

## ***Deductor Studio часть 1***

**Цель работы:** освоение базовых процедур создания и, наполнения данных хранилища Deductor Warehouse, выработка практических умений и навыков проектирования и создания хранилища данных в Deductor Studio; освоение технологии OLAP и способов представления данных с использованием многомерных кубов. Изучение инструментария Deductor по многомерному анализу данных, освоение умений и навыков получения данных из хранилища Deductor Warehouse (срезы из хранилища и OLAP-кубы).

### **Теоретическая часть**

#### **Общие сведения о Deductor Studio**

Deductor является аналитической платформой, предназначенной для создания законченных прикладных решений в области анализа данных. Реализованные в Deductor технологии позволяют на базе единой архитектуры пройти все этапы построения аналитической системы от консолидации данных до автоматического подбора моделей и визуализации полученных результатов.

Разработчиками представляются три типа аналитической платформы Deductor: Enterprise; Professional; Academic.

Версия Enterprise предназначена для корпоративного использования и имеет расширенный потенциал.

Версия Professional предназначена для небольших компаний и однопользовательской работы. В ней отсутствуют серверные компоненты, поддержка OLE, виртуальное хранилище, а традиционное хранилище данных можно создавать только на СУБД FireBird. Автоматизация выполнения сценариев обработки данных осуществляется только через пакетный режим.

Версия Deductor Academic предназначена только для образовательных целей. В ней ограничены возможности интеграции и автоматической обработки данных.

#### **Структура аналитической платформы Deductor**

Аналитическая платформа Deductor состоит из пяти частей:

- Warehouse – хранилище данных, консолидирующее информацию из разных источников;
- Studio – приложение, позволяющее пройти все этапы построения прикладного решения, рабочее место аналитика;
- Viewer – рабочее место конечного пользователя, одно из средств тиражирования знаний (т. е. когда построенные аналитиком модели используют пользователи, не владеющие технологиями анализа данных);
- Server – служба, обеспечивающая удаленную аналитическую обработку данных;
- Client – это динамическая библиотека для удаленной работы с Deductor Server. Обеспечивает доступ к серверу из сторонних приложений и управление его работой.

Вся работа по анализу данных в Deductor Studio сводится всего к 4 операциям:

- импорт данных;
- обработка данных;
- визуализация;
- экспорт данных.

Перечисленные выше действия реализуются с помощью четырех мастеров: "Мастера импорта", "Мастера обработки", "Мастера визуализации" и "Мастера экспорта".

### **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.

Deductor Warehouse реализует универсальное многомерное хранение, т.е. может содержать множество процессов с различным количеством измерений и фактов.

**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. Например, 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.

**Процесс** – совокупность измерений, фактов и атрибутов. По сути, процесс и есть "Куб".

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.

## **OLAP**

Данные из хранилища данных могут быть использованы для составления аналитической отчетности. В аналитической платформе Deductor для оперативного получения необходимых отчетов используется генератор отчетов (OLAP). Наиболее удобным инструментом для получения аналитической отчетности являются OLAP-кубы (On Line Analytical Processing - оперативная аналитическая обработка данных) (Рис.1). OLAP-кубы

представляют собой проекцию исходного куба данных на куб данных меньшей размерности. Такие проекции или срезы исходного куба представляются на плоскости в виде кросс-таблицы или кросс-диаграммы (Рис. 2). Кросс-таблица – удобное средство визуализации многомерных данных и получения необходимых форм отчетов. Кросс-таблица строится на основе многомерного представления в виде OLAP-куба и содержит измерения и факты, определенные при построении куба. При этом значения ячеек объединяются. Основной особенностью кросс-таблицы является то, что ее структура не является жестко определенной. Манипулируя заголовками измерений, пользователь может добиться, чтобы кросс-таблица выглядела наиболее информативно.

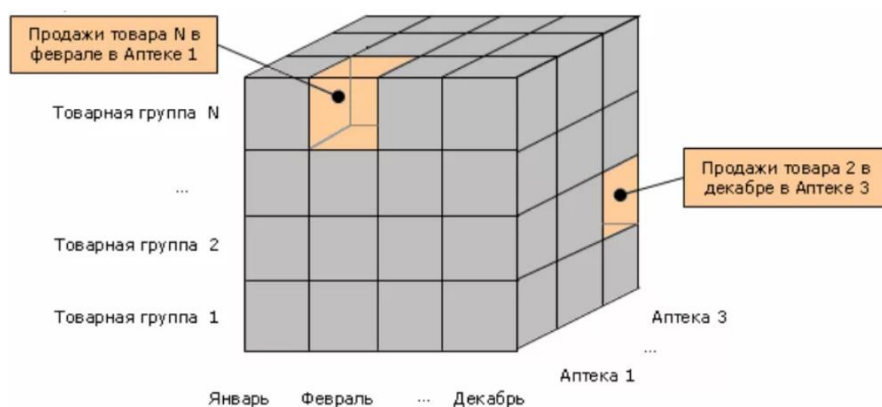


Рис. 1

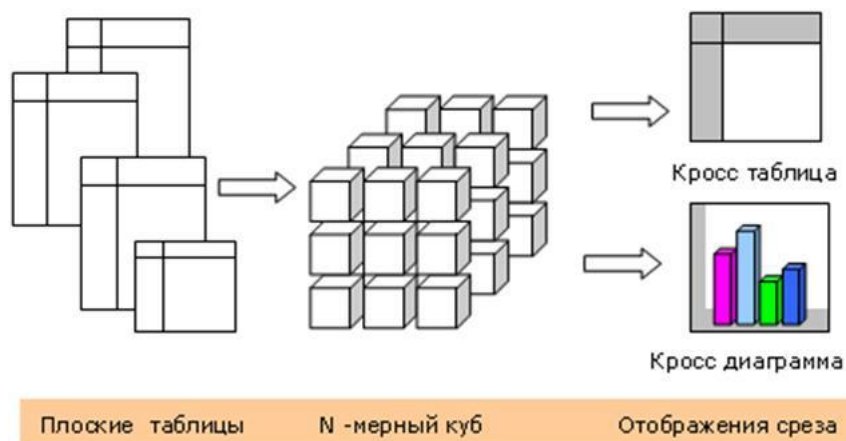




Рис. 2



## Подготовка к лабораторной работе

1. Скачать и установить Deductor <https://basegroup.ru/deductor/download>

### Практическая часть

1. Для создания нового хранилища данных нужно перейти во вкладку подключения, щелкнуть правой кнопкой мыши на «Подключения»  и выбрать «Мастер подключений».
2. Выбрать тип источника Deductor Warehouse и нажать «Далее».
3. Выбрать FireBird, указать путь к бд – D:\farma.gdb (не указывать C:\farma.gdb), логин – sysdba, пароль – masterkey, установить флаг «сохранять пароль» и нажать «Далее».
4. Выбрать последнюю версию (Warehouse 6), нажать «Создать файл базы данных с необходимой структурой метаданных».
5. Далее выбрать «Сведения» и «Метаданные» и нажать «Далее».
6. Указать имя, метку и описание нового хранилища. (Имя должно быть указано латинскими буквами) и нажать «Готово».
7. Проверить подключение, нажав «Настроить подключение»  и «Тест подключения», если подключение прошло успешно, можно продолжать.

### Проектирование структуры хранилища

1. Во вкладке подключения нажать «Открыть редактор метаданных» 
2. Нажать «Разрешить редактировать» 
3. Отметить узел «Измерения», нажать кнопку «Добавить» Имя – GR\_ID  
Метка – Группа.Код, Тип данных – целый
4. Повторить для всех остальных измерений, используя данные таблицы

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

5. Добавить текстовые атрибуты Группа.Наименование, Товар.Наименование и Отдел.Наименование.
6. Измерение Товар.Код должно ссылаться на измерение Группа.Код, добавьте эту ссылку. (Рис. 3)

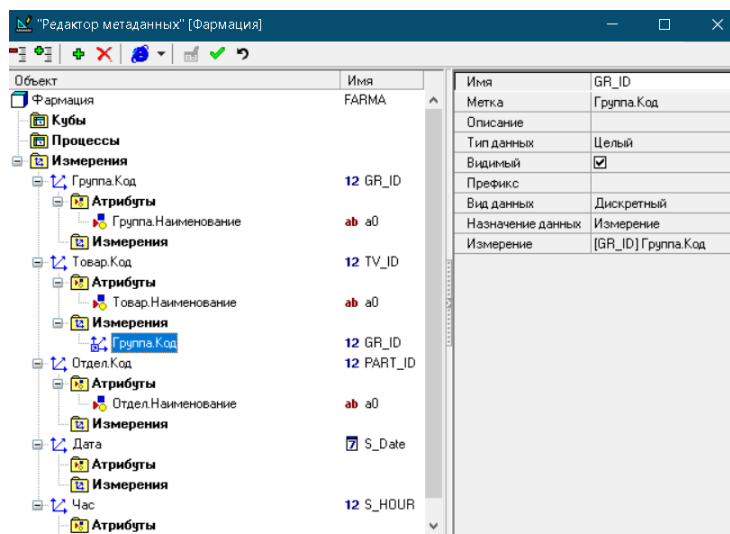


Рис. 3

7. Отметить узел «Процессы», нажать кнопку «Добавить», Метка – Продажи.
8. Добавить в процесс «Продажи» ссылки на измерения Дата, Отдел.Код, Товар.Код, Час.
9. Добавить в процесс «Продажи» два факта: Количество (целый) и Сумма (вещественный). Для этого выделите узел Факты и нажмите «Добавить». (Рис. 4)

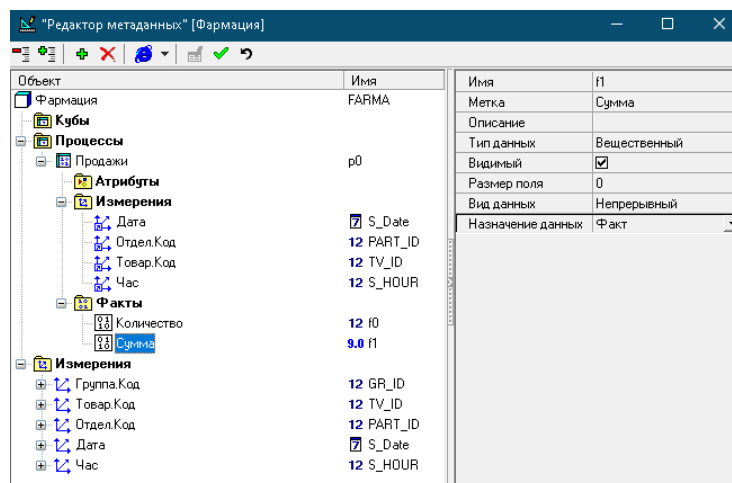



Рис. 4

10. Выйти из редактора метаданных, предварительно нажав кнопку «Принять изменения» .

## Загрузка информации


Информация содержится в четырех текстовых файлах, которые нужно импортировать в Deductor:

Groups.txt – товарные группы

Produces.txt - товары

Stores.txt - отделы

Sales.txt – продажи товаров по дням

1. Перейти на вкладку «Сценарии», нажать ПКМ по  «Сценарии» и запустить мастер импорта.
2. Выбрать источник информации Text, в следующем окне указать путь к файлу groups.txt (удобно разместить этот файл в одной папке со сценарием Deductor, чтобы указать относительный путь). Нажать «Далее», оставив настройки по умолчанию.
3. В окнах «Настройка форматов импорта из файла» и «Параметры импорта файла с разделителями» оставить настройки по умолчанию.
4. В окне «Импорт текстового файла» для Группа.Код указать тип данных - целый, вид данных – дискретный.
5. В следующем окне нажать «Пуск», затем «Далее».

6. Выбрать способ отображения данных – таблица. В следующем окне нажать «Готово».
7. Повторить эти действия для оставшихся трёх текстовых файлов. При импорте файла produces.txt указать параметры столбцов: Товар.Код – целый и дискретный, аналогично для Группа.Код. При импорте файла stores.txt указать параметры столбцов: Отдел.Код – целый и дискретный. При импорте файла sales.txt в окне «Настройка форматов импорта из файла» указать разделитель «.» (точка) для целой и дробной части числа, а также для компонентов даты. Для компонентов времени в качестве разделителя указать «:». Далее указать параметры столбцов: Дата – Дата/Время и дискретный, Сумма – вещественный и дискретный, Отдел.Код – целый и дискретный, аналогично для Товар.Код, Час и Количество.
8. Теперь необходимо загрузить данные из импортированных файлов в измерения. Выделить узел «Текстовый файл (groups.txt)», нажать ПКМ и вызвать мастер экспорта.
9. Выбрать Deductor Warehouse, в следующем окне указать созданное ранее хранилище.
10. Выбрать измерение Группа.Код и нажать «Далее».
11. Необходимо установить соответствие элементов объекта хранилища с полями источника данных. В нашем случае это произошло автоматически, т.к. имена полей в файле и метки в хранилище совпадают (Рис. 5).

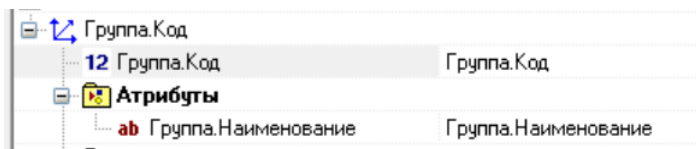


Рис. 5

12. Завершить экспорт файла. Больше ничего менять не требуется.
13. Прodelать аналогичные операции для двух других измерений – Отдел.Код и Товар.Код.



14. Прodelать аналогичные операции для процесса «Продажи». Для этого выбрать «Текстовый файл (sales.txt)» и запустить мастер экспорта. При этом в мастере экспорта появятся два дополнительных окна. В качестве измерения, по которым необходимо удалить данные из хранилища, выбрать Дата. В следующем окне выделить все параметры загрузки данных в хранилище.
15. Завершить экспорт файла. Больше ничего менять не требуется.

## Извлечение информации

1. Сделать выборку данных из процесса «Продажи». Для этого в мастере импорта выбрать тип источника данных – Deductor Warehouse, на следующем шаге – ваше хранилище данных, затем – процесс «Продажи».
2. Выбрать, какие измерения и атрибуты следует импортировать (Рис.6):

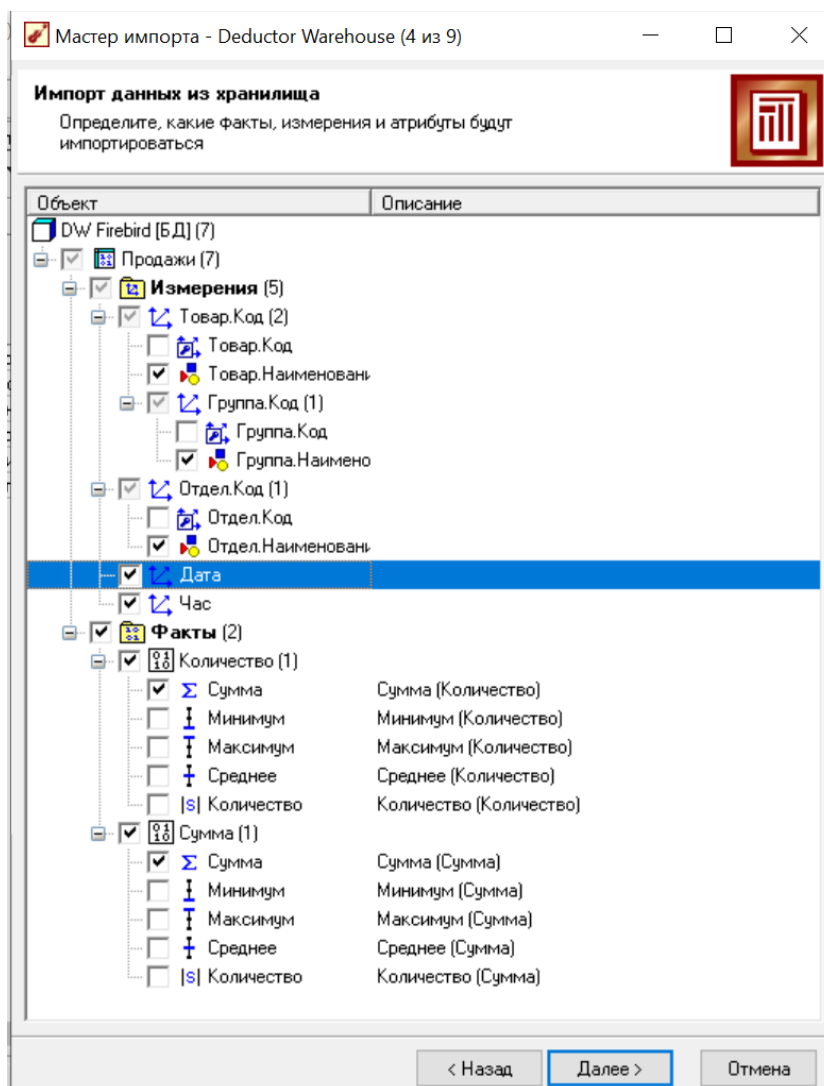


Рис.6

3. Определение среза оставить по умолчанию (Рис. 7).

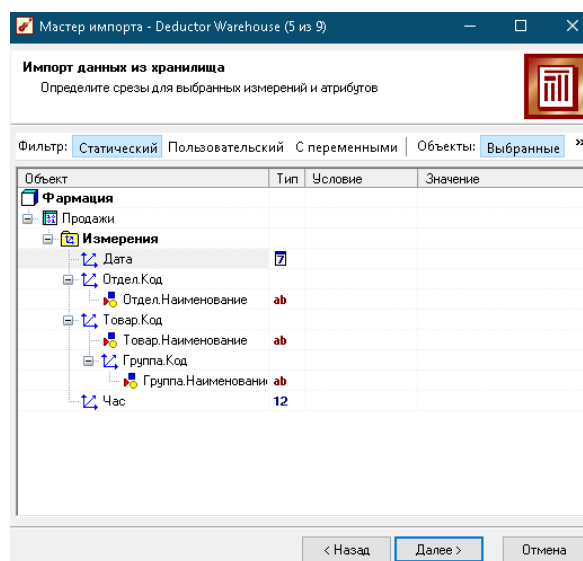


Рис.7

4. Завершить операцию импорта. Больше ничего менять не требуется.

## OLAP в Deductor

Технология OLAP представлена в Deductor в виде визуализатора «Куб» с двумя типами отчетов внутри него (кросс-таблица, кросс-диаграмма). Рассмотрим многомерный отчет, отражающий динамику сумм продаж по месяцам года в разрезе групп и аптек.

В хранилище данных измерение «Месяц» отсутствует, поэтому нужно разделить дату на год и месяц:

1. Выделить узел «Фармация: Продажи», ПКМ вызвать «Мастер обработки» (Рис. 8).

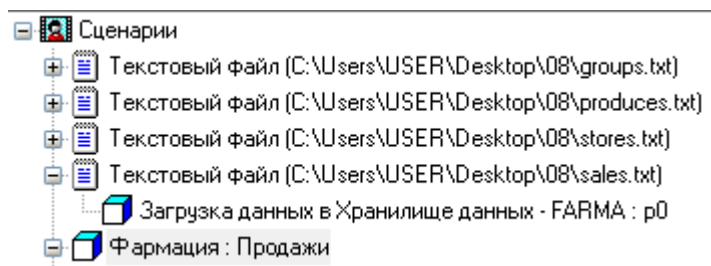


Рис. 8

2. Выбрать «Трансформация данных» → «Дата и время».

3. В окне «Преобразование даты и времени» выбираем «Год» и «Месяц» (Рис. 9)

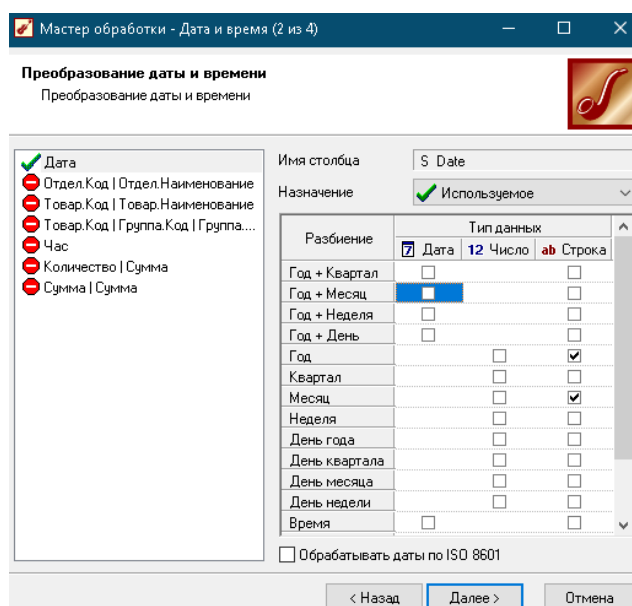


Рис. 9

4. В окне «Определение способа отображения» выбрать «Таблица» и завершить обработку.

### Создание кросс-таблицы

1. Выделить узел «Преобразование данных» и ПКМ вызвать «Мастер визуализации».
2. Выбрать «OLAP анализ» → «Куб».
3. Окно «Настройка назначений полей куба» оставить без изменений.
4. В окне «Настройка измерений» выбрать следующим образом (Рис.10):

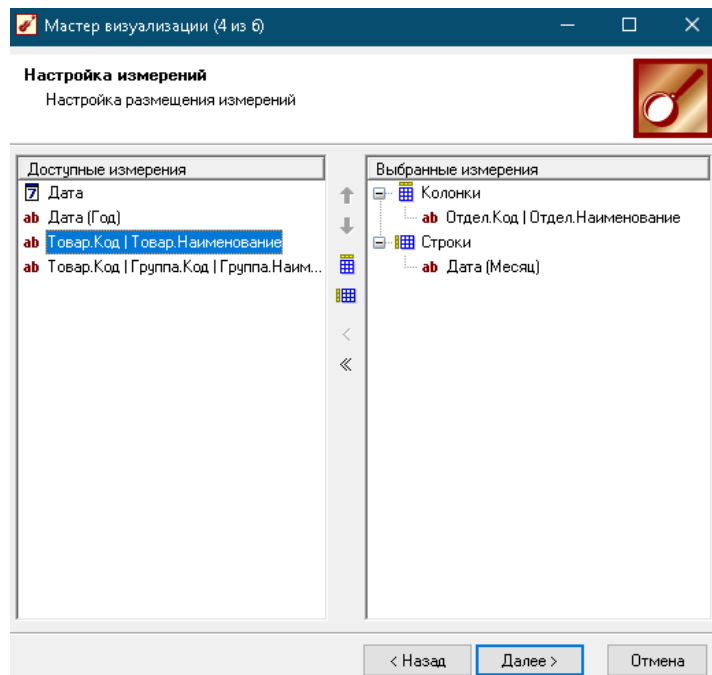


Рис. 10

5. В окне «Настройка фактов» выбрать следующим образом (Рис.11):

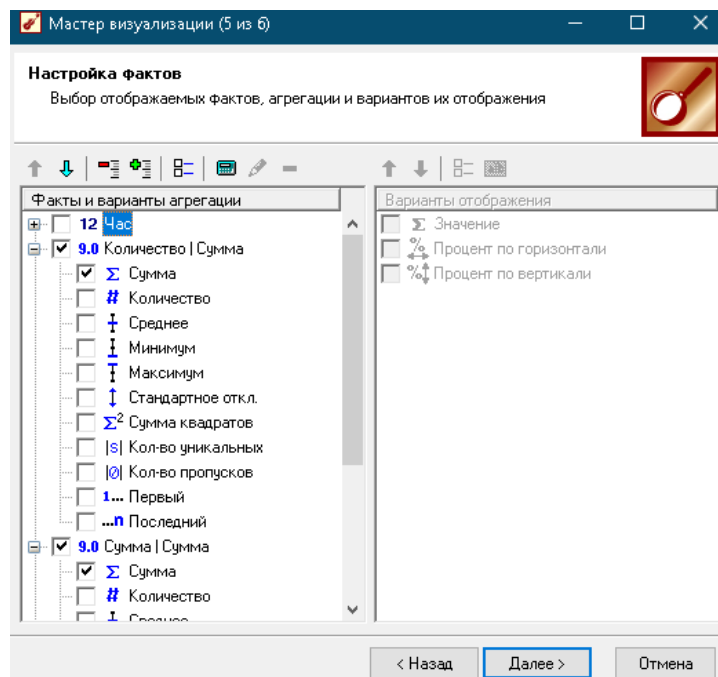


Рис. 11

6. Завершить визуализацию, результат представлен на Рис. 12:

Дата ▼		Дата [Год] ▼		Товар.Код   Товар.Наименование ▼		Товар.Код   Группа.Код   Группа.... ▼			
		Отдел.Код   Отдел.Наименование ▼							
		Аптека 1		Аптека 2		Аптека 3		Итого:	
Дата [Месяц] ▼	Σ	Количес	Σ Сумма	Σ Количес	Σ Сумма	Σ Количес	Σ Сумма	Σ Количес	Σ Сумма
01 Январь		573,00	33 284,05					573,00	33 284,05
02 Февраль		623,00	33 809,16					623,00	33 809,16
03 Март		534,00	32 241,17					534,00	32 241,17
04 Апрель		527,00	33 487,97	353,00	19 370,63			880,00	52 858,60
05 Май		449,00	22 377,64	224,00	10 759,29			673,00	33 136,93
06 Июнь		425,00	21 364,05	186,00	8 160,47			611,00	29 524,52
07 Июль		373,00	13 536,43	164,00	8 158,10			537,00	21 694,53
08 Август		312,00	14 324,56	227,00	10 764,90			539,00	25 089,46
09 Сентябрь		453,00	23 436,44	278,00	15 008,15			731,00	38 444,59
10 Октябрь		536,00	31 328,32	361,00	21 777,81	566,00	35 965,42	1 463,00	89 071,55
11 Ноябрь		588,00	33 413,86	281,00	16 416,76	603,00	46 399,28	1 472,00	96 229,90
12 Декабрь		591,00	32 596,49	350,00	21 365,74	655,00	43 121,26	1 596,00	97 083,49
Итого:		5 984,00	325 200,14	2 424,00	131 781,85	1 824,00	125 485,96	10 232,00	582 467,95

Рис. 12


В данной кросс-таблице представлены следующие измерения:

Отдел.Наименование – Аптека 1, Аптека 2, Аптека 3;

Дата(Месяц) – месяцы работы отделов;

Группа.Наименование – название групп лекарственных препаратов, присутствующих в продаже (причем данное измерение скрыто).

Фактами в данном случае будут стоимость(сумма) и количество проданных медикаментов.

С помощью  создайте кросс-диаграмму.

## Фильтрация

Определить товарные группы, приносящие 80% выручки

1. Перетаскивая измерения, изменить кросс-таблицу следующим образом (Рис.13):

Дата ▼		Дата (Год) ▼		Дата (Месяц) ▼		Товар.Код   Товар.Наименование ▼			
Отдел.Код   Отдел.Наименование ▼									
		Аптека 1		Аптека 2		Аптека 3		Итого:	
Товар.Код   Группа.Код   Группа..... ▼	Σ Количес	Σ Сумма	Σ Количес	Σ Сумма	Σ Количес	Σ Сумма	Σ Количес	Σ Сумма	
Антисептики и дезинфицирующие средства	2 340,00	93 707,56	1 222,00	46 125,89	624,00	32 070,39	4 186,00	171 903,84	
Биологически активные пищевые добавки	11,00	1 485,36	5,00	798,10	8,00	1 457,48	24,00	3 740,94	
Витамины и витаминоподобные средства	1 331,00	109 651,83	572,00	51 419,26	387,00	39 369,43	2 290,00	200 440,52	
Желчегонные средства и препараты желчи	76,00	1 950,42	32,00	865,58	26,00	638,74	134,00	3 454,74	
Иммуномодуляторы	1 150,00	90 621,24	366,00	26 953,69	508,00	42 635,45	2 024,00	160 210,38	
Местные анестетики	362,00	8 816,82	166,00	2 862,39	93,00	2 673,78	621,00	14 352,99	
Микро- и макроэлементы	450,00	8 319,45			123,00	2 295,49	573,00	10 614,94	
Общетонизирующие средства и адаптогены	264,00	10 647,46	61,00	2 756,94	55,00	4 345,20	380,00	17 749,60	
Итого:	5 984,00	325 200,14	2 424,00	131 781,85	1 824,00	125 485,96	10 232,00	582 467,95	

Рис.13

2. Нажать 

3. Настроить фильтр (Рис. 14):

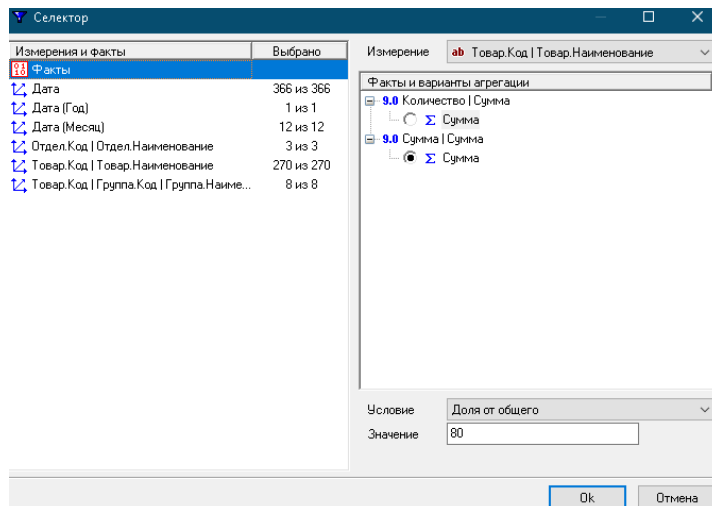


Рис. 14

## Отчеты

1. В меню «Вид» выбрать пункт «Отчеты».
2. Нажать «Добавить узел».
3. Выделить нужный узел дерева сценария, например, «Фармация: Продажи»

## Содержание отчета

1. Титульный лист
2. Цель лабораторной работы
3. Скриншоты этапов выполнения работы с пояснениями
4. Вывод

## Контрольные вопросы

1. К каким 4 операциям сводится вся работа в Deductor Studio?
2. Что такое куб, измерения и факты?
3. Как расшифровывается OLAP?

4. OLAP представлена в Deductor в виде визуализатора «Куб» с двумя типами отчетов внутри него, назовите их.
5. Что такое процесс, атрибут процесса?

### **Литература**

1. Кацко И.А., Паклин Н.Б. Практикум по анализу данных на компьютере / Под ред. Г.В. Гореловой. – М.: КолосС, 2009 – 278 с.: ил. – (Учебники и учеб. Пособия для студентов высш. учеб. заведений).
2. ПРАКТИКУМ [Р.002] Базовые навыки работы в Deductor Studio. [Электронный ресурс] : [сайт]. [2009]. URL: <https://docplayer.ru/28612323-Praktikum-p-002-bazovye-navyki-raboty-v-deductor-studio.html> (дата обращения: 07.11.2019).