

**Министерство образования и науки Российской Федерации
ФГАОУ ВО «Севастопольский государственный университет»**

**Институт информационных технологий
и управления в технических системах**

**Практическая работа №1
«Аналитическая система Deductor: Базовые навыки работы
в Deductor Studio»**

по дисциплине «Интеллектуальный анализ данных»
для студентов всех форм обучения направления подготовки
09.03.02 «Информационные системы и технологии»



Севастополь
2017

Цель: приобрести базовые навыки работы в Deductor Studio

Время: 2 часа

Краткие теоретические сведения

1. Развитие и назначение Deductor

Deductor – это аналитическая платформа, основа 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 построения моделей и визуализации полученных результатов.

Имеющиеся в ней средства позволяют на базе единой архитектуры реализовать все этапы построения аналитической системы – от создания хранилища данных до автоматического подбора моделей и визуализации полученных результатов (в виде кубов OLAP, таблиц, диаграмм, гистограмм, карт, графов и т.д.).

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** – клиент доступа к Deductor Server. Обеспечивает доступ к серверу из сторонних приложений и управление его работой.

Существует три типа варианта поставки платформы Deductor:

- Enterprise;
- Professional;
- Academic.

В зависимости от типа поставки набор доступных компонентов может различаться.

Версия Enterprise предназначена для корпоративного использования. В ней присутствуют:

- серверные компоненты Deductor Server и Deductor Client .
- интерфейс доступа к Deductor через механизм OLE Automation.
- традиционное хранилище данных Deductor Warehouse на трех СУБД: Firebird, MS SQL, Oracle.
- виртуальное хранилище данных Deductor Virtual Warehouse .

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.

Версии Professional и Enterprise требуют установки драйверов Guardant для работы с лицензионным ключом.

Версия Academic предназначена для образовательных и обучающих целей. Ее функционал аналогичен версии Professional за исключением:

- отсутствует пакетный запуск сценариев, т.е. работа в программе может вестись только в интерактивном режиме;
- отсутствует импорт из промышленных источников данных: 1С, СУБД, файлы MS Excel, Deductor Data File;

- некоторые другие возможности.

Вся работа в Deductor сводится к использованию нескольких мастеров – подключений, импорта, обработки, визуализации и экспорта. Платформа не имеет собственных средств ввода-вывода информации, но мастера импорта и экспорта обеспечивают взаимодействие с произвольными источниками и приемниками данных, для которых существуют стандартные механизмы доступа.

Обработка и визуализация в Deductor допускают любые манипуляции над набором данных – от самых простых, таких как сортировка, до весьма сложных (например, построение модели нейронной сети). Обработчик можно представить в виде «черного ящика», на вход которого подается исходный набор данных, а на выходе формируется преобразованный набор. Реализованные в Deductor обработчики позволяют осуществлять как анализ данных, так и различные операции с ними (очистка, слияние, объединение, фильтрация).

В процессе развертывания и использования аналитической платформы с ней взаимодействуют различные категории пользователей. Можно выделить четыре основные категории:

- аналитик;
- пользователь;
- администратор;
- программист.

Функции аналитика:

- создание в Deductor Studio сценариев – последовательности шагов, которую необходимо провести для получения нужного результата.
- построение, оценка и интерпретация моделей.
- настройка панели отчетов для пользователей Deductor Viewer.
- настройка сценария на поточную обработку новых данных.

Функции пользователя:

- просмотр готовых отчетов в Deductor Viewer.

Функции администратора:

- установка компонентов Deductor на рабочих местах и сервера ключей Guardant при необходимости.
- развертывание традиционного хранилища данных на сервере.
- контроль процедур регулярного пополнения хранилища данных.
- конфигурирование сервера Deductor Server.
- настройка пакетной и/или серверной обработки сценариев Deductor.
- оптимизация доступа к источникам данных, в том числе к хранилищу данных.

Функции программиста:

- интеграция Deductor с источниками и приемниками данных.
- вызов Deductor из внешних программ различными способами, в том числе взаимодействие с Deductor Server.

2. Начало работы с системой

Главное окно Deductor Studio

Главное окно Deductor Studio представлено на рис 1.

По умолчанию панель управления представлена одной вкладкой Сценарии. Кроме того, доступны еще две вкладки: Отчеты и Подключения. Сделать их видимыми можно следующими способами:

- главное меню: Вид → Отчеты и Вид → Подключения

- кнопки  и  на панели инструментов.

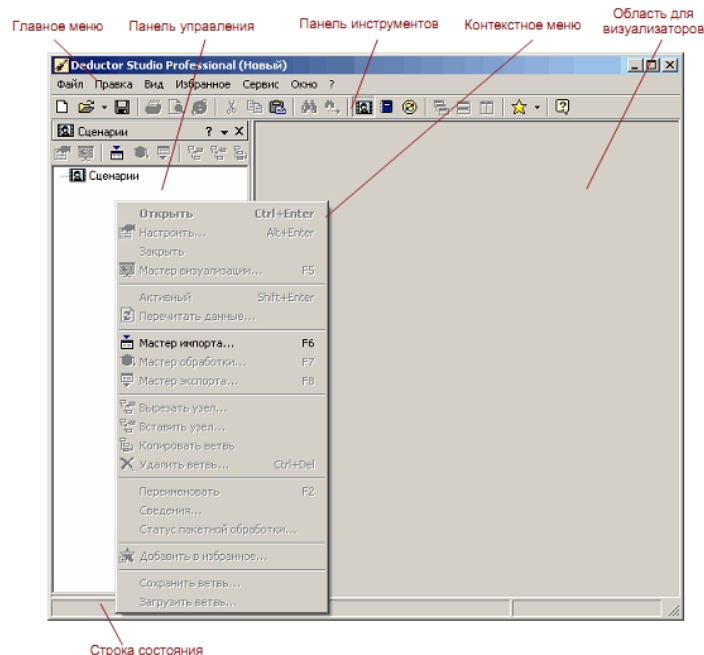


Рисунок 1. Главное окно Deductor Studio

Можно производить «drag & drop» манипуляции с вкладками, меняя их расположение и порядок.

При нажатии правой кнопки мыши на любой вкладке появляется контекстное меню (рис 2):

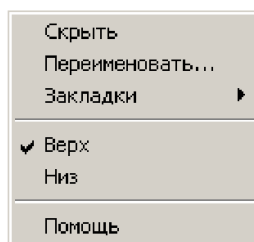


Рисунок 2. Контекстное меню для работы со вкладками

- Скрыть – делает вкладку невидимой;
- Переименовать – переименовывает название вкладки;
- Закладки – переключается на выбранную закладку;
- Верх/Низ – задает расположение названий вкладок: вверху либо внизу;
- Помощь – открывает раздел справки.

Понятие проекта

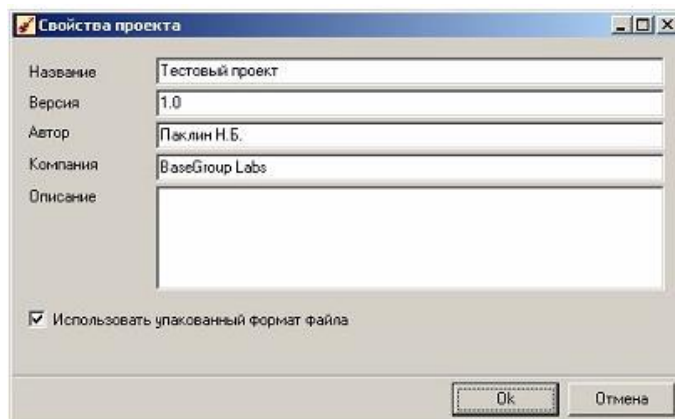
В Deductor Studio ключевым понятием является проект. Это файл с расширением*.ded, по структуре соответствующий стандартному xml-файлу. Он хранит в себе:

- последовательности обработки данных (сценарии);
- настроенные визуализаторы;
- переменные проекта и служебную информацию.

Замечание

По умолчанию файл проекта Deductor при сохранении запаковывается, что позволяет уменьшить его размер, поэтому просмотреть запакованный файл в виде xml невозможно. Для этого нужно снять опцию **Использовать упакованный формат файла** в диалоговом окне Свойства проекта (меню Файл → Свойства проекта...)

Каждый проект имеет авторские сведения: Название, Версия, Автор, Компания, Описание. Они заполняются в диалоговом окне Свойства проекта (меню Файл → Свойства проекта...).



Создать новый проект можно следующими способами:

- главное меню Файл → Создать;
- кнопка Создать новый проект на панели инструментов;
- клавиша Ctrl+N.

Открытие существующего проекта:

- главное меню Файл → Открыть;
- кнопка Открыть проект на панели инструментов;
- клавиша Ctrl+O.

Открыть проект можно еще одним способом: в главном меню Файл → История, найти имя проекта. Способ работает в том случае, если вы недавно открывали этот проект, и он сохранился в менеджере истории проектов.

В одной запущенной копии Deductor Studio можно открыть только один проект.

Для сохранения проекта под текущим именем нужно выбрать главное меню Файл →

Сохранить, нажать кнопку  или комбинацию Ctrl+S.

Для сохранения текущего проекта под другим именем: главное меню Файл → Сохранить как...

Понятие сценария и узла обработки

В Deductor Studio для аналитика основополагающим понятием является сценарий.

Сценарий представляет собой последовательность операций с данными, представленную в виде иерархического дерева. В дереве каждая операция образует узел, заголовок которого содержит: имя источника данных, наименование применяемого метода обработки, используемые при этом поля и т.д (рис 3). Кроме этого, слева от наименования узла стоит значок, соответствующий типу операции.

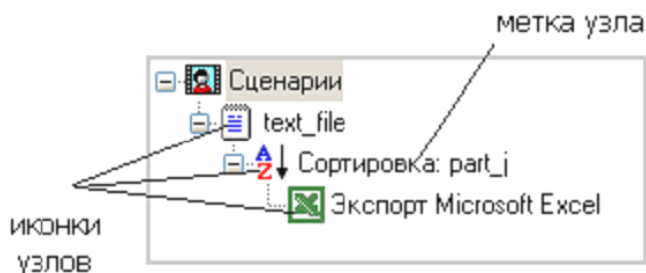


Рисунок 3. Дерево сценария

Если узел имеет подчиненные узлы, то слева от его названия будет расположен значок «+», щелчок по которому позволит развернуть узел, т.е. сделать видимыми все его подчиненные узлы, при этом значок «+» поменяется на «-». Щелчок по значку «-», наоборот, сворачивает все подчиненные узлы.

С помощью клавиш Ctrl+ ↑ и Ctrl+ ↓ можно перемещать узлы по дереву вверх-вниз в пределах подчинения родительскому узлу.

Сценарий состоит из ветвей.

Deductor не имеет собственных средств для ввода данных, поэтому сценарий всегда начинается с узла импорта из какого-либо источника. Любой вновь создаваемый узел импорта будет находиться на верхнем уровне (подчиненным главному узлу Сценарии).

Создание нового узла импорта осуществляется с помощью Мастера импорта. Вызвать мастер можно следующими способами:



- кнопка на панели инструментов закладки Сценарии;
- клавиша F6;
- контекстное меню Мастер импорта...

При вызове мастера импорта откроется окно первого шага мастера (рис 4).

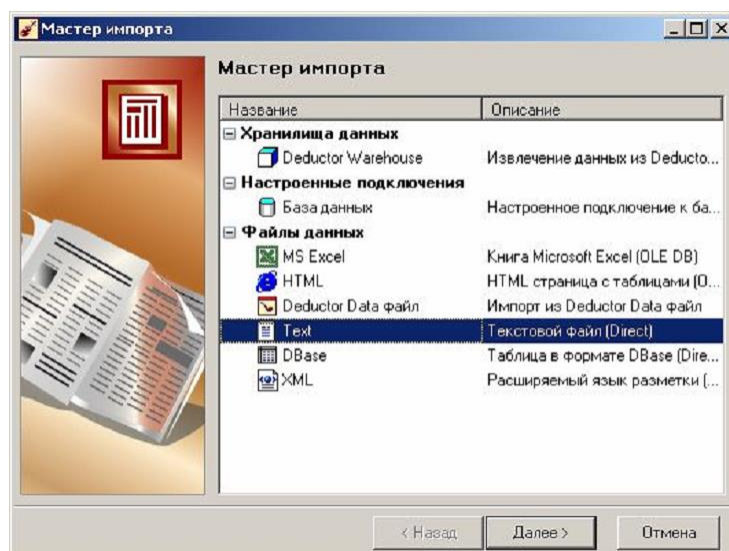


Рисунок 4. Мастер импорта. Шаг 1.

В нем все источники данных сгруппированы по следующим четырем категориям:

- хранилища данных;
- настроенные подключения;
- файлы данных;
- бизнес-подключения.


Некоторые категории могут отсутствовать в списке. Причиной этого может быть следующее:

- версия Deductor. Например, категории Настроенные подключения и Бизнес-подключения отсутствуют в версии Academic.
- в дереве подключений (вкладка Подключения) не зарегистрировано ни одного объекта из данной категории. Например, если не настроено ни одного подключения к хранилищу данных, то категория Хранилища данных будет отсутствовать.
- отключена «видимость» объекта или категории объекта.

Дальнейшие шаги мастера импорта будут зависеть от того, какой объект дерева категорий был выбран аналитиком.

К любому узлу импорта можно добавить узел обработки или узел экспорта, предварительно выделив узел импорта мышью. Новый узел будет добавлен как подчиненный к узлу импорта.

Создание нового узла обработки осуществляется с помощью Мастера обработки. Вызвать мастер можно следующими способами:

- кнопка  на панели инструментов закладки Сценарии;
- клавиша F7;
- контекстное меню Мастер обработки...

При вызове мастера обработки откроется окно первого шага мастера (рис 5).

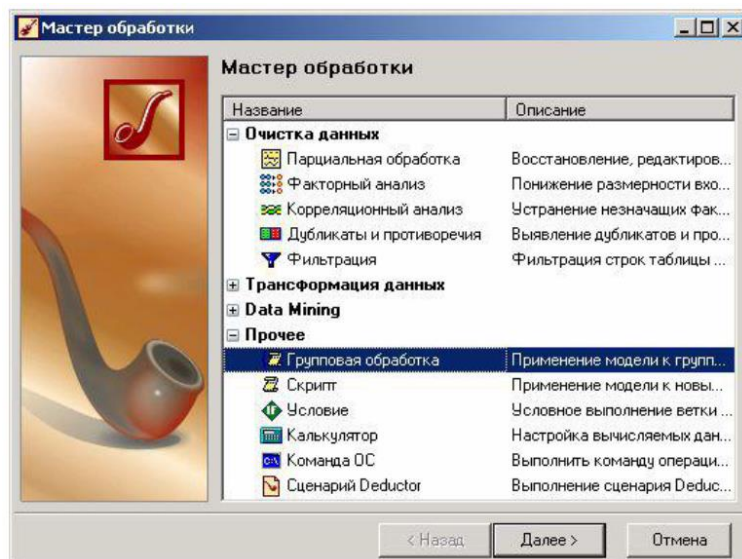


Рисунок 5 Мастер обработки.


В нем все обработчики сгруппированы по следующим четырем категориям:

- Очистка данных;
- Трансформация данных;
- Data Mining;
- Прочее.

Некоторые узлы могут отсутствовать в списке. Причины этого следующие:

- версия Deductor;
- отключена «видимость» объекта (или целой категории) объекта;
- узел «устарел» и в текущей версии Deductor его создание невозможно (допускается только его чтение и настройка).

Создание нового узла экспорта осуществляется с помощью Мастера экспорта. Вызвать мастер можно следующими способами:

- кнопка  на панели инструментов закладки Сценарии;
- клавиша F8;
- контекстное меню Мастер экспорта...

В нем все приемники данных сгруппированы по следующим 5 категориям:

- хранилища данных;
- базы данных;
- файлы;
- Web-серверы;
- прочее.

Причины отсутствия некоторых объектов или категорий мастера экспорта аналогичны тем, что перечислены при описании мастера импорта.

После узла экспорта невозможно добавить ни один узел.

Базовые операции над узлами сценария

Кроме команд вызова мастеров, к каждому узлу применимы базовые операции. Операции над узлами и ветками сценария можно выполнять следующими способами:

- кнопки панели инструментов на закладке Сценарии;
- контекстное меню;
- мышь.

Список доступных операций.

1. Открытие узла – узел запускается на выполнение, причем выполняются все родительские узлы, а справа открываются визуализаторы, настроенные для данного узла. В интерактивном режиме для каждого узла должен быть настроен хотя бы один визуализатор, например, Таблица или Сведения. Операция вызывается:

- двойной щелчок мышью на узле;
- клавиши Ctrl+Enter;
- контекстное меню Открыть.

2. Настройка узла – вызывается мастер импорта, мастер обработки или мастер экспорта, в зависимости от типа узла, для изменения параметров обработки, производимой в узле.

Операция вызывается:

- кнопка  ;
- клавиши Alt+Enter;
- контекстное меню Настроить....

3. Активация/деактивация узла – узел может быть либо активным, либо неактивным. Если узел неактивный, то, сделав его активным, выполнится сценарий для этого узла, но визуализаторы отображены не будут. Делая узел неактивным, закрываются все визуализаторы для него и для всех подчиненных узлов, а сам узел и подчиненные узлы превращаются в неактивные. Эта операция может быть использована для освобождения памяти. Операция активации/деактивации вызывается:

- клавиши Shift+Enter;
- контекстное меню Активный...

4. Перечитать данные узла – все узлы до корневого включительно будут закрыты, а затем выполнена ветка сценария от корневого до текущего узла. Операция вызывается:

- контекстное меню Перечитать данные...

5. Вырезать узел – удаляет текущий узел из сценария обработки. Все его потомки при этом перемещаются на один уровень вверх и начинают подчиняться родителю удаленного узла.

Операция вызывается:

- кнопка  ;
- контекстное меню Вырезать узел.

6. Вставить узел – вставляет перед текущим узлом сценария новый узел и вызывает для него мастер обработки. Вставить узел перед узлом импорта данных нельзя. Операция вызывается:

- кнопка  ;
- контекстное меню Вставить узел.

После вставки нового узла или удаления существующего узлы-потомки могут стать неработоспособными, в зависимости от обработки, выполняемой новым узлом.

7. Копировать ветвь – копирует ветвь сценария, начиная с текущего узла и включая все его потомки. Операция вызывается:

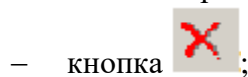


– кнопка

– контекстное меню Копировать ветвь;

– при помощи механизма drag & drop – выделив узел, и, удерживая нажатой клавишу Ctrl, указать курсором мыши на новый узел, который должен стать родителем старого. При этом переносимая ветка целиком скопируется в новое место.

8. Удалить ветвь – удаляет узел сценария и все его подузлы. Удаленная ветвь восстановлению не подлежит, поэтому к данной операции необходимо подходить с осторожностью. Операция вызывается:



– кнопка

– клавиши Ctrl+Del;

– контекстное меню Удалить ветвь.

9. Перенос ветви – переносит ветку сценария к новому узлу. Операция производится аналогично копированию ветви с помощью drag & drop без удерживания клавиши Ctrl.

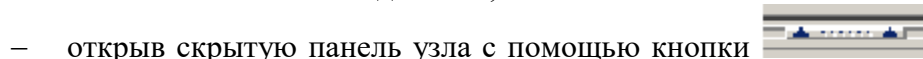
10. Переименовать – позволяет изменить метку текущего узла. Операция вызывается:

– клавиша F2;

– контекстное меню Переименовать...

11. Сведения – открывает диалоговое окно Сведения для текущего узла. В нем редактируется имя, метка и описание к узлу. Операция вызывается:

– контекстное меню Сведения...;



– открыв скрытую панель узла с помощью кнопки и нажать там одну из кнопок: Имя, Метка или Описание.

Имя узла может быть задано только латинскими символами, тогда как метка – любыми. Кроме того, имя узла должно быть уникально в пределах одного сценария. Как правило, необходимости в переименовании имен узлов не возникает.

12. Статус пакетной обработки – устанавливает статус пакетной обработки для узла.

13. Добавить в Избранное – текущий узел добавляется в список избранных узлов.

14. Сохранение ветви – вызывается стандартный диалог Сохранение, в котором можно указать путь и имя файла для сохранения ветви сценария, начинающейся с текущего узла.

Операция вызывается:

– контекстное меню Сохранить ветвь.

15. Загрузка ветви – вызывает стандартный диалог Открытие файла, в котором можно указать путь и имя файла, хранящего ветвь сценария. Загруженная ветвь сценария станет потомком текущего узла. Ветвь, начинающаяся с узла импорта данных, будет добавлена в проект как новая корневая ветвь. Операция вызывается:

– контекстное меню Загрузить ветвь.

По умолчанию ветвь сценария имеет расширение *.deb.

Импорт из текстовых файлов с разделителями.


Структурированный текстовый файл с разделителями – один из самых распространенных форматов хранения данных. Такой файл представляет собой обычный текстовый файл, столбцы данных в котором разделены однотипными символами-разделителями, например, символами табуляции, пробела, точки с запятой и т.д.

Процесс импорта данных из текстового с разделителями файла в мастере импорта (категория Текстовый файл (Direct)) содержит следующие шаги:

- указание имени файла;
- настройка параметров импорта;
- настройка импортируемых полей;

- запуск процесса импорта;
- выбор способа визуализации;
- задание сведений об узле.



На шаге Указание имени файла, нажав кнопку , необходимо выбрать имя текстового файла (расширения *.txt , *.csv), из которого следует выполнить импорт данных. После этого в поле «Имя файла» окна Мастера импорта появится имя выбранного файла и путь.

Допускается вручную ввести путь к файлу в строке поля Имя файла.

Имеется возможность использовать как абсолютные, так и относительные пути для файлов.

Они указываются относительно текущей директории Deductor. При открытии Deductor текущей директорией является директория файла проекта. Поэтому, если файл проекта и текстовые файлы располагаются в одной папке, то использование относительных путей в Мастере импорта позволит не перенастраивать узлы импорта при изменении расположения папки на жестком диске.

Здесь также доступны настройки:

- Начать импорт со строки – номер строки, начиная с которой будет делаться импорт данных из файла.
- флаг Первая строка является заголовком – установка флажка означает, что узел будет импортировать данные с учетом того, что все записи первой строки являются заголовками столбцов.
- Кодировка – ANSI (Windows) или ANCI (MS DOS).

На шаге Настройка параметров импорта нужно настроить параметры импорта данных из текстового файла, так как существует несколько форматов структурированных текстовых файлов. Доступные опции (рис 6):

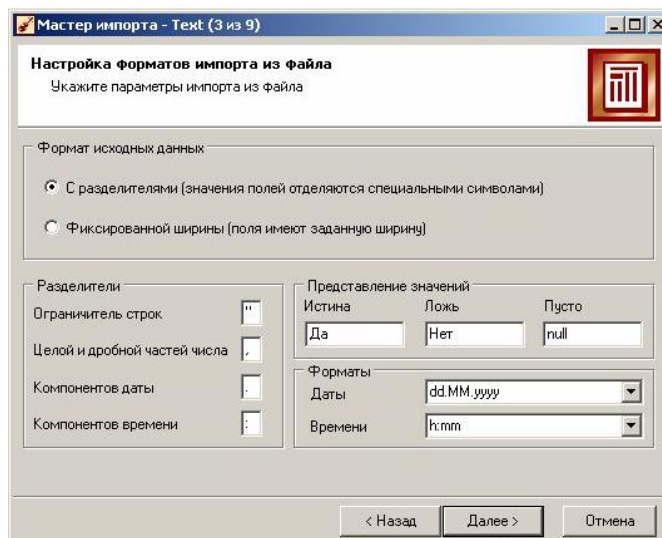


Рисунок 6. Настройка параметров импорта

- переключатель Формат исходных данных, который определяет символ-разделитель в файле (например: символ табуляции, пробел, запятая). Разделитель чаще всего присутствует. Если же нет, то нужно выбрать переключатель Фиксированной ширины (поля имеют заданную ширину) , а позже установить ширину каждого поля.
- Ограничитель строк – при задании данного параметра необходимо указать, какой именно ограничитель строкового значения нужно использовать при импорте данных из текстового файла. Обычно таким ограничителем является символ двойной кавычки " .

– Разделитель дробной и целой части числа – при задании данного параметра необходимо указать символ, разделяющий дробную и целую части в числовых значениях, содержащихся в файле.

– Разделитель компонентов даты – указывается символ, разделяющий компоненты даты в соответствующих значениях, содержащихся в файле.

– Разделитель компонентов времени – указывается символ, разделяющий компоненты времени в соответствующих значениях, содержащихся в файле.

– Форматы Даты/Времени – указываются форматы даты/времени, используемые в импортируемом файле.

– Представление значений – опция для полей логического типа, которое может принимать одно из трех значений – истина (true), ложь (false) и пустое значение (null). Определяет регламент записи в эти значения. Так, при настройках по умолчанию для любого логического поля значение Да будет восприниматься как истина, Нет – как ложь.

В качестве разделителей, представлений значений и форматов по умолчанию всегда предлагаются системные настройки операционной системы. Поэтому при импорте необходимо обращать внимание на их соответствие формату в импортируемом текстовом файле.

Следующее окно мастера зависит от установленного переключателя в флажке Формат исходных данных. Если был выбран формат С разделителями, то появится вкладка, на которой нужно явно указать символ-разделитель (по умолчанию – табуляция). Здесь же находится флаг Считать последовательные разделители одним – в случае последовательно идущих символов-разделителей они будут восприниматься за один. Такое бывает, например, когда символом-разделителем выступают несколько пробелов.

Предпросмотр текстового файла в виде таблицы внизу (загружаются только первые 10 строк) позволяет убедиться в корректности выбора настроек импорта даже не запуская его (рис 7).

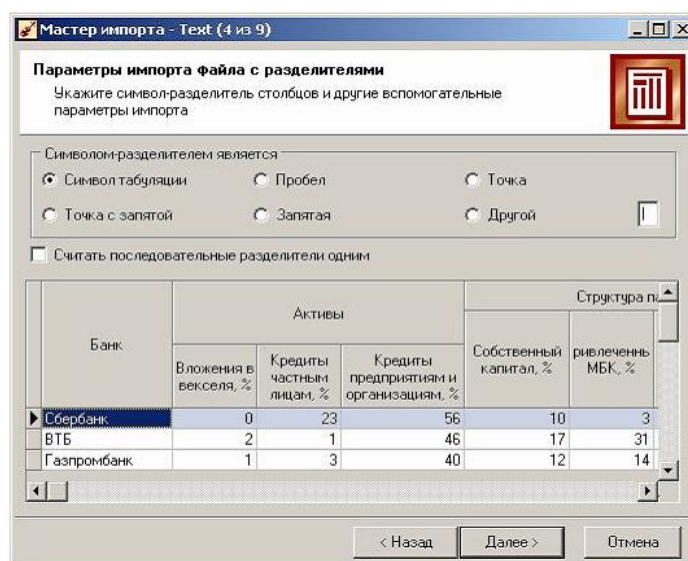
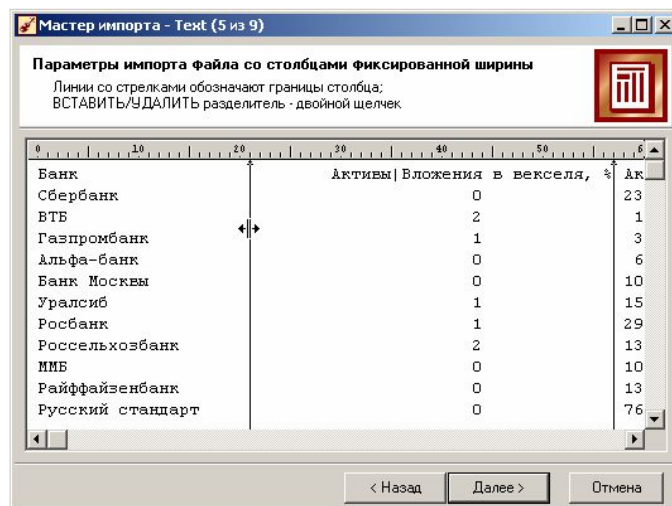


Рисунок 7 Предпросмотр текстового файла

Если был выбран флаг формат Фиксированной ширины, то появится вкладка, на которой нужно задать границы каждого поля. Создание, как и удаление маркера границы производится одним щелчком мыши. Двигая маркеры границ столбцов, можно изменять их, если они расставлены неправильно. Данные, распределенные по столбцам, показываются в области предварительного просмотра (рис 8).



На шаге Настройка параметров столбцов нужно настроить следующие параметры столбцов импортируемых данных, указав соответствующие значения в полях (рис 9).

Рисунок 9. Настройка параметров столбцов

Имя столбца – указывается имя, которое будет служить идентификатором столбца в последующих узлах. По умолчанию предлагается заголовок столбца из текстового файла, если на предыдущем шаге был установлен флажок Первая строка является заголовком.

Метка столбца – название, под которым данный столбец будет виден в визуализаторах. Допускаются любые символы, уникальность имен не обязательна.






Тип	Описание
 <i>логический</i>	данные в поле могут принимать только два значения – 0 или 1
 <i>дата/время</i>	поле содержит данные типа дата/время
 <i>вещественный</i>	числа с плавающей точкой
 <i>целый</i>	целые числа
 <i>строковый</i>	строки символов

Рисунок 10. Тип данных.

Узел импорта всегда пытается автоматически распознать тип данных по первой строке файла (если имеются заголовки, то по второй строке). Такой алгоритм срабатывает не всегда.

Вид данных – характер данных, содержащихся в столбце (рис 11) :

Вид	Описание
<i>непрерывный</i>	значения в столбце могут принимать любое значение в рамках своего типа
<i>дискретный</i>	данные в столбце могут принимать ограниченное число значений

Рисунок 11. Вид данных

Непрерывными могут быть только числовые данные. Дискретный характер носят, как правило, строковые данные, но не всегда. Дискретными могут быть назначены в зависимости от контекста решаемой задачи данные целого типа, реже – вещественного. Вид данных столбца влияет на:

- алгоритм расчета статистики по столбцу;
- работу аналитических алгоритмов.

Назначение – определяет порядок использования поля набора данных, полученного в результате импорта столбца (поля), при дальнейшей обработке импортированных данных (рис 12):










Назначение	Описание
 <i>первичный ключ</i>	поле будет использоваться в качестве первичного ключа
 <i>входное</i>	поле набора данных, построенное на основе столбца, будет являться входным полем обработчика (нейронной сети, дерева решений и т.д.)
 <i>выходное</i>	поле набора данных, построенное на основе столбца, будет являться выходным полем обработчика (например, целевым полем для обучения нейронной сети).
 <i>информационное</i>	поле содержит вспомогательную информацию, которую часто полезно отображать, но не следует использовать при обработке
 <i>измерение</i>	поле будет использоваться в качестве измерения в многомерной визуализации
 <i>атрибут</i>	поле содержит описание свойств или параметров некоторого объекта
 <i>факт</i>	значения поля будут использованы в качестве фактов в многомерной визуализации
 <i>транзакция</i>	транзакция – поле, содержащее идентификатор событий, происходящих совместно (одновременно); например, номер чека, по которому приобретены товары
 <i>элемент</i>	поле, содержащее элемент транзакции (событие).

Рисунок 12. Назначение

Изменить назначение группы столбцов одной операцией можно следующим образом:

- удерживая клавишу Shift , выделить мышкой или клавишами Ctrl+ ↓, Ctrl+ ↑ первый и последний столбцы группы столбцов и изменить их назначение;
- удерживая клавишу Ctrl , выделить мышкой только нужные столбцы и изменить их назначение.

Замечание

Установка назначения столбца набора данных при импорте не является обязательным действием (по умолчанию при импорте установлено назначение «информационное»). Однако это может снизить объем рутинных действий при последующем конструировании сценария. Например, при построении моделей (группа узлов обработки Data Mining) по умолчанию выходным полем, как правило, предлагается последнее поле, и, если это не так, придется каждый раз переопределять назначения полей в каждом новом узле.

На шаге Запуск процесса импорта стартует сам процесс импорта данных с ранее настроенными параметрами. Ход процесса импорта отображается с помощью индикатора. Если процесс импорта остановился, это сигнализирует о возможных ошибках при чтении данных. В этом случае появляется окно с сообщением об ошибке.

В случае возникновения ошибок несоответствия типов процесс импорта будет продолжен, но после его окончания будет отображен журнал регистрации ошибок с информацией о месте и причине их появления:

Для управления процессом импорта предусмотрены следующие кнопки:

- Пуск – запускает процесс в первый раз или возобновляет после паузы.
- Пауза – временно приостанавливает импорт.
- Стоп – останавливает процесс без возможности его продолжения.

На оставшихся двух шагах мастера импорта будет предложено выбрать визуализатор набора данных (по умолчанию предлагается Таблица) и задать сведения об узле.

Узел Настройка набора данных

Обработчик Настройка набора данных позволяет:

- изменить имя, метку, тип, вид и назначение полей текущего набора данных;
- изменить порядок следования столбцов в наборе данных;
- скрыть столбцы набора данных;
- задать опцию кэширования выходного набора.

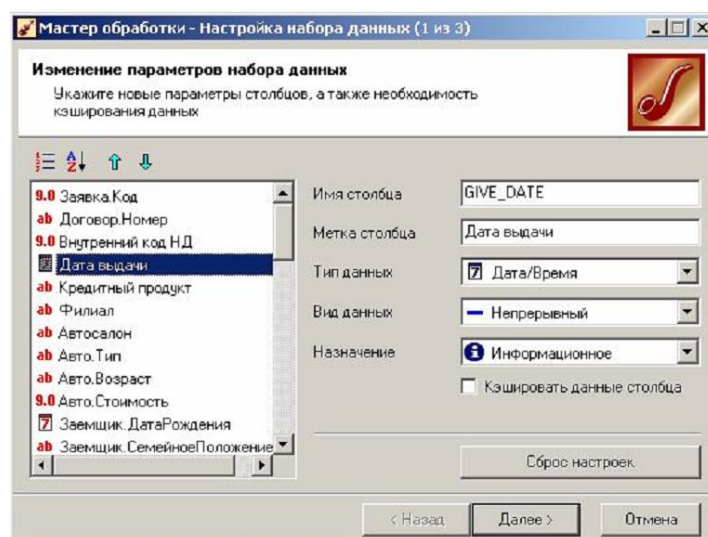



Рисунок 13. Настройка набора данных

Изменение имени или метки поля удобно в тех случаях, когда имена столбцов могут измениться в источнике данных или при перенастройке узлов верхних уровней. В этом случае в узле Настройка набора данных имя исходного столбца заменяется другим, на которое и настраиваются все дочерние узлы. После такой операции изменение имен полей на верхних уровнях не потребует перенастройки всех дочерних узлов в дереве сценариев.

Тип, вид и назначение можно изменить у нескольких столбцов одной операцией. Для этого достаточно их выделить, удерживая нажатой клавишу Ctrl или Shift.

Если параметры столбца были изменены, цвет иконки столбца меняется на красный. Для установки первоначальных параметров столбцов необходимо выделить столбец или список столбцов и нажать на кнопку Сброс параметров.

Чтобы скрыть столбец из набора данных, нужно задать ему назначение  Неиспользуемое.

Изменить порядок следования столбцов в наборе данных можно при помощи клавиш



Кэширование – это загрузка часто используемой информации в оперативную память для быстрого доступа к ней, минуя многократные считывания с жесткого диска. Кэширование может заметно повысить скорость работы сценария в ряде случаев (использование кэширования не входит в базовые навыки работы с Deductor).

Экспорт в текстовый файл

Выполняется при помощи мастера экспорта. В нем процесс экспорта данных в текстовый файл с разделителями (категория Файлы) содержит следующие шаги:

- настройка форматов экспорта;
- указание символа-разделителя столбцов;
- выбор экспортируемых полей;
- запуск процесса экспорта;
- выбор способа визуализации;
- задание сведений об узле.

На шаге Настройка параметров экспорта задаются параметры экспорта данных из текстового файла аналогично тем, что задавались в мастере импорта. Экспортироваться будут не все поля, а только те, у которых поднят флажок на шаге Выбор экспортируемых полей (рис 14):

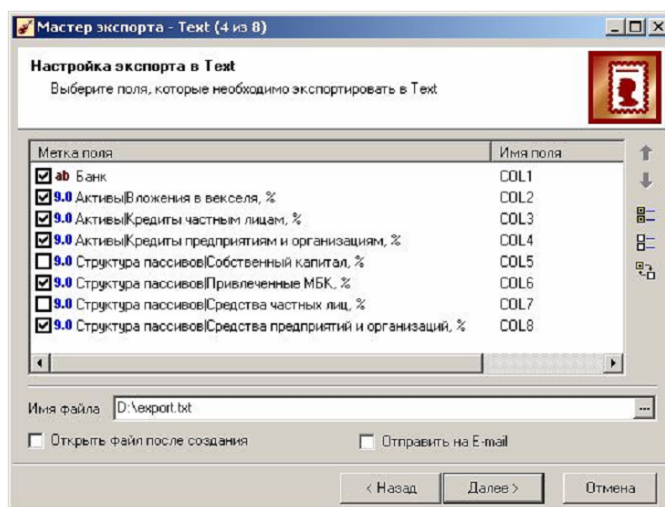


Рисунок 14. Настройка параметров экспорта

Здесь же задается имя файла экспорта. По умолчанию предлагается имя файла export.txt. Как и в случае с импортом, допускается использовать относительные пути.


Флажок Открыть после создания откроет текстовый файл программой – просмотрщиком.

Установка флага Отправить на E-mail позволит отправить файл на почтовый адрес в виде вложенного в письмо файла (доступно только в версии Enterprise).


Вкладка запуска процесса экспорта аналогична той, что присутствовала в мастере импорта.

Если файл с таким именем уже существует, возникнет окно с подтверждением о перезаписывании этого файла. На шаге выбора способа визуализации будет доступен только один визуализатор Прочее. Задание сведений об узле завершит работу мастера экспорта.

Узел «!»

При открытии сценария некоторые узлы могут иметь значок , а при попытке выполнить ветвь узла появится, например, следующее сообщение:



Значок  говорит о том, что выполнить данный узел невозможно. Причинами этого могут быть:

- Узел не поддерживается в текущей поставке Deductor. Например, узлы импорта из 1С не выполняются в Deductor Academic.
- Сценарий создан в более поздней версии (сборке) Deductor, а открыт в более ранней, и функционал такого узла еще не существовал в ранних версиях. Номер сборки можно проверить, открыв меню? → О программе.

Базовые визуализаторы

К каждому узлу сценария, который содержит структурированный набор данных, всегда предлагается несколько визуализаторов. Мастер визуализации в интерактивном пошаговом режиме позволяет выбрать и настроить наиболее удобный способ представления данных. В зависимости от выбранного способа будут настраиваться различные параметры, а Мастер, соответственно, будет содержать различное число шагов. Первый шаг мастера визуализации будет одинаков для всех видов, поскольку на нем и производится выбор визуализатора (рис 15).

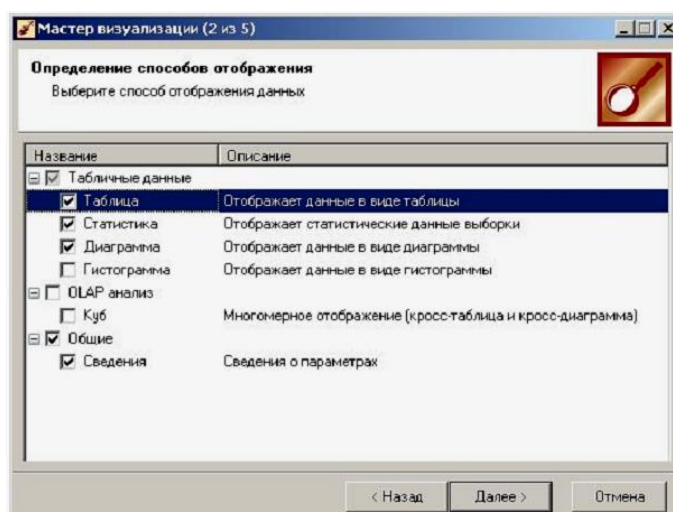



Рисунок 15. Мастер визуализации

Вызов мастера визуализации:

- кнопка  на панели инструментов закладки Сценарии;
- клавиша F5;
- контекстное меню Мастер визуализации...

Мастер визуализации запускается для выделенного узла сценария. Кроме того, этот мастер всегда является продолжением мастера обработки, т.е. активизируется при создании (настройке) любого узла.

Желаемые способы отображения следует пометить флажками. Одновременно может быть выбрано несколько визуализаторов, при этом каждый из них будет открыт в отдельном окне.

Замечание

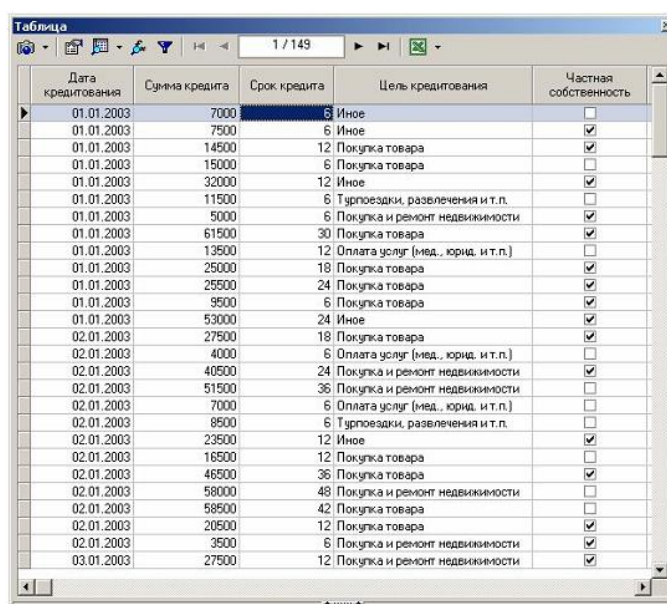
Если на первом шаге мастера визуализации одновременно выбрано несколько способов отображения данных, то все соответствующие шаги будут последовательно включены в общую процедуру настройки. Например, если выбраны Таблица и Диаграмма, то в мастер визуализации будут последовательно включены отдельные шаги для настройки таблицы и диаграммы.

Базовыми визуализаторами в Deductor являются следующие:

- Таблица;
- Статистика;
- Сведения.

Визуализатор Таблица

В таблице каждое поле набора данных размещается в отдельном столбце. Столбцы озаглавлены метками полей, а если метка не была задана, то именами полей. Ширину и порядок столбцов можно менять при помощи мыши.





Дата кредитования	Сумма кредита	Срок кредита	Цель кредитования	Частная собственность
01.01.2003	7000	3	Иное	<input type="checkbox"/>
01.01.2003	7500	6	Иное	<input checked="" type="checkbox"/>
01.01.2003	14500	12	Покупка товара	<input checked="" type="checkbox"/>
01.01.2003	15000	6	Покупка товара	<input type="checkbox"/>
01.01.2003	32000	12	Иное	<input checked="" type="checkbox"/>
01.01.2003	11500	6	Турпоездки, развлечения и т.п.	<input type="checkbox"/>
01.01.2003	5000	6	Покупка и ремонт недвижимости	<input checked="" type="checkbox"/>
01.01.2003	61500	30	Покупка товара	<input checked="" type="checkbox"/>
01.01.2003	13500	12	Оплата услуг (мед., юрид. и т.п.)	<input type="checkbox"/>
01.01.2003	25000	18	Покупка товара	<input checked="" type="checkbox"/>
01.01.2003	25500	24	Покупка товара	<input checked="" type="checkbox"/>
01.01.2003	9500	6	Покупка товара	<input checked="" type="checkbox"/>
01.01.2003	53000	24	Иное	<input checked="" type="checkbox"/>
02.01.2003	27500	18	Покупка товара	<input checked="" type="checkbox"/>
02.01.2003	4000	6	Оплата услуг (мед., юрид. и т.п.)	<input type="checkbox"/>
02.01.2003	40500	24	Покупка и ремонт недвижимости	<input checked="" type="checkbox"/>
02.01.2003	51500	36	Покупка и ремонт недвижимости	<input type="checkbox"/>
02.01.2003	7000	6	Оплата услуг (мед., юрид. и т.п.)	<input type="checkbox"/>
02.01.2003	8500	6	Турпоездки, развлечения и т.п.	<input type="checkbox"/>
02.01.2003	23500	12	Иное	<input checked="" type="checkbox"/>
02.01.2003	16500	12	Покупка товара	<input type="checkbox"/>
02.01.2003	46500	36	Покупка товара	<input checked="" type="checkbox"/>
02.01.2003	58000	48	Покупка и ремонт недвижимости	<input type="checkbox"/>
02.01.2003	58500	42	Покупка товара	<input type="checkbox"/>
02.01.2003	20500	12	Покупка товара	<input checked="" type="checkbox"/>
02.01.2003	3500	6	Покупка и ремонт недвижимости	<input checked="" type="checkbox"/>
03.01.2003	27500	12	Покупка и ремонт недвижимости	<input checked="" type="checkbox"/>

Рисунок 16. Таблица











В таблице можно настроить объединение заголовков столбцов. Например, есть два заголовка Продажи Сумма и Продажи Количество. Если переименовать (например, с помощью обработчика Настройка набора данных) метку первого столбца в Продажи|Сумма, а второй – в Продажи|Количество, то получим объединение заголовка в шапке таблицы.

Продажи	
Сумма	Количество


Символ«|» подсказывает визуализатору место в слове, где заканчивается общее название у двух заголовков.

Последовательное нажатие левой кнопкой мыши по заголовку активирует сортировку по данному столбцу в следующем порядке: сортировка по возрастанию  – сортировка по убыванию  – исходное состояние. Столбцы логического типа показываются в виде флажков.


В верхней части окна таблицы представлена панель инструментов, кнопки которой открывают доступ к следующим функциям.

	Функция	Горячая клавиша	Описание
	управление конфигурациями	–	Сохранение и восстановление конфигураций отображения таблицы
	настройка полей	<F11>	Позволяет настраивать видимость полей, отображаемых в таблице, а также задавать их формат и способ выравнивания
	способ отображения	<Ctrl+F12>	Переключение между отображением данных в виде таблицы или в виде формы
	статистика	–	Позволяет посмотреть статистику по текущим данным таблицы. Аналогично визуализатору <i>Статистика</i> , но открывается внизу таблицы, а не в отдельном окне
	фильтрация	<Ctrl+D>	Позволяет выполнять фильтрацию записей в таблице по заданным условиям
	первая запись	<Ctrl+PgUp>	Переход на первую запись набора данных
	предыдущая запись	<PgUp>	Переход на предыдущую запись набора данных
28 / 149	номер строки	–	Индикатор текущей записи
	следующая запись	<PgDn>	Переход на следующую запись набора данных
	последняя запись	<Ctrl+PgDn>	Переход на последнюю запись набора данных
	экспорт	–	Вызывается окно выбора файла для экспорта данных из таблицы в один из доступных текстовых форматов: <i>MS Excel</i> , <i>RTF</i> , <i>HTML</i> , <i>TXT</i> , <i>CSV</i> . По умолчанию предлагается экспорт в <i>MS Excel</i> . В версии <i>Academic</i> доступны не все форматы экспорта.

Однажды настроенный вид таблицы (к примеру, с различными фильтрами, форматами и видимостью столбцов и т.п.) можно сохранить, чтобы впоследствии быстро вернуться к нему.

Для этого в раскрывающемся по кнопке  списке нужно выбрать пункт Сохранить конфигурацию... и далее ввести ее название. Загрузить новую конфигурацию, можно, выбрав ее из списка конфигураций.

При вызове настройки полей появляется соответствующее диалоговое окно. В нем можно скрыть или сделать видимыми различные поля таблицы, определить способ выравнивания содержимого, ширину поля, а также задать формат отображения числовых данных и дат (рис 17).

Кнопка  переключает способ отображения набора данных, который может быть не только табличным, но и в виде формы. Это удобно, когда набор данных содержит большое количество столбцов.

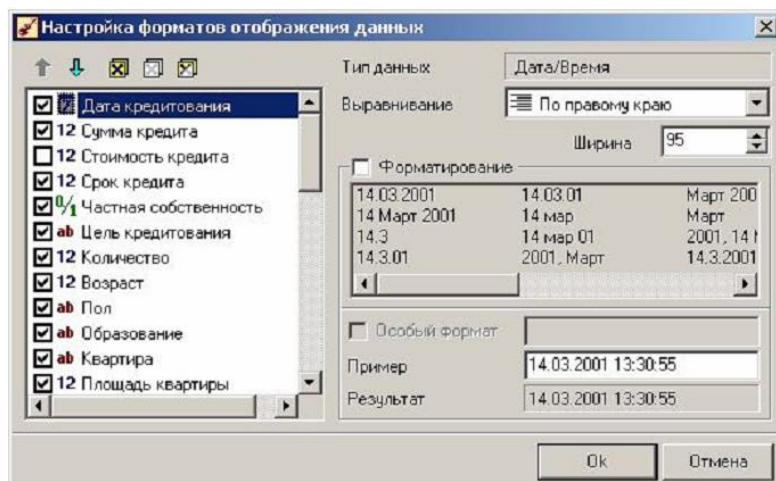





Рисунок 17.

Кнопка  открывает окно настройки условий фильтрации на набор данных. При включенном фильтре цвет кнопки меняется на , а цвет заголовков столбцов, которые участвуют в фильтре, изменяется на красный:

Дата (Год + Месяц)	Количество
2002-M01	355000
2002-M02	340000
2002-M03	405000
2002-M04	452000
2002-M05	464000
2002-M06	437000

Кнопка  открывает визуализатор Статистика, но не в отдельном вкладке, а в нижней части визуализатора Таблица.

Визуализатор Статистика

Статистика служит для отображения основных статистических характеристик набора данных конкретного узла.

Статистические характеристики отображаются в таблице по каждому полю выборки. В верхней части окна статистики отображается общее количество записей в наборе данных. Панель инструментов окна статистики позволяет управлять отображением статистических характеристик (среднее, минимум, максимум и т.п.) с помощью группы кнопок (рис 18).

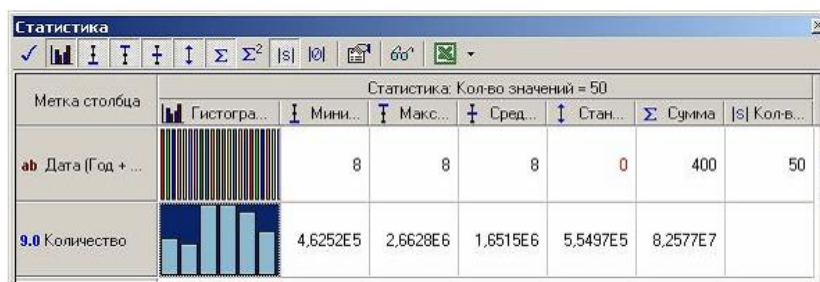



Рисунок 18. Статистика

Для полей дискретного типа, кроме прочих, всегда рассчитываются следующие статистические показатели:

- количество уникальных значений,
- количество пустых значений.

Просмотреть список уникальных значений можно следующими способами:

- двойной щелчок по ячейке Количество уникальных значений или по ячейке Гистограмма,

- кнопка  Обзор статистики.

Для поля непрерывного типа в обзоре статистики строится гистограмма распределения частот, она же в уменьшенном виде всегда показывается в соответствующем столбце (рис 19).

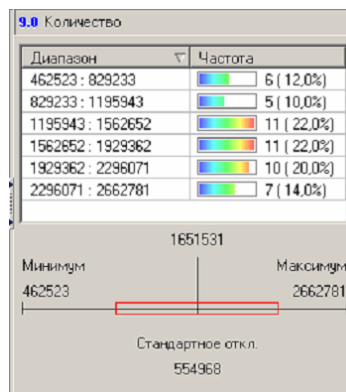


Рис 19.

Визуализатор Сведения

Визуализатор Сведения позволяет просмотреть все параметры, с которыми был выполнен тот или иной процесс преобразования данных, в результате которого была сформирована новая выборка: импорт, обработка одним из методов или экспорт. Такими параметрами являются время и длительность выполняемого процесса, условия остановки, наличие первичного ключа, ограничители столбцов, разделители целой и дробной частей чисел, элементов даты и т.д (рис 20).

Предусмотрено два вида представления описания: в виде дерева и текстовый. По умолчанию устанавливается вид дерева.

Узел	
Имя	146
Метка	Данные по продажам
Описание	
Объект	
Максимальное время выполнения	0
Время выполнения (мс)	31
Начало процесса	2007.09.03 11:31:45
Конец процесса	2007.09.03 11:31:46
Время выполнения	0:00:00
Процесс остановлен по условию останова	False
Процесс остановлен пользователем	False
Текстовый файл	..\Samples\TradeSales.txt
Добавить первичный ключ	False
Разделитель столбцов	Табуляция
Ограничитель строк	""
Считать последовательные разделители одним	False

Рисунок 20. Визуализатор Сведения

Визуализатор в основном предназначен для оперативного анализа текущих настроек узлов и для поиска возможных ошибок.

Визуализатор Сведения является единственно доступным для узлов экспорта.

Задание и порядок выполнения практической работы №1

Задание 1:

1. Скачать и установить Deductor Studio Academic

Задание 2:

1. Создать новый проект и сохранить его под именем **test.ded**. Не использовать упакованный формат файла!
2. Заполнить свойства проекта.
3. Просмотреть файл проекта через любой текстовый редактор.
4. Сделать видимой вкладку Подключения.

Задание 3.

1. Создать новый проект и сохранить его под именем **test2.ded**.
2. Создать и сохранить в любом текстовом редакторе файл следующего вида:
a,1,4.5,b,c,26/04/2007,d
a1,0,5,b1,c1,,d1
3. Импортировать его в Deductor, корректно настроив параметры импорта. Использовать относительный путь для файла. Метку узла переименовать в **Пример импорта файла**. В комментарии к узлу вписать: **Текстовый файл с разделителями-запятыми**.
4. Добавить к узлу узел **Настройка набора данных** и задать следующие метки к столбцам: Поле1, Поле2, Поле3 и т.д.
5. Экспортировать набор данных в текстовый файл с настройками, предлагаемыми по умолчанию.
6. Импортировать только что экспортированный файл в Deductor.
7. Присоединить к новому узлу импорта (путем копирования) предыдущую ветвь, начиная с узла **Настройка набора данных**.
8. Между экспортом и настройкой набора данных вставить еще один узел настройки, в котором изменить тип столбца Поле2 на логический.
9. Удалить только что вставленный узел.
10. Сохранить проект.

Задание 4.

1. В проекте, созданном в Задании 3, настроить следующие визуализаторы к любому узлу импорта: Таблица, Статистика. Перейти в режим формы и обратно. Имеются ли пропуски в записях?
2. В визуализаторе Таблица настроить, чтобы при отображении к значениям в Поле3 добавлялось слово «кг.». Сохранить конфигурацию визуализатора под названием «K1».
3. Сделать первые три столбца невидимыми. Сохранить конфигурацию визуализатора под названием «K2».
4. Вернуться к конфигурации K1.
5. В визуализаторе Таблица установить фильтр «Полеб = не пустой». Удалить фильтр.

Задание 5.

Ответить на контрольные вопросы

Контрольные вопросы

1. Из каких частей состоит Deductor?
2. Сколько категорий пользователей Deductor можно выделить?
3. Перечислите функции аналитика.
4. Кто обычно занимается проектированием и наполнением хранилища данных?
5. Сколько вкладок на панели управления Deductor Studio?
6. Что такое проект в Deductor Studio?
7. Какое расширение имеет файл проекта?
8. Как создать новый проект?
9. Как сохранить текущий проект под другим именем?
10. Как отредактировать свойства проекта?
11. Сколько проектов можно одновременно открыть в Deductor Studio?
12. Сколько мастеров имеется в Deductor Studio?
13. Какие характеристики набора данных показывает визуализатор Статистика?
14. Что означает красный заголовок столбца в визуализаторе Таблица?
15. Как обнаружить, имеются ли в столбце пропущенные значения?
16. Для чего предназначен визуализатор Сведения?
17. Как скрыть столбец в визуализаторе Таблица?