

ОРГАНИЗАЦИЯ ПЕРЕНОСА ИНФОРМАЦИИ ИЗ БАЗ ДАНЫХ В ИНФОРМАЦИОННЫЕ ХРАНИЛИЩА.

С.В.Малахов, О.В.Иванченко. **Організація переносу інформації з баз даних у інформаційне сховище.** У статті порушено одну з проблем, яка виникає при організації інформаційних сховищ — перенос інформації з баз даних у сховище даних. Розглянуто теоретичний аспект цієї проблеми — питання про умови формування сховищ на основі баз даних. Запропоновано відповідні критерії.

E.V.Malakhov, O.V.Ivanchenco. **The organization of information transferring from databases to datawarehouses.** The information transferring from databases to datawarehouse as one of the datawarehouses organization problems is being touched in the article. Some conditions of datawarehouses forming are based on databases as theoretical aspect of noted problem is being considered. Corresponding criterious are being suggested.

Создание информационных хранилищ, которые стали в последнее время одним из новых направлений развития баз данных, определило круг проблем, возникающих при их построении. В частности, необходимо решить задачи проектирования структуры информационного хранилища, определения критериев переноса данных в информационное хранилище и проблемы, связанные с качеством данных. От того, как будут разрешены эти вопросы, зависит успешная эксплуатация хранилища данных. Первый и последний аспекты проблемы организации хранилищ только упомянуты в данной работе, поскольку обсуждению этих вопросов можно посвятить отдельную статью. Рассмотрим основные критерии переноса данных в информационное хранилище.

Любую базу данных можно рассматривать как объектный архив, предназначенный для хранения атрибутов:

- *пассивных* объектов, не имеющих моделей состояний;
- *активных* объектов, подчиненных жизненному циклу.

В первом случае развитие базы данных заключается только в добавлении и удалении кортежей, соответствующих экземплярам объекта, точнее, их созданию и уничтожению. В этом случае база данных автоматически переходит в разряд информационных хранилищ полностью при выполнении следующего условия: *периодичность удаления кортежей из хранилища соизмерима со временем его существования*. Примером такого хранилища данных может служить электронный каталог библиотеки, “карточки” в который добавляются во время поступления новой литературы. Соответствующая информация формируется на основе данных базы данных отдела комплектации и учета фондов или поступает из внешних источников. Удаление кортежей из системы происходит только в случае списания литературы, которое производится следующим образом: газеты, общественно-политические журналы через 3-5 лет; технические журналы через 10 лет; книги в случае устаревания. Исключения составляют редкие случаи утери, а также износ литературы.

Во втором случае, кроме добавления и удаления кортежей происходит их модификация в зависимости от состояния экземпляра объекта, атрибуты которого содержат данный кортеж. В этом случае можно выделить следующие ситуации:

- *экземпляр* объекта переходит в заключительное состояние, в котором прекращает свое существование;
- *экземпляр* объекта переходит в состояние, в котором изменяются значения части существенных атрибутов;
- у *объекта* (в целом, как абстракции) меняется вектор атрибутов, и соответственно, модель состояний.

В первой ситуации перед уничтожением экземпляра объекта значения его атрибутов могут быть перенесены в информационное хранилище. Например, в библиотеке при оформлении заказа на литературу порождается экземпляр объекта “карточка заказа”. После выполнения заказа экземпляр уничтожается (соответствующая запись удаляется из базы данных “картотека заказов”), а связанные с ним данные переносятся в электронный каталог.

При изменении значений части существенных атрибутов оперируют динамическими базами данных. К существенным атрибутам можно отнести указательные и такое подмножество описательных атрибутов, которое отражает основные или сущностные характеристики объекта. Именно эти атрибуты в большинстве случаев содержат информацию о состоянии экземпляра объекта.

Для студента, например, как объекта подкласса класса “человек” существенными атрибутами являются, во-первых, наследуемые от этого класса “фамилия”, “имя”, “отчество” и “год рождения”, и, во-вторых, часть собственных, присущих только этому объекту, атрибутов, таких как “номер специальности”, “успеваемость по дисциплине” и т.п. Во время обучения данные об успеваемости изменяются, студент может перейти на другую специальность, но кроме того, может измениться его фамилия или семейное положение. Поэтому в качестве возможного критерия переноса данных в информационное хранилище *на уровне подкласса* предлагается рассматривать изменение тех существенных атрибутов, которые присущи только объекту подкласса, изменение же атрибутов, наследуемых от суперкласса, фиксируется *на уровне суперкласса*.

Необходимость в создании хранилища в случае изменения значений части существенных атрибутов существует, если оперативная информация по какой-либо причине представляет ценность для дальнейшего рассмотрения и анализа, например, при статистическом анализе, прогнозировании, выявлении специфических связей и т.п. Тогда в информационное хранилище вводится еще и третье измерение по времени или атрибут времени. Такое хранилище данных предназначено для хранения “снимков” значений атрибутов всех экземпляров объекта с отметкой моментов времени выполнения этих “снимков”. Временные точки для записи в информационное хранилище определяются его пользователем или разработчиком.

Например, при построении модели, представляющей динамику распределения банковских денежных ресурсов, информационной базой может выступать ежедневно пополняемое хранилище данных, которое включает, например, ежедневно рассчитываемый набор показателей, на основе: информации о сумме всех денежных средств, наличествующих в банке на начало текущей даты; информации о сумме клиентских остатков на начало текущей даты; сведений обо всех заключенных к текущей дате сделках. Такая информация дает возможность оценивать складывающиеся тенденции в развитии банка и облегчить выбор стратегических решений. Временной интервал выполнения “снимков”, в данном случае, — один день.

Другой пример — информационное хранилище университета, где среди прочей информации будут храниться и данные о студентах, окончивших это учебное заведение. Пополнение хранилища происходит ежегодно в конце обучения студентов, т.е. временной интервал — год. Возможен и другой вариант. Базы данных деканатов хранят данные об успеваемости студента в течении семестра и экзаменационной сессии. После сдачи экзаменов эти данные становятся постоянными и по предметам, изучение которых в данном семестре закончено, переносятся в хранилище. В этом случае снимки делаются посеместрово.

Если же информация в динамических базах данных после модификации не представляет интереса ни с какой точки зрения, то организация информационного хранилища на основе таких баз не имеет смысла.

Изменение вектора атрибутов объекта, наиболее вероятно, происходит при переходе объекта в другой домен либо при *миграции* экземпляра суперкласса. При этом меняется набор правил и линий поведения объекта и, как следствие, множество состояний, определяющих его жизненный цикл. При возникновении такой ситуации перед сменой модели состояний значения атрибутов экземпляра могут быть перенесены в информационное хранилище. Это можно пояснить и следующим образом: рассматривая миграцию экземпляра суперкласса между объектами подкласса, на уровне последнего можно разбить жизненный цикл суперкласса, даже если он замкнутый, на несколько разомкнутых моделей типа “рождение-смерть” по одной для каждого подкласса [1]. В этом случае для каждого такого подкласса в качестве условия переноса данных в информационное хранилище срабатывает критерий, соответствующий заключительному состоянию объекта. То есть в какой-то степени ситуацию с изменением вектора атрибутов и сменой модели состояний объекта можно свести к ситуации с уничтожением экземпляра объекта.

Расширим пример с информационным хранилищем университета. Когда студент оканчивает университет, уничтожается соответствующий экземпляр объекта подкласса. Перед тем, как это произойдет, информация о студенте-выпускнике, например, результаты последней сессии и защиты диплома, копируется в хранилище данных университета (см. рис.). С точки зрения суперкласса (в данном случае, объекта “человек”) это состояние всего лишь одно из возможных в его жизненном цикле. Поэтому информация об экземпляре суперкласса может быть перенесена в хранилище, которое соответствует уровню суперкласса и формируется на основании данных из ряда хранилищ, соответствующих его подклассам, в моменты миграции данного экземпляра суперкласса.

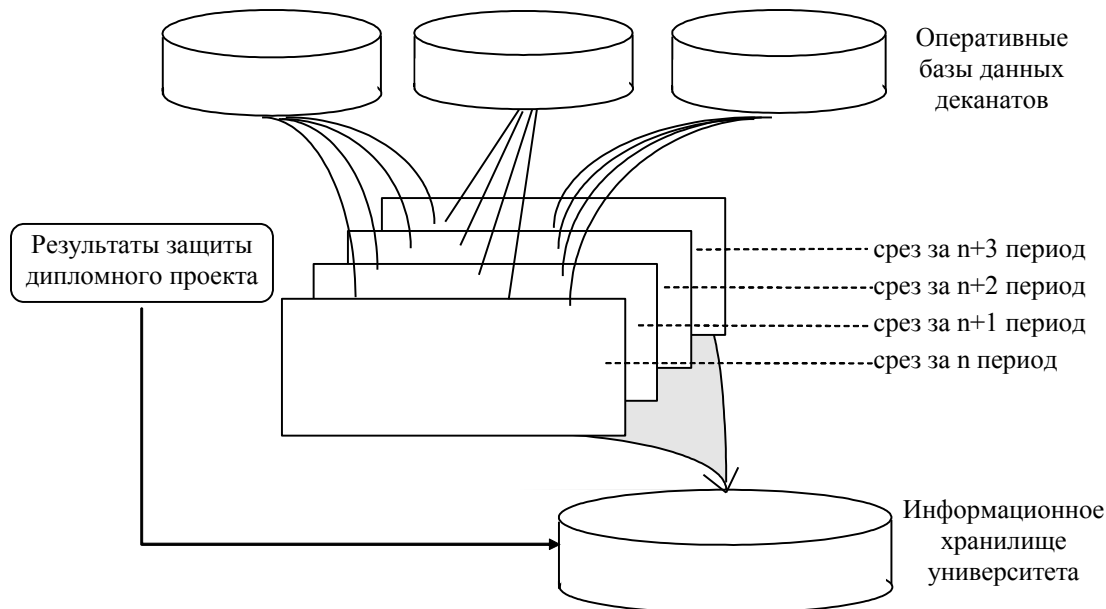


Схема формирования информационного хранилища университета.

Как видно из примеров, приведенных в статье, хранилище данных может быть организовано в соответствии с несколькими критериями. В частности, в примере, который проиллюстрирован рисунком, при формировании хранилища используются критерии, соответствующие как изменению части существенных атрибутов, так и миграции экземпляра объекта суперкласса.

В заключение следует отметить, что рассмотренные вопросы формирования информационных хранилищ являются основой или предпосылкой для решения следующих задач:

- пример формирования хранилищ данных как для объектов подклассов, так и для объекта их суперкласса приводит к необходимости рассмотреть построение иерархических или многоуровневых информационных хранилищ;
- введение третьего измерения по времени в случае изменения значений части существенных атрибутов требует перехода к многомерной структуре хранилищ данных и решения проблемы формирования такой структуры.

Литература

Шлеер С, Меллор С. Объектно-ориентированный анализ: моделирование мира в состояниях. — Киев: Диалектика, 1993.