электронный научно-технический журнал
ИНЖЕНЕРНЫЙ ВЕСТНИК
Издатель ФГБОУ ВПО "МГТУ им. Н.Э. Баумана". Эл No. ФС77-51036. ISSN 2307-0595

Объединение и обновление совместимых реляционных таблиц

 # 05, май 2014
авторы: Брешенков А. В., Белоус В. В.
УДК: 681.3.07

МГТУ им. Н.Э. Баумана
Breshenkov@rambler.ru

В работах [1-3] выполнен анализ проблем разработки методики формирования реляционных таблиц на основе использования информации табличного типа, рассмотрена задача приведения значений атрибутов существующей информации табличного вида к одному типу. Под информацией табличного вида понимаются информация, которая воспринимается как таблицы, но не удовлетворяет требованиям к реляционным таблицам. В данной статье рассматривается одна из задач преобразования существующей информации табличного вида в реляционные таблицы – задача объединения совместимых таблиц.

Во многих организациях используется следующий механизм сбора и обработки данных. В подразделениях организации ведутся базы данных(БД). В установленные даты данные из подразделений пересылаются по электронной почте в региональные центры. В установленные даты данные из региональных центров передаются в центр. Таким образом, в подразделениях организации накапливаются и обрабатываются данные, относящиеся только к конкретному подразделению. В региональных центрах накапливаются и обрабатываются данные, относящиеся к группе подразделений, входящих в региональный центр. В центре накапливаются и обрабатываются данные, относящиеся к группе региональных центров. Число уровней хранения данных может быть и большим, а также равным двум. Но от этого суть процесса передачи и приема данных не меняется. В любом случае в подразделениях необходимо выполнить экспорт основных таблиц, а в центрах необходимо выполнить импорт этих таблиц.

При таком подходе неизбежно многократное поступление одних и тех же данных в центр из регионов. Кроме того, повторно передаваемые данные (записи) могут быть час-

тично изменены, например, в каком-либо поле или группе полей могут быть перед повторной пересылкой данных изменены значения. Это порождает одновременно необходимость отсеивания уже принятых данных и необходимость модификации данных в центрах, если они уже ранее были получены.

Постановка задачи объединения совместимых таблиц.

Даны отношения

$A=(A_1,\ldots, A_i,\ldots, A_k)$, $i=\overline{1,k}$, где k – число атрибутов отношения A;

$B=(B_1,\ldots, B_j,\ldots, B_k)$, $j=\overline{1,k}$, где k – число атрибутов отношения B.

$(\forall A_i)(\exists B_j)(A_i = B_j)$

$SA=\{SA_1,\ldots, SA_m,\ldots, SA_n\}$ $m=\overline{1,n}$, где n – число записей отношения A; SA_m – m – я запись A.

$SB=\{SB_1,\ldots, SB_p,\ldots, SB_q\}$ $p=\overline{1,q}$, где q – число записей отношения B; SB_p – p – я запись B.

Для объединения совместимых таблиц необходимо получить $S = SA \cup SB$.

Самым разумным способом исключения дублирования записей при импорте представляется использование ключевых полей. Действительно, при попытке добавления новой записи, у которой значение ключевого поля равно значению ключевого поля записи, содержащейся в таблице, эта запись добавлена не будет. Это обусловлено особенностями первичных ключей, которые должны иметь уникальные значения. Во всех системах управления базами данных (СУБД) предусмотрены средства, которые реализуют невозможность формирования или добавления записей с одинаковыми значениями первичных ключей.

В качестве первичного ключа очень часто используется поле типа ”Счетчик”. При формировании новой записи в таком поле автоматически формируется уникальное значение. Это очень удобно, т.к. пользователю БД не нужно заполнять это поле и, кроме того, автоматически обеспечивается одно из требований к реляционным таблицам – уникальность записей.

Но в рассматриваемом случае использование ключевых полей типа ”Счетчик” неприемлемо. Дело в том, что в этом случае записи таблиц различных регионов будут иметь одинаковые значения ключевых полей, хотя по сути записи не совпадают. И при импорте записей из нескольких регионов в центральную БД значительная часть записей не будет

импортирована. В связи с этим значения ключевых полей должны формироваться таким образом, чтобы они были уникальны не только в рамках таблиц региональных БД, но и были уникальны в рамках всех аналогичных таблиц всех регионов. Один из самых простых способов формирования первичного ключа – это использование в качестве его значения конкатенации характерного обозначения региона, кода продавца и номера записи.

Рассмотрим реальный пример: предприятие, занимающееся поставкой оборудования, имеет представительства в нескольких регионах. Каждый продавец региона ведет собственную БД, в которой ведет учет поставки оборудования Заказчикам. Региональный руководитель собирает данные от продавцов и пересылает их в Центр. Продавцы формируют коммерческие предложения, каждое из которых имеет уникальное имя. Имя состоит из кода продавца, кода региона, и порядкового номера. На рис. 1 приведен фрагмент списка коммерческих предложений одного из продавцов.

OffersID	SAM_ID	Offers Number	Prospection	Код_зак	Код_Продавец	Код_агента	ФИО агента
38082	2665-7191	2665K38082	2665P4249	0	2665	0	Жарва Алла Викторовна
38087	2665-7289	2665K38087	2665P4254	0	2665	0	
38088	2665-7290	2665K38088	2665P4255	0	2665	0	
38089	2665-7291	2665K38089	2665P4256	0	2665	0	Сидоров Борис Иванович
38090	2665-7292	2665K38090	2665P4257	0	2665	0	
38091	2665-7293	2665K38091	2665P4258	0	2665	0	
38092	2665-7294	2665K38092	2665P4259	0	2665	0	

Рис. 1. Фрагмент списка коммерческих предложений

Уникальное имя записи в рамках всех регионов обеспечивается посредством поля ”Offers Number”.

По аналогии формируются ключевые поля и других таблиц, пересылаемых в региональные центры.

Импортирование одной из таблиц в промежуточную БД, которая отсылается в региональный центр, осуществляется посредством макроса, фрагмент которого приведен на рис. 2.

В данном случае в качестве источника для экспорта используется таблица ”Заказчики”, а в качестве приемника ”Заказчики_ex”. БД для экспорта располагается в папке c:\actmng\export.

Подобным образом экспортируются все основные таблицы, для каждой из которых сформирована макрокоманда ПреобразоватьБазуДанных.

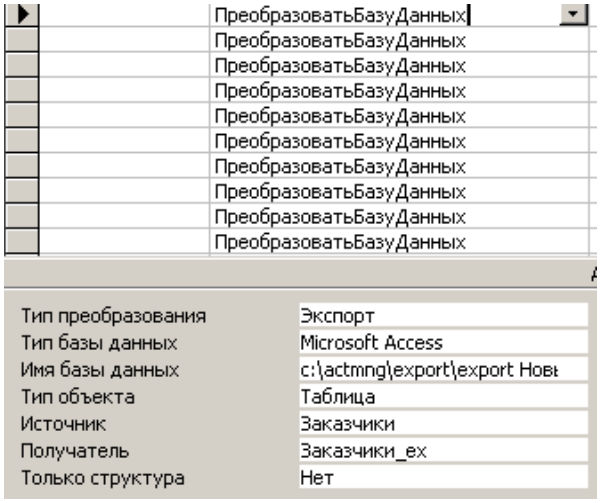


Рис. 2. Фрагмент макроса, обеспечивающего экспорт нескольких таблиц во вспомогательную БД

На рис. 3 приведен фрагмент содержимого таблицы, которая сформирована из записей, импортированных от двух продавцов. Как видно из рисунка, ключевое поле ”Offers Number” сформировано таким образом, что его содержимое в принципе не может повторяться у различных продавцов.

OffersID	SAM_ID	Offers Number	Prospection	Код_зак	Код_Продавец	Код_агента	ФИО агента
38082	2665-7191	2665K38082	2665P4249	0	2665	0	Жарва Алла Викторовна
38087	2665-7289	2665K38087	2665P4254	0	2665	0	
38088	2665-7290	2665K38088	2665P4255	0	2665	0	
38089	2665-7291	2665K38089	2665P4256	0	2665	0	Сидоров Борис Иванович
38090	2665-7292	2665K38090	2665P4257	0	2665	0	
38091	2665-7293	2665K38091	2665P4258	0	2665	0	
38092	2665-7294	2665K38092	2665P4259	0	2665	0	
38227	2662-7190	2662K38251	2662P4421	0	2662	0	
38228	2662-7411	2662K38252	2662P4484	7411	2662	0	Селянцев Константин Геор

Рис 3. Таблица с записями двух продавцов

Проиллюстрируем импорт данных, содержащих записи уже имеющихся в БД регионального центра. Эти записи наряду с новыми записями приведены на рис. 4.

OffersID	SAM_ID	Offers Number	Prospection	Код_зак	Код_Продавец	Код_агента	ФИО агента
38067	2662-7190	2662K38251	2662P4421	0	2662	0	
38230	2662-7411	2662K38252	2662P4484	7411	2662	0	Селянцев Константин Геор
38069	2665-7312	2665K38069	2665P4235	7312	2665	0	
38070	2665-7303	2665K38070	2665P4236	7304	2665	0	Сидоров Борис Иванович

Рис. 4. Данные, импортируемые в региональный центр

Запрос на добавление записей из временной таблицы в таблицу центра выглядит следующим образом:

INSERT INTO _КП SELECT [_КП_ex]. FROM _КП_ex;*

Здесь посредством конструкции *"SELECT [_КП_ex].* FROM _КП_ex"* выбираются значения всех полей временной таблицы *"_КП_ex",* и эти значения добавляются с помощью конструкции "*INSERT INTO _КП*" в таблицу БД центра с именем *"_КП"*. Следует отметить, что такое возможно в связи с тем, что с временной таблицей средствами Microsoft Access организована связь. Число подобных запросов соответствует числу импортируемых таблиц.

После выполнения запроса на импорт этих данных в БД регионального центра таблица с данными в центре примет вид рис. 5.

OffersID	SAM_ID	Offers Number	Prospection	Код_зак	Код_Продавец	Код_агента	ФИО агента
38082	2665-7191	2665K38082	2665P4249	0	2665	0	Жарва Алла Викторовна
38087	2665-7289	2665K38087	2665P4254	0	2665	0	
38088	2665-7290	2665K38088	2665P4255	0	2665	0	
38089	2665-7291	2665K38089	2665P4256	0	2665	0	Сидоров Борис Иванович
38090	2665-7292	2665K38090	2665P4257	0	2665	0	
38091	2665-7293	2665K38091	2665P4258	0	2665	0	
38092	2665-7294	2665K38092	2665P4259	0	2665	0	
38227	2662-7190	2662K38251	2662P4421	0	2662	0	
38228	2662-7411	2662K38252	2662P4484	7411	2662	0	Селянцев Константин Геор
38229	2665-7312	2665K38069	2665P4235	7312	2665	0	
38230	2665-7303	2665K38070	2665P4236	7304	2665	0	Сидоров Борис Иванович

Рис. 5. Таблица БД регионального центра после выполнения импорта данных продавца

Как видно из рисунка, в таблицу центра добавились только новые записи. Таким образом, в рассматриваемом примере при импорте данных обеспечивается фильтрация ранее импортированных данных. Это произошло благодаря тому, что поле "Offers Number" является ключевым, и записи с повторяющимися значениями ключевого поля не добавляются.

Кроме рассмотренной проблемы необходимо решить проблемы обновления записей таблиц региональных центров в том случае, если соответствующие записи были обновлены у продавцов.

Так как в региональном центре могут храниться записи нескольких десятков продавцов, то в средствах обновления должна быть предусмотрена возможность обновления записей непосредственно после выполнения их импорта. Кроме того, должна быть преду-

смотрена возможность обновления данных только того продавца, от которого импортированы данные. Так как состав полей, которые продавец может обновлять регламентирован, то в запросе на обновление нет необходимости задействовать все 250 полей таблицы. Тем более, что запрос на обновления большого количества полей в таблицах реального объема, будет выполняться долго (до нескольких минут).

На рис. 6 приведен бланк запроса на обновление полей тех записей, которые уже находятся в БД регионального центра.

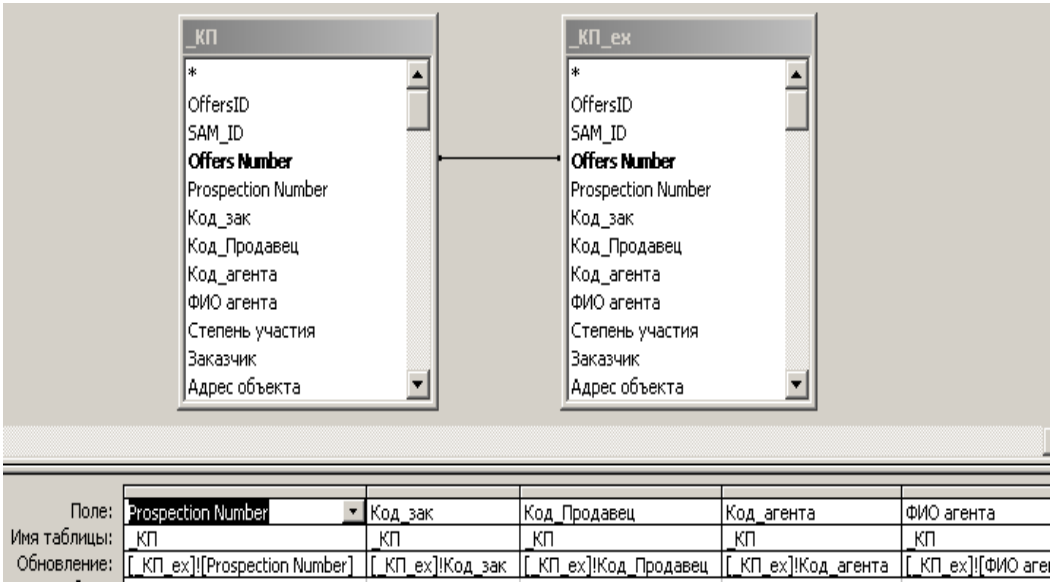


Рис. 6. Запрос на обновление записей, находящихся в БД регионального центра

Из рис. 6 видно, что обновляются только те записи, которые содержатся в БД регионального центра (таблица “*_КП*”) и которые, кроме того, импортируются от продавца (таблица “*_КП_ex*”). Это обеспечивается посредством сформированной связи ”один к одному” между этими таблицами по ключевым полям ”Offers Number”. Из рисунка также видно, что данные для обновления берутся из полей временной таблицы ”*_КП_ex*” (строка бланка запроса ”Обновление”). Таким образом, в ходе выполнения запроса перебираются пары одноименных записей из двух таблиц (записей с одинаковыми значениями ключевых полей). Одноименные записи таблицы ”*_КП*” обновляются посредством записей таблицы ”*_КП_ex*”. В связи с этим для каждой записи таблицы ”*_КП*” выбирается соответствующая запись таблицы ”*_КП_ex*”, а затем вместо старых значений полей таблицы ”*_КП*” заносятся соответствующие значения полей таблицы ”*_КП_ex*”. Важно отметить, что, несмотря на использование в рассмотренных таблицах более 200 полей, обновляются только 30 характерных полей.

Проиллюстрируем работоспособность предложенного способа обновления полей на примере. На рис. 6 представлены записи ”*_КІТ*”, находящейся в БД регионального центра.

На рис. 7 приведено содержимое таблицы продавца, которая экспортируется в БД регионального центра.

OffersID	SAM_ID	Offers Number	Prospection	Код_зак	Код_Продавец	Код_агента	ФИО агента
38067	2662-7190	2662K38251	2662P4421	111	2662	0	Селянцев Константин Гео
38230	2662-7411	2662K38252	2662P4484	222	2662	0	Селянцев Константин Гео
38069	2665-7312	2665K38069	2665P4235	333	2665	0	Сидоров Борис Иванович
38070	2665-7303	2665K38070	2665P4236	444	2665	0	Сидоров Борис Иванович

Рис. 7. Содержимое таблицы продавца, которая экспортируется в БД регионального центра

Как видно из рисунка, импортируются те же самые записи, которые уже были импортированы ранее. В связи с этим ожидается, что новых записей в БД центра не появится. Однако значения полей ”*Код_зак*” и ”*ФИО агента*” претерпели обновление, поэтому это обновление должно быть отражено в БД центра после импорта данной таблицы.

После выполнения импорта данной таблицы и выполнения запроса на обновление таблица БД регионального центра примет вид рис. 8.

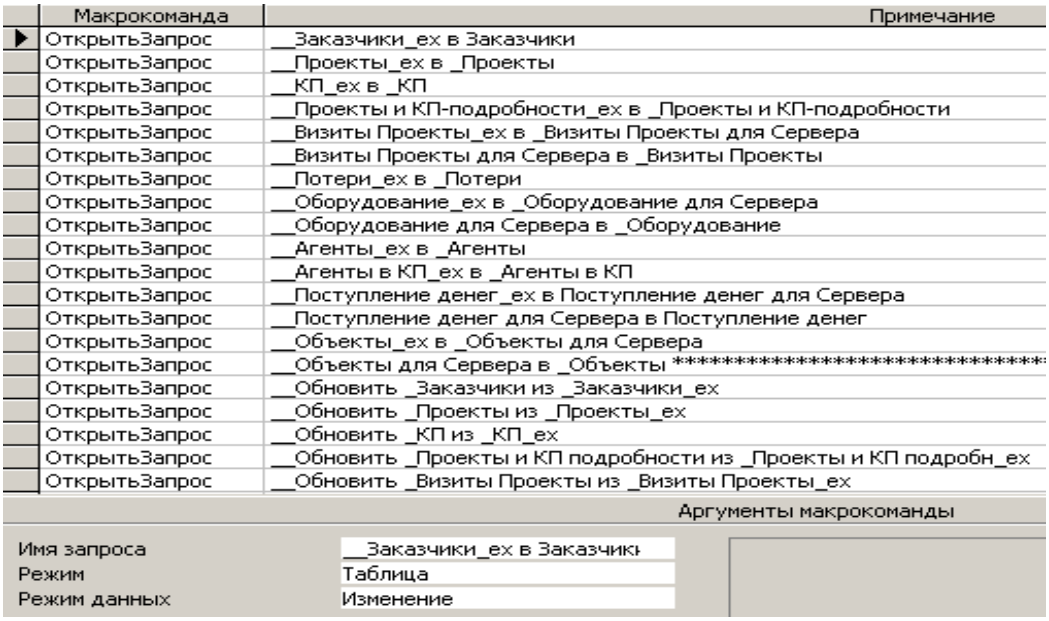
OffersID	SAM_ID	Offers Number	Prospection	Код_зак	Код_Продавец	Код_агента	ФИО агента
38082	2665-7191	2665K38082	2665P4249	0	2665	0	Жарва Алла Викторовна
38087	2665-7289	2665K38087	2665P4254	0	2665	0	
38088	2665-7290	2665K38088	2665P4255	0	2665	0	
38089	2665-7291	2665K38089	2665P4256	0	2665	0	Сидоров Борис Иванови
38090	2665-7292	2665K38090	2665P4257	0	2665	0	
38091	2665-7293	2665K38091	2665P4258	0	2665	0	
38092	2665-7294	2665K38092	2665P4259	0	2665	0	
38227	2662-7190	2662K38251	2662P4421	111	2662	0	Селянцев Константин Ге
38228	2662-7411	2662K38252	2662P4484	222	2662	0	Селянцев Константин Ге
38229	2665-7312	2665K38069	2665P4235	333	2665	0	Сидоров Борис Иванови
38230	2665-7303	2665K38070	2665P4236	444	2665	0	Сидоров Борис Иванови

Рис. 8. Содержимое таблицы регионального центра после импорта и обновления записей

Как видно из рисунка, одноименные записи в таблицу не добавились, что и ожидалось от импорта. А поля таблицы обновились значениями полей одноименных записей импортируемой таблицы.

Таким образом, предложенный подход импорта и обновления полей успешно работает.

Так как импортируется и обновляется несколько таблиц, оправданно объединить команды импортирования и обновления всех таблиц в рамках одной процедуры или в рамках одного макроса, как это реализовано на рис. 9.



Макрокоманда	Примечание
ОткрытьЗапрос	__Заказчики_ex в Заказчики
ОткрытьЗапрос	__Проекты_ex в _Проекты
ОткрытьЗапрос	__КП_ex в _КП
ОткрытьЗапрос	__Проекты и КП-подробности_ex в _Проекты и КП-подробности
ОткрытьЗапрос	__Визиты Проекты_ex в _Визиты Проекты для Сервера
ОткрытьЗапрос	__Визиты Проекты для Сервера в _Визиты Проекты
ОткрытьЗапрос	__Потери_ex в _Потери
ОткрытьЗапрос	__Оборудование_ex в _Оборудование для Сервера
ОткрытьЗапрос	__Оборудование для Сервера в _Оборудование
ОткрытьЗапрос	__Агенты_ex в _Агенты
ОткрытьЗапрос	__Агенты в КП_ex в _Агенты в КП
ОткрытьЗапрос	__Поступление денег_ex в Поступление денег для Сервера
ОткрытьЗапрос	__Поступление денег для Сервера в Поступление денег
ОткрытьЗапрос	__Объекты_ex в _Объекты для Сервера
ОткрытьЗапрос	__Объекты для Сервера в _Объекты **********
ОткрытьЗапрос	__Обновить _Заказчики из _Заказчики_ex
ОткрытьЗапрос	__Обновить _Проекты из _Проекты_ex
ОткрытьЗапрос	__Обновить _КП из _КП_ex
ОткрытьЗапрос	__Обновить _Проекты и КП подробности из _Проекты и КП подробн_ex
ОткрытьЗапрос	__Обновить _Визиты Проекты из _Визиты Проекты_ex

Аргументы макрокоманды

Имя запроса	__Заказчики_ex в Заказчики
Режим	Таблица
Режим данных	Изменение

Рис. 9. Фрагмент макроса для импорта и обновления таблицы

Список литературы.

1. Брешенков А.В., Бондарь М.Л. Построение системы репликации баз данных на основе протоколирования изменений (алгоритмы синхронизации и односторонней репликации) // Сб. трудов каф., посвященный 175-летию МГТУ им. Н.Э. Баумана. - М.: Эликс +, 2004. - С. 31-37.
2. Брешенков А.В., Мин Тхет Тин. Анализ проблем разработки методики формирования реляционных таблиц на основе использования информации табличного типа. Россия в XXI веке: проблемы, тенденции, перспективы // Материалы XIV Международного симпозиума "Уникальные феномены и универсальные ценности культуры", сборник научных статей / под общей ред. И.А. Архарова и И.Е. Моториной - М.: МГТУ им. Н.Э. Баумана, 2012 - С. 270-272.
3. Брешенков А.В., Белоус В. В. Приведение значений атрибутов существующей информации табличного вида к одному типу. Наука и образование. Инженерное образование: Эл. науч. издание. – 2010. (Номер гос. регистрации 0321000194)
4.