

## СПИСОК ЛИТЕРАТУРЫ

1. Карауш А.С. Модель тиражирования библиографических баз данных с использованием алгоритмических кодов записей // «EL-PUB2003»: Сб. тезисов и докл. VIII Междунар. конф. по электронным публикациям. – Новосибирск, 2003. – С. 14–15.
2. Карауш А.С., Копытков Д.Ю. Программное обеспечение для автоматической синхронизации баз данных системы «ИР-БИС» // Научные и технические библиотеки. – 2003. – № 10. – С. 88–91.
3. ГОСТ 7.14-98. Формат для обмена информацией. Структура записи. – Взамен ГОСТ 7.14-84; Введ. 01.01.99. – М.: Изд-во стандартов, 1998. – 4 с.
4. ГОСТ 7.19-85. Коммуникативный формат для обмена библиографическими данными на магнитной ленте. Содержание записи. – Взамен ГОСТ 7.19-79; Введ. 01.01.86. – М.: Изд-во стандартов, 1985. – 102 с.

УДК 004.75

## МОДЕЛЬ МЕЖКОРПОРАТИВНОГО ТИРАЖИРОВАНИЯ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМИЧЕСКИХ КОДОВ БЕЗ ЕДИНОГО СИНХРОНИЗИРУЮЩЕГО ВРЕМЕННОГО ИНТЕРВАЛА

Д.Ю. Копытков

Томский университет систем управления и радиоэлектроники

E-mail: piscetes@inbox.ru

*Рассмотрены проблемы использования модели межкорпоративного тиражирования данных. Предложена модель асинхронного тиражирования баз данных с использованием алгоритмических кодов и без единого синхронизирующего временного интервала, в течение которого работа с локальными данными запрещена.*

Тиражирование – технология, используемая в системах распределенных баз данных (БД), которая предусматривает поддержку копий некоторых фрагментов БД в нескольких узлах сети с целью приближения данных к месту их использования и сокращения тем самым сетевого трафика и/или повышения производительности системы. Механизм тиражирования очень важен, поскольку позволяет организации обеспечивать доступ пользователям к актуальным данным, когда они в этом нуждаются. Использование тиражирования позволяет достичь многих преимуществ, включая повышение производительности (в тех случаях, когда централизованный ресурс оказывается перегруженным), надежность хранения и доступность данных, наличие «горячей» резервной копии на случай восстановления, а также возможность поддержки мобильных пользователей и хранилищ данных.

Одним из механизмов тиражирования данных является асинхронный [1]. Данный механизм предусматривает обновление локальных БД системы после обновления исходной базы данных. Задержка в восстановлении согласованности данных может варьироваться от нескольких секунд до нескольких часов или даже дней. Однако, рано или поздно данные во всех копиях будут приведены в синхронное состояние.

Особенностью моделей асинхронного тиражирования является наличие периода времени, когда изменения в базе запрещены, что осложняет применение асинхронных моделей для участников, находящихся в разных часовых поясах.

В данной статье предложена модель асинхронного тиражирования данных без единого синхронизирующего интервала с применением алгоритма идентификации данных на основе алгоритмических кодов (алкодов).

Начальные данные и условные обозначения, применяемые далее при описании новой модели:  $D_n$  – БД  $n$ -го участника тиражирования;  $D_{n-t_1}$  – состояние БД  $n$ -го участника тиражирования в момент времени  $t_1$ ;  $T_{n-p}$ ,  $T_{n-b}$  – интервал времени, в течение которого  $n$ -ый участник тиражирования осуществляет или не осуществляет изменения в базе данных  $D_n$ ;  $\Delta_{n-(t_2-t_1)}$  – состояния записей, которые были подвергнуты изменениям в  $D_n$   $n$ -го участника тиражирования за время  $t_2-t_1$ ,  $\Delta_{x-y}$  – сделанные изменения.

Определим операции над состояниями баз данных:

1. Вычитание:  $D_x - D_y = \Delta_{x-y}$ .
2. Пересечение:  $D_x \cap D_y = REC_{x-y}$ , где  $REC_{x-y}$  – состояния записей, которые присутствуют в  $D_x$  и отсутствуют в  $D_y$ .
3. Суммирование:  $D_y + \Delta_{x-y} = D_x$ . Порядок слагаемых имеет значение, т. е.  $D_y + \Delta_{x-y} \neq \Delta_{x-y} + D_y$ . В случае  $D_y + \Delta_{x-y}$  состояния записей в  $D_y$  будут переписаны соответствующими состояниями записей, содержащихся в  $\Delta_{x-y}$ , а в случае  $\Delta_{x-y} + D_y$  – наоборот.

На рис. 1 схематически представлено время работы участника тиражирования с локальной БД, где в интервал времени  $t_2-t_1$  участник тиражирования совершает изменения в локальной БД, а в ин-

тервале  $t_3-t_2$  работа с БД запрещена. Таким образом, состояния баз данных в отрезках времени  $t_3$  и  $t_2$  равны, т. е.  $D_3=D_2$ . Сделанные изменения за отрезок времени  $t_2-t_1$  представлены в виде дельты  $\Delta_{t_2-t_1}$  равной разнице состояний  $D_2$  и  $D_1$ .

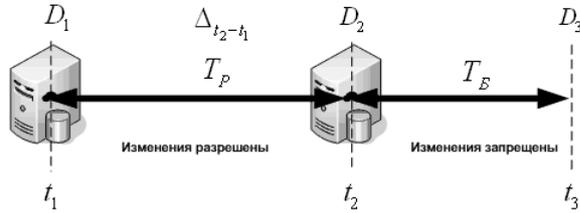


Рис. 1. Схема внесения изменений в  $D_n$  в течение дня

Рассмотрим работу модели на примере двух участников тиражирования (рис. 2). Каждый участник тиражирования выполняет следующий цикл действий:

1. редактирует данные локальной БД в рабочие интервалы времени  $T_{n-p}$ ;
2. создает дельты за интервал  $T_{n-p}$ ;
3. получает опубликованные дельты других участников тиражирования;
4. строит суммарную дельту на основе полученных опубликованных дельт других участников тиражирования и состояния записей локальной БД;

5. применяет суммарную дельту на локальную БД;
6. публикует суммарную дельту.

Необходимо отметить, что при построении суммарной дельты, участники тиражирования для идентификации данных, хранящихся в дельтах и локальных БД, используют алгоритмические коды (алкоды) [2, 3]. Использование алкодов в данной асинхронной модели необходимо вследствие того, что данные изменяются локально у каждого участника тиражирования, и до момента тиражирования ничего не известно о сделанных изменениях. Данные идентификаторы записей позволяют однозначно идентифицировать записи в дельте созданных другими участниками тиражирования и правильно соотнести к локальным записям участника, применяющего полученную дельту на БД.

Поэтапные действия каждого участника с учетом математических выкладок представлены в таблице.

Докажем, что состояния баз данных в шаге № 11 равны:

$$\begin{aligned}
 D_{1-6} = D_{2-6} &\rightarrow D_{1-5} + \Sigma_4 = D_{2-5} + \Sigma_4 \rightarrow \\
 \rightarrow D_{1-2} + \Sigma_3 + \Sigma_4 &= D_{2-1} + \Sigma_1 + \Sigma_3 + \Sigma_4 \rightarrow \\
 \rightarrow D_{1-2} + \Sigma_3 + \Sigma_4 &= D_{2-1} + \Sigma_1 + \Sigma_3 + \Sigma_4 \rightarrow \\
 \rightarrow D_{1-1} + \Sigma_3 \cap \Sigma_4 + \Sigma_4 &= D_{2-1} + \Sigma_4 \rightarrow \\
 \rightarrow D_{1-1} + \Sigma_4 &= D_{2-1} + \Sigma_4.
 \end{aligned}$$

Таблица. Последовательность действий участников тиражирования

№	T	Действия участника № 1 с	Действия участника № 2 с
1	$t_1$	$D_{1-1}=D_{2-1}$ .	$D_{1-1}=D_{2-1}$ .
2	$t_2-t_1$	$D_{1-2}=D_{1-1}$ .	Изменение состояния $D_{2-1}$ .
3	$t_3-t_2$	Изменение состояния $D_{1-2}$ .	$D_{2-2}-D_{2-1}=\Delta_{2-(t_2-t_1)}$ .
4	$t_4-t_3$	Изменение состояния $D_{2-1}$ .	1) Получение опубликованных дельт за интервал времени $t_4-t_3$ ; 2) $D_{2-3}=D_{2-2}+\Sigma(D_{2-2}\cap\Delta_{2-(t_3-t_2)};\Delta_{2-(t_3-t_2)})$ ; 3) $\Sigma(D_{2-2}\cap\Delta_{2-(t_3-t_2)};\Delta_{2-(t_3-t_2)})=\Sigma_1$ ; 4) опубликование суммарной дельты $\Sigma_1$ .
5	$t_5-t_4$	$D_{1-3}-D_{1-2}=\Delta_{1-(t_5-t_4)}$ .	Изменение состояния $D_{2-3}$ .
6	$t_6-t_5$	1) Получение опубликованных дельт за интервал времени $t_6-t_5$ ; 2) $D_{1-4}=D_{1-3}+\Sigma(D_{1-3}\cap\Sigma_1;\Delta_{1-(t_6-t_5)};\Sigma_1)$ ; 3) $\Sigma(D_{1-3}\cap\Sigma_1;\Delta_{1-(t_6-t_5)};\Sigma_1)=\Sigma_2$ ; 4) опубликование суммарной дельты $\Sigma_1$ .	Изменение состояния $D_{2-3}$ .
7	$t_7-t_6$	Не редактирование данных $D_{1-4}$ .	$D_{2-4}-D_{2-3}=\Delta_{2-(t_7-t_6)}$ .
8	$t_8-t_7$	Не редактирование данных $D_{1-4}$ .	1) Получение опубликованных дельт за интервал времени $t_8-t_7$ ; 2) $D_{2-5}=D_{2-4}+\Sigma(D_{2-4}\cap\Sigma_2;\Delta_{2-(t_8-t_7)};\Sigma_2)$ ; 3) $\Sigma(D_{2-4}\cap\Sigma_2;\Delta_{2-(t_8-t_7)};\Sigma_2)=\Sigma_3$ ; 4) опубликование суммарной дельты $\Sigma_3$ .
9	$t_9-t_8$	Не редактирование данных $D_{1-4}$ .	Не редактирование данных $D_{2-5}$ .
10	$t_{10}-t_9$	1) Получение опубликованных дельт за интервал времени $t_{10}-t_9$ ; 2) $D_{1-6}=D_{1-5}+\Sigma(D_{1-5}\cap\Sigma_3;\Sigma_3)$ ; 3) $\Sigma(D_{1-5}\cap\Sigma_3;\Sigma_3)=\Sigma_4$ ; 4) опубликование суммарной дельты $\Sigma_4$ .	Не редактирование данных $D_{2-5}$ .
11	$t_{11}-t_{10}$	Не редактирование данных $D_{1-6}$ .	1) Получение опубликованных дельт за интервал $t_{11}-t_{10}$ ; 2) $D_{2-6}=D_{2-5}+\Sigma(D_{2-5}\cap\Sigma_4;\Sigma_4)$ ; 3) $\Sigma(D_{2-5}\cap\Sigma_4;\Sigma_4)=\Sigma_4$ ; 4) опубликование суммарной дельты $\Sigma_4$ .

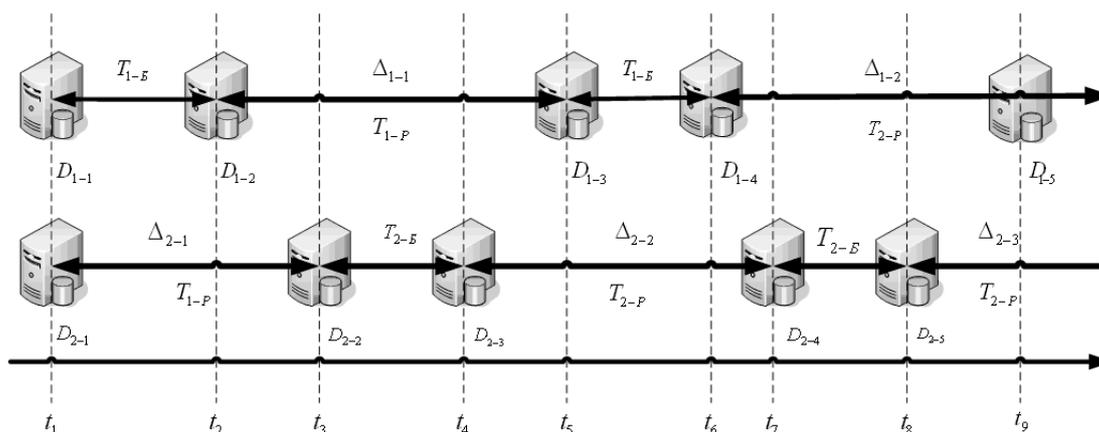


Рис. 2. Работа модели тиражирования на примере двух участников

Таким образом, видно, что полная синхронизация данных [1] двух участников будет завершена после нередктирования данных каждым из участников в рабочий интервал. Следовательно, период синхронизации БД участников можно представить в виде следующей формулы:

$$T_{\text{синхр.}} = (\max(T_{1-P}, \dots, T_{n-P}) + \max(T_{1-B}, \dots, T_{n-B})) \times (N+1), \quad (1)$$

где  $T_{\text{синхр.}}$  – период синхронизации данных;  $\max(T_{1-P}, \dots, T_{n-P})$  – функция максимума от аргументов;  $N$  – количество интервалов  $T_{x-P}$ ,  $T_{y-B}$ .

Из (1) следует, что  $T_{\text{синхр.}} \rightarrow 0$ , при  $\max(T_{1-P}, \dots, T_{n-P}) \rightarrow 0$  и  $\max(T_{1-B}, \dots, T_{n-B}) \rightarrow 0$  одновременно. Т. е. весь цикл участника тиражирования должен выполняться незамедлительно при каждом изменении состояния локальной БД каждого участника.

Таким образом, синхронные модели тиражирования данных являются частными случаями данной асинхронной модели, которые выполняют условия  $\max(T_{1-P}, \dots, T_{n-P}) \rightarrow 0$  и  $\max(T_{1-B}, \dots, T_{n-B}) \rightarrow 0$  в виде распределенных транзакций.

#### Заключение

Предложена модель асинхронного тиражирования баз данных с использованием алгоритмических кодов и без единого синхронизирующего временного интервала, в течение которого работа с локальными данными запрещена. Предлагаемая модель позволяет синхронизировать данные между межкорпоративными участниками тиражирования без согласования единого интервала времени и единства типов баз данных. Полная синхронизация данных между всеми участниками тиражирования будет завершена после нередктирования данных каждым из участников в рабочий интервал времени.

#### СПИСОК ЛИТЕРАТУРЫ

1. Когаловский М.Р. Энциклопедия технологий баз данных. – М.: Финансы и статистика, 2002. – 800 с.
2. Карауш А.С., Копытков Д.Ю. Программное обеспечение для автоматической синхронизации баз данных системы «ИР-

БИС» // Научные и технические библиотеки. – 2003. – № 10. – С. 88–91.

3. Карауш А.С. Модель тиражирования библиографических баз данных с использованием алгоритмических кодов записей // «EL-PUB2003»: Сб. тезисов и докл. VIII Междунар. конф. по электронным публикациям. – Новосибирск, 2003. – С. 14–15.