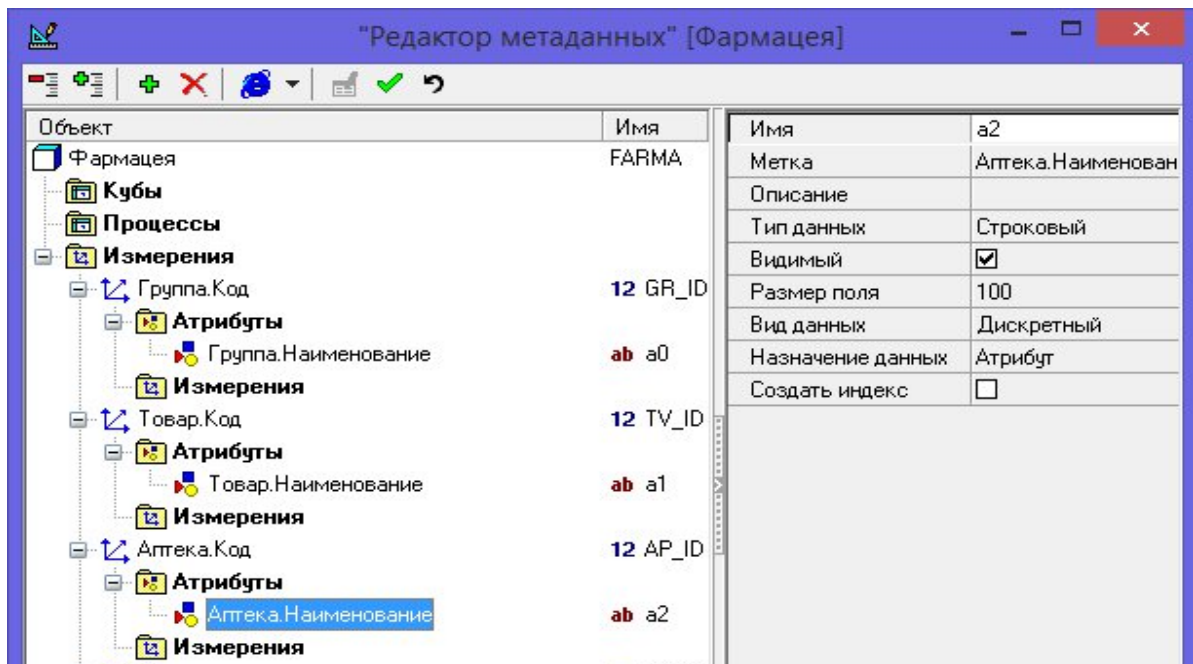
Таблица. параметры изменений

Измерение	Имя	Метка	Тип данных
Код группы	GR ID	Группа.Код	Целый
Код товара	TV ID	Товар.Код	Целый
Код аптеки	AP ID	Аптека.Код	Целый
Дата	DATA	Дата	Дата/Время
Час	HOUR	Час	Целый

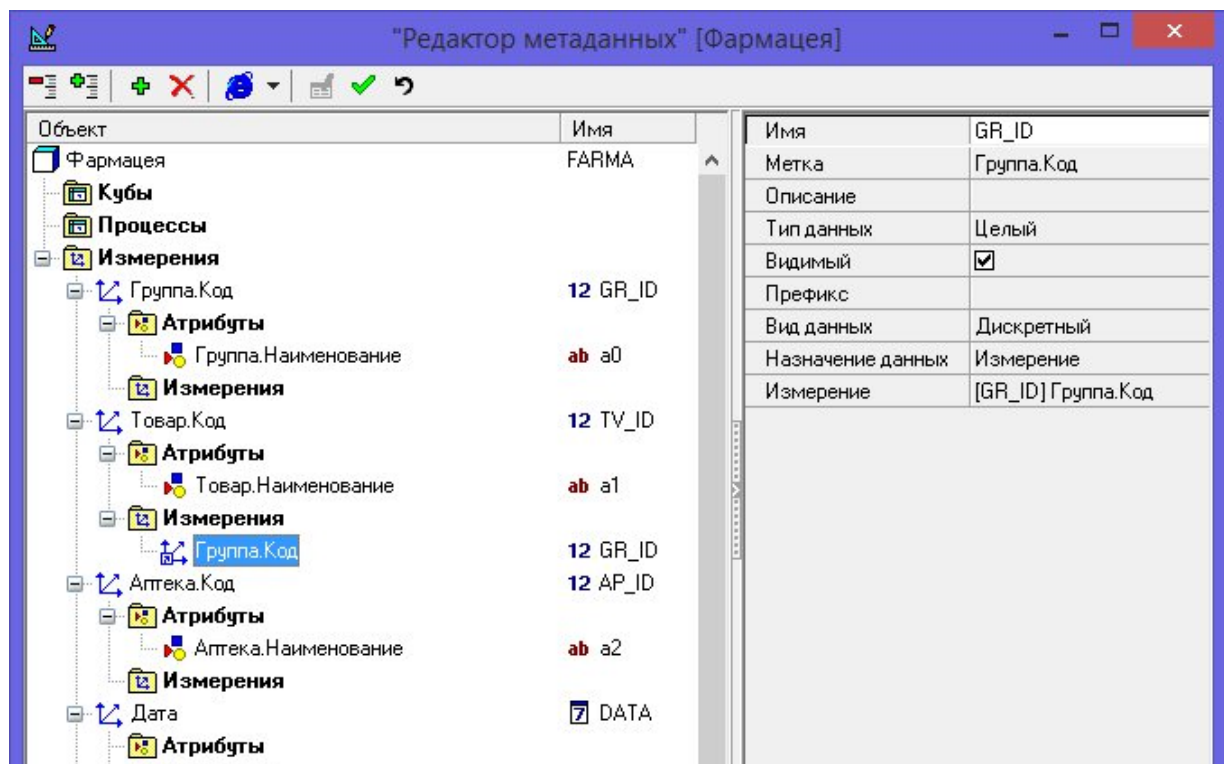
В результате структура метаданных хранилища будет содержать пять измерений.



К каждому измерению, кроме *Дата* и *Час*, добавьте текстовый атрибут (это делается нажатием кнопки *Добавить*). Для измерения *Группа.Код* это будет *Группа.Наименование*, для измерения *Товар.Код* – *Товар.Наименование*, для измерения *Аптека.Код* – *Аптека.Наименование*.

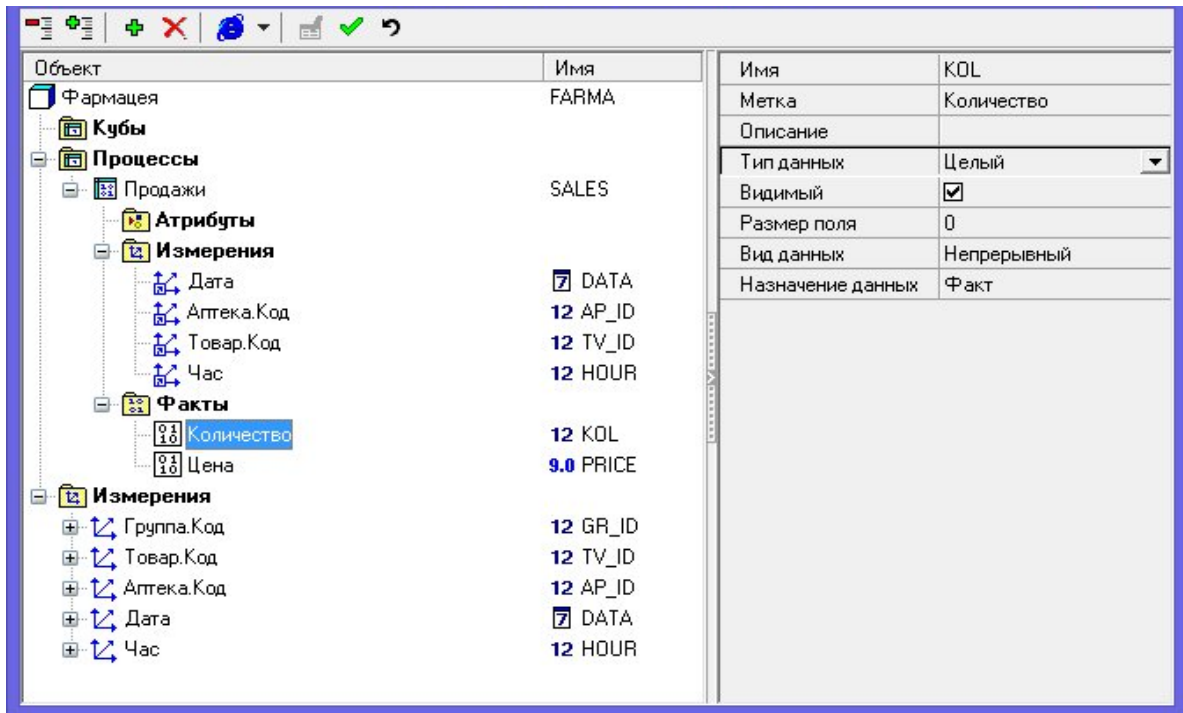



Каждое измерение может ссылаться на другое измерение, реализуя тем самым иерархию измерений. В нашем случае измерение *Товар.Код* ссылается на измерение *Группа.Код*. Эту ссылку и установите путем простого добавления.



После того как все измерения и ссылки созданы, приступайте к формированию процесса ("снежинки"). Назовите его *Продажи* и добавьте четыре существующих измерения: *Дата*, *Аптека.Код*, *Товар.Код*, *Час*.

Кроме них, в процессе участвуют два факта: *Количество* и *Цена*, причем первый — целочисленный, а второй — вещественный.



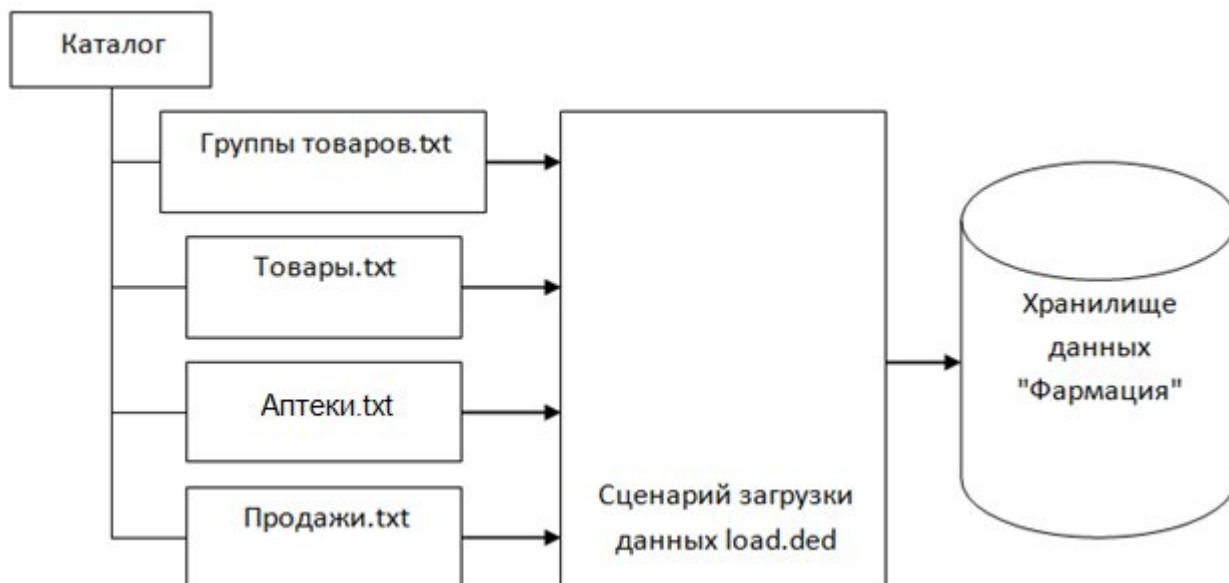
Зафиксируйте изменения структуры ХД, нажав кнопку  *Принять изменения*. На этом проектирование структуры и метаданных закончено.

### Наполнение хранилища данных (загрузка данных)

После создания структуры хранилища 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 написать соответствующий сценарий в Deductor Studio. Он должен выполнять следующие функции:

- Импорт данных в Deductor Studio из базы данных или предопределенных файлов;
- Операционную предобработку данных, например, очистку или преобразование формата;
- Загрузку данных в измерения и процессы хранилища Deductor Warehouse.

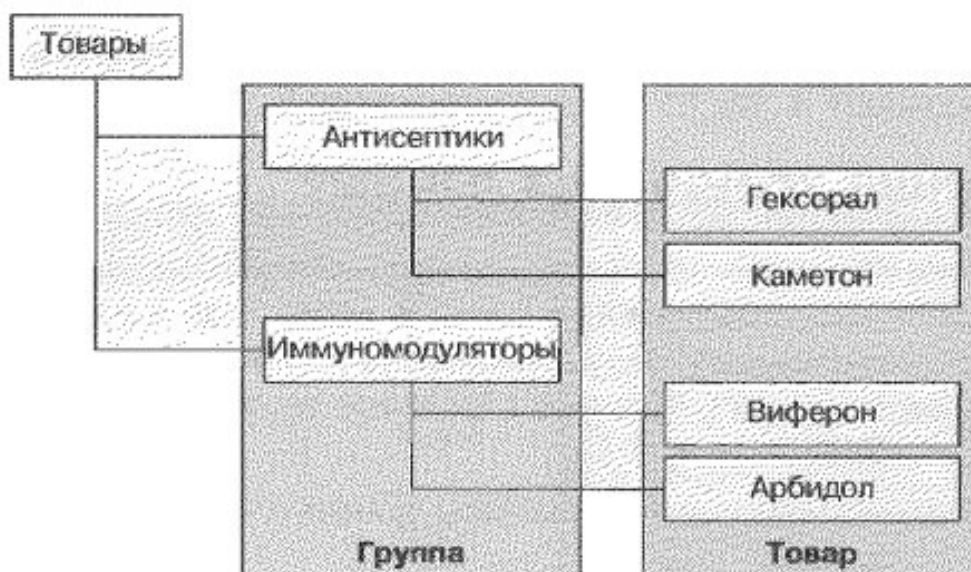
В нашем примере исходными данными для ХД служат четыре текстовых файла: *Группы товаров.txt*, *Товары.txt*, *Аптеки.txt*, *Продажи.txt*. Поэтому сценарий должен быть настроен на использование этих файлов в качестве источников данных.



При создании сценария необходимо строго придерживаться следующих правил:


- Первыми загружаются все измерения, имеющие атрибуты. Только после загрузки всех измерений загружаются данные в процесс.
- Измерения нужно загружать, начиная с самого верхнего уровня иерархии и спускаясь ниже. Это крайне важно: в противном случае иерархия не будет создана.
- Допускается не загружать отдельные измерения, не имеющие атрибутов и не состоящие в иерархии измерений. Значения таких измерений можно создавать во время загрузки в процесс с помощью специальной операции.

Поясним второе правило. Измерение Группа находится в иерархии выше измерения Товар, поэтому последовательность загрузки измерений будет следующая: Группа, Товар.

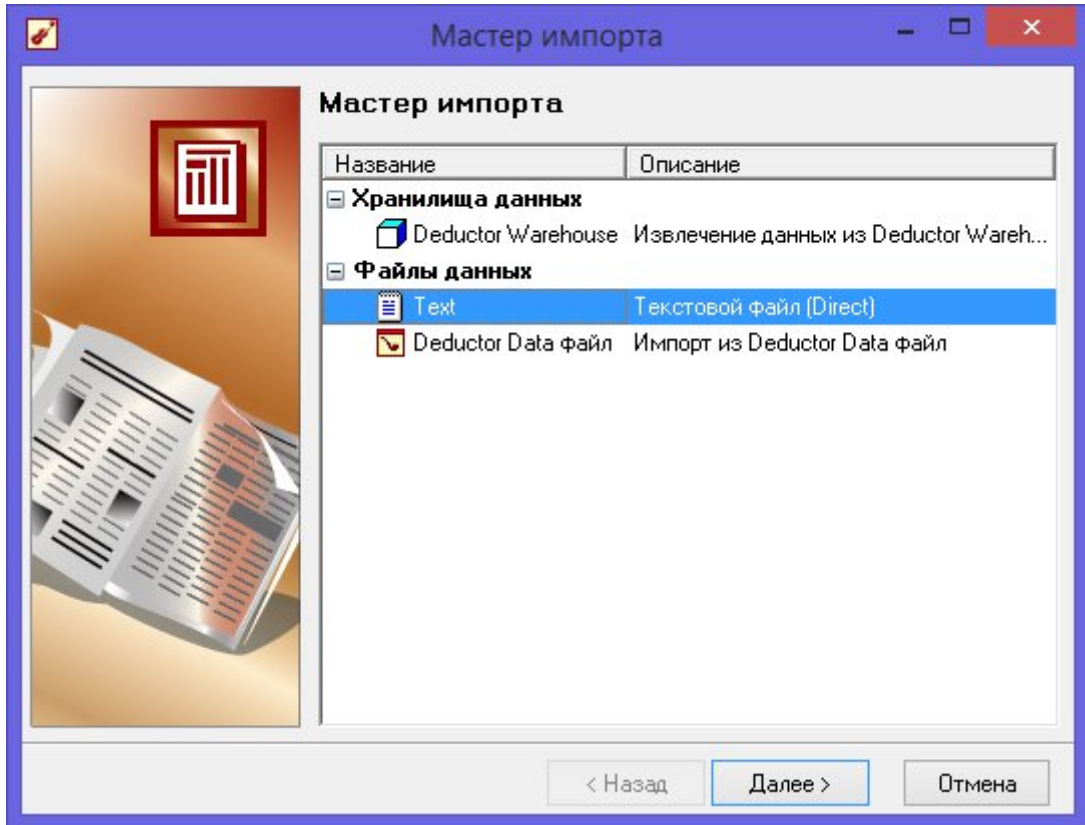


Импортируйте все четыре текстовых файла в Deductor в следующем порядке:

*Группы товаров.txt, Товары.txt, Аптеки.txt, Продажи.txt.*

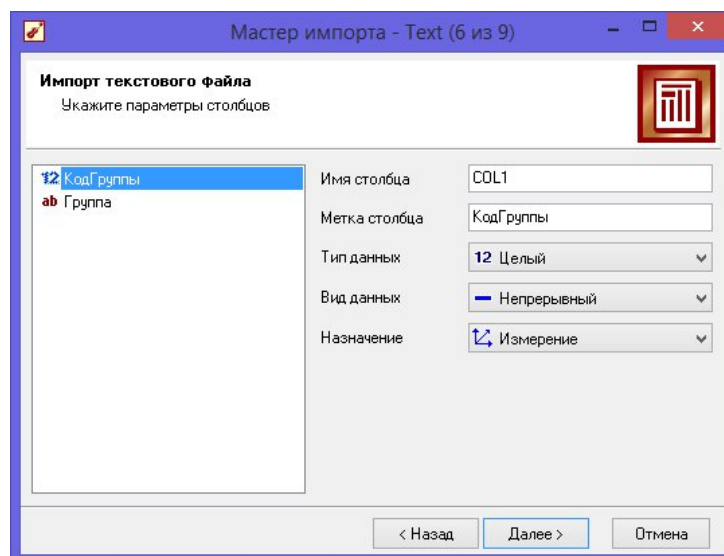
Для этого перейдите на вкладку *Сценарий* и вызовите *Мастер импорта*.  Выберите тип источника

– текстовый файл и нажмите кнопку *Далее*.



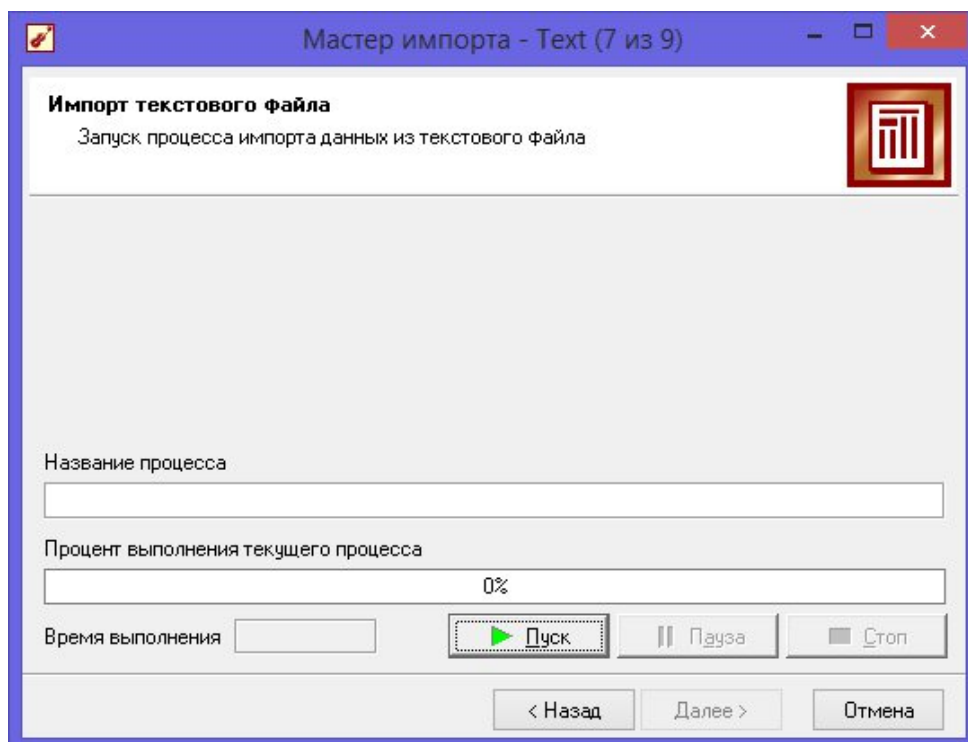
В следующем окне мастера укажите путь к первому текстовому файлу, нажав в поле *Имя файла* кнопку с тремя точками. В окне *Открыть* выделите имя *Группы товаров* и нажмите кнопку *Открыть*.

Последовательно нажимаем кнопку *Далее* в окнах мастера импорта, пока не откроется следующее окно импорта.

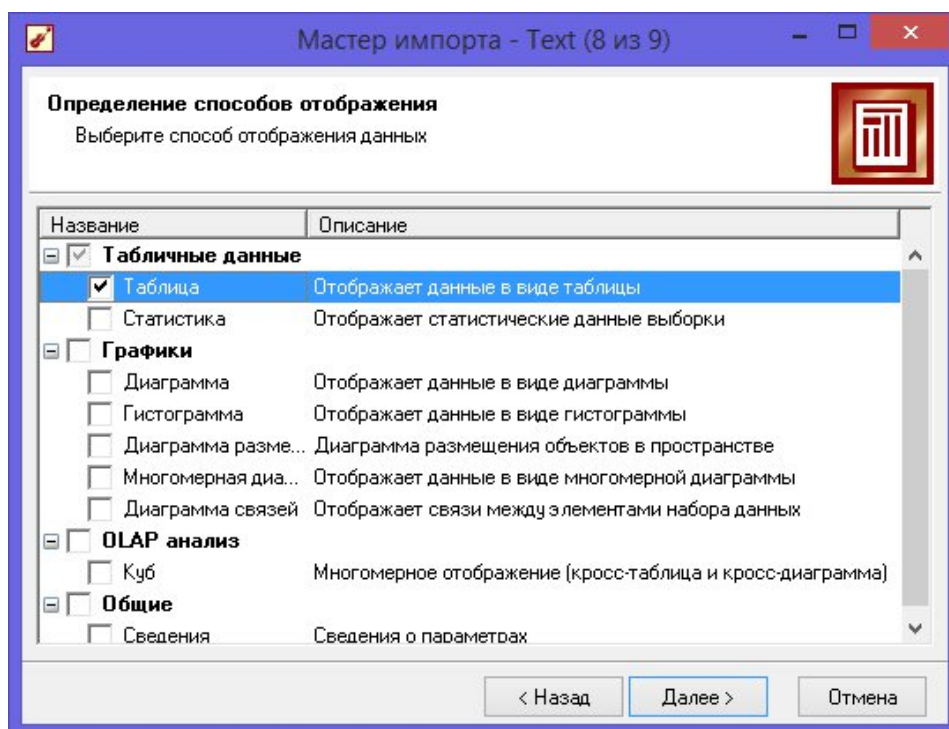


Здесь необходимо установить в полях «Тип данных» и «Назначение» те значения, которые были установлены в структуре ХД.

В противном случае экспорт в ХД будет невозможным. Нажимаем кнопку *Далее*.

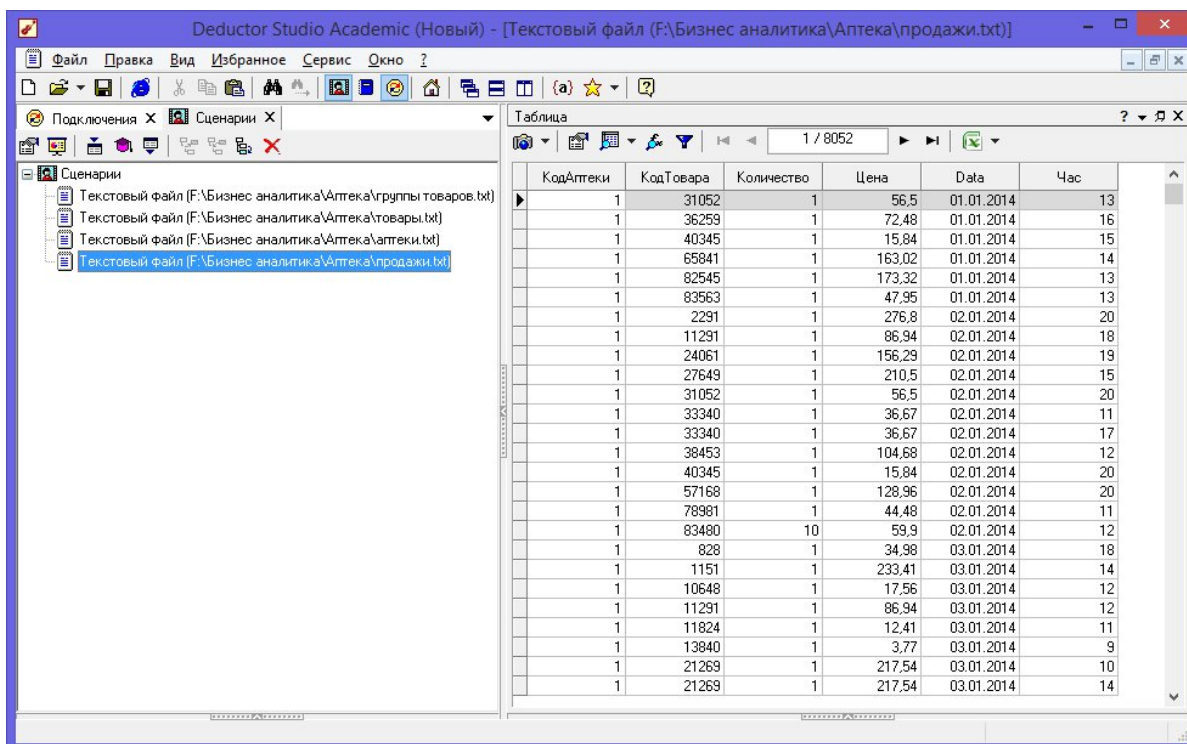


Нажимаем *Пуск*. После удачного завершения процесса нажимаем *Далее*.



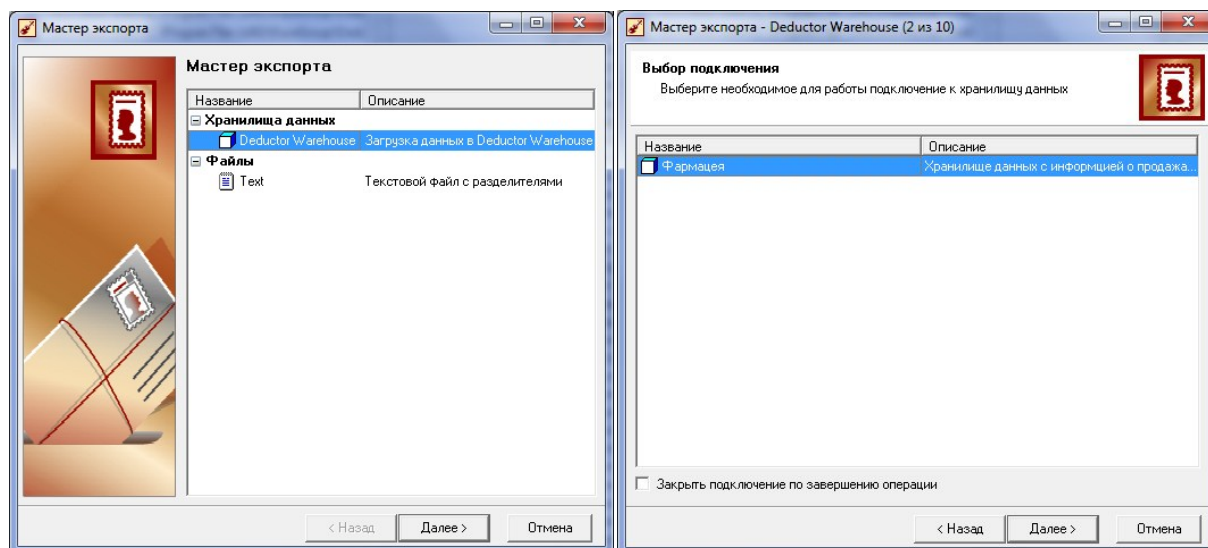
Последовательно нажимая кнопку *Далее* завершите процедуру импорта первого файла.

Аналогичным образом экспортируйте файлы *Товары*, *Аптеки* и *Продажи*. При необходимости порядок следования веток сценария можно менять с помощью клавиш CTRL+↑ и CTRL+↓.

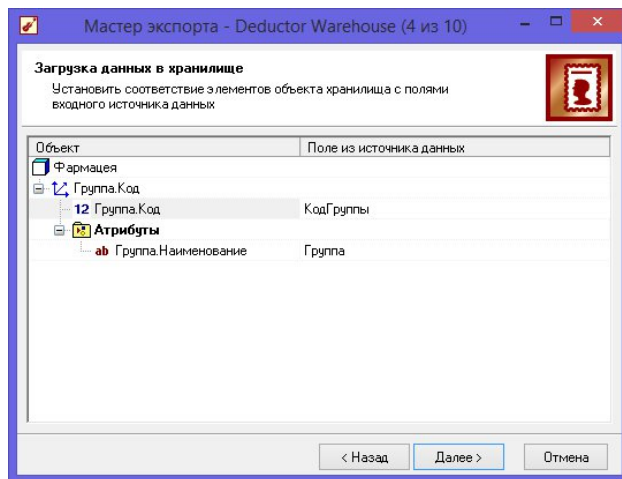
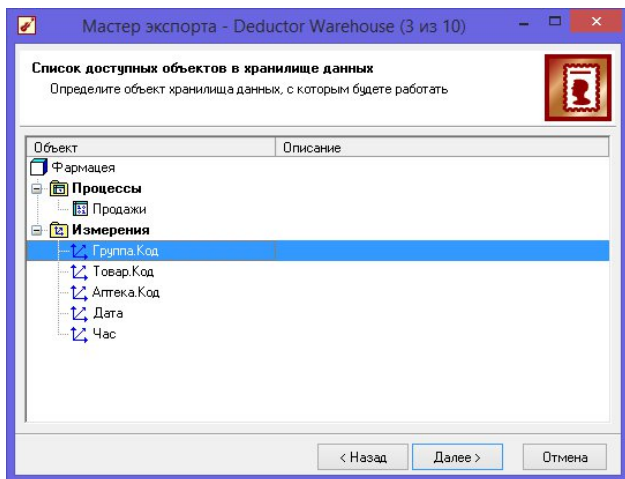


КодАптеки	КодТовара	Количество	Цена	Дата	Час
1	31052	1	56,5	01.01.2014	13
1	36259	1	72,48	01.01.2014	16
1	40345	1	15,84	01.01.2014	15
1	65841	1	163,02	01.01.2014	14
1	82545	1	173,32	01.01.2014	13
1	83563	1	47,95	01.01.2014	13
1	2291	1	276,8	02.01.2014	20
1	11291	1	86,94	02.01.2014	18
1	24061	1	156,29	02.01.2014	19
1	27649	1	210,5	02.01.2014	15
1	31052	1	56,5	02.01.2014	20
1	33340	1	36,67	02.01.2014	11
1	33340	1	36,67	02.01.2014	17
1	38453	1	104,68	02.01.2014	12
1	40345	1	15,84	02.01.2014	20
1	57168	1	128,96	02.01.2014	20
1	78981	1	44,48	02.01.2014	11
1	83480	10	59,9	02.01.2014	12
1	828	1	34,98	03.01.2014	18
1	1151	1	233,41	03.01.2014	14
1	10648	1	17,56	03.01.2014	12
1	11291	1	86,94	03.01.2014	12
1	11824	1	12,41	03.01.2014	11
1	13840	1	3,77	03.01.2014	9
1	21269	1	217,54	03.01.2014	10
1	21269	1	217,54	03.01.2014	14

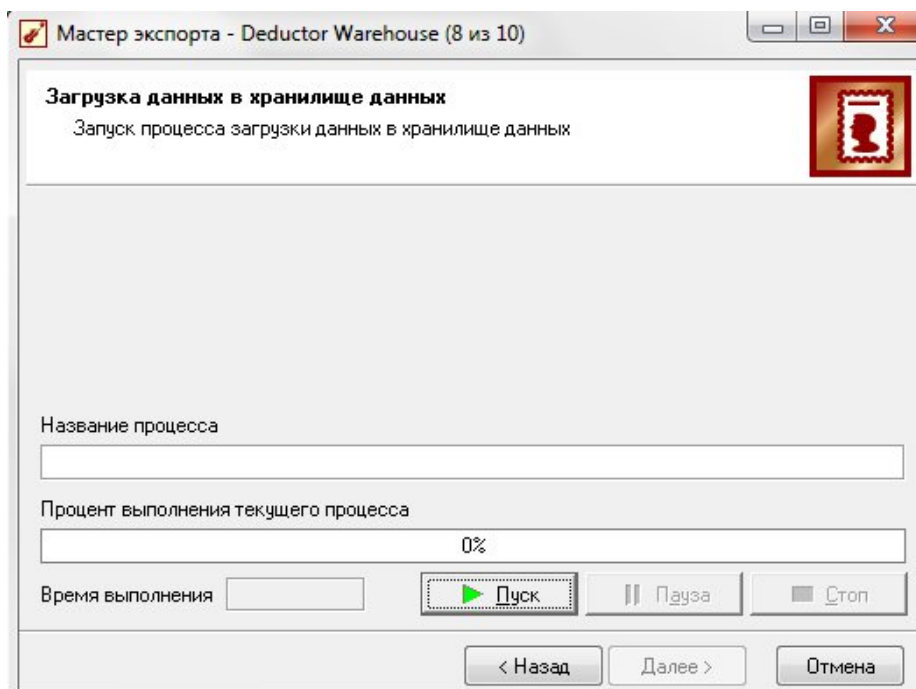
Рассмотрим последовательность загрузки данных в измерение на примере первого измерения *Группа.Код*. Выбрав первый узел сценария, вызовите *Мастер экспорта*. Из списка типа приемников выберите *Deductor Warehouse*. В следующем окне из списка доступных хранилищ выберите нужное (*Фармацев*).



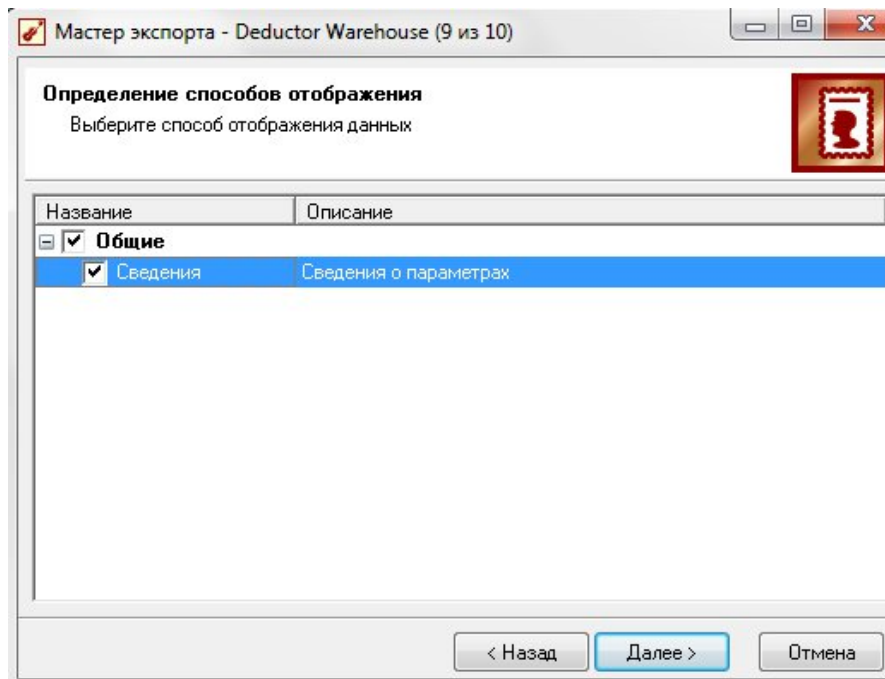
Далее следует указать, в какое именно измерение будет загружаться информация – это *Группа.Код*. Осталось установить соответствие элементов объекта в хранилище данных с полями входного источника данных (то есть таблицы *Группы товаром.txt*). В случае, когда имена полей и (или) метки в семантическом слое хранилища данных совпадают, делать ничего не нужно.



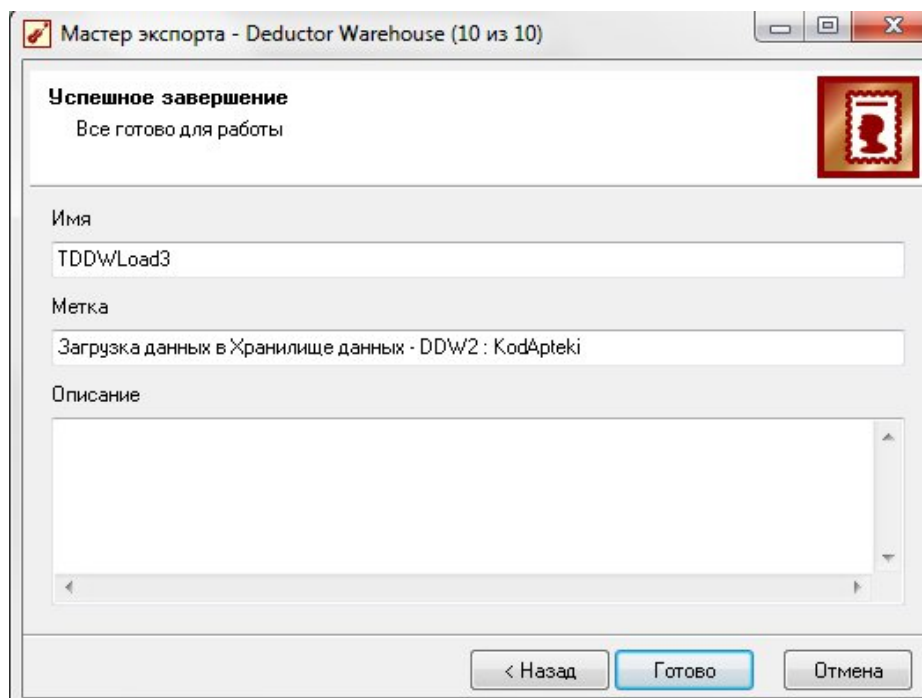
Нажать кнопку *Далее*.



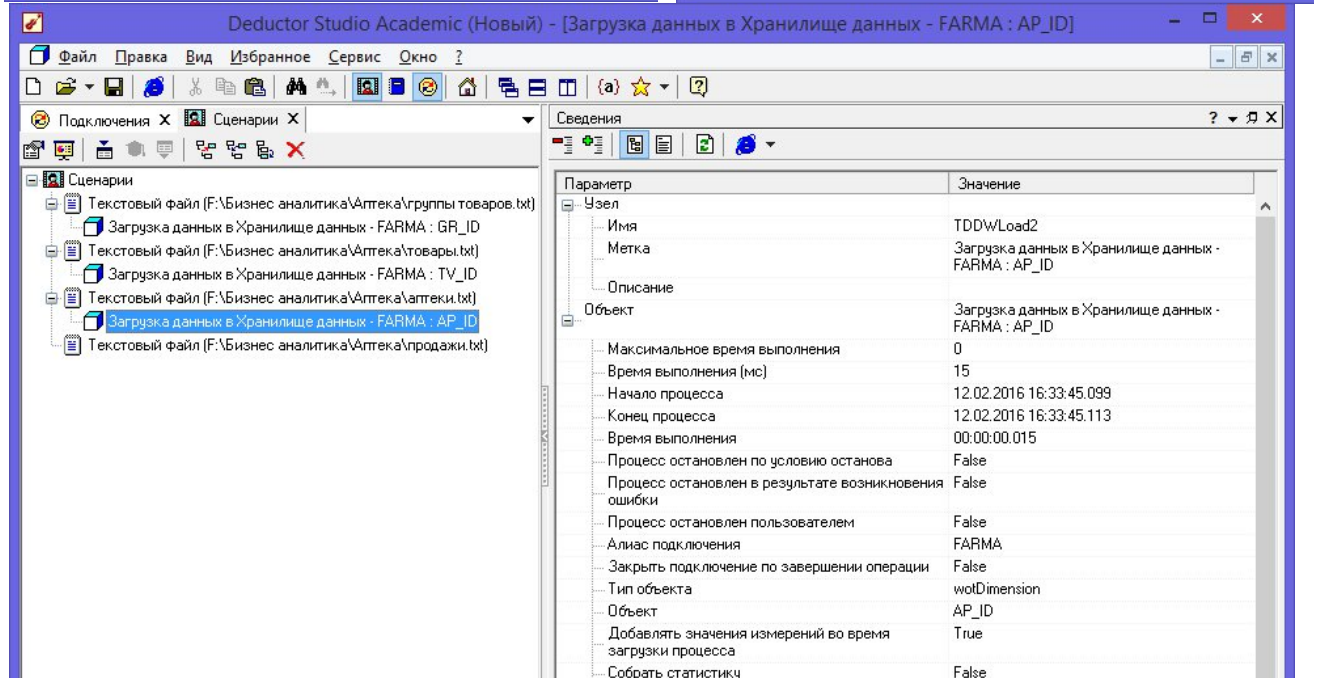
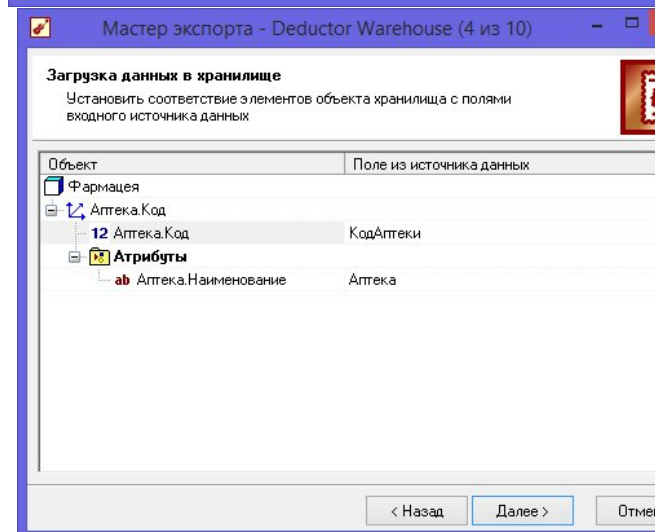
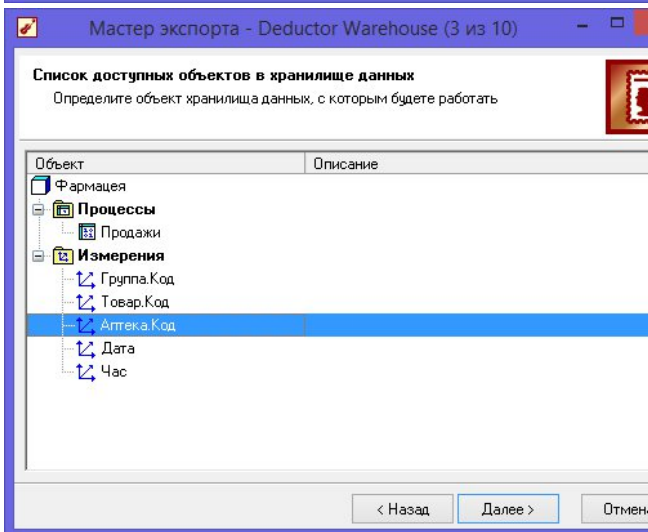
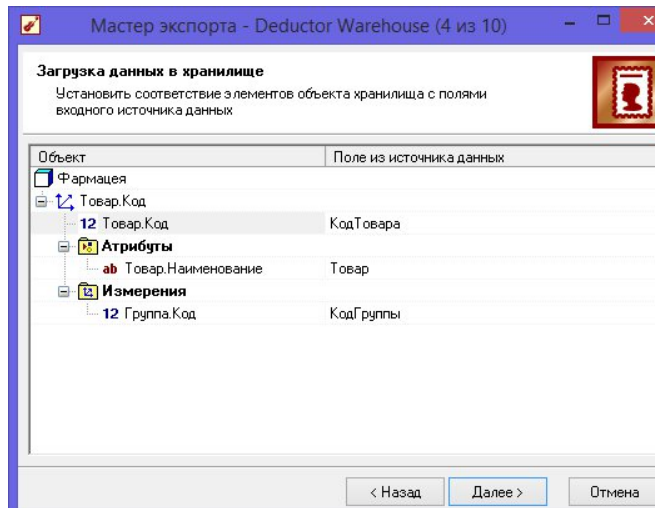
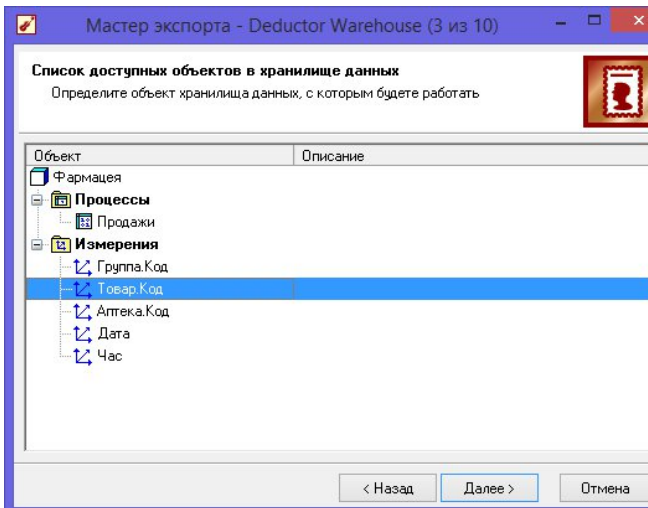
Нажать кнопку *Пуск* и затем *Далее*



Это окно мастера экспорта оставьте без изменения. Нажать Далее.



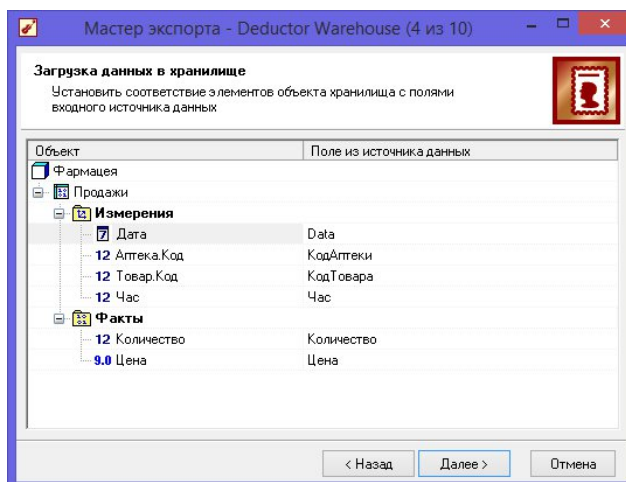
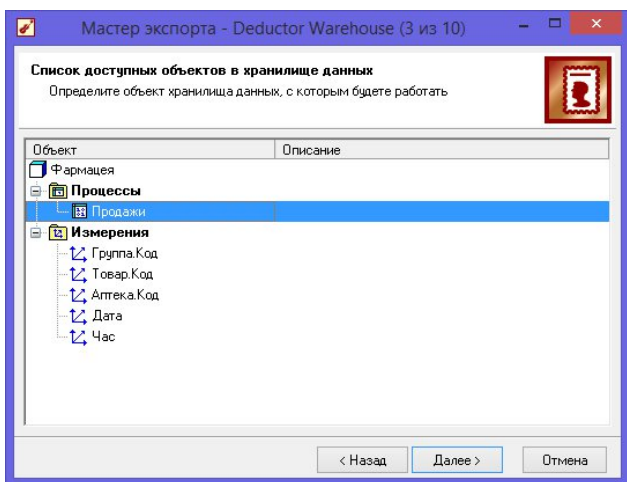
Нажать Готово. В результате загрузятся в измерение данные. При этом старые данные, если они были, обновятся. Прделавав аналогичные действия еще для двух измерений – *Товары.Код* и *Аптеки.Код*, получим следующий сценарий.



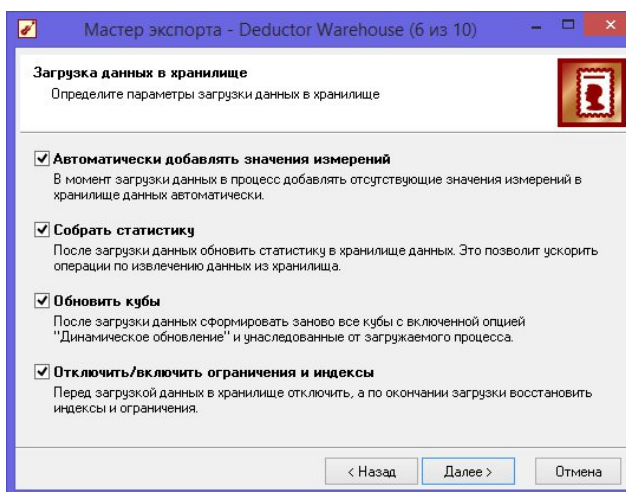
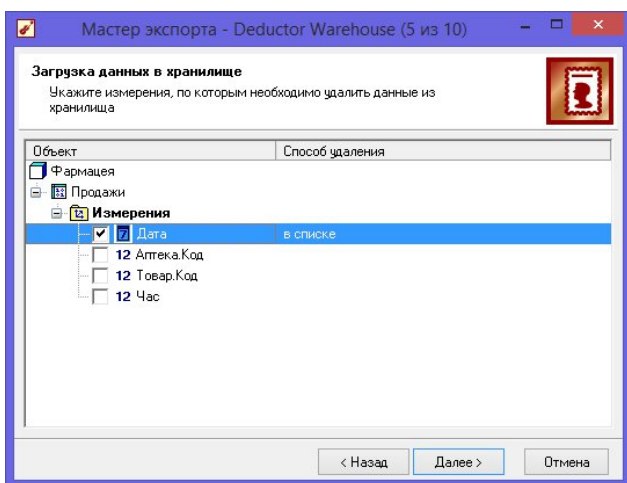
Загрузка измерений на этом заканчивается, несмотря на то, что остались еще два измерения – *Дата* и *Час*. Они не имеют атрибутов и не участвуют в иерархии, поэтому их значения можно загрузить на этапе экспорта в процесс.

**Загрузим данные в процесс Продажи.** В отличие от загрузки измерений в мастере экспорта появятся два специфических шага. На одном из них нужно задать параметры контроля непротиворечивости данных в хранилище – указать измерения, по которым следует удалять данные из хранилища. Выбирается действие, выполняемое в ситуации, когда в процесс загружается информация, которая совпадает по значениям из некоторых измерений. Может быть два варианта: удалить старые данные и загрузить новые или запретить удаление и оставить то, что было загружено ранее.

Выделите последний текстовый файл и запустите *Мастер экспорта*. На третьем шаге выделите процесс *Продажи* и нажмите кнопку *Далее*. На четвертом шаге устанавливается соответствие элементов объекта хранилища с полями входного источника данных.



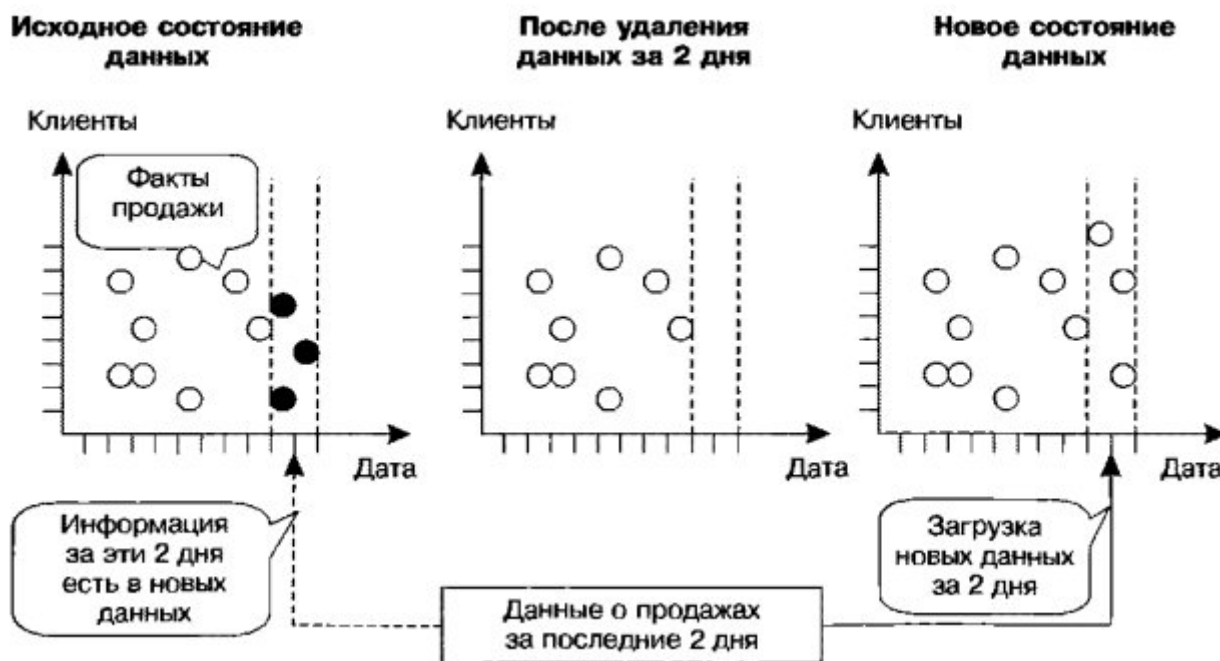
Затем начинается загрузка данных в хранилище. На пятом шаге укажите измерение, по которому будут удаляться данные из хранилища, установив флажок *Дата*. На следующей странице мастера лучше оставить настройки по умолчанию. Флажок *Автоматически добавлять значения измерений* позволяет "на лету" добавлять новые значения в существующие измерения. Эту операцию нужно применять осторожно, так как можно очень быстро засорить хранилище ненужными данными.



Поясним операцию удаления на примере. Допустим, в хранилище имеется процесс с двумя измерениями: Клиент и Дата. Необходимо загрузить в хранилище данные о продажах за последние два дня.

Если в наборе данных, который мы загружаем, имеются все сведения о продажах за эти два дня, то можно указать: «Удалять данные по измерению и выбрать таким измерением Дата».

Программа определит, что по измерению Дата в исходных данных всего два значения, а потом удалит из хранилища в процессе Продажи всю информацию за эти два дня и загрузит новую.



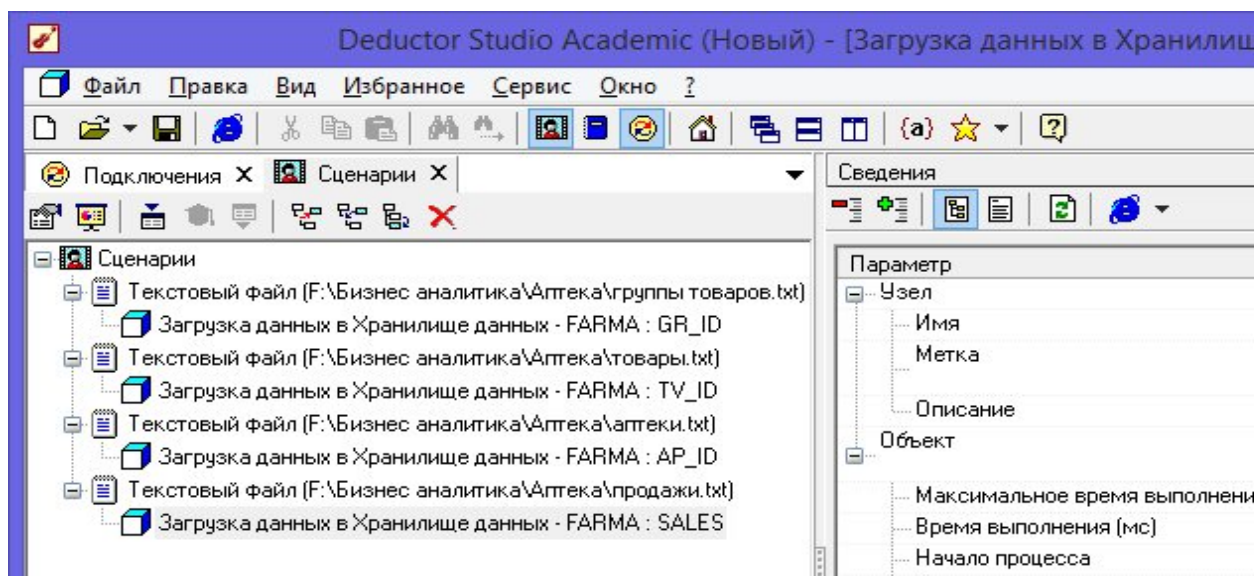
Подобный способ загрузки удобен еще и тем, что позволяет избежать коллизий, например, когда в хранилище имеются некорректные данные за какой-то период. В таком случае лучше все данные за этот период удалить, а после загрузить новые корректные сведения.

На следующем шаге мастера можно задать любой вариант агрегации данных. Выберите что указано на рисунке ниже.



При загрузке данных в хранилище происходит их группировка по измерениям, поэтому необходимо указывать варианты агрегации атрибутов и фактов процесса.

По окончании работы мастера получим следующий сценарий



В результате всех выполненных действий будет:

- создано и наполнено хранилище данных;
- написан сценарий загрузки (пополнения) данных из источника и ХД;
- продуман контроль непротиворечивости данных в ХД.

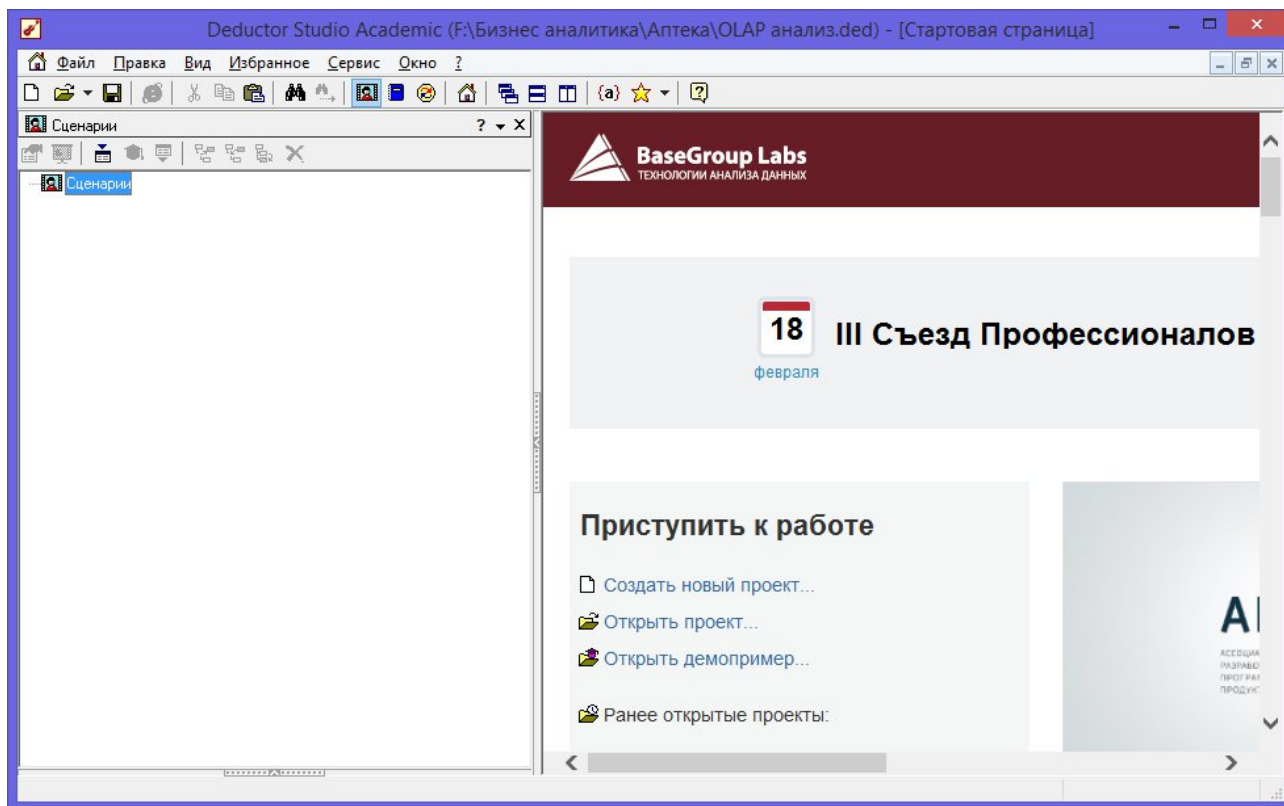
Полученный сценарий загрузки привязан не к данным непосредственно, а к их структуре, то есть в нем смоделирована последовательность действий, которые нужно выполнять для загрузки данных в ХД: имена файлов-источников, соответствие полей и т.д. Один раз созданный сценарий впоследствии применяется для пополнения хранилища данными. Как правило, эти процедуры проводятся по регламенту в нерабочее время (например, ночью) с использованием пакетного или серверного режима.

Сохранить сценарий загрузки в файле проекта. Имя файла произвольное, расширение будет добавлено программой автоматически (<имя1>.ded). **Этот файл необходимо сдавать преподавателю на проверку.**

## 2. OLAP анализ

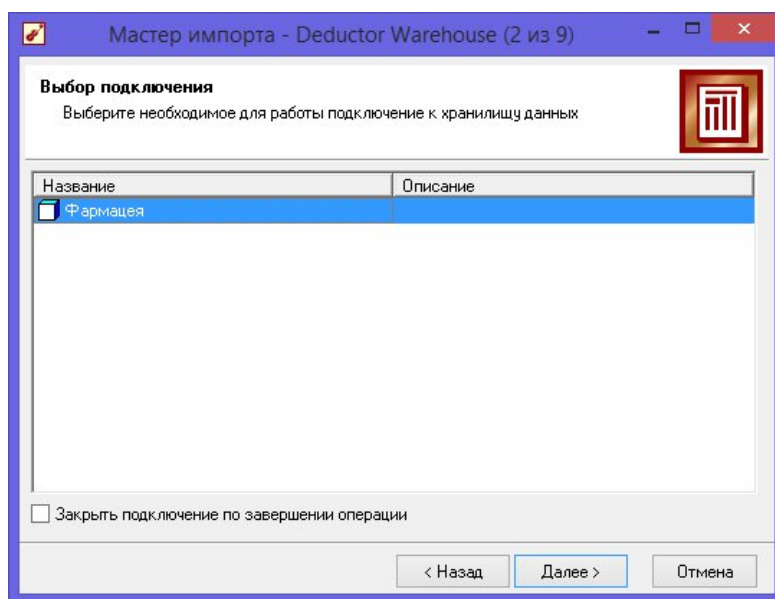
Построим отчет, отражающий динамику сумм продаж по дням недели и времени работы аптек в разрезе групп товаров и аптек.

Создадим новый проект.



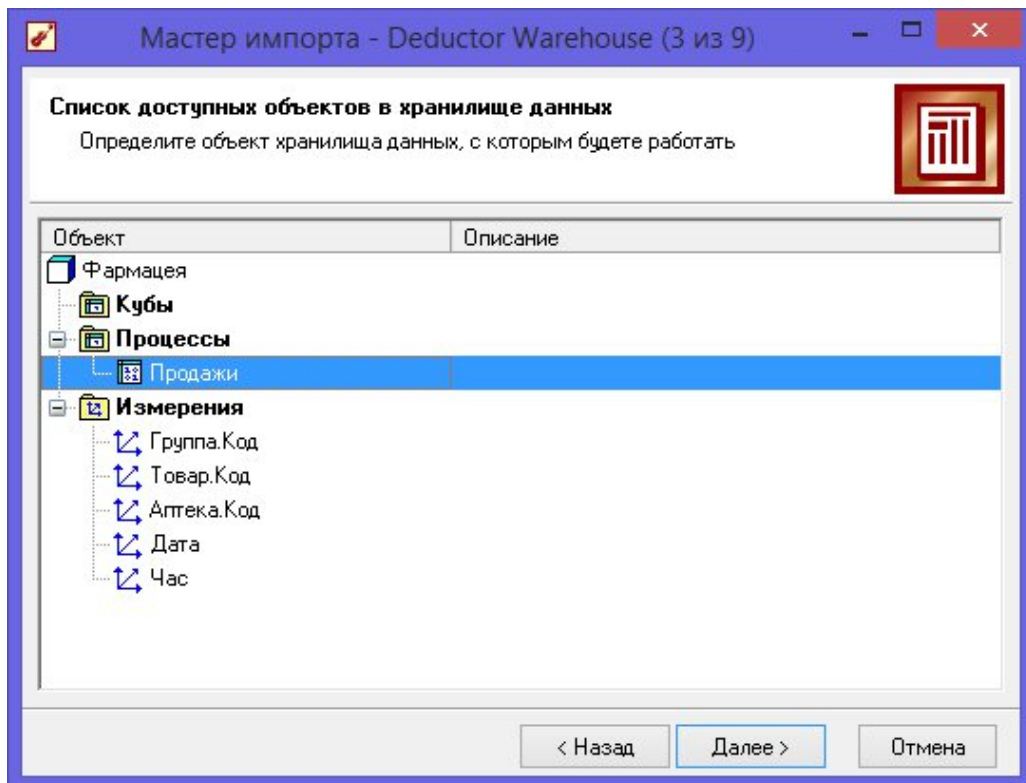
### 2.1. Импорт данных из хранилища (выгрузка)

Для импорта данных из хранилища нажмите на кнопку *Мастер импорта* (📄) и выберите из списка хранилище с именем *Фармацевт*.

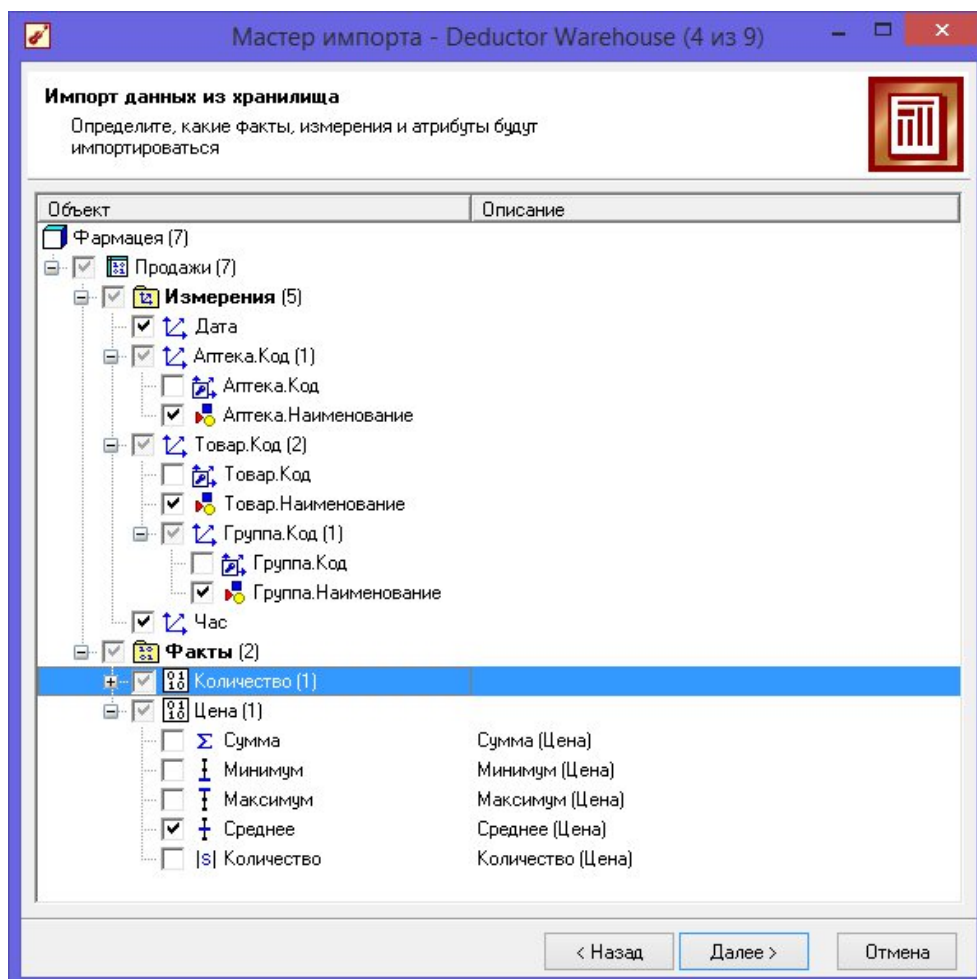


Нажмите *Далее*.

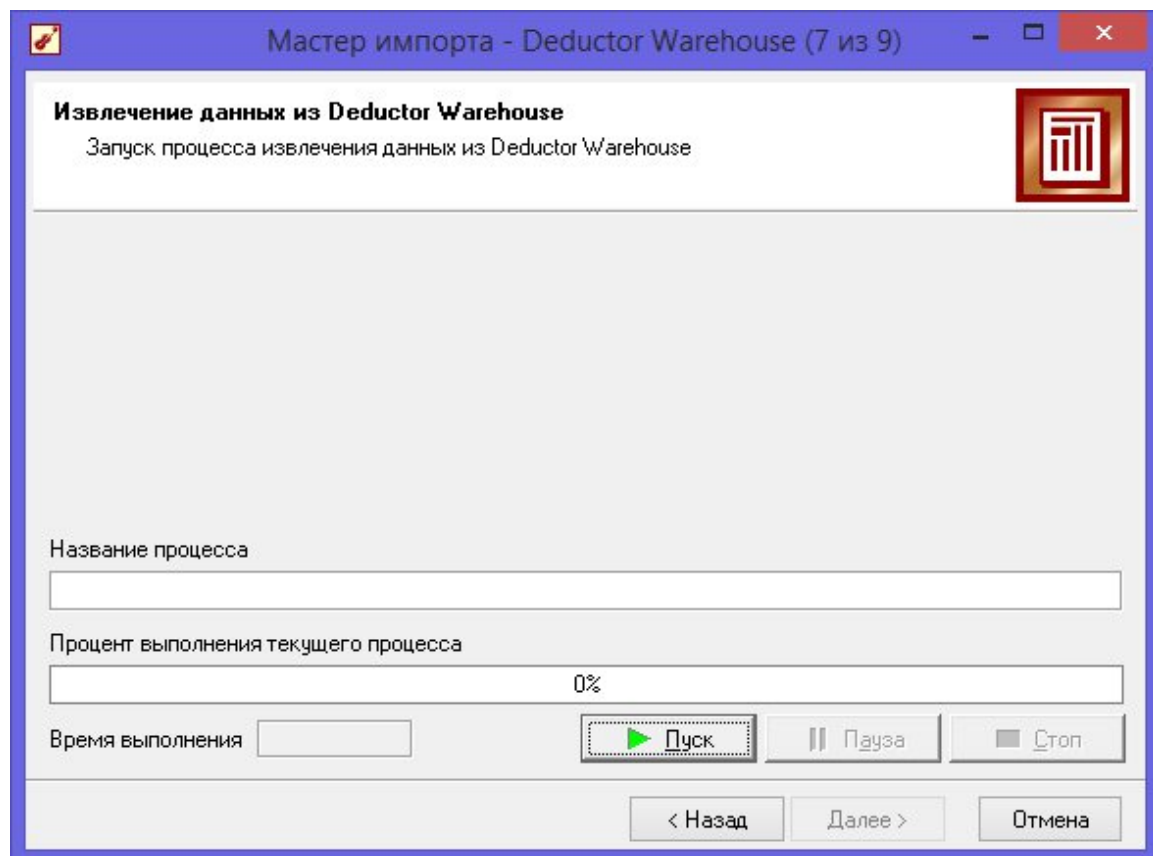
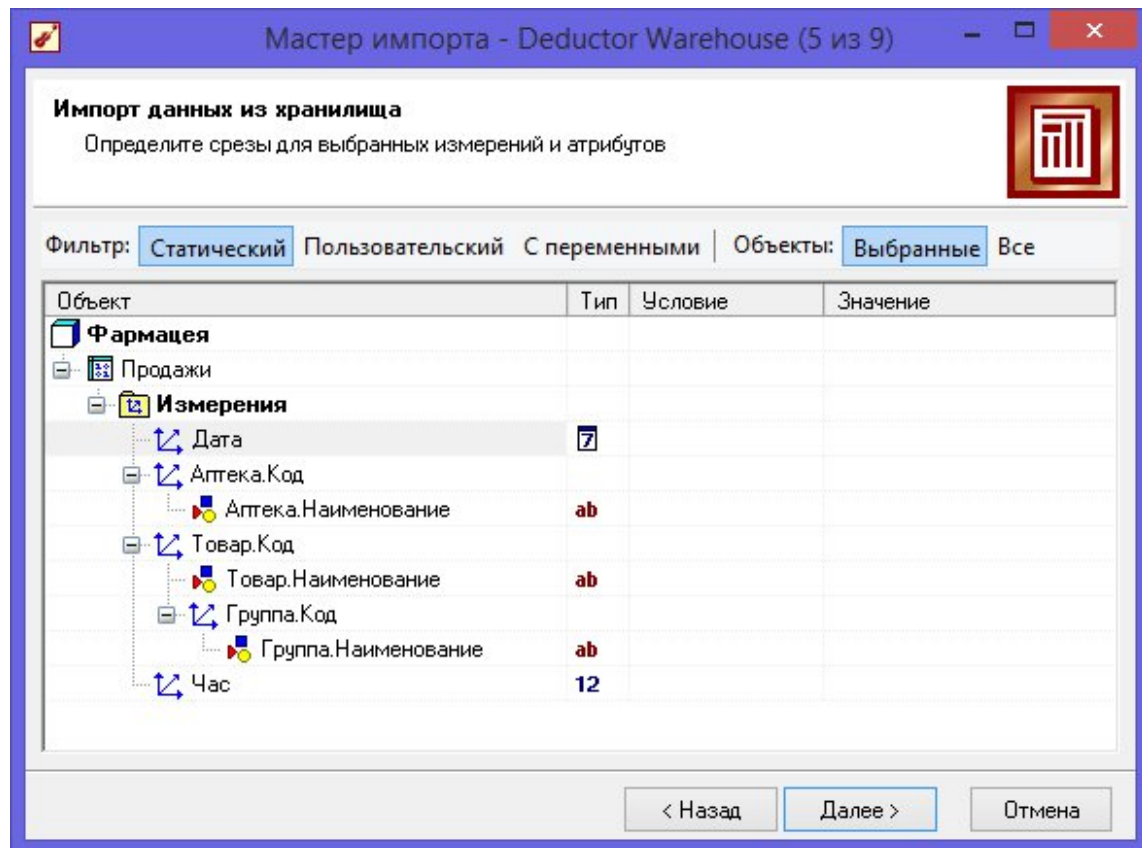
Выберите процесс *Продажи* и нажмите *Далее*.



Включите атрибуты, измерения и факты процесса которые необходимо импортировать, как показано ниже.

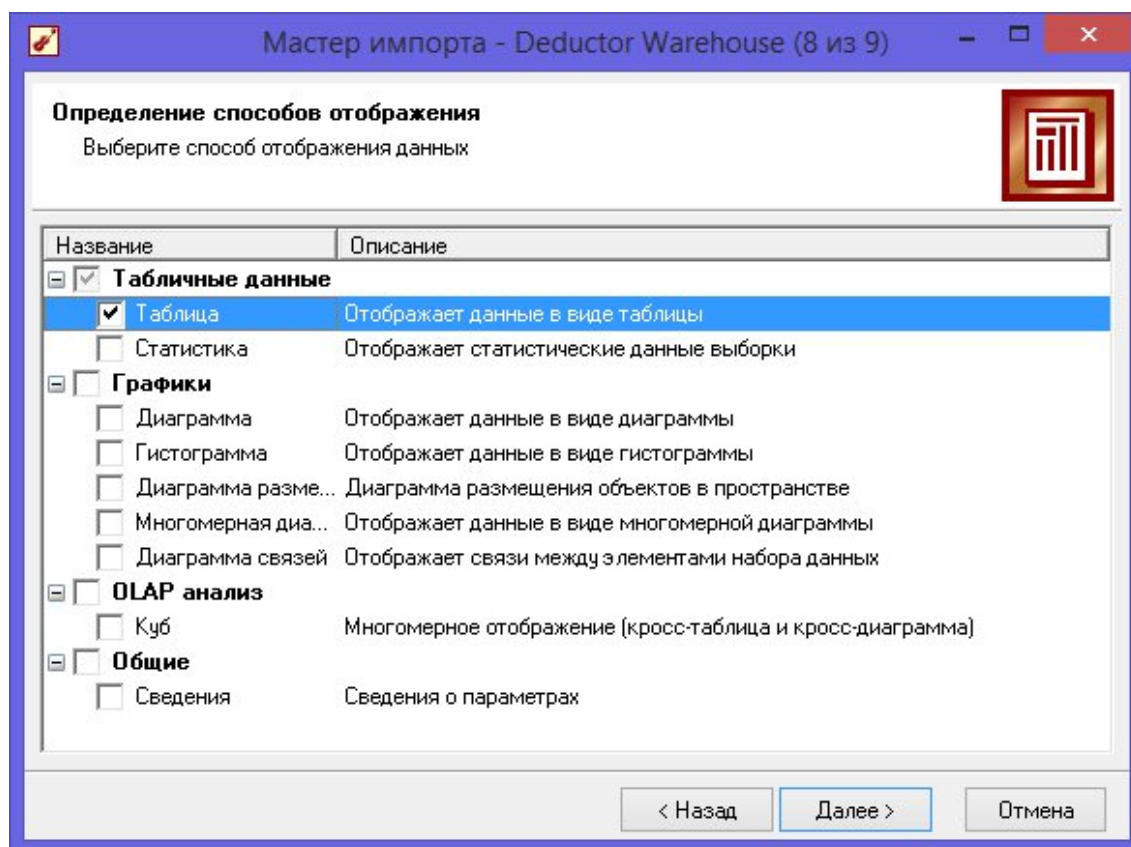


Срезы измерений и атрибутов будем формировать на этапе управления кубом *OLAP*.

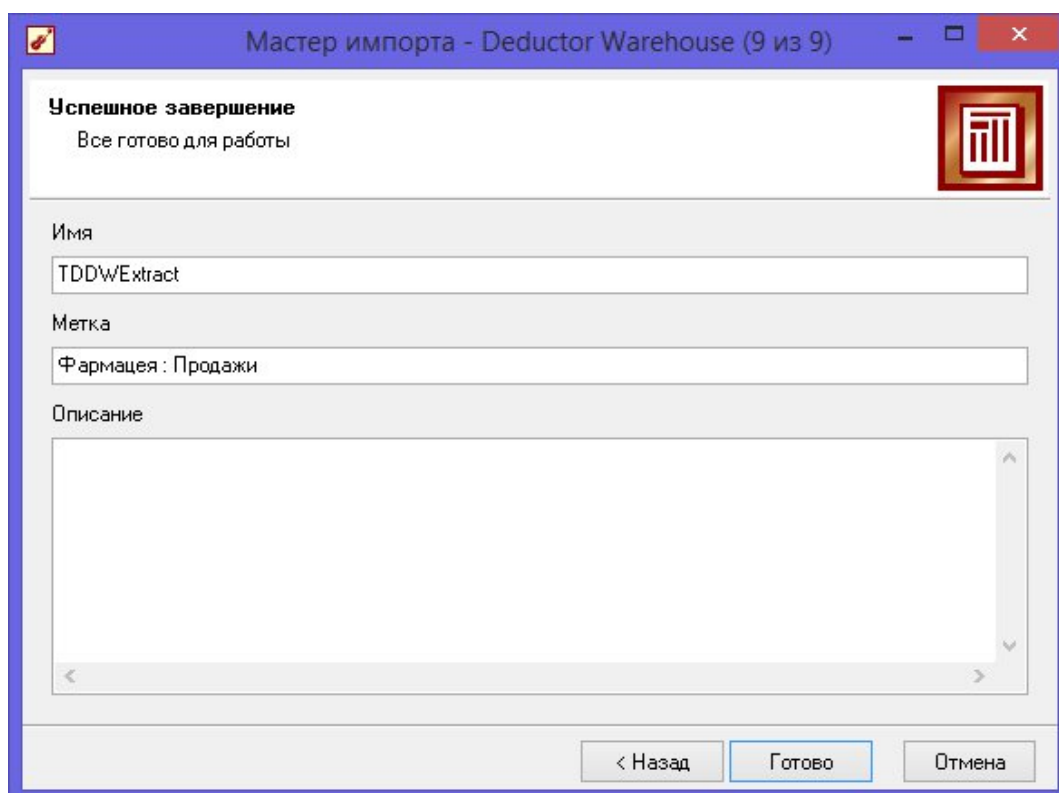


Нажмите на кнопку *Пуск*.

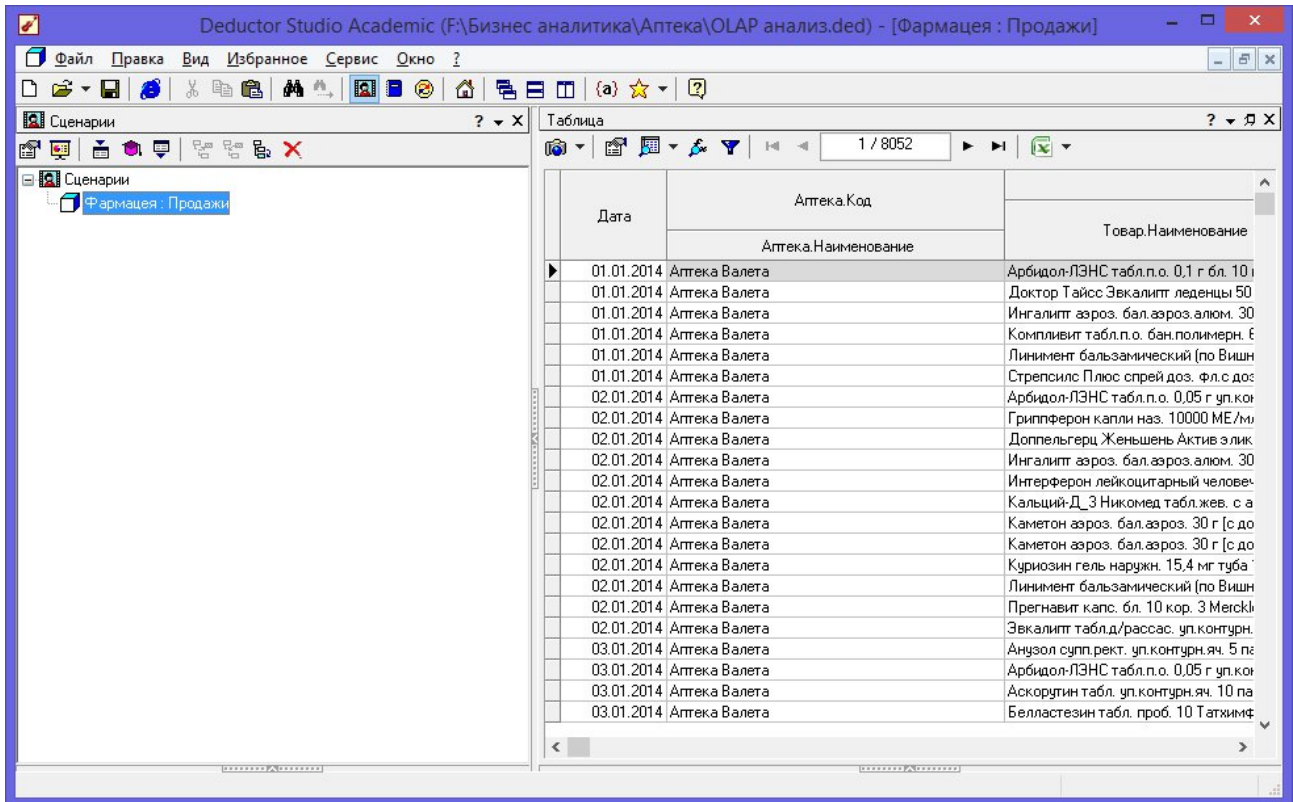
Выберите способ отображения данных *Таблица*.



Нажмите на кнопку *Готово* для завершения импорта данных из процесса *Продажи* хранилища данных.



## Результат импорта данных из процесса *Продажи*.

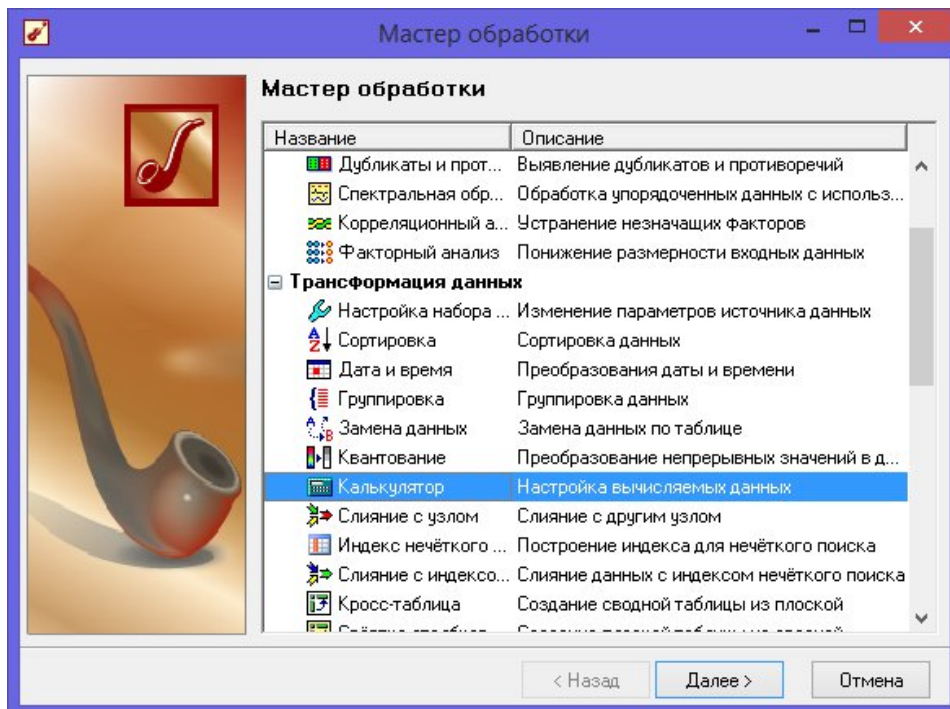


The screenshot shows the Deductor Studio Academic interface. The main window displays a table with the following data:

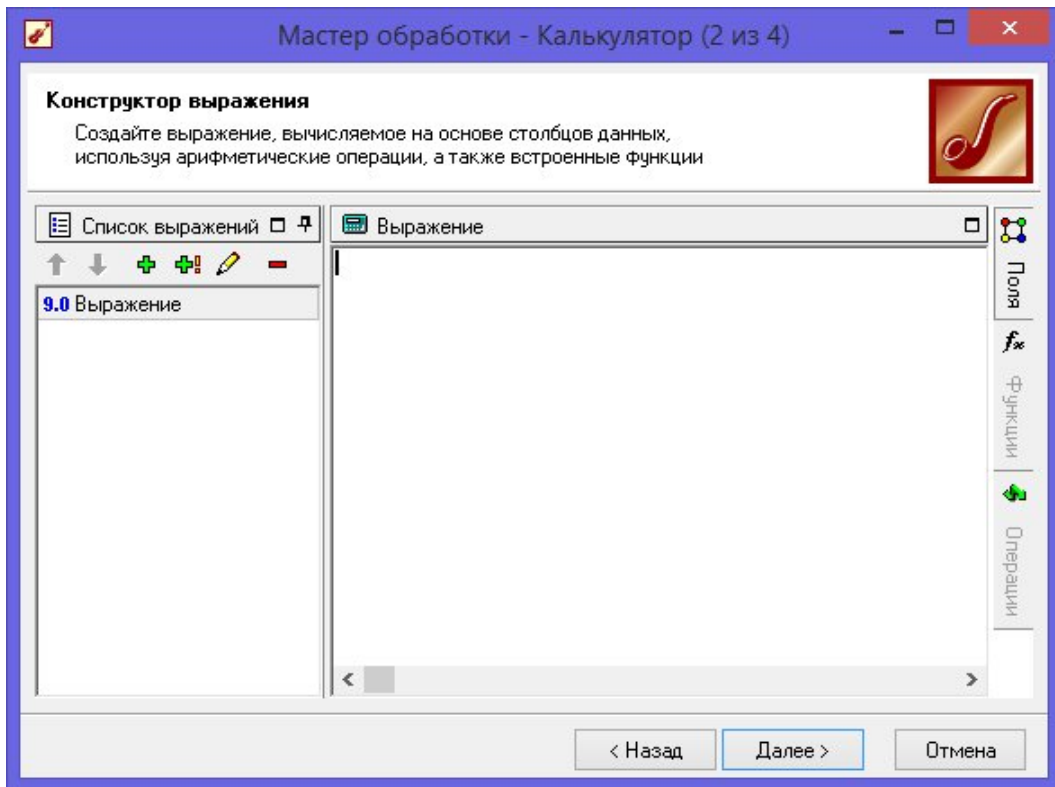
Дата	Аптека.Код	Товар.Наименование
	Аптека.Наименование	
01.01.2014	Аптека Валета	Арбидол-ЛЭНС табл.п.о. 0,1 г бл. 10
01.01.2014	Аптека Валета	Доктор Тайсс Эвкалипт леденцы 50
01.01.2014	Аптека Валета	Ингалипт аэроз. бал.аэроз.алюм. 30
01.01.2014	Аптека Валета	Компливит табл.п.о. бал.полимерн. 6
01.01.2014	Аптека Валета	Линимент бальзамический (по Вишн
01.01.2014	Аптека Валета	Стрепсилс Плюс спрей доз. фл.с доз
02.01.2014	Аптека Валета	Арбидол-ЛЭНС табл.п.о. 0,05 г уп.ко
02.01.2014	Аптека Валета	Гриппферон капли наз. 10000 МЕ/мл
02.01.2014	Аптека Валета	Допельгерц Женьшень Актив элик
02.01.2014	Аптека Валета	Ингалипт аэроз. бал.аэроз.алюм. 30
02.01.2014	Аптека Валета	Интерферон лейкоцитарный челове
02.01.2014	Аптека Валета	Кальций_Д_3 Никомед табл.жев. с а
02.01.2014	Аптека Валета	Каметон аэроз. бал.аэроз. 30 г [с до
02.01.2014	Аптека Валета	Каметон аэроз. бал.аэроз. 30 г [с до
02.01.2014	Аптека Валета	Куриозин гель наружн. 15,4 мг туба
02.01.2014	Аптека Валета	Линимент бальзамический (по Вишн
02.01.2014	Аптека Валета	Прегнавит капсул. бл. 10 кор. 3 Merck
02.01.2014	Аптека Валета	Эвкалипт табл./рассас. уп.контурн.
03.01.2014	Аптека Валета	Анузол супп.рект. уп.контурн.яч. 5 п
03.01.2014	Аптека Валета	Арбидол-ЛЭНС табл.п.о. 0,05 г уп.ко
03.01.2014	Аптека Валета	Аскорутин табл. уп.контурн.яч. 10 па
03.01.2014	Аптека Валета	Белластезин табл. проб. 10 Татхимф

## 2.2. Подготовка данных для аналитической отчетности.

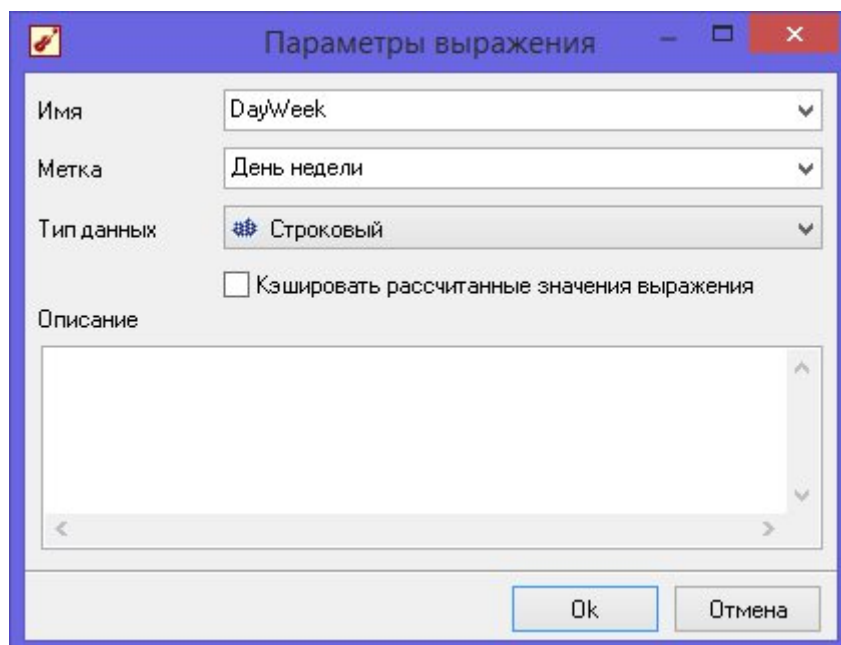
Подготовка разреза – дни недели. Выделите из поля *Дата* день недели. Для этого нажмите на кнопку *Мастер обработки* (🔧). В разделе *Трансформация данных* выберите опцию *Калькулятор*.



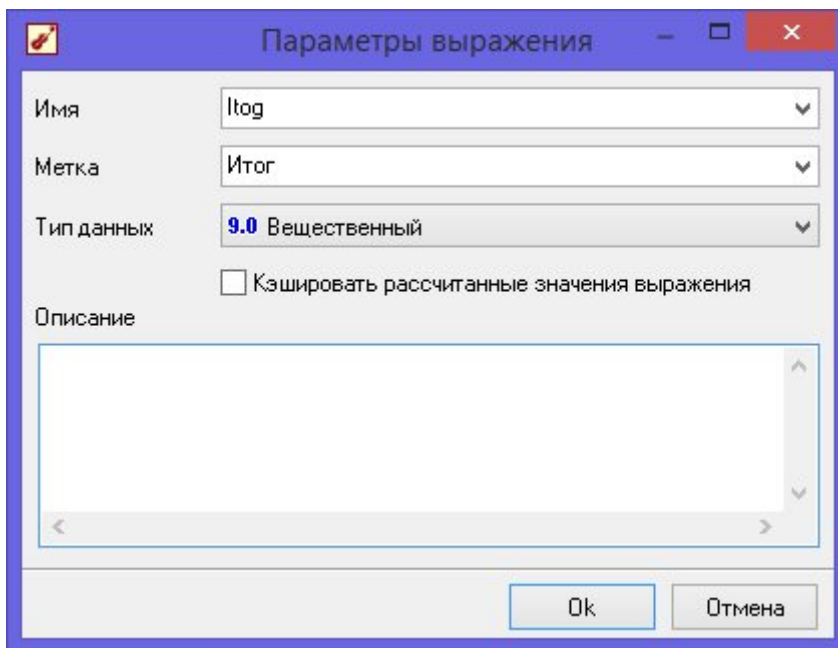
В окне *Конструктор выражения* нажмите на кнопку *Редактировать параметры выражения* (✎).



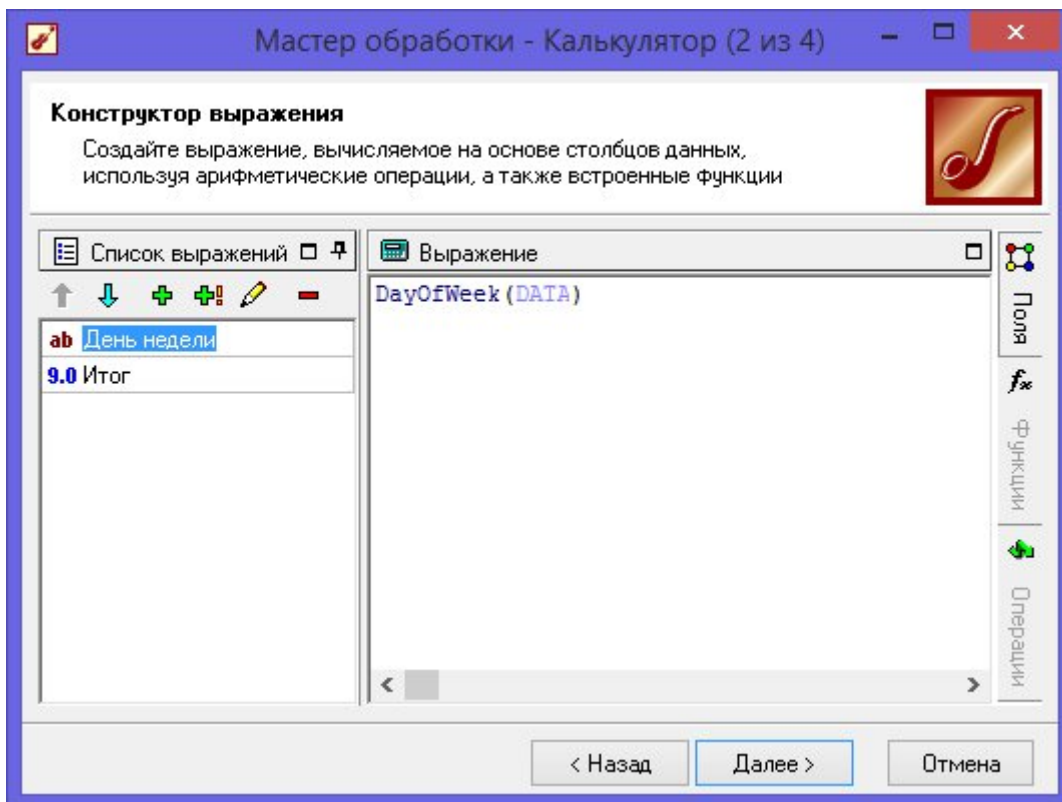
Задайте имя и метку (*День недели*). Нажмите *ОК*.



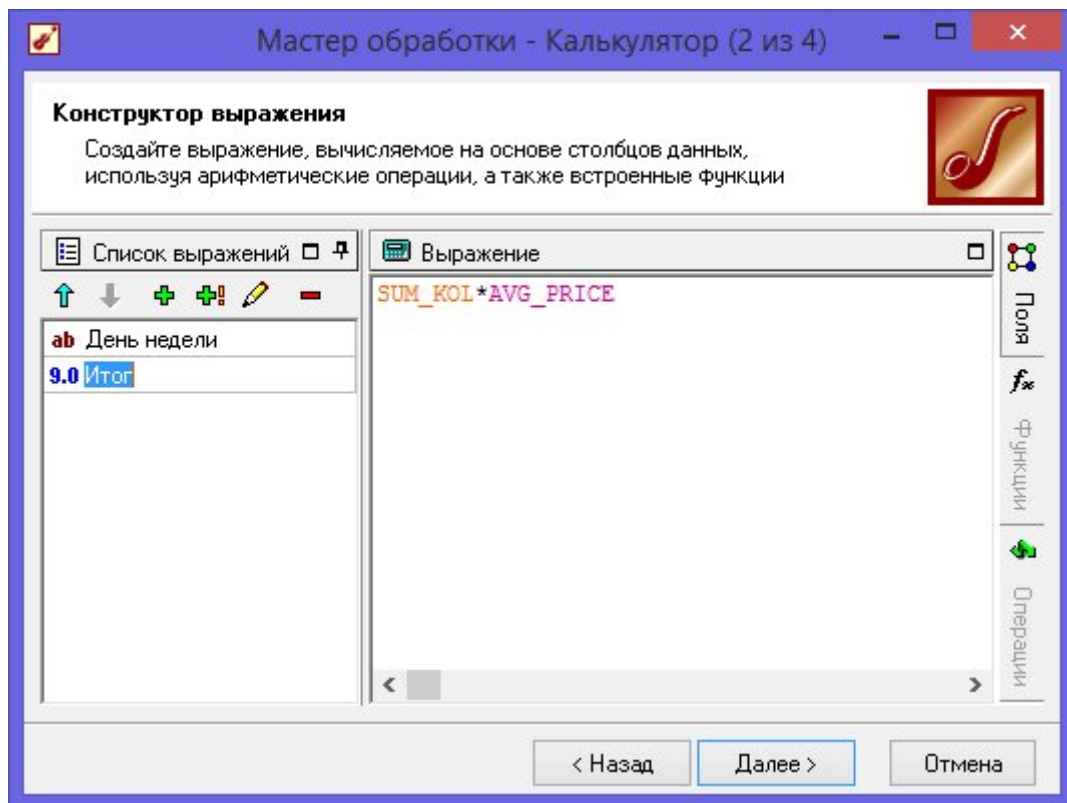
Задайте второе выражение. Для этого нажмите на кнопку *Добавить выражение* (+).  
Задайте имя и метку (*Итог*). Нажмите *ОК*.



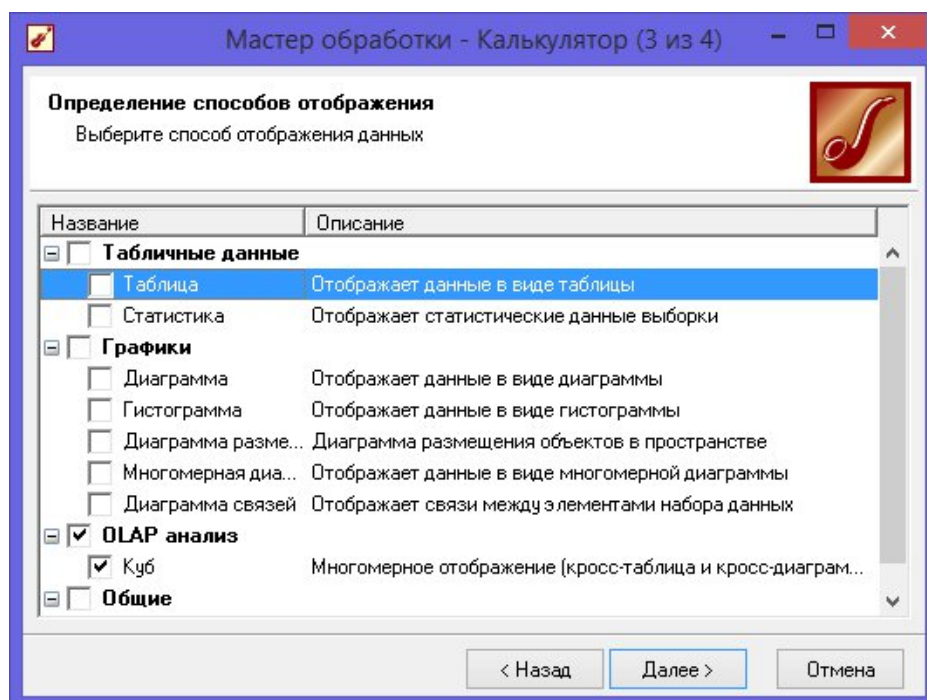
В окне *Конструктор выражения* перейдите на первое выражения и задайте для него выражение, используя вкладки *Функции* и *Поля* (справа).



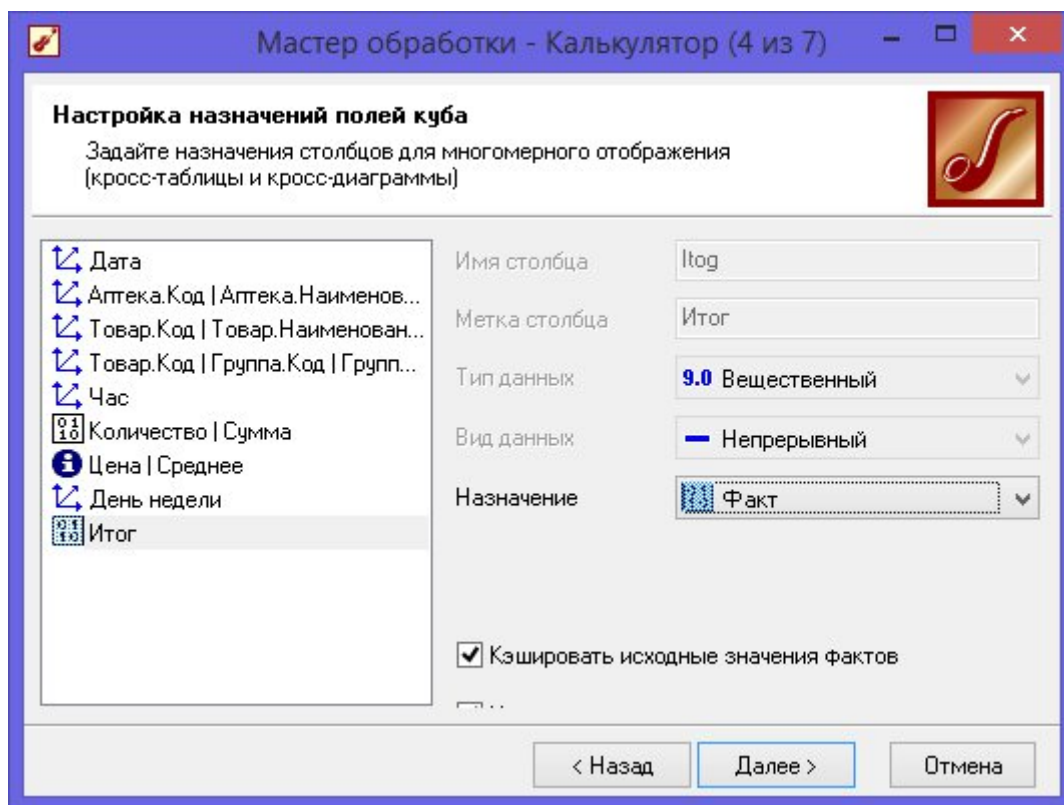
Перейдите на второе выражение и задайте произведение  $SUM\_KOL * AVG\_PRICE$  используя вкладку *Поля*.



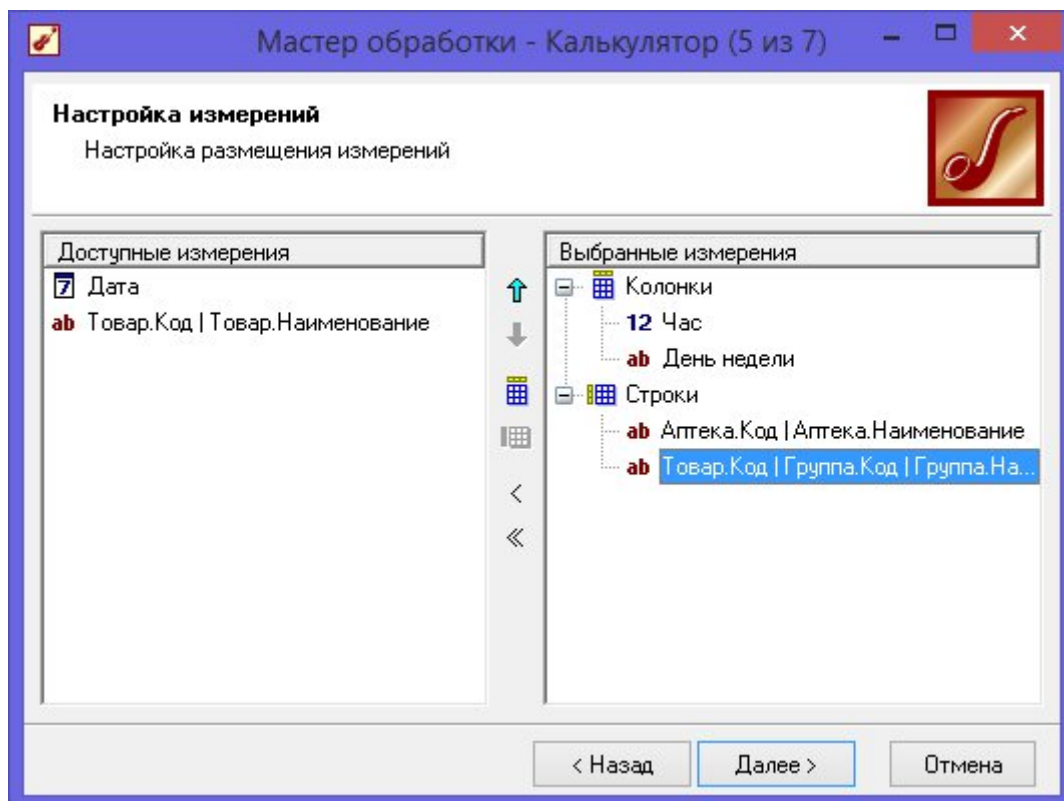
Выберите способ отображения данных *Куб*.



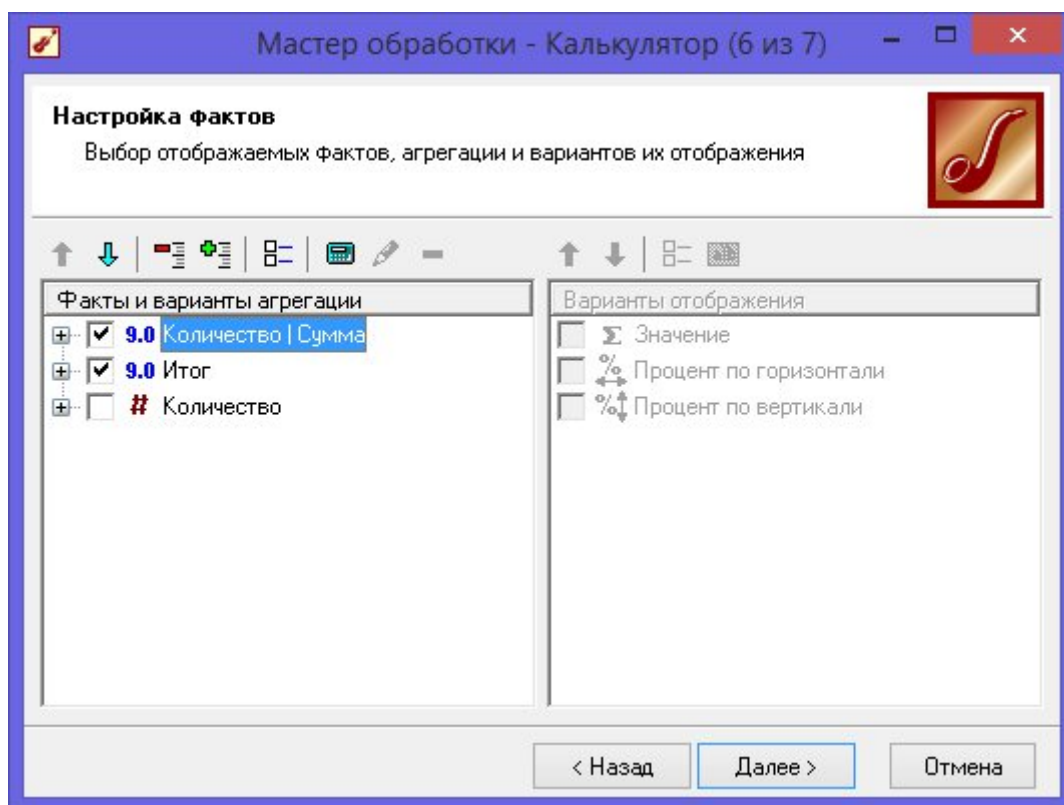
Задайте назначения столбцов для многомерного отображения (устанавливается в строке *Назначение*).



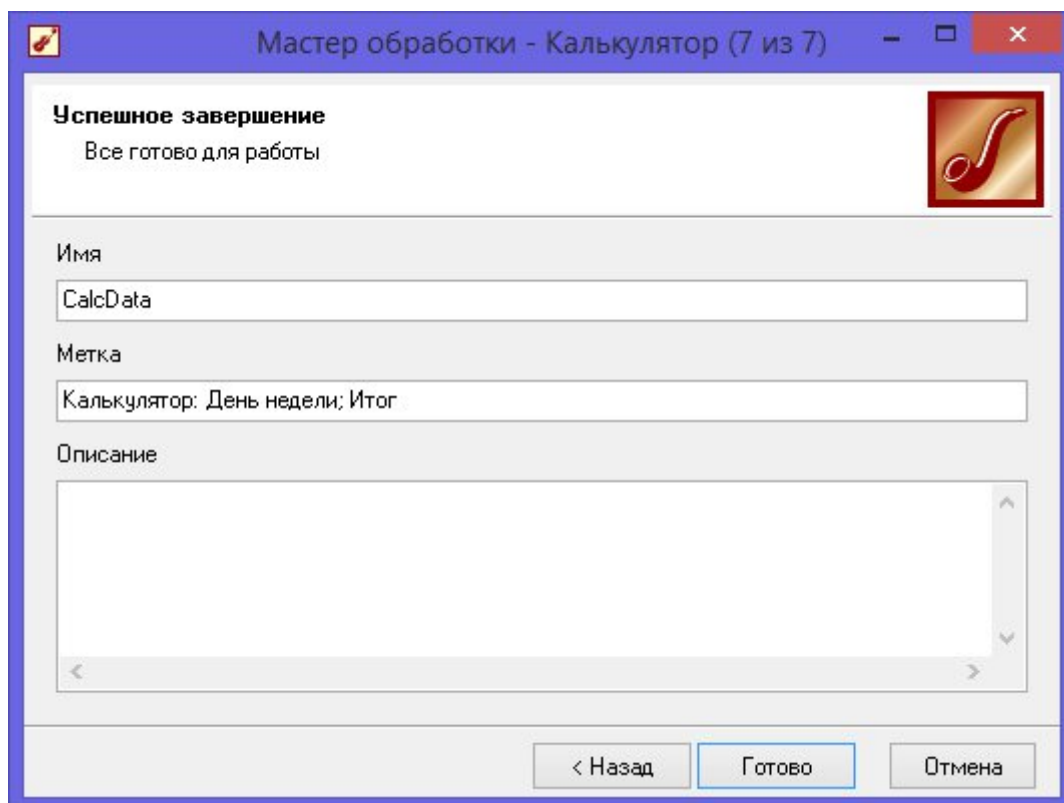
Выполните настройку размещения измерений по строкам\столбцам как показано ниже.



Подключите факты *Количество* и *Итог* для отображения в *Кубе*.



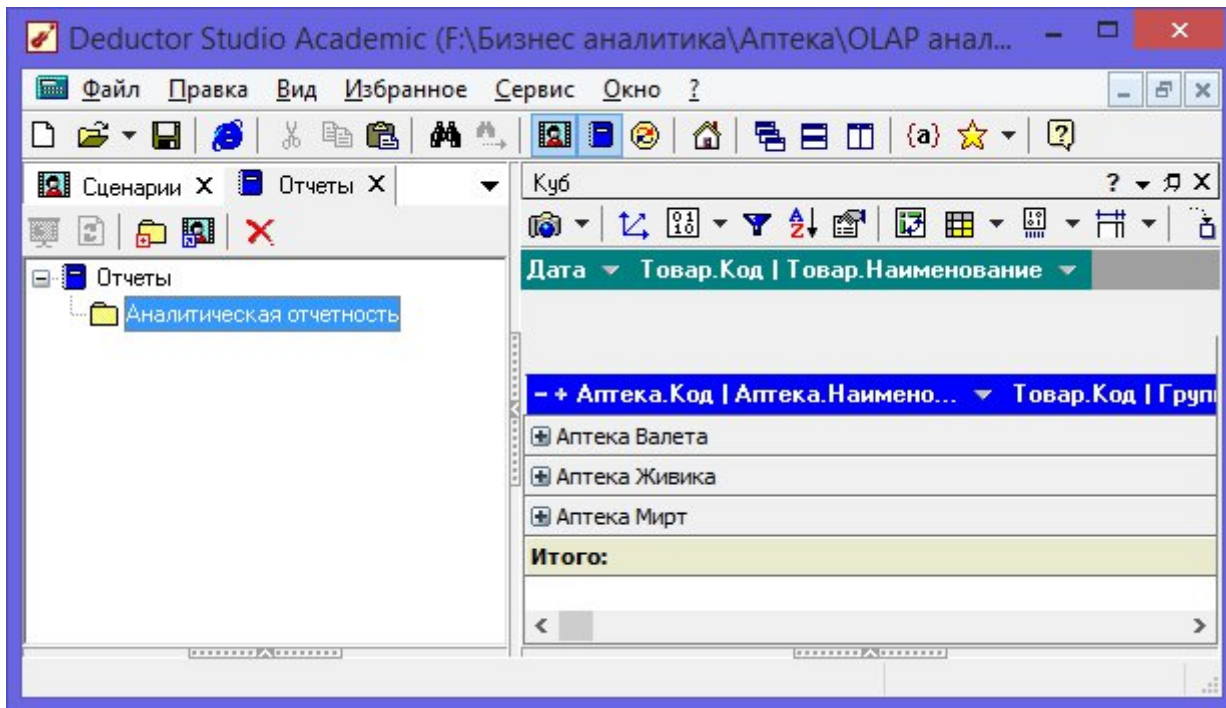
Нажмите на кнопку *Готово* для завершения построения *Куба-OLAP*.



### 2.3. Аналитическая отчетность.

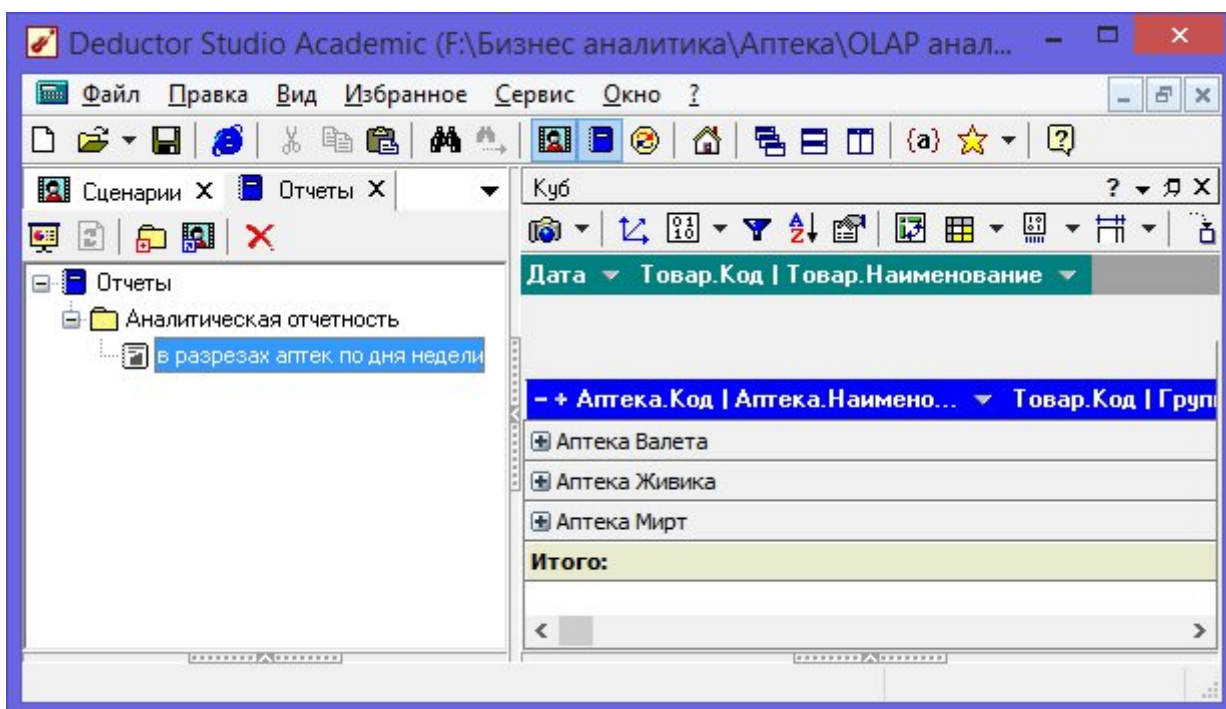
Переключитесь на вкладку *Отчеты* в меню *Вид*.

Создайте папку *Аналитическая отчетность*. Для этого нажмите на кнопку *Добавить папку* (📁). Переименуйте папку.

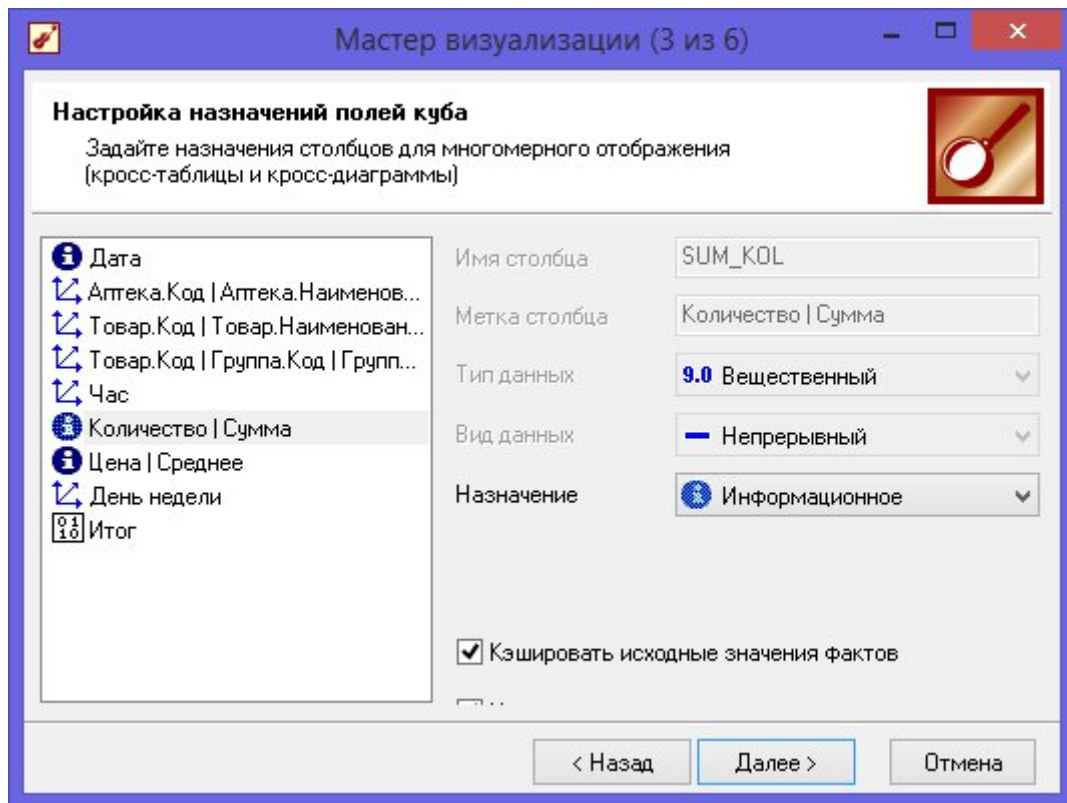


Нажмите на кнопку *Добавить узел* (📁). Укажите на узел сценария *Калькулятор: День недели; Итог* и нажмите *Выбрать*.

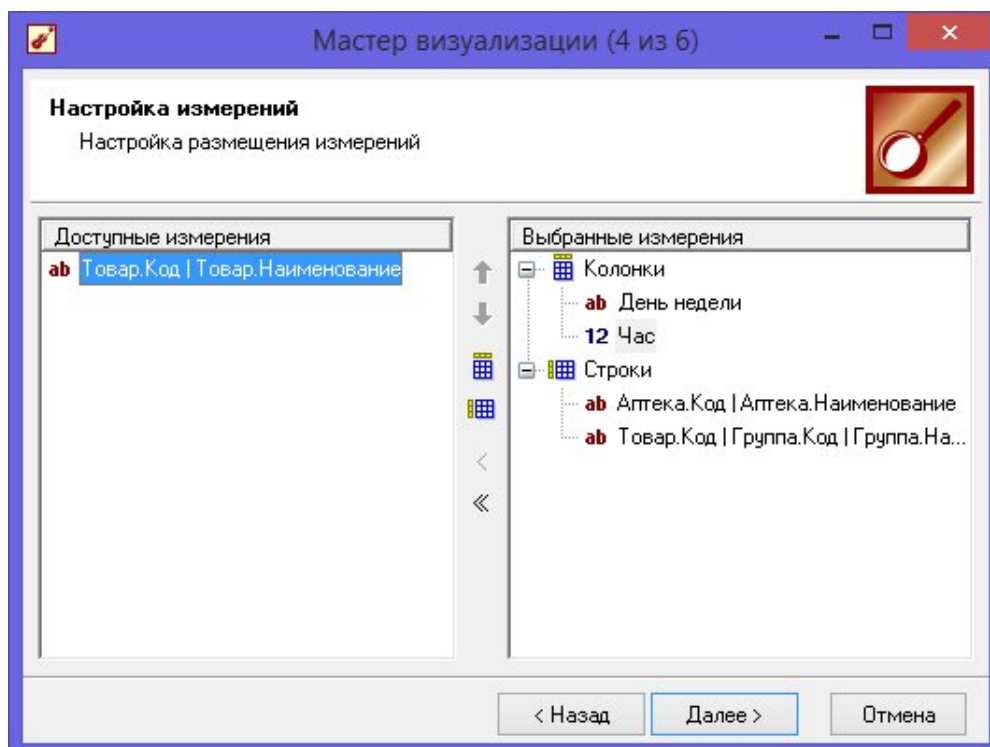
Получим узел отчетности. Переименуйте его: *в разрезах аптек по дням недели*.



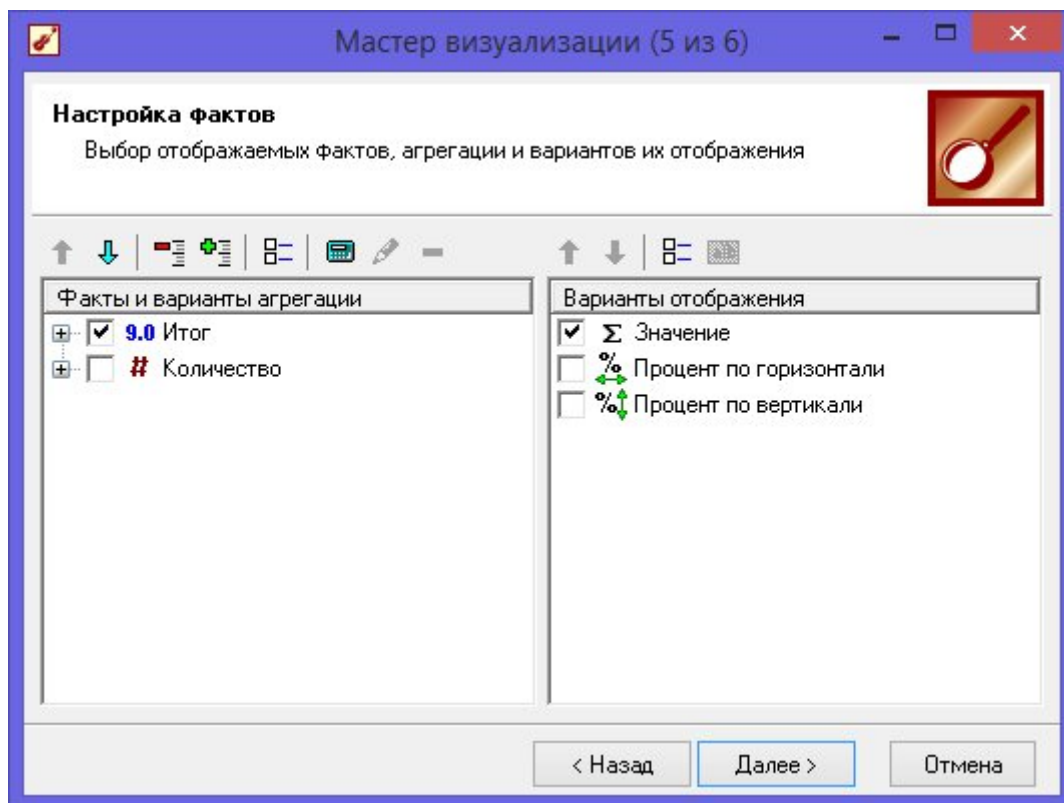
Настройте куб для отображения нужного среза. Для этого нажмите на кнопку *Мастер визуализации* (🔍). Задайте назначение столбцов для многомерного отображения.



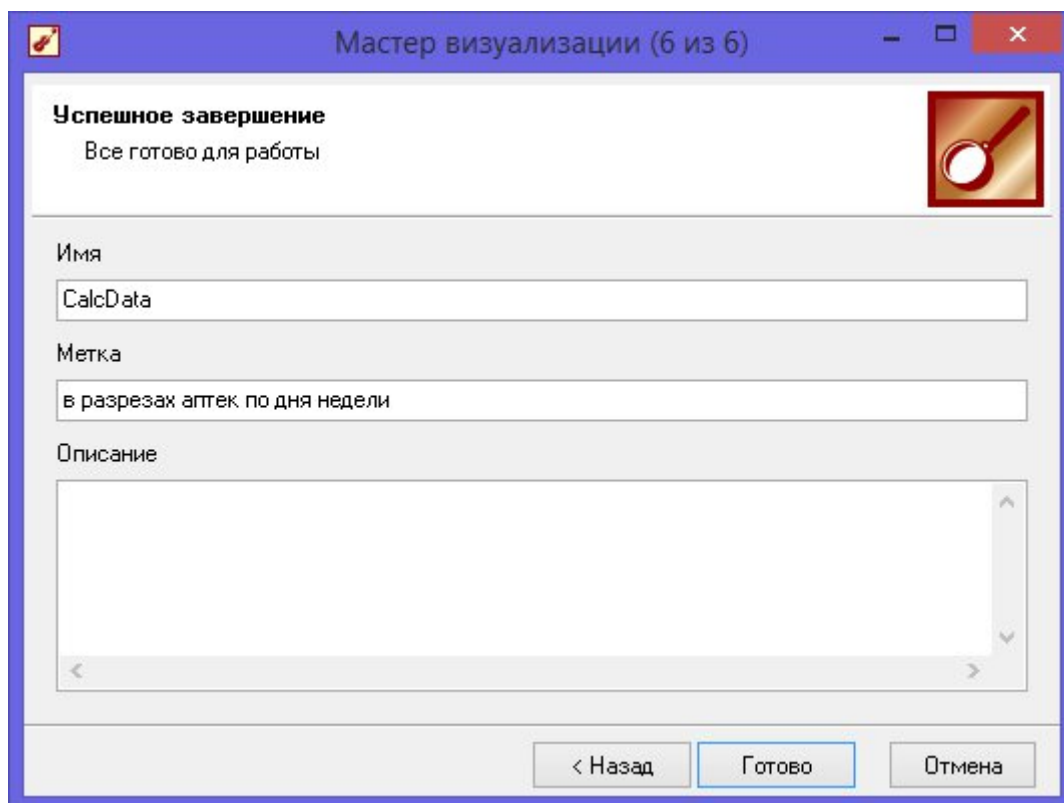
Выполните настройку размещения измерений как показано ниже.





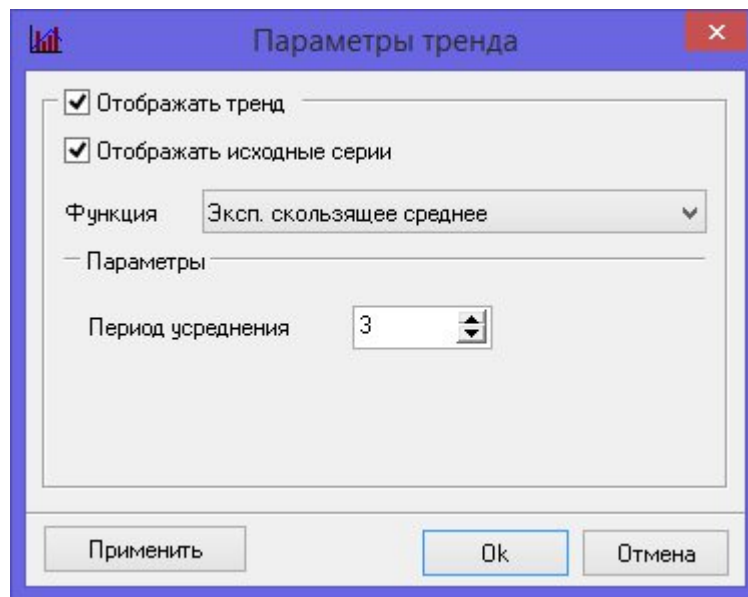
Включите флажок *Итог*.



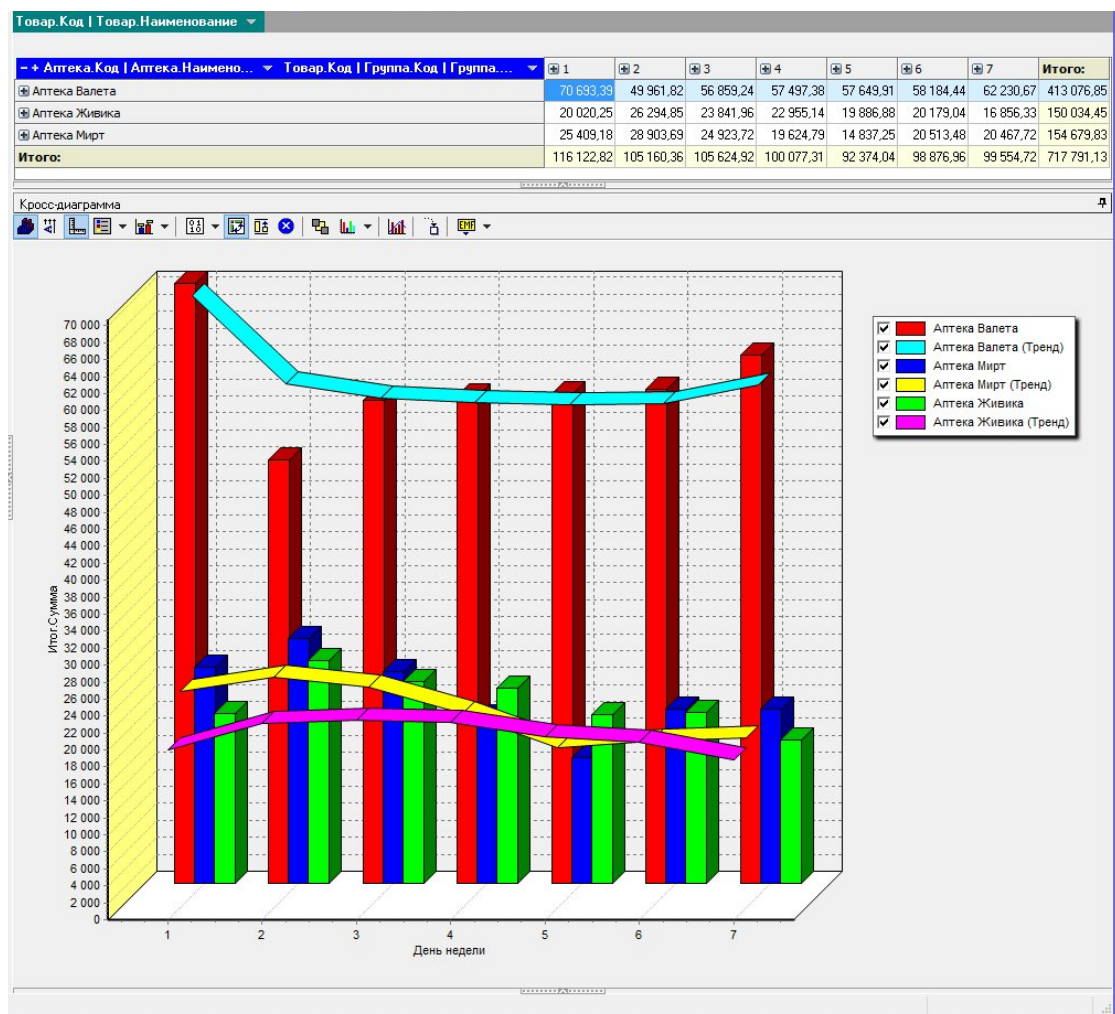
Нажмите на кнопку *Готово*.




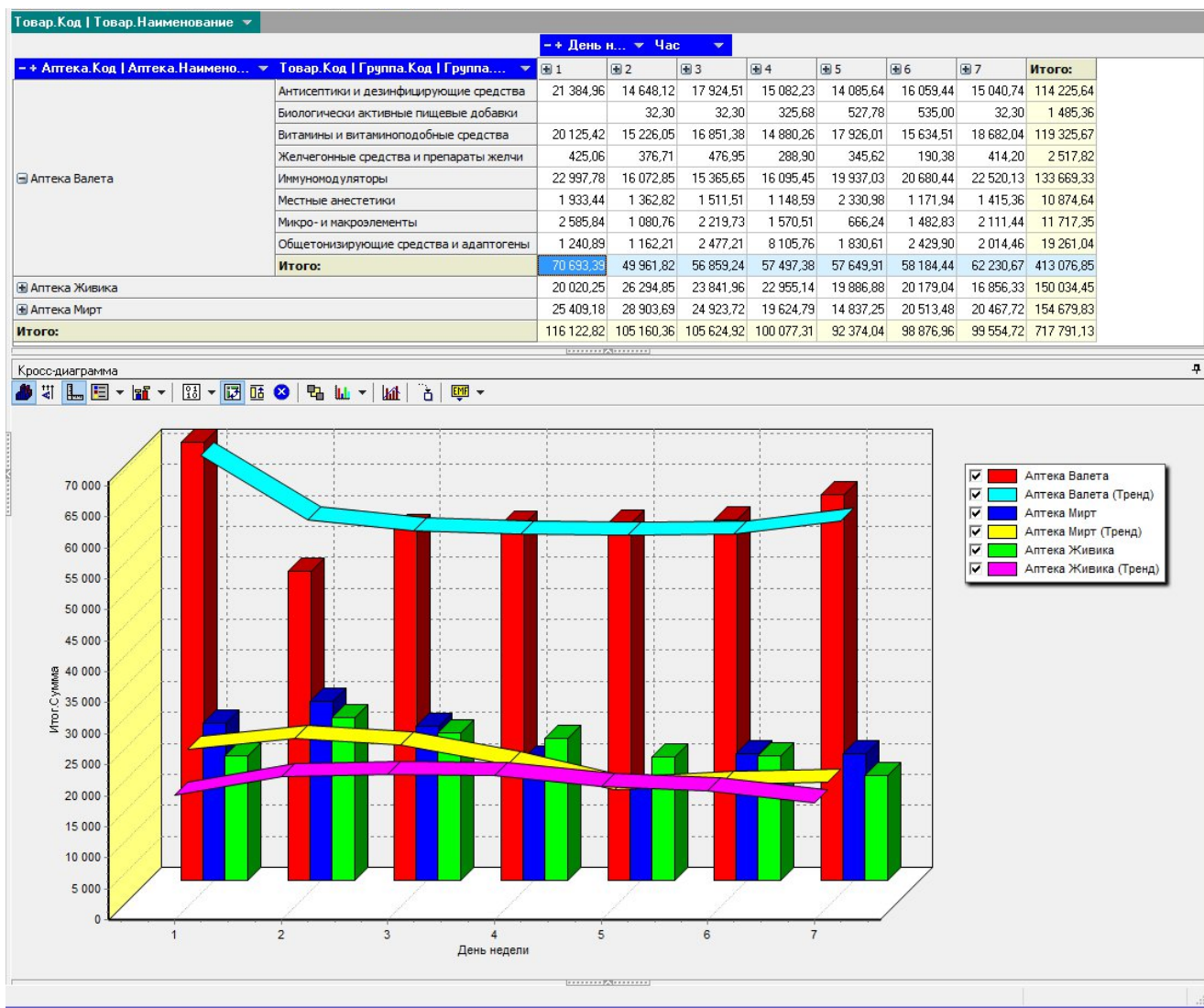
На инструментальной панели нажмите на кнопку *Показать/Скрыть кросс-диаграмму* (  ). Разместите диаграмму снизу к кубу-OLAP. Нажмите на кнопку *Параметры тренда* (  ). Установите параметры тренда как показано ниже.



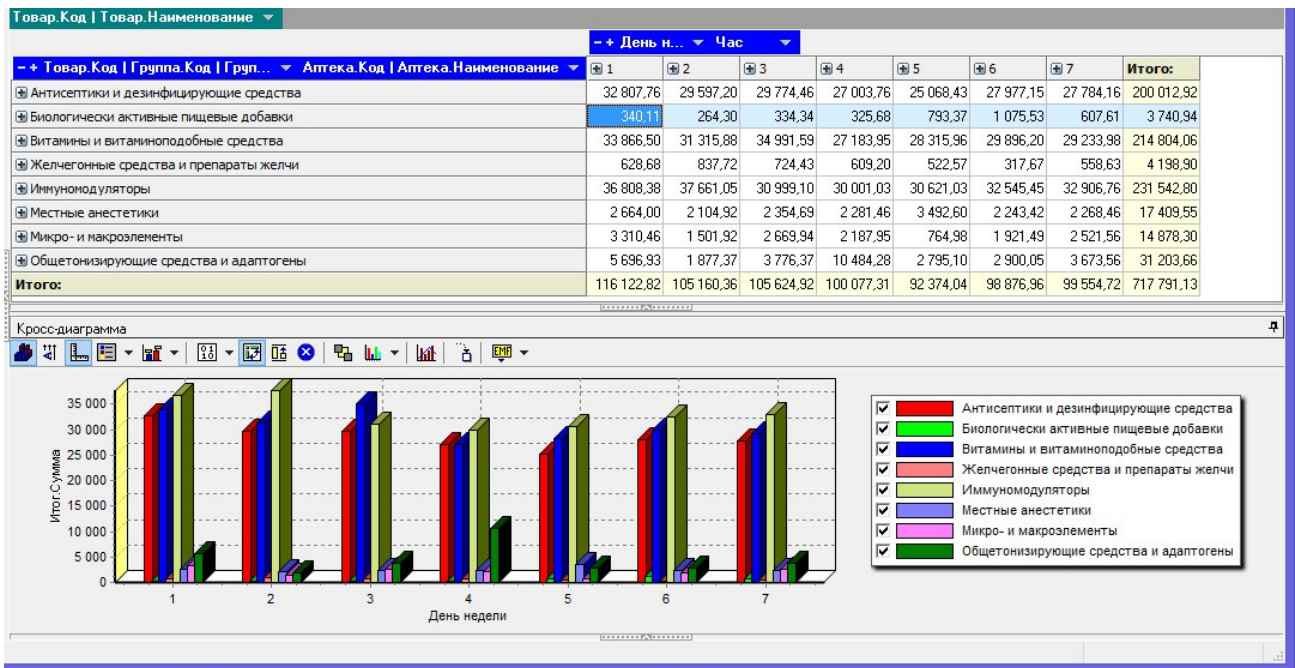
Результат аналитической отчетности по выбранному срезу.



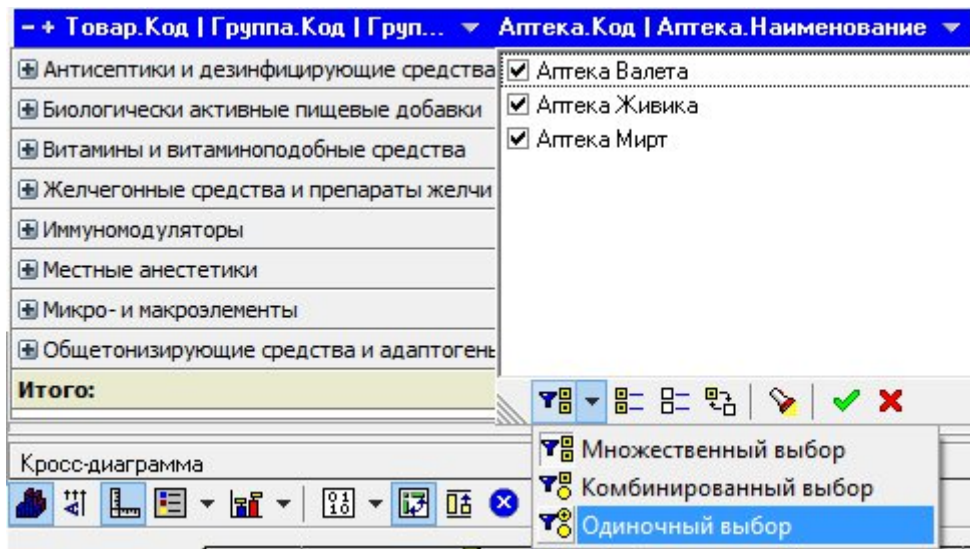
Раскройте поле *Товар.Код* для аптеки Валета (нажмите на кнопку ). В кубе отобразятся итоги продаж по каждой товарной группе за каждый день недели.



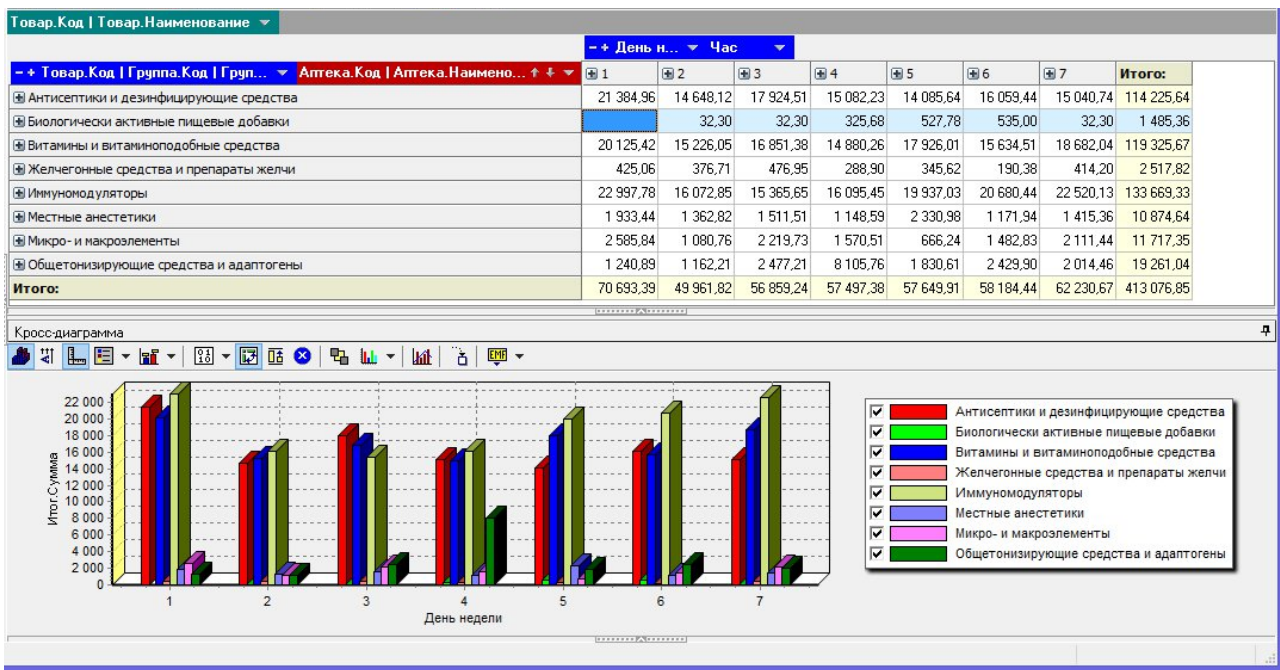
Для отображения итогов продаж по товарным группам по всем аптекам перетащите поле *Товар.Код* товара левее поля *Аптека.Код*



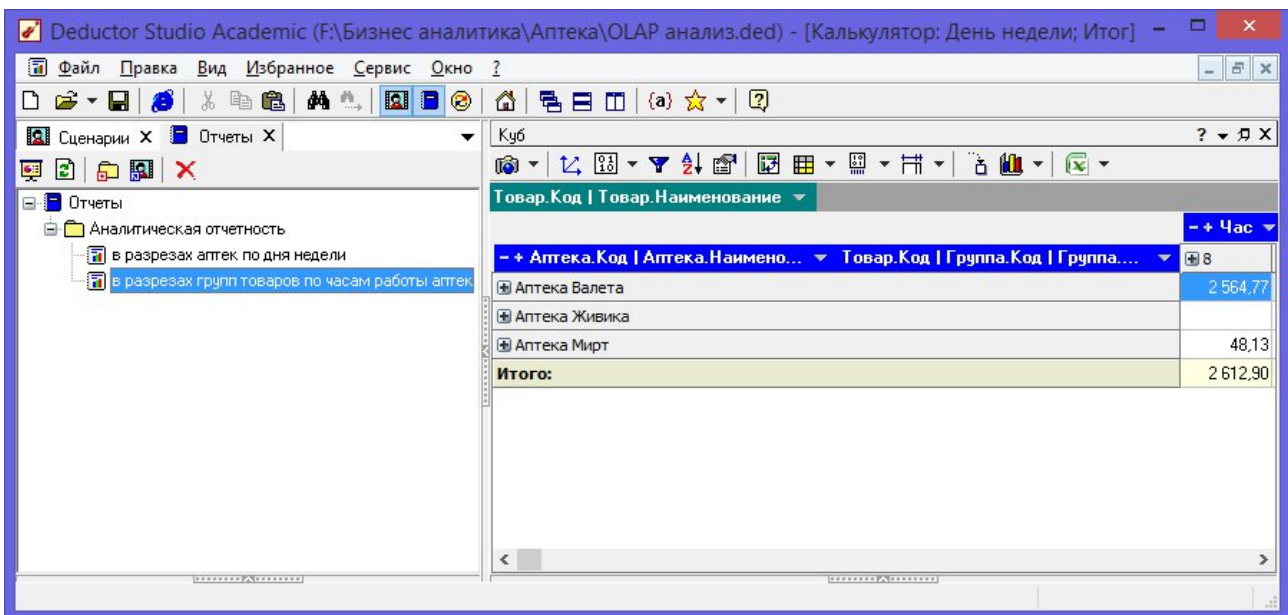
Для отображения итогов продаж по группам товаров для каждой аптеки в отдельности установите фильтр для поля *Аптека.Код*, т.е. режим одиночный выбор.



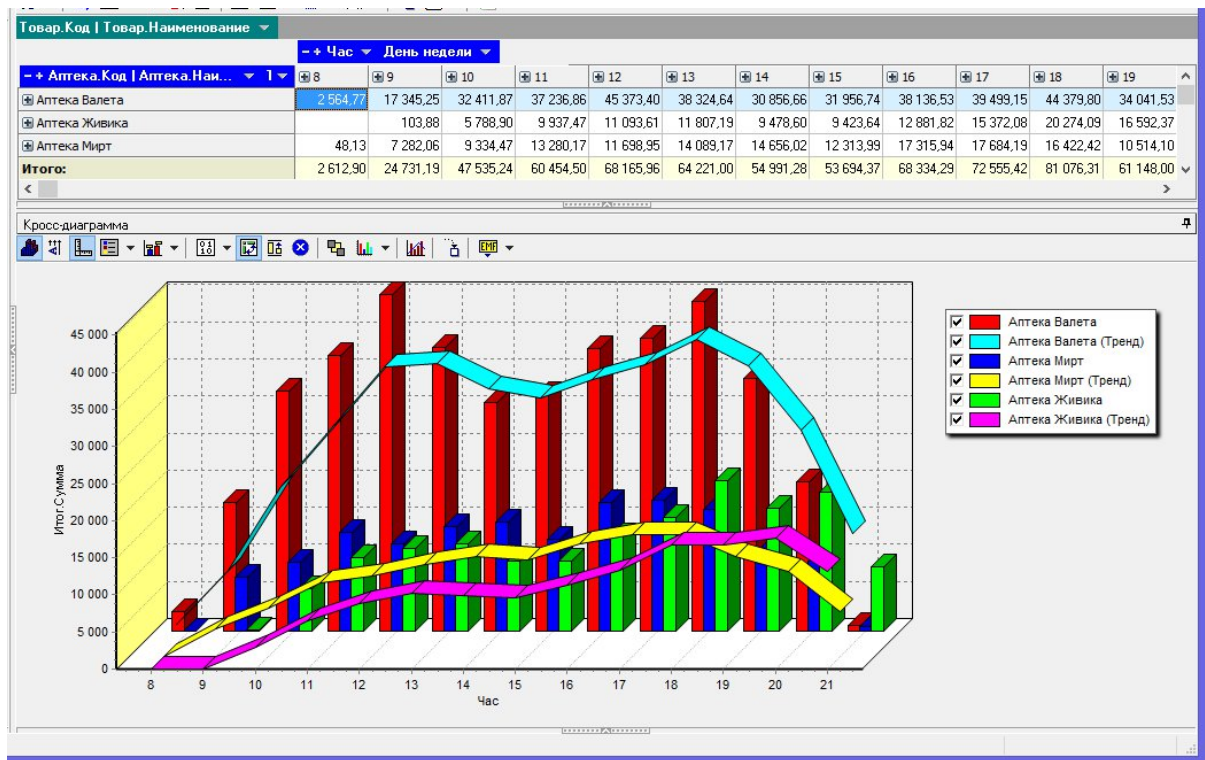
В поле *Аптека.Код* на красном фоне появятся стрелки сканирования. Можно последовательно просматривать итоги продаж по каждой аптеке.



**Задание.** *Создайте аналитическую отчетность по итогам продаж в разрезе почасовой работы.*

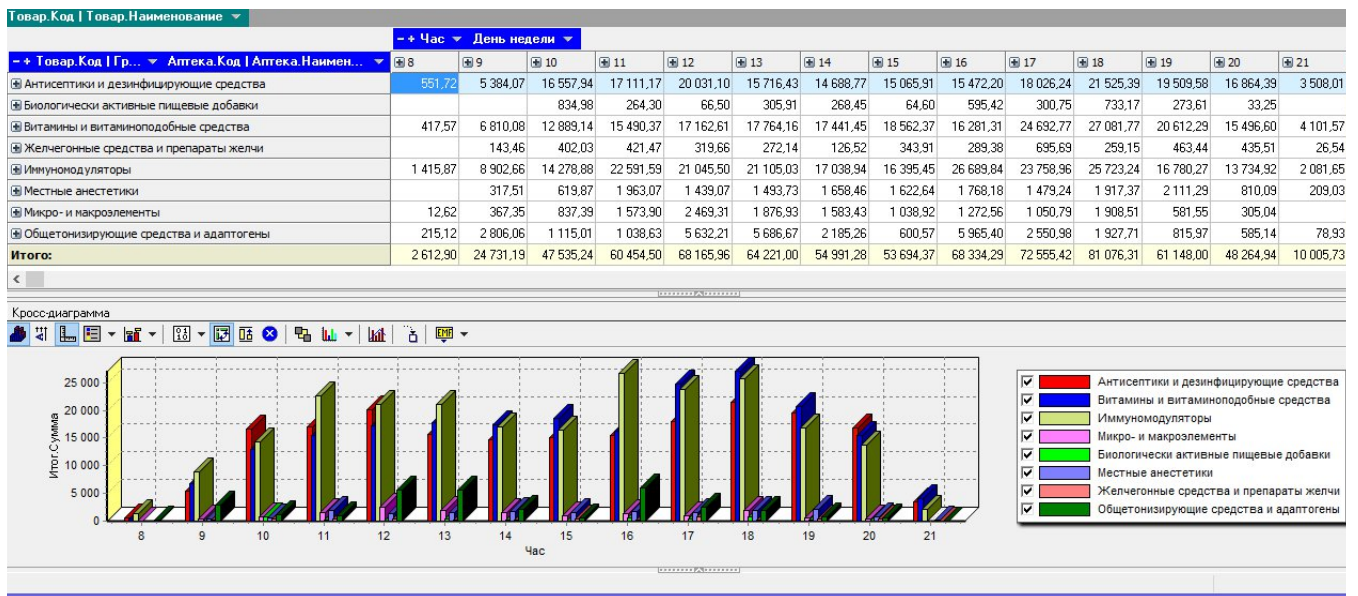


Настройки куба-OLAP в разрезах аптек по часам работы.



Утренние и вечерние часы малоприбыльные.

Перестройте куб. Отобразите итоги продаж по товарным группам для каждой аптеки в отдельности по часам работы.



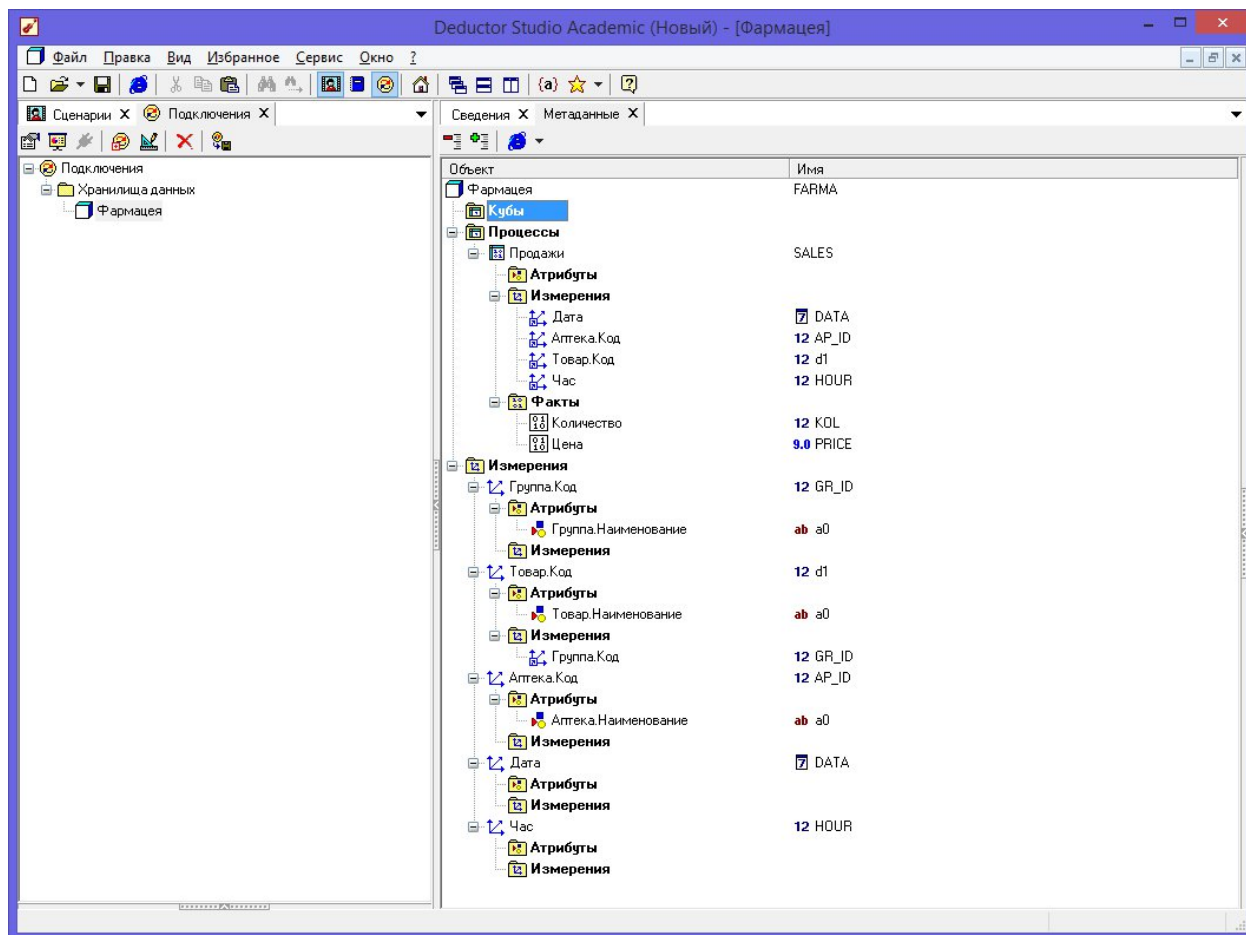
На основании проведенного исследования сделайте рекомендации по оптимизации работы аптечной сети.

В завершение сохранить проект. Имя файла проекта произвольное. Файл проекта имеет расширение: <имя 2>.ded. Этот файл необходимо сдавать преподавателю на проверку.

## Пример выполнения задания

### Вариант 00

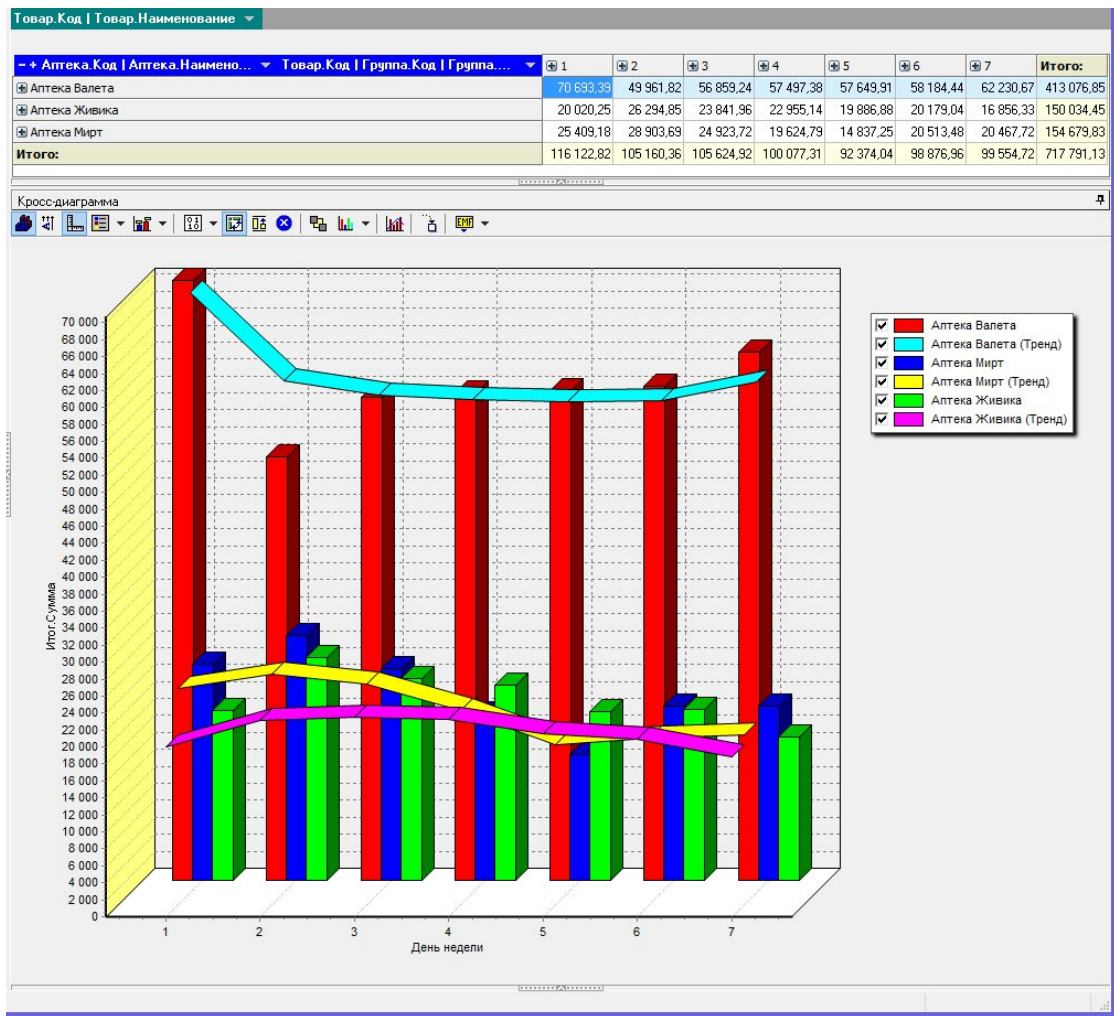
#### 1. Создать Хранилище Данных (ХД).



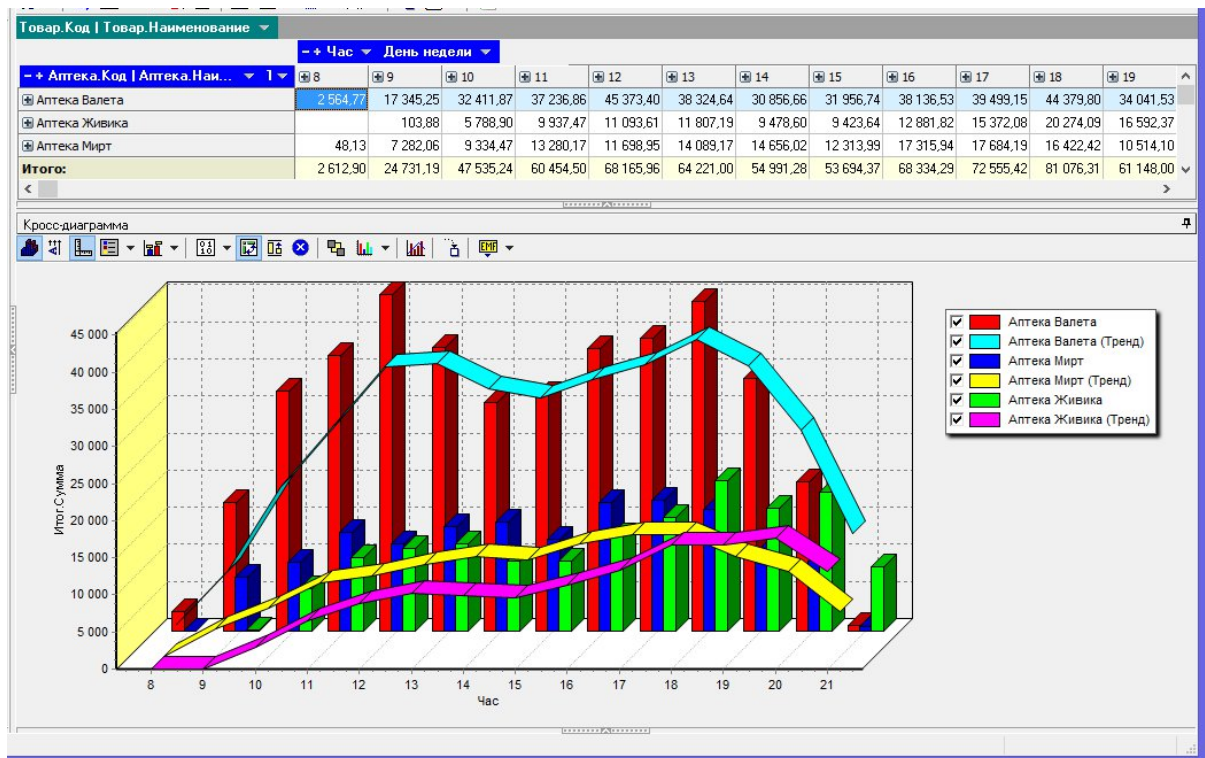
#### 2. OLAP анализ.

Создать отчетности по итогам продаж в разрезе аптек:

- по дням недели,



- почасовой работы.



На основе проведенного исследования сделайте рекомендации по оптимизации работы аптечной сети.

Анализ диаграмм продаж лекарственных средств в сети аптек показывает:

- в каких-то аптеках спад продаж выпадает на середину недели, в каких-то на начало и конец недели;

- утренние и вечерние часы малоприбыльные.

Рекомендации: рассмотреть вопрос о предоставлении скидок различным категориям граждан.

## ЛИТЕРАТУРА

1. Акофф Р. Искусство решения проблем: пер. с англ.- М.: Мир, 1982.- 224 с.
2. Андерсен Бьёрн. Бизнес-процессы. Инструменты совершенствования: пер. с англ. - М.: РИА «Стандарты и качество», 2003.- 272 с
3. Бариленко В.И. Основы бизнес-анализа : учебное пособие. — М. : КНОРУС, 2014. — 272 с.
4. Бьер Майк. Интеллектуальное ведение сопровождения бизнеса: пер. с англ.- М.: КУДИЦ-Образ, 2005. – 240 с.
5. Елиферов В.Г., Репин В.В. Бизнес-процессы: Регламентация и управление: Учебник. - М.: ИНФРА-М, 2005. - 319 с
6. Компания BaseGroup Labs. *Deductor. Руководство аналитика. Версия 5.3.* б.м. : BaseGroup Labs, 2013.
7. Конрад Карлберг. Бизнес-анализ с использованием Excel. Решение бизнес-задач, 4-е издание - М.: «Вильямс», 2013. – 576 с.
8. Майк Ротер, Джон Шук. Учись видеть бизнес-процессы. Практика построения карт потоков создания ценности: пер. с англ., 4-е изд. - Альпина Паблишер; Москва, 2015. – 144 с.
9. Паклин Н.Б. Орешков В.И. Бизнес-аналитика: от данных к знаниям, - Санкт-Петербург : Питер, 2013.
10. Перерва А.Д., Иванова В.А. Путь аналитика. Практическое руководство IT-специалиста. – СПб.: Питер, 2012. – 304 с.
11. Расиел И. Инструменты McKinsey : Лучшая практика решения бизнес-проблем: пер. с англ. Ю.Е. Корнилович. — 2-е изд. — М. : Манн, Иванов и Фербер, 2009. — 224 с.
12. Тим Кларк, Александр Остервальдер, Ив Пинье. Твоя бизнес-модель: Системный подход к построению карьеры. – М.: Альпина Паблишер, 2013. – 260 с.
13. Фрэнкс Билл. Укрощение больших данных. Как извлекать знания из массивов информации с помощью глубокой аналитики: пер. с англ.- М.: Манн, Иванов и Фербер, 2014.