

Использование трехмерных кубов данных в реализации системы бизнес-анализа

Б.А. Черныш
А.В. Мурыгин

Ссылка для цитирования

Черныш Б.А., Мурыгин А.В. Использование трехмерных кубов данных в реализации системы бизнес-анализа // Программные продукты и системы. 2023. Т. 36. № 2. С. 237–244. doi: 10.15827/0236-235X.142.237-244

Информация о статье

Поступила в редакцию: 29.11.2022

После доработки: 15.02.2023

Принята к публикации: 20.02.2023

Аннотация. Бизнес-анализ является одним из ключевых инструментов управления, позволяющих получать достоверную картину текущего состояния дел на предприятии по всем направлениям деятельности. Для обеспечения этого процесса в любой компании в качестве показателей ее работы используются различные данные. Источником данных прежде всего являются интегрированные информационные системы. Эти системы могут использовать либо имеющиеся в их составе инструменты бизнес-анализа, либо специализированные решения, позволяющие выполнять сложные аналитические задачи по заданной постановке. В статье рассматриваются особенности обоих подходов, их преимущества и недостатки, приводятся примеры существующих на рынке зарубежных и отечественных продуктов для бизнес-анализа. Предлагается способ построения трехмерных кубов с использованием содержащихся в системе данных на примере модуля бизнес-анализа разработанной авторами интегрированной информационной системы SciCMS. Описываются используемые методики, исходные требования и ограничения. Проведена формализация задач, рассмотрен математический аппарат построения многомерных моделей данных на основе информации из фиксированного набора нормализованных таблиц реляционной БД. Представлены примеры SQL-запросов и выходных данных. В ряде случаев (работа с нереляционной СУБД, необходимость в наличии заранее рассчитанных агрегатных значений, сложность и высокая стоимость прямых SQL-запросов и т.д.) применение описанного способа построения многомерных кубов невозможно. Решением данной проблемы в SciCMS является собственный модуль импорта и трансформации данных на основе библиотеки с открытым исходным кодом. В статье обобщены основные достоинства и недостатки предлагаемого подхода, перспективы его использования на отечественных предприятиях.

Ключевые слова: многомерное представление, БД, BI, OLTP, OLAP, бизнес-анализ, аналитическая обработка, трехмерный куб, реляционная схема, нормализация, нормальная форма, интегрированная информационная система, GraphQL, SciCMS

При эксплуатации любой *интегрированной информационной системы* (ИИС) рано или поздно возникает необходимость в построении различных аналитических отчетов на основании имеющихся данных, их визуализации и составлении прогнозов. Это обусловлено прежде всего постановкой аналитических задач: анализ ключевых показателей деятельности, маркетинговый и финансово-экономический анализ, анализ сценариев, моделирование, прогнозирование и т.д. Решение этого класса задач возможно с помощью OLAP-инструментария. В более широком смысле данный класс систем именуют системами бизнес-анализа (BI – Business Intelligence), или BI-системами. С помощью BI-инструментов разрозненные данные, хранящиеся в системе, преобразуются в структурированную информацию, описанную в терминах соответствующей предметной области. Эта информация, в свою очередь, позволит принимать обоснованные управленческие решения как руководителям предприятий, так и узким отраслевым специалистам, аналитикам, экономистам и т.д. Эффект от применения инстру-

ментария бизнес-анализа может быть выражен как прямо (за счет снижения трудозатрат на технические операции и расходов на поддержку разрозненных БД), так и косвенно (за счет повышения эффективности бизнес-процессов компании).

Существующие решения и подходы в формировании многомерных моделей

Как правило, для реализации требований бизнес-анализа применяют специализированные программные средства, не являющиеся непосредственно частью ИИС, данные которой используют для аналитики. Такое разделение обусловлено прежде всего тем, что обычные транзакционные БД (OLTP), на которых построено подавляющее большинство ИИС, оптимизированы под быстрое внесение данных, но не под быстрое построение сложной отчетности (в отличие от OLAP-систем). OLAP-база обычно имеет специальную архитектуру и содержит предварительно просчитанные агрегатные данные, что и обеспечивает высокую скорость выполнения запросов.

Платой за это становится необходимость синхронизировать OLTP и OLAP. Поскольку подобный процесс, как правило, периодичен, между появлением данных в оперативной базе и аналитическими данными имеется запаздывание. Данный подход реализован практически во всех современных BI-системах: Tableau, Qlik, Microsoft BI, Oracle BI, SAP BW и др. [1]. Существенным недостатком имеющих коммерческих BI-решений является узкий круг поддерживаемых источников данных. Так, например, продукты Microsoft, кроме собственной СУБД, в качестве источника могут использовать БД Oracle и Teradata. Oracle BI имеет еще более узкое ограничение и работает лишь с собственной линейкой СУБД. Отечественный рынок средств бизнес-аналитики также достаточно обширен [2]. В России широкое распространение получила система Галактика BI, построенная на платформе Microsoft BI и расширяющая функционал базовой платформы компонентами и сервисами, ориентированными на бизнес-задачи управления предприятием [3]. Среди лидеров рынка можно также выделить такие продукты, как Форсайт, Loginom, Visiology, Polymatica, Visary BI, Luxms BI, ProceSet, Modus BI [4].

Другим подходом в решении задач оперативной аналитической обработки данных является встраивание OLAP-инструментария непосредственно в OLTP-хранилище. Это достигается как использованием специальных элементов данных (например, регистров оперативного учета в платформе «1С: Предприятие» [5]), так и оптимальной (с точки зрения выполнения OLAP-запросов) структуризацией схемы данных. В этом случае функционал BI является частью системы и, разделяя с ней общие данные, позволяет формировать на их основе многомерные представления для построения отчетов, графиков, диаграмм, витрин данных и пр. При разработке информационной системы SciCMS [6] именно второй подход лег в основу модуля аналитики, являющегося частью ядра системы.

Подход в построении трехмерных кубов, используемый в SciCMS

Модуль аналитики в SciCMS позволяет использовать данные любых сущностей для построения многомерного OLAP-куба [7]. Каждый из атрибутов сущности может служить измерением в этом кубе. Особую роль играют

атрибуты с типами date/time/datetime/timestamp (дата/время) и location (местоположение). Атрибут первого типа образует временное измерение на кубе и отвечает на вопрос «когда» относительно значений фактов. OLAP-куб ограничен одной шкалой времени независимо от числа задействованных в нем сущностей. При этом временные оси каждой из этих сущностей должны иметь одинаковый тип (из допустимых типов даты/времени). Атрибут типа location содержит данные о местоположении (широту, долготу и текстовую метку измерения) и отвечает на вопрос «где» при анализе значений фактов. Аналогично шкале времени куб может содержать только одно измерение местоположения. При настройке параметров куба ось фактов определяется атрибутами, называемыми метриками и отвечающими на вопрос «что» относительно самих значений этих атрибутов.

На рисунке 1 показано диалоговое окно настройки панели визуализации с использованием пользовательского интерфейса клиентской части системы SciCMS Client. Взаимодействие с сервером осуществляется по протоколу GraphQL. Итоговая настроенная панель показана на рисунке 2. Панель визуализации может использоваться как непосредственно из приложения, так и в качестве встроенного элемента на любой веб-странице.

Идея использования трехмерных кубов с атрибутами «что-где-когда» не нова и успешно применяется в BI-системах (например Luxms BI [4]). При этом измерения времени и местоположения, как и сами метрики, хранятся в собственных таблицах. Таким образом, каждый набор данных включает набор из трех специально подготовленных таблиц. Заполнение таблиц осуществляется путем настраиваемого импорта из других источников.

Методика, примененная в системе SciCMS, предлагает другой подход: использование уже имеющихся сущностей (построенных на реляционных таблицах) с выделением в них атрибутов даты/времени, местоположения и метрик. Поскольку SciCMS уже хранит все метаданные сущностей, для настройки параметров куба достаточно выбрать нужные атрибуты в качестве осей. Это дает возможность динамически (в зависимости от параметров визуализации) изменять требуемые оси куба. При правильной индексации столбцов таблиц, задействованных в построении куба, время запросов сопоставимо с выборкой

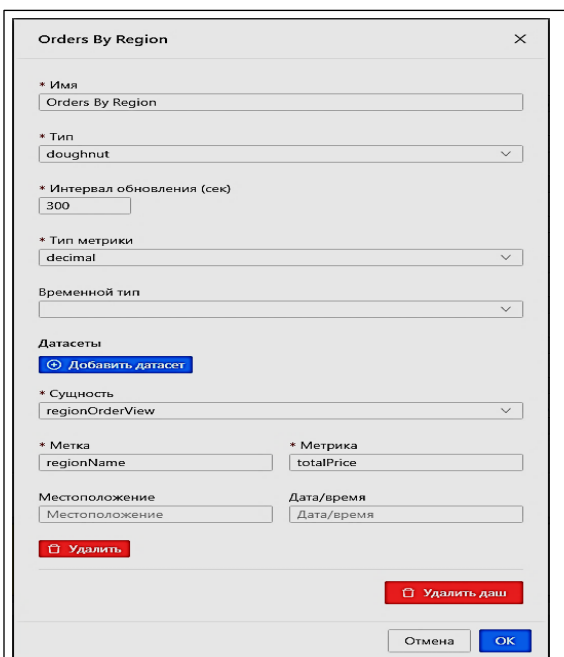


Рис. 1. Диалоговое окно настройки визуализации

Fig. 1. A visualization settings dialog

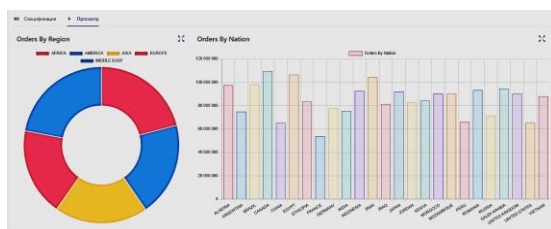


Рис. 2. Настроенная панель визуализации

Fig. 2. A customized visualization panel

данных из специально подготовленных структур в Luxms BI. Кроме непосредственного представления данных реляционных таблиц, в SciCMS имеются сущности «только для чтения», содержащие произвольный SQL-запрос к БД. Такой подход позволяет избежать необходимости создания дополнительных представлений в СУБД (если такая операция является излишней или запрещена администратором БД).

Важной особенностью предлагаемого механизма построения OLAP-кубов является то, что система SciCMS может быть подключена к неограниченному числу источников данных одновременно. Таким образом, любая таблица (представление, SQL-запрос) из любого источника данных напрямую может быть задействована в построении куба (без необходимости экспорта и преобразования).

Отсутствие дополнительных таблиц и исключение необходимости экспорта являются важными преимуществами, так как наряду с экспортом внешние системы бизнес-анализа требуют трансформации исходных данных в собственный формат хранения. В совокупности настройка параметров экспорта и преобразования данных, как и сам процесс экспорта (по требованию или в заданное время по расписанию), могут требовать существенных затрат времени пользователя (возможно, с привлечением сотрудников, сопровождающих BI-систему) и вычислительных ресурсов предприятия.

Если сущности представляют данные реляционных таблиц, которые соответствуют правилам нормальных форм, можно утверждать, что схема обладает свойством соединения без потерь информации в соответствии с теоремой из [8]: пусть σ – декомпозиция отношения R , образованная схемами отношений в третьей нормальной форме, и пусть также X – ключ R . Тогда $\tau = \sigma \cup \{X\}$ – декомпозиция R , такая, что все составляющие ее схемы отношений находятся в третьей нормальной форме. Эта декомпозиция сохраняет зависимости и сохраняет свойство соединения без потерь.

На примерах можно показать, как данное свойство позволяет сформировать многомерное (гиперкубическое) представление любой конфигурации и размерности. Во всех примерах используется БД TPC-H, специально разработанная для проведения оценочных тестов систем принятия решений [9]. Схема данных представлена на рисунке 3.

Перед рассмотрением примеров проведем формализацию задачи [10]. Пусть задана схема БД $\mathcal{R} = \{R_1, R_2, \dots, R_k\}$, полученная в результате нормализации отношений [8, 11]. Отношения R_i определены на множестве атрибутов $U = \{A_1, A_2, \dots, A_n\}$. Пусть $[R_i]$ – схема отношения (множество атрибутов, на которых определено отношение R_i). Предположим, что схема \mathcal{R} является редуцированной [11], то есть не существует таких двух отношений, что $[R_i] \subseteq [R_j]$ при $i \neq j$. Кортеж $t[X]$ – совокупность значений атрибутов $A_j \in X \subseteq [R_i]$, заданных в кортеже $t \in R_i$. Неопределенное значение NULL атрибута A_j в кортеже t : $t[A_j] = \text{NULL}$ не равно любому другому значению, в том числе другому неопределенному значению.

Многомерное представление будем задавать в виде совокупности размерностей $\{D_1, D_2, \dots, D_d\}$, где D_k – множество расширенных

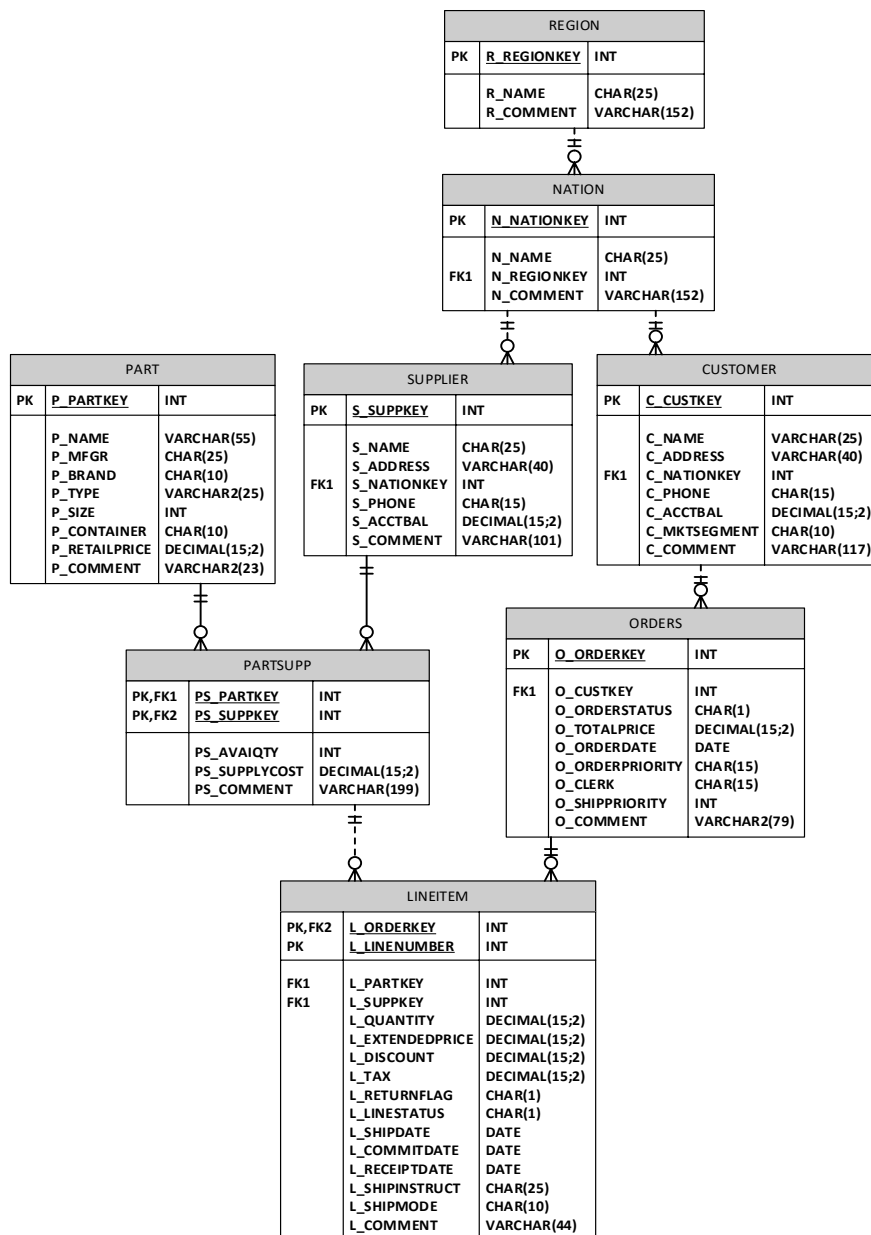


Рис. 3. Схема данных TPC-H

Fig. 3. TPC-H data schema

имен атрибутов: $R_i, A_j, A_\phi \in [R_i]$; M – множество мер, также заданных в виде расширенных имен атрибутов. Значения D_k являются значениями координат гиперкуба, значения M будут располагаться в рабочей области гиперкуба. Для каждой размерности задается ограничение в виде логической формулы F_k .

Пример 1. Необходимо найти данные заказов (код, дата, общая сумма, имя клиента) клиентов из Бразилии и Японии, включающие позиции от поставщика с кодом 54, оплата которых была выполнена с 02.02.1998 по

28.10.1998, с доставкой авиа- и ж/д транспортом, при этом товары были доставлены до 29.11.1998.

С целью упрощения постановки и реализации в примерах не используемые в выборке атрибуты опускаются. Таким образом, в рамках описанной схемы БД имеем следующее подмножество атрибутов: A_1 – код заказа, A_2 – код позиции заказа, A_3 – код клиента, A_4 – код страны, A_5 – код поставщика, A_6 – дата заказа, A_7 – общая сумма заказа, A_8 – имя клиента, A_9 – название страны, A_{10} – дата платежа,

A_{11} – дата доставки, A_{12} – способ доставки. Здесь существуют следующие функциональные зависимости: $DEP = \{A_1 \rightarrow A_3A_6(A_7), A_2 \rightarrow A_1A_5A_{10}A_{11}A_{12}, A_3 \rightarrow A_4A_8, A_4 \rightarrow A_9\}$. На основании данного подмножества получаем упрощенную схему отношений: заказы = $R_1(A_1, A_3, A_6, A_7)$, позиции заказов = $R_2(A_2, A_1, A_5, A_{10}, A_{11}, A_{12})$, клиенты = $R_3(A_3, A_4, A_8)$, страны = $R_4(A_4, A_9)$, где жирным выделены ключевые атрибуты отношений. Одно из возможных представлений гиперкуба приведено на рисунке 4.

Атрибуты измерений выделены жирным шрифтом, атрибуты фактов – курсивом, значения атрибутов – обычным шрифтом. Схема гиперкуба (рис. 4) может быть представлена в следующем виде: $\{R_1.A_6(R_1.A_7)\} \times \{R_2.A_5\{R_2.A_{10}\{R_2.A_{11}\{R_2.A_{12}\}\}\}\} \times \{R_3.A_8\} \times \{R_4.A_9\}$, где $D_1 = \{R_1.A_6\}$, $D_2 = \{R_2.A_5, R_2.A_{10}, R_2.A_{11}, R_2.A_{12}\}$, $D_3 = \{R_3.A_8\}$ и $D_4 = \{R_4.A_9\}$ – измерения, $M = \{R_1.A_7\}$ – факты. Логическое ограничение: $F = ((R_4.A_9 = \text{'BRAZIL'} \vee R_4.A_9 = \text{'JAPAN'}) \wedge R_2.A_5 = 54 \wedge R_2.A_{10} \leq \text{'02.02.1998'} \wedge R_2.A_{10} \geq \text{'28.10.1998'} \wedge R_2.A_{11} < \text{'28.11.1998'} \wedge (R_2.A_{12} = \text{'AIR'} \vee R_2.A_{12} = \text{'RAIL'}))$.

Соответствующий SQL-запрос выглядит следующим образом:

```
select O.O_ORDERKEY, O.O_ORDERDATE,
O.O_TOTALPRICE, C.C_NAME from ORDERS O
  left join LINEITEM L on O.O_ORDERKEY =
= L.L_ORDERKEY
  left join CUSTOMER C on O.O_CUSTKEY =
= C.C_CUSTKEY
  left join NATION N on C.C_NATIONKEY =
= N.N_NATIONKEY
where L.L_SUPPKEY = 54
and L.L_RECEIPTDATE between
TO_DATE('02.02.1992', 'dd.mm.yyyy')
and TO_DATE('28.10.1998', 'dd.mm.yyyy')
and L.L_SHIPDATE < TO_DATE
('28.11.1998', 'dd.mm.yyyy')
and L.L_SHIPMODE in ('AIR', 'RAIL')
and N.N_NAME in ('BRAZIL', 'JAPAN');
```

Пример 2. Необходимо найти данные оплат поставленных товаров поставщику с кодом 75 для клиентов из Европы.

Задано минимальное подмножество атрибутов: A_1 – код позиции заказа, A_2 – код заказа, A_3 – код клиента, A_4 – код страны, A_5 – код региона, A_6 – код поставщика, A_7 – номер позиции, A_8 – дата платежа, A_9 – количество товара в позиции, A_{10} – цена, A_{11} – название страны, A_{12} – название региона. На данном множестве атрибутов существуют следующие функциональные зависимости: $DEP = \{A_1 \rightarrow A_2A_6A_7A_8(A_9A_{10}), A_2 \rightarrow A_3, A_3 \rightarrow A_4, A_4 \rightarrow$

$\rightarrow A_5A_{10}, A_5 \rightarrow A_{11}\}$. Получаем схему отношений: позиции заказов = $R_1(A_1, A_2, A_6, A_7, A_8, A_9, A_{10})$, заказы = $R_2(A_2, A_3)$, клиенты = $R_3(A_3, A_4)$, страны = $R_4(A_4, A_5, A_{11})$, регионы = $R_5(A_5, A_{12})$. Одно из возможных представлений гиперкуба приведено на рисунке 5.

Схема гиперкуба (рис. 5) может быть представлена следующим образом: $\{R_1.A_6\{R_1.A_7(R_1.A_9)\}(R_1.A_{10})\} \times \{R_4.A_{11}\} \times \{R_5.A_{12}\}$, где $D_1 = \{R_1.A_6, R_1.A_7\}$, $D_2 = \{R_4.A_{11}\}$ и $D_3 = \{R_5.A_{12}\}$ – измерения, $M_1 = \{R_1.A_9\}$ и $M_2 = \{R_1.A_{10}\}$ – факты. Логическое ограничение: $F = (R_1.A_6 = 75 \wedge R_5.A_{11} = \text{'EUROPE'})$.

Пример соответствующего SQL-запроса:

```
select L.L_LINENUMBER, L.L_RECEIPTDATE,
L.L_QUANTITY * L.L_EXTENDEDPRI as cost,
N.N_NAME
from LINEITEM L
  left join ORDERS O on L.L_ORDERKEY =
= O.O_ORDERKEY
  left join CUSTOMER C on O.O_CUSTKEY =
= C.C_CUSTKEY
  left join NATION N on
C.C_NATIONKEY = N.N_NATIONKEY
  left join REGION R on
N.N_REGIONKEY = R.R_REGIONKEY
where L.L_SUPPKEY = 75 and R.R_NAME =
= 'EUROPE';
```

Примечательно, что в SciCMS описанные механизмы можно использовать двумя способами. В первом случае при использовании сущностей, содержащих произвольный запрос, гиперкуб строится напрямую средствами SQL. Второй вариант опирается на графовую природу протокола GraphQL и позволяет выполнять построение кубов необходимой размерности на схеме, обладающей свойством соединения без потерь информации, используя исключительно API GraphQL. Возможен также смешанный вариант, при котором запрос GraphQL на сущностях «только для чтения» с произвольными запросами комбинируется с получением связанных сущностей из реальных таблиц.

Несмотря на перечисленные возможности системы, в ряде случаев все же могут потребоваться механизмы импорта и преобразования данных. В качестве примера можно привести работу с нереляционной СУБД, необходимость в наличии заранее рассчитанных агрегатных значений, сложность и высокую стоимость прямых SQL-запросов и т.д. Для применения в подобных обстоятельствах SciCMS также имеет собственный модуль импорта и трансформации данных, основанный на фреймворке с открытым исходным кодом Apache Spark [12].

Код поставщика	54			
Страна	Бразилия		Япония	
Способ доставки	Авиа	Ж/д	Авиа	Ж/д
Дата заказа	<i>Сумма</i>	<i>Сумма</i>	<i>Сумма</i>	<i>Сумма</i>
21.01.1998	74400,72	37642,97	215242,29	135068,75
04.02.2021	176100,92	25692,07	136280,88	330490,79
03.07.1998	36784,63	99554,29	225136,87	183673,68

Рис. 4. Форма данных заказов

Fig. 4. An order data form

Код поставщика	75			
Регион	Европа			
Страна	Германия	Россия	Великобритания	Франция
Номер позиции	<i>Общая сумма</i>	<i>Общая сумма</i>	<i>Общая сумма</i>	<i>Общая сумма</i>
11	21735.2	167477	15585.12	362409.75
12	3300102.47	3780782.67	152926.38	3881625
13	154364	3460320	696736	35362.75

Рис. 5. Форма оплаты поставщику

Fig. 5. A payment form to a supplier

Выводы

Рынок ВІ-систем предлагает широкий выбор как зарубежных, так и отечественных решений. Однако далеко не все они соответствуют требованиям интеграции с существующими системами. Основным недостатком систем этого класса является необходимость хранения и обслуживания собственных структур данных, а также требования к трансформации и экспорту исходных данных в их внутренний формат. Это, в свою очередь, влечет существенные накладные расходы как на настройку операций, так и на их выполнение и сопровождение.

Эффективным решением перечисленных проблем является предлагаемый способ построения трехмерных кубов, реализованный в ИИС SciCMS. Его суть – в использовании уже имеющихся сущностей (построенных на реляционных таблицах) с выделением в них атрибутов даты/времени, местоположения и метрик. Поскольку SciCMS уже хранит все метаданные сущностей, для настройки параметров куба достаточно выбрать нужные атрибуты в качестве осей. Это дает возможность динамически (в зависимости от параметров визуализации) изменять требуемые оси куба. Важной особенностью предлагаемого механизма построения OLAP-кубов является то, что система SciCMS может быть подключена к неограниченному числу источников данных одновре-

менно. Таким образом, любая таблица (представление, SQL-запрос) из любого источника данных напрямую может быть задействована в построении куба (без необходимости экспорта и преобразования).

Основным требованием к реляционным таблицам (лежащим в основе сущностей) является наличие у них свойства соединения без потери информации. Кроме того, столбцы, участвующие в формировании гиперкуба, должны быть надлежащим образом проиндексированы на уровне СУБД. На базе приведенного математического аппарата разобраны примеры задач с соответствующими SQL-запросами и выходными данными, показывающие, как наличие данного свойства позволяет сформировать многомерное представление любой конфигурации и размерности.

Сочетание в SciCMS сущностей, основанных на реальных таблицах и на произвольных запросах, обеспечивает возможность построения сложных OLAP-структур данных в сочетании с быстродействием транзакционной СУБД. В данном случае GraphQL может применяться не только как протокол клиент-серверного взаимодействия, но и как средство компоновки измерений и фактов на клиентской стороне для построения сложных отчетов и визуализаций.

Ввиду своей открытости, а также в соответствии с реализуемым в нашей стране масштаб-

ным проектом по импортозамещению в сфере информационно-коммуникационных технологий данный инструментарий может использо-

ваться широким кругом отечественных предприятий как эффективная альтернатива западным решениям.

Список литературы

1. Ткачева Н.В., Сухомлинова М.И., Северина Ю.Н., Шевченко А.Н. Сравнительная оценка BI-продуктов и платформ как инструмента разработки информационно-аналитических систем // Регион: Системы, экономика, управление. 2014. № 4. С. 235–240.
2. Куликова Г.А. Оценка динамики рынков средств бизнес-аналитики в современных условиях // Вестн. ОКСУ. Информационные технологии. 2021. № 2. С. 76–81. doi: 10.52374/78548596_2021_18_2_76.
3. Кохно П.А., Проколова Т.В. Современный уровень автоматизации систем принятия решений предприятиями оборонно-промышленного комплекса // Научн. вестн. оборонно-промышленного комплекса России. 2017. № 1. С. 40–53.
4. Ильяшенко В.М. Обзор российских систем бизнес-аналитики: вызовы и возможности // Развитие науки и практики в глобально меняющемся мире в условиях рисков: сб. матер. X Междунар. науч.-практич. конф. 2022. С. 109–113.
5. Гончарова Ю.А., Андреева Н.М. Чтение данных периодического регистра сведений «IC:Предприятие 8.1» методами табличной модели информационной базы // Актуальные проблемы авиации и космонавтики: матер. конф. 2012. № 1. С. 393–394.
6. Черныш Б.А., Мурыгин А.В. Динамическая схема GraphQL в реализации интегрированной информационной системы // Программные продукты и системы. 2022. Т. 35. № 4. С. 561–566. doi: 10.15827/0236-235X.140.644-653.
7. Бухонов Д.О., Сергеева О.О., Говоров П.Ю., Дурманов В.В. и др. Применение технологии OLAP-кубов для анализа данных // Вестн. ДИТИ. 2021. № 2. С. 48–54.
8. Гарсиа-Молина Г., Ульман Дж., Уидом Дж. Системы баз данных. Полный курс; [пер. с англ.]. М.: Вильямс, 2017. 1088 с.
9. Липина А.С., Леготин Д.Л. Использование метода анализа иерархий для проверки производительности СУБД // Обучение фрактальной геометрии и информатике в вузе и школе в свете идей академика А.Н. Колмогорова: матер. Междунар. науч.-методич. конф. 2016. С. 117–119.
10. Зыкин С.В., Мосин С.В., Полуянов А.Н. Технология раздельного формирования многомерных данных // Вестн. ДГТУ. 2016. № 2. С. 121–128. doi: 10.12737/19696.
11. Harrington J.L. Relational Database Design and Implementation. Morgan Kaufmann Publ., 2016, 712 p. doi: 10.1016/c2015-0-01537-4.
12. Трофимцов Е.В. Анализ данных с использованием фреймворка Apache Spark // Актуальные проблемы прикладной математики, информатики и механики: сб. тр. Междунар. науч. конф. 2022. С. 1071–1076.

Using three-dimensional data cubes in the implementation of a business intelligence system

**Boris A. Chernysh
Aleksandr V. Murygin**

For citation

Chernysh, B.A., Murygin, A.V. (2023) 'Using three-dimensional data cubes in the implementation of a business intelligence system', *Software & Systems*, 36(2), pp. 237–244 (in Russ.). doi: 10.15827/0236-235X.142.237-244

Article info

Received: 29.11.2022

After revision: 15.02.2023

Accepted: 20.02.2023

Abstract. Business analysis is one of the key management tools that allows getting a reliable picture of the current business situation in an enterprise in all areas of its activity. To ensure this process in any company, there are various data used as its performance indicators. The data source is primarily integrated information systems (IIS) of various types (ERP – Enterprise Resource Planning, CRM – Customer Relationship Management, MES – Manufacturing Execution System, etc.) These systems either incorporate business analysis tools (BI – Business Intelligence) or use specialized solutions that allow performing complex analytical tasks according to a given formulation. This article discusses the features of both approaches, their advantages and disadvantages, provides examples of foreign and domestic products for business analysis existing on the market. The authors propose a method for constructing three-dimensional cubes using the data contained in this system on the example of the BI-module developed by the authors of the IIS SciCMS. There are descriptions of the

used methods and algorithms, the initial requirements and limitations. The authors have carried out the formalization of tasks and considered the mathematical apparatus for constructing multidimensional data models based on information from a fixed set of normalized tables of a relational database. There are examples of SQL queries and output data. In some cases (working with a non-relational DBMS, the need for pre-calculated aggregate values, the complexity and high cost of direct SQL queries, etc.), the described method for building multidimensional cubes may not be applicable. The solution to this problem in SciCMS is its own data import and transformation module based on an open source library. The article summarizes the main advantages and disadvantages of the proposed approach, the prospects for its use in domestic enterprises.

Keywords: multidimensional representation, database, BI, OLTP, OLAP, business intelligence, analytical processing, three-dimensional cube, relational schema, normalization, normal form, integrated information system, GraphQL, SciCMS

Reference List

1. Tkacheva, N.V., Sukhomlinova, M.I., Severina, Yu.N., Shevchenko, A.N. (2014) 'Comparative appraisal of BI-products and platforms as an instrument of analytical systems engineering', *Region: Systems, Economy, Management*, (4), pp. 235–240 (in Russ.).
2. Kulikova, G.A. (2021) 'Assessment of the dynamics of the markets of business intelligence tools in modern conditions', *Bul. of the Educ. Consortium Central Russ. Univ. Inform. Technol.*, (2), pp. 76–81. doi: 10.52374/78548596_2021_18_2_76 (in Russ.).
3. Kokhno, P.A., Prokopova, T.V. (2017) 'Modern level of automation of systems of decision-making of defense industry complex by the enterprises', *Sci. Bull. of the Military-Industrial Complex of Russia*, (1), pp. 40–53 (in Russ.).
4. Ilyashenko, V.M. (2022) 'Overview of Russian business intelligence systems: Challenges and opportunities', *Proc. 10th Intl. Sci.-Pract. Conf. Development of Sci. and Pract. in a Globally Changing World Under the Conditions of Risks*, pp. 109–113 (in Russ.).
5. Goncharova, Yu.A., Andreeva, N.M. (2012) 'Reading the data of the periodic information register "1C: Enterprise 8.1" using the methods of the information base tabular model', *Proc. Current Problems of Aviation and Astronautics*, (1), pp. 393–394 (in Russ.).
6. Chernysh, B.A., Murygin, A.V. (2022) 'A GraphQL dynamic schema in integrated information system implementation', *Software & Systems*, 34(4), pp. 561–566. doi: 10.15827/0236-235X.140.644-653 (in Russ.).
7. Bukhonov, D.O., Sergeeva, O.O., Govorov, P.Yu., Durmanov, V.V. et al. (2021) 'Application of OLAP cube technology for data analysis', *Vestn. of DIET*, (2), pp. 48–54 (in Russ.).
8. Garcia-Molina, H., Ullman, J.D., Widom, J. (2002) *Database Systems: The Complete Book*, NJ: Prentice Hall Publ., 1248 p. (Russ. ed.: Moscow, 2017, 1088 p.).
9. Lipina, A.S., Legotin, D.L. (2016) 'The use of the analytic hierarchy process for performance testing DBMS', *Proc. Conf.*, pp. 117–119 (in Russ.).
10. Zykin, S.V., Mosin, S.V., Poluyanov, A.N. (2016) 'Technology of separate generation of multidimensional data', *Vestn. of DSTU*, (2), pp. 121–128. doi: 10.12737/19696 (in Russ.).
11. Harrington, J.L. (2016) *Relational Database Design and Implementation*, Morgan Kaufmann Publ., 712 p. doi: 10.1016/c2015-0-01537-4.
12. Trofimtsov, E.V. (2022) 'Data analysis using the Apache Spark framework', *Proc. Intern. Sci. Conf. on Current Problems of Applied Mathematics, Informatics and Mechanics*, pp. 1071–1076 (in Russ.).

Авторы

Черныш Борис Александрович¹,
аспирант, borisblack@mail.ru
Мурыгин Александр Владимирович¹, д.т.н.,
профессор, зав. кафедрой информационно-
управляющих систем, avm514@mail.ru

Authors

Boris A. Chernysh¹,
Postgraduate Student, borisblack@mail.ru
Aleksandr V. Murygin¹, Dr.Sc. (Engineering),
Professor, Head of Department of Information
and Control Systems, avm514@mail.ru

¹ Сибирский государственный университет науки и технологий им. академика М.Ф. Решетнева, г. Красноярск, 660037, Россия

¹ Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, 660037, Russian Federation