

Е. Ю. ПОНОМАРЕВА, студентка НТУ «ХПИ»,
Н. И. БЕЗМЕНОВ, канд. техн. наук, профессор НТУ «ХПИ»,
В. Г. БОРИСОВ, канд. техн. наук, доцент НТУ «ХПИ»

РАЗРАБОТКА ЭЛЕМЕНТОВ БИЗНЕС-СИСТЕМЫ КОММЕРЧЕСКОГО МЕДИЦИНСКОГО УЧРЕЖДЕНИЯ

У статті описана розробка системи підтримки прийняття рішень, що дозволяє досить швидко завантажувати і аналізувати дані через багатогранну призму різних умов за короткий проміжок часу. Business Intelligence система розроблена для аналізу даних мережі комерційних медичних установ. За основу узят модель обліку лікарняних рахунків, яка використовується в таких розвинених країнах як США та Великобританія.

В статье описана разработка системы поддержки принятия решений, позволяющая предельно быстро загружать и анализировать данные через многогранную призму различных условий за короткое время. Business Intelligence система разработана для анализа данных сети коммерческих медицинских учреждений. За основу взята модель учета больничных счетов, которая используется в таких развитых странах как США и Великобритания.

In the article has been described development of system of support decision-making. This system allows maximum quickly to load and to analyses information through the many-sided prism of different terms in short time. Business Intelligence system is developed for data analysis of network commercial medical establishments. The model of taking stock of hospital accounts, which has been used in such developed countries as the USA and Great Britain was took as base.

Введение. С нынешним развитием ИТ-индустрии простой автоматизации повседневных процессов недостаточно. Каждый день возникают все новые задачи, для решения которых необходима целая команда квалифицированных специалистов в различных направлениях разработки программного обеспечения. Одной из таких задач является предоставление работникам учреждения программного продукта, позволяющего значительно упростить анализ данных.

Интеллектуальные системы принятия решений (*Business Intelligence*) – программное обеспечение, позволяющее компаниям, используя свои многочисленные базы данных (БД), донести информацию в простом и доступном виде до своих работников, менеджмента и бизнес-партнеров. Эти системы уже используются тысячами компаний при поиске путей повышения доходов, сокращения расходов, перераспределения ресурсов и повышения эффективности оперативной деятельности.

С ростом проблем, с которыми сталкивается отрасль здравоохранения, становятся все актуальней частные коммерческие медицинские организации и клиники, которые должны сохранять лидерство, постоянно находясь хотя бы на один шаг впереди конкурентов. Бизнес-разведка (системы поддержки принятия решений) может помочь медицинским учреждениям, больницам и

плательщиками увеличить конкурентные преимущества путем разработки «умных» бизнес-решений и создание благоприятных деловых решений.

Среди многочисленных способов использования коммерческими медицинскими организациями платформы *Business Intelligence* для принятия «умных» решений во множестве различных задач в медицинской отрасли можно выделить следующие: финансовый анализ; качество исполнения и анализ рисков; маркетинговый анализ; анализ обслуживания пациентов; анализ эксплуатационных характеристик и стоимости управления.

Постановка задачи. Целью работы является разработка элементов бизнес системы коммерческого медицинского учреждения, используя технологии *Business Intelligence (BI)*.

Описание предметной области. Разработана BI-система для анализа данных коммерческих медицинских учреждений (поликлиник). За основу взята модель учета больничных счетов, которая используется в таких развитых странах как США и Великобритания. Необходимо четко понимать все детали и специфику предоставляемых клиентом данных. Условно эта модель может быть представлена как совокупность трех основных финансовых компонентов:

1. *Учетные записи (Accounts)*. В учетных записях содержится информация о пациенте, поставленных ему диагнозах, пройденных им процедурах, датах прибытия и выписки из больницы, врачах, обследовавших пациента, финансовом состоянии лечения и т.д.
2. *Транзакции (Transactions)*. Транзакции – это документированные финансовые операции, подтверждающие оплату больничных услуг.
3. *Начисления (Charges)*. Начисления (сборы) – это документированные финансовые операции, связанные с запросами в страховые компании об оплате (начислениях) больничных услуг.

Определившись со спецификой данных, необходимо разработать реляционную базу данных для их хранения, которая будет соответствовать двум основным требованиям – сохранность данных и возможность их быстрого извлечения. Для этого нужно пересмотреть принятые правила создания реляционных баз данных и исключить компоненты, которые не позволяют добиться максимально возможной скорости чтения данных, заменив их функции, с помощью внедрения более рациональных решений, учитывающих специфику проблемы.

Описание структуры и технологий реализации. Для реализации данной системы была предусмотрена многоуровневая серверная архитектура (см. рис 1).

Для данного решения часть клиента выполнена в виде *WEB*-приложения, что предоставляет гибкий доступ к данным различным пользователям без привязки к конкретному рабочему месту. Обработанные данные представлены в виде отчетов, данных с наложением различных фильтров,

страниц с графиками, страниц с подытоженными данными, а также в полностью развернутом (детальном) виде.

Сервер *Business Intelligence* представляет собой сервер с развернутым продуктом *Microsoft Analysis Services*, при помощи которого спроектировано многомерное хранилище данных. Реляционная база данных (БД) может находиться как на одном сервере с *Analysis Services*, так и на независимом сервере, что выгодней по производительности и обычно проще и целесообразней в реализации. Реляционная БД представляет собой периодически пополняемое хранилище данных с четкой денормализованной архитектурой, запрограммированным процессом Извлечения-Преобразования-Загрузки данных (*ETL process – extraction, transformation, loading*), который полностью или частично обрабатывает при пополнении хранилища данными.

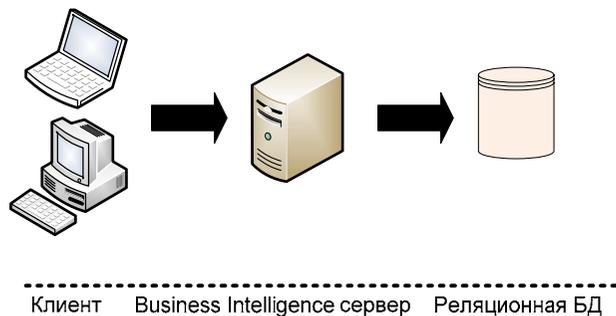


Рис. 1. Многоуровневая серверная архитектура системы бизнес-аналитики

Проектирование ETL модуля. В данной реализации под *ETL* модулем подразумевается набор *sql*-скриптов и/или программного обеспечения работающего в комплексе, на вход которого поступает набор клиентских данных в определенном формате, а результатом его работы являются эти же данные, но очищенные и трансформированные в требуемый формат и загруженные в *SQL* базу данных (см. рис. 2).

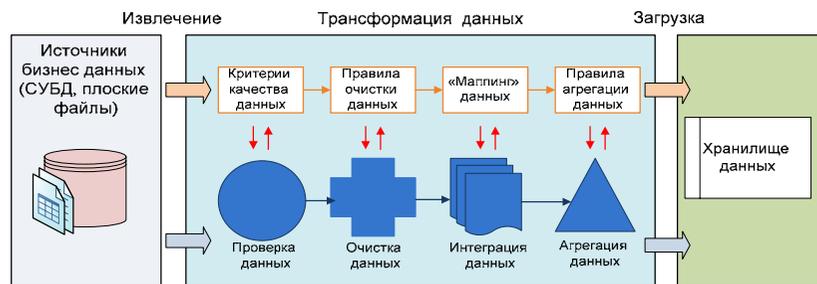


Рис. 2. Обобщенная схема ETL процесса

ETL процесс условно можно разделить на несколько функциональных этапов:

- извлечение данных;
- трансформация данных;
- загрузка данных.

На этапе извлечения выполняется либо запрос к внешней базе данных, либо разбор плоского файла (текстовые файлы, файлы Microsoft Office Excel и т. д.). Во время загрузки происходит сохранение расширенных, очищенных данных в конечных таблицах фактов и многомерной базе данных. Этап трансформации в свою очередь является самым объемным и выполняется в несколько шагов:

1. **Проверка данных.** На данном шаге выполняется проверка данных на их пригодность к последующему использованию в процессе – выполняются проверки, согласно критериям качества, описываемых метаданными.
2. **Очистка данных.** Этот этап позволяет избавиться от «мусора» в данных, а также повысить их качество.
3. **Интеграция данных** представляет собой обновление справочников (словарей), выполнение калькуляций, расчет вычисляемых полей, а также расстановку ключей в таблицах фактов («маппинг»).
4. **Агрегация данных.** Вычисление агрегированных данных.

ETL модуль может повышать качество данных, но никогда не должен никоим образом искажать смысл обрабатываемой информации. Его код должен быть качественным и оптимальным.

Проектирование многомерной базы данных. *SQL* база данных является хранилищем преобразованных клиентских данных и источником данных для *OLAP* уровня и *web*-приложения. Термин «*OLAP*» неразрывно связан с термином «хранилище данных» (*Data Warehouse*). *OLAP (online analytical processing – аналитическая обработка в реальном времени)* представляет собой совокупность средств многомерного анализа данных, накопленных в хранилище.

При проектировании многомерной базы данных использовалась «схема звезды» (см. рис. 3).

Данная схема представляет собой логическую организацию реляционного хранилища данных, при которой модель данных состоит из двух типов таблиц: таблицы фактов, расположенной в «центре» звезды, и несколько таблиц измерений, образующих «лучи». Каждая таблица фактов содержит детальные данные и внешние ключи на таблицы измерений.

Выводы. В данной системе применительно к медицинским учреждениям и коммерческим организациям был реализован специальный набор измерений и метрик, таких как текущий денежный баланс, структурное отделение, ФИО лечащих врачей, диагнозы, процедуры, страховые планы, применяемые

лекарства и медицинские принадлежности, даты поступления и выписки и т.д. При пересечении этих измерений в различных комбинациях можно без особого труда проанализировать следующее: какие лекарства больше всего используются, на какие моменты приходится пик посещаемости каких из отделений, какова текущая задолженность клиентов перед поликлиникой, по какой из страховых компаний максимальная задолженность и т. д.



Рис. 3. Упрощенный пример «схемы звезда»

Таким образом, применение системы BI в медицинской отрасли имеет далеко идущие перспективы и позволяет анализировать большие объёмы информации, заостряя внимание пользователей лишь на ключевых факторах эффективности, моделируя исход различных вариантов действий, отслеживая результаты принятия тех или иных решений.

Список литературы: 1. Kimball R.. The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses. John Wiley&Sons, 1996 – 367 p. 2. Larson B. Delivering Business Intelligence with Microsoft SQL Server 2005. McGraw-Hill/Osborne, 2006. – 792 p. 3. Thomsen E. OLAP Solutions: Building Multidimensional Information Systems. Wiley Computer Publishing, 1997. – 608 p. 4. John C. Hancock; Toreen R.. Practical Business Intelligence with SQL Server 2005. Addison-Wesley Professional, 2006. – 432 p. 5. Hanson C. W. Healthcare Informatics. New York: McGraw-Hill, 2006. – 227 p. 6. Спирли Э. Корпоративные хранилища данных. Планирование, разработка, реализация. Том.1: Пер. с англ. М.: Вильямс, 2001. – 400 с. 7. Корнеев В.В., Куреев А.Ф., и др. Базы данных. Интеллектуальная обработка информации. – М.: Нолидж, 2001. – 352 с.

Поступила в редколлегию 12.01.09

А.А. ПАВЛОВ, д-р техн. наук, проф. каф. АСОИУ НТУУ «КПИ»,
А.В. ЧЕХОВСКИЙ, студент каф. АСОИУ НТУУ «КПИ»

ПОСТРОЕНИЕ МНОГОМЕРНОЙ ПОЛИНОМИАЛЬНОЙ РЕГРЕССИИ. АКТИВНЫЙ ЭКСПЕРИМЕНТ С ОГРАНИЧЕНИЯМИ

Обґрунтовується можливість зведення задачі побудови багатовимірної поліноміальної регресії к послідовності одновимірних регресійних задач в умовах обмеженого активного експерименту і вирішенню відповідних систем алгебраїчних рівнянь з постійними коефіцієнтами.

Обосновывается возможность сведения задачи построения многомерной полиномиальной регрессии к последовательности одномерных регрессионных задач в условиях ограниченного активного эксперимента и решению соответствующих систем алгебраических уравнений с постоянными коэффициентами.

In the article there has been substantiated the possibility of solving a multi-dimensional polynomial regression problem as a sequence of one-dimensional regression problems in a limited experiment and solving the relevant systems of algebraic equations with constant coefficients.

Проблема эффективного восстановления полиномиальной регрессии по настоящее время является актуальной [1-4]. Использование нормированных ортогональных полиномов Форсайта, с учётом рекуррентной формулы их построения [3] позволяет предложить эффективный метод построения многомерной полиномиальной регрессии. Он базируется на следующем анализе построения одномерной полиномиальной регрессии с помощью нормированных ортогональных полиномов Форсайта [3]:

Рассмотрим одномерную модель

$$Y(x) = \theta_0 + \theta_1 x + \dots + \theta_r x^r + E, \quad (1)$$

где x – детерминированная переменная, значение которой в экспериментах исследователь может задавать произвольно. $\theta_j, j = \overline{0, r}$ – неизвестные коэффициенты, r может быть избыточно большим, E – случайная величина с произвольным распределением, σ_E^2 неизвестна, либо существует её верхняя оценка σ^2 . По результатам экспериментов $(x_i, y_i, i = \overline{1, n})$ нужно найти истинное значение r и оценить значение коэффициентов $\theta_j, j = \overline{0, r}$.

Были построены таблицы для оценки дисперсий $\theta_j, j = \overline{0, r}$ для случая, когда значения детерминированного аргумента $x_i, i = \overline{1, n}$ равномерно распределены внутри отрезка с концами $(-a, a), a > 0$.