

УДК 004.652.8

В.М. ЛЕВЬКИН, Н.В. ЧЕРНЕНКО, А.В. ПЕТРИЧЕНКО*Харьковский национальный университет радиоэлектроники, Украина***ИНТЕГРАЛЬНЫЙ КРИТЕРИЙ ОЦЕНКИ ЦЕЛЕСООБРАЗНОСТИ ИНТЕГРАЦИИ
НЕСКОЛЬКИХ БАЗ ДАННЫХ НА ОСНОВЕ ФУНКЦИОНАЛЬНОЙ
ОБЕСПЕЧЕННОСТИ СУБД**

В данной работе исследована проблема, сложившаяся в ходе децентрализованной автоматизации организации, рассмотрены возможные пути решения. Для устранения выявленных недостатков предлагается произвести интеграцию баз данных отдельных разнородных информационных систем в единую базу данных организации. Проведен анализ возможных подходов к интеграции баз данных. Традиционные подходы требуют участия высококвалифицированных специалистов и значительных временных затрат. Для минимизации рисков, связанных с процессом интеграции, предложен интегральный критерий оценки целесообразности интеграции нескольких баз данных на основе функциональной обеспеченности систем управления базами данных.

Ключевые слова: база данных, интеграция, информационная система, система управления базами данных, интегральный критерий, функциональная обеспеченность.

Введение

Активное развитие информационных технологий и отсутствие единой централизованной политики автоматизации деятельности организации привели к тому, что на одном объекте автоматизации могут функционировать несколько информационных систем. Данные системы могут быть разнородными по модели данных используемых баз данных, по технологиям разработки графических интерфейсов пользователей, по технологиям передачи данных и т.д. Сложившаяся ситуация усугубляется тем фактом, что статистика успешности реализации проектов информационных систем говорит о высокой доли проектов, которые в полной мере не удовлетворяют потребности пользователей [1].

По мере функционирования на одном объекте управления нескольких информационных систем происходит накопление информации в их базах данных. Со временем, при увеличении информации в базе данных, коэффициент полезного действия использования ряда информационных систем на одном объекте управления снижается. Это может быть связано с формированием некорректных управляющих воздействий на основании неактуальной, а зачастую, и противоречивой информации, которая выявляются на стыке предметных областей унаследованных информационных систем. Данный факт вызван тем, что целостность данных осуществляется на уровне системы управления базами данных. Сложной задачей является осуществление целост-

ности данных на «стыке» систем, особенно по мере роста объемов хранимых данных.

Как показывает анализ данного вида проблемы, решением является интеграция баз данных информационных систем [2, 3]. Особое место отводится базам данных в виду того, что информационные системы направлены на удовлетворение потребностей пользователей в информации, а непосредственным инструментом хранения и работы с данными являются базы данных. Анализ статистики применения различных моделей данных показывает, что в подавляющем большинстве случаев используются реляционная модель данных или производная от нее – объектно-реляционная модель данных.

Задачам интеграции уделено много внимания в научных трудах [2 – 4]. Начиная с 70-х годов, велись работы по структурной интеграции различных моделей данных. Часть трудов посвящены вопросам интеграции данных [3, 4], некоторые – вопросам интеграции ИС [2]. Существуют также работы, посвященные интеграции баз данных, в которых рассматривается как интеграция структурной составляющей, так и интеграция данных [5]. Независимо от предлагаемых подходов к интеграции, все авторы сходятся во мнении, что это дорогостоящий процесс, требующий участия высококвалифицированных специалистов [2 – 6].

Однако в ходе анализа данной проблемной области не было выявлено подходов или решений, которые бы на этапах, предшествующих процессу интеграции, позволили бы оценить целесооб-

разность проведения работ по ее реализации. Вместе с тем существует риск достижения необходимого результата. Для минимизации возможных рисков в ходе интеграции баз данных унаследованных информационных систем предлагается разработать критерий, который позволит учитывать особенности систем управления базами данных унаследованных информационных систем.

Постановка задачи исследования

Целью проводимых исследований является разработка интегрального критерия оценки целесообразности интеграции нескольких баз данных на основе функциональной обеспеченности систем управления базами данных.

Как отмечалось выше, существуют методы, обеспечивающие интеграцию БД на уровне модели хранения данных, однако они не учитывают в полной мере функциональные особенности СУБД. В данном случае речь не идет о какой-нибудь внештатной ситуации. Дело в том, что в случае соответствия нескольких СУБД одному стандарту, их реализация все равно будет отличаться потому, что разработчики, соблюдая соответствие их продукта стандартам, все равно наполняют свои продукты особенными возможностями, которые выгодно выделяют их продукт среди конкурентов. Речь не идет о том, что какие-то СУБД лучше – они по-своему хороши, они разные.

В качестве примера можно привести следующую ситуацию. Практически все реляционные (или объектно-реляционные) СУБД на сегодняшний момент соответствуют стандарту SQL 92 ANSI/ISO или его производным. Однако, например, механизмы блокировки данных могут отличаться. Запросы, выполняемые в одной СУБД, формируют ожидаемый результат, но те же запросы в иной СУБД могут формировать абсолютно другой результат. Например, `SELECT * FROM SM_TABLE` (для СУБД с поддержкой блокировки на чтение) будет иметь тот же эффект что и `SELECT * FROM SM_TABLE FOR UPDATE` (для СУБД, не поддерживающей блокировку на чтение) [2].

Исходя из выше сказанного, можно сделать вывод: прежде чем переходить к интеграции на уровне модели данных необходимо ответить на вопросы: «А существует ли возможность интеграции БД при определенной функциональной обеспеченности СУБД и насколько широко применяются ее особенности, сколько может стоить такая интеграция? Целесообразно ли ее проводить?».

Для ответа на эти вопросы предлагается некоторый критерий.

Разработка интегрального критерия оценки целесообразности интеграции нескольких баз данных на основе функциональной обеспеченности систем управления базами данных

Критерий оценки целесообразности интеграции на основании функциональной обеспеченности СУБД имеет следующий вид:

$$K_{\text{оц.ц.}} = \max_{0 \leq R \leq 100} (R),$$

где R – интегральный нормированный показатель, описывается выражением (4)

Формирование критерия оценки целесообразности интеграции нескольких баз данных с учетом функциональной обеспеченности СУБД представим в виде алгоритма.

Алгоритм формирования введенного критерия можно представить в виде следующих этапов:

1. Определение перечня показателей интеграции БД IND (на основе экспертных оценок).
2. Определение группы показателей G_{IND} (на основе экспертных оценок).
3. Для каждого показателя выбор группы показателей (из определенных на ш.2), к которой он относится.

4. Определение весовых коэффициентов групп показателей Kg_i ($\sum_{i=1}^I Kg_i = 1$, i – порядковый номер

группы показателей), определяющие степень влияния (важности реализации) каждой из групп показателей на интегральный критерий. Принцип расстановки коэффициентов следующий: чем более критична реализация показателей определенной группы в ходе интеграции нескольких БД, тем выше весовой коэффициент данной группы.

5. Определение коэффициентов степени важности реализации каждого из показателей K_j

($\sum_{j=1}^J K_j = 1$, j – порядковый номер показателя).

6. Определение затрат для выполнения каждого из показателей $IND_{ij}=[0,100]$, $IND_{ij} \in Z$ в рамках своей группы при интеграции нескольких БД, где i – порядковый номер группы, j – порядковый номер показателя в рамках группы. Для определения показателя используют модели оценки трудозатрат и расходования материальных ресурсов. Принцип определения затрат следующий: чем больше потребуются таких затрат, тем ниже значение показателя: 0 – коренное изменение, потребует максимальных трудозатрат; 100 – полное соответствие по данному показателю.

7. В случае если существуют такие показатели, реализация которых будет определять возможность интеграции нескольких БД (критичные показатели – IND_{CRk}), их необходимо выделить в отдельную группу. Возможность реализации интеграции нескольких БД по таким показателям определяется выражением следующего вида:

$$P = \prod_{k=1}^K IND_{CRk}, \quad (1)$$

где IND_{CRk} – критичные показатели, принимают следующие значения:

$$IND_{CRk} = \begin{cases} 0, & k=1,2,3... \\ 1, & \end{cases}$$

Возможность реализации войдет в одну группу показателей, а стоимость реализации – в другую.

8. Определение возможности реализации (R) интеграции нескольких БД в единую по формуле:

$$R = P * \sum_{i=1}^I (Kg_i * \sum_{j=1}^J K_j * IND_{ij}), \quad (2)$$

где P – возможность реализации процесса интеграции, рассчитанная по формуле (1),

Kg_i – весовой коэффициент группы показателей, $i=1,2,3...$

K_j – весовой коэффициент показателя в группе, $j=1,2,3...$

IND_{ij} – нормированное значение i-го показателя в j-й группе.

В случае если есть критичные показатели, то формула (2) приобретает вид:

$$R = \prod_{k=1}^K IND_{CRk} * \sum_{i=1}^I (Kg_i * \sum_{j=1}^J K_j * IND_{ij}) \quad (3)$$

где IND_{CRk} – критичные показатели, принимают следующие значения:

$$IND_{CRk} = \begin{cases} 0, & k=1,2,3... \\ 1, & \end{cases}$$

Kg_i – весовой коэффициент группы показателей, $i=1,2,3...$

K_j – весовой коэффициент показателя в группе, $j=1,2,3...$

IND_{ij} – нормированное значение i-го показателя в j-й группе.

Если критичных показателей не выявлено в ходе анализа функционирования БД, то формула (3) принимает вид:

$$R = \sum_{i=1}^I (Kg_i * \sum_{j=1}^J K_j * IND_{ij}) = \begin{cases} R_1 > 90, \\ R_2 \geq 70, \\ R_3 < 70, \\ R_4 = 0, \end{cases} \quad (4)$$

где Kg_i – весовой коэффициент группы показателей, $i=1,2,3...$

K_j – весовой коэффициент показателя в группе, $j=1,2,3...$

IND_{ij} – нормированное значение i-го показателя в j-й группе.

Для оценки целесообразности интеграции нескольких баз данных в единую необходимо проанализировать результат, который получается в ходе формирования критерия оценки целесообразности интеграции баз данных на основании функциональной обеспеченности систем управления базами данных. В качестве результата получается вещественное число в диапазоне [0,100] в виду нормирования значений показателей, весовых коэффициентов групп показателей и каждого показателя в группе. Интерпретация полученного результата следующая: чем ближе R к 100, тем меньше стоимость реализации интеграции БД. На основании экспертных оценок было предложено 3 диапазона значений. В случае попадания результирующего значения R в первый диапазон ($R > 90$) интеграция БД возможна, для второго диапазона ($R \geq 70$) – возможна, но со значительными затратами, для третьего диапазона – не рекомендуется (есть вероятность что и невозможна).

В случае если $R=0$ интеграция нескольких БД невозможна. В случае интеграции идентичных СУБД (включая версию) нет необходимости оценки целесообразности интеграции СУБД. Процесс интеграции снижается на уровень схемы данных.

Для демонстрации применения предложенного критерия оценки целесообразности интеграции нескольких баз данных на основании функциональной обеспеченности СУБД рассмотрим следующий пример. Пусть на одном объекте автоматизации функционируют две информационные системы, в которых в качестве СУБД используются Oracle и MS Access. Формирование введенного критерия необходимо проводить согласно предложенному алгоритму:

1. Определение показателей.

$IND = \{$ модель хранения данных, назначение БД, специализированные возможности хранения данных, специализированные возможности манипулирования данными, возможность легального использования СУБД, стоимость владения БД, количество поддерживаемых пользователей, уровень безопасности хранения данных, уровень отказоустойчивости, наличие адаптеров для подключения к СУБД в рамках используемой технологии $\}$.

Примеры показателей.

Модель хранения данных во многом определяет возможность интеграции БД и в некоторых случаях может стать критичной в данном процессе. В качестве примеров моделей данных выступают: реляционная, объектно-реляционная, объектная, иерархическая, сетевая.

«Назначение БД» – это БД для хранения оперативной информации, хранилище данных, БД для систем реального времени.

«Специализированные возможности хранения данных». Для крупных производителей СУБД свойственно иметь в своих продуктах встроенные специфические типы хранения данных, которые предоставляют большее удобство использования таких СУБД и высокую безопасность хранения данных. Примеры: Nested tables – вложенные таблицы (структуры данных в рамках атрибута таблицы – представляют объектные возможности СУБД), BLOB – Binary Large Object, LOB – Large Object для СУБД ORACLE позволяют хранить изображения, большие объемы информации в закодированном виде.

«Специализированные возможности манипулирования данными» могут использоваться для переноса бизнес-логики информационной системы на уровень СУБД. Для СУБД ORACLE примером могут выступать аналитические функции.

«Уровень безопасности хранения данных» предполагает различные возможности предоставления доступа к данным на уровне авторизации СУБД, выделения системных и объектных привилегий, использования ролей пользователей.

«Уровень отказоустойчивости» предполагает применение всевозможных технологий резервного копирования и восстановления БД.

«Наличие адаптеров» для подключения к СУБД в рамках используемой технологии может повлиять на выбор той или иной СУБД.

2. Определение групп показателей G_{IND}

G_{IND} {критичные, важные, незначительные}

Критичные – такие показатели, реализация которых будет определять возможность интеграции нескольких БД. В случае нереализуемости интеграции БД по данным показателям нет возможности интеграции в целом.

Важные – показатели, которые в значительной мере определяют успех интеграции нескольких БД, но в случае их невыполнения интеграция возможна.

Незначительные – показатели, несоответствие БД которым не вызывает сложностей при интеграции, но их выполнение позволяют в полной мере удовлетворить потребности пользователей.

3. Установление принадлежности каждого из показателей группам показателей.

К критичным показателям следует отнести следующие: модель хранения данных, назначение БД.

К важным относятся показатели, учитывающие специализированные возможности хранения данных, специализированные возможности манипулирования данными, возможность легального использования СУБД, стоимость владения БД, количество

поддерживаемых пользователей, уровень безопасности хранения данных, уровень отказоустойчивости

В группу незначительных показателей необходимо включить показатель наличия адаптеров для подключения к СУБД в рамках используемой технологии.

Как было отмечено выше, на объекте автоматизации существует 2 информационных системы, в качестве СУБД для них используются лицензионные версии СУБД MS Access 2002 и СУБД ORACLE 10G Standard Edition. При интеграции таких СУБД отдаются следующие предпочтения: сохранность и безопасность хранимых данных, высокая производительность, большое количество обрабатываемых запросов, стоимость интеграции.

В рассмотренном случае в качестве базовой СУБД (на основе которой необходимо реализовать интеграцию) выбрана СУБД ORACLE, в виду того, что она позволяет одновременно обрабатывать сотни тысяч запросов пользователей, имеет клиент-серверную архитектуру, множество простых типов данных и на объекте автоматизации уже функционирует лицензионная копия данной СУБД.

4. Определение весовых коэффициентов для групп {важные, незначительные}: G_{IND1} : $K_{g1}=0,95$, G_{IND2} : $K_{g2}=0,05$. Такие величины определены основании экспертных оценок. В первую группу входят: реализация важных и стоимость реализации критичных показателей (в случае их совместимости $P=1$), во вторую группу входят незначительные показатели: наличие адаптеров для подключения к СУБД в рамках используемой технологии. Для СУБД ORACLE не составляет особого труда поиска возможных адаптеров для любых современных сред программирования.

5. Определение коэффициентов степени важности реализации каждого из показателей K_{ij} :

- модель хранения данных: $K_{11}=0,25$;
- назначение БД: $K_{12}=0,25$;
- специализированные возможности хранения данных: $K_{13}=0,1$;
- специализированные возможности манипулирования данными: $K_{14}=0,1$;
- возможность легального использования СУБД: $K_{15}=0,05$;
- стоимость владения БД: $K_{16}=0,05$;
- количество поддерживаемых пользователей: $K_{17}=0,05$;
- уровень безопасности хранения данных: $K_{18}=0,1$;
- уровень отказоустойчивости: $K_{19}=0,05$;
- наличие адаптеров для подключения к СУБД в рамках используемой технологии: $K_{21}=1$.

6. Определение трудозатрат на интеграцию по введенным показателям IND_{ij} :

– модель хранения данных: $IND_{11}=100$ (идентичные модели),

– назначение БД: $IND_{12}=100$ (идентичное назначение),

– специализированные возможности хранения данных: $IND_{13}=100$ (в СУБД MS Access нет никаких специфических возможностей, которые бы не поддерживались СУБД Oracle),

– специализированные возможности манипулирования данными: $IND_{14}=100$ (в СУБД MS Access нет никаких специфических возможностей, которые бы не поддерживались СУБД Oracle),

– возможность легального использования СУБД: $IND_{15}=100$,

– стоимость владения БД: $IND_{16}=80$ (возрастет нагрузка на администратора БД и возможно потребуется дополнительная вакансия),

– количество поддерживаемых пользователей: $IND_{17}=75$ (необходимо ввести дополнительно набор пользователей, определить их права и привилегии),

– уровень безопасности хранения данных: $K_{18}=100$,

– уровень отказоустойчивости: $K_{19}=100$,

– наличие адаптеров для подключения к СУБД в рамках используемой технологии: $IND_{21}=85$ (могут потребоваться трудозатраты на поиск и адаптацию существующих аналогов).

7. Определение возможности интеграции по критическим показателям: $IND_{CR1}=1$ – существует возможность интеграции идентичных моделей хранения данных, $IND_{CR2}=1$ – существует возможность интеграции одинаковых по назначению БД.

8. Определение возможности реализации (R) интеграции нескольких БД в единую (формула 3):

$$R = (IND_{CR1} * IND_{CR2}) * (K_{g1} * (K_{11} * IND_{11} + K_{12} * IND_{12} + K_{13} * IND_{13} + K_{14} * IND_{14} + K_{15} * IND_{15} + K_{16} * IND_{16} + K_{17} * IND_{17} + K_{18} * IND_{18} + K_{19} * IND_{19}) + K_{g2} * (K_{21} * IND_{21}))$$

и, подставив числовые значения, получим:

$$R = 1 * (0,95 * (25 + 25 + 10 + 10 + 5 + 4 + 3,75 + 10 + 5) + 0,05 * 85) = 92,625 + 4,25 = 96,875.$$

Таким образом, полученное значение данного критерия 96,875 входит в первый диапазон значений ($R > 90$). В данном случае целесообразно проводить интеграцию и она не потребует значительных дополнительных затрат.

Выводы

В статье предложен интегральный критерий оценки целесообразности интеграции нескольких баз данных на основе функциональной обеспеченности систем управления базами данных. Данный критерий позволяет оценить возможность интеграции нескольких БД (учитывая функциональность СУБД) на первоначальном этапе, что позволяет снизить риски и стоимость такого процесса из-за того, что на раннем этапе можно определить невозможность такой интеграции или предварительно сформировать стоимость и сроки проведения работ. Применение данного критерия в информационной технологии интеграции реляционных баз данных позволяет минимизировать риски получения неприемлемого результата.

Литература

1. Кале, В. Внедрение SAP R/3. Руководство для менеджеров и инженеров [Текст] / В.Кале. - М.: Компания АйТи. - 2006. - 511 с.
2. Ziegler, P. Three Decades of Data Integration - All Problems Solved? [Text] / Patrick Ziegler, Klaus R. Dittrich // 18th IFIP World Computer Congress (WCC 2004), Building the Information Society, Kluwer Academic Publishers. - 2004. - Vol. 1. - P. 3-12.
3. Koch, C. Dissertation Data integration against. Multiple Evolving Autonomous Schemata [Электронный ресурс] / С. Koch. - Режим доступа: http://www.csd.uoc.gr/~hy562/Papers/thesis_final.pdf. - 18.10.2011 г.
4. Кулик, Ю.А. Методика создания контекстно-независимой системы интеграции данных [Текст] / Ю.А. Кулик, А.С. Ковалев // Радиоэлектронные и компьютерные системы. - 2011. - № 1. - С. 46-51.
5. Левыкин, В.М. Метод синтеза единой структурной составляющей реляционной модели данных [Текст] / В.М. Левыкин, В.А. Филатов, Н.В. Черненко // Управляющие системы и машины. - 2011. - № 4. - С. 3-8.
6. Stein, Robert E. Re-Engineering the Manufacturing System: Applying the Theory of Constraints [Text] / Robert E. Stein. - Marcel Dekker, Inc. - 2003. - 384 p.
7. Кайт Т. Oracle для профессионалов [Текст]: пер. с англ. / Т. Кайт. - СПб.: ООО «ДиаСофтЮП», 2003. - 672 с.

Поступила в редакцию 20.10.2011

Рецензент: д-р техн. наук, проф., проф. кафедры системотехники В.В. Бескоровайный, Харьковский национальный университет радиоэлектроники, Харьков

ІНТЕГРАЛЬНИЙ КРИТЕРІЙ ОЦІНКИ ДОЦІЛЬНОСТІ ІНТЕГРАЦІЇ ДЕКІЛЬКОХ БАЗ ДАНИХ НА ОСНОВІ ФУНКЦІОНАЛЬНОЇ ЗАБЕЗПЕЧЕНОСТІ СКБД

В.М. Левикін, М.В. Черненко, О.В. Петриченко

В даній роботі проведено дослідження проблеми, що склалася під час проведення децентралізованої автоматизації організації, було розглянуто існуючі методи рішення даних проблем. Для подолання таких недоліків було запропоновано провести інтеграцію баз даних окремих різнорідних інформаційних систем в єдину базу даних організації. Проведено аналіз існуючих підходів до інтеграції баз даних. Традиційні підходи потребують участі висококваліфікованих спеціалістів та значних часових витрат. Для мінімізації ризиків, що пов'язані з процесом інтеграції, запропоновано інтегральний критерій оцінки доцільності інтеграції декількох баз даних на основі функціональної забезпеченості систем керування базами даних.

Ключові слова: база даних, інтеграція, інформаційна система, система керування базами даних, інтегральний критерій, функціональна забезпеченість.

AN INTEGRATION CRITERION FOR ASSESSING EXPEDIENCY OF DATABASES INTEGRATION BASED ON THE FUNCTIONAL PROVIDING OF DBMS

V.M. Levykin, N.V. Chernenko, A.V. Petrichenko

In this article research of problem which formed as result of organization's decentralized automation was conducted. The review of existing methods for solving these problems was realized. An integration of heterogeneous databases of different information systems was proposed as problem-solving. The analysis of existing approaches to the databases integration. However existing approaches need participation of qualified specialists and serious and time costs. An integration criterion based on the functional providing was proposed for minimize of integration risks.

Key words: database, integration, information system, database management system, integration criterion, functional providing.

Левыкин Виктор Макарович – д-р техн. наук, проф., проф. кафедри інформаційних управляючих систем, Харківський національний університет радіоелектроніки, Харків, Україна, e-mail: iyc@kture.kharkov.ua

Черненко Николай Владимирович – аспірант кафедри інформаційних управляючих систем, Харківський національний університет радіоелектроніки, Харків, Україна, e-mail: iyc@kture.kharkov.ua

Петриченко Александр Вячеславович – канд. техн. наук, доц., н.с. кафедри інформаційних управляючих систем, Харківський національний університет радіоелектроніки, Харків, Україна, e-mail: iyc@kture.kharkov.ua