

ТЕХНОЛОГИЯ РАЗРАБОТКИ БАЗ ДАННЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

В. М. Грачев², В. И. Есин¹, Н. Г. Полухина³, С. Г. Рассомахин¹

Рассматривается проблема создания необходимой технологии, позволяющей осуществлять эффективную разработку корпоративных баз данных. В качестве решения предлагается новая технология, основывающаяся на семантической модели “объект-событие”, использовании универсальной модели данных, языка модели данных, специального программного инструментария разработчика базы данных. Раскрываются принципиальные отличительные особенности предлагаемой технологии и ее преимущества.

Ключевые слова: база данных, информационная система, модель данных “объект-событие”, универсальная модель данных.

Постановка задачи. Для решения своих бизнес-задач значительному числу институтов, корпораций, предприятий, компаний необходимо создание информационных систем (ИС) и их основы – базы данных (БД), которые обеспечивали бы надежное, безопасное одновременное хранение актуальных, достоверных, целостных больших объемов структурированных и неструктурированных данных из разных существенно отличающихся предметных областей (ПрО) и их обработку. Например, данные в области изучения, освоения и использования ресурсов мирового океана, космоса, медицины, значительного объема данные ускорительных экспериментов, данные о конструкциях, техническом состоянии агрегатов и трубопроводов, их координатах, проектной, технической и эксплуатационной документации, поставках, контрактах и расчетов за газ, капитальном строительстве и т. д. – для ИС газовой промышленности и т. д. При этом заранее

¹ Харьковский национальный университет им. В. Н. Каразина, 61022, Харьков, пл. Свободы, 4.

² Национальный исследовательский ядерный университет “МИФИ”, 115409 Россия, Москва, Каширское шоссе, 31.

³ ФИАН, 119991 Россия, Москва, Ленинский пр-т, 53; e-mail: polukhina@sci.lebedev.ru.

неизвестно, какие данные, когда, какого типа, структуры и т. д. могут понадобиться и использоваться в будущем.

Известные сегодня технологии проектирования баз данных такого класса, как правило, ориентированы на индивидуальную разработку уникальной, имеющей свою собственную модель данных, базы данных для конкретной предметной области. То есть при существующих технологиях база данных разрабатывается уникальным способом, как правило, с “нуля”, хотя и с помощью известных подходов: например, поэтапного проектирования с использованием концептуальной, логической и физической моделей данных (рис. 1). И в этом случае, чтобы удовлетворить в полной мере потребности организаций в функциональности проектируемых БД, их качестве (требование к которым, как правило, в процессе создания только повышаются) потребуется разработка различных БД с различными моделями данных, различный инструмент моделирования и использования данных базы, что, в совокупности, влечет за собой значительные объективные затраты (не всегда даже прогнозируемые) временных, финансовых, людских ресурсов, что крайне нежелательно для заказчика. Поэтому разработчика, как правило, заставляют работать в жестко ограниченных рамках финансовых и временных ресурсов.

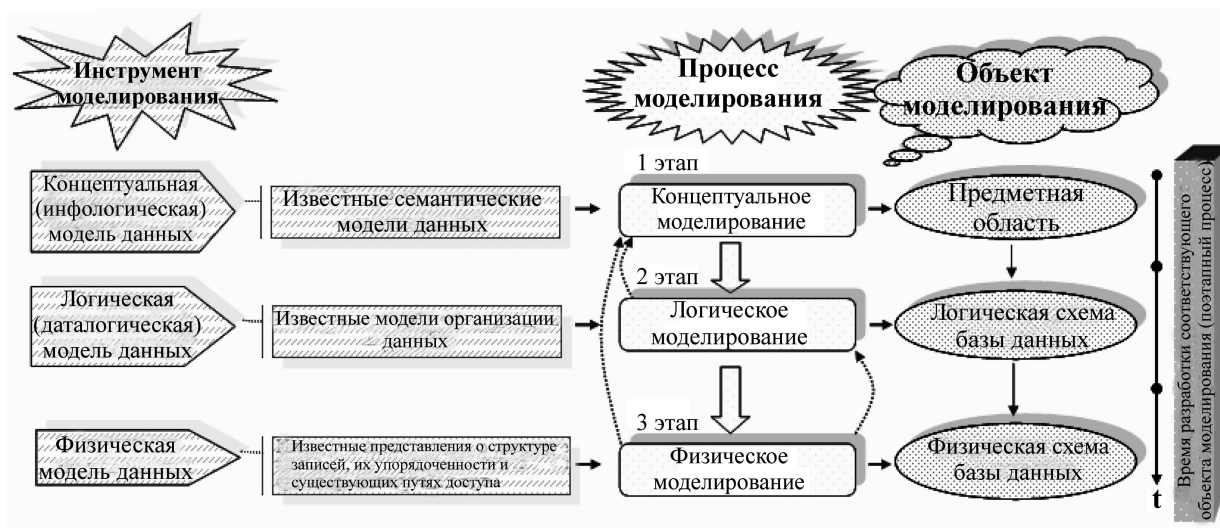


Рис. 1: Традиционная технология разработки баз данных.

В связи с этим возникает объективная необходимость разработки такой технологии, которая: а) позволяла бы при заданных ограничениях на выделяемые временные и финансовые ресурсы качественно проектировать и использовать любого уровня слож-

ности новые базы данных для различных информационных систем; б) способствовала бы использованию уже существующей схемы БД и приложений, с ней работающих; в) позволяла бы проводить реинжиниринг существующих БД информационных систем, построенных на различных принципах, платформах и содержащих значительный объем разнообразных данных, то есть повышала бы эффективность процесса разработки различных баз данных ИС.

Технология разработки баз данных. В результате проведенного анализа классических методов проектирования баз данных, требований, предъявляемых к корпоративным БД информационных систем, а также используемых для моделирования объектов различных предметных областей различных моделей данных была разработана такая технология. Она основана на использовании семантической модели «объект-событие» [1], универсальной модели данных (УМД) [2, 3], языка модели данных [4] и специального программного инструментария разработчика БД. Как совокупность специальных разработанных методов данная технология теоретически не привязана к конкретным программным решениям и платформам, хотя сегодняшняя ее реализация связана с определенными системами и платформами. Основные ее принципиальные отличительные особенности (в сравнении с традиционной), представлены на рис. 2.

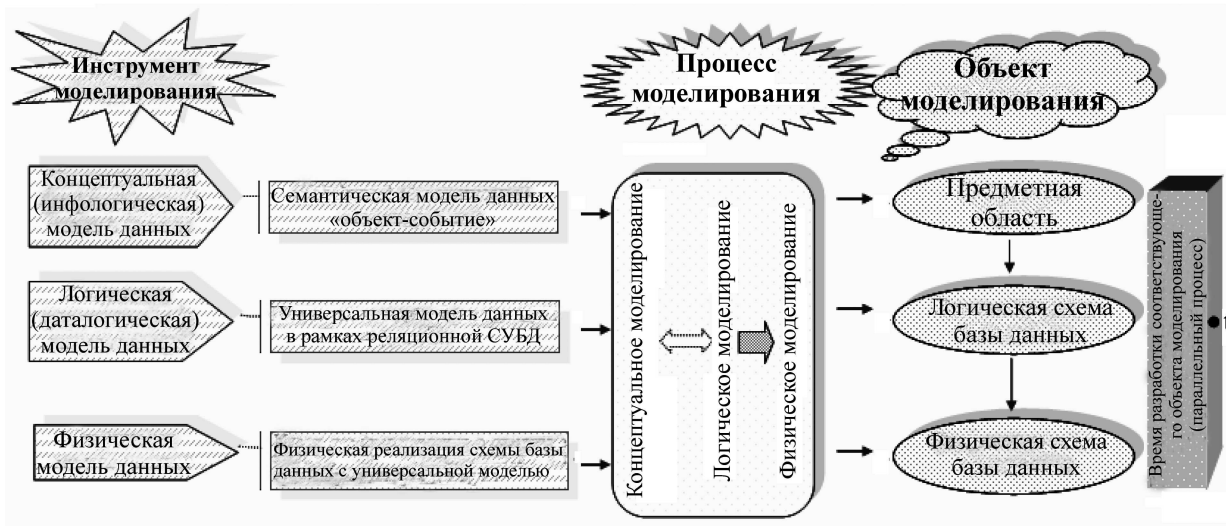


Рис. 2: Предлагаемая технология разработки баз данных.

Особо следует отметить тот факт, что процесс эксплуатации разрабатываемой БД при первом подходе может начаться только лишь после 3 этапа (физического моделирования). В то время как при втором подходе – он может начаться ввиду заранее инстал-

лируемой схемы БД с универсальной моделью, позволяющей на концептуальном уровне адекватно моделировать любые объекты и события любой ПрО в терминологии модели “объект-событие”, сразу же. То есть схему базы данных (ее интенционал) изменять не надо. Она есть инвариант. При этом процесс формирования данных, их модификации, извлечения из БД может осуществляться как с помощью специального разработанного программного инструментария разработчика БД, так и с помощью приложений пользователя, созданного им самостоятельно или по его заказу. Для формирования, модификации, извлечения данных из БД может быть использован как структурный язык запросов (SQL), так и специально разработанный в рамках технологии язык модели данных (ЯМД), позволяющий обращаться к БД с УМД в терминологии, близкой к естественному языку.

Схема БД с УМД, обладая достоинствами реляционных систем управления БД (СУБД), включает и преимущества, присущие перспективным OBDA-системам [5]. А именно она позволяет: обеспечивать более абстрактное представление данных, чем при использовании традиционных моделей данных, проверять качество данных, документировать терминологию организации.

Благодаря инвариантности структуры схемы БД с УМД по отношению к набору данных существенно упрощается и удешевляется процесс проектирования базы данных ИС и приложений, с ней работающих, с повышением их качества. Так как, *во-первых*, процесс проектирования БД сводится не к разработке каждый раз новых объектов ее схемы, что требует достаточно высокой квалификации специалистов и длительного времени, а к инсталляции стандартной схемы БД со структурой УМД и дальнейшим ее наполнением элементами конкретной предметной области. *Во-вторых*, благодаря использованию ЯМД (который во многом стал возможным в виду инвариантности структуры схемы БД с УМД), специального программного инструментария разработчика БД с УМД, библиотеки стандартных программ и методики их использования процесс формирования базы данных существенно облегчается. При этом, если потребуется создание собственного программного обеспечения для работы с базой данных, прикладные программисты даже невысокой квалификации могут справиться очень легко с этой задачей. Им не обязательно вдаваться в подробности модели базы данных, тонкости и особенности диалектов языка SQL. Программистам достаточно знать один из языков программирования и структуру программного интерфейса языка модели данных. Многие сложные для них вопросы, связанные со спецификой работы с БД, например, с организацией работы приложения с БД в многопользовательском режиме (и не только)

решаются благодаря возможностям, реализованным в интерпретаторе ЯМД. А особенность любого программного проекта заключается в том, что людские ресурсы вносят основной вклад в его стоимость (оплата труда программистов различного уровня квалификации может отличаться в разы). И, *в-третьих*, специалисты предметной области, понимающие простые и понятные операторы ЯМД, могут более тесно и эффективно взаимодействовать с разработчиками БД и программистами, ставя им конкретные задачи, согласовывая и контролируя их действия на протяжении всего жизненного цикла эксплуатируемой ИС. Известно, что основная проблема крупных проектов состоит в том, что они требуют координации между большим количеством групп, которым приходится больше общаться между собой [6]. С ростом размера проекта число коммуникационных связей между людьми растет в квадратичной зависимости от количества участников проекта. Что, в свою очередь (наряду с некоторыми другими факторами), обуславливает экспоненциальный рост трудоёмкости.

Проведенные с помощью использования конструктивной модели стоимости *СОСОМО II*, а также некоторых других методов, средств, экспертных оценок [6, 7] расчеты показали, что в целом использование предлагаемой технологии для разработки БД рассматриваемого класса для любой ПрО позволяет уменьшить в (4–20) раз трудозатраты и более чем в 2.5 раза время разработки.

Эффективность (как продуктивность использования ресурсов для достижения требуемого качества) процесса разработки различных баз данных ИС, можно формализовать в виде:

$$E = \sum_q k_q \times \frac{S_q}{P_q}, \quad (1)$$

где $S_q = \sqrt{\frac{1}{2}P_q (\frac{1}{2}P_q - k_a a_q) (\frac{1}{2}P_q - k_b b_q) (\frac{1}{2}P_q - k_t t_q)}$ – “нормированная” площадь, характеризующая соответствующее q -е ($q = 1 \dots Q$) качество (функциональность, применимость, переносимость, надежность и т. д.), и $P_q = k_a a_q + k_b b_q + k_t t_q$ – периметр треугольника ограничений проекта (рис. 3). Здесь k_q – некоторые положительные коэффициенты, учитывающие важность (приоритет) q -го “качества” (численные значения которых были получены с помощью метода анализа иерархий [8, 9]); k_a, k_b, k_t – некоторые положительные нормирующие коэффициенты для соответствующего ресурса (k_a – для финансовых затрат – a_q (на q - качество); k_b – для объема работ – b_q ; k_t – для времени работ – t_q), как показали расчеты. Эффективность E можно увеличить за счет использования новой технологии относительно традиционного подхода разработка БД, более чем в (4.5–40) раз.



Рис. 3: Треугольник ограничений проекта.

Выводы. 1. Проведенные в соответствии с определенными показателями исследования качества БД с УМД показали высокую степень ее функциональной пригодности, уровня защищенности, адекватности отображения объектов реального мира, способности к взаимодействию с различными операционными, прикладными системами и компонентами информационных систем, отличную прозрачность ее архитектуры, достаточно полную и корректную документацию.

2. Благодаря инвариантности структуры схемы БД с УМД по отношению к набору данных, использованию ЯМД, специального программного инструментария разработчика БД с УМД и методики их применения упрощается, ускоряется и удешевляется процесс проектирования любой базы данных ИС. Процесс проектирования БД сводится к установке стандартной схемы БД со структурой УМД и дальнейшему ее наполнению элементами конкретной предметной области с помощью созданного программного инструментария разработчика. Это в целом позволяет уменьшить более чем в (4–20) раз трудозатраты и более чем в 2.5 раза время, необходимое для разработки соответствующей БД.

3. Использование схемы БД с УМД, языка модели данных, специального программного инструментария, библиотеки стандартных программ и методики их применения позволяет повысить эффективность разработки БД относительно традиционной технологии их создания более чем в (4.5–40) раз.

ЛИТЕРАТУРА

- [1] В. И. Есин, *Семантическая модель данных “объект-событие”*. Вісник Харківського національного університету (Х., Харьковський національний університет ім. В. Н. Каразіна, № 925, 65, 2010).

- [2] В. И. Есин, *Универсальная модель данных и ее отличительные особенности*. Вісник Харківського національного університету (Х., Харьковський національний університет ім. В. Н. Каразіна, № 960, 141, 2011).
- [3] В. И. Есин, Ю. А. Пергаменцев, *Технология проектирования модели предприятия на основе универсальной модели данных* <http://www.citforum.ru/database/articles/udm/>.
- [4] В. И. Есин, М. В. Есина, *Язык для универсальной модели данных. Системы обработки інформації* (Х., Харківський університет Повітряних Сил, № 5(95), 193, 2011).
- [5] М. Р. Когаловский, *Системы доступа к данным, основанные на онтологиях* в: Второй Симпозиум “Онтологическое моделирование: состояние и направления исследований и применения”, Казань, Россия, 2010, <http://www.cemi.rssi.ru/mei/articles/kogalov10-02.pdf>.
- [6] С. Макконнелл, *Сколько стоит программный проект* (М., “Русская Редакция”, СПб. : Питер, 2007).
- [7] В. В. Липаев, *Качество программных средств. Методические рекомендации* Под общей ред. проф. д.т.н. А. А. Полякова (М., Янус-К, 2002).
- [8] Т. Саати, *Принятие решений. Метод анализа иерархий* (М., Радио и связь, 1993).
- [9] Томас Л. Саати, *Принятие решений при зависимостях и обратных связях: Аналитические сети*. Пер. с англ. Науч. ред. А. В. Андрейчиков, О. Н. Андрейчикова. (М., Издательство ЛКИ, 2008).

Поступила в редакцию 24 апреля 2014 г.