

SECTION 4. Computer science, computer engineering and automation.

Nazgul Bolatbekovna Nauryzbayeva

Methodist

Taraz Technical Institute, Kazakhstan

nazgul.nauryzbaeva.1980@mail.ru

Tamara Nikolayevna Rubleva

graduate student in the direction: computer science computer engineering

Senior teacher of Department "Applied Informatics",

Institute of New Information Technologies,

Kyrgyz State University of Construction, Transport and Architecture, Kyrgyzstan

BUILDING A MULTIDIMENSIONAL DATABASE INIT KSUCTA

Abstract: *In this article the technology of multidimensional databases is considered. In multidimensional databases data are considered as cubes which are generalization of spreadsheets on any number of measurements. An example of application of this approach on the example of educational institution is reviewed.*

Key words: *the database, the cube, tables, dimensions, educational institution.*

УДК 681.3.019:378.1

ПОСТРОЕНИЕ МНОГОМЕРНЫХ БАЗ ДАННЫХ ИНИТ КГУСТА

Аннотация: *В данной статье рассмотрена технология многомерных баз данных. В многомерных базах данных данные рассматриваются как кубы, которые являются обобщением электронных таблиц на любое число измерений. Рассмотрен пример применения данного подхода на примере учебного заведения.*

Ключевые слова: *база данных, куб, таблицы, измерения, учебное заведение.*

Многомерные базы данных – технология, которая предлагает не только высокую производительность и простоту использования, но и обеспечивает возможности, необходимые для разработки, расширения и быстрого развертывания бизнес-приложений при сокращении ИТ-затрат. Технология многомерных баз данных – ключевой фактор интерактивного анализа больших массивов данных с целью поддержки принятия решения. Подобные базы данных трактуют данные как многомерные кубы, что очень удобно именно для их анализа [1].

Многомерные модели рассматривают данные либо как факты с соответствующими численными параметрами, либо как текстовые измерения, которые характеризуют эти факты. В университете, к примеру, выпуск методической разработки – это факт, количество разработок – параметр, а тип выпускаемой методической разработки, год и кафедры – измерения. Запросы агрегируют значения параметров по всему диапазону измерения, и в итоге получают такие величины, как общее количество выпускаемых методических разработок за выбранный промежуток времени.

Многомерные базы данных рассматривают данные как кубы, которые являются обобщением электронных таблиц на любое число измерений. Кубами легко управлять, добавляя новые значения измерений.

На рис. 1 показан куб, содержащий данные по выпуску методических разработок с дополнительным измерением – «Год». В соответствующих ячейках хранятся данные количестве выпускаемых разработок. В примере можно обнаружить «факт» – непустую

ячейку, содержащую соответствующие числовые параметры – для каждой комбинации время, разработка и кафедра, где был выпущена, по крайней мере, одна разработка. В ячейке размещаются числовые значения, связанные с фактом – в данном случае, это выпуск методических разработок – единственный параметр [2].

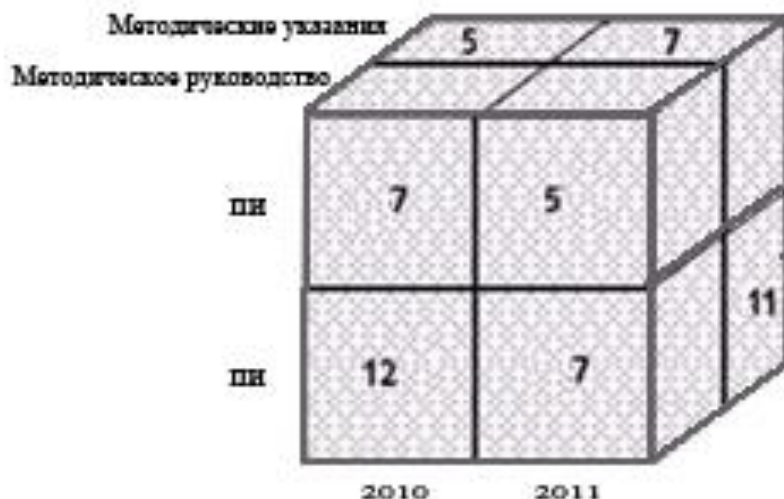


Рисунок 1 - Пример куба, содержащего данные о выпуске методических разработок.

В общем случае куб позволяет представить только два или три измерения одновременно, но можно уменьшить размерность куба, агрегировав некоторые размерности. К примеру, рассматривая выпуск методических разработок по кафедрам и по годам, мы агрегируем информацию для каждого сочетания кафедры и год. Так, на рис. 1, сложив поля 7 и 11, получаем общее количество разработок, выпущенных на кафедре ПИ в 2011 году.

Измерения

Измерения – ключевая концепция многомерных баз данных. Измерения используются для выбора и агрегирования данных на требуемом уровне детализации. Измерения организуются в иерархию, состоящую из нескольких уровней, каждый из которых представляет уровень детализации, требуемый для соответствующего анализа [2,3]. Иногда бывает полезно определять несколько иерархий для измерения. Например, модель может определять время как в финансовых годах, так и в календарных.

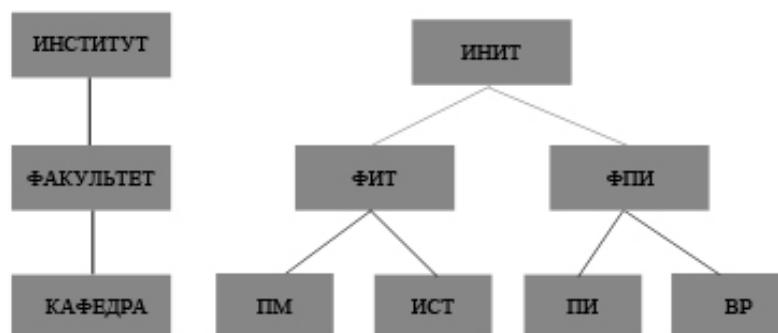


Рисунок 2 - Пример схемы измерений местоположения. Каждое значение размерности является частью значения институт.

На рис. 2 показана схема «Местоположение» для данных по количеству выпускаемых методических разработок. Из трех уровней измерений местоположения самый низкий уровень – «Кафедра». Значения уровня «Кафедра» группируются в значения на уровне «Факультет», к примеру, ПИ и ИСиТ находятся в ИНИТ. Уровень Институт представляет все измерения.

Факты

Факты представляют субъект – некий шаблон или событие, которые необходимо проанализировать. В большинстве многомерных моделей данных факты однозначно определяются комбинацией значений измерений; факт существует только тогда, когда ячейка для конкретной комбинации значений не пуста. Каждый факт обладает некоторой гранулярностью, определенной уровнями, из которых создается их комбинация значений измерений. Например, гранулярность факта в кубе, представленном на рис. 1 – это (Год x Методическая разработка x Кафедра). (Год x Тип x Кафедра) и (Учебное полугодие x Методическая разработка x Кафедра).

Хранилища данных, как правило, содержат следующие три типа фактов [3].

- **События (event)** моделируют события реального мира, при этом каждый факт представляет определенный экземпляр изучаемого явления. Примерами могут служить выпуск методических разработок.

- **Мгновенные снимки (snapshot)** моделируют состояние объекта в данный момент времени, такие как уровни выпуска методических разработок.

- **Совокупные мгновенные снимки (cumulativesnapshot)** содержат информацию о выпуске методических разработок за определенный отрезок времени. Например, совокупное количество по выпуску методических указаний за 2010 год, можно легко сравнить с показателями за 2011 год.

Параметры

Параметры состоят из двух компонентов:

- численная характеристика факта, например, количество выпускаемых разработок;
- формула, обычно простая агрегативная функция, скажем, сумма, которая может объединять несколько значений параметров в одно.

- **Аддитивные параметры** могут содержательным образом комбинироваться в любом измерении. Например, имеет смысл суммировать общее количество для выпускаемых методических разработок, местоположения и времени.

- **Полуаддитивные параметры**, которые не могут комбинироваться в одном или нескольких измерениях. Например, можно определить количество публикаций за пять лет.

Запросы

Многомерная база данных естественным образом предназначена для определенных типов запросов [4 - 11].

- **Запросы вида slice-and-dice** осуществляют выбор, сокращающий куб. К примеру, можно рассмотреть сечение куба на рис. 1, приняв во внимание только те ячейки, которые касаются выпуска методических указаний, а затем еще больше сократить его, оставив ячейки, относящиеся только к 2010 году.

- **Запросы вида drill-down и roll-up** – взаимнообратные операции, которые используют иерархию измерений и параметры для агрегирования. Например, свертка от уровня «Кафедра» до уровня «Факультет» на рис. 2 агрегирует значения для ПИ и ИСиТ в одно значение – ИНИТ.

- **Запросы вида ranking** [4]. возвращает только те ячейки, которые появляются в верхней или нижней части упорядоченного определенным образом списка, например, три самых рекомендуемых методических разработки по кафедре ПИ в 2011 году.

Реализация

Многомерные базы данных реализуют в двух основных формах.

- Системы многомерной оперативной аналитической обработки (MOLAP) хранят данные в специализированных многомерных структурах.
- Реляционные системы OLAP (ROLAP). Для хранения данных используют реляционные базы данных.

В исследовании выбрана система ROLAP, так как она наиболее приемлема для наращивания ядра с использованием ее лучшей масштабируемости по сравнению с MOLAP.

Для хранения информации выбрана схема «звезда», при которых данные хранятся в таблицах фактов и таблицах измерений. Как показано на рис. 3, на каждое измерение отводится одна таблица. Таблица измерений содержит ключевой столбец, по одному столбцу для каждого уровня измерений с текстовыми описаниями значений этого уровня, и по одному столбцу для каждого свойства уровня в измерении.

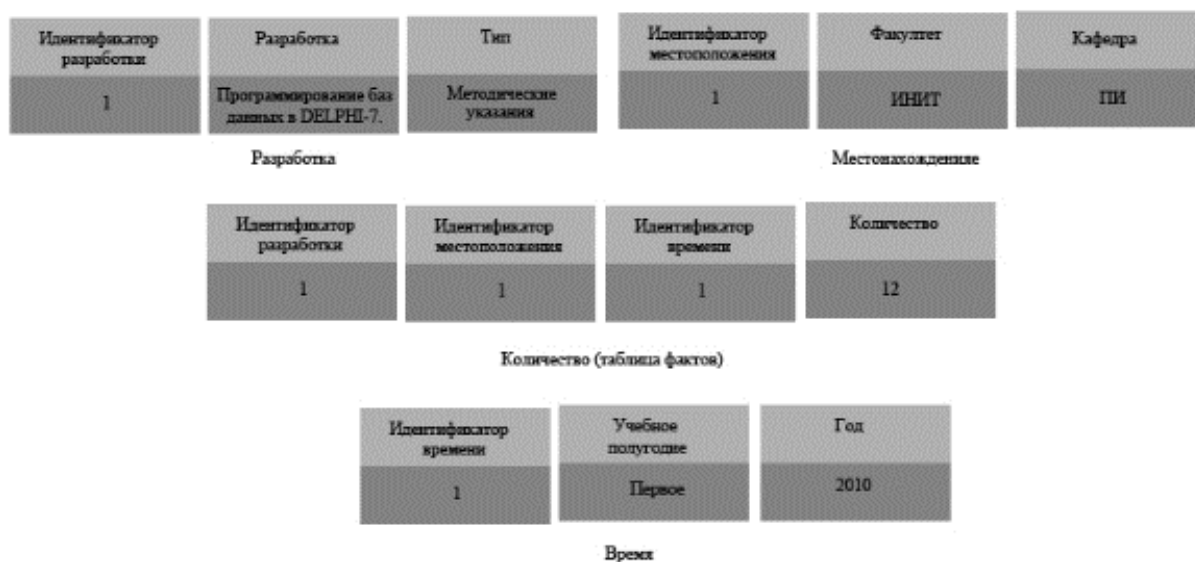


Рисунок 3 - Схема «звезда» для куба выпуска методических разработок.

Информация со всех уровней в измерении хранится в одной таблице измерений, например, названия выпускаемых разработок и их тип хранятся в таблице «Разработки».

Таблица фактов в схеме «звезда» в нашем примере содержит год издания для одной конкретной разработки и соответствующие значения измерений. Она включает столбец внешнего ключа для каждого из трех измерений: разработка, кафедра и год. Таблицы измерений имеют соответствующие ключевые столбцы и по одному столбцу для каждого уровня измерений, например, «Идентификатор местоположения», «Кафедра» и «Факультет».

Заключение

Использование технологии OLAP помогает объединить разнородные данные из комплекса исходных информационных систем, в едином хранилище, обеспечить их поддержание в актуальном состоянии и предоставить инструментарий для поддержки принятия оперативных управленческих решений и сравнения показателей выпуска методических разработок по институтам, факультетам и кафедрам за различные периоды времени.

References

1. Кристиан Йенсен, Торбен Бэч. Технология многомерных баз данных. Открытые системы №1,2002
2. Миронов А.А., Мордвинов В.А., Скуратов А.К. [Семантико-энтропийное управление OLAP и модели интеграции xOLAP в SemanticNET \(ONTONET\)](#). Информатизация образования и науки №2, 2009. С. 21-30.
3. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Методы и модели анализа данных: OLAP и DataMining. – СПб.: БХВ-Петербург, 2004. – 336 с.
4. Кудрявцев Ю.А. OLAP технологии: обзор решаемых задач и исследований // Бизнес-информатика. – 2008.
5. Ларсон Б. Разработка бизнес-аналитики в Microsoft SQL Server 2005. – СПб.: Питер, 2008. – 684 с.
6. Вильямс - Программирование баз данных Microsoft SQL Server 2005. Базовый курс.2007. – 831 с.
7. Крамаренко Т. А. Рекомендации по выбору механизмов доступа приложений баз данных на С++ для СУБД MySQL в средах Code GEAR RAD Studio 2009 и Visual Studio 2005 / Т. А. Крамаренко, Ю. Л. Тихонов // Вісн. Східноукр. нац. ун-ту імені Володимира Даля. – Луганськ, 2010. – № 10 (152). – Ч. 2. – С. 107 – 114.
8. Сеппа Д. Microsoft ADO.NET / Д. Сеппа : Пер. с англ. – М. : „Русская Редакция”, 2003. – 640 с.
9. Постолиит А. В. Visual Studio .NET: разработка приложений баз данных / А. В. Постолиит. – СПб. : БХВ-Петербург, 2003. – 544 с.
10. Баженова И. Ю. Разработка распределенных приложений баз данных / И Ю. Баженова. – М. : МГУ им. М. В. Ломоносова, 2006. – 201 с.
11. Шумаков П. В. ADO.NET и создание приложений баз данных в среде Microsoft Visual Studio .NET. Руководство разработчика с примерами на С# / П. В. Шумаков. – М : ДИАЛОГ-МИФИ, 2003. – 528 с.