

**МНОГОМЕРНЫЕ БАЗЫ ДАННЫХ. ОПЕРАТИВНАЯ
АНАЛИТИЧЕСКАЯ ОБРАБОТКА ДАННЫХ****MULTIDIMENSIONAL DATABASES.
ONLINE ANALYTICAL PROCESSING**

Макалада оперативдүү аналитикалык маалыматтарды иштен чыгуу контекстинде көп өлчөмдүү талдоонун колдонулушу каралган. Маалыматтардын топтомун талап кылынган деңгээлде тактоо, талдоо жана салыштыруу маселелери каралган суроолордун актуалдуулугун шарттады.

Ачкыч сөздөр: OLAP, көп ченемдүү базалар, көп ченемдүү кубдар, фактылар таблицасы, өлчөө.

В данной статье рассмотрена оперативная аналитическая обработка данных в контексте анализа с применением многомерных баз данных. Возможность агрегирования данных на требуемом уровне детализации, анализа и сравнения количественных показателей и обусловили актуальность этого вопроса.

Ключевые слова: OLAP, многомерные базы, многомерные кубы, таблица фактов, измерение.

This article is devoted to online analytical processing in context of using analysis of the multidimensional databases. The ability data of aggregate at the required level of detail of analysis and comparison of quantitative indicators and led to actuality of the matter.

Keywords: OLAP, multidimensional databases, multidimensional cubes, fact table, dimension.

Первую многомерную систему – OLAP Server (оперативная обработка данных), предназначенную для массового рынка, корпорация Microsoft выпустила в 1998 году и теперь такие системы становятся распространенными продуктами.

Применение многомерных баз данных это 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 [1].

Системы на основе многомерных баз данных идеально подходят для потребностей учебных заведений в силу их разветвленной организационной структуры и необходимости оптимальной организации их в базы данных.

В основе концепции OLAP лежит принцип многомерного представления данных.

По Кодду, многомерное концептуальное представление (multidimensionalconceptualview) представляет собой множественную перспективу, состоящую из нескольких независимых измерений, вдоль которых могут быть проанализированы определенные совокупности данных. Каждое измерение включает направления консолидации данных, состоящие из серии последовательных уровней обобщения, где каждый вышестоящий уровень соответствует большей степени агрегации данных по соответствующему измерению [2].

OLAP предоставляет удобные быстродействующие средства доступа, просмотра и анализа деловой информации. Пользователь получает естественную, интуитивно

понятную модель данных, организовав их в виде многомерных кубов (Cubes). Осями многомерной системы координат служат основные атрибуты анализируемого бизнес-процесса.

В данной статье рассматривается применение технологии многомерных баз данных для отдела магистратуры и аспирантуры КГУСТА. На рис.1. представлен «куб» для указанной базы данных.

	ИНИТ	ИП	ИАД	ИСЭМ	ИЗЭС	ИНТРАСКОМ
Магистрант	63	10				
Аспирант	3	1				
Докторант	4	2				
Соискатель	16	2				

Рис.1. Пример куба

Например, для анализа контингента отдела магистратуры и аспирантуры это могут быть категория обучающихся, кафедры. В качестве одного из измерений используется время. На пересечениях осей – измерений (Dimensions) – находятся данные, количественно характеризующие процесс – меры (Measures). Это могут быть количество обучающихся магистрантов, аспирантов, докторантов и соискателей. Пользователь, анализирующий информацию, может «разрезать» куб по разным направлениям, получать сводные (например, по учебным годам) или, наоборот, детальные (по учебным полугодиям) сведения и осуществлять прочие манипуляции.

В качестве мер в трехмерном кубе, изображенном на рис. 1 использованы количество обучающихся магистрантов, соискателей, аспирантов и докторантов, а в качестве измерений – время, институты, кафедры. Измерения представлены на определенных уровнях группировки: обучающиеся группируются по категориям, кафедры – по факультетам, факультеты по институтам, а данные о времени совершения операций – по учебным годам.

В общем случае куб позволяет представить только два или три измерения одновременно, но можно уменьшить размерность куба, агрегировав некоторые размерности.

Для визуализации данных, хранящихся в кубе, применяются, как правило, привычные двумерные, т. е. табличные, представления.

Двумерное представление куба можно получить, «разрезав» его поперек одной или нескольких осей (измерений): фиксируются значения всех измерений, кроме двух, и получается обычная двумерная таблица. В горизонтальной оси таблицы (заголовки столбцов) представлено одно измерение, вертикальной (заголовки строк) – другое, а в ячейках таблицы – значения мер. При этом набор мер фактически рассматривается как одно из измерений – например, выбираем для показа одну меру (и тогда можем разместить в заголовках строк и столбцов два измерения), либо показываем несколько мер (и тогда одну из осей таблицы займут названия мер, а другую – значения единственного «неразрезанного» измерения).

На рис. 2. изображен двумерный срез куба для одной меры – количество обучающихся и двух «неразрезанных» измерений – институт и время.

Значения, «откладываемые» вдоль измерений, называются членами или метками (members). Метки используются как для «разрезания» куба, так и для ограничения (фильтрации) выбираемых данных - когда в измерении, остающемся «неразрезанным»,

нас интересуют не все значения, а их подмножество, например, три кафедры. Значения меток отображаются в двумерном представлении куба как заголовки строк и столбцов.

	инит	иип	иад	исэм	изэс	ИНТРА СКОМ
Магистрант	63	10				
Аспирант	3	1				
Докторант	4	2				
Соискатель	16	2				

Рис. 2. Двумерный срез куба для одной меры

Метки могут объединяться в иерархии, состоящие из одного или нескольких уровней (levels). Например, метки измерения «Направление» естественно объединяются в иерархию с уровнями:

Институт
 Факультет
 Кафедра
 Направление

В соответствии с уровнями иерархии вычисляются агрегатные значения, например, количество обучающихся в Институте или на Факультете. В одном измерении можно реализовать более одной иерархии - скажем, для времени: {Учебный год, Учебное полугодие}.

Детализацию рассматриваемого куба можно проводить, используя ключевую концепцию многомерной базы данных – измерение.

Так, измерение Студент может определяться направлением консолидации, состоящим из уровней обобщения «Институт-кафедра-студент». Измерение Время может даже включать два направления консолидации – «Учебный год – учебное полугодие». В этом случае становится возможным произвольный выбор желаемого уровня детализации информации по каждому из измерений. Операция спуска (drillingdown) соответствует движению от высших ступеней консолидации к низшим; напротив, операция подъема (rollingup) означает движение от низших уровней к высшим (рис. 3) [3].



Рис. 3. Измерения и направления консолидации 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. Иногда бывает полезно определять несколько иерархий для измерения.

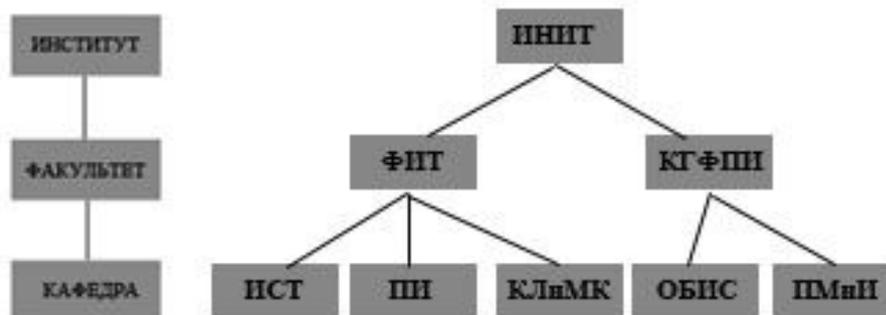


Рис. 4. Пример схемы измерений местоположения

Каждое значение размерности является частью значения институт. На рис.4 показана схема «Местоположение» для данных по количеству обучающихся в отделе магистратуры и аспирантуры КГУСТА. Из трех уровней измерений местоположения самый низкий уровень – «Кафедра». Значения уровня «Кафедра» группируются в значения на уровне «Факультет», к примеру, ПИ и ИСиТ, КлиМК находятся в ФИТ. Уровень Институт представляет все измерения.

Кроме того существует еще одна важная концепция – факты, которые представляют собой 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 (Учебный год x Категория обучающегося x Кафедра) или (Учебное полугодие x Категория обучающегося x Институт).

В настоящее время многомерные базы данных реализуют в двух основных формах.

1. Системы многомерной оперативной аналитической обработки (MOLAP) хранят данные в специализированных многомерных структурах.
2. 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. 5, 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.

Информация со всех уровней в измерении хранится в одной таблице измерений, например, названия выпускаемых разработок фамилии обучающихся и их категории хранятся в таблице «Студенты».

Таблица фактов в схеме «звезда» в нашем примере содержит учебный год поступления для одной категории обучающихся и соответствующие значения измерений. Она включает столбец внешнего ключа для каждого из трех измерений: категория, кафедра и год. Таблицы измерений имеют соответствующие ключевые столбцы и по одному столбцу для каждого уровня измерений, например, «Идентификатор местоположения», «Кафедра» и «Факультет» [4].



Рис.5. Схема «звезда» для куба категории обучающихся

Особенности рассматриваемой технологии разработки многомерных баз данных основаны на следующих преимуществах OLAP:

- согласованность исходной информации и результатов анализа. При наличии OLAP системы всегда есть возможность проследить источник информации и определить логическую связь между полученными результатами и исходными данными. Снижается субъективность результатов анализа.
- проведение многовариантного анализа. Применение OLAP системы позволяет получить множество сценариев развития событий на основе набора исходных данных. За счет инструментов анализа можно смоделировать ситуации по принципу «что будет, если».
- управление детализацией. Детальность представления результатов может изменяться в зависимости от потребности пользователей. При этом нет необходимости осуществлять сложные настройки системы и повторять вычисления. Отчет может содержать именно ту информацию, которая необходима для принятия решений [5].

Заключение

В результате рассмотрения схемы сбора информации отдела магистратуры и аспирантуры, а также учитываемых показателей, содержащихся в отчётных документах, было проведено их разделение по временным признакам. Это позволяет сохранить всю собираемую информацию в многомерную базу данных, имеющую три измерения: время, обучающиеся, параметры.

Причина использования OLAP для обработки запросов — это скорость [5]. OLAP делает мгновенный снимок (snapshot) реляционной БД и структурирует ее в пространственную модель для запросов.

Использование технологии OLAP помогает объединить разнородные данные из комплекса исходных информационных систем, в едином хранилище, обеспечить их поддержание в актуальном состоянии [6]. Кроме того, применение указанной технологии, предоставляет обширный инструментарий для проведения качественного анализа данных и сравнения показателей количества обучающихся в отделе магистратуры КГУСТА по институтам, факультетам и кафедрам за различные периоды времени.

Список литературы

1. Кристиан Йенсен. Технология многомерных баз данных [Текст] / Кристиан Йенсен, Торбен Бэч // Открытые системы. - 2002. - №1. – С.45-50.
2. Codd E. F., Codd S. B., Salley C. T. Providing OLAP (On-Line Analytical Processing) to User-Analysts: An IT Mandate. - E. F. Codd & Associates, 1993.– 20 с.
3. Миронов А.А. Семантико-энтропийное управление OLAP и модели интеграции xOLAP в SemanticNET (ONTONET)[Текст] / А.А. Миронов, В.А. Мордвинов, А.К. Скуратов // Информатизация образования и науки. - 2009. - №2. -с. 21-30.
4. Барсегян А.А. Методы и модели анализа данных: OLAP и DataMining[Текст] / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. – СПб.: БХВ-Петербург, 2004. – 336 с.
5. Кудрявцев Ю.А. OLAP технологии: обзор решаемых задач и исследований[Текст] / Ю.А. Кудрявцев // Бизнес-информатика. – 2008. С. 66-70.
6. Гудков А. А. Применение новых информационных технологий анализа информации в сфере образования [Текст] / А. А. Гудков // Известия Пензенского государственного педагогического университета им. В.Г. Белинского. - 2007. - № 7. - С. 201-203.