

Учебный курс

Хранилища данных

Лекция 3

Создание куба в SQL Server 2005

Лекции читает

Кандидат технических наук, доцент

Перминов Геннадий Иванович

Цель:

методом SQL Server Business Intelligence Development Studio создать реляционное хранилище для предметной области «Индекс РТС».

Задачи:

- 1. построить хранилище для индекса РТС с простыми измерениями;
 - 1.1. создать решение для индекса РТС;
 - 1.2. произвести подключение к источнику данных;
 - 1.3. создание Data Source View;
- 2. произвести подключение к хранилищу дополнительной размерности типа «Время и Дата»;
 - 2.1. добавить измерение Date;
 - 2.2. произвести расчет измерения Date;
 - 2.3. произвести просмотр измерения Date;
- 3. произвести подключение к хранилищу размерности, созданной из таблицы фактов;
 - 3.1. добавить измерение Fact_Dim;
 - 3.2. рассчитать размерность из таблицы фактов;
 - 3.3. осуществить просмотр размерности из таблицы фактов.

Работа с ХД в новой версии сервера разделена на несколько шагов:

- создания решения (solution);
- подключение к источнику данных (data source);
- создание Data Source View;
- создание кубов и размерностей (рассматривается в теме 2);
- развертывание базы (deployment);
- администрирование базы.

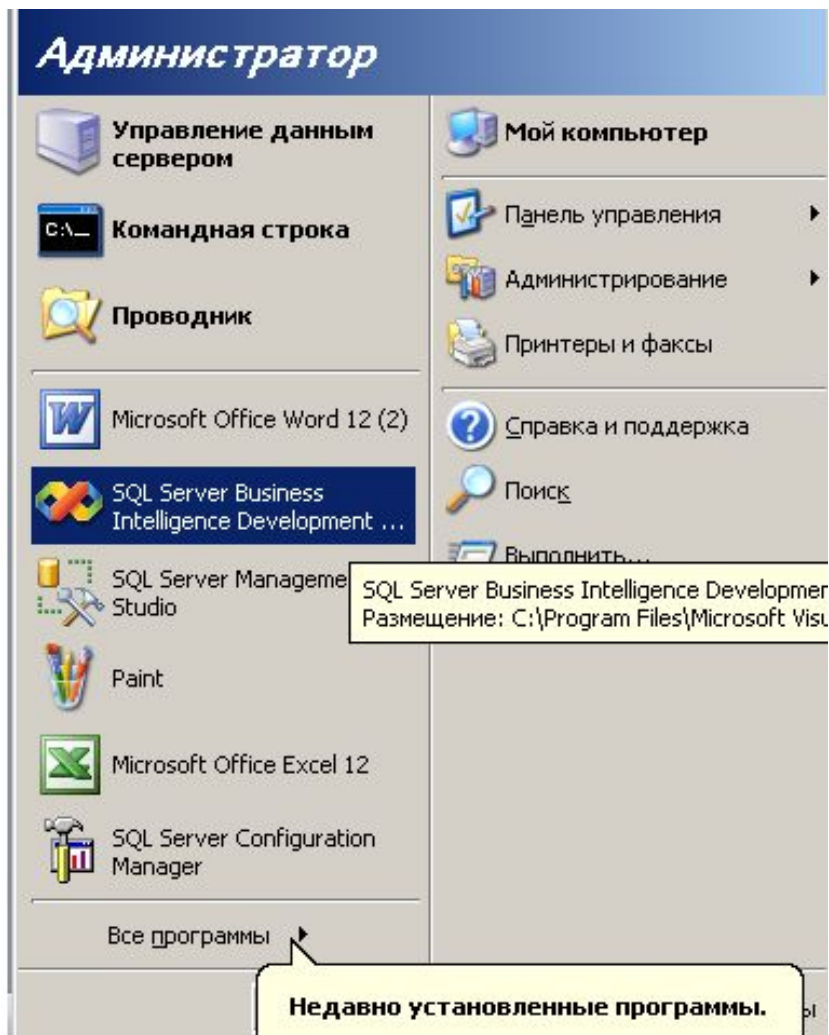
Для выполнения этих действий есть два различных инструментария:

- 1) разработка производится с использованием **SQL Server Business Intelligent Development Studio** (Среда разработки интеллектуальных систем предприятия);
- 2) в то время как сопровождение и администрирование выполняется с помощью **SQL Server Management Studio** (Среда сопровождения SQL-сервера).

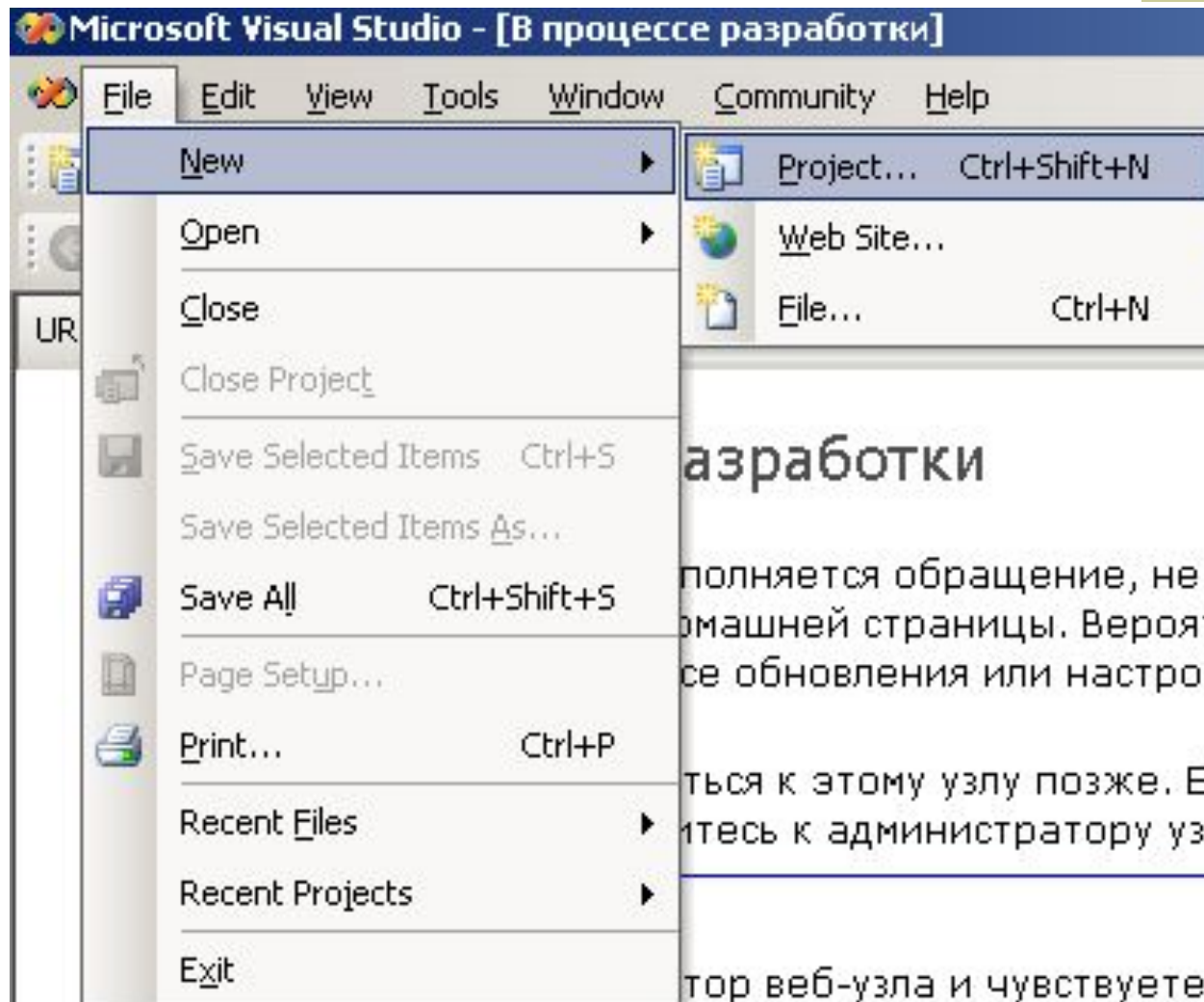


1. Создание решения

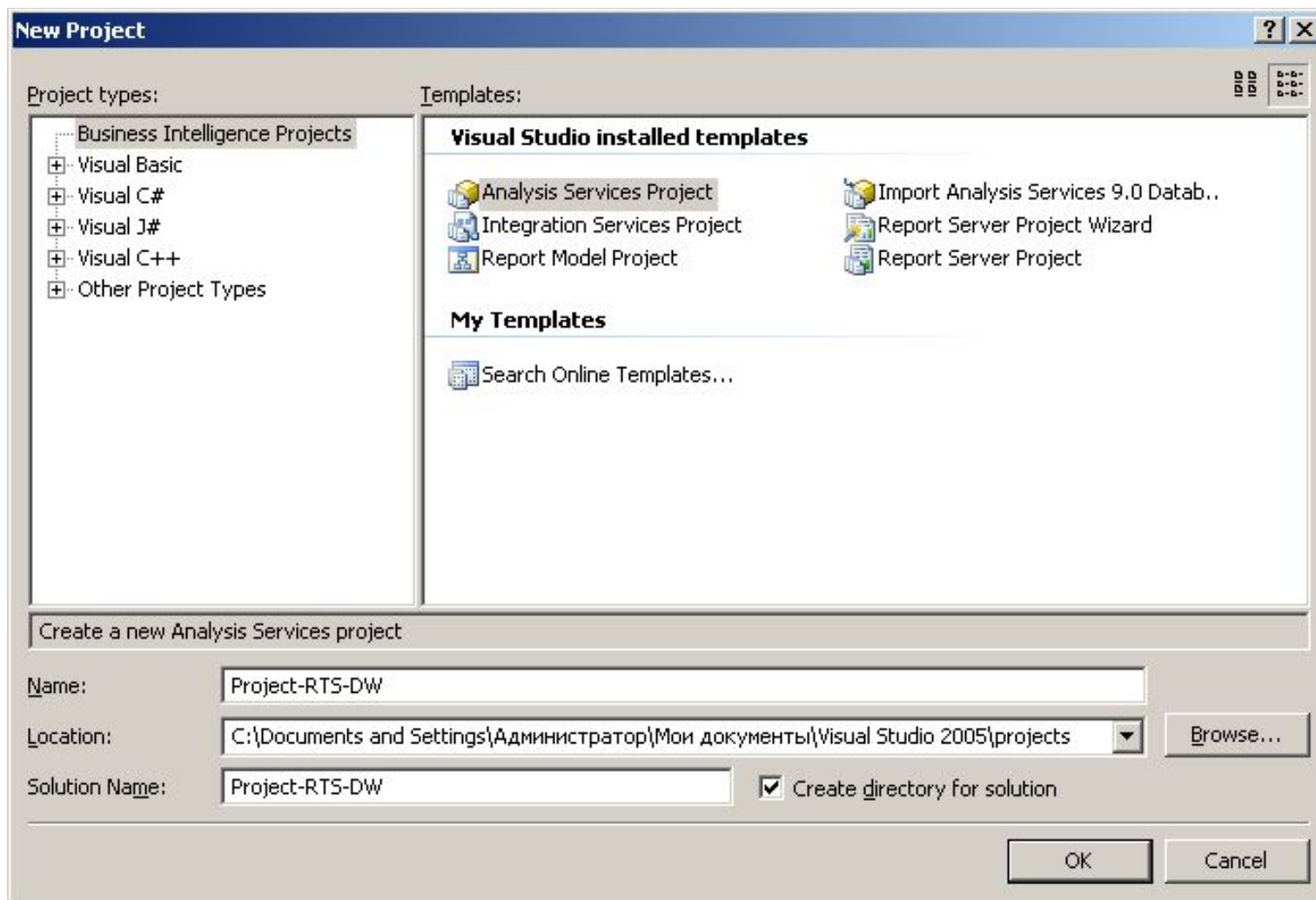
1.1. Вызов SQL серверной среды интеллектуальных систем предприятия




1.2. Команда создания нового проекта



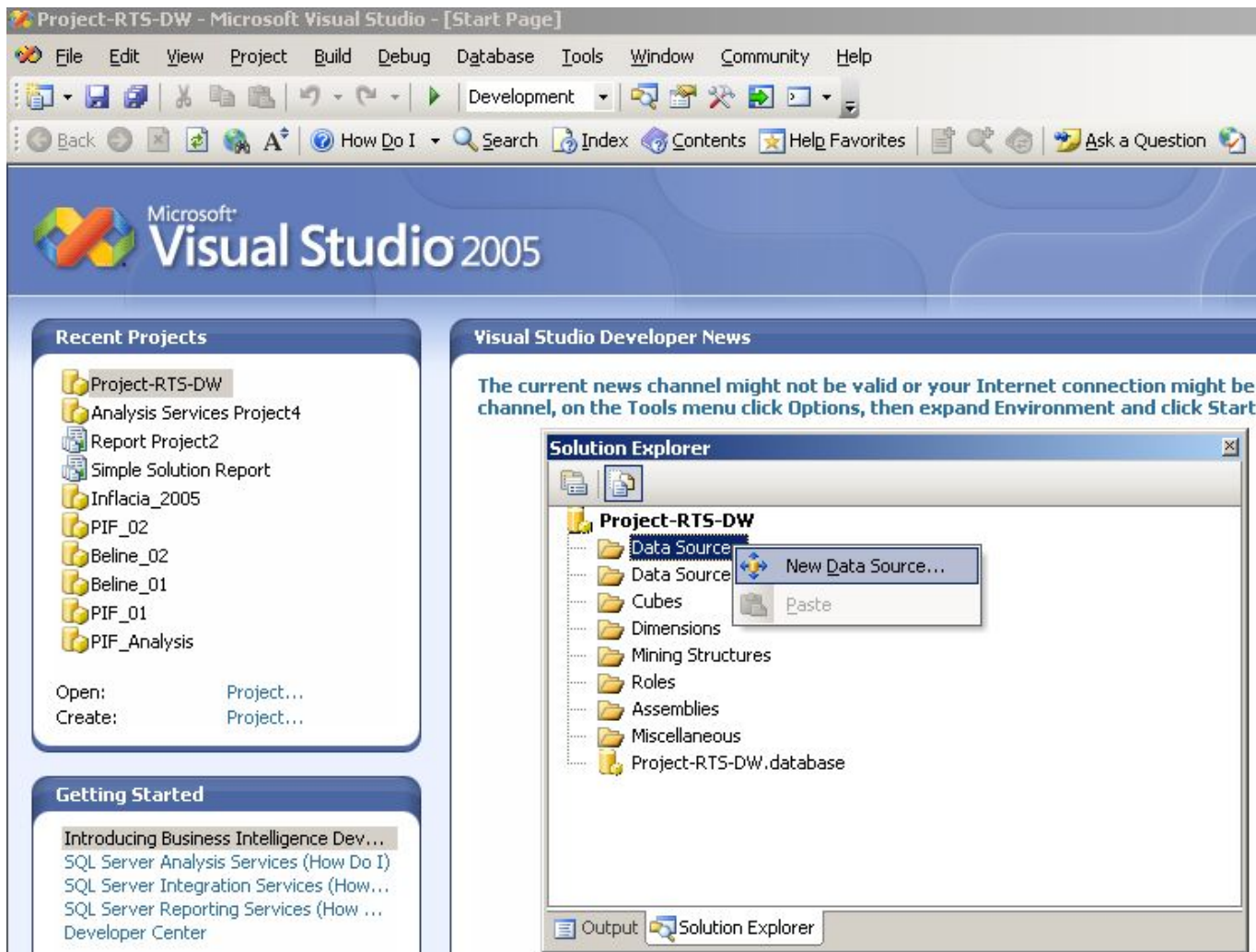
1.3. Выбор шаблона «Проект аналитических систем»



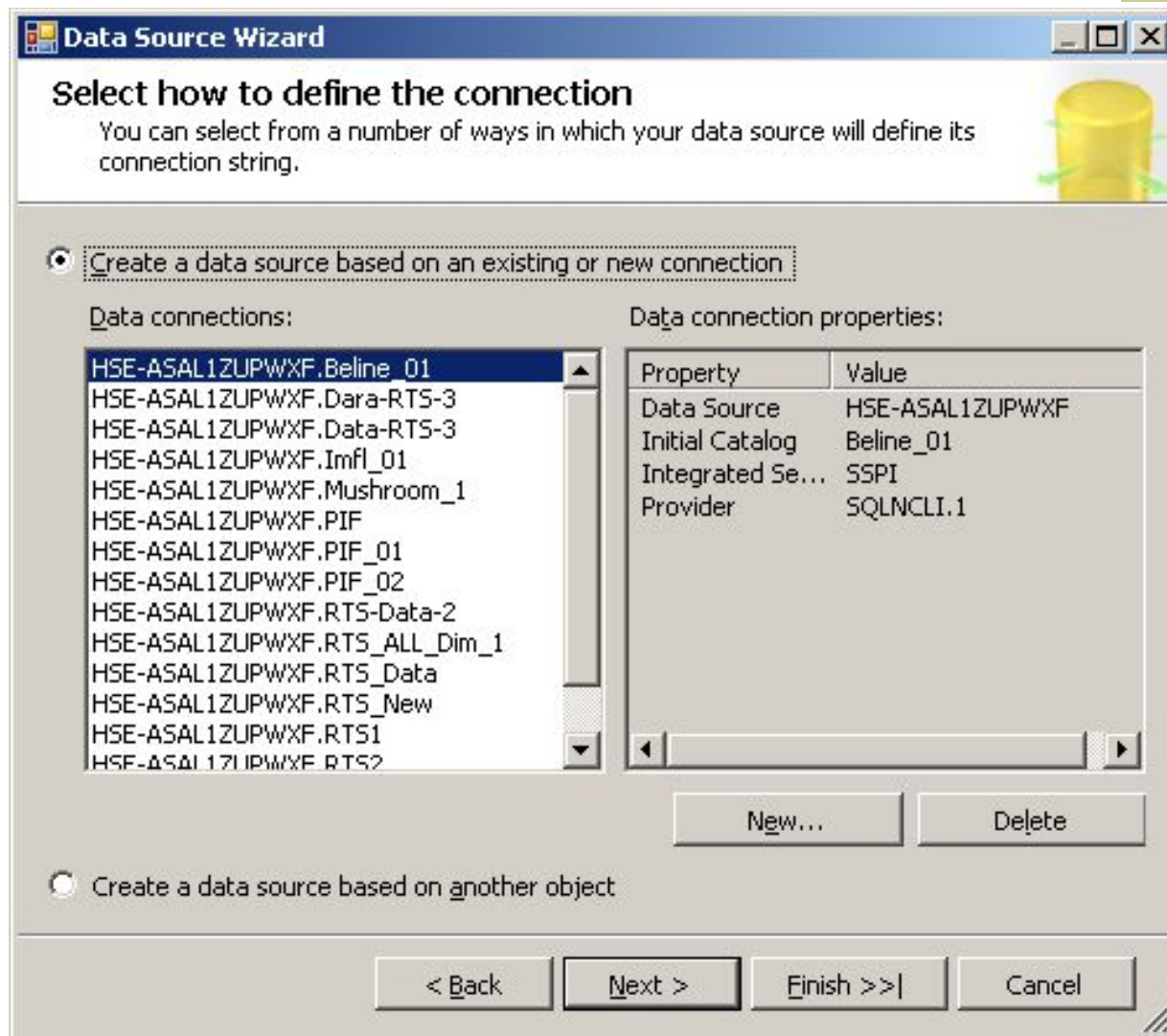


2. Подключение к источнику данных (Data Source)

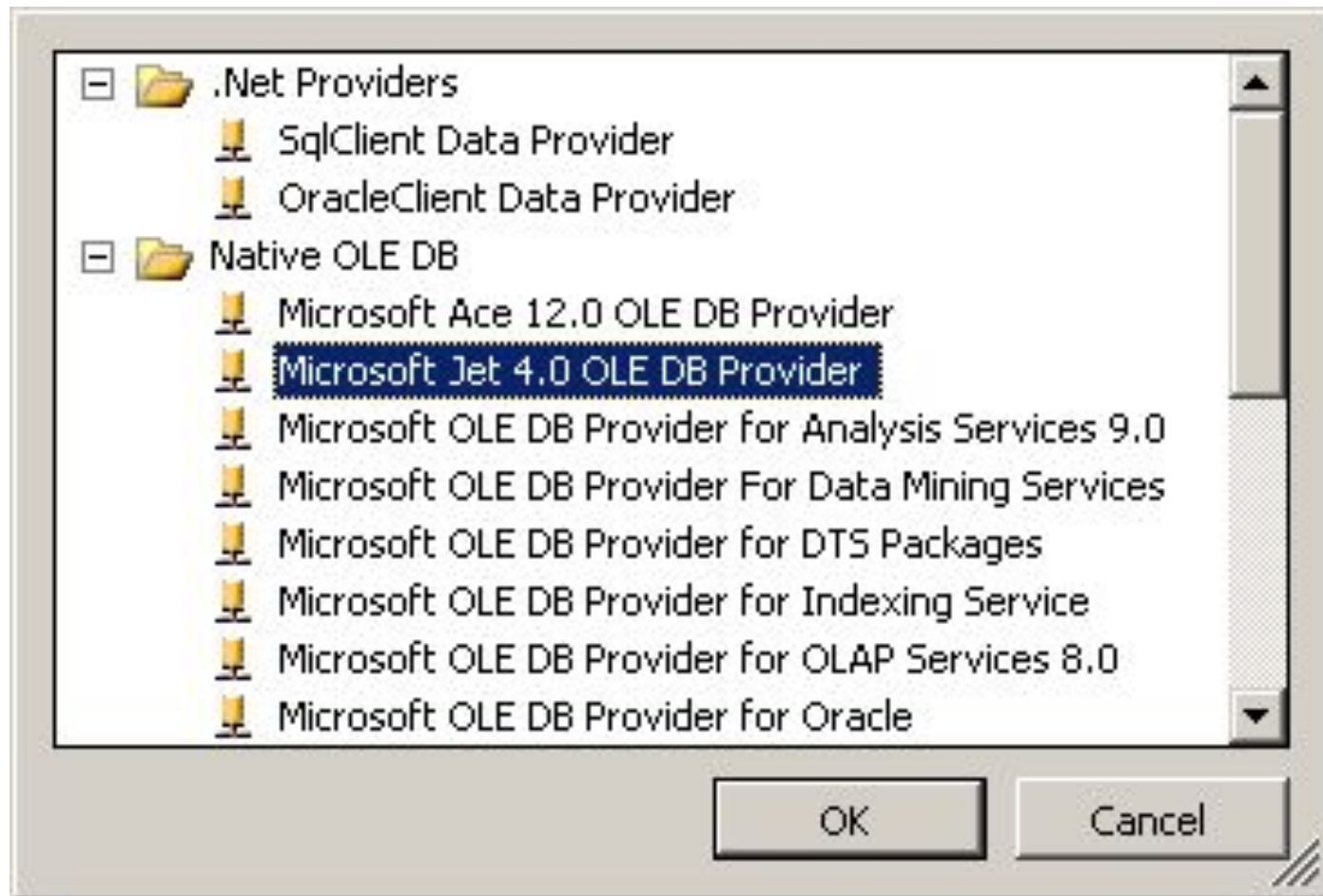
2.1. Команды выбора нового источника данных



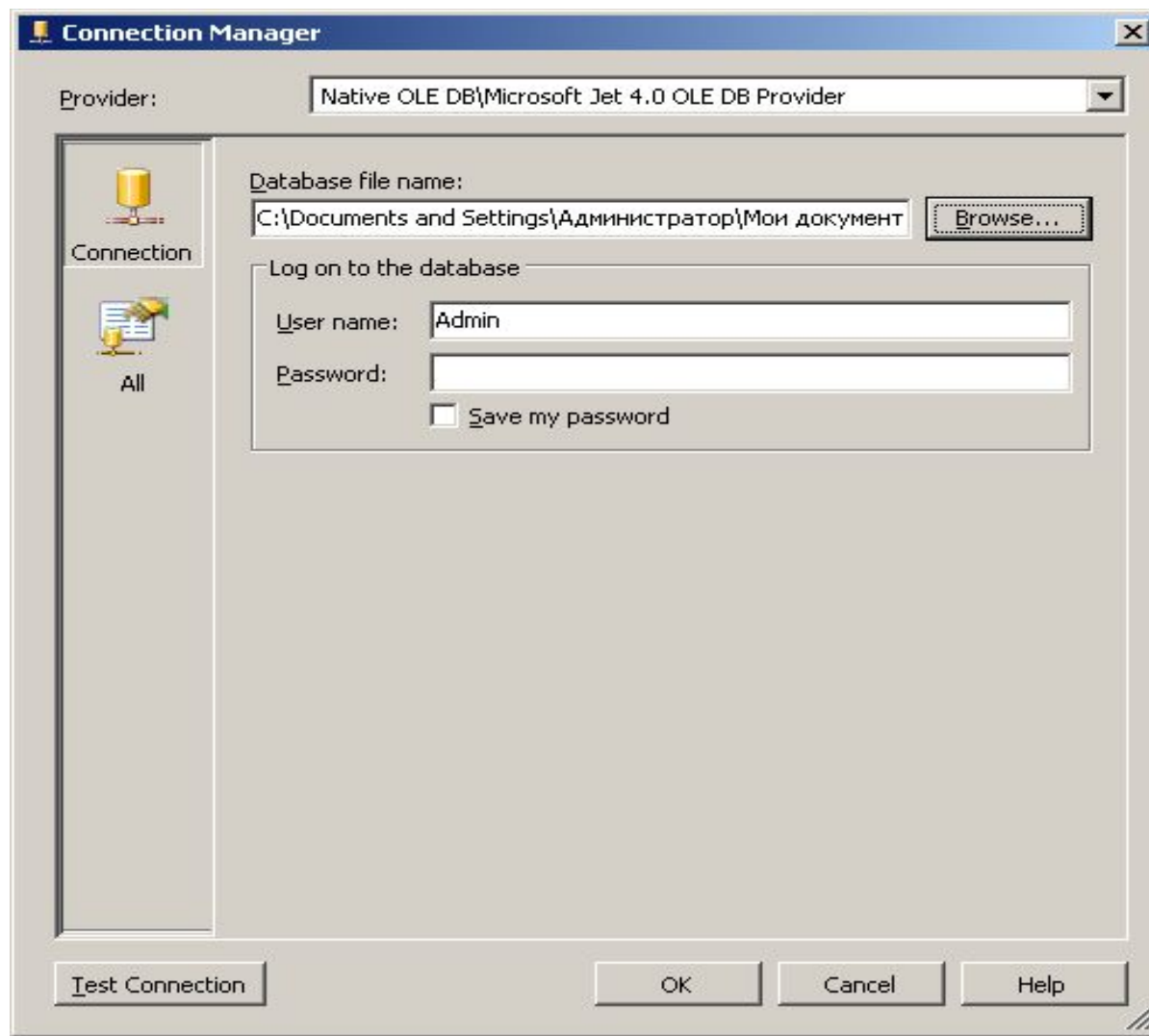
2.2. Список существующих источников данных



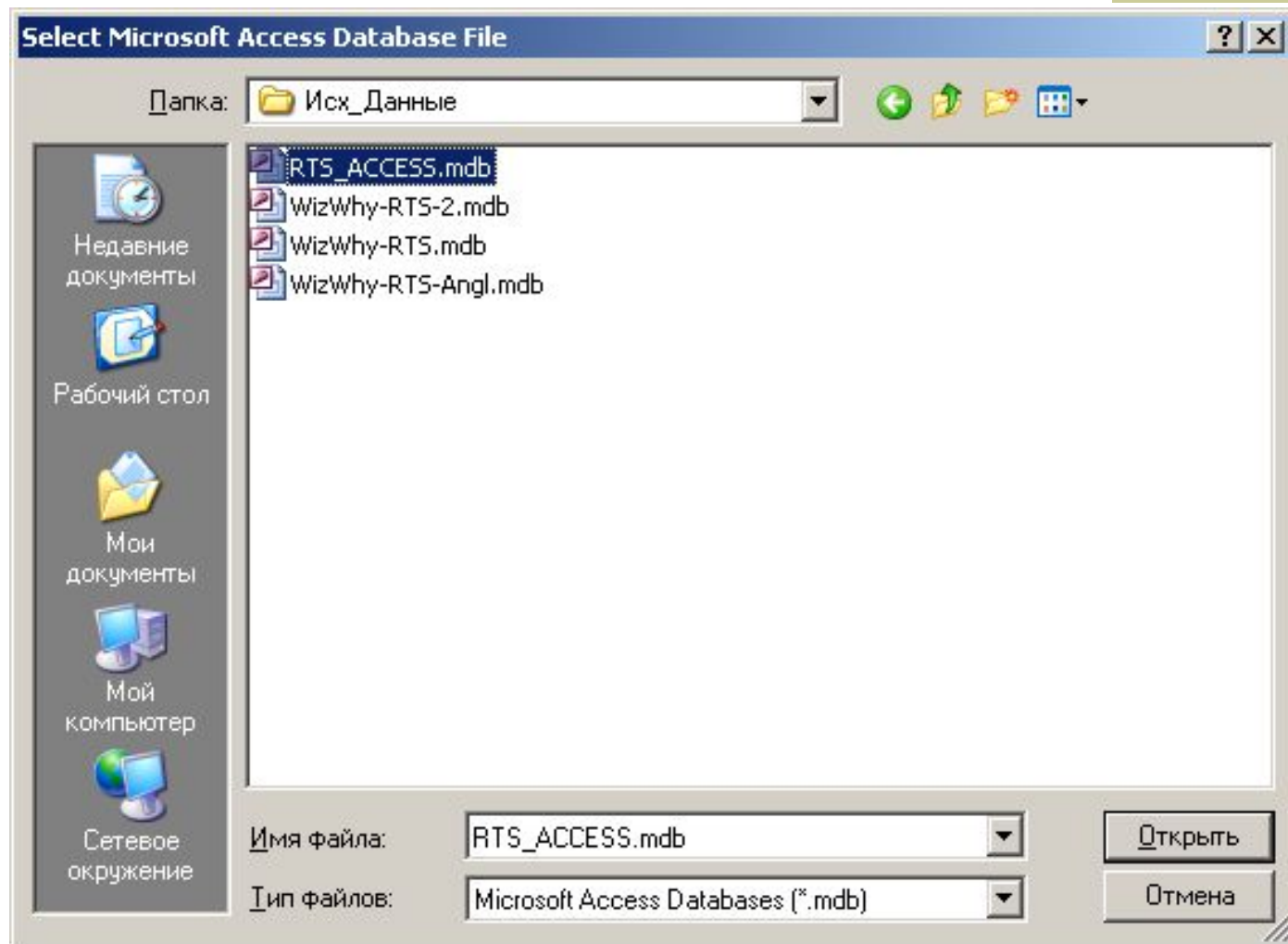
2.3. Выбор провайдера для реляционной базы данных типа Access



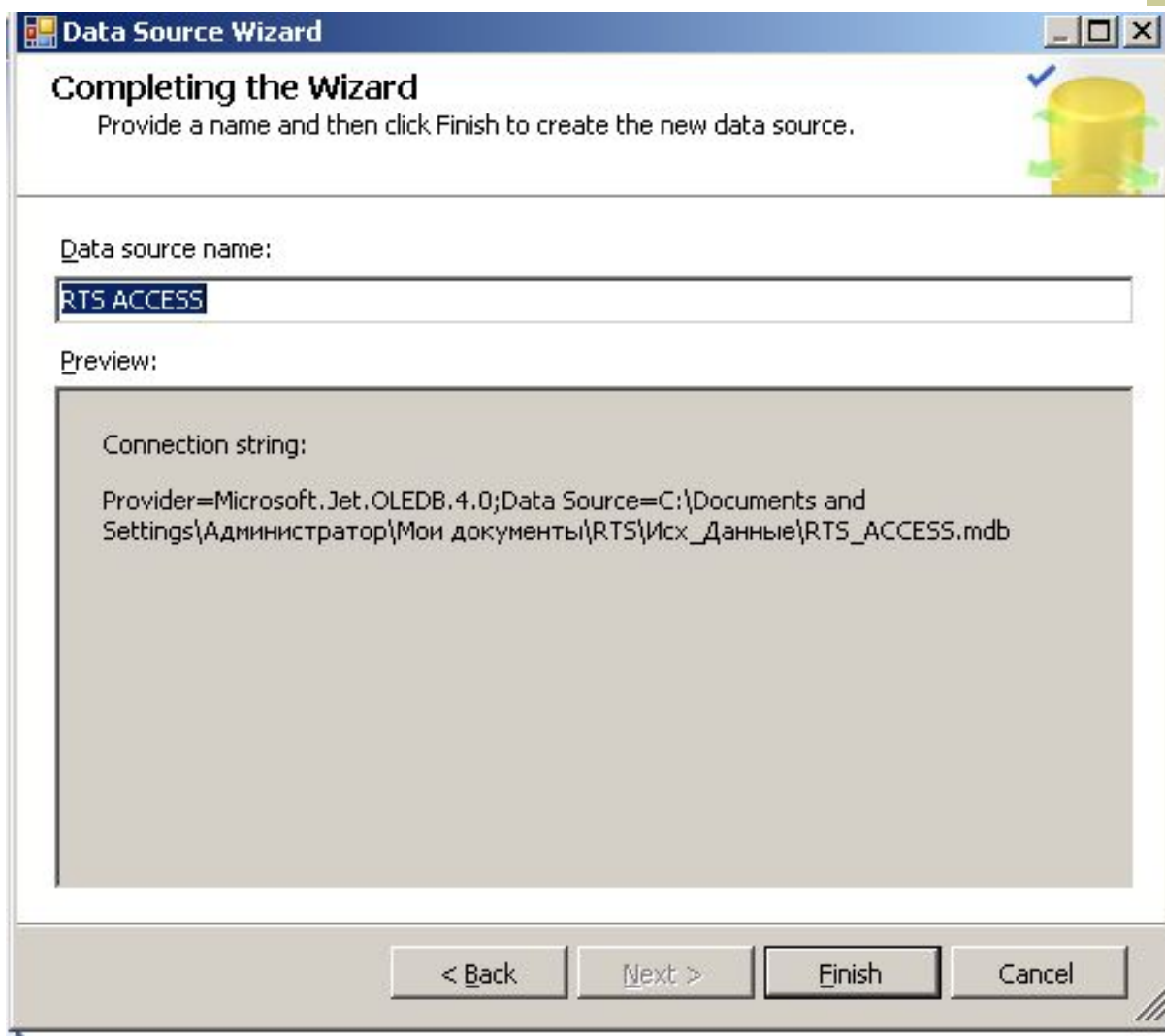
2.4. Поиск имени источника данных (Базы данных - Access)



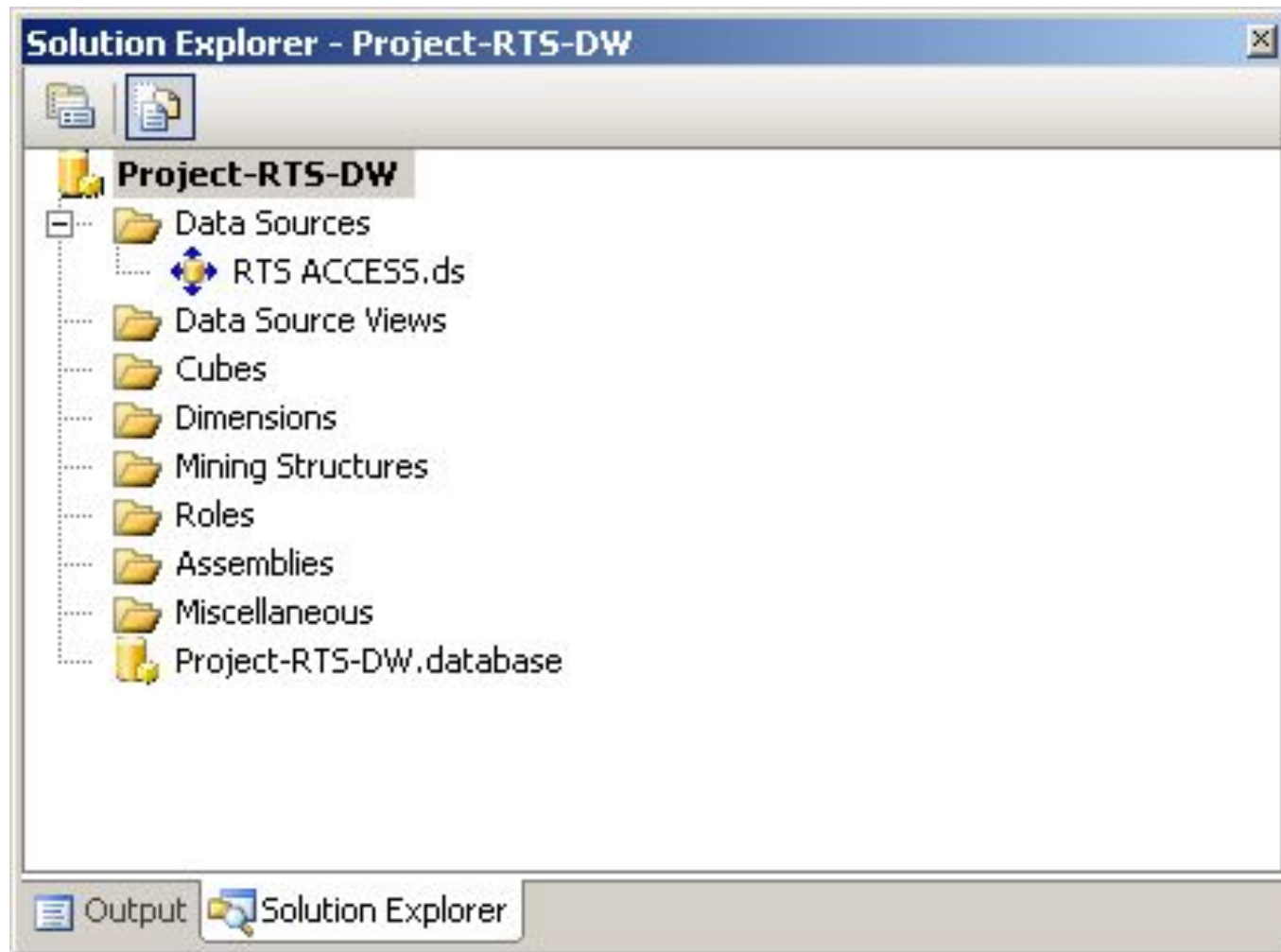
Стандартное окно поиска файла



2.6. Окончание подключения к источнику данных. Создание имени источника



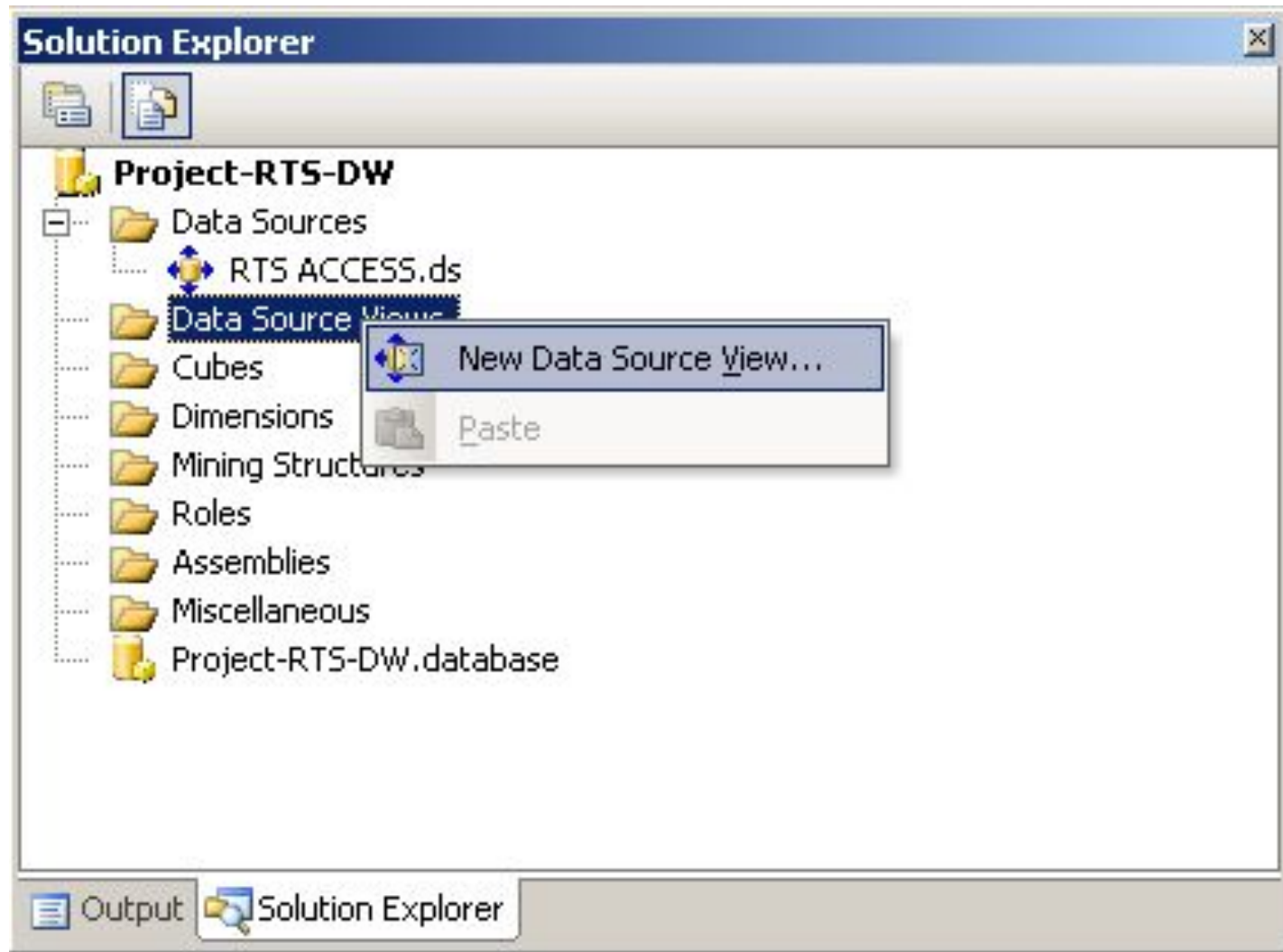
2.7. Внешний вид папки Solution (Решение)



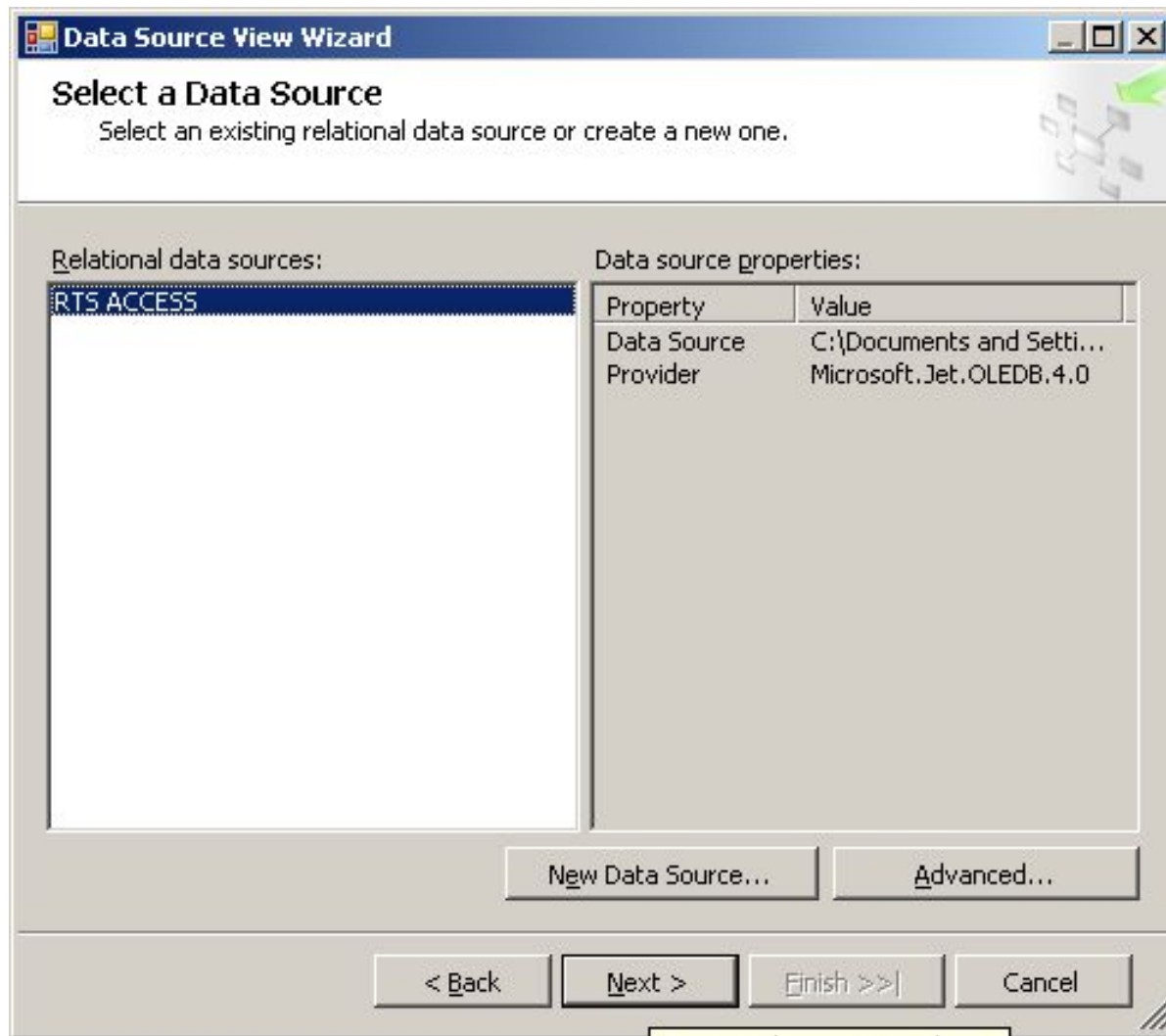
3. Создание Data Source View

Под Data Source View понимается срез источника, который будет использоваться для заполнения хранилища, при этом в него могут входить как таблицы, так и представления (view) реляционной базы – источника данных.

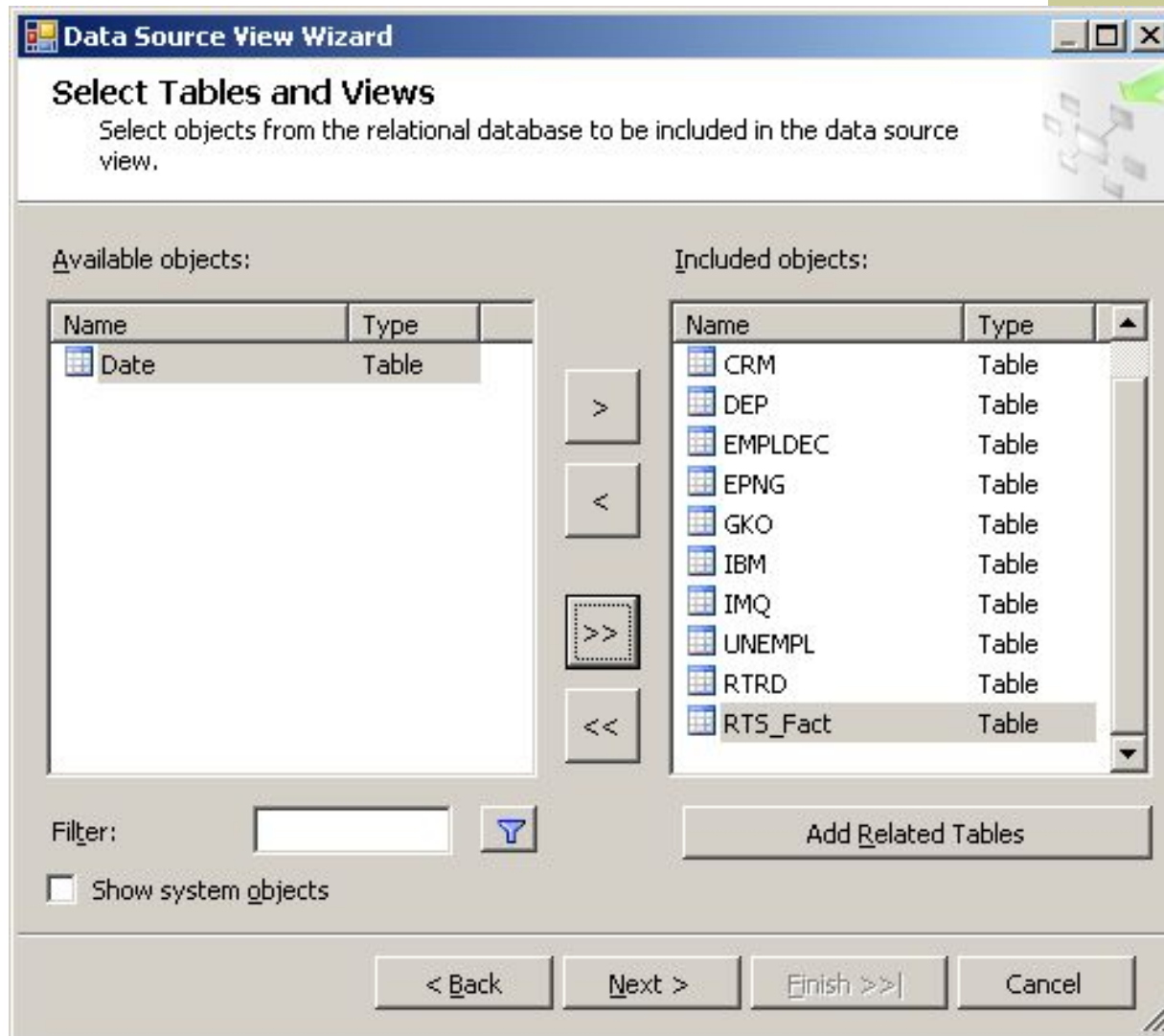
3.1. Команда на построения среза источника данных



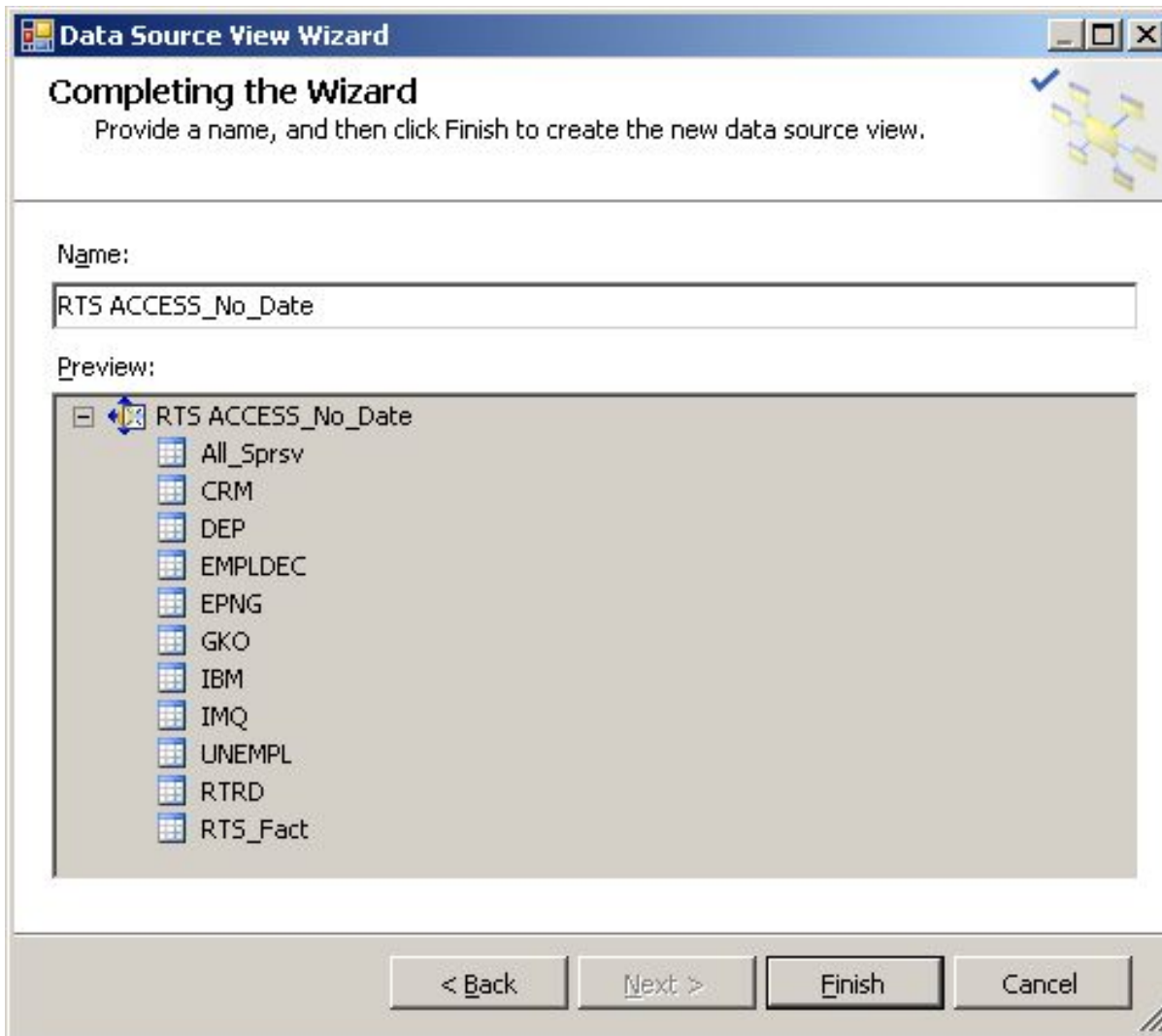
3.2. Выбор подключаемого источника данных



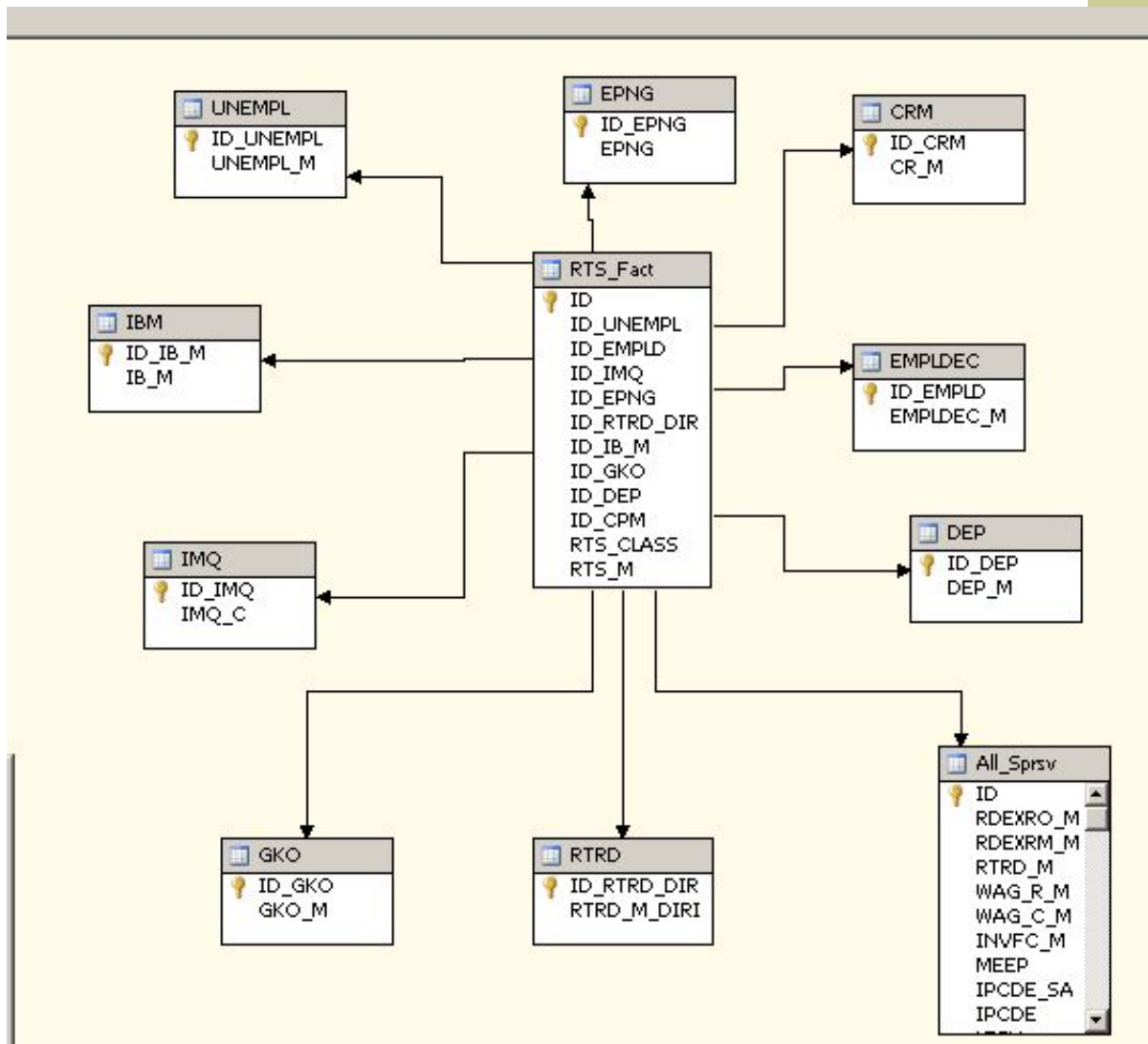
3.3. Выбор таблиц для включения в срез



3.4. Ввод имени подготовленного среза



3.5. Вид представления среза в дизайнере



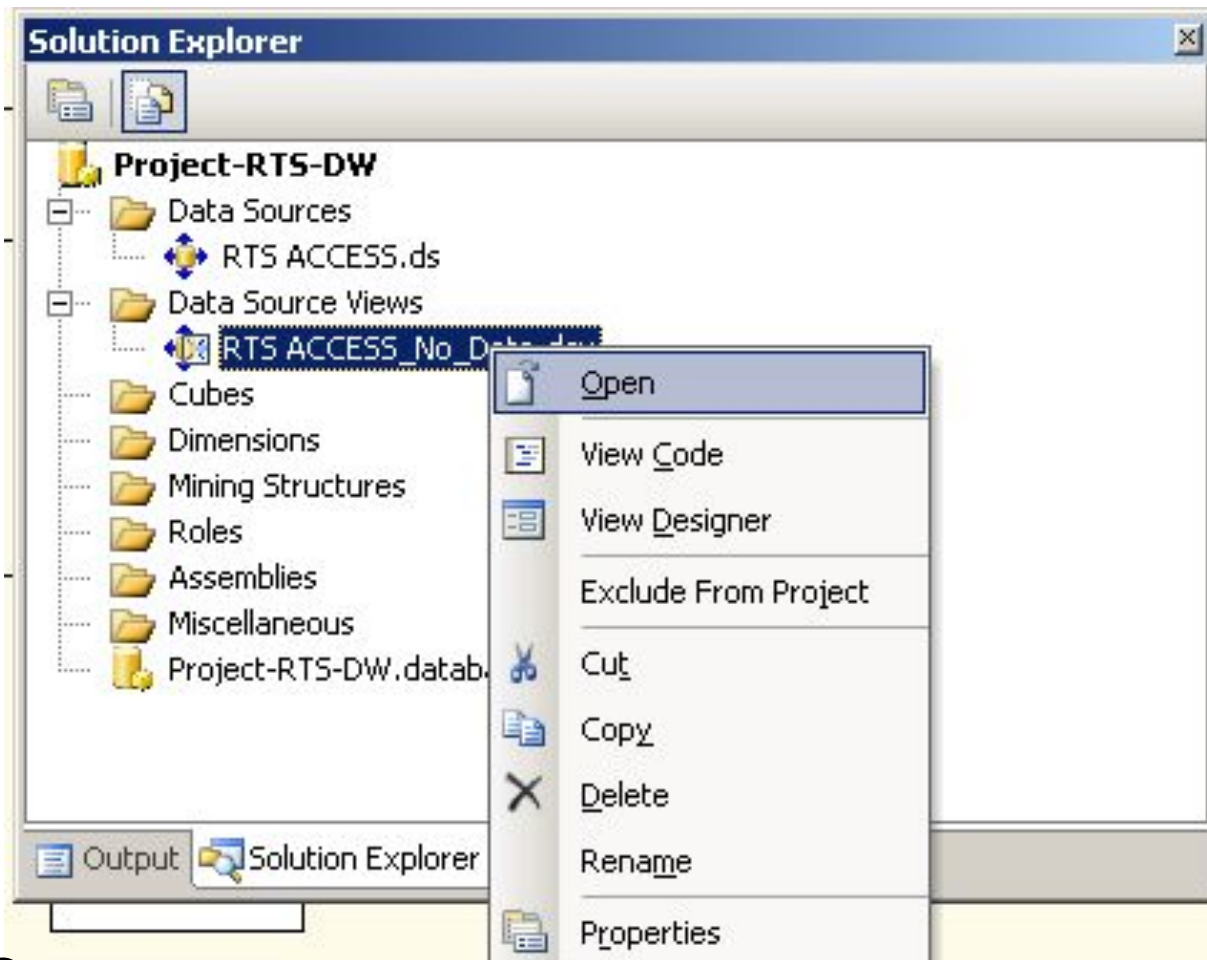
4. Подключение дополнительных размерностей

4.1. Дополнительные виды размерностей

С точки зрения своих возможностей размерности в рассматриваемой версии OLAP-сервера (SQL Server 2005) могут быть:

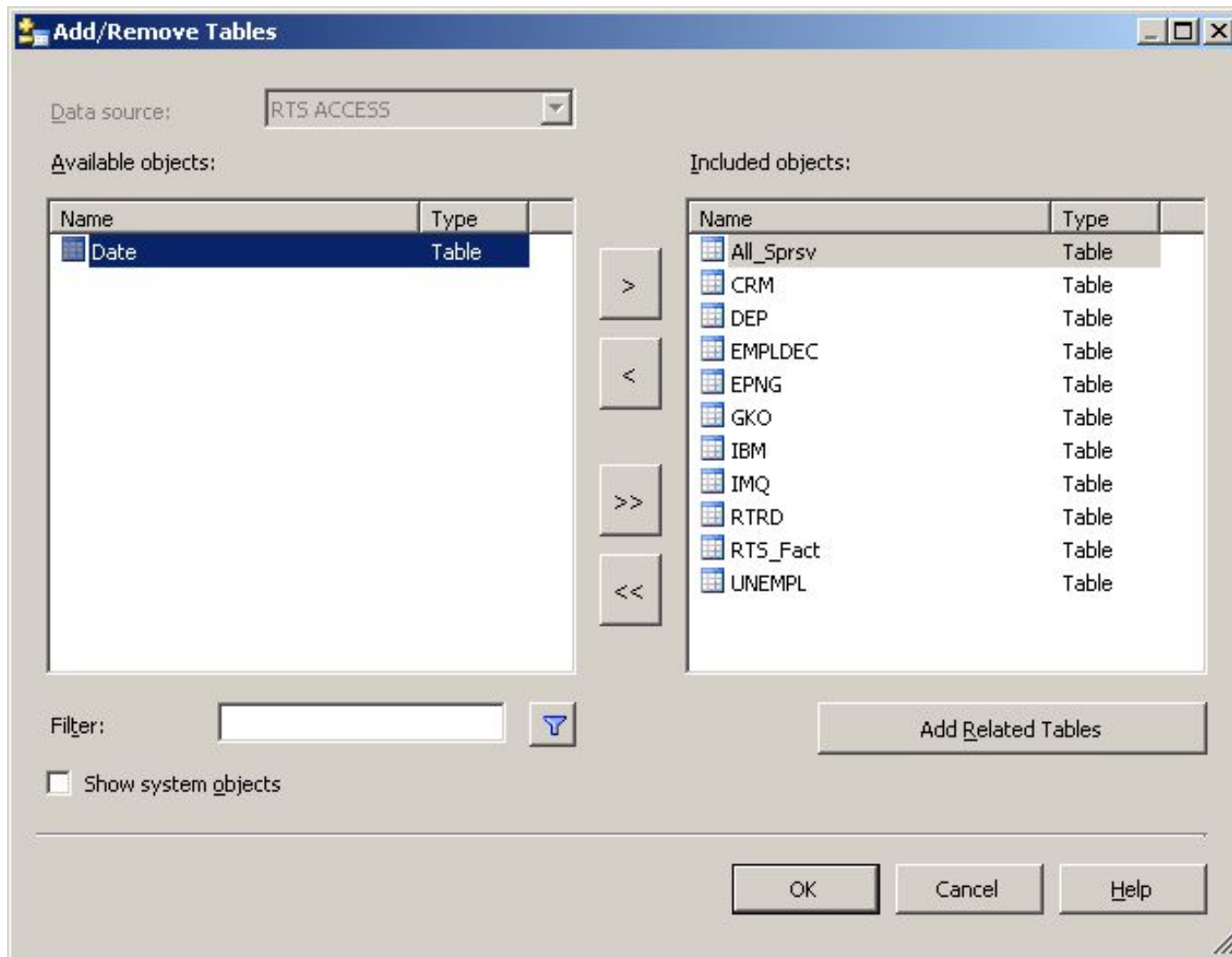
- регулярными (Regular);
- из таблицы фактов (Fact Dimension).
- ссылочными (Reference);
- многие-ко-многим (Many-to-Many);

4.2. Подключение размерности типа Дата

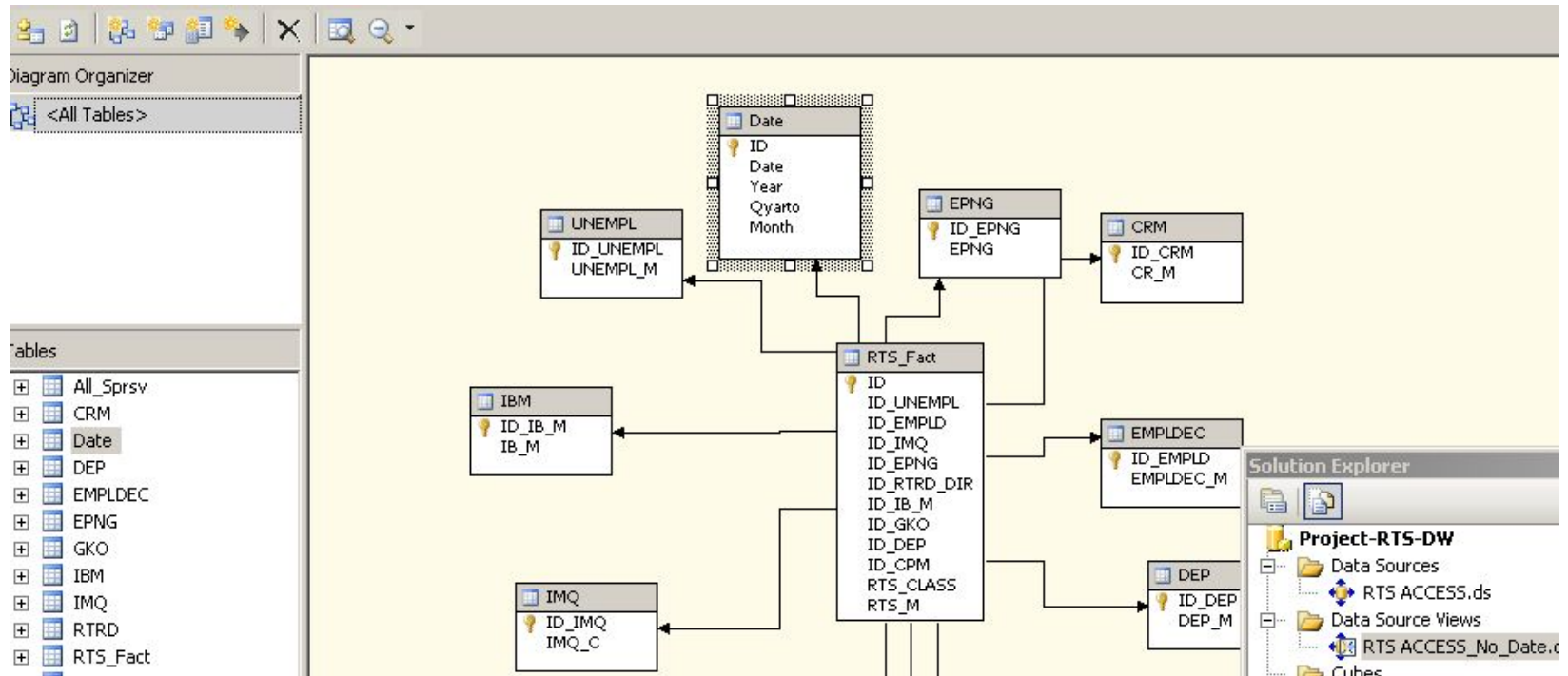


Открытие среза

Вид окна «Добавление таблицы»

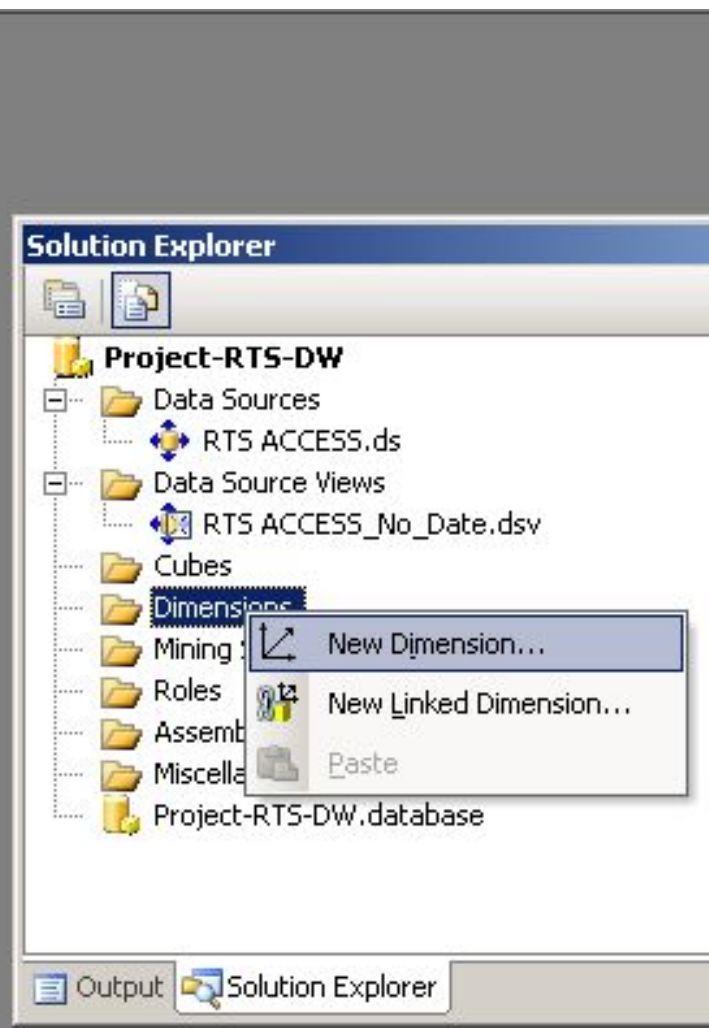


Вид дизайнера с добавленной таблицей Date

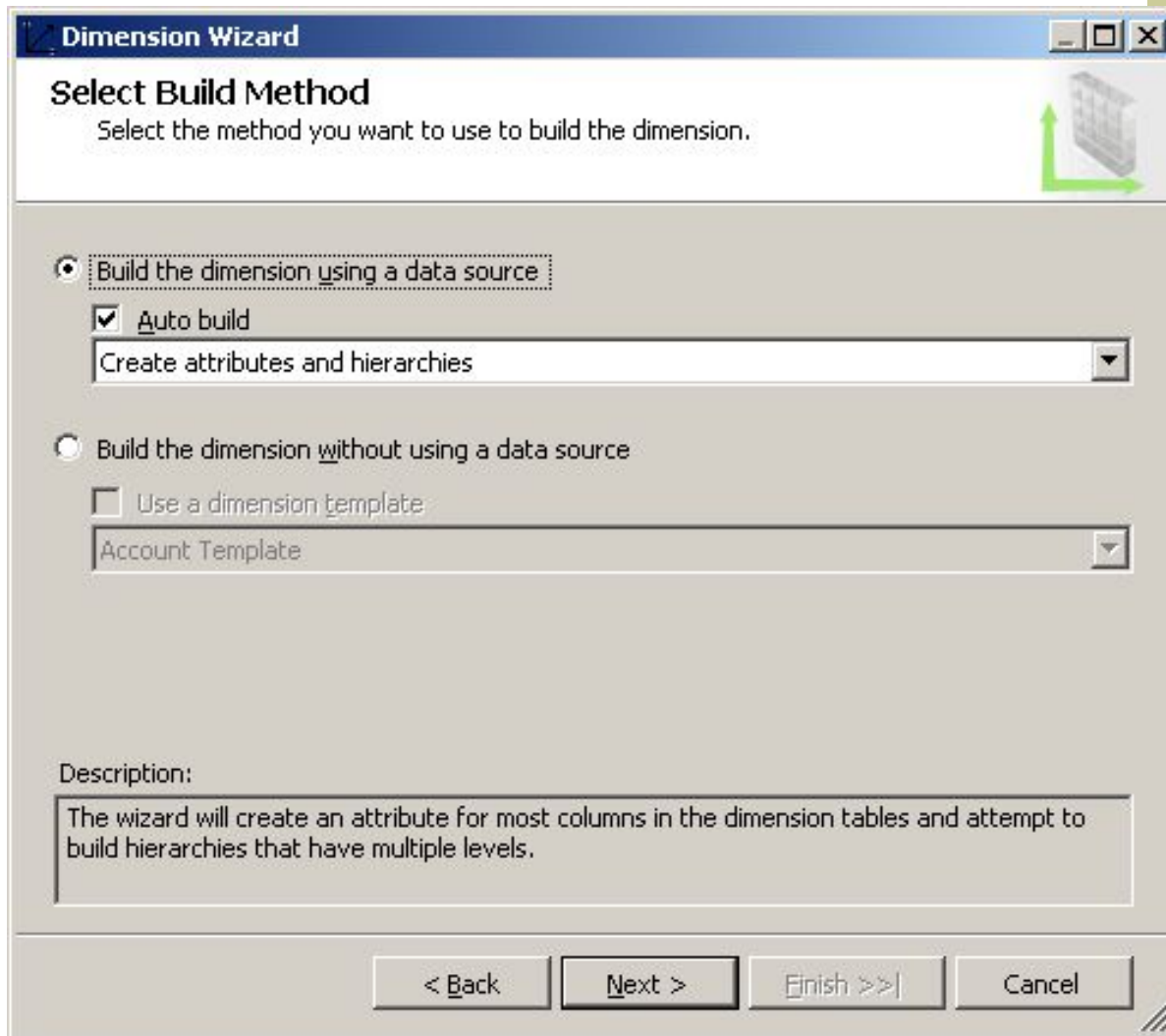


Вызов команды на построение нового измерения

Quarto	Month
	9
	10
	11
	12
	1
	2
	3
	4
	5
	6
	7
	8
	9
	10
	11
	12
	1
	2
	3
	4
	5



Выбор метода построения измерения



Dimension Wizard

Select Build Method
Select the method you want to use to build the dimension.

Build the dimension using a data source

Auto build
Create attributes and hierarchies

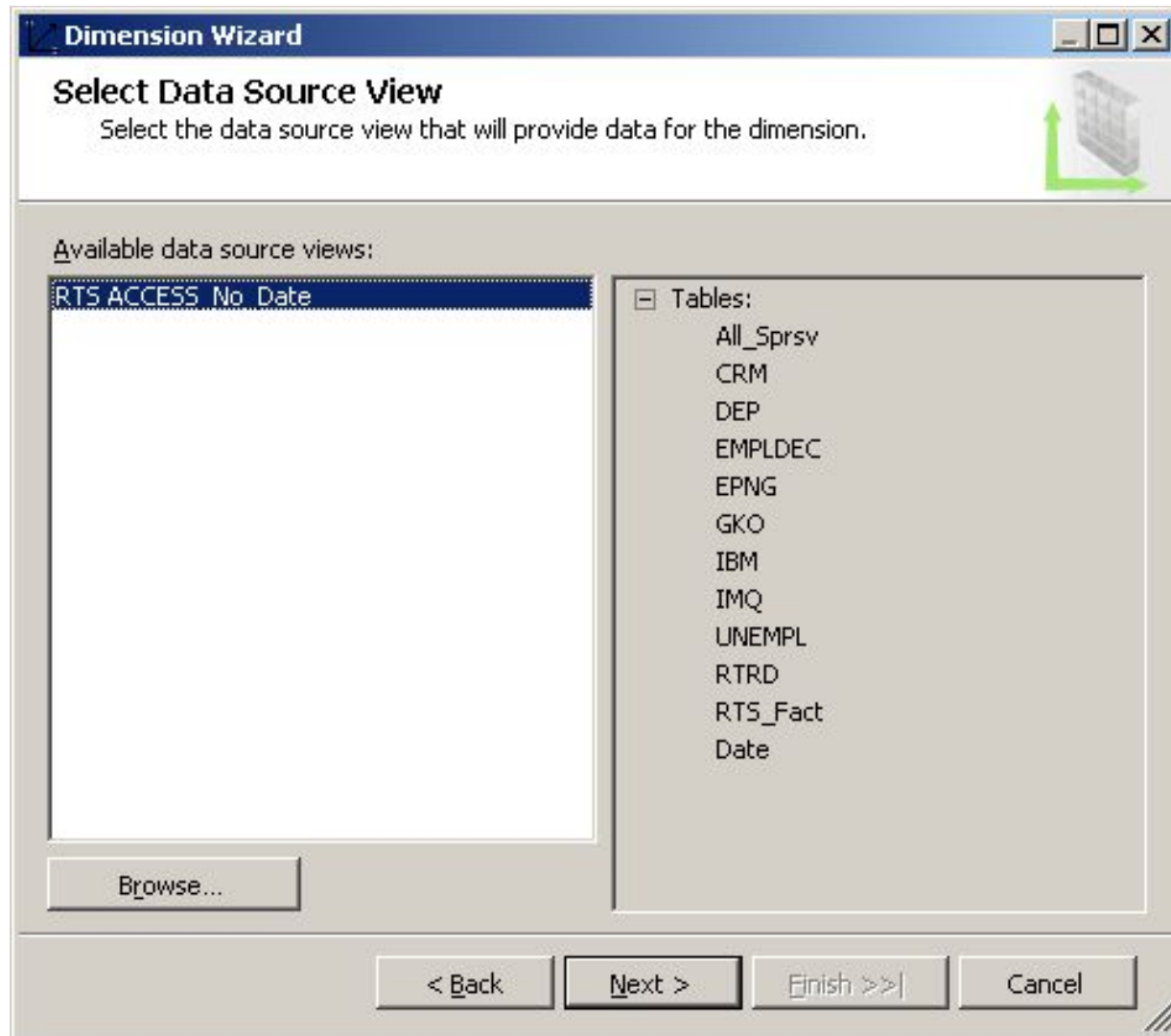
Build the dimension without using a data source

Use a dimension template
Account Template

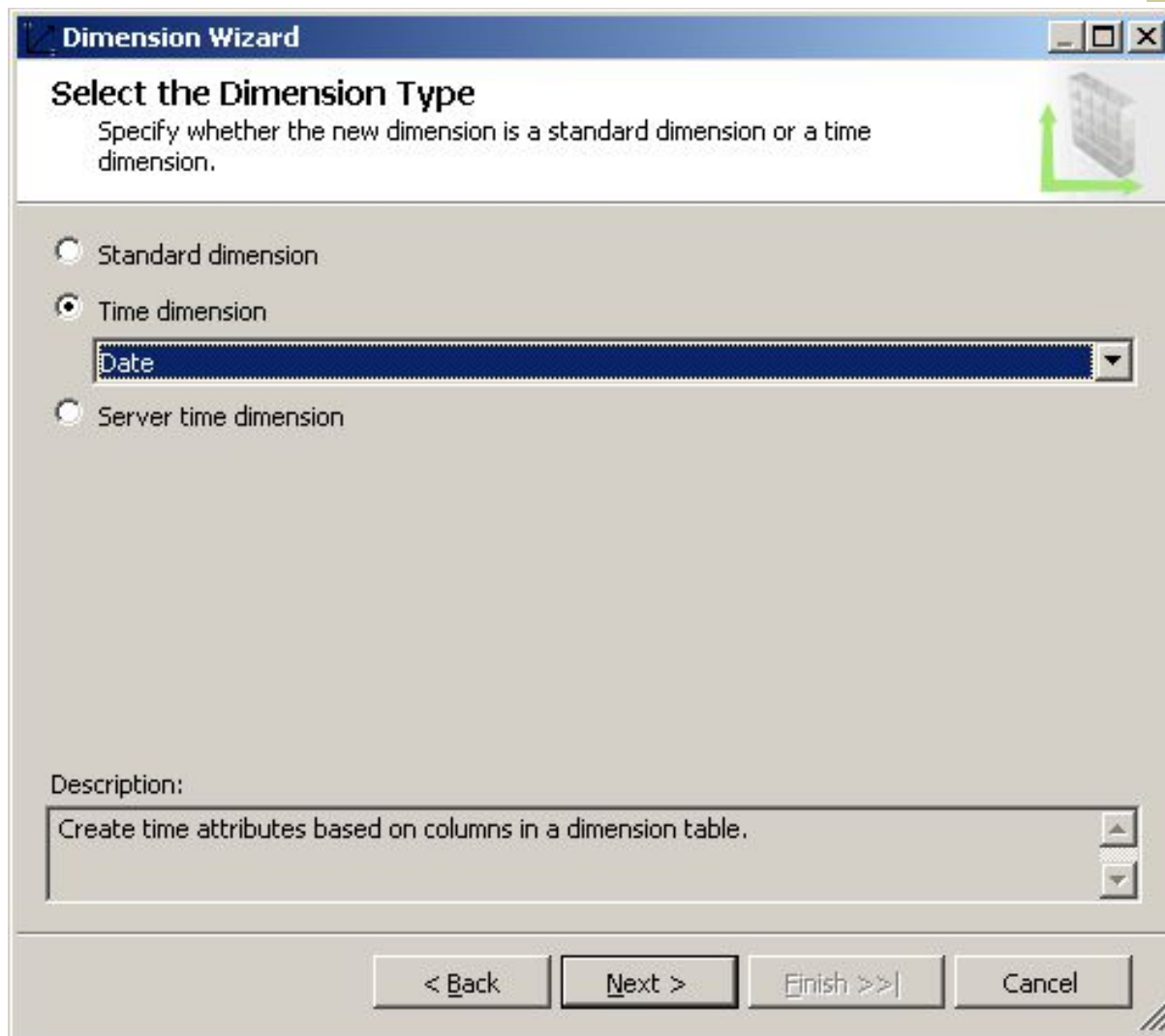
Description:
The wizard will create an attribute for most columns in the dimension tables and attempt to build hierarchies that have multiple levels.

< Back Next > Finish >>| Cancel

Выбор представления



Выбор типа измерения



Dimension Wizard

Select the Dimension Type
Specify whether the new dimension is a standard dimension or a time dimension.

Standard dimension

Time dimension

Date

Server time dimension

Description:
Create time attributes based on columns in a dimension table.

< Back Next > Finish >> | Cancel

Установка соответствия периодов

Dimension Wizard

Define Time Periods

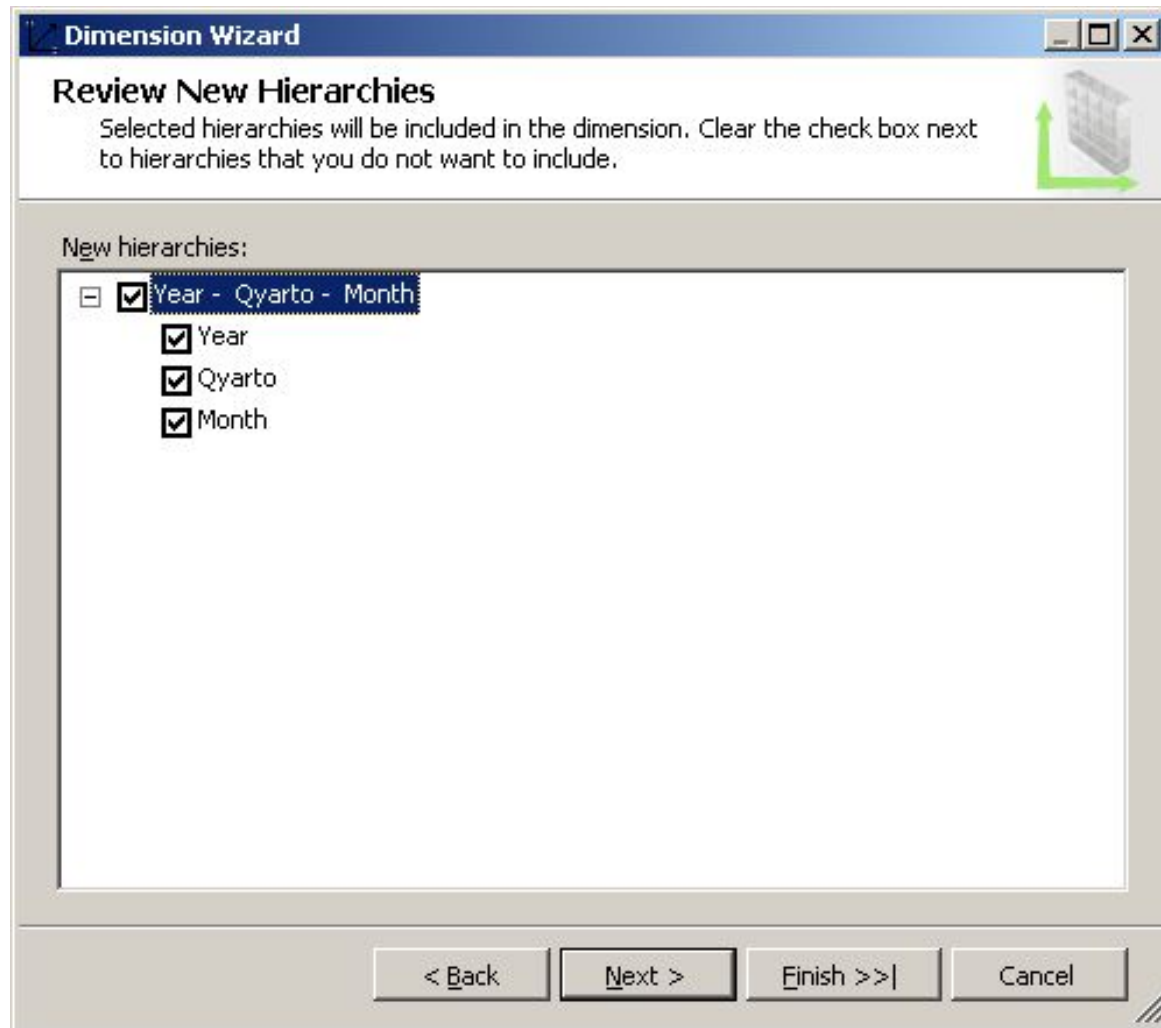
Select columns that represent time elements.

Time table columns:

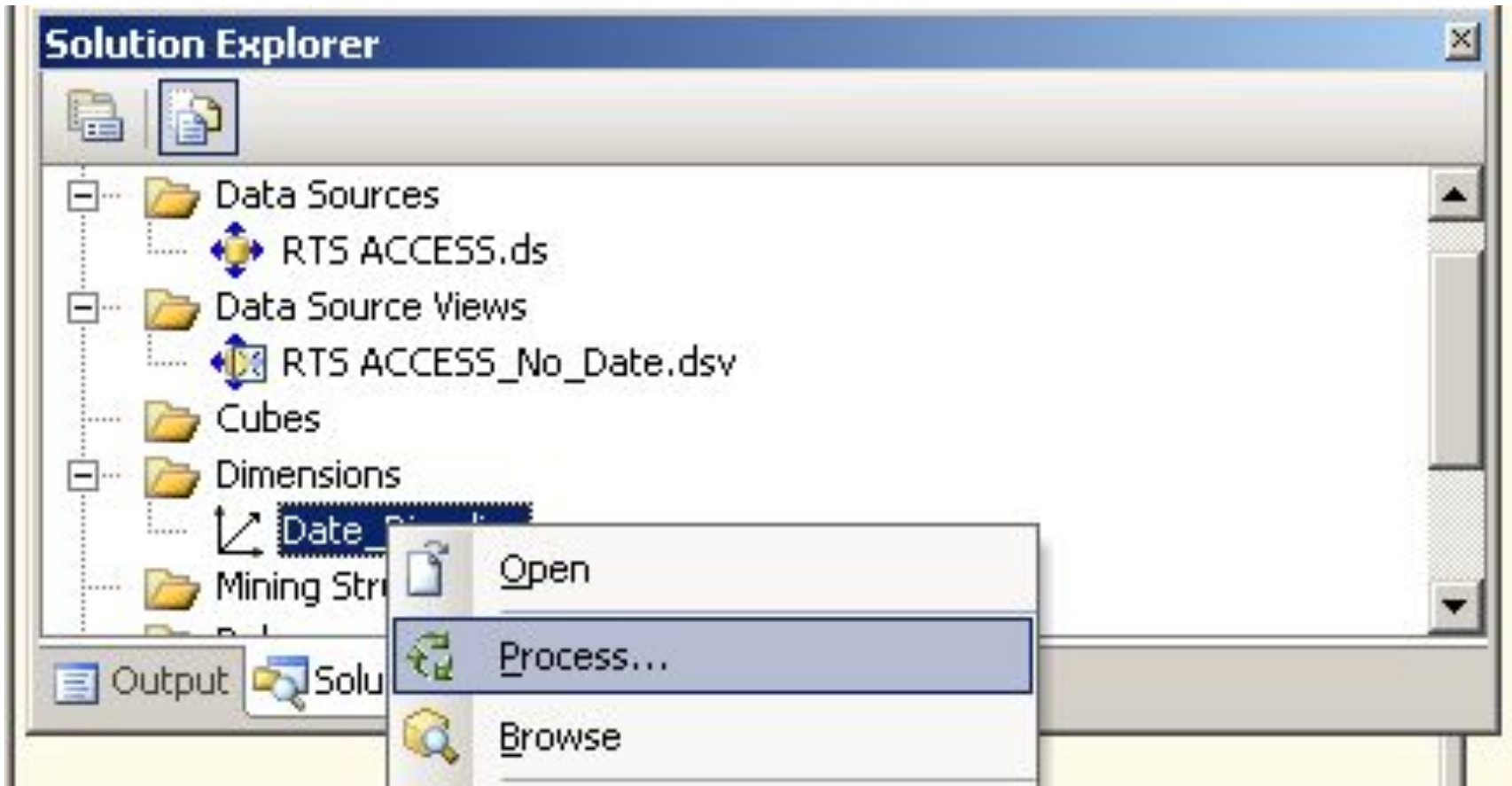
Time Property Name	Time Table Columns
Year	Year
Half Year	
Quarter	Qyarto
Trimester	
Month	Month
Date	ID
Ten Days	Date
Week	Month
Hour	
Minute	
Second	
Undefined Time	

< Back Next > Finish >>| Cancel

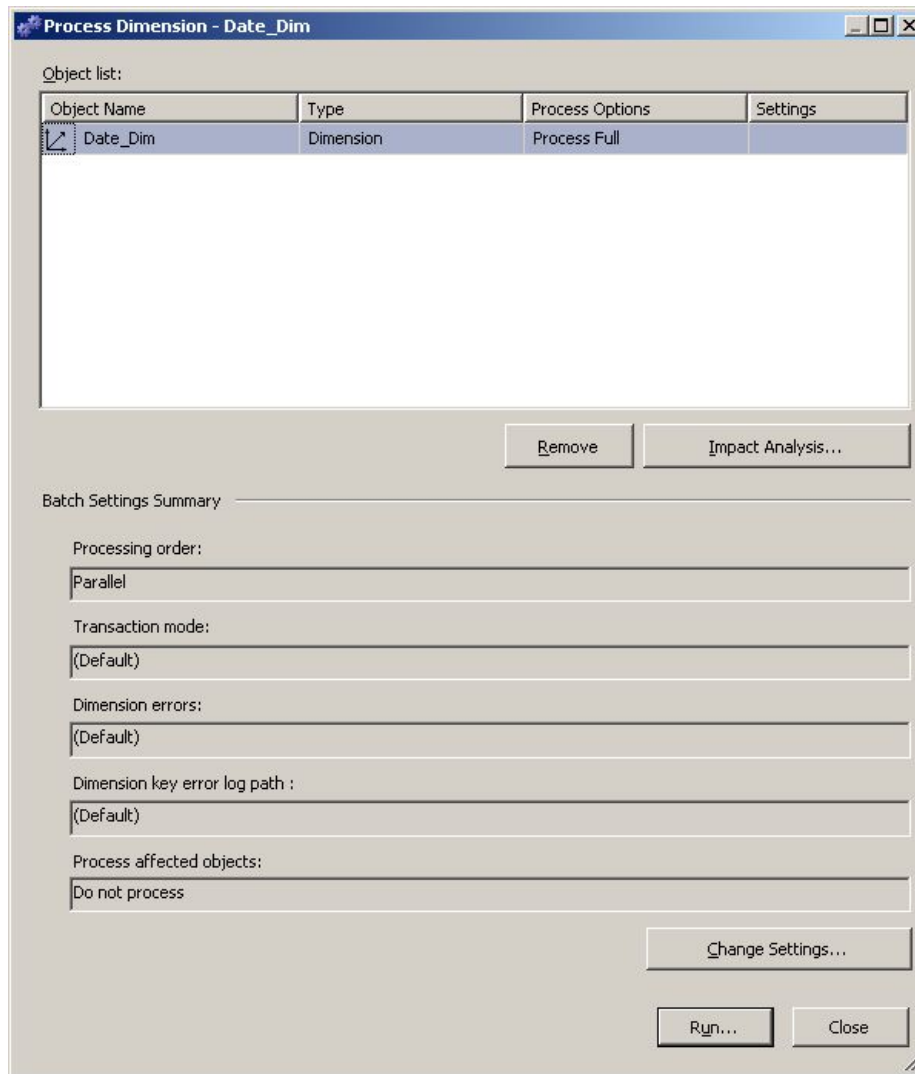
Иерархии подключаемого измерения



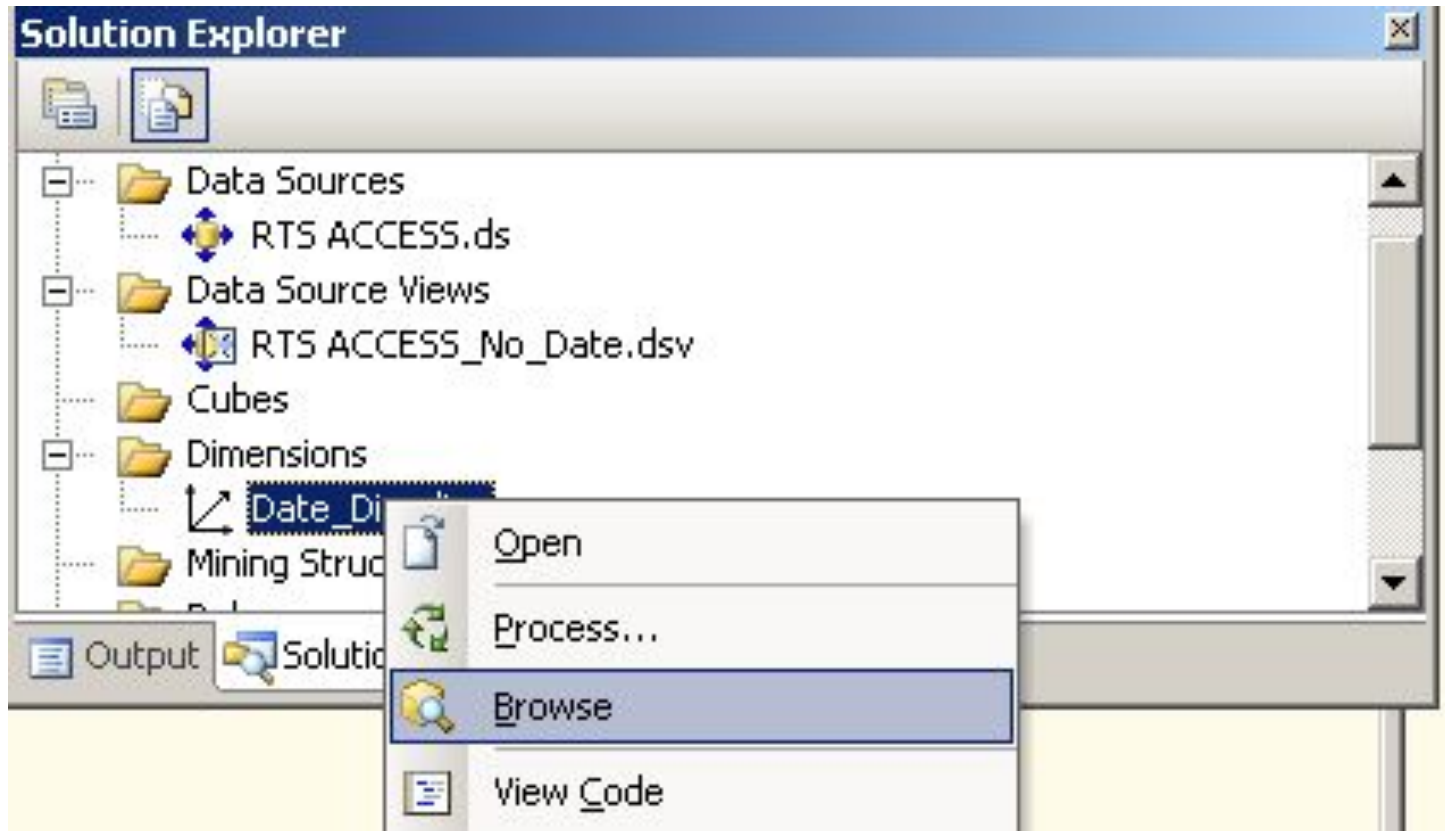
4.3. Расчет измерения Date



Процесс расчета измерения



4.3. Просмотр измерения Date



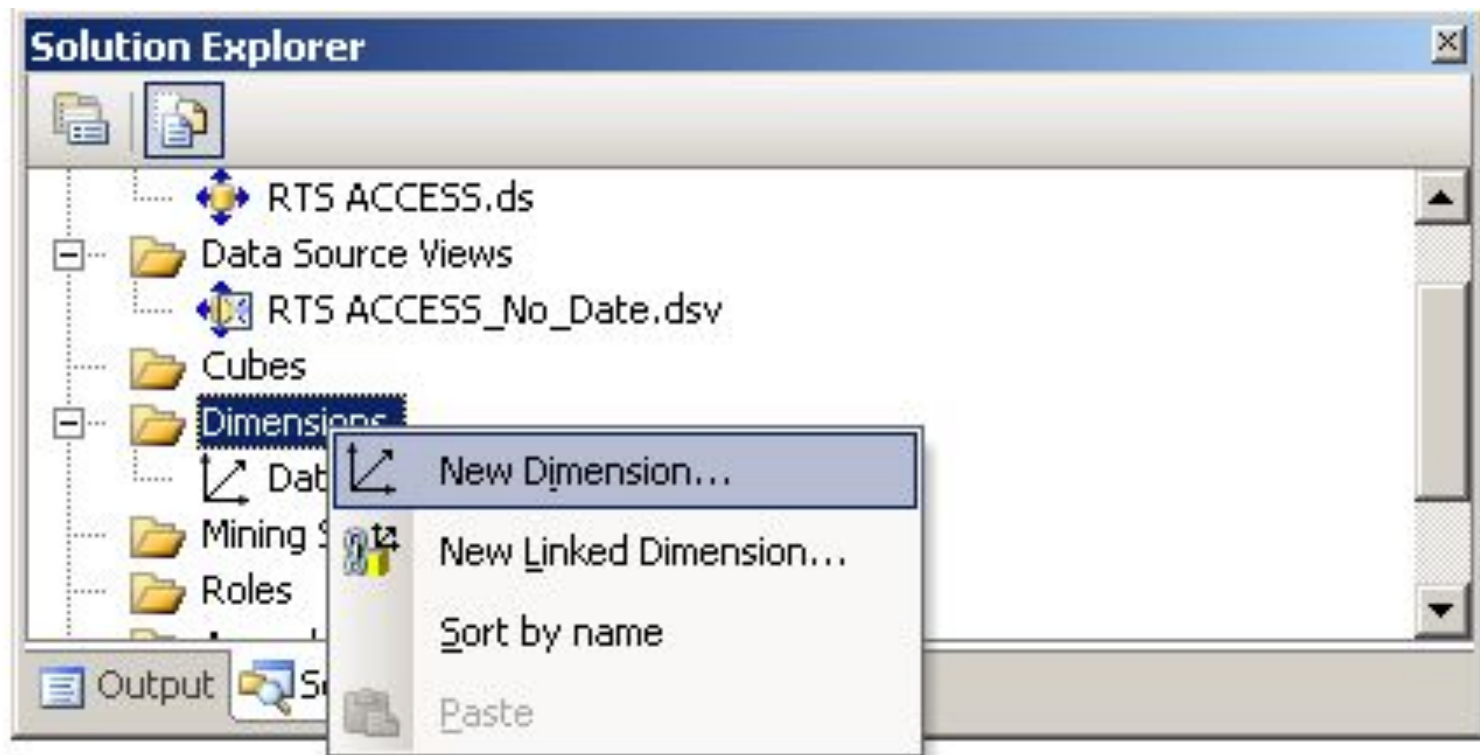
Команда на просмотр измерения

Просмотр размерности Дата

The screenshot displays a BI tool interface with the following components:

- Dimension Structure:** Shows a hierarchy of 'Year - Qyarto - Month'. The 'Current level' is set to '(All)'.
- Dimension Hierarchy:** A tree view showing the 'Date' dimension structure. The root is 'All', followed by years from 1995 to 2006. Each year is expanded to show quarters (3, 4) and months (1, 2). For 1996, months 4, 5, and 6 are also visible.
- Solution Explorer:** A window showing the project structure. It includes folders for 'Data Sources', 'Data Source Views', 'Cubes', 'Dimensions', and 'Mining Structures'. The 'Date_Dim.dim' file is highlighted under the 'Dimensions' folder.

4.4. Добавление размерности из таблицы фактов



Добавление новой размерности

Выбор типа измерения

Dimension Wizard

Select the Dimension Type
Specify whether the new dimension is a standard dimension or a time dimension.

Standard dimension

Time dimension

All_Sprsv

Server time dimension

Description:
Define a dimension based on one or more dimension tables. The structure of the data defines the attributes and hierarchies in the dimension.

< Back Next > Finish >> Cancel

Выбор таблицы и её атрибутов для измерения

Dimension Wizard

Select the Main Dimension Table
Select the main table and one or more key columns that relate to the fact table.

Main table:
RTS_Fact

Key columns:

- ID_IMQ
- ID_EPNG
- ID_RTRD_DIR
- ID_IB_M
- ID_GKO
- ID_DEP
- ID_CPM
- RTS_CLASS
- RTS_M

Column containing the member name (optional):
ID

< Back Next > Finish >>| Cancel

Атрибуты, включаемые в измерение

Dimension Wizard

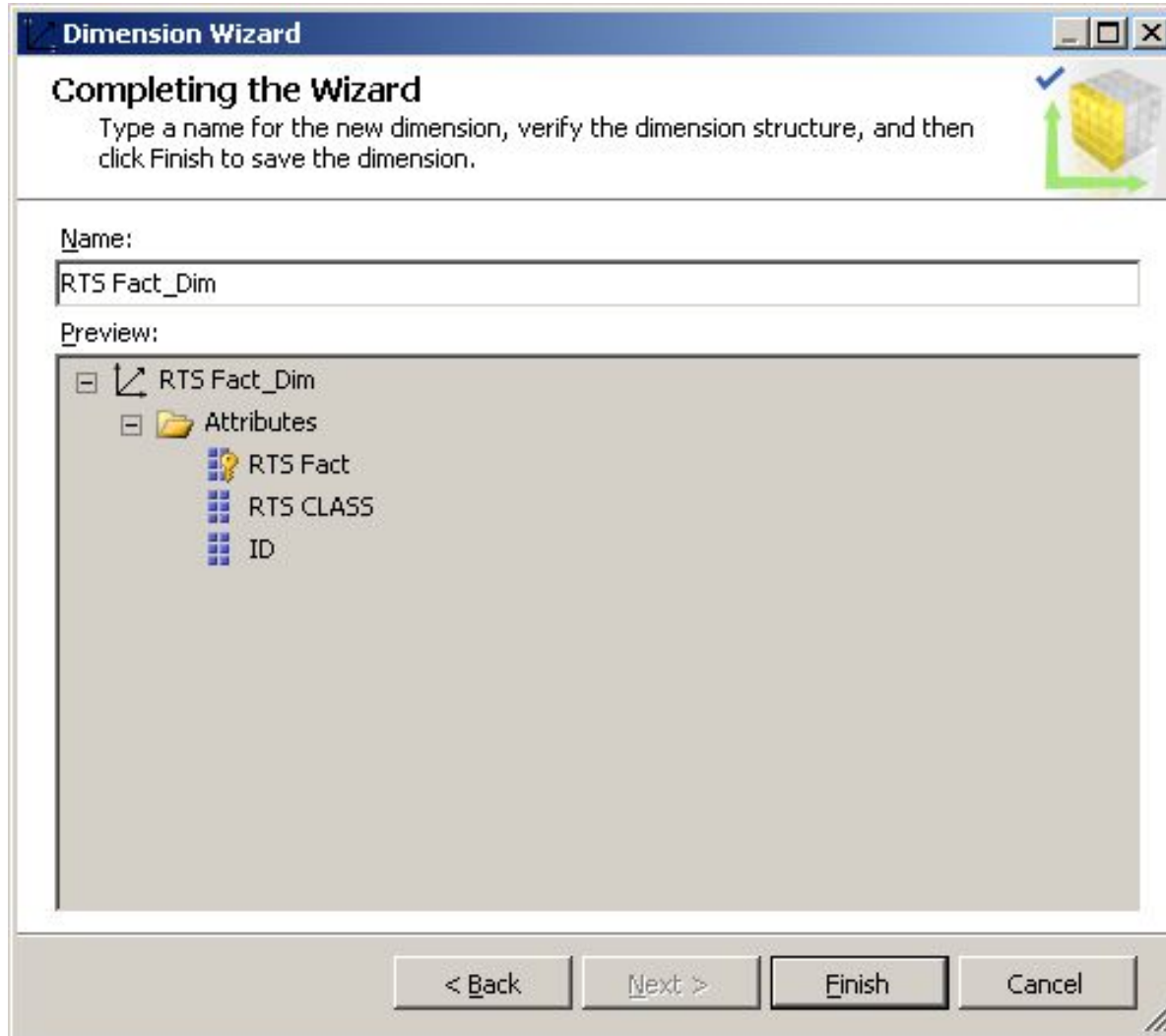
Select Dimension Attributes
Select attributes that you want to include in the dimension.

Dimension attributes:

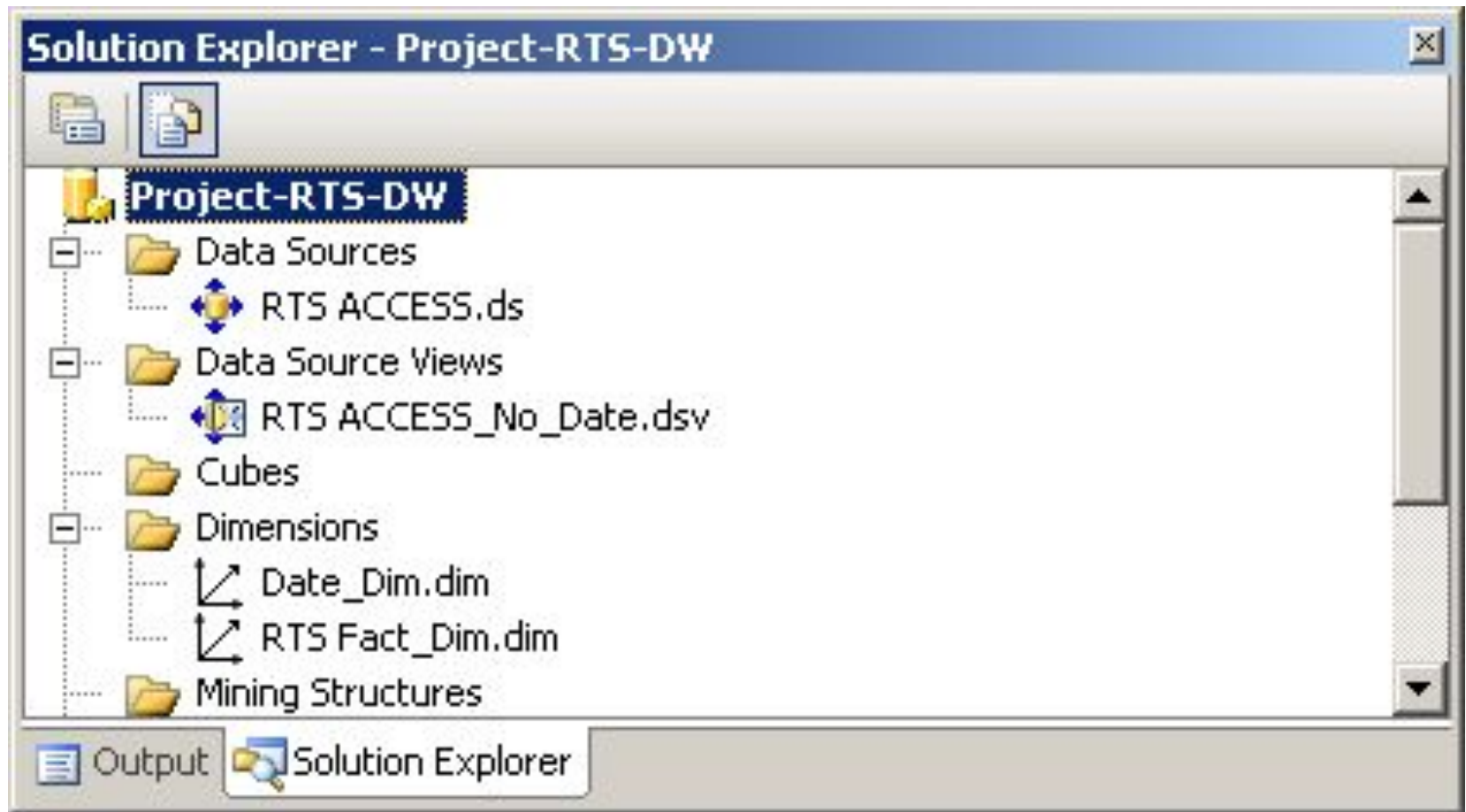
<input type="checkbox"/>	Attribute Name	Attribute Key Column	Attribute Name Column
<input checked="" type="checkbox"/>	RTS M	RTS_M	RTS_M
<input type="checkbox"/>	ID EMPLD	ID_EMPLD	ID_EMPLD
<input type="checkbox"/>	ID GKO	ID_GKO	ID_GKO
<input type="checkbox"/>	ID UNEMPL	ID_UNEMPL	ID_UNEMPL
<input checked="" type="checkbox"/>	ID	ID	ID
<input type="checkbox"/>	ID EPNG	ID_EPNG	ID_EPNG
<input type="checkbox"/>	ID DEP	ID_DEP	ID_DEP
<input type="checkbox"/>	ID CPM	ID_CPM	ID_CPM
<input type="checkbox"/>	ID IB M	ID_IB_M	ID_IB_M
<input type="checkbox"/>	ID RTRD DIR	ID_RTRD_DIR	ID_RTRD_DIR
<input type="checkbox"/>	ID IMQ	ID_IMQ	ID_IMQ

< Back Next > Finish >> | Cancel

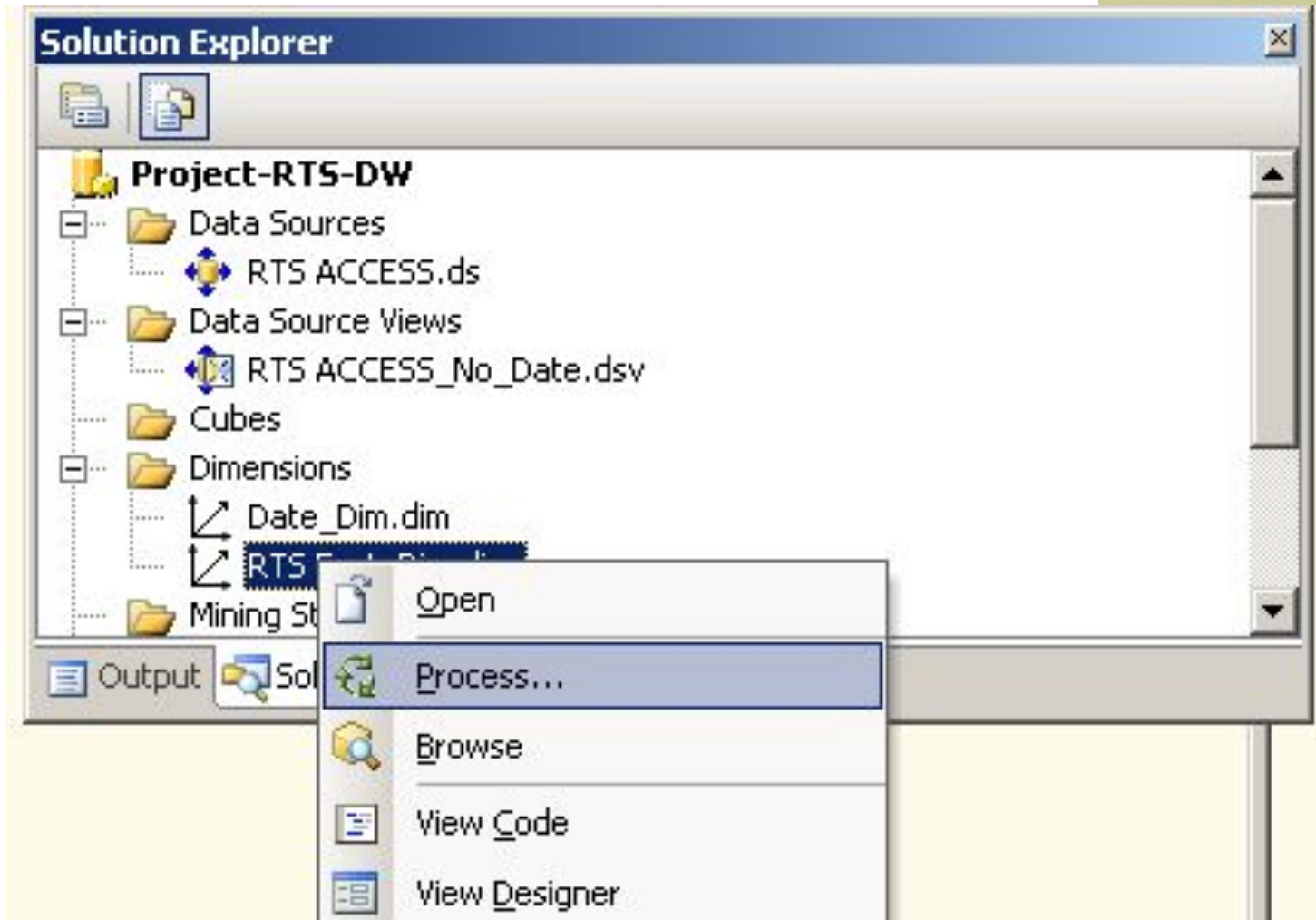
Состав измерения и его имя



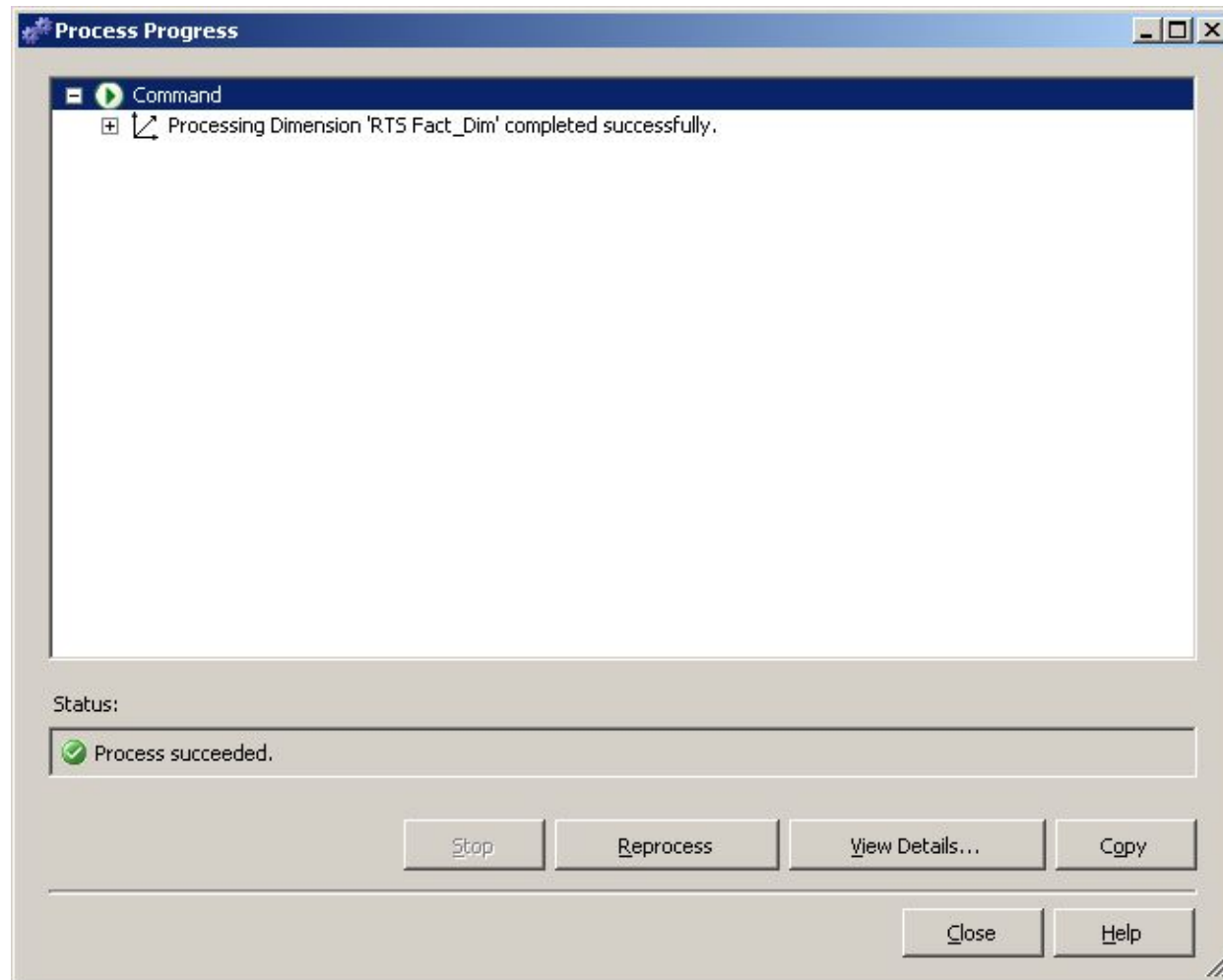
Появилось добавленное измерение



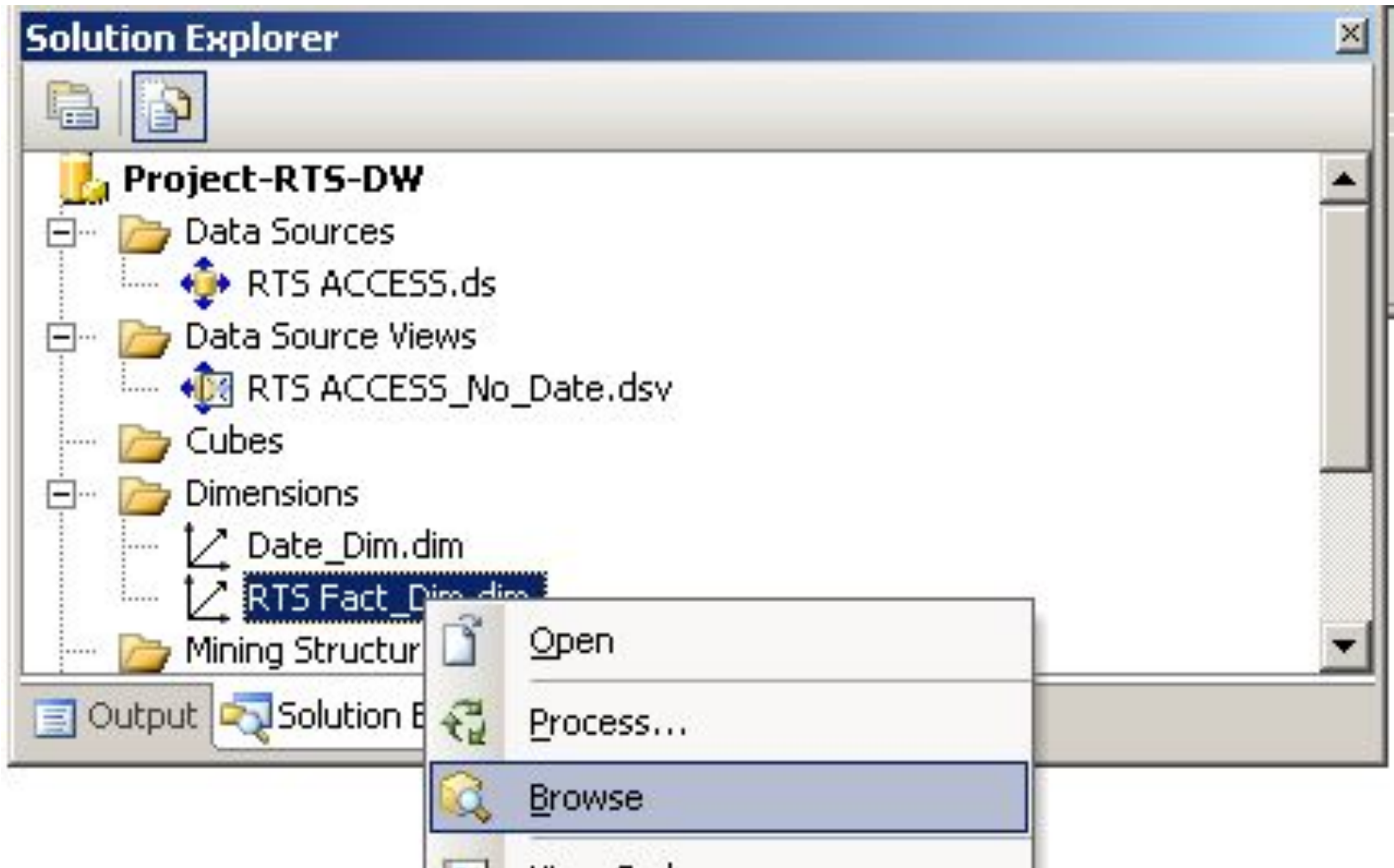
4.5. Расчет размерности из таблицы фактов



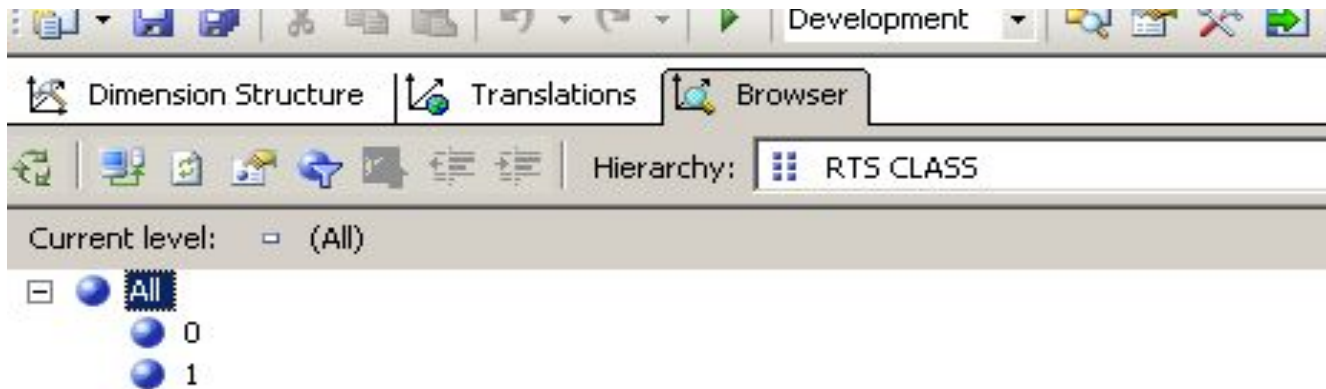
Расчет закончен



4.6. Просмотр размерности из таблицы фактов



Просмотр добавленной размерности, созданной из таблицы фактов



5. Создание куба

5.1. Особенности показателей в кубах данных MS SQL Server 2005

Система агрегирования представляет собой внутренний механизм, обеспечивающий управляемую «свертку» данных по иерархиям размерностей.

Если не управлять агрегированием, то оно выполняется по умолчанию, то есть свертка показателей производится с использованием функции суммирования снизу вверх (в нашем примере — от дней до лет). Управляя этим механизмом, мы можем:

- изменять функции агрегирования, используемые для свертки данных;
- изменять порядок свертки (выполнять ее не по всем размерностям);
- определять, на каких уровнях иерархии будет производиться свертка данных и т. д.

Показатели могут быть:

- аддитивными (additive);
- полуаддитивными (semiadditive);
- неаддитивными (nonadditive).

Аддитивные показатели

- **Аддитивные показатели**, также называемые полноаддитивными, агрегируются со всеми размерностями, включенными в группы показателей. Говоря другими словами (в терминах Microsoft), они пересекаются с любым членом размерности, то есть выбор любого члена любой размерности приводит к пересчету агрегатов показателей.

Полуаддитивные показатели

- **Полуаддитивные показатели** агрегируются относительно некоторых (не всех) размерностей.
- Например, показатель, определяющий количество товара на руках, может агрегироваться по географическому признаку, но не агрегироваться по остальным размерностям, поскольку к ним не должен относиться. О таких показателях говорят, что они не пересекаются с некоторыми размерностями.
- Скажем, агрегирование такого показателя по размерности «время» просто бессмысленно, поскольку нас не интересует, сколько товара было на руках в какой-то период времени.

Неаддитивные показатели

- **Неаддитивные показатели** не агрегируются по размерностям, но могут быть посчитаны для любой ячейки куба.
- Например, подсчитываемый показатель, возвращающий процент дохода, не может быть агрегирован из значений процентов своих дочерних ячеек других размерностей.

5.2. Особенности сохранения кубов

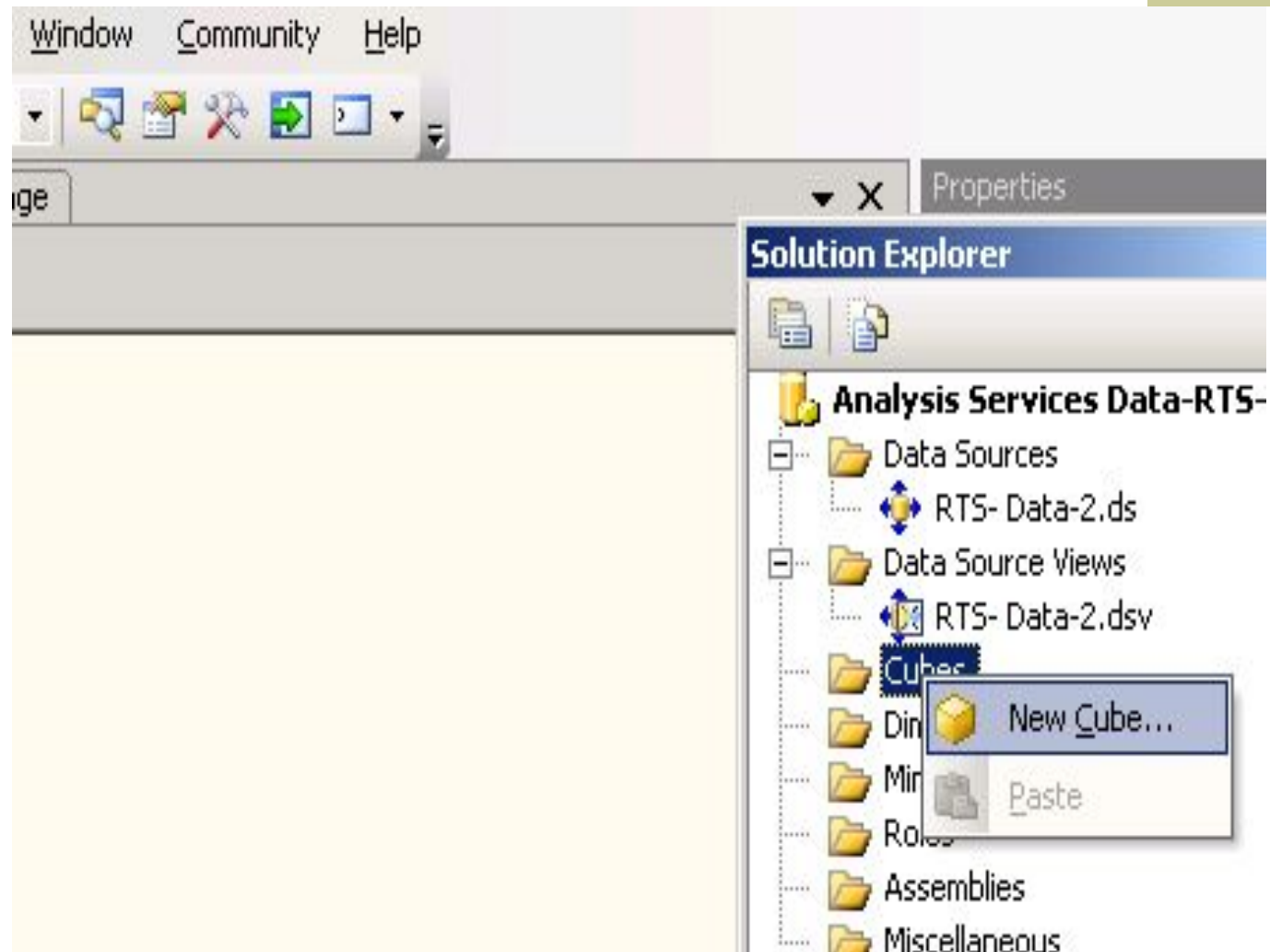
- пространство на диске не выделяется под пустые ячейки;
- выполняется сжатие данных;
- сложный механизм агрегирования определяет, каким образом будет произведена свертка данных на каждом уровне иерархии размерностей, чтобы при этом не пострадало качество.

5.3. Подключение простых размерностей

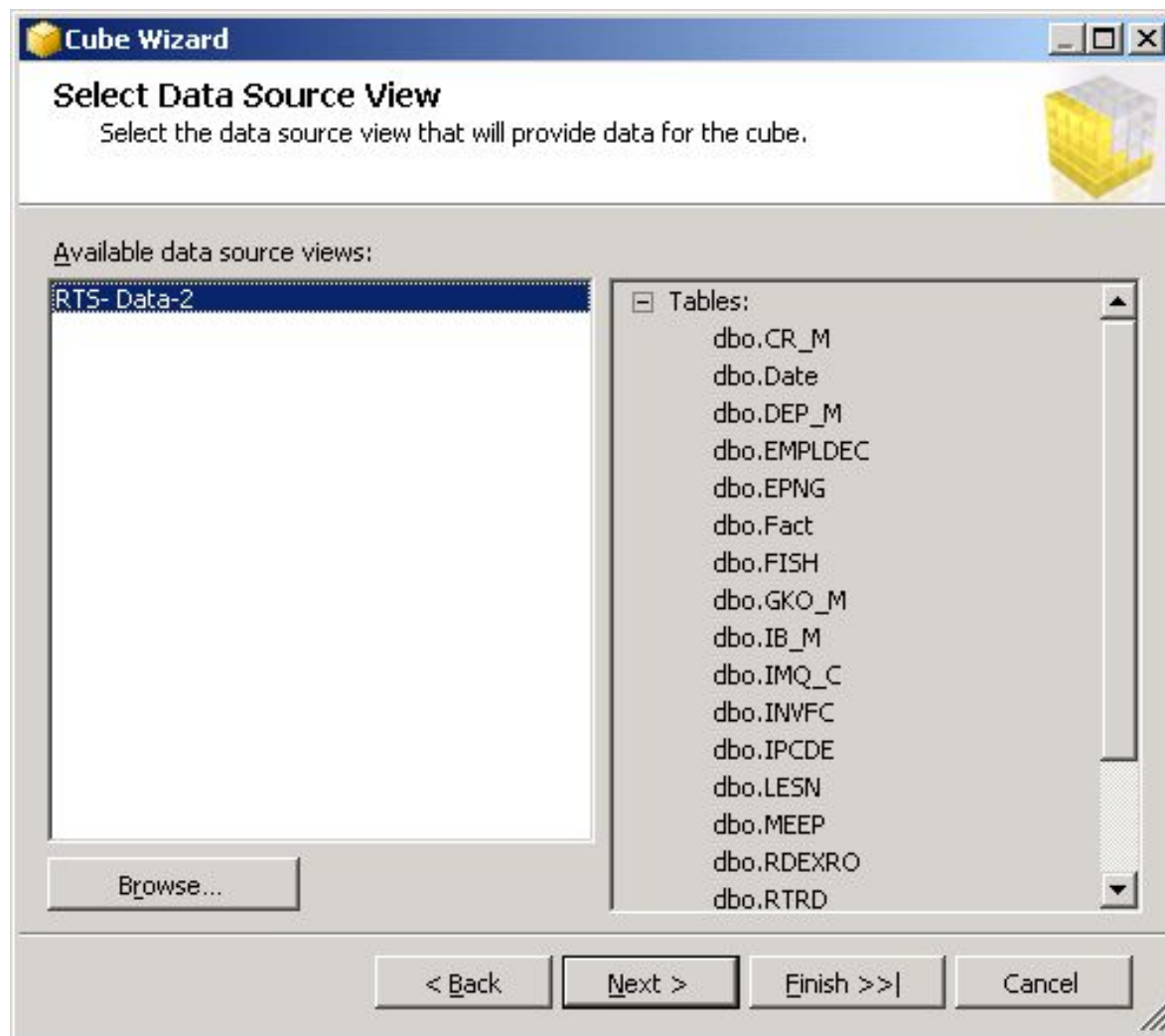
К простым размерностям относятся:

- размерности, состоящие из одной таблицы (сбалансированные и несбалансированные);
- размерности, состоящие из нескольких таблиц (сбалансированные и несбалансированные);
- размерности типа «родитель-потомок».

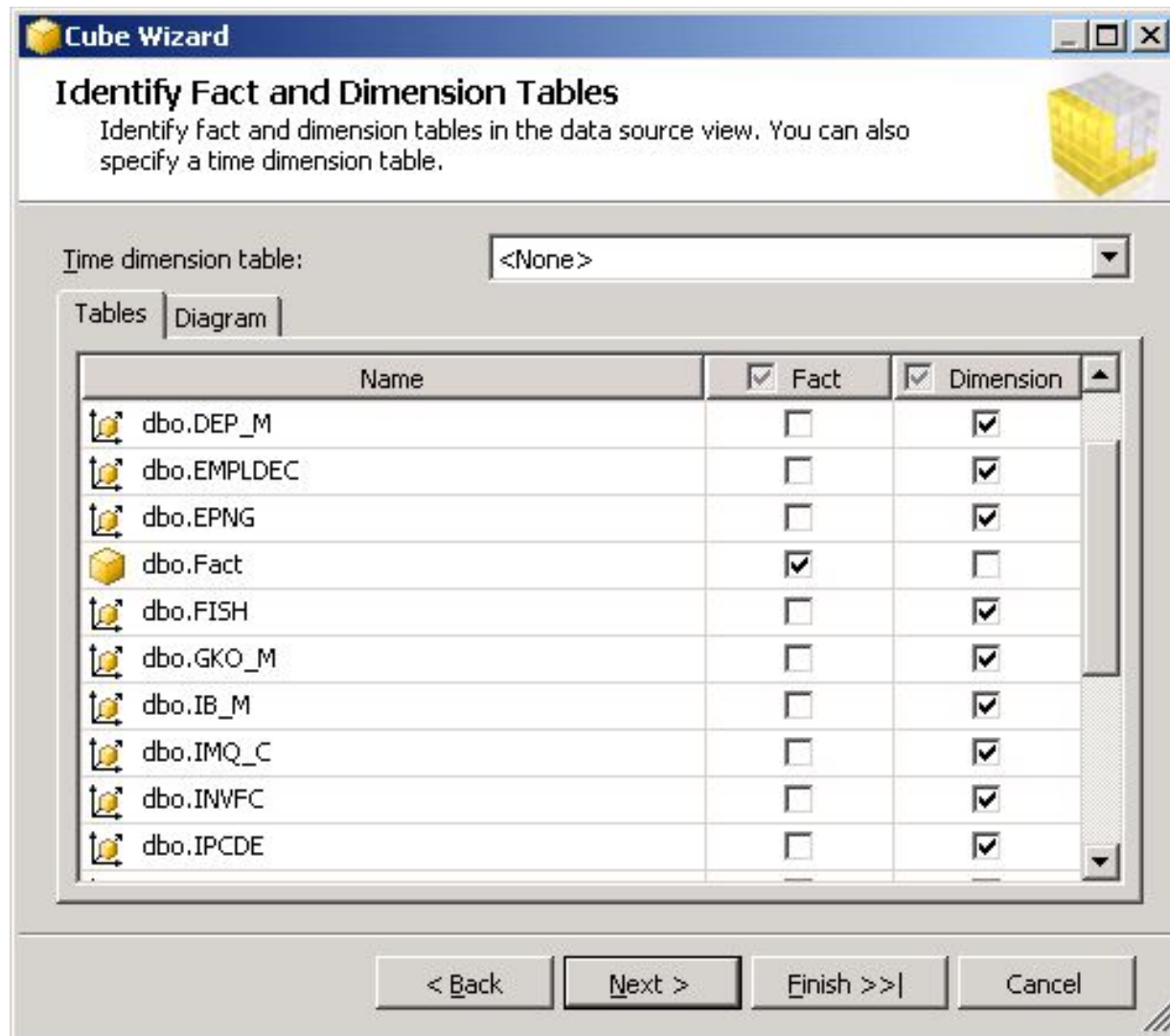
Команда на создание нового куба



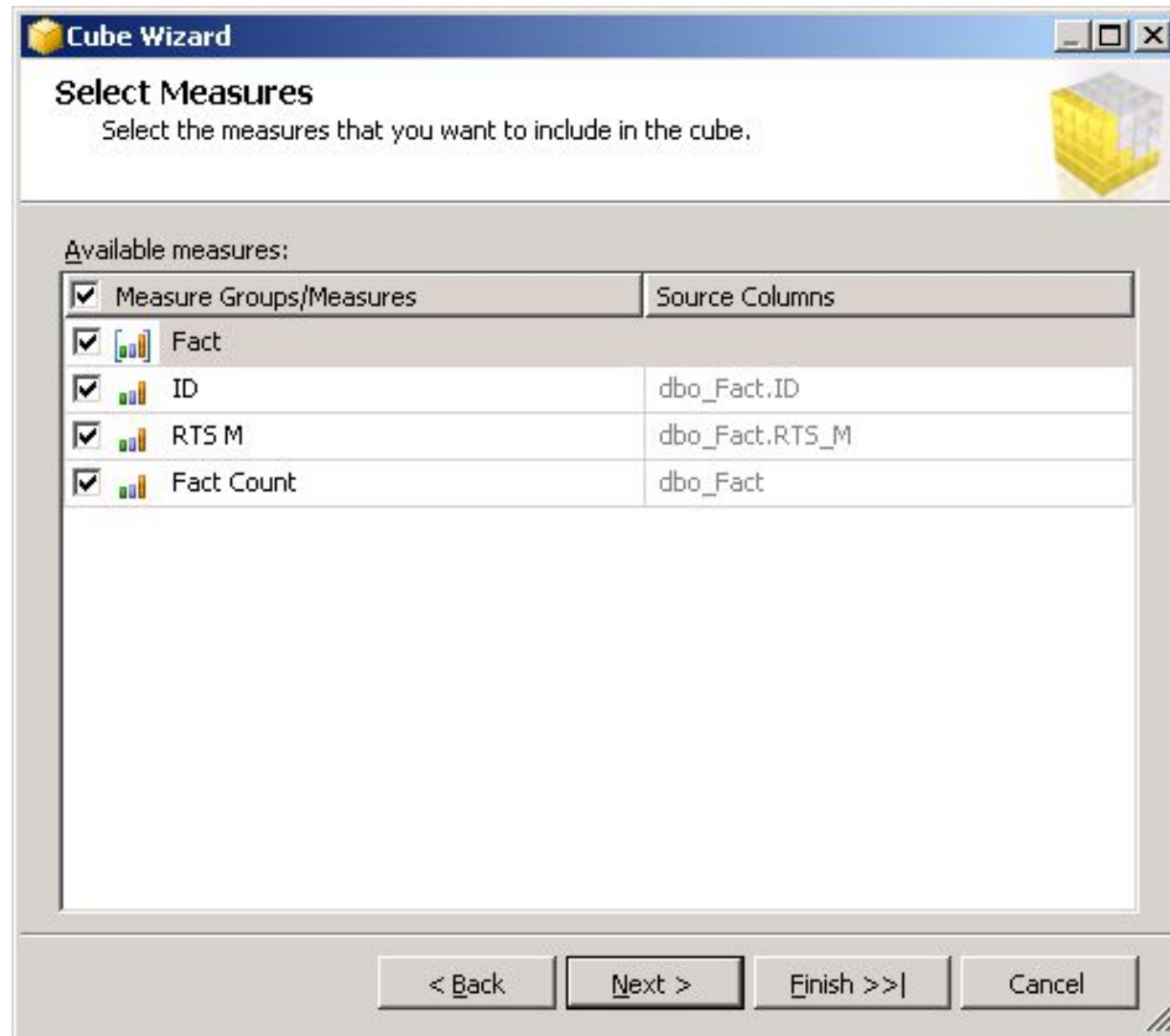
Выбор источника данных в виде реляционного хранилища



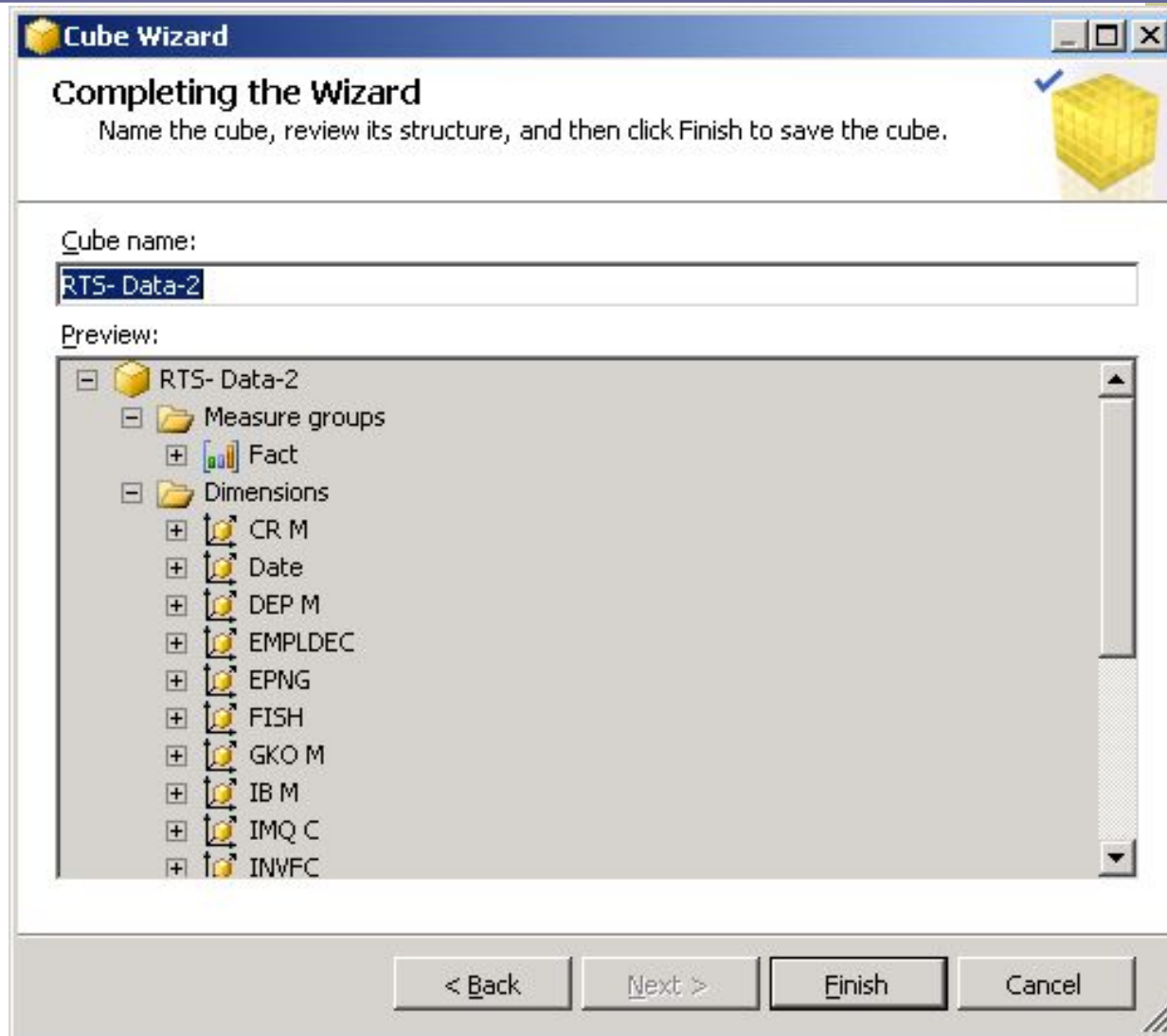
Распределение таблиц среза по измерениям и фактам куба



Окно определения фактов



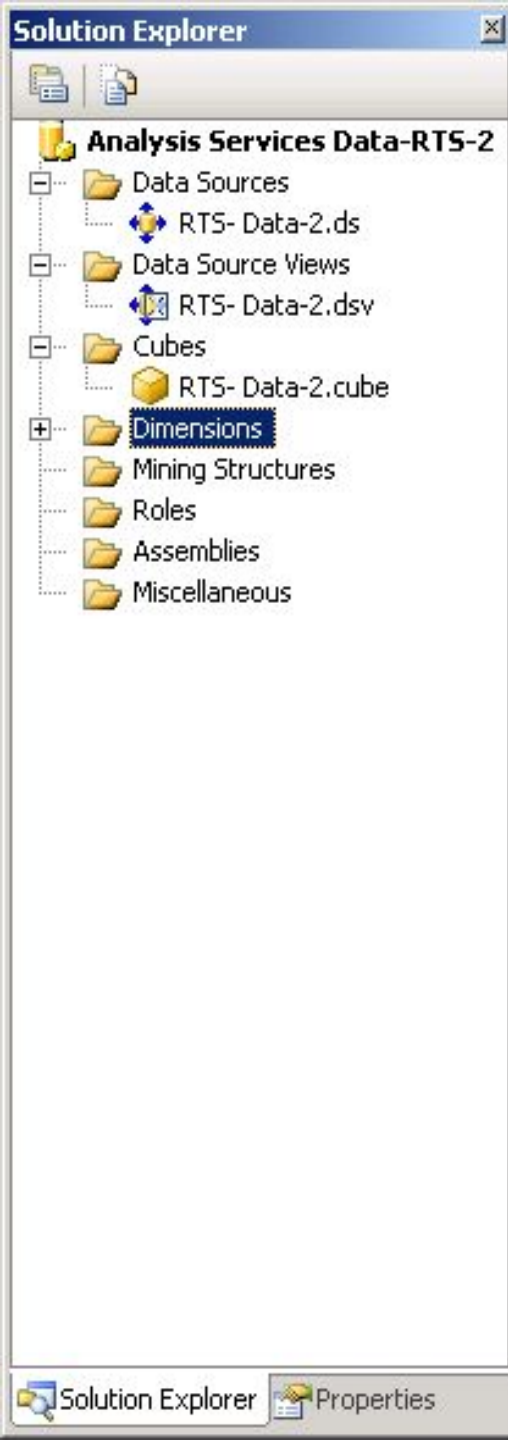
Сохранение куба



5.5. Расчет (процессинг) куба

- В панели **Solution Explorer** (Проводник решений) в проекте **Analysis Services RTS-Data-2** на кубе **RTS-Data-2** (Индекс РТС) щелкните правой кнопкой мыши и затем выберите **Process** (Процессинг) из контекстного меню, произойдет процессинг куба.

- В панели **Process Cube — Main Sale Measures** (Процессинг куба — Главные показатели сбыта) щелкните **Run** (Выполнить); выполнится процессинг.



Куб создан по имени RTS-Data-2.cube



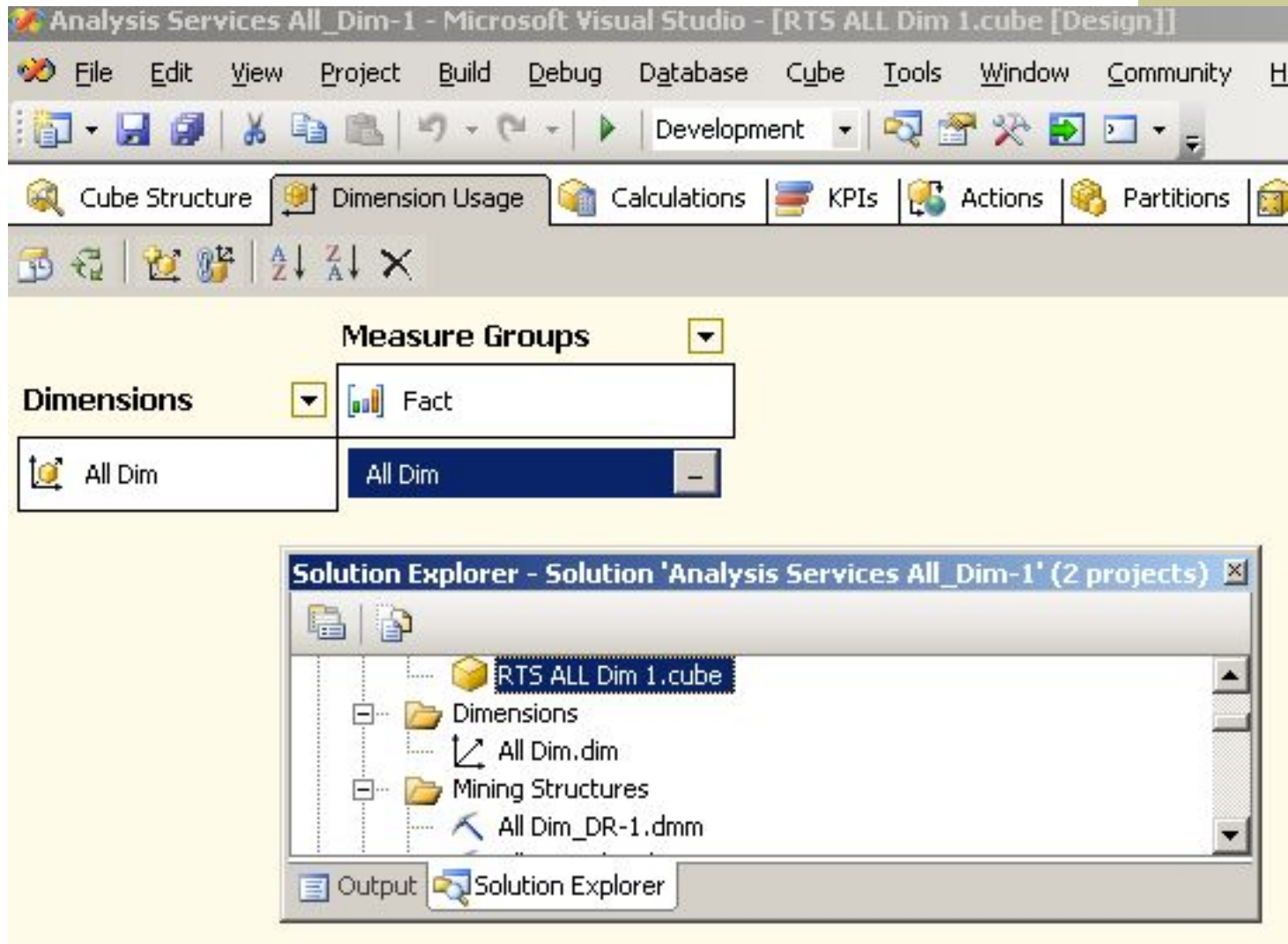
6. Подключение размерности типа «Время и Дата»

6.1. Измерение «Дата» можно подключить двумя способами:

- 1) при построении куба указать, что это временное измерение и
- 2) добавить измерение к уже построенному кубу.

Мы демонстрируем второй способ.

6.2. Вид вкладки Dimension Usage (Использование размерности)



6.3. Добавление нового измерения

Measure Groups

Dimensions

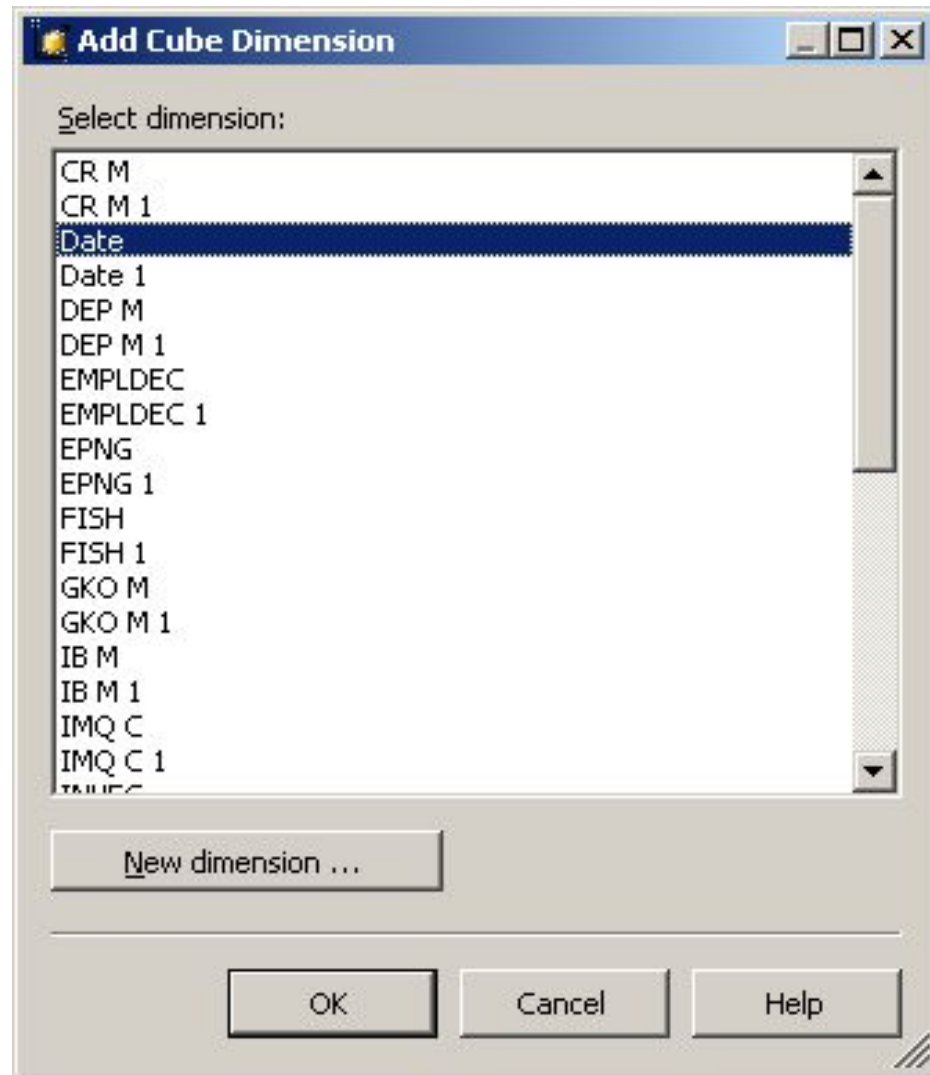
Fact

DEP M 1	DEP M
EMPLDEC 1	EMPLDEC
EPNG 1	EPNG
FISH 1	FISH
GKO M 1	GKO M
IB M 1	IB M
IMQ C 1	IMQ C
INVFC 1	INVFC
IPCDE 1	IPCDE
LESN 1	LESN
MEEP 1	
RDEXRO	
RTRD 1	
UNEMPL	
WAG C M	

Context menu for MEEP 1:

- Add Cube Dimension...
- New Linked Object...
- Cut
- Copy
- Paste
- Delete
- Rename

6.4. Подключаемое измерение (Дата)





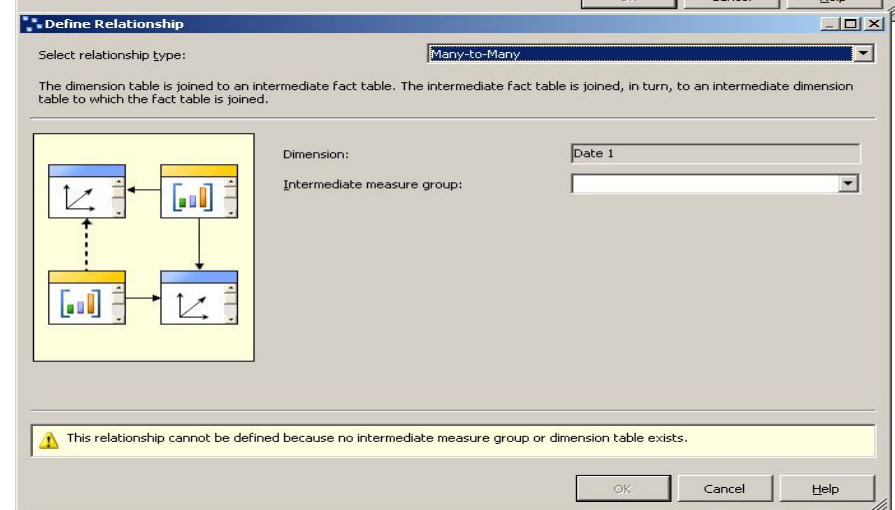
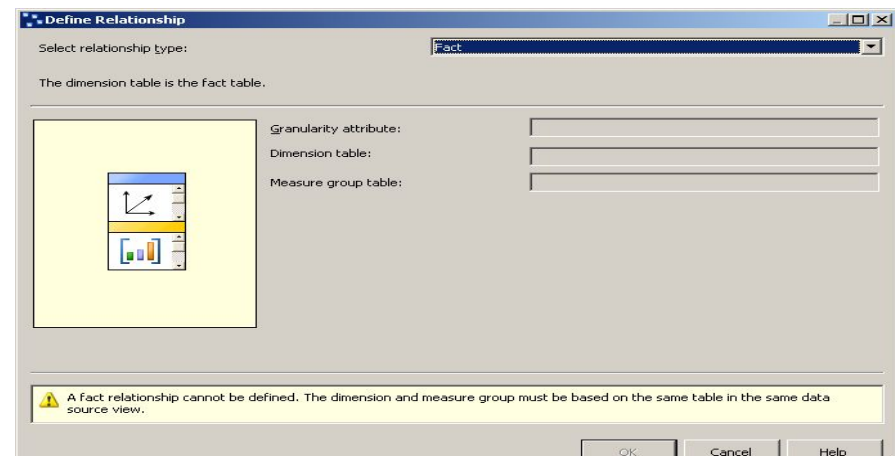
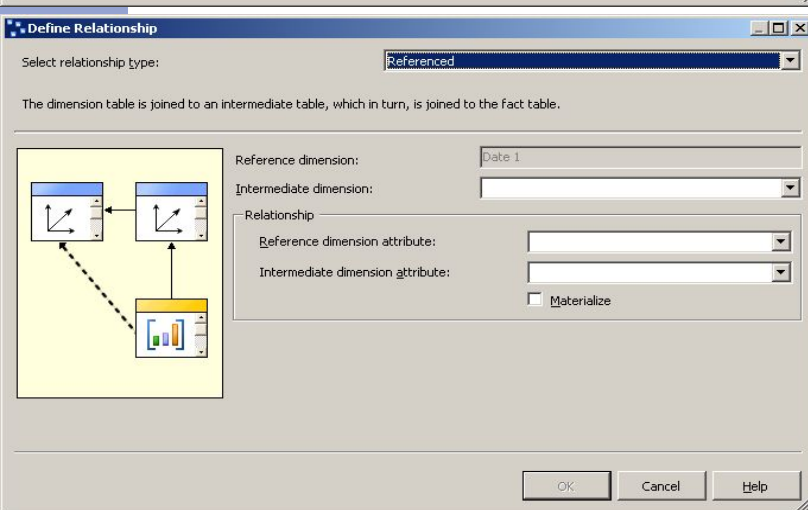
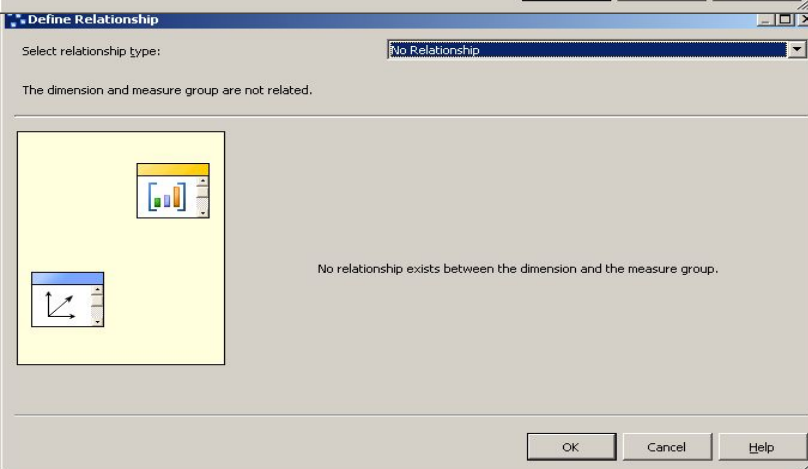
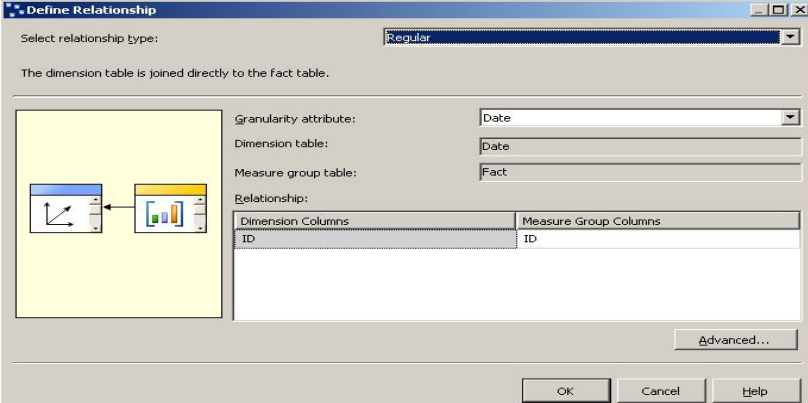
Measure Groups ▼

Dimensions ▼ Fact

EMPLDEC 1	EMPLDEC
EPNG 1	EPNG
FISH 1	FISH
GKO M 1	GKO M
IB M 1	IB M
IMQ C 1	IMQ C
INVFC 1	INVFC
IPCDE 1	IPCDE
LESN 1	LESN
MEEP 1	MEEP
RDEXRO 1	RDEXRO
RTRD 1	RTRD
UNEMPL 1	UNEMPL
WAG C M 1	WAG C M
WAG R M 1	WAG R M
Date	

6.5. В измерениях появилась Date (Дата)

6.6. Выбор типа СВЯЗИ



7. Подключение к кубу размерности, созданной из таблицы фактов

7.1. Команда на добавление новой размерности

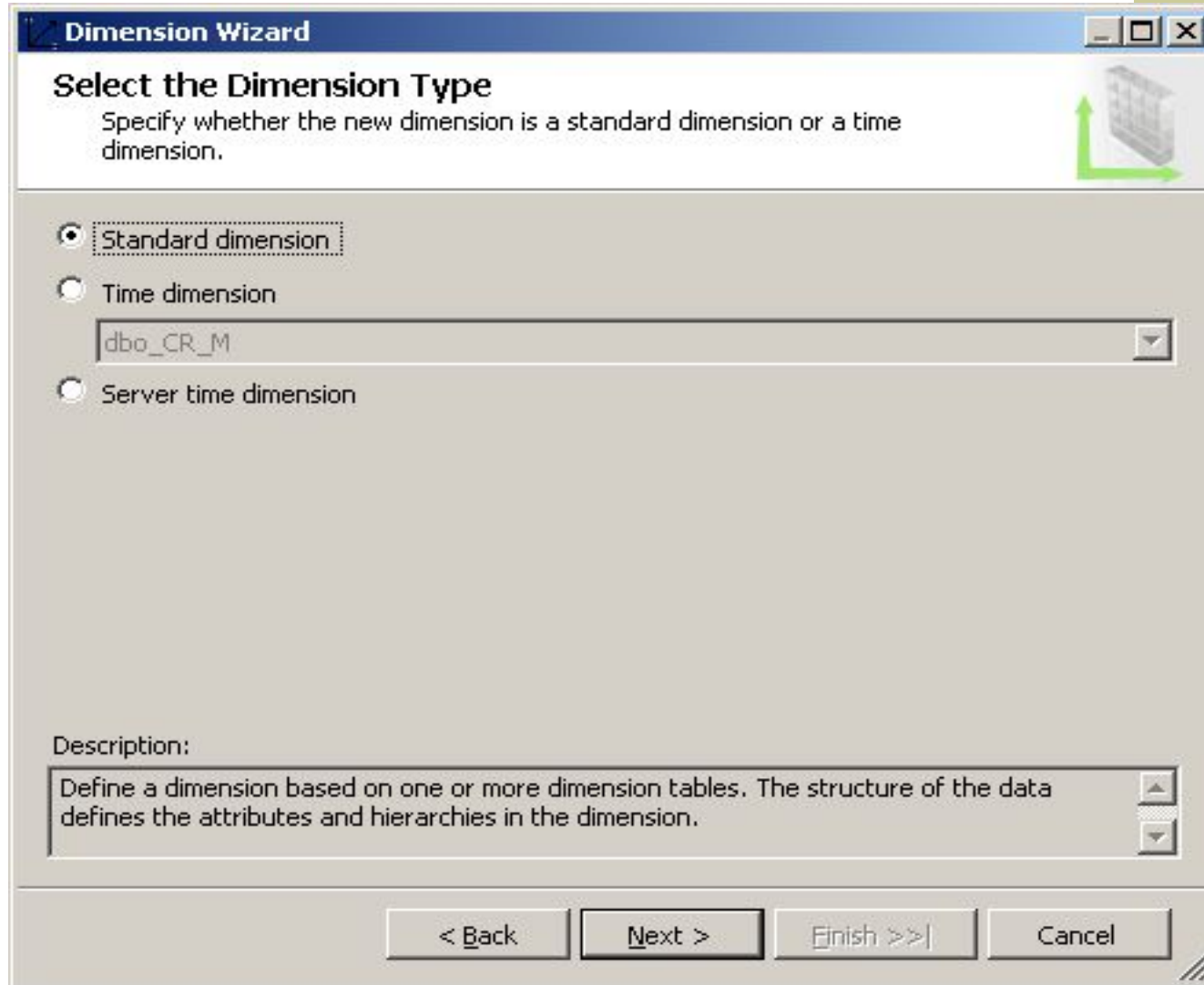
The screenshot displays the 'Dimension Usage' tab in Microsoft SQL Server Enterprise Manager. The main area shows a table of dimension usage for a 'Fact' measure group. A context menu is open over the 'Fact' measure group, with the 'Add Cube Dimension...' option selected. The 'Solution Explorer' on the right shows the project structure for 'Analysis Services RTS-4(HI)'.

Dimensions	Measure Groups
DEP M 1	DEP M
EMPLDEC 1	EMPLDEC
EPNG 1	EPNG
FISH 1	FISH
GKO M 1	GKO M
IB M 1	IB M
IMQ C 1	IMQ C
INVFC 1	INVFC
IPCDE 1	IPCDE
LESN 1	LESN

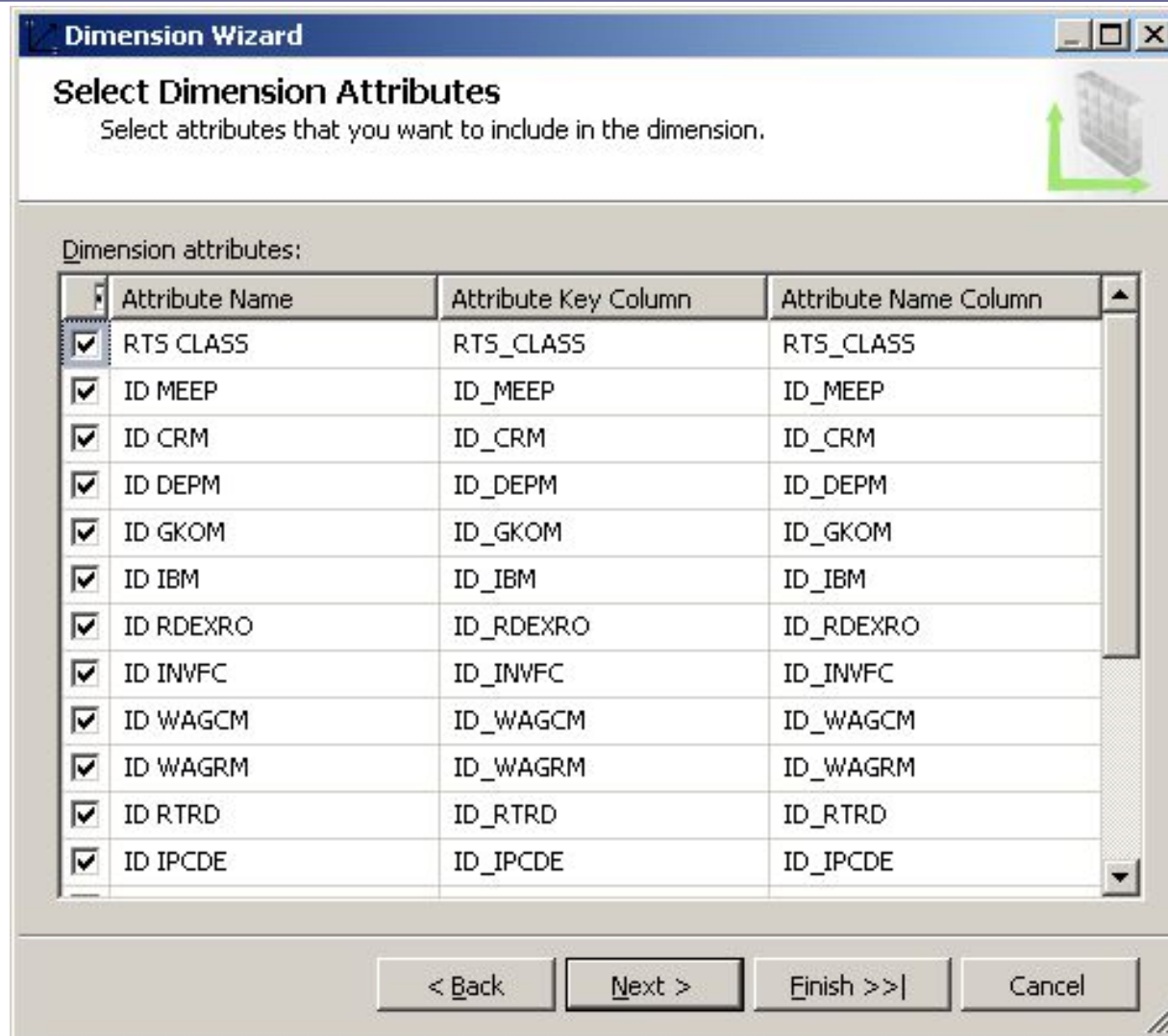
Solution Explorer

- Analysis Services RTS-4(HI)
 - Data Sources
 - Data-RTS-2
 - Data Source Views
 - Data-RTS-5
 - RTS- Data-2
 - Cubes
 - Data-RTS-5
 - Data-RTS-6
 - RTS- Data-2
 - RTS- Data-2-No_Data

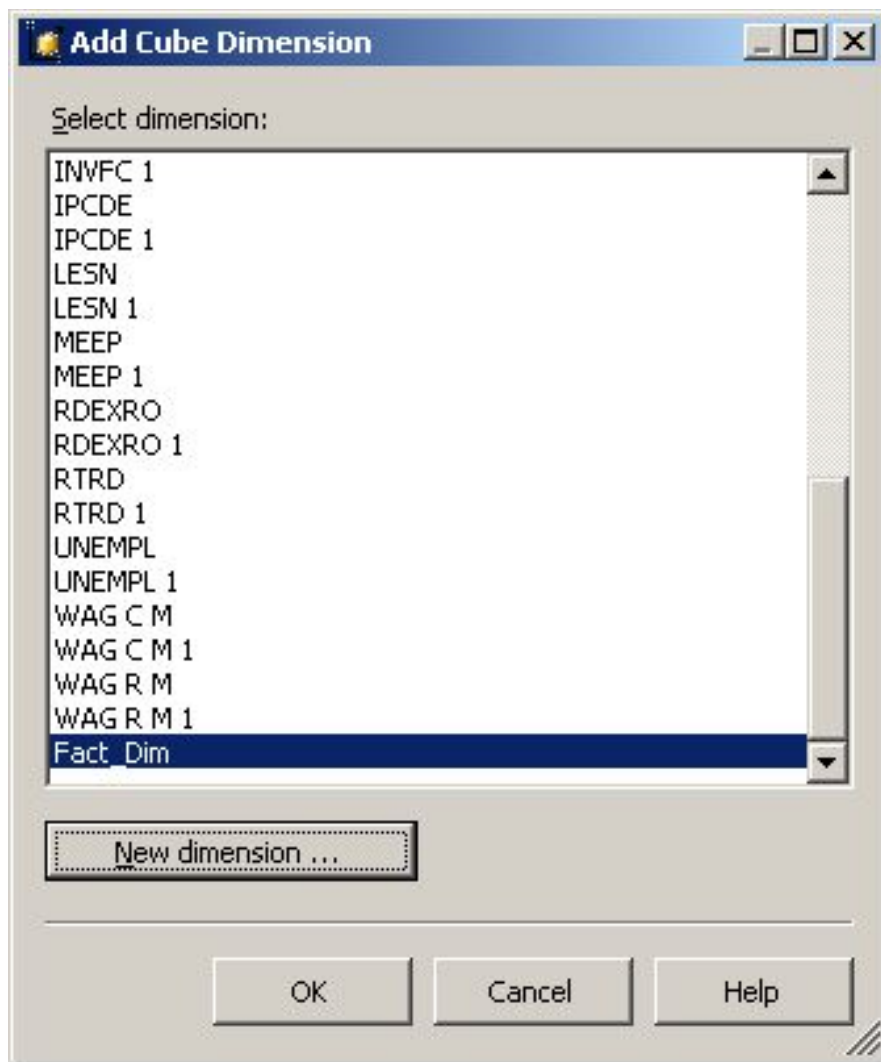
7.2. Выбор типа нового измерения



7.3. Выбор необходимых атрибутов



7.4. В списке измерений появилось вновь созданное (Fact Dim)

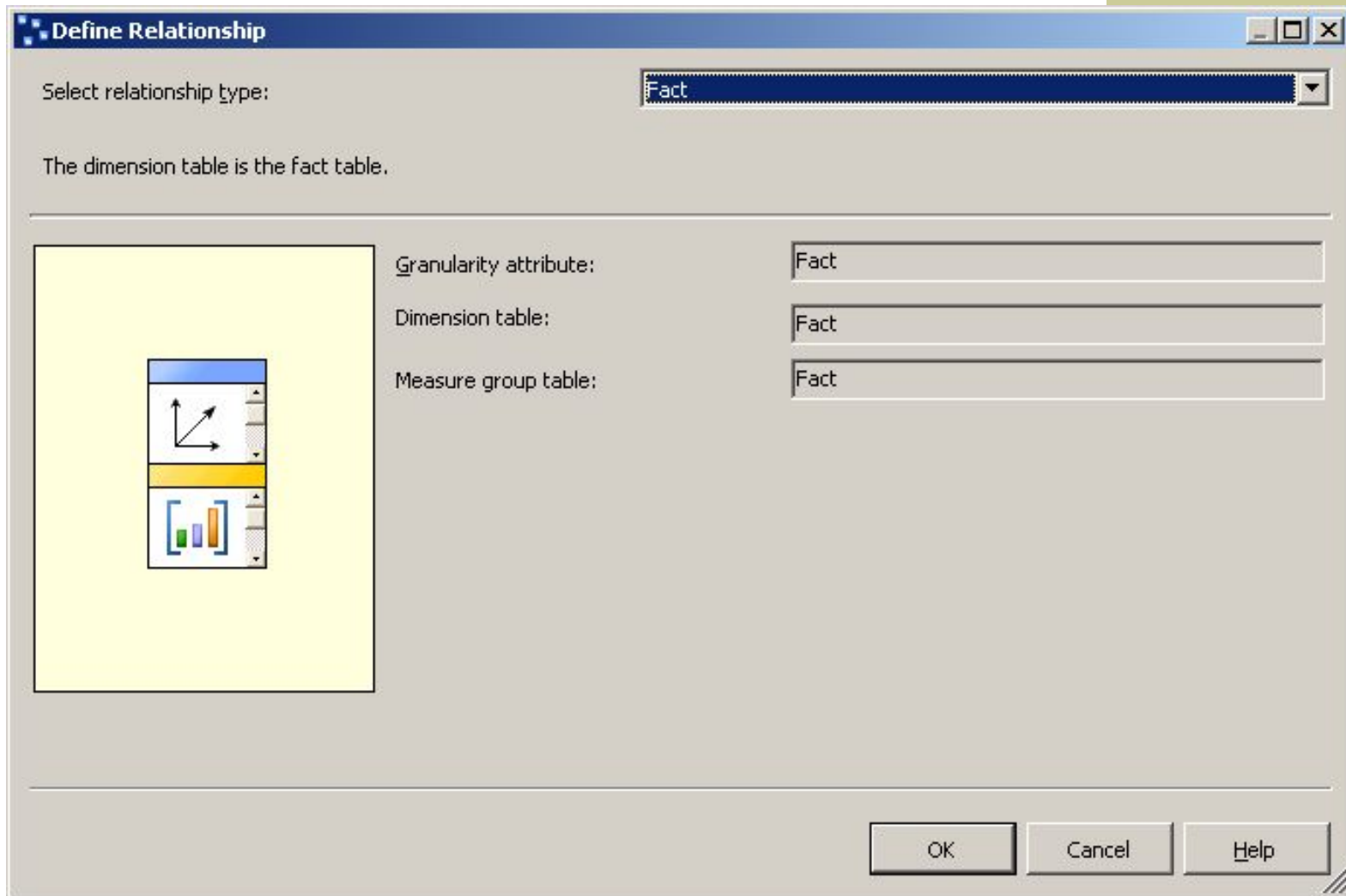


The screenshot shows the 'Dimensions' pane in the Analysis Services Design view. The 'Measure Groups' dropdown is set to 'Fact'. The list of dimensions and their associated fact tables is as follows:

Dimensions	Fact
EMPLDEC 1	EMPLDEC
EPNG 1	EPNG
FISH 1	FISH
GKO M 1	GKO M
IB M 1	IB M
IMQ C 1	IMQ C
INVFC 1	INVFC
IPCDE 1	IPCDE
LESN 1	LESN
MEEP 1	MEEP
RDEXRO 1	RDEXRO
RTRD 1	RTRD
UNEMPL 1	UNEMPL
WAG C M 1	WAG C M
WAG R M 1	WAG R M
Fact_Dim	Fact

7.5. Панель Analysis Services RTS-Data-2.cube [Design] (Построитель (дизайнер) куба)

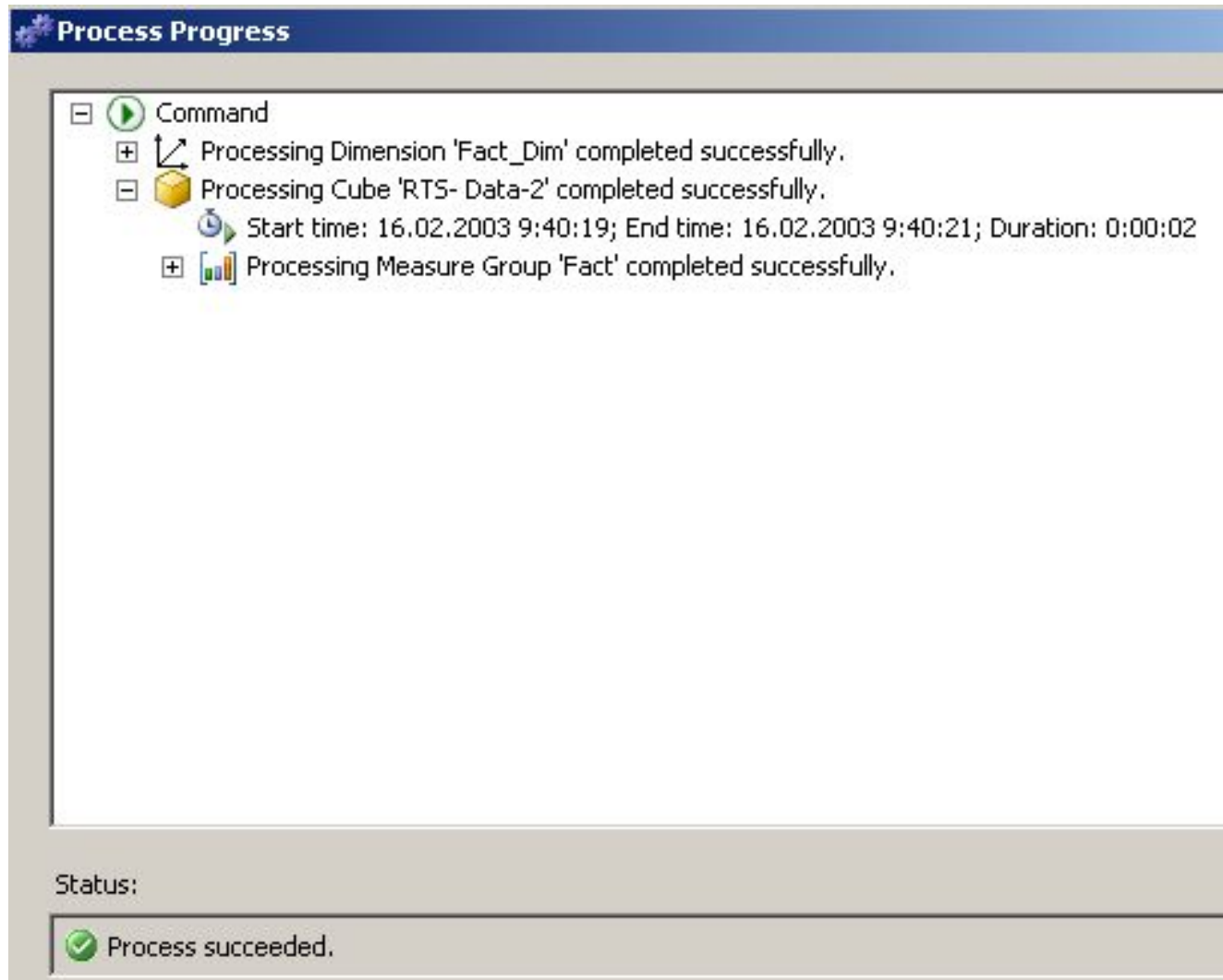
7.6. Выбор связи нового измерения и факта



Measure Groups	
Dimensions	Fact
EMPLDEC 1	EMPLDEC
EPNG 1	EPNG
FISH 1	FISH
GKO M 1	GKO M
IB M 1	IB M
IMQ C 1	IMQ C
INVFC 1	INVFC
IPCDE 1	IPCDE
LESN 1	LESN
MEEP 1	MEEP
RDEXRO 1	RDEXRO
RTRD 1	RTRD
UNEMPL 1	UNEMPL
WAG C M 1	WAG C M
WAG R M 1	WAG R M
Fact_Dim	Fact

7.7. После выбора связи пересечение между Fact_Dim и Fact должно стать определенным

7.8. Процессинг обновленного куба



The screenshot displays the 'Process Progress' window with a tree view of the execution process. The root node is 'Command', which is expanded to show three sub-items: 'Processing Dimension 'Fact_Dim' completed successfully.', 'Processing Cube 'RTS- Data-2' completed successfully.', and 'Processing Measure Group 'Fact' completed successfully.'. The second item is further expanded to show a clock icon and the text 'Start time: 16.02.2003 9:40:19; End time: 16.02.2003 9:40:21; Duration: 0:00:02'. At the bottom of the window, a 'Status:' bar shows a green checkmark icon and the text 'Process succeeded.'

Process Progress

- [-] [▶] Command
 - [+] [↗] Processing Dimension 'Fact_Dim' completed successfully.
 - [-] [📦] Processing Cube 'RTS- Data-2' completed successfully.
 - [🕒] Start time: 16.02.2003 9:40:19; End time: 16.02.2003 9:40:21; Duration: 0:00:02
 - [+] [📊] Processing Measure Group 'Fact' completed successfully.

Status:

[✔] Process succeeded.