

УДК 681.3.069

© Меньшикова Л.В.
Menshikova L.

МЕТОДЫ РЕИНЖИНИРИНГА ИНФОРМАЦИОННО-АНАЛИТИЧЕСКОЙ СИСТЕМЫ КРУПНОГО ПРЕДПРИЯТИЯ (часть 2)

LARGE-SCALE COMPANIES BI-SYSTEM RE-ENGINEERING METHODS (path 2)

Аннотация. Рассмотрены методы реинжиниринга ИАС предприятия, а именно метод синтеза унаследованных систем; метод прямой итерационной разработки информационно-аналитических систем; метод перехода к единой централизованной системе на основе единого корпоративного хранилища данных. Для каждого из методов сформулированы архитектурные и организационные принципы, необходимые для успешного развития реинжиниринга в каждом из вариантов.

Annotation. Described BI-systems re-engineering methods: a legacy systems synthesis method, a direct interactive BI development methods and way to the single centralized architecture system based on unitary data warehouse. Formulated architecture and organization principles which necessary to success re-engineering for each of the methods.

Ключевые слова. Информационно-аналитическая система, проект информатизации крупномасштабного предприятия, внедрение информационно-аналитической системы.

Key words. BI-systems, IT-projects of the large-scale companies, information and analytical systems implementation.

Данная статья является продолжением статьи с аналогичным названием, опубликованной ранее, в которой были рассмотрены метод синтеза унаследованных систем и метод прямой итерационной разработки информационно-аналитических систем.

Метод перехода к единой централизованной системе на основе единого корпоративного хранилища данных

Преобразование противоречивой и разрозненной информационной среды любого предприятия в структурированную и планомерно развивающуюся среду – это первый шаг на пути к открытой и гибкой архитектуре корпоративного хранилища данных (КХД), которая является залогом успешной последующей разработки любой системы поддержки принятия решения. Архитектура КХД является отправной точкой, определяющей возможность непротиворечивого и своевременного принятия реше-

ний на основе сформированной модели данных, отвечающей описанию предметной области поддержки принятия решений. Именно от нее в конечном итоге зависит устойчивость и надежность СППР, а также скорость работы всей системы в целом. При разработке архитектуры хранилища данных сначала необходимо сформулировать основополагающие принципы, которые должны соблюдаться в ходе всего проекта. Это, во-первых, некоторые организационные принципы, а, во-вторых, архитектурные принципы.

Принципы создания информационно-аналитических систем на основе хранилищ данных

Опыт построения КХД на крупных многофилиальных предприятиях показывает, что решающее значение для успешного хода данного процесса имеют организационные меры:

- постоянная заинтересованность и поддержка со

Меньшикова Лариса Валерьевна – кандидат физико-математических наук, доцент МИРЭА; экономический советник Департамента информационных систем Банка России, тел./факс: +7(495)753-94-20.

Menshikova Larisa – candidate of physico-mathematical science, associate professor of MIREA; economical adviser of Bank of Russia, tel./fax: +7(495)753-94-20.

стороны спонсора проекта, при этом все спорные вопросы должны решаться административным методом в минимально короткие сроки;

- централизация сбора данных, причем данные должны собираться по возможности неконсолидированные;
- проведение работы с участием бизнес-подразделений, досконально знающих предметную область.

Создание хранилища, вообще говоря, состоит из следующих этапов: анализа бизнес-задач в части определения необходимых для их решения данных, разработки модели данных (включающий агрегацию и консолидацию данных, а также устранение противоречий в них), разработки собственно базы данных на основе построенной модели, первоначальной загрузки данных в КХД (разовой) и организации процессов загрузки данных в КХД, а также выгрузки данных из КХД в витрины данных (наличие этого этапа зависит от типа концептуальной архитектуры).

Создание модели данных КХД возможно только при непосредственном участии специалистов предметной области и никакие внешние консультанты, имеющие опыт создания СППР и КХД, но не знающие досконально особенности деятельности данного конкретного предприятия, на котором проводится автоматизация, здесь не помогут. Аналогично не решат всех проблем и типовые системы аналитики и шаблоны хранилищ данных. Они могут только помочь в этом процессе и сократить сроки создания системы и затраты на ее реализацию.

Следование вышеуказанным организационным принципам построения КХД позволит наиболее эффективно решить многие проблемы, возникающие при построении хранилища, такие как отсутствие эффективного партнерства между ИТ и бизнесом, связанное с тем, что эти две категории сотрудников рассматривают все бизнес-процессы предприятия со своей позиции, часто говорят об одном и том же понятии разными терминами и не понимают до конца проблем, связанных не с их деятельностью. Кроме того, отсутствие ясности у конечного пользователя, какой именно результат он хочет получить и почему этого нельзя добиться с использованием той или иной технологии – еще одна причина появления противоречий между бизнес-подразделениями и ИТ-структурами. Централизация сбора данных позволит минимизировать задержку данных при сборе и частично устранить несогласованные измерения и факты.

К *архитектурным принципам* относятся принципы построения данных, принципы работы механизма запросов, принципы построения рабочих хранилищ, принципы построения метаданных, принципы масштабируемости, принципы управления хранилищем.

Основанная на вышеуказанных принципах архитектура хранилища данных обеспечивает правила и подходы к построению концептуальной и логической моделей и их физической реализации.

При любом выборе концептуальной архитектуры должны быть решены следующие вопросы: движения данных от источников к хранилищу; распределения и обработки данных для хранилища; описания метаданных, проектирования модели и собственно хранилища; обработки запросов и представления информации.

Собственно архитектура хранилища данных делится на три части: архитектуру доступа, архитектуру данных и технологическую архитектуру. *Архитектура доступа* представляет собой описание способов, с помощью которых конечные пользователи получают доступ к данным из хранилища, а также к данным, помещаемым в хранилище. *Архитектура данных* – это описание структуры и жизненного цикла данных. *Технологическая архитектура* – это описание технологических деталей хранилища и его взаимодействий с унаследованными системами и существующими технологиями. Фактически же формулировка архитектурных принципов зависит от среды, в которой должна быть построена архитектура.

Если придерживаться вышеуказанных организационных принципов, то сформулированные выше архитектурные принципы нельзя рассматривать в качестве полного множества принципов для произвольного проекта, кроме того, конкретные принципы должны соответствовать конкретному проекту и методологии разработки и корпоративной культуре предприятия. Нужно учитывать следующие факторы: величину проектного бюджета, количество ИТ-специалистов, выделенных для участия в проекте, время, выделенное на выполнение проекта, структуру организации и опыт организации в области СППР и хранилищ данных.

Итак, после того как обозначены архитектурные принципы, которые должны соблюдаться в ходе проекта построения КХД, выбирается концептуальная архитектура хранилища, наиболее полно им соответствующая. Затем идет этап построения логической модели, который служит началом процесса перехода архитектуры из абстрактного мира концептуальных моделей в физический мир анализа аппаратного обеспечения, масштабируемости, времени ответа и пропускной способности. При проектировании этот уровень абстракции используется для отображения концептуальной модели на модель, в которой определение функциональных возможностей предваряет определение требований к аппаратному обеспечению. Выбор концептуальной архитектуры храни-

лица данных основан на последовательности компромиссов, учитывающих масштабы организации, требования и финансовые ограничения. Выбор логической модели также основан на генеральном плане развития организационной ИТ-архитектуры и принятых в организации стандартах в области ИТ. Хранилище данных должно адаптироваться и взаимодействовать со всеми остальными корпоративными системами.

При оценке времени и ресурсов, необходимых для построения КХД, необходимо учесть следующие факторы: количество и типы пользователей КХД; расположение пользователей и количество мест их расположения; количество и типы источников данных; количество и типы баз данных-источников; количество и типы баз данных КХД; частоту загрузки данных в хранилище; выделенное время подготовки и загрузки данных; необходимое время подготовки и загрузки данных; сложность типичных незапланированных запросов к хранилищу; сложность стандартных отчетов.

Объем начального проекта определяется только исходя из решения наиболее важных бизнес-задач и консолидации данных в наиболее востребованных отчетах. Ни в коем случае не ставится задача одномоментно упорядочить все данные, имеющиеся в организации.

Перед началом работ необходимо провести расчет трудоемкости и других ресурсных затрат на создание КХД в следующих *трех вариантах*:

- реконструкции существующей модели данных;
- разработки модели КХД с использованием какого-либо промышленного шаблона рассматриваемой предметной области;
- разработки модели КХД с нуля под заказ.

Итак, сначала необходимо провести анализ данных, которые будут храниться в КХД, провести их нормализацию, идентификацию и реструктуризацию для исключения повторов и агрегированных данных одновременно с самими данными. В результате структура хранилища должна быть такова, что данные должны быть организованы по принципу вложенности. Это самый длительный по времени период, не приносящий к тому же видимых результатов бизнес-подразделениям. Его можно сократить путем включения в первый этап только наиболее важных отчетов.

Мировая практика показывает, что работы, связанные с разработкой модели данных, являются наиболее сложными и трудоемкими, а поэтому для сокращения срока работ по построению этой модели и снижения риска непостроения единой согласованной оптимальной модели КХД, лучшей практикой является использование какого-

либо промышленного шаблона модели данных, ориентированного на конкретную предметную область модели в качестве альтернативы созданию собственной модели данных «с нуля» под заказ или собственными силами.

Кроме того, мировой опыт построения таких систем говорит о том, что нужно разделять данные, используемые для оперативной обработки и для решения задач анализа. Это позволяет применять структуры данных, которые удовлетворяют требованиям их хранения с учетом использования в вышеуказанных двух кардинально отличающихся системах. Такое разделение позволяет оптимизировать структуру данных под задачи, для выполнения которых они используются. Но информация для проведения аналитических исследований в хранилище данных будет поступать и из транзакционных систем, возможно в некотором агрегированном виде, поэтому при построении всего КХД следует использовать промышленную модель данных, адекватную данной предметной области, которая будет объединяющей для данных OLTP и OLAP – приложений. И еще - в целях построения структуры данных, согласованной в части использования единой терминологии, необходимо иметь единый глоссарий терминов, взаимосвязь между которыми определена единым образом. Поэтому, если выбранный промышленный шаблон позволяет еще и разработать, и вести единый глоссарий терминов, однозначно увязанный с моделью данных, это еще один аргумент использовать промышленный шаблон для построения модели данных.

С учетом вышесказанного *основополагающие принципы, которые следует соблюдать при построении КХД*, сводятся к следующим:

- наличию поддержки спонсора (заказчика), с самого начала заручиться поддержкой спонсора проекта. Заказчик должен уяснить, что построение КХД - процесс длительный и не приносящий скорых и явных результатов, но совершенно необходимый для формирования непротиворечивой информационной среды предприятия;
- поэтапному построению КХД (сначала под одну задачу, затем под вторую и т.д.), так как невозможно одномоментно консолидировать все данные, имеющиеся в организации;
- интеграции и дедубликации существующей информации, то есть в КХД не должно быть несогласованных измерений и фактов;
- централизации процесса сбора данных;
- хранению по возможности единственной версии «правды» и сбору только атомарных данных (в случае, если хранятся агрегированные данные – запрету drill-down);
- очистке и аудиту данных перед их загрузкой в КХД;

- построению модели данных с непосредственным участием бизнес-пользователей и с использованием того или иного промышленного шаблона;
- предоставлению данных пользователю в удобных форматах (из-за форматов, которые не устраивают конечного пользователя можно вообще его лишиться);
- быстрой доставке данных (время реакции должно быть меньше 5 секунд, иначе пользователь будет избегать таких запросов).

OLAP и OLTP

Дадим сначала определение этих понятий.

OLAP (online analytical processing) – оперативная аналитическая обработка данных (либо интерактивный анализ данных), которая реализована в рамках приложения, поддерживающего многомерное представление данных.

OLTP (online transaction processing) – оперативная обработка транзакционных данных, которая реализована в рамках приложения, работающего в режиме реального времени.

Определим связь этих понятий с основными понятиями реляционных и многомерных систем управления баз данных (СУБД): *первичный ключ, отношение или схема данных, кортеж, атрибут, тип данных*. В классической модели СУБД, если при определении переменной отношения явно не указывается ее первичный ключ, то по умолчанию первичным ключом считается полный набор атрибутов заголовка отношения или схемы. В этом случае такая переменная отношения может принимать любое значение, соответствующее заголовку, и первичный ключ не играет роль ограничения. Следовательно, OLTP – приложения используют реляционные СУБД, а OLAP-системы лучше реализуются в рамках многомерной модели.

В связи с этим разные требования предъявляются в транзакционных и аналитических системах к вычислительным комплексам:

1. Форматы данных, ориентированные на надежное хранение данных, отличаются от форматов данных, которые целесообразно использовать для систем с быстрым информационным обслуживанием.

2. В случае, если каналы связи не могут удовлетворить требования к передаче информации в автоматизированных системах, либо если требования безопасности таковы, что информацию следует делить по регионам, надлежит использовать витрины данных, которые являются подмножеством БД.

3. В случае, если используются ВД, то наполнение ВД происходит в часы спада активности пользователей и

возникает *дополнительная проблема* – репликация изменений за день и увязывание его с данными в основной БД.

4. Требования к надежности OLAP-систем и OLTP-систем, а также требования к скорости их восстановления – различны.

Остановимся на истории эволюции понятий OLAP и OLTP. История OLAP начинается в 1993 г., когда была опубликована статья «Обеспечение OLAP (оперативной аналитической обработки) для пользователей-аналитиков». Первоначально казалось, что разделения транзакционных и аналитических систем (OLTP – OLAP) вполне достаточно. Однако вскоре выяснилось, что OLAP – системы очень плохо справляются с ролью посредника между различными транзакционными системами-источниками данных и клиентскими приложениями. Стало ясно, что необходима среда хранения аналитических данных. И поначалу на эту роль претендовали единые базы данных, в которые предлагалось копировать исходную информацию из источников данных. Эта идея оказалась не вполне жизнеспособной, поскольку транзакционные системы разрабатывались, как правило, без единого плана, и содержали противоречивую и несогласованную информацию.

Так появились хранилища данных, предназначенные только для надежного хранения информации, и системы извлечения, очистки и загрузки данных. OLAP-системы работали поверх хранилищ данных.

Вскоре выяснилось, что хранилища данных накапливают настолько важную для организации информацию, что всякий несанкционированный доступ в хранилище чреват серьезными финансовыми потерями. Кроме того, ориентированные на надежное хранение форматы данных плохо сочетаются с требованиями быстрого информационного обслуживания. Территориальная распределенность и организационная структура предприятия также требуют специфического подхода к информационному обслуживанию каждого пользователя или группы пользователей. Все это привело к тому, что практически в любой OLAP-системе есть свой OLTP-сегмент, а количество промежуточных слоев между ними растет с каждым годом, и возникает рассогласование этих уровней.

Целесообразность использования каждого из этих промежуточных слоев напрямую зависит от того, какие задачи стоят перед информационно-аналитической системой и какое ПО в ней используется, то есть от архитектуры АС.

OLTP и OLAP-системы выдвигают разные требования к программно-аппаратному комплексу в части про-

изводительности, пропускной способности, доступности комплекса, моделям данных, организации хранения данных, способов доступа к данным, пиковым нагрузкам, объемам обрабатываемых данных и методам обработки.

Основные отличия OLTP и OLAP-систем состоят в следующем:

- для транзакционной обработки характерно большое количество чтений и записей в базу данных. Аналитическая обработка может потребовать всего несколько обращений к БД;
- длина записей в OLTP обычно не превышает 1000 символов. Аналитический запрос может потребовать мегабайты данных за одно обращение для анализа;
- количество пользователей транзакционной системы может достигать несколько тысяч человек. Число аналитиков значительно меньше;
- нагрузка на транзакционные системы распределяется более или менее равномерно во времени. Нагрузка на аналитические системы, как правило, максимальна в конце отчетных периодов (месяца, квартала, года);
- транзакционная обработка осуществляется в основном применительно к текущим данным. Аналитические вычисления производятся над «историческими» данными;
- данные в транзакционных системах могут обновляться тогда, как в аналитических системах данные должны только добавляться, и попытка внесения изменений задним числом должна вызывать, по меньшей мере, настороженность.

Управление бизнес-процессами

Сначала проводится диагностика ключевых бизнес-процессов, в ходе которой необходимо выяснить, какие процессы – основные, какая у них логика следования, оптимально ли организована логика каждого процесса и насколько эффективен каждый бизнес-процесс для предприятия.

Для этого используют методологию управления бизнес-процессами BPM (Business Process Management). При этом целью является повышение качества исполнения бизнес-процессов, снижение операционных расходов, сокращение временных затрат, получение возможности контроля результатов деятельности для повышения качества управления, а также непрерывное совершенствование внутренних бизнес-процессов.

Можно порекомендовать соблюдать следующие требования при выборе BPM-системы:

- поддержку задач "человек-человек" и удобство интерфейса пользователя;
- поддержку организационной структуры и ролевых групп;

- возможность переназначения заданий, оперативного вмешательства в процесс и обработки исключительных ситуаций;
- возможность управления логикой процесса с рабочего места пользователя;
- удобство использования и администрирования;
- присутствие графических средств разработки моделей бизнес-процесса;
- поддерживаемые архитектуры и стандарты;
- производительность и масштабируемость;
- способность обслуживать многочисленные, продолжительные и распределённые процессы;
- понятный интерфейс настройки и возможность минимального участия ИТ-специалистов во внедрении и поддержке;
- возможность информирования в реальном времени по отклонениям показателей процесса;
- поддержку сервис-ориентированной архитектуры (SOA Service Oriented Architecture);
- присутствие шаблонов бизнес-процессов, на основании которых могут быть разработаны новые процессы;
- невысокую совокупную стоимость владения.

Построить существующий или планируемый бизнес-процесс можно с помощью какого-либо специального инструмента.

Сначала выбирается бизнес-процесс, имеющий наибольшее значение для организации. Не важно, описан он или нет, главное — чтобы процесс был автоматизирован в одной или нескольких информационных системах в любой из существующих технологий.

Затем на основании данных из транзакционных систем с помощью этого средства восстанавливается фактическая модель бизнес-процесса (графические схемы всех вариантов его реального выполнения), проводится ретроспективный анализ результативности процесса и его участников. При этом необходимо выгрузить данные, восстановить на их основе фактический бизнес-процесс и настроить несколько типовых разрезов для анализа процесса. После этого бизнес-аналитики совместно с представителями ИТ дают рекомендации о возможностях оптимизации восстановленного процесса.

Выводы

Рассмотрены методы реинжиниринга ИАС предприятия, а именно метод синтеза унаследованных систем; метод прямой итерационной разработки информационно-аналитических систем; метод перехода к единой централизованной системе на основе единого корпоративного хранилища данных. Для каждого из трех

методов сформулированы архитектурные и организационные принципы, необходимые для успешного развития реинжиниринга в каждом из трех вариантов.

Литература

1. Меньшикова Л.В. Принципы построения корпоративного хранилища данных // Двойные технологии. – 2007.-№4(41).-с.73-76.
2. Menshikova L.V. "Development of Informational and Analytical Systems with Acquisition of Feedback from Users in the Framework of Large-scale Projects/Abstracts of First International Specialized Symposium Space and Global Security of Humanity", Limassol-Cyprus: International Academy of Astronautics, Russian Academy of Cosmonautics, International Association "Znanie", 2009, p.57.
3. Спирли Эрик. Корпоративное хранилище данных. Концепция Инмана// Издательский дом «Вильямс». – 2001г. -396 с.
4. Ralph Kimball, Laura Reeves, Margy Ross, Warren Thornthwaite. The Data Warehouse Lifecycle Toolkit// Wiley computer publishing. – 1998.-771 с.

Материал поступил в редакцию 20. 10. 2011 г.