

УДК 338.262.7

**ПРОЕКТИРОВАНИЕ ХРАНИЛИЩ И ВИТРИН ДАННЫХ ДЛЯ АНАЛИТИКА****<sup>1</sup>Курьян И.С., <sup>1</sup>Рябцева Л.В., <sup>2</sup>Попова Н.С., <sup>3</sup>Курьян И.С.**<sup>1</sup>*Кузбасский гуманитарно-педагогический институт ФГБОУ ВО «Кемеровский государственный университет», Новокузнецк, e-mail: kuryan.ivan@mail.ru;*<sup>2</sup>*Таштагольский техникум горных технологий и сферы обслуживания, Таштагол, e-mail: pp\_goose@mail.ru;*<sup>3</sup>*Сибирский государственный индустриальный университет, Новокузнецк, e-mail kuryan.ilya@mail.ru*

В данной статье представлены результаты проведенного научного исследования, целью которого является выявление проблем проектирования хранилищ и витрин данных для аналитика на предприятии и способов их решения. Объектом исследования авторами выбрано ООО «УК «СОЮЗ-СЕРВИС»», занимающееся экспертизой промышленных зданий и сооружений, систем электроснабжения, электрооборудования, тепло-снабжения, управлением и автоматизацией промышленных процессов на территории Кузбасса и соседних регионов. Для обработки и анализа данных, полученных в ходе исследования деятельности предприятия, использован метод интерактивного анализа массивов больше чем с двумя размерностями, которые представлены в виде матрицы со строками и столбцами данных (OLAP-куб) с применением реляционных систем управления базами данных (СУБД). Одной из существенных задач при проектировании многомерных массивов данных является выбор учитываемых фактов. На основании анализа специфики бизнес-процессов предприятия авторами выделены следующие виды учитываемых фактов: транзакционные факты (например, регистрация дефектов зданий и сооружений, выявленных в ходе обследования); моментальные снимки – срез массива данных в выбранный момент времени (например, количество дефектов и замечаний за месяц и за год); факты, основанные на документообороте предприятия, такие как: договоры на обследование предприятий или технического задания главного инженера проекта; факты, связанные с состоянием зданий и сооружений (например, факт ремонта или развития дефекта в ходе эксплуатации объекта, факт сноса здания или его реконструкции). Результатом исследования является внедрение в ООО «УК «СОЮЗ-СЕРВИС» спроектированной и протестированной на базе данного предприятия аналитической системы (OLAP), позволяющей наблюдать и анализировать результаты протекания бизнес-процессов на предприятии.

**Ключевые слова:** бизнес, информатика, информация, анализ, OLAP, бизнес-процесс, исследование, куб, IT, оптимизация

**DESIGNING DATA WAREHOUSES AND STOREFRONTS FOR AN ANALYST****<sup>1</sup>Kuryan I.S., <sup>1</sup>Ryabtseva L.V., <sup>2</sup>Popova N.S., <sup>3</sup>Kuryan I.S.**<sup>1</sup>*Kuzbass Humanitarian Pedagogical Institute of Kemerovo State University, Novokuznetsk, e-mail: kuryan.ivan@mail.ru;*<sup>2</sup>*Tashtagol Technical College of Mining Technologies and Service Sector, Tashtagol, e-mail: pp\_goose@mail.ru;*<sup>3</sup>*Siberian State Industrial University, Novokuznetsk, e-mail: kuryan.ilya@mail.ru*

This research paper presents the results of the research study, which aims to discover the problems of designing data warehouses and data marts for the analyst in the enterprise and ways to solve them. The authors selected SOYUZ-SERVICE LLC as the object of the study, which is engaged in the examination of industrial buildings and structures, power supply systems, electrical equipment, heat supply, management and automation of industrial processes in Kuzbass and neighboring regions. To process and analyze the data obtained during the research of the enterprise, the method of interactive analysis of arrays is used more than with two dimensions, which are presented in the form of a matrix with rows and columns of data (OLAP cube) using relational database management systems (DBMS). One of the essential tasks in designing multidimensional data arrays is to choose the facts to consider. Based on the analysis of the specifics of the business processes of the enterprise, the authors identified the following types of facts taken into account: transactional facts, for example, can be cited, registration of defects in buildings and structures identified during the survey; snapshots – a slice of the data array at a selected point in time, for example, the number of defects and remarks per month and per year; facts based on the document flow of the enterprise, such as contracts for the survey of enterprises or the technical assignment of the chief engineer of the project; facts related to the condition of buildings and structures, for example, the fact of repair or development of a defect during the operation of the facility, the fact of demolition of the building or its reconstruction. The result of the study is the introduction of an analytical system (OLAP) designed and tested on the basis of this enterprise in LLC UK SOYUZ-SERVICE, which allows you to observe and analyze the results of business processes at the enterprise.

**Keywords:** business, computer science, information, analysis, OLAP, business process, research, cube, IT, optimization

Хранилища и витрины данных являются фундаментом для разработки и использования систем поддержки принятия решений. Системы, спроектированные на основе

технологии хранилищ и витрин данных, позволяют эффективно анализировать клиентскую базу, продажи и доходы предприятия. Хранилище данных – это удобно структу-

рированный массив данных организации, для обработки и хранения которого используется единый аппаратно-программный комплекс, дающий возможность осуществлять: оперативный доступ к текущей информации и к статистике прошлых периодов; многомерную аналитику данных; прогнозирование и статистическую выборку в рамках заданной нормативно-справочной информации.

В работе любой организации используются операционные системы, являющиеся источником данных для производственного и управленческого учета. В качестве примера можно привести системы бухгалтерского учета, банковские и биллинговые системы. Применение ETL-приложений (программ для хранения и обработки информации) позволяет осуществлять преобразование данных из источников с последующей их загрузкой для хранения. Для построения хранилища формируется заданная структура и обычно выбирается определенная модель данных. Модель данных представляет собой описание всех сущностей и объектов базы данных корпоративного хранилища данных и включает в себя концептуальную, логическую и физическую модели базы данных.

При формировании концептуальной модели происходит выделение сущностей и прописываются их взаимосвязи. В рамках логической модели все сущности относятся к каким-либо бизнес-областям, детально описываются, структурируются их взаимосвязи. Физическая модель базы данных подразумевает определение ее структуры на уровне отдельных таблиц, полей, индексов, партиций. В дальнейшем хранилище данных используется как источник для отчетности: формируются области анализа и витрины данных [1].

На уровне описания принципов построения хранилищ данных структура построения выглядит достаточно простой для внедрения и реализации на предприятии. Но, основываясь на опыте, полученном в ходе проектирования и внедрения хранилищ данных, а также на разборе примеров построения схожих хранилищ, можно отметить, что на всех этих этапах создания обычно возникает ряд серьезных проблем как теоретического, так и практического характера. При этом наиболее существенные вопросы и сложности появляются именно при эксплуатации и сопровождении хранилищ и витрин данных. В связи с этим проектирование хранилищ и витрин данных представляется актуальной проблемой для аналитика на предприятиях различных отраслей промышленности.

Целью исследования является выявление проблем проектирования хранилищ и витрин данных для аналитика на предприятии и способов их решения. Объектом исследования авторами выбрано ООО «УК «СОЮЗ-СЕРВИС»», занимающееся экспертизой промышленных зданий и сооружений, систем электроснабжения, электрооборудования, теплоснабжения, управлением и автоматизацией промышленных процессов в Кемеровской области – Кузбассе.

#### **Материалы и методы исследования**

В качестве метода проектирования хранилищ и витрин данных в исследовании использован метод Кимбола. Метод проектирования Кимбола широко используется в Bottom-up дизайне. Принцип проектирования по методу Кимбола основывается на объединении разрозненных витрин данных, которые были спроектированы для аналитика с целью обеспечения быстрого доступа к единому хранилищу данных, необходимых для решения определенных задач предметной области. Подобные витрины данных являются, с одной стороны, базами для формирования пользовательской отчетности, а с другой – хранилищем информации.

В качестве типа витрины данных предлагается использовать схему «звезда». Сутью данной структуры является необходимость наличия таблицы фактов, которая напрямую связана с таблицей измерений. Запросы, генерируемые в данном типе витрин данных, служат элементарным сопоставлением отдельного взятого факта с каждой из таблиц измерений.

Поскольку отличительными особенностями схемы «звезда» являются сравнительно малая избыточность данных и высокая в сравнении с нормализованными структурами производительность, для более эффективного процесса проектирования хранилища данных рекомендуется использовать промышленные (профессиональные) СУБД и инструменты класса OLAP/Reporting, которые раскрывают преимущества схемы «звезда» для значительного ускорения выполнения запросов.

Для обработки и анализа исходных данных в работе использован метод интерактивного анализа многомерных массивов данных (OLAP-куб).

#### **Результаты исследования и их обсуждение**

В результате анализа существующих публикаций по хранению и использованию баз данных было установлено, что существует два больших класса IT-систем: оперативная система обработки транзакций

в реальном времени (OLTP) и оперативная система аналитической обработки информации (OLAP) [2–5].

OLTP – система, позволяющая осуществлять обработку бизнес-процессов в реальном времени, такая система служит неотъемлемой частью ведения бизнеса [6]. При выходе ее из строя нарушается работа ключевых бизнес-процессов, что для крупного бизнеса является неприемлемым. Однако аналитические системы обработки информации позволяют анализировать результаты работы каждого бизнес-процесса, используя методики интеллектуального анализа данных.

База данных OLAP представляет собой агрегированные данные, которые хранятся в многомерных массивах данных. Для этого используются реляционные базы данных, в которых информация хранится независимо и одновременно [6]. Одним из методов представления многомерных массивов данных является OLAP-куб (рис. 1).

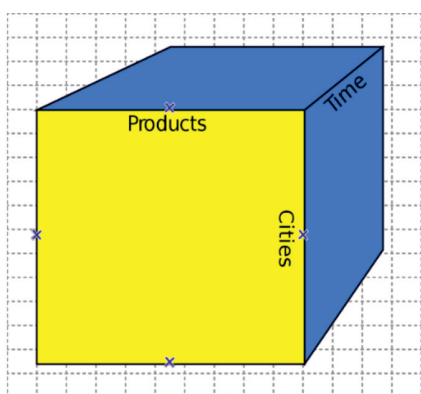


Рис. 1. Схематическое представление OLAP-куба

В OLAP-кубе значениями элементов массива служат меры куба, а индексами массива являются измерения или оси куба. Выражение осей и мер куба можно представить в виде формулы 1.

$$W:(x, y, z) \rightarrow Wxyz, \quad (1)$$

где  $x, y, z$  – измерения;  
 $W$  – мера.

Для исследуемого предприятия ООО «УК «СОЮЗ-СЕРВИС»» были выбраны следующие измерения:

- время: в разрезе года, квартала, месяца;
- место положение объекта: Прокопьевск, Киселевск, др.;
- шифр объекта;
- место образования дефекта: стена, отстойник, колонна, связи опор;
- фактор появления дефекта – влага, рабочая нагрузка, ошибки в строительстве.

На рис. 2 показано использование метода OLAP-куб на исследуемом предприятии.

Список измерений, как видно на рис. 2, представляет собой совокупность атрибутов реляционных таблиц, а именно местонахождение, год, место образования дефекта, фактор.

После анализа особенностей текущего управленческого, бухгалтерского и производственного учета на предприятии, а также с учетом сформулированных руководством стратегических задач и оперативных потребностей были определены требования к организации хранения информации:

- объединение показателей производства предприятия согласно плану;
- обеспечение данных о плановых и фактических показателях производства в подробном представлении, а также в историческом разрезе (доступ к архивной документации);

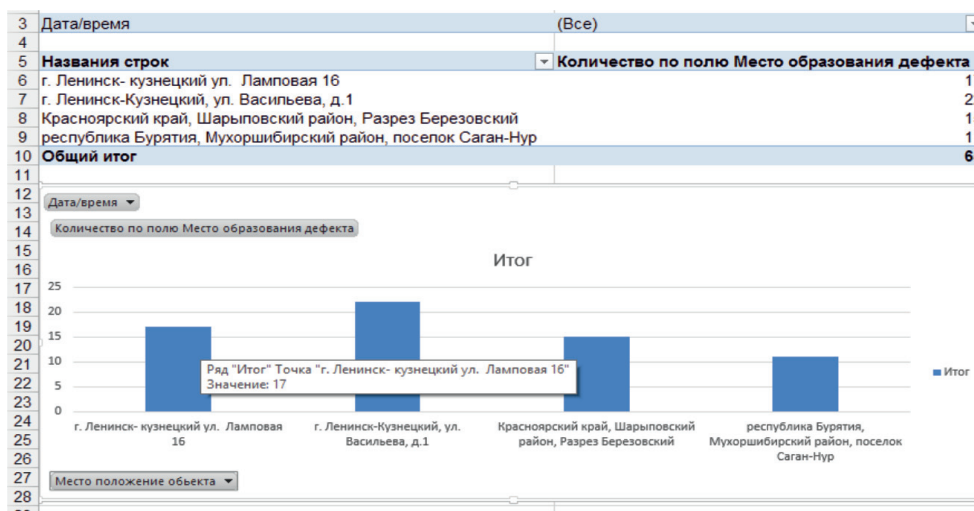


Рис. 2. OLAP-куб на основе данных предприятия ООО «УК «СОЮЗ-СЕРВИС»»

- деление всех показателей на секции, имеющие показатели для каждого временного периода;
- объединение: месяцев – по кварталам, кварталов – по годам;
- сортировка показателей по дате и типу (план или факт).

Инструментом для объединения и структурирования информации из различных источников в общую базу знаний служит витрина данных. Витрина данных, по сути, является так называемым срезом данных из базы данных, относящихся к отдельным аспектам деятельности организации [7].

Вся информация, которая необходима для построения витрины данных, содержится в виде таблиц в БД, пример таких данных представлен на рис. 2.

Список ключевых мер (факторов), необходимых руководителю организации, представляет собой:

- показатель планового значения;
- показатель фактического значения;
- разность планового и фактического значений;
- необходимое количество показателей для конкретной службы.

В соответствии с проектируемой моделью для решения поставленных нами задач необходимо использовать определенную иерархическую систему [8].

Для дальнейшего создания, построения и наглядности работы витрины данных на рассматриваемом производстве необходимо смоделировать OLAP-куб, в основе которого лежит многомерная модель данных. Данную задачу можно решить с помощью пакета прикладных программ Analysis Services (AS), который предлагает инструменты проведения анализа данных. Еще одним удобным средством создания

аналитических решений является семейство прикладных программ Microsoft SQL Server и Microsoft SQL Management Studio. При этом процесс создания базы данных с помощью СУБД MS SQL Server Management Studio начинается с подключения СУБД к заранее установленному локальному серверу.

Спроектированная авторами витрина данных используется в среде Microsoft SQL. Она была реализована в виде реляционной базы данных, что представляет собой только часть реализации процесса проектирования витрины данных.

В базе данных создаются таблицы, в которые вносятся информация, получаемая из соответствующих отделов. Для формирования таблицы необходимо заполнить три столбца: «Имя столбца», «Тип данных», «Разрешить значение NULL». После этого таблицы внутри базы данных можно считать готовыми к работе. На их основе формируется основная таблица, столбцы которой представляют собой описанные выше измерения OLAP-куба.

Следующим этапом действий по проектированию баз и витрин данных является создание связей в реляционной базе данных, которые можно представить в виде ER диаграммы (рис. 3).

ER диаграмма – это диаграмма типа «сущность – связь», в которой описывается процесс взаимодействия таблиц внутри базы данных [9]. Основными элементами диаграммы являются «сущность», а также «первичный» и «внешний» ключи. Сущностью является таблица внутри базы данных. Первичным ключом называется столбец, который определяет таблицу одним единственно верным способом. Внешними ключами таблицы соединяются между собой.

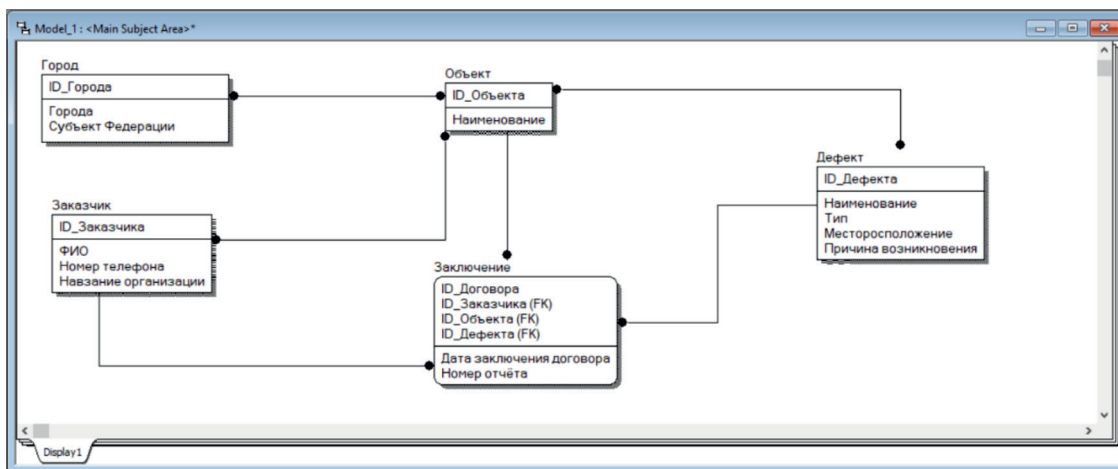


Рис. 3. Схема связи между таблицами



Следует отметить, что выполнение описанных этапов проектирования базы и витрины данных является фундаментальной, но не единственной сложной задачей для формирования полноценной среды для бизнес-аналитики.

### Выводы

Таким образом, по результатам проведенного авторами исследования можно сделать следующие выводы.

1. OLTP-системы, служащие для автоматизации бизнес-процессов компании, позволяют значительно повысить скорость и качество принятия управленческих решений, обладают высокой степенью многообразия, определяемого сферой бизнеса и функциональным направлением внутри него.

2. Хранилище данных является основой и ядром аналитической системы.

3. Основной задачей для проектирования витрин данных служит выбор типа фактов, необходимых для решения поставленных задач. При анализе предметной области деятельности предприятия ООО «УК «СОЮЗ-СЕРВИС»» было выделено несколько фактов:

– регистрация дефектов сооружения – транзакционный факт;

– состояние объектов в определенный момент времени и количество дефектов за месяц или за год – моментальный снимок;

– обследование на основе договоров или технического задания главного инженера проекта – факты, основанные на документообороте предприятия;

– факт ремонта или развития дефекта в ходе эксплуатации объекта, факт сноса здания или его реконструкции эти факты связывает состояние объекта.

Для успешной бизнес-аналитики недостаточно сформировать централизованную и удобно структурированную информацию, крайне важным являются также возможности ее эффективной визуализации.

Традиционные отчеты, даже построенные на основе единого хранилища, как правило, лишены гибкости. В них невозможно повысить или снизить степень детализации, «свернуть» данные или представить их «под другим углом». И, чем больше вариантов представления данных обеспечивает система, тем шире поле аналитических возможностей и формулирования бизнес-идей, которые влекут за собой все большие требования к удобству и оперативности визуализации информации.

Комплексное использование предлагаемых в исследовании подходов и методов проектирования хранилища и витрины данных позволит эффективно решать стоящие перед аналитиком задачи в условиях информатизации и цифровизации экономики.

### Список литературы

1. Батьков В.О. Анализ проблем современных хранилищ данных // Труды Международного симпозиума «Надежность и качество». 2013. Т. 1. С. 259–261.
2. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям. СПб.: Питер, 2013. 704 с.
3. Черненко В.М., Толочко С.И. Анализ информационных систем и определение понятий информационная система поддержки оперативных решений // Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение. 2011. Спецвып. № 5. С. 69–80.
4. Сарка Д., Лак М., Йеркич Г. Microsoft SQL – Server 2012. Реализация хранилищ данных. Учебный курс Microsoft. М.: Высшая школа, 2020. 816 с.
5. Спирли Э. Корпоративные хранилища данных. Планирование, разработка и реализация. М.: ИД «Вильямс», 2017. 400 с.
6. Хоббс Л. Oracle 9iR2. Разработка и эксплуатация хранилищ баз данных / Пер. с англ. М.: КУДИЦ-ОБРАЗ, 2014. 592 с.
7. Бергер А., Горбач И.В. Microsoft SQL Server 2018 Analysis Services. OLAP и многомерный анализ данных. СПб.: ВХБ-Петербург, 2017. 928 с.
8. Толдыкина Е.В., Кудинов А.В. Разработка архитектуры и программных средств витрин данных для предприятия нефтегазовой отрасли // Известия Томского политехнического университета (Известия ТПУ). 2009. № 5. С. 79–85.
9. Ларсон Б. Разработка бизнес-аналитики в Microsoft SQL Server 2005. СПб.: Питер, 2018. 684 с.