

УДК 681.3

ТРАНЗАКТНАЯ ОБРАБОТКА ИНФОРМАЦИИ В РАСПРЕДЕЛЕННЫХ БАЗАХ ДАННЫХ

Мехтиева Тахир¹, Алиев Алекпер²¹Институт Кибернетики НАН Азербайджана, Азербайджан²Бакинский Государственный Университет, Азербайджан**Аннотация**

Работа посвящена транзактной обработке информации в распределенных базах данных. Приводится модель распределенной системы, позволяющая решать многие теоретические задачи управления параллелизмом транзакций. Рассматриваются две основные модели параллельного управления транзакциями: синтаксическая модель и семантическая модель. На базе семантической модели предлагается алгоритм блокировки управления параллелизмом транзакций в распределенных базах данных. Далее приводится пример на базе распределенной банковской системы.

Примером распределенной системы (РС) может служить распределенная база данных (РБД), представляющая собой совокупность логически связанных локальных баз данных, географически распределенных и соединенных компьютерными сетями и потоков прикладных процессов, которые могут одновременно использовать несколько баз данных как единое целое. Конечной целью создания РС обработки информации является интеграция информационных и вычислительных ресурсов, а также средств коммуникации и оргтехники и т.п. целого региона и даже мира для своевременного предоставления возможностей их потребления независимо от географического расположения пользователей [1].

Ниже принята простая модель РС, которую можно описать следующим набором предположений:

- РС является совокупностью узлов $S = \{S_i | i = \overline{1, N}\}$ и системы передачи данных, связывающей любую пару узлов;
- В узлах РС хранится и обрабатывается прикладными процессами информация, организованная в виде баз данных;
- Прикладные процессы взаимодействуют друг с другом только путем обмена сообщениями, которые могут либо инициировать содержательную обработку информации, либо выполнять чисто управленческие функции;
- Система передачи данных (сообщений) надежна;
- Процессоры, функционирующие в узлах РС, надежны.

Несмотря на простоту, введенная модель позволяет решать многие теоретические задачи управления, достаточно близкие к задачам, возникающим в реальных практически важных РС. Каждая прикладная программа, предполагающая взаимодействие с РБД, содержит одну или более элементарных действий (чтение, запись и т.д.) и переводящих ее из одного состояния в другое. В результате выполнения некоторые из этих действий могут временно нарушить целостность РБД. Для предотвращения таких нежелательных последствий действия над объектами РБД группируются в транзакции – атомарные (неделимые) операции, которые являются единицами целостности РБД и переводят ее из одного непротиворечивого состояния в другое непротиворечивое состояние. В настоящее время общепринятым методом обработки информации в РС является транзактная обработка. Это означает, что весь процесс обработки состоит из множества заданий, называемых транзакциями. Существуют различные модели транзакций, чаще всего имеют в виду традиционные транзакции, характеризующиеся четырьмя классическими свойствами: атомарности, согласованности, изолированности, долговечности. Транзакция хорошо сформирована, если она: блокирует объект прежде, чем получить доступ к нему; не блокирует объект, который уже заблокирован; до своего полного завершения снимает блокировки с каждого заблокированного им объекта.

Главной проблемой является обеспечение целостности информации в РБД. Проблема обеспечения целостности РБД существенно усложняется при параллельном выполнении транзакций. Поэтому необходимы средства, обеспечивающие сериализуемость множества параллельных транзакций, то есть эквивалентность результата выполнения данной совокупности параллельных транзакций некоторому последовательному выполнению того же множества транзакций. Система управления транзакциями должна так работать в РС, чтобы, начав работать над целостной РС, она оставляла после себя также целостную РС. РС непротиворечива, если значения ее объектов удовлетворяют всем ограничениям целостности. Ограничения целостности – это множество формальных правил, устанавливающих и регламентирующих связи между объектами РС. Они представляют собой некоторые предикаты, которым должны удовлетворять корректные данные.

Ввиду фундаментальной важности указанные проблемы привлекли внимание ведущих зарубежных и отечественных специалистов в области распределенной обработки информации.

Один из методов решения задачи сериализуемости основан на классическом результате и называется метод "блокировки с двухфазным жизненным циклом транзакций". В указанном методе блокировка на ресурсы устанавливается постепенно при выполнении транзакций, и только после терминации (завершения) транзакции все ресурсы разблокируются.

Этот метод позволяет выполнять вместе только совместимые операции. Однако совместное выполнение транзакций может привести к тупиковой ситуации, когда две или более транзакции одновременно блокируют некоторые ресурсы и каждой из них для полного завершения нужен ресурс, заблокированный другой транзакцией. Разрешить такую ситуацию можно с помощью отката одной или нескольких транзакций, вызвавших конфликт. В результате оставшаяся транзакция получает освобожденный ресурс и завершает свою отработку. Отметим, что этот метод предполагает значительную централизованность РС. Такая централизация не всегда эффективна, так как, например, возможна сильная загруженность центрального узла, что снижает общую производительность РС.

Существуют две основные модели параллельного управления транзакциями: синтаксическая модель и семантическая модель. Каждая параллельно управляемая модель имеет критерий, который может быть интерпретирован как множество корректных планов. В случае синтаксической модели параллельного управления используются моноверсионная сериализуемость и мультиверсионная сериализуемость. В синтаксической модели предполагается, что транзакция - это частично упорядоченное множество операций чтения и записи. Критерий корректности плана основан на сериализуемости. Когда говорят, что алгоритм управления параллелизмом является корректным с учетом сериализуемости, имеется в виду, что алгоритм сохраняет целостность базы данных для любого множества ограничений целостности, для любого множества вычислений, выполненных транзакциями. В случае семантической модели параллельного управления используется мультиуровневая атомарность. В семантической модели управления параллелизмом предполагается, что алгоритм управления параллелизмом, кроме синтаксической информации, также владеет семантической информацией о вычислениях, произведенных транзакциями. Эта дополнительная семантическая информация о транзакциях позволяет смягчить условия ограничения критерия сериализуемости, а также позволяет использовать общий критерий корректности, для которого сериализуемость есть частный случай. В случае семантической параллельно управляемой модели задача вычисления корректности плана не может быть поставлена так же, как и в случае синтаксической модели.

В работе предлагается алгоритм блокировки, который расширяет классическую модель транзакций за рамки предположения о том, что транзакция - это частично упорядоченное множество простых операций чтения и записи элементов данных. Этот алгоритм позволяет выполнять параллельных транзакций. В работе множество корректных планов определено мультиуровневым атомарным критерием. Мультиуровневая атомарность описывает множество корректных, семантически целостных планов, как планов не имеющих циклов в определенных отношениях, описывающих зависимости среди шагов транзакций. В работе приводится пример на базе распределенной банковской системы.

Список использованных источников:

1. Мурадова Г.И., Алиев А.А. Модель и алгоритм отказоустойчивой системы управления транзакциями в распределенных базах данных // Системы управления и информационные технологии, Москва, №3.1(37), 2009, с.174-180.