

УДК 004.651:004.94-043.86

95. ИСПОЛЬЗОВАНИЕ OLAP-СИСТЕМ В АНАЛИТИЧЕСКОЙ ОБРАБОТКЕ ДАННЫХ

Герасименя В.В., Герман А.А., студентки группы 377901

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Нагулевич Р.С. – старший преподаватель, магистр экономических наук

Аннотация. В работе рассматривается роль OLAP как инструмента трансформации сырых данных в управленческие решения. Раскрывается архитектура OLAP-систем: источники данных, сервер и клиентские приложения. Формализованы понятия многомерной модели: гиперкуб, измерения, меры, операции среза и вращения. Особое внимание уделено трехуровневой структуре многомерности: представление, обработка, хранение. Проанализированы подходы к хранению данных: MOLAP, ROLAP, HOLAP. Данный инструментариум обеспечивает переход от реактивной отчетности к проактивному стратегическому прогнозированию.

Ключевые слова. OLAP-системы, многомерная аналитическая обработка данных, многомерная модель данных, гиперкуб, хранилище данных, витрины данных, измерения, меры, операции среза и вращения, клиент-серверная архитектура, MOLAP, ROLAP, HOLAP, поддержка принятия решений, бизнес-аналитика.

В современном мире хозяйствующие субъекты аккумулируют значительные массивы данных, известные как Big Data, однако сами по себе эти сведения не приносят пользы. Ключевая задача заключается в извлечении из них значимой информации для поддержки принятия управленческих решений. Простых отчетов, формируемых операционными системами, для этого недостаточно, поскольку они отвечают на вопрос «Что произошло?», но не позволяют установить причинно-следственные связи и не дают возможности прогнозировать последствия.

В условиях цифровой трансформации финансово-банковской деятельности (ФБД) особую значимость приобретают инструменты, позволяющие оперативно обрабатывать растущие объемы транзакционных данных. В данном исследовании рассмотрена концепция использования OLAP-технологий в финансово-банковской деятельности как ключевого инструмента аналитической обработки данных, обеспечивающего многомерный анализ и поддержку принятия управленческих решений. OLAP представляет собой ключевой компонент организации хранилищ данных; концепция данной технологии была описана в 1993 году Эдгаром Коддом. OLAP выступает связующим звеном в экосистеме данных: он использует подготовленную информацию из хранилищ и предоставляет удобный интерактивный инструмент для её исследования человеком-аналитиком. Результаты OLAP-анализа, в свою очередь, могут стать основой для более глубокого изучения закономерностей с помощью технологий интеллектуального анализа данных (Data Mining).

Цель OLAP заключается в предоставлении пользователю возможности за приемлемое время, например, за секунды проводить сложный ad-hoc анализ данных (нерегламентированные запросы, формулируемые пользователем в процессе исследования для проверки возникающих гипотез) с различных точек зрения, называемых измерениями. Ключевые принципы OLAP, сформулированные Коддом, включают многомерное концептуальное представление данных, предполагающее организацию данных в виде гиперкубов; интуитивное манипулирование данными, реализуемое через операции среза, вращения, детализации и свертки; доступность, обеспечивающую быстрый отклик системы; интерактивный анализ, допускающий формулирование нерегламентированных запросов; а также многопользовательскую поддержку. Архитектура OLAP-системы строится по принципу клиент-сервер и включает несколько ключевых компонентов. Источником в OLAP-системах выступает сервер, поставляющий данные для анализа; им может служить хранилище данных (Data Warehouse), набор таблиц или иные системы.

Хранилище данных представляет собой реляционную базу данных, спроектированную для хранения консолидированной исторической информации. Основная таблица хранилища содержит числовые показатели (меры). В его состав входят виртуальное хранилище данных, основанное на репозитории метаданных, описывающих источники и процедуры доступа к данным; витрины данных (Data Mart) – облегченные, тематически ориентированные подмножества хранилища, предназначенные для конкретных отделов или задач, использование которых повышает безопасность, упрощает проектирование и снижает нагрузку на основное хранилище; а также глобальное хранилище данных (современный подход, при котором единое хранилище служит источником для всех витрин данных,

обеспечивая согласованность информации). Хранилищам данных присущи следующие свойства: предметная ориентация, интегрированность, инвариантность во времени, неразрушаемость (стабильность) и минимизация избыточности.

Хранилище данных служит поставщиком информации для многомерной базы данных, основными классами объектов которой являются измерения и показатели. К измерениям относятся множества значений, по которым происходит индексация данных, например, время, профессии и другие; при этом измерения представляют собой оси анализа: время, профессия, ФИО. Прикладной частью OLAP-системы выступает OLAP-сервер, который выполняет всю работу (в зависимости от модели системы) и хранит информацию, к которой обеспечивается активный доступ; данный компонент осуществляет вычисления, хранит многомерные данные и обрабатывает запросы. Клиентское приложение предоставляет пользователю интерфейс для удаленного доступа к данным, формулирования сложных запросов, генерации отчетов и визуализации результатов.

Основными понятиями многомерной модели данных являются гиперкуб, измерение, метки, ячейка и мера. Гиперкуб данных представляет собой «контейнер» для n -мерной структуры данных ($n > 2$), внутри которого упорядоченно хранятся агрегированные бизнес-данные (см. рисунок 1). Гиперкуб содержит одно или более измерений и представляет собой упорядоченный набор ячеек. Каждая ячейка определяется одним и только одним набором значений измерений – меток. Ячейка может содержать данные – меру – либо оставаться пустой. Измерением называется множество меток, образующих одну из граней гиперкуба; измерение выступает осью анализа (например, время, регион, товар, клиент) и может иметь иерархию (год разделяется на кварталы, кварталы разделяются на месяцы, а месяцы делятся на дни). Метка представляет собой конкретное уникальное значение внутри одного измерения (например, «Минск», «Март 2026», «Инженер»), являясь точкой на оси. Если измерение – это ось координат (например, ось X «Города»), то метка – это конкретная точка на этой оси (например, «Минск», «Брест», «Гомель»). Ячейка определяется как минимальная и неделимая точка в многомерном пространстве, образуемая пересечением одной метки из каждого измерения; в ней хранится анализируемое значение, и она может быть пустой. Мера представляет собой числовой показатель, который помещается в ячейку и является предметом анализа (например, объем продаж, прибыль).

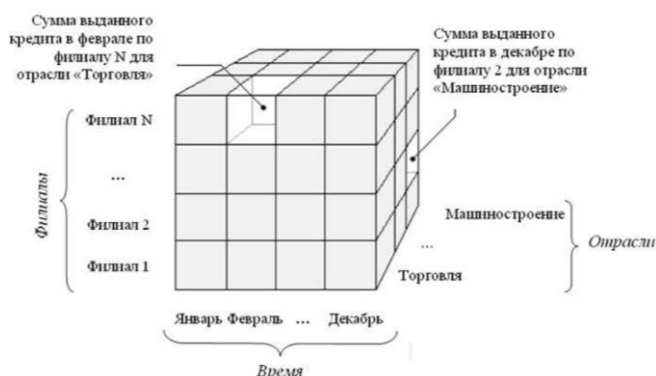


Рисунок 1 – Пример гиперкуба данных

Операция среза представляет собой выделение подмножества данных путем фиксации одной или нескольких меток. Например, запрос «Показать сумму выданного кредита по филиалу N за февраль для отрасли «Торговля» фиксирует метки «Филиал N», «Февраль» и «Торговля», «отрезая» от общего куба нужный «кусок» данных. Если представить данные о кредитовании в виде куба с тремя измерениями (филиалы, время, отрасли), где внутри на пересечении содержатся числа, например, сумма выданного кредита, то при фиксации двух или трех измерений множество фиксированных меток будет {Филиал N, Февраль, Торговля}, а множество фиксированных измерений – {Филиал, Время, Отрасль}. Пересечение гиперплоскостей, образуемое при фиксации меток, представляет собой срез – наиболее тонкий «кусок» данных. Операция вращения заключается в изменении порядка представления (визуализации) измерений, что обеспечивает возможность визуализации данных в форме, наиболее комфортной для их восприятия. В терминах рассматриваемой модели данных вращение означает смену последовательности фиксации меток при построении среза; результатом вращения для двумерного среза (таблицы) выступает замена столбцов на строки, а строк на столбцы.

Многомерный анализ данных может осуществляться с помощью различных программных решений, которые условно делятся на клиентские и серверные OLAP-средства. Клиентские OLAP-инструменты представляют собой приложения, устанавливаемые непосредственно на компьютер пользователя и

выполняющие вычисления локально; примером таких средств служат сводные таблицы в Microsoft Excel. Принцип работы клиентских OLAP-средств заключается в следующем: приложение загружает исходные данные (из локальной базы данных или файлов), самостоятельно вычисляет агрегатные данные (если источником выступает настольная СУБД), отправляет SQL-запросы на сервер (в случае хранения данных в серверной СУБД), сохраняет результаты вычислений в кэше оперативной памяти и обеспечивает интерфейс для многомерного анализа. Серверные OLAP-системы представляют собой специализированные серверные приложения, которые берут на себя всю вычислительную нагрузку. Архитектурные особенности таких систем включают выполнение сервером всех вычислений агрегатных данных, хранение многомерных кубов в специализированном формате, обслуживание запросов от множества клиентских приложений и обеспечение актуальности данных через механизмы обновления.

Многомерность в OLAP приложениях может быть разделена на три уровня, каждый из которых выполняет специфическую функцию в обеспечении полноценного аналитического процесса.

Первый уровень представляет собой многомерное представление данных, то есть средства конечного пользователя, обеспечивающие многомерную визуализацию и манипулирование данными. Данный уровень выступает в роли интерфейса взаимодействия аналитика с системой и характеризуется максимальной абстракцией от физической структуры хранения информации. Пользователь воспринимает данные не как набор таблиц реляционной базы, а как естественно многомерную структуру, гиперкуб, что позволяет ему сосредоточиться на содержательном анализе, а не на технических аспектах организации данных. Визуализация осуществляется в форме, наиболее удобной для восприятия, такой как многомерные кубы, двумерные таблицы, диаграммы различных типов. Манипулирование данными реализуется через интуитивно понятные операции среза, то есть фиксации отдельных значений измерений, вращения, означающего изменение порядка отображения измерений, и детализации, представляющей собой переход от обобщенных показателей к более детальным. Первый уровень обеспечивает доступность аналитических возможностей для пользователя, не требующего глубоких технических знаний.

Второй уровень представляет собой многомерную обработку, включающую средство формулирования многомерных запросов, то есть специализированный язык, и процессор, способный обработать и выполнить такой запрос. Традиционный реляционный язык SQL оказывается здесь непригодным, поскольку он ориентирован на работу с плоскими таблицами и не обладает нативной поддержкой многомерных операций, таких как срез, вращение, детализация и свертка. Для формулирования многомерных запросов используются специализированные языки, например MDX, которые позволяют описывать выборки данных в терминах измерений, иерархий и показателей. Процессор запросов, понимающий многомерную семантику, принимает такие запросы, оптимизирует их выполнение и возвращает результат в структурированном виде. В некоторых архитектурах данный уровень также может включать механизм трансляции многомерных запросов в SQL запросы к реляционным источникам данных, что позволяет интегрировать OLAP средства с существующими реляционными хранилищами без их полной замены.

Третий уровень представляет собой многомерное хранение, то есть физическую организацию данных для ускорения запросов. Данный уровень является необязательным, но важным компонентом, поскольку он определяет, где и в каком формате физически размещаются исходные и агрегатные данные. В зависимости от выбранного подхода, такого как MOLAP, ROLAP или HOLAP, данные могут храниться в многомерных структурах, оптимизированных для быстрого выполнения операций по любому измерению, либо оставаться в реляционных таблицах с использованием специализированных схем, например звезда или снежинка, либо комбинировать оба способа, размещая агрегаты в многомерном хранилище, а детальные данные в реляционном. Физическая организация данных на этом уровне напрямую влияет на производительность запросов, скорость загрузки и обновления информации, а также на масштабируемость системы. Оптимальный выбор архитектуры хранения позволяет достичь баланса между скоростью отклика системы, объемом занимаемого дискового пространства и гибкостью управления данными.

Существует три основных подхода к физическому хранению многомерных данных. MOLAP (Multidimensional OLAP) предполагает хранение исходных и агрегатных данных 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. Эти системы обеспечивают полный цикл OLAP-обработки; они либо включают в себя, помимо серверного компонента, собственный интегрированный клиентский интерфейс, либо используют для связи с пользователем внешние программы работы с электронными таблицами. Преимуществами MOLAP являются максимальная

производительность запросов и оптимизированная структура для OLAP-операций; недостатками для данной системы являются значительное увеличение объема хранения и длительное время загрузки и обновления.

ROLAP (Relational OLAP) предполагает, что исходные данные остаются в той же реляционной базе данных, где они изначально и находились, а агрегатные данные помещаются в специально созданные для их хранения служебные таблицы в той же базе данных. Преимуществами ROLAP выступают возможность работы с большими объемами детальных данных и использование мощностей реляционных СУБД; недостатки у этой системы являются зависимость производительности от качества SQL-оптимизации и меньшая эффективность для сложных многомерных запросов.

HOLAP (Hybrid OLAP) представляет собой гибридный подход, при котором исходные данные остаются в той же реляционной базе данных, где они изначально находились, а агрегатные данные хранятся в многомерной базе данных. Преимуществами данного подхода являются баланс производительности и гибкости, а также оптимальное использование ресурсов. Компонентами HOLAP выступают специализированный (не SQL) язык запросов и процессор запросов, понимающий многомерную семантику. Особенности данного подхода являются непригодность SQL для нативных многомерных запросов, необходимость трансляции многомерных запросов в SQL (в гибридных системах) и обеспечение семантики операций OLAP (свертка, детализация и другие).

Результаты исследования демонстрируют, что OLAP-системы представляют собой необходимый инструмент трансформации сырых данных из хранилищ в многомерную модель гиперкуба, обеспечивающую удобство аналитического познания. Благодаря реализации интуитивных операций (срез, вращение) и применению различных архитектурных подходов (клиентские и серверные средства, MOLAP, ROLAP, HOLAP) данные системы предоставляют бизнес-аналитикам и руководителям возможность самостоятельного, оперативного и гибкого исследования информации, выявления тенденций, установления причин проблем и моделирования сценариев. Разработанный инструментарий позволяет осуществить переход от реактивного управления, основанного на отчетах о прошлом, к проактивному и обоснованному принятию решений, что создает основу для формирования ключевого конкурентного преимущества в современной цифровой экономике.

Список использованных источников:

1. Бекаревич, Ю. Б. Особенности оперативного анализа данных при использовании табличной модели / Ю. Б. Бекаревич, Н. В. Пушкина // Проблемы экономики и менеджмента. – 2015. – № 2 (42). – С. 120–127.
2. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Методы и модели анализа данных OLAP и DataMining, 2004.
3. Ильяшенко, О. Ю. Роль BI-систем в совершенствовании процессов обработки и анализа бизнес информации / И. В. Ильин // Наука и бизнес: пути развития – 2017.

UDC 004.651:004.94-043.86

USING OLAP SYSTEMS IN ANALYTICAL DATA PROCESSING

Gerasimenya V.V., German A.A

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Nagulevich R.S. – Senior Lecturer, Master of Economics

Annotation. This paper examines the role of OLAP as a tool for transforming raw data into management decisions. It describes the architecture of OLAP systems, including data sources, servers, and client applications. The concepts of a multidimensional model are formalized: hypercube, dimensions, measures, slice and rotate operations. Particular attention is paid to the three-level structure of multidimensionality: presentation, processing, and storage. Data storage approaches are analyzed: MOLAP, ROLAP, and HOLAP. This toolkit facilitates the transition from reactive reporting to proactive strategic forecasting.

Keywords. OLAP systems, multidimensional analytical data processing, multidimensional data model, hypercube, data warehouse, data marts, dimensions, measures, slice and rotate operations, client-server architecture, MOLAP, ROLAP, HOLAP, decision support, business analytics.