

УДК 004.655.3

МАНИПУЛИРОВАНИЕ ДАННЫМИ И МНОГОМЕРНЫЙ АНАЛИЗ

В.В. Герасимов¹, К.К. Дегтярь², Н.В. Карпенко³

¹ассистент кафедры электронных вычислительных машин, Днепропетровский национальный университет им. Олесь Гончара, г. Днепр, Украина, e-mail: gerasimov@dsu.dp.ua

²студент группы КИ-16с-2, кафедра электронных вычислительных машин, Днепропетровский национальный университет имени Олесь Гончара, г. Днепр, Украина, e-mail: kostya9304@gmail.com

³кандидат физико-математических наук, доцент кафедры электронных вычислительных машин, Днепропетровский национальный университет имени Олесь Гончара, г. Днепр, Украина, e-mail: karpenko_nadija@mail.ru

Аннотация. Рассмотрены возможности платформы фирмы Microsoft SQL Integration Services, в частности её служба Analysis Services. Рассмотрены кубы данных, их поддержка и те преимущества, которые они дают в итоге. Кратко описана технология MDX, необходимая для работы с многомерными хранилищами данных.

Ключевые слова: MDX, SQL, многомерные хранилища данных, Analysis Services, интеграция.

DATA MANIPULATION AND MULTIDIMENSIONAL ANALYSIS

V.V. Gerasimov¹, K.K. Degtyar², N.V. Karpenko³

¹assistant, Computer systems engineering department, Dnepr National University named after Oles' Honchar, Dnepr, Ukraine, e-mail: gerasimov@dsu.dp.ua

²student, Computer systems engineering department, Dnepr National University named after Oles' Honchar, Dnepr, Ukraine, e-mail: kostya9304@gmail.com

³Ph.D. in Physics and Mathematical Sciences, Associate Professor, Computer systems engineering department, Dnepr National University named after Oles' Honchar, Dnepr, Ukraine, e-mail: karpenko_nadija@mail.ru

Abstract. The possibilities of Microsoft SQL Integration Services, its service Analysis Services are considered. Cubes of data, their support and those advantages which they give thus are considered. The technology MDX necessary for operation with multivariate data stores is briefly described.

Keywords: MDX, SQL, multidimensional data warehousing, Analysis Services, integration.

Введение. В сегодняшних реалиях очень трудно представить крупное предприятие, которое не использует базы данных для хранения и дальнейших манипуляций с данными. Не играет роли направление, в котором ра-

ботает предприятие, будь то сеть магазинов, ресторанов или же направления, не связанные с коммерцией. В большинстве случаев базы обширны и данные в них исчисляются сотнями гигабайт информации. Для работы с такими объёмами данных чаще всего используются многомерные хранилища данных, которые позволяют значительно ускорить выборку данных, что положительным образом сказывается на работе предприятия в целом и его управлении в частности.

Цель работы. Рассмотреть возможности MS Integration services и её службы Analysis Services. Обосновать использование новой технологии MDX и создание многомерных хранилищ данных.

Материал и результаты исследований. Без возможности импорта и экспорта информации база данных лишь большое хранилище. Именно поэтому для MS SQL Server была создана платформа Integration Services. Службы Integration Services — это платформа для построения решений по интеграции и преобразованию данных уровня предприятия. Службы Integration Services используются при решении сложных бизнес-задач путем копирования и загрузки файлов, отправки электронных сообщений в ответ на события, обновления хранилищ данных, очистки и интеллектуального анализа данных, а также управления объектами и данными MS SQL Server. При решении сложных бизнес-задач пакеты могут работать как отдельно, так и совместно с другими пакетами. Службы Integration Services могут извлекать и преобразовывать данные из ряда таких источников как файлы XML-данных, неструктурированные файлы и источники реляционных данных, и затем загружать эти данные в один или несколько реляционных объектов.

В состав служб Integration Services входит широкий набор встроенных задач и преобразований, средства для построения пакетов, а также служба Integration Services для выполнения пакетов и управления ими. С помощью графических инструментов служб Integration Services можно создавать готовые решения, не написав ни единой строки кода, либо программировать подробную объектную модель служб Integration Services для программного создания пакетов и создания в программном коде пользовательских задач и других объектов пакета [1].

По своей структуре Integration Services является ядром, на котором базируются другие компоненты MS SQL Server (рис. 1).

Основными компонентами MS SQL Server являются:

- Database Engine;
- Reporting Services;
- Analysis Services.

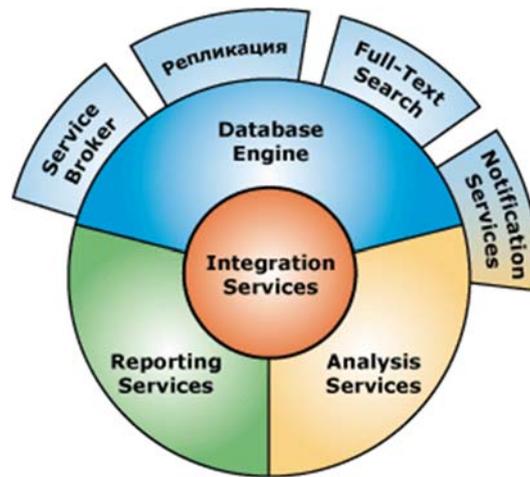


Рисунок 1 – Схема служб Microsoft SQL Server

Рассмотрим данные компоненты более подробно. Database Engine — это основная служба хранения, обработки и защиты данных. Компонент Database Engine обеспечивает средства управляемого доступа и быстрой обработки транзакций, удовлетворяющие требованиям большинства приложений по обработке данных в организации. Кроме того, компонент Database Engine предоставляет разнообразные механизмы, поддерживающие высокий уровень доступности.

Reporting Services (SSRS) обеспечивает функциональные возможности формирования веб-отчетов в рамках предприятия, позволяя создавать отчеты на основе данных из множества источников, публиковать их в разных форматах, централизованно управлять безопасностью и подписками.

Analysis Services (SSAS) — это компонент, предназначенный для создания многомерных баз данных и управления ими. На сегодняшний день это наиболее современная из трех версий подобного программного обеспечения от Microsoft. Analysis Services предоставляет мощные инструменты для разработки, создания, тестирования и развёртывания многомерных баз данных. Основная задача SSAS — создание кубов данных, измерений или так называемых осей куба, а также управление ими и обеспечение поддержки для пользователей BI-проектов. Analysis Services поддерживает язык MDX, с помощью которого возможно запрашивать требуемые данные. MDX предназначен для формирования запросов к многомерным базам данных подобно тому, как язык SQL служит для создания запросов к реляционным базам данных. Язык MDX является компонентом OLE DB для спецификации OLAP и поддерживается другими производителями средств бизнес-аналитики. Программное обеспечение Analysis Services компании Microsoft предлагает определённые расширения, позволяющие добиться большего от своих многомерных баз данных или так называемых кубов.

Основой многомерной базы данных является куб. Каждый куб обычно содержит более двух измерений. Куб представляет собой многомерную структуру данных, из которой можно запрашивать бизнес-информацию. Кубы создаются из фактических данных и измерений. Куб может содержать фактические данные из одной или нескольких таблиц фактов и чаще всего содержит несколько измерений. Любой конкретный куб обычно имеет доминирующий предмет анализа. Например, можно создать куб Sales (Продажи), с помощью которого будете проводиться анализ объема продаж по регионам.

Конечной целью создания кубов является минимизация времени обработки запросов, извлекающих требуемую информацию из фактических данных. Для реализации этой задачи кубы обычно содержат предварительно вычисленные итоговые данные, называемые агрегациями. Возвращение запросом уже существующих агрегированных данных выполняется почти мгновенно в сравнении с выполнением неподготовленных запросов, которые самостоятельно вычисляют итоговые данные. Фактически "сердцем" средств бизнес-анализа является возможность запрашивать информацию, которая вполне может включать гигабайты или даже терабайты предварительно вычисленных данных, и при этом получать немедленный ответ с сервера.

Говоря о многомерных базах данных невозможно не упомянуть о MDX. Подобно тому, как SQL (Structured Query Language — язык структурированных запросов) представляет собой язык создания запросов для извлечения данных из реляционных баз данных, MDX (Multi-Dimensional expressions — язык многомерных выражений) является языком запросов, используемым для извлечения данных из многомерных баз данных. Точнее, MDX используется для запрашивания данных из баз данных OLAP с помощью Analysis Services и поддерживает два особых режима. При использовании в качестве выражений MDX позволяет определять многомерные объекты и данные для вычисления значений, а также управлять ими. Как язык запросов он используется для извлечения данных из баз данных Analysis Services. Изначально MDX был разработан компанией Microsoft и был введен вместе с Analysis Services 7.0 в 1998 году [2].

Язык SQL изначально был спроектирован не для программистов, а для аналитиков и поэтому имеет синтаксис, напоминающий естественный язык, но он со временем все больше усложнялся и теперь мало кто из аналитиков хорошо умеет им пользоваться, если умеет вообще. Он стал инструментом программистов. Язык запросов MDX разработанный Mosha Pasumansky в корпорации Майкрософт, тоже изначально был ориентирован на аналитиков [3].

SQL – мощный инструмент, который обеспечивает пользователям, программам и вычислительным системам доступ к информации, содержащейся в реляционных базах данных.

Основные достоинства языка SQL заключаются в следующем:

- стандартность – использование языка SQL в программах стандартизировано международными организациями;
- независимость от конкретных СУБД – все распространенные СУБД используют SQL, т.е. реляционную базу данных можно перенести с одной СУБД на другую с минимальными доработками;
- возможность переноса с одной вычислительной системы на другую – СУБД может быть ориентирована на различные вычислительные системы, однако приложения, созданные с помощью SQL, допускают использование как для локальных БД, так и для крупных многопользовательских систем;
- реляционная основа языка – SQL является языком реляционных БД, поэтому он стал популярным тогда, когда получила широкое распространение реляционная модель представления данных. Табличная структура реляционной БД хорошо понятна, а потому язык SQL прост для изучения;
- возможность создания интерактивных запросов – SQL обеспечивает пользователям немедленный доступ к данным, при этом в интерактивном режиме можно получить результат запроса за очень короткое время без написания сложной программы;
- возможность программного доступа к БД – язык SQL легко использовать в приложениях, которым необходимо обращаться к базам данных. Одни и те же операторы SQL употребляются как для интерактивного, так и программного доступа, поэтому части программ, содержащие обращение к БД, можно вначале проверить в интерактивном режиме, а затем встраивать в программу;
- обеспечение различного представления данных – с помощью SQL можно представить такую структуру данных, что тот или иной пользователь будет видеть различные их представления. Кроме того, данные из разных частей БД могут быть скомбинированы и представлены в виде одной простой таблицы, а значит, представления пригодны для усиления защиты БД и ее настройки под конкретные требования отдельных пользователей;
- возможность динамического изменения и расширения структуры БД – язык SQL позволяет манипулировать структурой БД, тем самым обеспечивая гибкость с точки зрения приспособленности БД к изменяющимся требованиям предметной области;

- поддержка архитектуры клиент-сервер: SQL – одно из лучших средств для реализации приложений на платформе клиент-сервер. SQL служит связующим звеном между взаимодействующей с пользователем клиентской системой и серверной системой, управляющей БД, позволяя каждой из них сосредоточиться на выполнении своих функций [4].

Язык MDX выполняет те же функции, что и SQL, и в целом похож на него, но SQL не способен работать с многомерными хранилищами данных.

Работа с аналитическими массивами данных подразумевает только их чтение и не подразумевает запись. Значит, в языке MDX нет предложений для изменения данных, а есть только одно предложение выборки — select.

В OLAP из многомерных кубов можно делать срезы – когда данные фильтруются по одной или нескольким осям, или проекции – когда по одному или нескольким осям куб «схлопывается», агрегируя данные [3].

Использование языка MDX не ограничено авторскими правами на продукт Analysis Services. Этот язык используется для извлечения информации из баз данных OLAP, так как основан на стандартах отрасли. Язык является частью спецификации OLEDB для OLAP, финансируемой Microsoft, поддерживается многими другими провайдерами OLAP, включая Intelligence Server компании Microstrategy, Essbase Server компании Hyperion и Enterprise BI Server от SAS. Некоторые компании стремятся расширить стандарт и обеспечить дополнительную функциональность, и вследствие этого расширения MDX, действительно, разработаны индивидуальными производителями. В расширениях MDX реализованы функции, которых в стандарте нет, но предполагается, что все компоненты таких расширений MDX разработаны согласно стандарту. Analysis Services предлагает несколько расширений к стандарту MDX, определяемых спецификацией OLEDB для OLAP.

Хотя язык запросов MDX использует синтаксис, подобный синтаксису языка SQL, они значительно отличаются [2].

Например, если необходимо узнать о продажах сувенирных рукавиц в розничном магазине в Ванкувере, Британская Колумбия, в феврале, когда проводились Олимпийские игры 2010 года, необходимо написать запрос MDX, который будет выглядеть так:

```
SELECT
[Store].[Store Country].[Canada].[Vancouver] ON COLUMNS,
[Product].[All Products].[Clothing].[Mittens] ON ROWS
FROM [Sales]
WHERE ([Measures].[Unit Sales], [Date].[2010].[February])
```

Однако, если есть необходимость написать это в SQL, понадобится более громоздкий SQL-запрос, примерно такой:

```
SELECT SUM(Sales.[Unit Sales])
FROM (Sales INNER JOIN Stores
ON Sales.StoreID = Stores.StoreID)
```

```
INNER JOIN Products
ON Sales.ProductID = Products.ProductID
WHERE Stores.StoreCity = 'Vancouver'
AND Products.ProductName = 'Mittens'
AND Sales.SaleDate BETWEEN '01-02-2010' AND '28-02-2010'
```

Ещё более многословным становится выражение SQL, которое сравнивает продажи за февраль 2010 года с показателями предыдущего месяца или предыдущего года. Напротив, MDX упрощает создание очень коротких лаконичных запросов, которые более понятны для чтения и понимания, что, в свою очередь, уменьшает количество ошибок. Это позволяет написать более общие выражения, например, для сравнения предыдущих периодов, которые на каждом уровне развёртки во временной иерархии (например, день, месяц, квартал, год) показывают или сравнивают с предыдущей единицей времени на этом уровне [5].

OLAP системы поддержки принятия решений обычно обладают средствами предоставления пользователю агрегатных данных для различных выборок из исходного набора в удобном для восприятия и анализа виде (таблицы, диаграммы и т.п.). Традиционный подход сегментирования исходных данных использует выделение из исходных данных одного или нескольких многомерных наборов данных (нередко называемый гиперкубом или метакубом), оси которых содержат атрибуты, а ячейки – агрегируемые количественные данные. Вдоль каждой оси атрибуты могут быть организованы в виде иерархий, представляющих различные уровни их детализации. Благодаря такой модели данных пользователи могут формулировать сложные запросы, генерировать отчёты, получать подмножества данных [3].

Выводы. Служба MS Analysis Services предоставляет все необходимые инструменты для создания и поддержки многомерных хранилищ данных. Можно абсолютно точно утверждать, что если количество записей в хранилище данных исчисляется тысячами записей, то не стоит использовать многомерность, так как запросы будут проводиться с минимумом затрат времени, но если записей миллиарды, то многомерность предоставит возможность минимизации времени обработки запросов, извлекающих требуемую информацию из фактических данных. Технология MDX является универсальным языком для работы с многомерными хранилищами данных. MDX позволяет сравнительно просто выполнять весьма сложные вычисления связанными с крестотабличными запросами, реализовать которые в SQL было бы весьма трудоемко ввиду отсутствия в нем специальных функций, необходимых для работы с “кубами”.

ЛИТЕРАТУРА

1. Службы SQL Server Integration Service [Электронный ресурс]. Режим доступа: [https://msdn.microsoft.com/ru-ru/library/ms141026\(v=sql.120\).aspx](https://msdn.microsoft.com/ru-ru/library/ms141026(v=sql.120).aspx)
2. Сивакумар Харинатх, Стивен Куинн SQL Server Analysis Services 2005 и MDX: Пер. с англ. – М.: ООО “И.Д. Вильямс”, 2008. – 848 с.
3. Введение в многомерный анализ [Электронный ресурс]. Режим доступа: <https://habrahabr.ru/post/126810/>
4. Преимущества языка SQL [Электронный ресурс]. Режим доступа: http://life-prog.ru/1_1971_preimushchestva-yazika-SQL.html
5. Why MDX? [Электронный ресурс]. Режим доступа: <http://www.simba.com/blog/why-mdx/>

УДК 004.4'22

КЛЮЧЕВЫЕ ВОЗМОЖНОСТИ SPARX ENTERPRISE ARCHITECT и VISUAL PARADIGM

В.В. Герасимов¹, А.Ю. Золотаревский²

¹ассистент кафедры электронных вычислительных машин, Днепропетровский национальный университет им. Олеся Гончара, г. Днепр, Украина, e-mail: gerasimov@dsu.dp.ua

²студент группы КИ-16с-2; кафедра электронных вычислительных машин, Днепропетровский национальный университет им. Олеся Гончара, г. Днепр, Украина, e-mail: arlidor@yahoo.com

Аннотация. В работе рассмотрены ключевые возможности двух инструментов для построения диаграмм UML, призванные облегчить разработку программного обеспечения на всех этапах его создания.

Ключевые слова: UML, моделирование, диаграммы, архитектура.

KEY FEATURES SPARX ENTERPRISE ARCHITECT & VISUAL PARADIGM

V.V. Gerasimov¹, A.Y. Zolotarevskiy²

¹assistant, Department of Computer, Oles Honchar Dnepr National University, Dnepr, Ukraine, e-mail: gerasimov@dsu.dp.ua

²student, Department of Computer, Oles Honchar Dnepr National University, Dnepr, Ukraine, e-mail: arlidor@yahoo.com

Abstract. The paper discusses the key features of two tools for constructing UML diagrams, designed to facilitate the development of software at all stages of its development.

Keywords: UML, modeling, diagrams, architecture.