

П.Е. Котиков¹
А.А. Нечай²

P.E. Kotikov
A.A. Neschay

РЕПЛИКАЦИЯ ДАННЫХ МЕЖДУ СЕРВЕРАМИ БАЗ ДАННЫХ В СРЕДЕ ГЕОИНФОРМАЦИОННЫХ СИСТЕМ

DATA REPLICATION BETWEEN DATABASE SERVERS IN THE ENVIRONMENT OF GEOINFORMATION SYSTEMS

Данная статья посвящена проблеме организации репликации данных между серверами баз данных в распределенных гетерогенных базах данных геоинформационных систем. Показаны некоторые перспективные решения.

Ключевые слова: геоинформационная система, гетерогенные базы данных, репликация.

This article deals with the problem of data replication between database servers in distributed heterogeneous databases of geographic information systems. Some promising solutions are shown.

Keywords: geographic information system, heterogeneous database, replication.

Распределенные базы данных известны сложностью организации репликации в них. Вместе с тем, многие предлагаемые решения для геоинформационных систем (ГИС) основаны именно на их использовании. Как убедительно показано в ряде работ [2–4], такой подход дает ряд существенных преимуществ. Во всех случаях репликация необходима. Несомненно, без использования репликаций в распределенных базах данных невозможно обеспечить ни целостности, ни непротиворечивости данных. Таким образом, речь идет об основе безопасности и защиты данных.

Характерно мнение [1]: «За репликацией будущее. Сейчас улучшились каналы, и на время ее актуальность отпала. Но необходимость репликации, проблемы ее осуществления будут всегда. Будут математические решения, технические проекты, новые методы организационных решений...».

Известны виды репликационных архитектур:

- единая база данных с удаленным доступом на рабочих местах;
- единая база данных с обновляемыми копиями на других серверах;
- распределенная база данных.

¹ Кандидат технических наук, доцент Военно-космической академии им. А.Ф. Можайского.

² Преподаватель Военно-космической академии им. А.Ф. Можайского.

В геоинформационных системах предполагается применение, в основном, распределенных решений. Распределенная база данных (БД) является наиболее сложным видом архитектуры с обновляемыми БД. В этом виде вся БД разделена на секторы, каждый из которых расположен на удаленном сервере и таблицы которого реплицируются на другие серверы (то есть, общая БД состоит из таблиц удаленных баз данных). Такая архитектура удобна возможностью существования разнородных БД и низкой загрузкой сети. Неудобство – в установке и дальнейшем развитии системы.

Репликация позволяет распространять многочисленные копии одной и той же информации по различным базам данных в рамках локальной сети использующей организации. Известно, что репликация имеет следующие преимущества:

- конечные данные приближаются к пользователям;
- снижается влияние сред OLAP (Online Analytical Processing – аналитическая обработка в реальном времени) на производительность системы;
- повышается использование служб OLTP (Online Transaction Processing – оперативная обработка транзакций);
- уменьшается вероятность конфликтов между различными серверами, работающими с одними и теми же данными;

– серверы баз данных управляют реплицированными данными автономно, то есть можно создавать собственные правила, хранимые процедуры и представления, обрабатывающие свои копии данных.

Всё это чрезвычайно существенно для ГИС.

Возможны два способа репликации: собственно репликация и распространение транзакций. Оба способа репликации позволяют поддерживать различные копии данных свежими. В одной среде возможно также одновременное использование обоих методов. Значительный интерес представляет применение метода распространения транзакций. В чем его особенности?

Распространенные транзакции исключают автономность, не допускают задержек, не гарантируют транзакционной целостности базы данных. Каждый подписчик обязательно должен иметь постоянное соединение с издателем и всеми другими подписчиками, поэтому автономность подписчикам не требуется, а любые задержки обновления весьма нежелательны. Как правило, этот тип репликации используется в том случае, когда каждый подписчик обрабатывает версию базы данных в реальном времени.

Преимущества этого типа репликации уже используются в системах резервирования данных большого объема. Расположенные на удаленных серверах подписчики могут переложить часть работы на сервер публикации. Транзакционная целостность при этом гарантируется.

Как видим, есть и недостатки и явные преимущества. Как это работает?

Процесс репликации управляется пятью различными агентами. Каждый агент решает отдельную задачу. В репликации участвуют следующие агенты.

1. *Агент рассылки* передает информацию из рассылаемой базы данных подписчикам.

2. *Агент чтения протокола* управляет журналом транзакций всех публикуемых баз данных. Агент использует журнал в процессе репликации. Когда агент чтения протокола находит транзакции, являющиеся частью публикации, он копирует их в рассылаемую базу данных, откуда агент рассылки отправляет их подписчикам.

3. *Агент слияния* объединяет изменения, сделанные на различных серверах.

4. *Агент снимков* передает снимки данных перед началом репликации. Необходимость передачи обусловлена тем, что подписчик, у которого отсутствуют снимки обновленных данных, не может выполнить транзакции. Агент снимков используется в различных типах репликации снимком.

5. *Агент чтения очереди* читает из очереди SQL Server сообщения каждого подписчика и посылает транзакции подписчикам для выполнения.

Возможны варианты – репликация слиянием и репликация снимком.

При репликации слиянием агент слияния выполняется как на сервере рассылки, так и на сервере каждого подписчика. Если данные рассылаются принудительно, то агент слияния находится на сервере рассылки, а если запрашиваются, то – на серверах подписчиков. Процесс репликации состоит из следующих этапов.

1. Агент снимков (находящийся на сервере рассылки) создает исходный снимок данных и посылает его подписчикам. Как указывалось ранее, перед началом репликации (за исключением репликации снимком) подписчики согласуют данные с издателями.

2. Для проведения репликации слиянием на сервере рассылки создается рабочая папка рассылки.

3. Начинается репликация.

4. Агент слияния запрашивает у подписчиков изменения и отправляет их на сервер публикации.

5. Агент слияния получает сообщения о конфликтах данных, возникших при обновлении данных на сервере публикации, и выполняет соответствующие действия.

Чтобы репликация слиянием проходила нормально, в структуру таблиц и базы данных рассылки должны быть внесены изменения, которые позволяют серверу правильно реагировать на конфликты данных.

1. В рабочую папку рассылки добавляются системные таблицы. Они отслеживают изменения, используемые при согласовании и разрешении конфликтов данных.

2. На серверах публикации и подписки сервер создает триггеры, участвующие в репликации слиянием. Эти триггеры срабатывают, когда в одной из реплицируемых таблиц происходят изменения. Информация об изменениях записывается в системные таблицы, добавленные в рабочую папку рассылки. Сохраненные изменения позволяют отслеживать изменения каждой строки или столбца всех изменяемых таблиц.

3. Для каждой строки реплицируемых таблиц сервер создает новый столбец. Затем для идентификации каждой строки в это поле записывается глобальный уникальный идентификатор или глобальный уникальный идентификатор строки. Это делается для того, чтобы при обновлении данных в различных местах базы данных эти обновления можно было различить.

Репликация слиянием лучше всего подходит для случаев, когда вероятность изменений одной и той же записи в разных местах невелика. В репликации слиянием проводится горизонтальное выделение таблиц. Если позволить обновлениям одних и тех же данных в разных местах, то будут происходить частые конфликты данных. Ясно, что такое решение в ГИС мало кого устраивает.

В репликации снимком копия всей статьи или публикации передается сервером публикации подписчику. При репликации снимком с обновляющимися подписчиками обновления выполняются как на сервере публикации, так и у подписчика, однако при согласовании поступившая реплицируемая статья полностью накладывается на данные подписчика, заменяя несоответствия. Агент слияния не участвует в репликации снимком, однако в нем участвует агент рассылки. Если осуществляется подписка по запросу, то агент рассылки находится на сервере подписчика, а если принудительная подписка, то – на сервере рассылки. Выбор типа подписки (по запросу или принудительная) зависит от многих факторов, в первую очередь – от загруженности сервера рассылки и от желаемого способа управления подписками. Если нужно, чтобы подписки управлялись централизованно, то выбирается принудительная подписка с рассылкой с сервера рассылки, а если предпочтительнее передать управление подписчикам, то – по запросу с рассылкой через агента рассылки на сервере подписчика.

В транзакционной репликации в базу данных рассылки копируются только транзакции, созданные в публикуемой базе данных. Затем обновленные данные сразу же после изменения передаются базам данных подписчиков, что существенно уменьшает время задержки. База данных подписчиков рассматривается как «только для чтения», поскольку такой тип репликации является односторонним. Данные изменяются только сервером публикации.

В транзакционной репликации не используется агент слияния. Вместе с тем, должна существовать некоторая основа, к которой применяются транзакции. Как и в репликации снимком, в данном случае используется агент рассылки. Если реализуется подписка по запросу, то агенты рассылки находятся на серверах подписки, а если принудительная подписка, то – на сервере рассылки. Чтобы во всех местах в одно и то же время обновлялись одни и те же данные, следует использовать именно распространенные транзакции. Это исключает возможность двойного присоединения (закрепления) данных.

Чтобы база данных сервера могла быть опубликована в Internet, необходимо несколько изменить конфигурацию системы.

1. Компьютеры с агентом рассылки и агентом слияния, а также все соединенные с ними компьютеры должны связываться посредством соединений TCP/IP.

2. Серверы публикации и рассылки должны иметь прямое сетевое соединение с Internet. Это требование обусловлено не только соображениями безопасности, но и необходимостью решения проблем с временем задержки.

В целом, и с Интернетом всё возможно.

Отдельное место занимает вопрос о применении синхронной репликации. Синхронная репликация рассматривается как политика тиражирования данных в распределенных базах данных, предусматривающая модификацию всех копий измененных транзакцией данных до того момента, когда будет произведена фиксация этой транзакции. Поддержка синхронной симметричной репликации явилась новой возможностью еще в версии Oracle 7.0. Тем более, что это возможно и в современных версиях. В приложениях, где необходимо иметь несколько обновляемых копий одной таблицы в разных базах данных и где целостность данных имеет первостепенную важность, все копии нужно изменять в масштабах транзакции. Синхронная симметричная репликация позволяет достичь этой цели, хотя при значительном объеме обновлений при этом рост нагрузки на сеть и сервер могут быть существенными. Несмотря на это, синхронная симметричная репликация – всё же наилучшее средство, позволяющее осуществлять эффективную выборку последних данных на любом из узлов. Она обеспечивает абсолютную надежность и более высокую готовность, чем использование одной главной таблицы, потому что для запросов необходимо наличие только локальной копии. Производительность обработки запросов оптимальна, так как все запросы можно выполнять локально.

Синхронная репликация порождает целый ряд сложностей, связанных с проблемами возможной недоступности одной или нескольких баз данных или недоступностью части сети. Но это отдельный вопрос.

Таким образом, учитывая особенности работы с данными в ГИС, несомненно перспективным видится применение транзакционной репликации в интересах обеспечения высокой надежности и высокой степени уверенности в отсутствие двойного закрепления за потребителями. Наилучшим средством эффективной выборки данных на любом из узлов может быть

синхронная симметричная репликация, обеспечивающая также полную надежность и высокую готовность.

Вместе с тем, сопоставляя все рассмотренные обстоятельства построения репликации баз данных в гетерогенной среде ГИС, важно отметить, что ключевое значение для правильного решения проектанта может иметь фрагментационный анализ, который и позволит решить, как разбить или реплицировать данные между узлами.

Литература

1. Гаршин И.К. Практика работы с Oracle: генерация, администрирование, репликация. – М. : ПОЛТЕКС, 2000. – 250 с.
2. Колбина О.Н. Современные и теоретические аспекты управления распределёнными базами данных / О.Н. Колбина, Е.П. Истомин // Информационные технологии и системы: управление, экономика, транспорт // Право : сб. науч. тр. – Вып. 1(9). – СПб. : ООО «Андреевский издательский дом», 2011.
3. Колбина О.Н. Применение распределённых баз данных в геоинформационных системах прогнозирования георисков / О.Н. Колбина, Е.П. Истомин, Е.М. Зоринова // Сборник трудов Международной научно-практической конференции «Инфогео-2013». – СПб. : ООО «Андреевский издательский дом», 2013.
4. Разработка и развитие методов, моделей и систем геоинформационного управления пространственно-распределёнными объектами : отчет о НИР / Е.П. Истомин, А.Г. Соколов, О.Н. Колбина. – СПб. : Российский государственный гидрометеорологический университет, 2013. – 102 с.
5. Разработка автоматизированных систем обработки информации и управления (АСО-ИУ). – Режим доступа: <http://3ys.ru/razrabotka-avtomatizirovannykh-sistem-obrabotki-informatsii-i-upravleniya-asoiu.html> (дата обращения: 25.01.15).
6. Лохвицкий В.А. Подход к построению системы автоматизированной интеграции информации в базу данных для её своевременной актуализации / В.А. Лохвицкий, С.В. Калиниченко, А.А. Нечай // Мир современной науки. – 2014. – № 2 (24). – С. 8–12.
7. Нечай А.А. Выявление недеklarированных возможностей аппаратно-программного обеспечения / А.А. Нечай // Экономика и социум. – 2014. – № 1–2 (10). – С. 457–460.
8. Нечай А.А. Специфика проявления уязвимостей в автоматизированных системах управления критически важными объектами / А.А. Нечай, П.Е. Котиков // Современные тенденции в образовании и науке : сборник научных трудов по материалам Международной научно-практической конференции : в 14 ч. – Тамбов, 2014. – С. 96–97.
9. Вепрев С.Б. Скрытый метод выявления утечек инсайдерской информации / С.Б. Вепрев, П.И. Гончаров // Вестник Российского нового университета. – 2014. – № 4. – С. 152–155.