

Практическая работа № 3

Создание многомерных баз данных

Цель работы: изучение приемов создания многомерных баз данных и OLAP-кубов систем поддержки принятия решений корпоративных информационных систем.

Задание: используя утилиту Analysis Manager СУБД MS SQL Server создать OLAP-куб и многомерное хранилище данных для анализа продаж в различных разрезах:

- 1) описать источники данных для OLAP-куба;
- 2) создать временные и регулярные измерения;
- 3) сформировать OLAP-куб и многомерное хранилище;

3.1 Основные положения

Теоретически OLAP-куб, созданный с помощью аналитических служб Microsoft, может содержать все данные из таблицы фактов плюс агрегатные значения для тех групп записей из этой таблицы, которые соответствуют верхним уровням иерархии измерений. При необходимости можно производить динамическое обновление куба, если в таблицу фактов были добавлены новые записи, а также выбрать, будут ли данные с нижних уровней иерархии храниться в самом кубе, что соответствует способу хранения данных Multidimensional OLAP, или они будут считываться из таблицы фактов хранилища данных, что соответствует способам хранения данных Relational OLAP и Hybrid OLAP. С точки зрения пользователя различий между этими способами хранения нет, не считая разницы в производительности обращающихся к этим кубам приложений.

Аналитические службы сохраняют агрегатные данные только для простейших агрегатных функций (сумм, числа записей, максимальных и минимальных значений). Однако в случае необходимости можно создавать так называемые вычисляемые члены (calculated members) для получения других типов агрегатных значений (средних, средневзвешенных, смещенных и несмещенных дисперсий и т.д.). При этом, помимо применения встроенных средств создания агрегатных данных, Analysis Services позволяет

использовать для вычисления агрегатных данных функции VBA или Excel, а также создавать собственные.

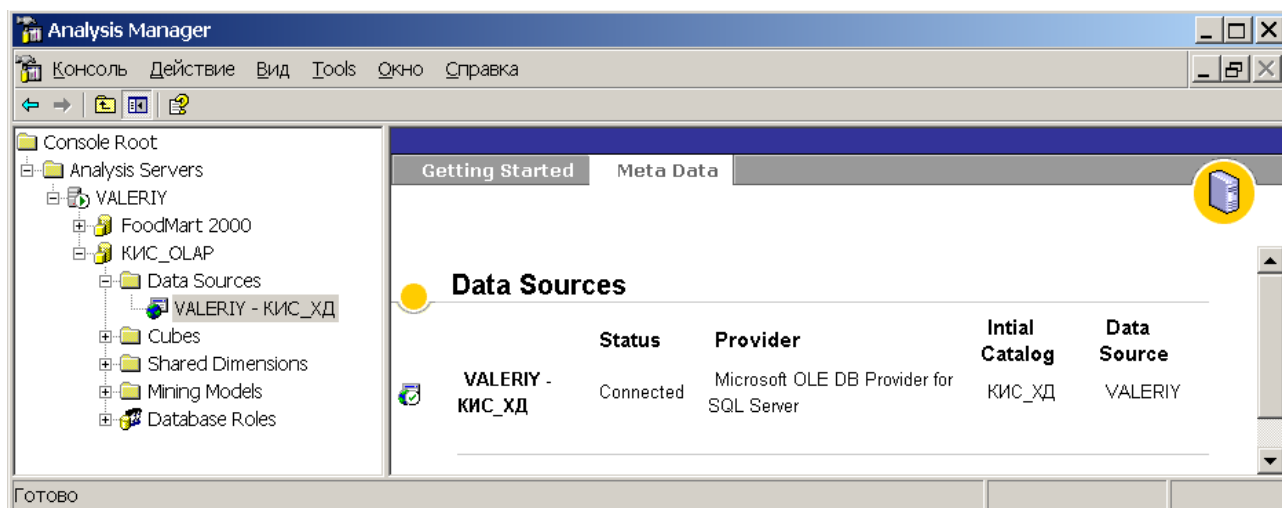
Так, для создания нескольких кубов, имеющих одинаковые измерения, можно сгруппировать их в одну многомерную базу данных, а сами эти измерения поместить в библиотеку (library), сделав их коллективными, то есть общедоступными для всех кубов, содержащихся в базе данных (shared dimensions). Можно также создавать измерения, принадлежащие только одному кубу (private dimensions). И, наконец, аналитические службы Microsoft позволяют создавать так называемые виртуальные кубы (virtual cubes), которые в определенной степени являются аналогами представлений (view) реляционных СУБД. Виртуальные кубы не содержат данных, но позволяют представить в виде единого куба данные из нескольких кубов, имеющих хотя бы одно общее коллективное измерение.

3.2 Порядок выполнения работы

3.2.1 Создание многомерных баз данных и описание источников данных

Для создания многомерного OLAP-куба на основании ранее созданного хранилища данных КИС_ХД следует запустить утилиту Analysis Manager, зарегистрировать в нем OLAP-сервер, выбрав пункт Register Server из контекстного меню элемента Analysis Servers в левой части главного окна Analysis Manager. Затем нужно соединиться с OLAP-сервером, выбрав пункт Connect контекстного меню соответствующего элемента.

Поскольку OLAP-кубы хранятся в многомерных базах данных, создадим таковую, выбрав пункт New Database из контекстного меню элемента, соответствующего OLAP-серверу, и введем имя базы данных и ее описание. Прежде чем создавать OLAP-кубы, необходимо описать источники исходных данных для них. Для описания источника данных выберем из контекстного меню элемента Data Sources пункт New Data Sources, заполним поля стандартной диалоговой панели Data Link Properties (Свойства связи с данными), в качестве провайдера данных указав OLE DB Provider for SQL Server и выбрав базу данных КИС_ХД.

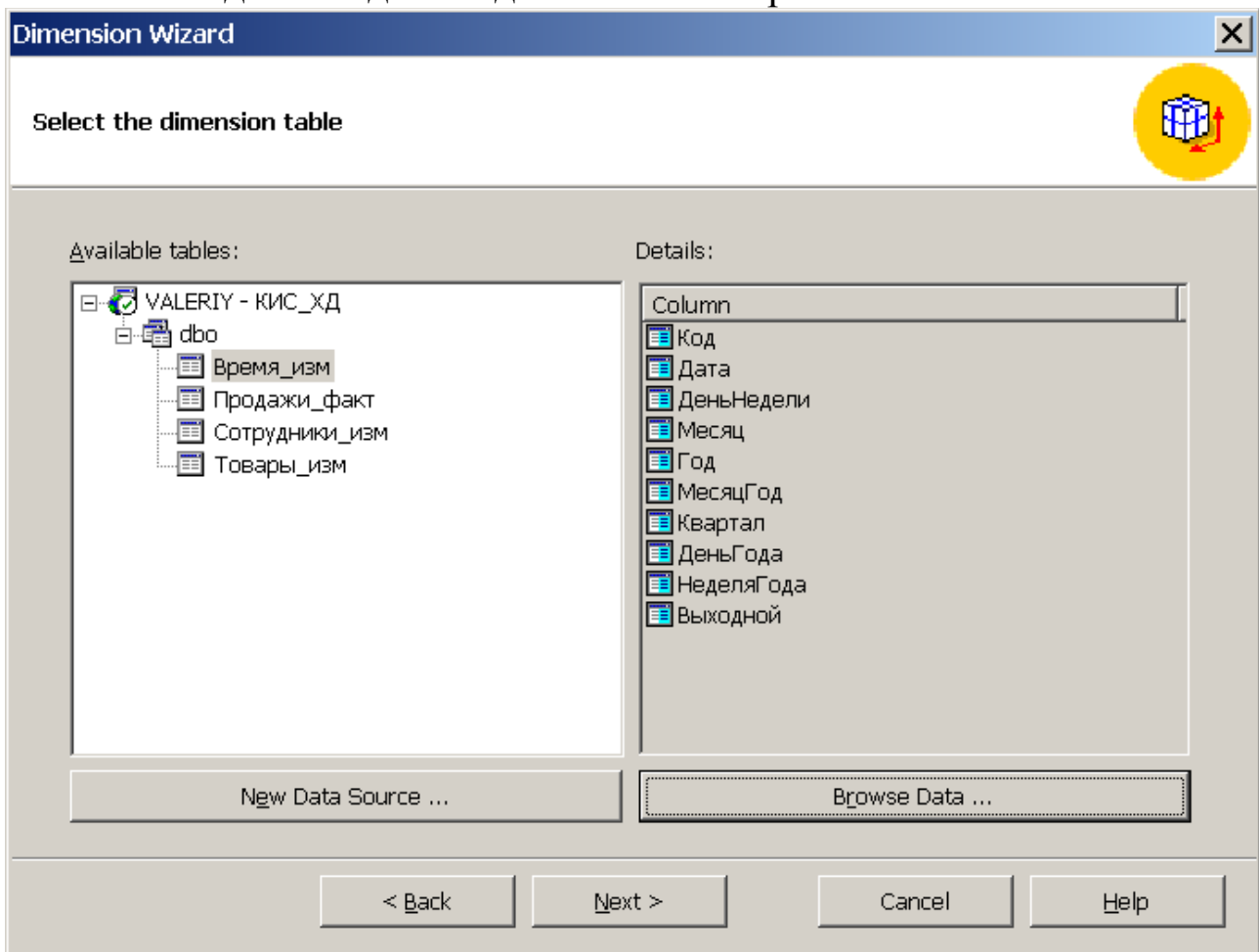


3.2.2 Создание измерения типа «дата/время»

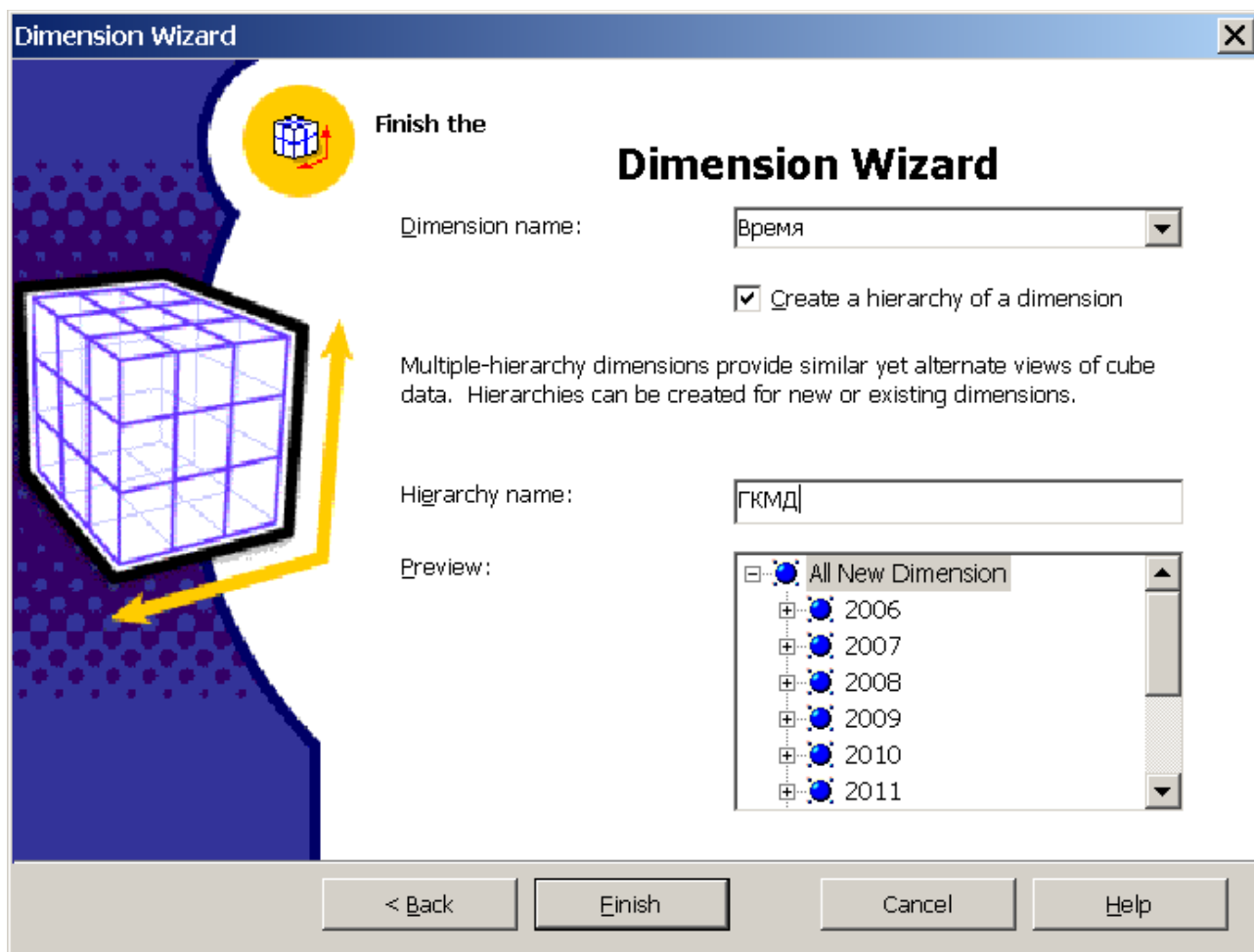
В Microsoft SQL Server Analysis Services измерения делятся на коллективные (shared dimensions), частные (private dimensions) и виртуальные (virtual cubes). Коллективные измерения это измерения, которые могут быть использованы одновременно в нескольких кубах. Такие измерения основаны на стандартных данных, применимых при анализе различных областей деятельности компании (н-ер, списки сотрудников). Коллективные измерения принадлежат самой многомерной базе данных и не зависят от того, какие кубы имеются в многомерной базе данных и есть ли они там вообще. Частные измерения принадлежат конкретному кубу и создаются вместе с ним. Они применяются в том случае, когда данное измерение имеет смысл только в одной конкретной предметной области. Создать как коллективное, так и частное измерение можно двумя способами: с помощью соответствующего мастера и с помощью редактора измерений. Виртуальные кубы, которые в определенной степени являются аналогами представлений (view) реляционных СУБД. Виртуальные кубы не содержат данных, но позволяют представить в виде единого куба данные из нескольких кубов, имеющих хотя бы одно общее коллективное измерение.

Создадим коллективное измерение, основанное на таблице хранилища данных Время_изм, воспользовавшись мастером создания измерений (Dimension wizard). Запустить его можно с помощью команды New Dimension/Wizard из контекстного меню элемента Shared Dimensions. В первую очередь следует выбрать, на основании чего создается измерение. Поскольку исходное хранилище данных основано на схеме «звезда», следует выбрать в мастере опцию Star

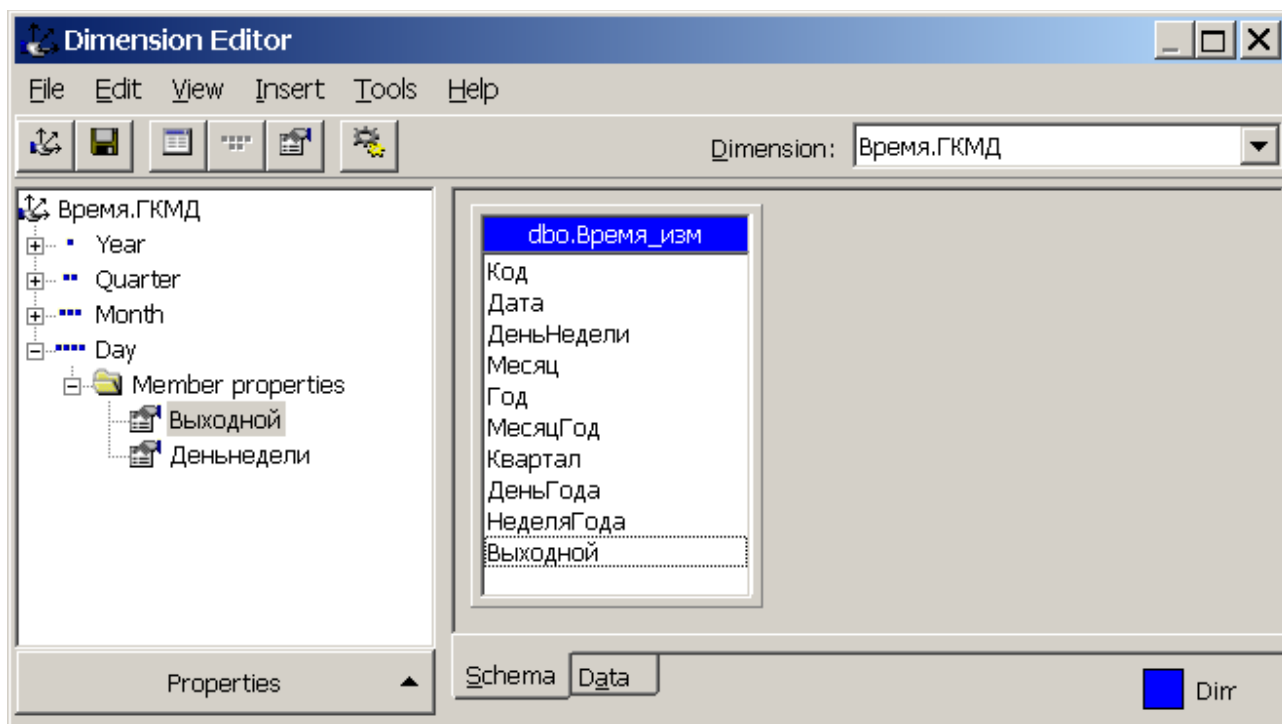
Schema: a single dimension table, а затем имя таблицы, служащей источником данных для создаваемого измерения.



Иерархия данных в измерениях, основанных на данных типа «дата/время», подчиняется определенным стандартным правилам – ведь время измеряется в годах, месяцах, днях, часах, минутах независимо от того, какую предметную область мы анализируем. Поэтому измерения в OLAP-средствах обычно делятся на стандартные (не имеющие отношения ко времени) и временные. Поскольку наше измерение относится к последним, в диалоговой панели Select the dimension type выберем опцию Time Dimension и в качестве колонки, в которой содержатся данные типа «дата/время», укажем поле Дата.



Теперь необходимо выбрать уровни иерархии измерений (например, решить, интересна ли нам информация о часах и минутах, нужны ли нам номера недель года и т.д.), а также определить, когда начинается год с точки зрения данного измерения, поскольку во многих странах начало финансового года не совпадает с началом года календарного. В нашем случае выберем уровни Year, Quarter, Month, Day и согласимся с тем, что год начинается 1 января.



Далее нам предстоит выбрать, является ли измерение изменяющимся (changing dimension). В изменяющихся измерениях можно перемещать члены измерений между уровнями без перерасчета данных измерения. Однако временные измерения не делают изменяющимися – никто не перемещает месяцы из одного года в другой. В заключение требуется указать имя будущего измерения (н-ер, Время) и, если есть необходимость, создать иерархию в измерении и также задать ее имя (н-ер, ГКМД). Дело в том, что при необходимости можно создать еще одно измерение, основанное на тех же данных, с тем же именем, но с другой иерархией, например Year, Week, Day; в этом случае мы имеем разное представление одних и тех же данных.

Создание измерения заканчивается запуском редактора измерений Dimension Editor. В нем при необходимости можно внести изменения в структуру измерения, например, добавив дополнительные уровни или свойства членов измерения. Так, если планируется анализ зависимость продаж от дня недели или сравнивать продажи в выходные и будние дни, нужно перенести в раздел Member Properties уровня День поля ДеньНедели, Выходной исходной таблицы. Теперь можно сохранить созданное измерение и закрыть редактор измерений. Повторим все указанные действия, выбрав при этом другую иерархию Year, Week, Day, и назовем вновь созданное измерение Время.ГНД.

3.2.3 Создание регулярного измерения

Следующее коллективное измерение создадим с помощью редактора измерений. Запустить его можно с помощью команды New Dimension/Editor из контекстного меню элемента Shared Dimensions. Далее в диалоге Select the dimension table выберем в качестве источника данных таблицу Товары_изм. В редакторе для создания двух уровней иерархии – Производитель и Наименование – перенесем мышью соответствующие имена полей в левую часть. Дадим измерению имя Товар в поле Name вкладки Basic панели Properties. Поскольку переносить продукты из категории вполне естественно, сделаем это измерение изменяющимся соответствующее свойство доступно на вкладке Advanced. Аналогично создадим измерение Сотрудники.

Dimension Editor

File Edit View Insert Tools Help

Товары

Товары

- Производитель
- Наименование

Properties

Basic Advanced

Changing True
 Write-enabled False

Changing
 Indicates that the dimension is optimized to handle data changes efficiently at the cost of some performance.

Schema Data

Dimension Editor

File Edit View Insert Tools Help

Сотрудники

Сотрудники

- Отдел
- Регион

Properties

Basic Advanced

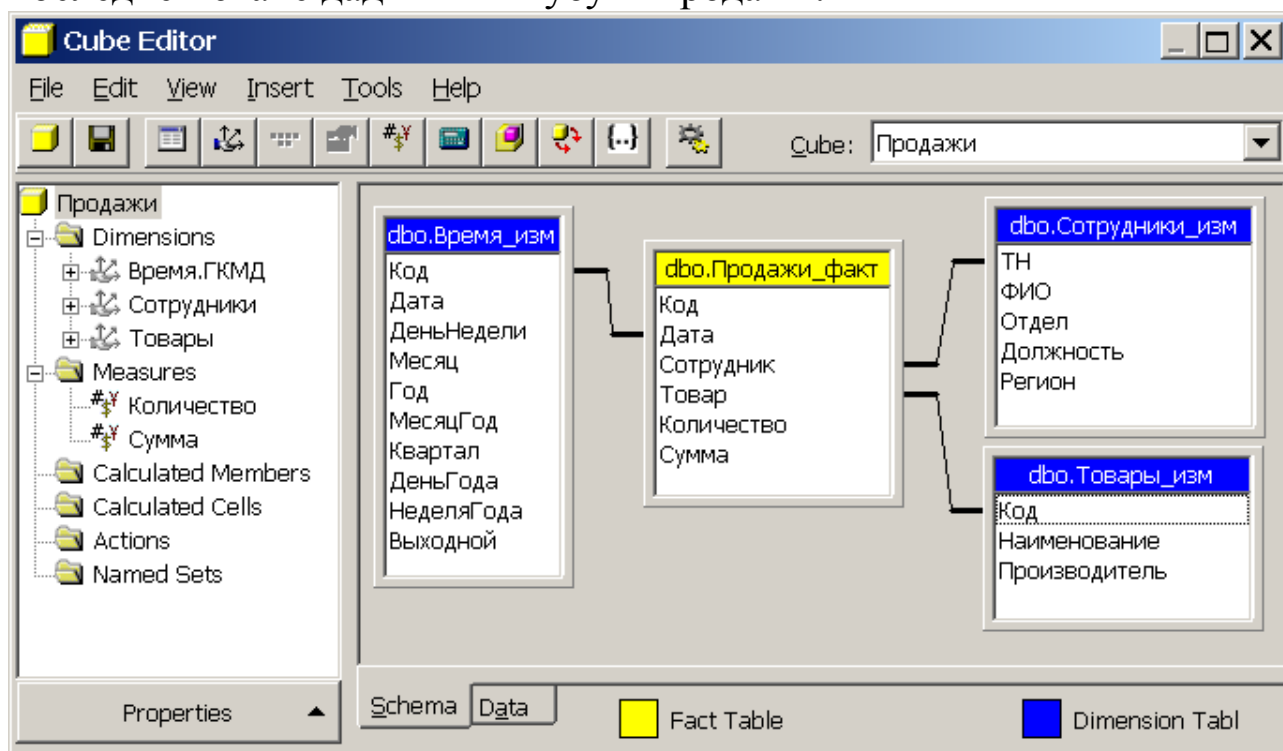
Name Сотрудники
 Description

Dimension
 Dimensions categorize the data of your cube in a hierarchical way. The database dimension

Schema Data

3.2.4 Создание OLAP-кубов

Как и измерение, куб можно создать с помощью соответствующего мастера, запускаемого командой New Cube/Wizard контекстного меню элемента Cubes, или непосредственно в редакторе кубов. В качестве примера создадим куб, основанный на хранилище данных КИС_ХД и использующий созданные выше измерения. Первое, что следует сделать после запуска мастера – выбрать таблицу фактов. В нашем случае это таблица Продажи_факт. Далее из таблицы фактов следует выбрать поля, на основе которых вычисляются меры куба (данные которых подлежат обработке с помощью агрегатных функций): Количество и Сумма. Следующим шагом будет выбор коллективных измерений, используемых в этом кубе, а также создание, при необходимости, недостающих частных измерений. Выберем коллективные измерения Время.ГКМД, Сотрудники, Товары. На последнем этапе дадим имя кубу – Продажи.



Таким образом, мы определили метаданные куба. По окончании работы мастера будет запущен редактор кубов, в котором при необходимости можно внести исправления в определение куба. Добавим к нашему кубу вычисляемые значения, то есть значения, которые не хранятся в самом кубе, а вычисляются «на лету». Примером такого значения может быть дополнительная мера, вычисленная на основе уже имеющихся. При вычислениях можно использовать функции Analysis Services, выражения VBA, а также

собственные библиотеки функций (последние следует зарегистрировать в Analysis Services). Для создания вычисляемых выражений следует выбрать раздел Calculated Members и из контекстного меню выбрать опцию New Calculated Member. После этого будет запущен построитель выражений (Calculated Member Builder), в котором можно создавать и редактировать выражения, перетаскивая мышью имена измерений и их уровней, мер, имена функций. Например, перенесем в поле для выражения имена мер [Measures].[Сумма] и [Measures].[Цена], поставим между ними знак «/», а в качестве значения имени (Member Name) введем СредняяЦена.

Calculated Member Builder

Parent dimension: Measures

Parent member: Change...

Member name: СредняяЦена

Value expression: [Measures].[Сумма]/[Measures].[Количество] Check

Data

- Продажи
 - Measures
 - MeasuresLevel
 - Количество
 - Сумма
 - Время
 - Сотрудники
 - Товары

Functions

- (All)
- Array
- Dimension
- Hierarchy
- Level
- Logical
- Member
- Numeric
- Other
- Set
- String
- Tuple

Insert

9	8	7	/
6	5	4	*
3	2	1	-
0	.	%	+
()	{	}

Register...

Measure: Количество

[Measures].[Количество]

OK Cancel Help

В результате мы получили еще одну меру – значение средней цены за единицу товара. Теперь можно сохранить куб.

3.2.5 Создание многомерного хранилища данных

Процесс создания куба на этом не завершен – было создано только его определение. В настоящее время применяются три способа хранения исходных и агрегатных данных:

MOLAP (Multidimensional OLAP) исходные и агрегатные данные хранятся в многомерной базе данных. Хранение данных в многомерных структурах позволяет манипулировать данными как многомерным массивом, благодаря чему скорость вычисления агрегатных значений одинакова для любого из измерений. Однако в этом случае многомерная база данных оказывается избыточной, так как многомерные данные полностью содержат исходные реляционные данные.

ROLAP (Relational OLAP) исходные данные остаются в той же реляционной базе данных, где они изначально и находились. Агрегатные же данные помещают в специально созданные для их хранения служебные таблицы в той же базе данных.

HOLAP (Hybrid OLAP) исходные данные остаются в той же реляционной базе данных, где они изначально находились, а агрегатные данные хранятся в многомерной базе данных.

Некоторые OLAP-средства поддерживают хранение данных только в реляционных структурах, некоторые только в многомерных. Однако большинство современных серверных OLAP-средств поддерживают все три способа хранения данных. Выбор способа хранения зависит от объема и структуры исходных данных, требований к скорости выполнения запросов и частоты обновления OLAP-кубов.

Для данного примера вполне подойдет хранение всех данных в многомерной базе данных (MOLAP), так как объем исходных данных невелик. Однако в других случаях следует оценить, какой способ хранения наиболее выгоден для данной задачи. Еще один вопрос, который следует решить при создании многомерного хранилища данных – сколько агрегатов следует хранить? Агрегаты – это заранее вычисленные агрегатные данные, соответствующие ячейкам куба. Чем их больше, тем быстрее выполняются запросы к многомерному хранилищу и тем больше объем самого хранилища. Поэтому в общем случае требуется некое их количество, позволяющее осуществить разумный баланс между компактностью и производительностью.

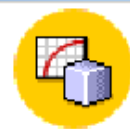
Для определения количества агрегатов и их вычисления следует запустить Storage Design wizard – мастер создания многомерного хранилища. Для этого в редакторе кубов следует выбрать пункт меню Tools/Design Storage. В начале следует указать способ хранения данных – MOLAP, ROLAP или HOLAP (в нашем примере MOLAP).

Затем выбрать, какова должна быть производительность при выполнении запросов (либо будущий максимальный объем хранилища). После этого можно нажать на кнопку Start и получить зависимость производительности от объема хранилища.

Storage Design Wizard



Set aggregation options



Set an aggregation option, and then click Start.

Aggregations are precalculated summaries of data that make querying a cube faster.

Aggregation options

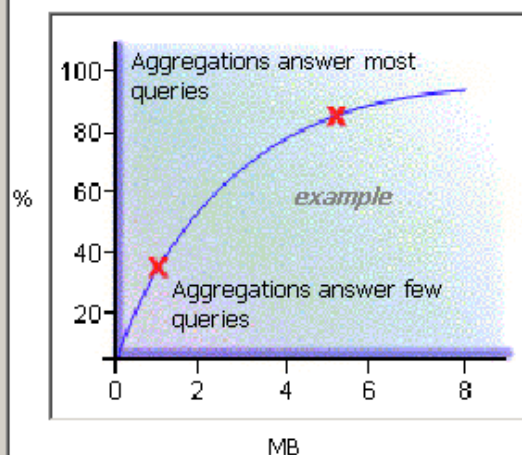
- ☒ Estimated storage reaches MB
- ☐ Performance gain reaches %
- ☐ Until I click Stop

Start

Stop

Reset

Performance vs. Size



< Back

Next >

Cancel

Help

Storage Design Wizard [X]

Set aggregation options [Icon]

Set an aggregation option, and then click Start.

Aggregations are precalculated summaries of data that make querying a cube faster.

Aggregation options

☒ Estimated storage reaches MB

☐ Performance gain reaches %

☐ Until I click Stop

Performance vs. Size

Size (MB)	Performance (%)
0,0	0
0,05	40
0,1	80
0,2	95
0,3	98
0,4	100
0,5	100

35 aggregations designed (0,4 MB, 100%)

На последнем шаге мастера, необходимо вычислить агрегатные данные, выбрав опцию Process now (или в редакторе кубов команду Tools/Process Cube). Просмотр готового куба осуществляется в редакторе на закладке Data. В результате можно получить различные двухмерные сечения куба, перемещая имена измерений на горизонтальную и вертикальную оси, а также скрывая и раскрывая уровни.

