

УДК 044.03; 681.518:061

М.В. Евланов, Е.В. Корнеева

Харьковский национальный университет радиоэлектроники, Харьков

ФОРМАЛИЗОВАННОЕ ОПИСАНИЕ ИНФОРМАЦИОННОЙ МОДЕЛИ БИЗНЕС-ПРОЦЕССА

Рассмотрено формализованное представление информационной модели бизнес-процесса, основанной на концепции хранилищ данных. На основе теоретико-множественных моделей, описывающих реляционные базы данных, разработаны модели, описывающие элементы хранилища данных, создаваемых в реляционных и объектно-реляционных СУБД. Определены основные условия существования хранилища данных, построенного на основе предлагаемого формализованного описания.

Ключевые слова: мониторинг бизнес-процессов, хранилище данных, таблица фактов, связь «один – ко многим», составной первичный ключ.

Введение

Основные проблемы мониторинга бизнес-процессов предприятия. В настоящее время информационно-аналитические системы (ИАС) успешно эксплуатируются на многих предприятиях различных типов. Основная концепция ИАС заключается в преобразовании данных, накопленных в информационных системах (ИС) оперативного управления предприятием, в информацию, которая может использоваться для принятия управленческих решений. Традиционные ИАС обычно работают с историческими данными, подготовленными для анализа с использованием специализированной информационной технологии (ИТ), к которым относят, например, OLAP, Data Mining, Reporting, Data Warehouse и т.п. Однако в настоящее время возникает все большая потребность в специализированных ИТ оперативного мониторинга бизнес-активности (Business Activity Monitoring, BAM) предприятия.

Концепция BAM была предложена аналитической компанией Gartner. В соответствии с этой концепцией BAM рассматривается как предоставление доступа в режиме, близком к режиму реального времени, к наиболее важным индикаторам производительности бизнеса с целью увеличения скорости и эффективности бизнес-операций [1]. Типичный сценарий BAM упрощенно выглядит следующим образом: лицо, принимающее бизнес-решения, оценивает происходящие на предприятии экономические события, используя специализированное программное обеспечение, основанное на наборе правил фильтрации событий и уведомления пользователей о возникающих событиях. Для решения задач BAM используется целый ряд ИТ. Так, например, для получения информации о состоянии процессов и событиях, происходящих в бизнес-процессах (БП), используются ИС управления БП, для доступа к разнообразным источникам данных и интегрирован-

ным приложениям используются ИТ интеграции корпоративных приложений.

Основными задачами BAM являются [1]:

- измерение фактических параметров БП и контроль их отклонений от заданных;
- контроль исполнения регламентов;
- обнаружение нештатных ситуаций;
- персонализированный контроль за исполнителями БП;
- контроль нагрузки на сотрудников и загрузки оборудования.

BAM тесно связан с интеграцией корпоративных приложений: для полноценного обеспечения мониторинга бизнес-событий в реальном времени на крупном предприятии, необходимо интегрировать все используемые ИС.

От того, насколько оперативно сотрудники предприятия могут отреагировать на бизнес-события, во многом зависит эффективность работы предприятия. Решение задач BAM обеспечивает возможность оперативно отслеживать и обрабатывать важнейшие бизнес-события (новую сделку, изменение цепочки поставок, изменение законодательства, запрос на доработку ИС и т.д.). Внедрение и эксплуатация систем BAM способствует принятию быстрых, обоснованных и четких управленческих решений, результатами которых являются конкурентные преимущества, получаемые предприятием. Поэтому разработка специализированных ИТ BAM является актуальной научно-прикладной задачей.

Анализ подходов к построению хранилищ данных для мониторинга бизнес-процессов. Современные ИТ мониторинга бизнес-активности используют единое пространство данных. Это пространство представляется как одно или несколько хранилищ данных (ХД), содержащих хронологическую информацию о ходе выполнения БП предприятия. Под ХД понимается предметно-ориентированный, интегрированный, неизменяемый, поддерживающий хронологию набор данных, ор-

ганизированный для целей поддержки принятия решений [2].

В настоящее время существуют такие подходы к реализации ХД [2]:

- формирование и эксплуатация физического ХД, данные в которое импортируются из оперативных источников данных (ОИД);
- формирование и эксплуатация виртуального ХД, которое представляет собой совокупность аналитических запросов к используемым ОИД;
- формирование и эксплуатация витрин нескольких витрин данных (ХД, содержащих только тематически объединенные данные), данные в которые импортируются из ОИД;
- формирование и эксплуатация физического ХД и совокупности витрин данных, причем данные в ХД импортируются из ОИД, а данные в витрины данных импортируются из ХД.

В этом случае разрабатываемое ХД БП можно представить как гиперкуб. Целостность этого гиперкуба определяется общими правилами проверки целостности данных в ХД, которые реализованы к настоящему времени в большинстве СУБД.

Для реализации гиперкуба ХД и витрин данных в настоящее время используются следующие способы [2]:

- multivariate OLAP, предполагающий использование специализированных многомерных БД для реализации многомерной модели данных;
- relational OLAP, предполагающий реализацию многомерной модели данных на основе реляционных БД;
- hybrid OLAP, предполагающий совместное использование многомерных и реляционных БД для реализации многомерной модели данных.

Использование в качестве оперативных источников данных реляционных и объектно-реляционных БД ИС, эксплуатируемых на предприятии, определяет необходимость первоочередного использования способа relational OLAP в процессе создания ХД или витрин данных.

Выделение нерешенной части проблемы и постановка задачи исследования. Для решения проблемы построения ИТ ВАР в работе [3] предложена обобщенная категорная модель, описывающая взаимодействие основных компонент данной ИТ. Эта модель имеет вид

$$L^{IT} = [L^{ISO}, L^{IM}, L^{AM}, L^{VM}, F_{IM}^{ISO}, F_{ISO}^{IM}, F_{AM}^{ISO}, F_{ISO}^{AM}, F_{VM}^{ISO}, F_{ISO}^{VM}, F_{AM}^{IM}, F_{VM}^{AM}], \quad (1)$$

где L^{IT} – категория структурированных множеств, описывающая ИТ мониторинга БП; L^{ISO} – подкатегория, описывающая изоморфные алфавиты языков описания БП; L^{IM} – подкатегория, описывающая информационную модель БП; L^{AM} – подкатегория, описывающая аналитическую модель БП; L^{VM} –

подкатегория, описывающая визуальную модель БП; F_{IM}^{ISO} – функтор, отображающий подкатегорию L^{ISO} в подкатегорию L^{IM} ; F_{ISO}^{IM} – функтор, отображающий подкатегорию L^{IM} в подкатегорию L^{ISO} ; F_{AM}^{ISO} – функтор, отображающий подкатегорию L^{ISO} в подкатегорию L^{AM} ; F_{ISO}^{AM} – функтор, отображающий подкатегорию L^{AM} в подкатегорию L^{ISO} ; F_{VM}^{ISO} – функтор, отображающий подкатегорию L^{ISO} в подкатегорию L^{VM} ; F_{ISO}^{VM} – функтор, отображающий подкатегорию L^{VM} в подкатегорию L^{ISO} ; F_{AM}^{IM} – функтор, отображающий подкатегорию L^{IM} в подкатегорию L^{AM} и устанавливающий правила преобразования информационной модели БП в аналитическую модель; F_{VM}^{AM} – функтор, отображающий подкатегорию L^{AM} в подкатегорию L^{VM} и устанавливающий правила преобразования аналитической модели БП в визуальную модель.

Предложенное теоретико-категорная модель ИТ ВАР отличается от традиционных категорных моделей ИС и ИТ. Отличие это заключается в изначальном выделении подкатегории L^{ISO} , обеспечивающей изоморфность алфавитов языков моделирования БП. В разрабатываемой ИТ можно выделить три группы таких языков:

- языки информационного моделирования БП;
- языки аналитического моделирования БП;
- языки визуального моделирования БП.

Реализацию подкатегории L^{ISO} предполагается осуществить на основе контекстно-свободных грамматик [4]. За счет этого возможно, в отличие от существующих подходов, достичь следующих преимуществ:

- изначальное согласование информационной, аналитической и визуальной моделей БП, каждая из которых соответствует принятым стандартам;
- использование в процессе интеграции различных моделей БП преимуществ каждого способа моделирования;
- возможность корректировки информационной, аналитической или визуальной модели БП по результатам формирования и использования двух других моделей БП.

Особенностью использования ИТ ВАР, основанной на теоретико-категорной модели (1), является отказ от изначального проектирования ХД, описывающих хронологию выполнения исследуемого БП. Модель (1) предполагает реализацию возможности проектирования ХД по результатам концептуального моделирования данных на основе сформированной визуальной или математической модели БП. Такой подход позволяет формировать ХД с уче-

том особенностей конкретных БП, чьи визуальные или математические модели сформированы по результатам изучения деятельности конкретного предприятия. В этом случае разрабатываемое ХД БП можно представить как гиперкуб, измерения которого определяются выделенными элементарными работами БП.

Однако для решения этой проблемы необходимо формализованное описание ХД, на основе которого становится возможным автоматизированная разработка ХД с учетом особенностей других моделей БП. Между тем, в настоящее время проблеме формализованного описания ХД уделяется недостаточно внимания. Реализованы только отдельные элементы такого описания, которые предполагают изначальное участие человека на всех стадиях концептуального, логического и физического проектирования ХД.

В связи с этим задача разработки моделей ХД, которые позволили бы формализовать процессы формирования описаний ХД на основе некоторых общих языков моделирования, решаемая в рамках создания ИТ ВАРМ в соответствии с моделью (1), является актуальной как с теоретической, так и с практической точек зрения.

Разработка формализованного описания информационной модели бизнес-процесса

Формализованное описание схемы данных, формируемой на основе реляционной модели данных, рассмотрено в работе [5]. Основываясь на этой работе, ХД можно формально описать кортежем вида

$$ХД = \langle \{E_i\}, \{R_j\}, \{D_l\} \rangle, \quad (2)$$

где $\{E_i\}$ – множество сущностей ХД, $i = 1, t$; $\{R_j\}$ – множество связей, определенных на сущностях множества $\{E_i\}$ и элементах этих сущностей, $j = 1, n$; $\{D_l\}$ – множество доменов, определенных в ХД, $l = 1, p$.

Множество всех сущностей $\{E_i\}$ ХД можно описать в виде кортежа

$$\{E_i\} = \langle E_F, \{E_D\} \rangle \quad (3)$$

или

$$\{E_i\} = \langle E_F, \{E_D\}, \{E_S\} \rangle, \quad (4)$$

где E_F – таблица фактов; $\{E_D\}$ – множество измерений, каждое из которых состоит из одной таблицы; $\{E_S\}$ – множество измерений, каждое из которых состоит из нескольких связанных таблиц.

Выражение (3) описывает основные сущности схемы «звезда», выражение (4) – схемы «снежинка». Эти схемы являются типовыми схемами реализации способа relational OLAP [2].

Каждая сущность из множества $\{E_i\}$ может быть описана кортежем вида [6]

$$E_i = \langle n_{E_i}, \text{Tit}_{E_i}, \{e_{E_i}^k\} \rangle, \quad (5)$$

где n_{E_i} – имя сущности E_i ; Tit_{E_i} – заголовок сущности E_i ; $\{e_{E_i}^k\}, k = 1, 2, \dots$ – тело сущности E_i ;

$e_{E_i}^k$ – k-й экземпляр сущности E_i .

Заголовок Tit_{E_i} сущности E_i в свою очередь может быть описан кортежем вида [6]

$$\text{Tit}_{E_i} = \langle \text{atr}_{E_i}^m \rangle \subseteq \text{Atr}_{E_i}, \quad (6)$$

где Atr_{E_i} – множество атрибутов, используемых для формирования заголовков всего множества сущностей E ; $\text{atr}_{E_i}^m$ – m-й атрибут, который используется для описания заголовка сущности E_i .

Каждый атрибут $\text{atr}_{E_i}^m$ сущности E_i может быть описан кортежем вида

$$\text{atr}_{E_i}^m = \langle n_{E_i}^m, D_{E_i}^m, T_{E_i}^m \rangle, \quad (7)$$

где $n_{E_i}^m$ – имя атрибута $\text{atr}_{E_i}^m$; $D_{E_i}^m$ – домен атрибута $\text{atr}_{E_i}^m$, причем $D_{E_i}^m \in \{D_l\}$; $T_{E_i}^m$ – тип атрибута $\text{atr}_{E_i}^m$.

Каждый экземпляр сущности $e_{E_i}^k$ может быть описан кортежем вида

$$e_{E_i}^k = \langle \text{val}_{E_i}^{km} \rangle, \quad (8)$$

где $\text{val}_{E_i}^{km}$ – значение, которое принимает в экземпляре сущности $e_{E_i}^k$ атрибут $\text{atr}_{E_i}^m$, используемый для описания заголовка сущности E_i (при этом $\text{val}_{E_i}^{km} \in D_{E_i}^m$).

Связью $R_j \in \{R_j\}$ в ХД принято называть именovanную значащую ассоциацию между двумя сущностями или сущности с самой собой. В общем случае связь $R_j \in \{R_j\}$ между сущностями E_F и E_D может быть описана кортежем вида [5]

$$R_j = \langle n_{R_j}, \text{Atr}_{E_F}^{R_j}, \text{Atr}_{E_D}^{R_j}, \text{Pow}_{E_F}^{R_j}, \text{Pow}_{E_D}^{R_j}, S_{E_F}^{R_j}, S_{E_D}^{R_j} \rangle, \quad (9)$$

где n_{R_j} – имя связи R_j ; $\text{Atr}_{E_F}^{R_j} \subseteq \text{Atr}_{E_F}$ – подмножество атрибутов сущности E_F , участвующих в образовании связи R_j ; $\text{Atr}_{E_D}^{R_j} \subseteq \text{Atr}_{E_D}$ – подмножество атрибутов сущности E_D , участвующих в образовании связи R_j ; $\text{Pow}_{E_F}^{R_j}$ – мощность связи R_j для

сущности E_F (количество экземпляров $e_{E_F}^k \in B_{E_F}$, которые участвуют в образовании связи R_j); $Pow_{E_D}^{R_j}$ – мощность связи R_j для сущности E_D (количество экземпляров $e_{E_D}^k \in B_{E_D}$, которые участвуют в образовании связи R_j); $S_{E_F}^{R_j}$ – степень участия сущности E_F в связи R_j , которая определяет обязательность участия экземпляров $e_{E_F}^k \in B_{E_F}$ в связи R_j ; $S_{E_D}^{R_j}$ – степень участия сущности E_D в связи R_j , которая определяет обязательность участия экземпляров $e_{E_D}^k \in B_{E_D}$ в связи R_j .

В ХД по сравнению с БД существует только один тип связи, образующий множество $\{R_j\}$, определяемый как тип «один – ко многим» R^{om} . Тогда условие принадлежности связи R_j к данному типу связей, выделенному по количеству участвующих в образовании экземпляров сущностей, определяется следующим образом

$$R_j \in R^{om} \text{ при условии } \begin{cases} Pow_{E_D}^{R_j} = 1; \\ Pow_{E_F}^{R_j} = k, k = 1, p. \end{cases} \quad (10)$$

Чтобы получить детальное описание связей типа «один – ко многим», введем понятие первичного ключа PK_{E_D} для таблицы измерений E_D . Такой ключ будет описываться выражением [5]

$$PK_{E_D} = Atr_{E_D}^{PK} \subseteq Atr_{E_D}, \quad (11)$$

где $Atr_{E_D}^{PK} = \langle atr_{E_D}^{mPK} \rangle$ – подмножество атрибутов, образующих первичный ключ для сущности E_D .

Элементы данного подмножества выделяются по следующему условию [5]

$$\begin{cases} val_{E_D}^{kPK} = \{val_{E_D}^{kmPK}\} \subseteq \langle val_{E_D}^{km} \rangle; \\ val_{E_D}^{kPK} \neq val_{E_D}^{lPK} \text{ если } k \neq l; \\ val_{E_D}^{kPK} = val_{E_D}^{lPK} \text{ если } k = l \end{cases} \quad (12)$$

при учете следующих ограничений

$$val_{E_D}^{kmPK} \neq \emptyset; r(Atr_{E_D}^{PK}) \rightarrow \min, \quad (13)$$

где $val_{E_D}^{kPK}$ – значение первичного ключа для экземпляра сущности $e_{E_D}^k$; $val_{E_D}^{kmPK}$ – значение m -го атрибута, который участвует в образовании первичного ключа, присутствующее в экземпляре сущности $e_{E_D}^k$; $r(Atr_{E_D}^{PK})$ – функция, определяющая мощность

подмножества атрибутов $Atr_{E_D}^{PK}$, образующих первичный ключ сущности E_D .

В ХД по сравнению с моделью реляционной БД отсутствует понятие внешнего ключа. В схеме ХД, кроме первичных ключей, как упоминалось выше, есть понятие уникального составного ключа SK_{E_F} таблицы фактов E_F , объединяющего первичные ключи таблицы измерений. Такой ключ можно описать выражением

$$SK_{E_F} = Atr_{E_F}^{SK} \subseteq Atr_{E_F}, \quad (14)$$

где $Atr_{E_F}^{SK} = \langle atr_{E_F}^{mSK} \rangle$ – подмножество атрибутов, образующих составной ключ для сущности E_F .

Элементы данного подмножества выделяются по следующему условию

$$\begin{cases} \exists E_F, \text{ причем для } e_{E_F}^k \in B_{E_F} \exists e_{E_D}^l \in B_{E_D}; \\ val_{E_F}^{kPK} = val_{E_D}^{lPK}; \\ \forall val_{E_D}^{lPK} = val_{E_F}^{tSK}, \end{cases} \quad (15)$$

где $val_{E_F}^{kPK}$ – значение первичного ключа для экземпляра сущности $e_{E_F}^k$; $val_{E_D}^{lPK}$ – значение первичного ключа для экземпляра сущности $e_{E_D}^l$; $val_{E_F}^{tSK}$ – значение составного ключа для экземпляра сущности $e_{E_F}^t$.

При этом необходимо учесть также, что таблица фактов, как правило, содержит уникальный составной ключ, объединяющий первичные ключи таблицы измерений PK_{E_D} . Поэтому правильность формирования составного ключа таблицы фактов определяется условием

$$E_F \cap E_D = PK_{E_D}. \quad (16)$$

С учетом сказанного выше выражение (9) для связи типа R^{om} можно преобразовать следующим образом

$$R_j^{om} = \langle n_{R_j}, SK_{E_F}^{R_j}, PK_{E_D}^{R_j}, Pow_{E_F}^{R_j}, Pow_{E_D}^{R_j}, S_{E_F}^{R_j}, S_{E_D}^{R_j} \rangle, \quad (17)$$

при условии

$$\begin{cases} Pow_{E_F}^{R_i} \geq 1; \\ Pow_{E_D}^{R_i} = 1. \end{cases} \quad (18)$$

Тогда обобщенное формализованное описание информационной модели БП в виде подкатегории L^{IM} будет иметь вид

$$L^{IM} = \left[\langle n_{E_i}, \langle n_{E_i}^m, D_{E_i}^m, T_{E_i}^m \rangle, \langle val_{E_i}^{km} \rangle, \langle n_{R_j}, SK_{E_F}^{R_j}, PK_{E_D}^{R_j}, Pow_{E_F}^{R_j}, Pow_{E_D}^{R_j}, S_{E_F}^{R_j}, S_{E_D}^{R_j} \rangle \right] \quad (19)$$

при обязательном выполнении условий (12), (13) (15), (16) и (18).

Объектами данной подкатегории являются описания сущностей, атрибутов и их доменов, определяющие структуру и наполнение таблиц ХД или витрин данных. Морфизмами данной подкатегории являются описания связей между отдельными сущностями ХД или витрин данных. Такое декларативное описание информационной модели автоматизированного БД позволяет формировать ХД или витрины данных не только путем явного определения метаданных (описаний сущностей, атрибутов и связей), но и путем формирования метаданных на основе описаний визуальных или математических моделей автоматизированного БП.

Выводы и перспективы дальнейших исследований

Предлагаемое в данной статье формализованное описание информационной модели БП, в основе которого лежит концепция ХД, является дальнейшим усовершенствованием предложенной в работе [5] одним из авторов статьи теоретико-множественной модели данных, основанной на положениях реляционной модели данных Кодда. Однако, в отличие от классического реляционного представления, предлагаемое теоретико-множественное представление позволяет не ограничиваться реляционными языками конструирования сущностей и связей ХД, а использовать многообразие языков, основанных на контекстно-свободных грамматиках. Следует отметить, что в общем случае математические модели таких грамматик базируются на аппарате теории автоматов с магазинной памятью [7], что позволит упростить решение задачи сопряжения языков конструирования и запросов к ХД БП и языков визуального и аналитического моделирования БП.

В качестве примера языка, чья грамматика является контекстно-свободной, можно указать язык XML. Широкое распространение этого языка в современных ИТ и инструментальных средствах разработки ИТ позволяет упростить дальнейшие работы по созданию ИТ ВМ.

Список литературы

1. Черняк, Л. На пути к предприятию, управляемому в реальном времени / Л. Черняк // Открытые системы. – 2002. – № 12. – С. 43-47.
2. Барсегян, А.А. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, Н.И. Холод. – СПб.: БХВ-Петербург, 2008. – 384 с.
3. Євланов, М.В. Формалізований опис технології візуального моделювання автоматизованих бізнес-процесів промислового підприємства / М.В. Євланов, В.О. Антонов, Є.В. Корнєєва // Вісник Академії митної служби України. – 2007. – № 3 (35). – С. 95-100.
4. Корнєєва Е.В. Обобщенная модель автоматизированного бизнес-процесса / Е.В. Корнєєва // 11-й міжнародний молодіжний форум „Радіоелектроніка і молодь у XXI ст.”: зб. матеріалів форуму. – Х.: ХНУРЕ, 2007. – С. 413.
5. Левыкин, В.М. Параллельное проектирование информационного и программного комплексов информационной системы // В.М. Левыкин, М.В. Евланов, В.С. Сугробов // Радиотехника. – 2006. – Вып. 146. – С. 89-98.
6. Дейт, К.Дж. Введение в системы баз данных / К.Дж. Дейт. – 8-е изд.: пер. с англ. – М.: Издательский дом «Вильямс», 2006. – 1328 с.
6. Хопкрофт, Дж. Э. Введение в теорию автоматов, языков и вычислений / Дж.Э. Хопкрофт, Р. Мтовани, Дж.Д. Ульман. – М.: Издательский дом «Вильямс», 2002. – 528 с.

Поступила в редколлегию 25.09.2009

Рецензент: д-р техн. наук, проф. В.П. Пугатин, Харьковский национальный технический университет сельского хозяйства им. Петра Василенко, Харьков.

ФОРМАЛІЗОВАНИЙ ОПИС ІНФОРМАЦІЙНОЇ МОДЕЛІ БІЗНЕС-ПРОЦЕСУ

М.В. Євланов, Є.В. Корнєєва

Розглянуто формалізоване представлення інформаційної моделі бізнес-процесу, заснованої на концепції сховища даних. На основі теоретико-множинних моделей, що описують реляційні бази даних, розроблені моделі, що описують елементи сховища даних, що створюються в реляційних і об'єктно-реляційних СУБД. Визначені основні умови існування сховища даних, побудованого на основі пропонованого формалізованого опису.

Ключові слова: моніторинг бізнес-процесів, сховище даних, таблиця фактів, зв'язок «один – до багатьом», складений первинний ключ.

FORMALIZING DESCRIBE OF INFORMATION MODEL OF BUSINESS-PROCESS

M.V. Evlanov, E.V. Korneeva

Formalize presentation of informative model of business-process, based on conception of depositories of information is considered. On the basis of theoretical-plural models, describing relational databases, models, describing the elements of depository of information, created in реляційних and object-relational database, are developed. The basic terms of existence of depository of information, built on the basis of the offered формалізованого description are certain.

Keywords: monitoring of business-process, depository of information, table of facts, connection «one – to to many», component primary cursor.