

П.Е. Котиков¹
А.А. Нечай²

P.E. Kotikov
A.A. Nechay

**РЕШЕНИЕ ПРОБЛЕМЫ УПРАВЛЕНИЯ
ПАРАЛЛЕЛЬНЫМ ВЫПОЛНЕНИЕМ
ТРАНЗАКЦИЙ В РАСПРЕДЕЛЕННЫХ
БАЗАХ ДАННЫХ ДЛЯ УСТРАНЕНИЯ
ОПАСНОЙ ПРОТИВОРЕЧИВОСТИ**

**SOLUTION TO THE PROBLEM
OF CONCURRENCY CONTROL
IN DISTRIBUTED DATABASE
TRANSACTIONS TO ELIMINATE
DANGEROUS INCONSISTENCY**

Данная статья посвящена проблеме организации распределенных запросов и транзакций в распределенных гетерогенных базах данных геоинформационных систем. Показаны некоторые сложности управления параллельным выполнением и возможные решения.

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

This article deals with the problem of distributed queries and distributed transactions across heterogeneous databases of geographic information systems. Some complexities of concurrency management and their possible solutions are presented.

Keywords: geographic information system, heterogeneous database, transaction.

Одной из наиболее ценных возможностей при применении гетерогенных распределенных баз данных (БД) является возможность выполнения распределенных запросов [1; 4]. Такую возможность – выполнять распределенный запрос – предоставляет только полностью распределенная база данных, поскольку здесь возможно:

1) разбить базу данных на несколько фрагментов;

2) ссылаться на один или более таких фрагментов из одного запроса, используя прозрачность фрагментации;

3) обеспечить возможность в одном запросе обращаться к физически разделенной таблице.

Во всех случаях прозрачность транзакций остается свойством системы управления распределенной базы данных, которое гарантирует, что все транзакции БД будут обеспечивать целостность и непротиворечивость базы данных. В случае распределенной базы данных геоинформационной системы, транзакции БД могут обновлять данные, хранящиеся на различных компьютерах, объединенных в сеть. Свойство

прозрачности транзакций будет гарантировать, что транзакция будет завершена только в том случае, если на всех сайтах базы данных, вовлеченных в транзакцию, будут завершены все части транзакции. В системах распределенных баз данных для управления транзакциями и обеспечения целостности и непротиворечивости базы данных, как правило, используется сложный механизм распределенных транзакций и распределенных запросов. Также сложным становится само управление параллельным выполнением в распределенной среде. Несмотря на известность этого факта, на практике он не привлекает должного внимания, вплоть до возникновения проблем при проектировании [2; 4].

Вне зависимости от того, является ли транзакция распределенной или нет, она формируется на базе одного или нескольких запросов. Основное отличие нераспределенных транзакций от распределенных состоит только в том, что последние могут обновлять или запрашивать данные на нескольких удаленных сайтах сети. Распределенная транзакция (distributed transaction) позволяет ссылаться на несколько различных локальных или удаленных сайтов. При этом каждый простой запрос может ссылаться только на один удаленный сайт, а транзак-

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

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

ция в целом может ссылаться на несколько сайтов, поскольку каждый запрос сам по себе может ссылаться на различные сайты. Распределенный запрос (distributed request) позволяет получать данные от нескольких удаленных сайтов. Поскольку каждый запрос может получать доступ к данным, расположенным более чем на одном сайте, транзакция может получать доступ к нескольким сайтам [3].

Размещение и разбиение данных должно быть прозрачно для конечного пользователя. Свойство прозрачности транзакций гарантирует, что все транзакции можно рассматривать как централизованные, а также обеспечивает их сериализуемость, т.е. одновременное выполнение транзакций. Это значит, что независимо от того, являются они распределенными или нет, база данных будет переходить из одного устойчивого состояния в другое.

Исключительно важным становится механизм управления параллельным выполнением из-за возможной противоречивости [5]. Управление параллельным выполнением становится особенно значимым в среде распределенной базы данных, поскольку многоместные (на нескольких сайтах) и многопроцессорные операции с большей вероятностью могут привести к противоречивости данных и тупикам, чем одноместные (выполняющиеся на одном сайте) системы [6]. Компонент TP (процессор транзакций) системы управления распределенной БД должен гарантировать, что все части транзакции на всех сайтах будут завершены до того, как последний оператор COMMIT завершит всю транзакцию в целом. Предположим, что каждая операция транзакции подтверждалась локальным процессором данных (DP), но один из DP не смог записать результаты транзакции. Это может привести к противоречивому состоянию БД и неизбежным проблемам с целостностью, поскольку невозможно отменить уже записанные данные.

Решение проблемы состоит в использовании протокола двухфазного подтверждения транзакции (two-phase commit protocol) [3; 5]. Централизованной базе данных необходим только один процессор данных (DP). Все операции с базой данных проводятся на одном сайте, и последовательность операций сразу становится известна. Распределенные базы данных позволяют транзакциям осуществлять доступ к данным на нескольких сайтах. В этих случаях завершающий оператор COMMIT не должен выполняться до тех пор, пока каждый сайт не завершит свою часть транзакции. Протокол двухфазного подтверждения транзакции требует, чтобы каждая

запись в журнале транзакций процессора данных выполнялась до фактического обновления фрагмента. В этом и есть его главная роль. Протокол двухфазного подтверждения транзакции требует применения протокола «выполнить-отменить-повторить» и протокола упреждающей записи. Протокол DO-UNDO-REDO используется процессором данных для отката транзакций назад (roll back) или отката транзакций вперед (roll forward) на основе записей. Чтобы гарантировать, что операции DO-UNDO-REDO смогут обеспечить корректное выполнение операций при крахе системы, используется протокол упреждающей записи. Протокол упреждающей записи (write-ahead protocol) принуждает фиксировать в журнале запись данных для постоянного хранения перед фактическим выполнением этой операции. Протокол двухфазного подтверждения транзакции определяет операции между двумя типами узлов: узел-координатор (coordinator) и один или более подчиненных узлов-субординаторов (subordinates), или когорт (cohort). Протокол реализуется в две фазы.

Фаза 1. Подготовка

1. Координатор посылает сообщение PREPARE TO COMMIT (подготовка к завершению) всем субординаторам.

2. Субординаторы получают сообщение, записывают информацию в журнал транзакций в соответствии с протоколом упреждающей записи и посылают координатору уведомление YES/PREPARED TO COMMIT (да/завершение подготовлено) или NO/NOT PREPARED (нет/завершение не готово).

3. Координатор убеждается, что все узлы готовы к завершению, или в противном случае отменяет действие.

Если все узлы сообщили, что они готовы к завершению (PREPARED TO COMMIT), транзакция переходит в фазу 2. Если один или более узлов отвечают, что они не готовы (NO или NOT PREPARED), координатор распространяет среди всех субординаторов сообщение ABORT (прекращение).

Фаза 2. Последний оператор COMMIT

1. Координатор оповещает всех субординаторов, рассылая сообщение COMMIT, и ожидает ответа.

2. Каждый субординатор, получив сообщение COMMIT, обновляет базу данных в соответствии с протоколом DO.

3. Субординаторы отвечают координатору сообщением COMMITTED (завершено) или NOT COMMITTED (не завершено).

Если один или более субординаторов не вы-

полнили операцию завершения, координатор рассылает сообщение ABORT и тем самым инициирует операцию UNDO (отмену всех изменений).

Цель протокола двухфазного подтверждения транзакции состоит в обеспечении корректного завершения всеми узлами своих частей транзакции; в противном случае транзакция отменяется. Если один или более узлов не выполняют операцию завершения, то необходимая информация по восстановлению БД будет находиться в журнале транзакций, и база данных может быть восстановлена с помощью протокола DO-UNDO-REDO.

Таким образом, реализация возможности организации распределенных запросов к распределенным базам данных ГИС связана не только с безусловным выполнением требований прозрачности, но и с необходимостью решения проблемы управления параллельным выполнением транзакций в распределенной среде.

Возможным решением является использование протокола двухфазного подтверждения транзакций.

Литература

1. Колбина О.Н. Современные и теоретические аспекты управления распределёнными базами данных / О.Н. Колбина, Е.П. Истомин // Информационные технологии системы: управление, экономика, транспорт, право : сборник научных трудов. – СПб. : ООО «Андреевский издательский дом», 2011. – Вып. 1(9).
2. Колбина О.И. Применение распределённых баз данных в геоинформационных системах прогнозирования георисков / Е.П. Истомин, О.Н. Колбина, Е.М. Зоринова // Сборник трудов Международной научно-практической конференции «Инфогео-2013». – СПб. : ООО «Андреевский издательский дом», 2013.
3. Миков А.И., Замятина Е.Б. Распределенные системы и алгоритмы / А.И. Миков, Е.Б. Замятина. – М. : ИНТУИТ, 2008. – 287 с.
4. Разработка и развитие методов, моделей и систем геоинформационного управления пространственно-распределенными объектами: отчет о НИР / Е.П. Истомин, А.Г. Соколов, О.Н. Колбина. – СПб. : Российский государственный гидрометеорологический университет, 2013. – 102 с.
5. Роб П., Коронел К. Системы баз данных: проектирование, реализация и управление. – 5-е изд., перераб. и доп. / пер. с англ. – СПб. : БХВ-Петербург, 2004. – 1040 с.
6. Лохвицкий В.А. Подход к построению системы автоматизированной интеграции информации в базу данных для её своевременной актуализации / В.А. Лохвицкий, С.В. Калинин, 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 частях. – Тамбов, 2014. – С. 96–97.
9. Клименко И.С. К исследованию феномена информации / И.С. Клименко, Л.В. Шарапова // Вестник Российского нового университета. – 2014. – Вып. 4. – С. 141–148.