

Практическая работа № 2

Заполнение хранилищ данных с помощью Data Transformation Services

Цель работы: изучение приемов переноса, копирования и преобразования данных в SQL-серверных СУБД.

Задание: используя утилиту Data Transformation Services (DTS) СУБД MS SQL Server создать пакет для переноса данных из оперативной БД в хранилище:

- 1) описать источники данных;
- 2) описать потоки данных, порядок и условия их выполнения;
- 3) описать правила преобразования данных;
- 4) создать правила выполнения пакета по расписанию.

2.1 Основные положения

Проектирование хранилища и создание базы данных соответствующей структуры – лишь первый шаг к созданию хранилища данных. Далее следует позаботиться о том, чтобы таблицы этого хранилища были заполнены данными, соответствующими текущему состоянию оперативной базы данных.

Data Transformation Services (DTS) – это набор служб SQL Server, предназначенных для организации импорта, экспорта, преобразования данных и переноса их между любыми источниками, доступными через интерфейсы OLE DB. С помощью DTS можно копировать структуры данных и сами данные из одной базы данных в другую, создавать средства для переноса данных, встроенные в приложения, а также пополнять хранилища данных из разнообразных источников, которые вовсе не обязательно должны быть базами данных SQL Server.

Для заполнения хранилища данных обычно требуется создать и выполнить так называемый пакет DTS (DTS package), содержащий описание последовательности всех действий, которые следует выполнить при переносе данных (включая преобразование типов

данных, выполнение SQL-запросов и т.д.). Такой пакет можно выполнить с помощью SQL Server Enterprise Manager или утилиты dtsrun, сохранить его в службах метаданных Meta Data Services (в прежних версиях SQL Server это хранилище называлось репозитарием) либо в виде структурированного файлового хранилища. Также возможно программное выполнение DTS-пакетов с помощью свойств и методов соответствующих объектов SQL DMO – для этого можно автоматически сгенерировать код на языке Visual Basic. В SQL Server 2000 также поддерживается возможность сохранения DTS-пакетов в формате XML. Рассмотрим процесс создания пакета DTS, заполняющего хранилище «КИС_ХД» данными из оперативной базы данных «КИС».

2.2 Порядок выполнения

2.2.1 Описание источников данных


Создать пакет DTS можно с помощью соответствующего редактора — DTS package editor. Для его запуска следует с помощью SQL Server Enterprise Manager соединиться с сервером, содержащим хранилище данных, найти в разделе Data Transformation Services элемент Meta Data Service Packages и выбрать опцию New Package из его контекстного меню.

Далее требуется описать базу данных, в которой находится хранилище данных. Для этого необходимо перенести на рабочее пространство редактора пакетов DTS пиктограмму Microsoft OLE DB Provider for SQL Server с палитры Data tool в левой части окна редактора. После этого появится диалоговая панель Connection Properties для описания источников данных OLE DB. В этом окне присвоим этому источнику данных имя КИС, укажем в качестве источника данных базу данных КИС и параметры доступа к ней (например, Use Windows NT authentication). Так как в хранилище данных содержится четыре таблицы, и потребуется четыре отдельных операции по их заполнению, то сделаем еще три копии этого же источника данных, повторив данную операцию три раза, но при этом отметив в диалоговых панелях Connection Properties опцию Existing Connection и выбирая имеющийся источник данных КИС. Тем же способом опишем приемник исходных данных – базу данных КИС_ХД, присвоим ему имя КИС_ХД и создадим еще одну его копию.

2.2.2 Описание потоков данных и последовательности выполнения задач

Execute SQL Task Properties

General

 You can run SQL code on the selected connection. You must select a connection and then provide the SQL code to execute.

Description: Очистка хранилища данных

Existing connection: КИС_ХД

Command time-out: 0

SQL statement:

```
TRUNCATE TABLE Продажи_факт
DELETE FROM Время_изм
DELETE FROM Сотрудники_изм
DELETE FROM Товары_изм
```

Parameters... Build Query...

Browse... Parse Query

OK Отмена Справка

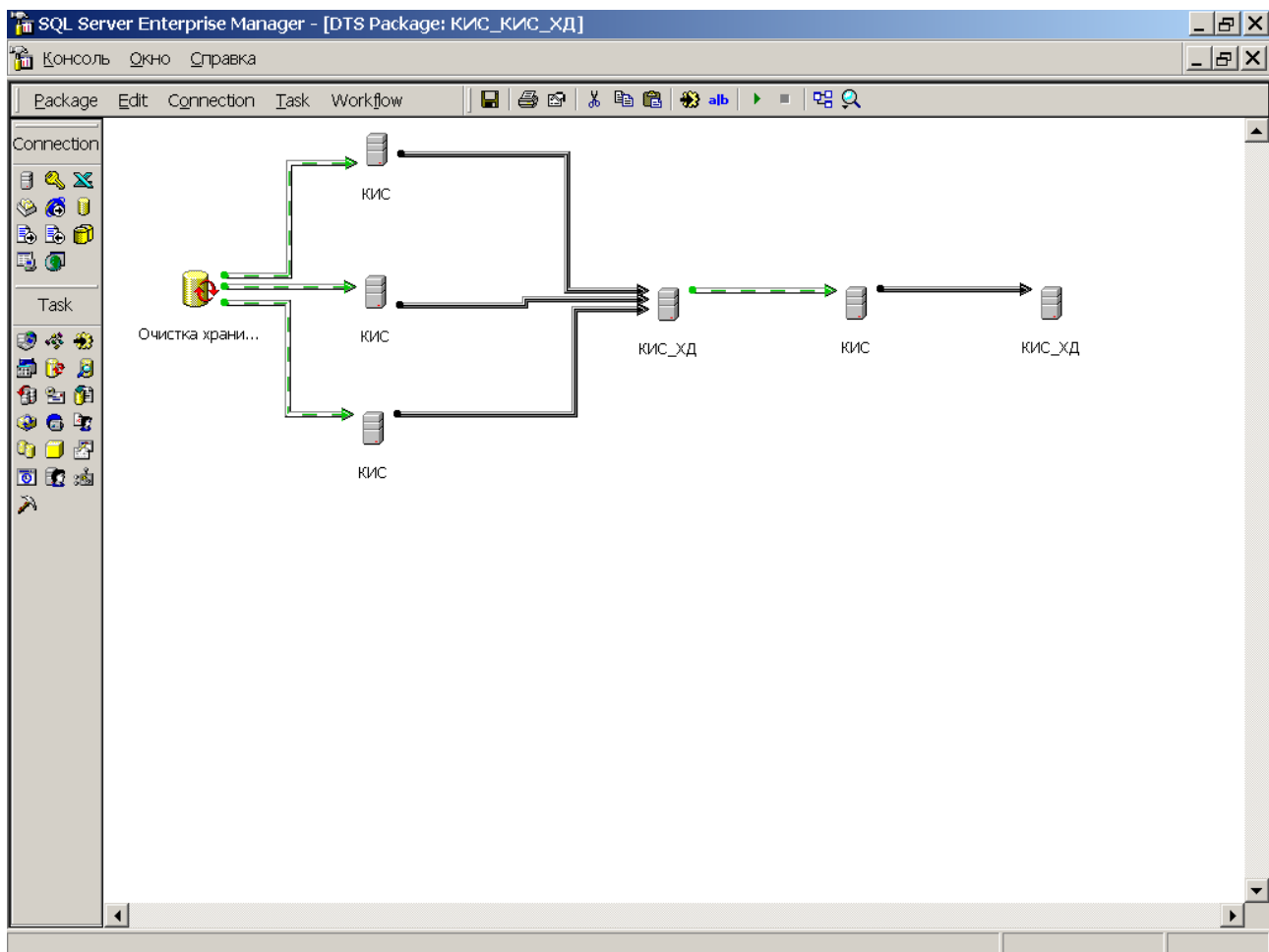
Перед заполнением таблиц в хранилище данных они будут полностью очищаться. Для этой цели следует перенести в рабочее пространство редактора пиктограмму Execute SQL Task. При этом на экране появится диалоговая панель Execute SQL Task Properties, в которой заполним поля Description (описание задачи) и SQL Statement (сюда добавим операторы для удаления данных из всех таблиц хранилища данных). При большом объеме данных удаление обычно не применяется – в этом случае к уже существующим данным добавляются новые.

Далее следует определить, какие потоки данных нужны для заполнения хранилища. С этой целью с помощью щелчков мыши при

нажатой клавише Ctrl следует выбрать один источник данных КИС и один из приемников данных КИС_ХД. Когда обе пиктограммы будут выделены, следует выбрать опцию Transform Data Task из контекстного меню источника данных КИС_ХД, и тогда пиктограммы окажутся соединенными стрелкой, соответствующей одной из задач преобразования и переноса данных. Далее повторим эту же операцию с двумя другими экземплярами источника данных КИС и с тем же самым экземпляром приемника данных КИС_ХД.

Таким образом, были созданы задания для переноса данных в три таблицы измерений нашего хранилища. Эти задачи могут выполняться параллельно, ведь таблицы измерений в нашем хранилище не связаны друг с другом. Однако они могут быть выполнены только после полной очистки всего хранилища. Чтобы описать это условие (такие условия определяются словосочетанием precedence constraint), следует одновременно выбрать пиктограмму Execute SQL Task и одну из уже задействованных пиктограмм источника данных КИС, а затем из контекстного меню источника данных КИС выбрать опцию Workflow/On Success. Появившаяся зеленая пунктирная стрелка между пиктограммами означает, что перенос данных в соответствующую таблицу изменений будет осуществлен только после успешного завершения очистки хранилища. Далее следует повторить это действие с оставшимися экземплярами источника данных КИС.

Задача заполнения данными таблицы фактов может быть выполнена только после того, как будут заполнены все таблицы измерений. Поэтому сначала следует создать задание Transform Data Task для оставшихся экземпляров КИС и КИС_ХД и указать условие ее выполнения Workflow/On Success только в случае успешного заполнения таблиц измерений.



2.2.3 Описание преобразования данных

Далее следует задать параметры выполнения всех четырех заданий Transform Data Task, т.е. описать, откуда берутся и как преобразовываются данные при переносе из оперативной базы данных КИС в хранилище КИС_ХД. Начнем с таблицы Сотрудники_изм. Для ее заполнения следует открыть окно свойств одного из заданий, соответствующих задачам заполнения таблиц измерений (т.е. свойств одной из трех стрелок). В появившемся окне заполнить поле Description, и выбрав опцию SQL Query, ввести текст SQL-запроса:

```
SELECT Сотрудники.ТН, Сотрудники.Фамилия+' '+Сотрудники.Имя+' '+Сотрудники.Отчество AS ФИО,
Отделы.Наименование AS Отдел, Должности.Должность,
Регионы.Регион
FROM Сотрудники INNER JOIN Отделы ON Сотрудники.Отдел =
Отделы.Код
INNER JOIN Должности ON Сотрудники.Должность =
Должности.Код
INNER JOIN Регионы ON Отделы.Регион = Регионы.Код
```

результат которого должен быть помещен в таблицу Сотрудники_изм (выбирается в качестве приемника данных на вкладке Destination), и на вкладке Nransformations проверим соответствие столбцов источника и приемника данных.

Transform Data Task Properties



Source

Destination

Transformations

Lookups

Options



Enter a table name or the results of a query as a data source.

Description:

Перенос сотрудников

Connection:

КИС

☐ Table / View:

[КИС].[dbo].[ProbeTimestamp]

☒ SQL query:

```
SELECT Сотрудники.ТН, Сотрудники.Фамилия + ' ' + Сотрудники.Имя  
+ ' ' + Сотрудники.Отчество AS ФИО, Отделы.Наименование AS Отдел,  
Должности.Должность, Регионы.Регион FROM Сотрудники INNER  
JOIN Отделы ON Сотрудники.Отдел = Отделы.Код INNER JOIN  
Должности ON Сотрудники.Должность = Должности.Код INNER  
JOIN Регионы ON Отделы.Регион = Регионы.Код
```

Parameters...

Preview...

Build Query...

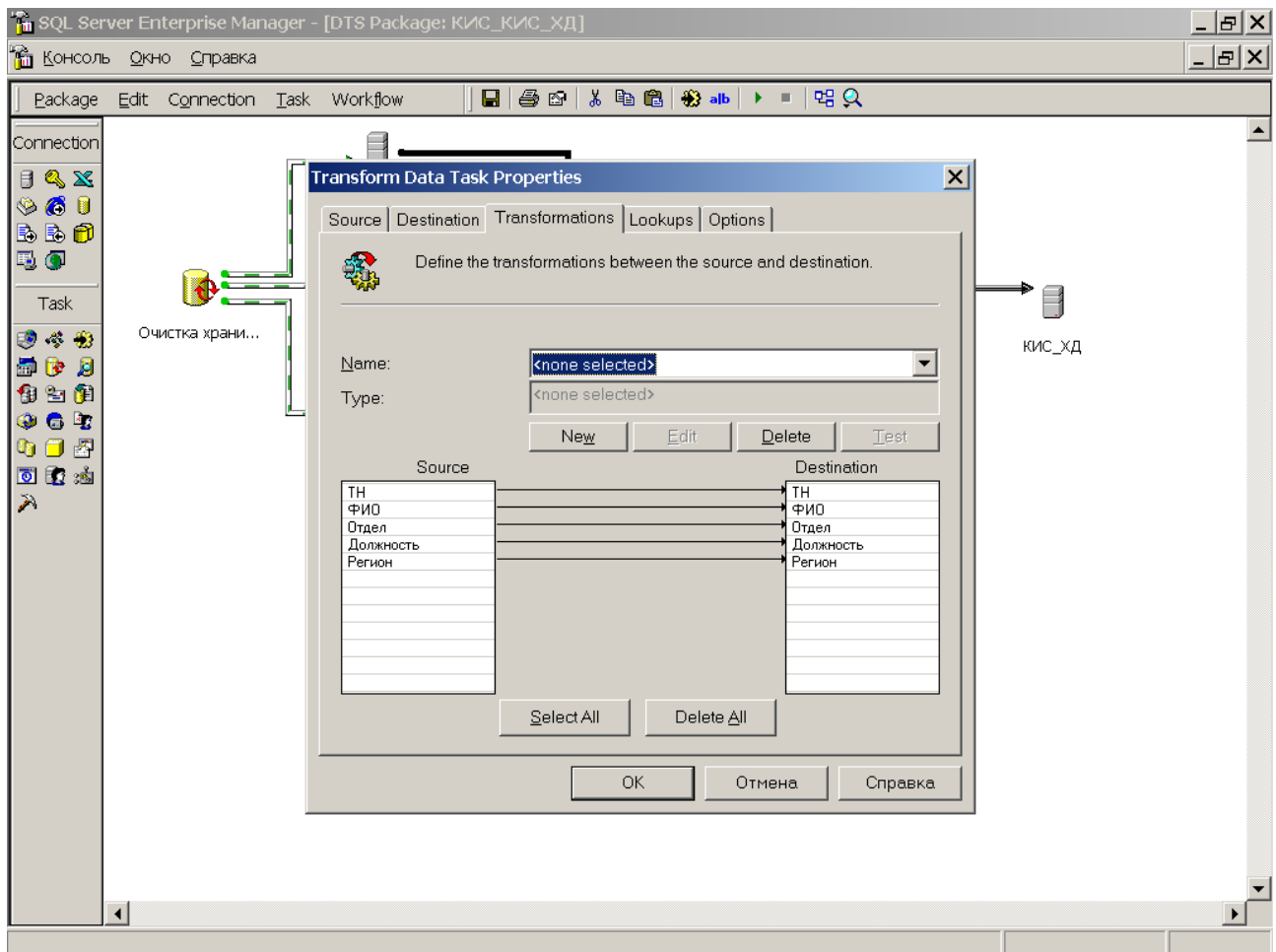
Browse...

Parse Query

OK

Отмена

Справка



Аналогично для следующего задания Transform Data Task, целью которого является заполнение таблицы Товары_изм параметры выполнения будут следующие:

Transform Data Task Properties



Source

Destination

Transformations

Lookups

Options



Enter a table name or the results of a query as a data source.

Description:

Перенос товаров

Connection:

КИС

☐ Table / View:

[КИС].[dbo].[ProbeTimestamp]

☒ SQL query:

```
SELECT Товары.Код, Товары.Наименование,  
        Производители.Наименование AS Производитель  
FROM Товары INNER JOIN Производители  
ON Товары.Производитель = Производители.Код
```

Parameters...

Preview...

Build Query...

Browse...

Parse Query

OK

Отмена

Справка

Transform Data Task Properties

Source | Destination | **Transformations** | Lookups | Options

Define the transformations between the source and destination.

Name: <none selected>

Type: <none selected>

New Edit Delete Test

Source	Destination
Код	Код
Наименование	Наименование
Производитель	Производитель

Select All Delete All

OK Отмена Справка

Для заполнения таблицы Время_изм следует открыть окно свойств последнего из заданий, соответствующих задачам заполнения таблиц измерений. Текст SQL-запроса, результат которого должен быть помещен в таблицу Время_изм:

```
SELECT DISTINCT Продажи.Дата AS Дата, DateName(dw,
Продажи.Дата) AS ДеньНедели,
DatePart(mm, Продажи.Дата) AS Месяц, DatePart(yy, Продажи.Дата)
AS Год,
DatePart(qq, Продажи.Дата) AS Квартал, DatePart(dy, Продажи.Дата)
AS ДеньГода,
case DatePart(dw, Продажи.Дата)
when (6) then 'Д'
when (7) then 'Д'
else 'Н' end AS Выходной,
```

```
DateName(month, Продажи.Дата)+' '+DateName(year, Продажи.Дата)
AS МесяцГод,
DatePart(wk, Продажи.Дата) AS НеделяГода
FROM Продажи WHERE Продажи.Дата IS NOT NULL
```

Transform Data Task Properties



Source

Destination

Transformations

Lookups

Options



Enter a table name or the results of a query as a data source.


Description:

Заполнение измерения времени

Connection:

КИС

☐ Table / View:

 [КИС].[dbo].[ProbeTimestamp]

☒ SQL query:

```
SELECT DISTINCT Продажи Дата AS Дата, DateName(dw, Продажи Дата) AS
DatePart(mm, Продажи Дата) AS Месяц, DatePart(yy, Продажи Дата) AS
DatePart(qq, Продажи Дата) AS Квартал, DatePart(dy, Продажи Дата) AS
case DatePart(dw, Продажи Дата)
when (6) then 'Д'
when (7) then 'Д'
else 'H' end AS Выходной,
DateName(month, Продажи Дата) AS М, DateName(year, Продажи Дата) AS
Г
```

Parameters...

Preview...

Build Query...

Browse...

Parse Query

OK

Отмена

Справка

Transform Data Task Properties

Source Destination Transformations Lookups Options

Define the transformations between the source and destination.

Name: <none selected>

Type: <none selected>

New Edit Delete Test

Source	Destination
Дата	Код
ДеньНедели	Дата
Месяц	ДеньНедели
Год	Месяц
Квартал	Год
ДеньГода	МесяцГод
Выходной	Квартал
МесяцГод	ДеньГода
НеделяГода	НеделяГода
	Выходной

Select All Delete All

OK Отмена Справка

Завершив работу с таблицами измерений, займемся преобразованием данных для таблицы фактов. В данном случае исходный набор данных, преобразуемый в таблицу Продажи_факт, представляет собой результат следующего запроса:

```

SELECT КИС_ХД.dbo.Время_изм.Код As Дата,
КИС_ХД.dbo.Сотрудники_изм.ТН,
КИС_ХД.dbo.Товары_изм.Код As Товар,
КИС.dbo.Продажи.Количество, КИС.dbo.Продажи.Сумма
FROM КИС.dbo.Продажи INNER JOIN КИС_ХД.dbo.Время_изм
ON КИС_ХД.dbo.Время_изм.Дата=КИС.dbo.Продажи.Дата
INNER JOIN КИС_ХД.dbo.Сотрудники_изм
ON КИС_ХД.dbo.Сотрудники_изм.ТН=КИС.dbo.Продажи.Сотрудник
INNER JOIN КИС_ХД.dbo.Товары_изм
ON КИС_ХД.dbo.Товары_изм.Код=КИС.dbo.Продажи.Товар

```

Transform Data Task Properties



Source

Destination

Transformations

Lookups

Options



Enter a table name or the results of a query as a data source.

Description:

Загрузка продаж

Connection:

КИС

☐ Table / View:

[КИС].[dbo].[ProbeTimestamp]

☒ SQL query:

```
SELECT КИС_ХД.dbo.Время_изм.Код AS Дата,  
       КИС_ХД.dbo.Сотрудники_изм.ТН,  
       КИС_ХД.dbo.Товары_изм.Код AS Товар,  
       КИС.dbo.Продажи.Количество, КИС.dbo.Продажи.Сумма  
FROM КИС.dbo.Продажи INNER JOIN КИС_ХД.dbo.Время_изм  
ON КИС_ХД.dbo.Время_изм.Дата=КИС.dbo.Продажи.Дата  
INNER JOIN КИС_ХД.dbo.Сотрудники_изм  
ON КИС_ХД.dbo.Сотрудники_изм.ТН=КИС.dbo.Продажи.Сотрудник
```

Parameters...

Preview...

Build Query...

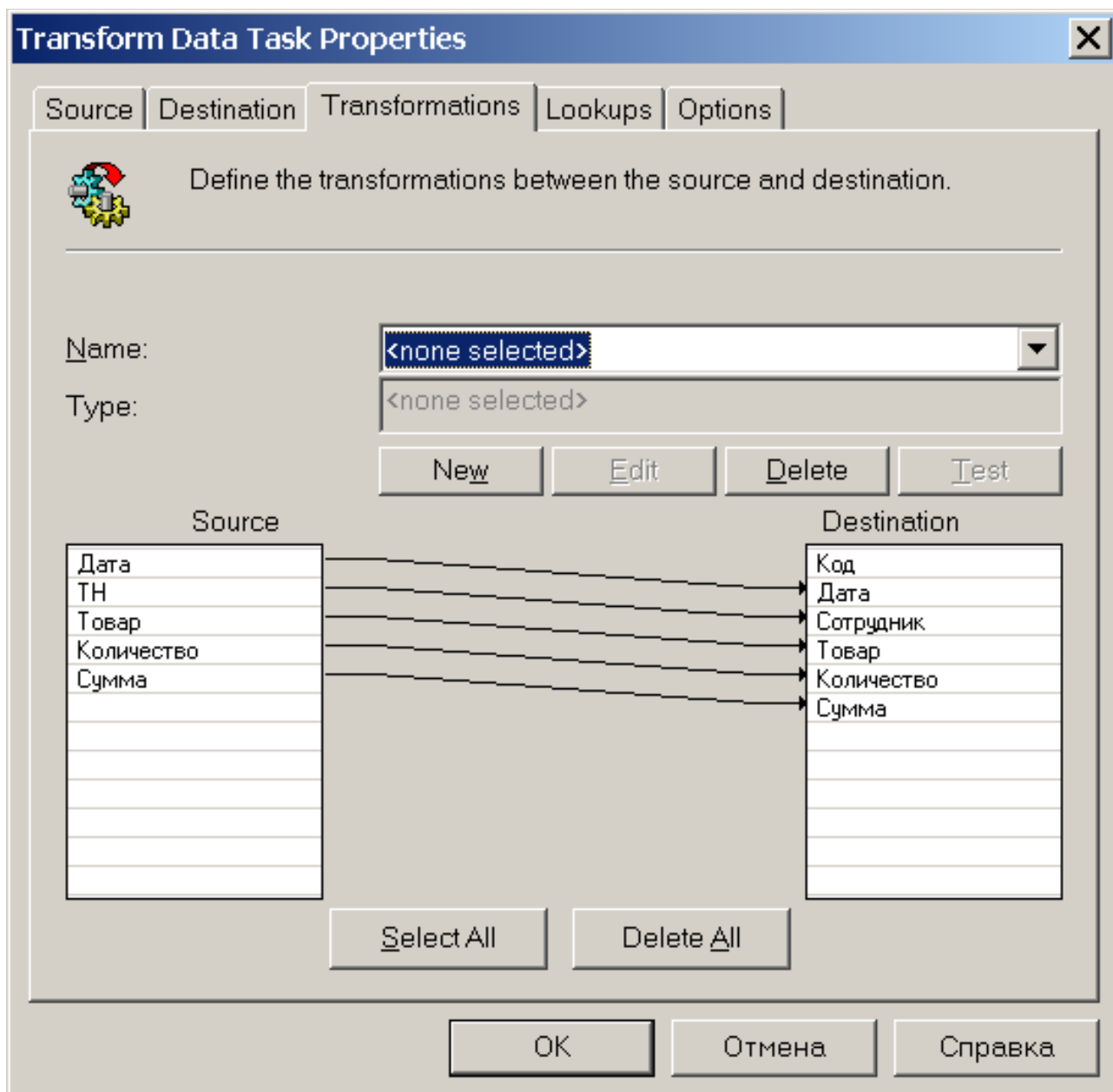
Browse...

Parse Query

OK

Отмена

Справка



Таким образом, были описаны все задания Transform Data Task, необходимые для заполнения хранилища данными.

2.2.4 Запуск на выполнение пакетов DTS

Созданный пакет DTS следует сохранить, выбрав опцию Package/Save из меню редактора пакетов DTS. Выполнить его можно, выбрав пункт меню Package/Execute. После этого начнется процесс преобразования данных и заполнения ими таблиц хранилища данных. Для того чтобы данные в хранилище соответствовали текущему или недавнему состоянию оперативной базы данных, можно создать расписание, согласно которому будет автоматически выполняться данный пакет. Для этого следует выбрать его в Enterprise Manager и опцию Schedule Package из контекстного меню. Далее следует выбрать нужный режим обновления данных в диалоговой панели Edit Recurring Job Schedule. Для запуска пакета по расписанию необходимо, чтобы был запущен

SQL Server Agent – служба, инициирующая выполнение различных заданий по расписанию.

Edit Recurring Job Schedule

Job name: NW_Load

Occurs

- ☒ Daily
- ☐ Weekly
- ☐ Monthly

Daily

Every 1 day(s)

Daily frequency

- ☐ Occurs once at: 0:00:00
- ☒ Occurs every: 1 Hour(s) Starting at: 0:00:00 Ending at: 23:59:59

Duration

Start date: 03.06.2001

- ☐ End date: 03.06.2001
- ☒ No end date

OK Cancel Help