

УДК 528.92

DOI: 10.33764/2411-1759-2021-26-1-98-109

ПРИМЕНЕНИЕ МЕХАНИЗМА ПРЕДПРОЦЕССОРНОЙ ОБРАБОТКИ РАЗНОРОДНЫХ ДАННЫХ В ГЕОИНФОРМАЦИОННЫХ СИСТЕМАХ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЯ

Ольга Николаевна Колбина

Российский государственный гидрометеорологический университет, 192007, Россия, г. Санкт-Петербург, Воронежская улица, 79, кандидат технических наук, доцент кафедры прикладной информатики, тел. (812)633-01-86, e-mail: olya_kolbina@mail.ru

Евгений Петрович Истомин

Российский государственный гидрометеорологический университет, 192007, Россия, г. Санкт-Петербург, Воронежская улица, 79, доктор технических наук, профессор, зав. кафедрой прикладной информатики, тел. (812)633-01-86, e-mail: biom@bk.ru

Наталья Владимировна Яготинцева

Российский государственный гидрометеорологический университет, 192007, Россия, г. Санкт-Петербург, Воронежская улица, 79, кандидат технических наук, доцент кафедры прикладной информатики, тел. (812)633-01-86, e-mail: yagotintceva@yandex.ru

Марсель Равильевич Вагизов

Санкт-Петербургский государственный лесотехнический университет им. С. М. Кирова, 194021, Россия, г. Санкт-Петербург, Институтский пер., 5, кандидат технических наук, доцент кафедры информационных систем и технологий, тел. (812)633-01-86, e-mail: bars-tatarin@yandex.ru

Проблема автономности локальных баз данных в распределенных информационных системах широко известна и давно решается специалистами, однако однозначного и всеми признанного подхода к решению данной проблемы еще нет. Не является исключением и процесс работы с распределенными базами данных при создании геоинформационных систем поддержки принятия решений, где первостепенным критерием работы системы должна оставаться нерушимость структуры и полноты локальных баз данных. Цель исследования авторов заключается в повышении автономности локальных баз данных в геоинформационных системах поддержки принятия решения на основе применения предпроцессорной обработки разнородных данных. В данной статье предложены функциональные подходы и конкретные решения по физическому и логическому представлению структуры геоинформационной системы по средствам реализации механизма предпроцессорной обработки, что позволит соединять и преобразовывать данные еще до начала их отображения в приложении. Таким образом, лицо, принимающее решение, при работе в геоинформационной системе поддержки принятия решения будет иметь полноценный доступ к актуальной информации с возможностью фиксации полученных результатов, путем сохранения созданной по его запросам виртуальной базы данных.

Ключевые слова: базы данных, транзакции, распределенные геоинформационные системы, разнородные данные, предпроцессорная обработка, системы управления данными, виртуальный процессор, лицо, принимающее решение, информационное пространство, глобальные тупики

Введение

Внедрение информационных технологий в человеческую деятельность привело к развитию и усовершенствованию информационных систем, телекоммуникационных сетей и микропроцессорной техники. В последнее десятилетие особенно бурно начали разви-

ваться и применяться геоинформационные системы (ГИС), которые направлены на решение проблем прокладки маршрута, оптимизацию деятельности различных организаций, нахождение интересующих объектов на карте, поддержку принятия решения и др.

Геоинформационные системы поддержки принятия решения встречаются в различных

предметных областях, будь то мониторинг экологической безопасности, управление водными объектами, а также в медицине, когда речь идет о быстром реагировании служб скорой помощи и быстром принятии решения, приоритетности вызовов.

Зачастую необходимая информация для принятия объективного решения хранится в разных информационных системах, созданных для нужд определенной структуры или отдела, без учета перспективы дальнейшего использования данных и интеграции с другими ИС. Например, для принятия управленческого решения, при выбросе вредных веществ в атмосферу, необходимо получить и проанализировать информацию о концентрации вредных веществ (базы данных газоанализатора), о метеорологической обстановке (базы данных метеорологической станции) и о численности населения (база данных паспортного стола) в районе выброса. Объединение и концентрация всей необходимой информации в одном геопространстве делают прозрачным процесс управления и повышают эффективность поддержки принимаемых решений. Успешность реализации подобных систем будет зависеть от трех составляющих процессов обработки геоданных, а именно, обработка должна быть сквозной, поверхностной и согласованной [1]. В свою очередь, необходимо обеспечить доступ к пространственно-распределенной информации, которая является оперативной, а сам доступ поддерживает свойства достоверности и полноты. Геоинформационная система, обладающая вышеперечисленными качествами, позволит лицам, принимающим решения (ЛПР), обоснованно выполнять стратегические и управленческие оперативные действия. Из-за территориальной удаленности источников данных, различной структуры баз данных (БД) и форматов хранящейся в ней информации не всегда может быть реализовано такое геоинформационное решение.

Системы управления данными при работе с количеством баз данных больше одной должны удовлетворять требованиям целостности, согласованности данных и учитывать степень многослойности информационного пространства. Проблема заключается в техно-

логической сложности перехода от централизованного управления унифицированными данными к распределенной обработке данных в многоформатной информационной среде, включая автономные системы работы в локальной сети разнородных баз данных, которые не предоставляют возможности управлять производительностью глобальных транзакций.

Механизмы формирования единого информационного пространства при построении ГИС поддержки принятия решения

Решением сохранения полноты и структуры баз данных при единовременном обращении нескольких пользователей может стать попытка обеспечения независимости и распределенности систем баз данных при условии их разнородности.

Существует большое количество технологий, обеспечивающих устранение проблем разнородности и распределенности баз данных, однако однозначное сохранение независимости источников информации ни одна предоставить не может. Чаще всего в распределенных системах применяется технология, позволяющая лишь выполнять операцию чтения из источника данных, что приводит к отсутствию конфликтов между прямыми транзакциями. Технология реплицирования, или ДАО-технология, может быть успешно применена и для реализации географической информационной системы поддержки принятия решения, однако развитие и увеличение количества потребляемой информации пользователем подталкивает разработчиков к усовершенствованию механизмов работы с распределенными гетерогенными базами данных [2].

Повсеместная цифровизация информации направляет разработчиков к созданию информационных систем, которые используют как можно больше данных в своей работе, тем самым повышается качество принятия решения управленцем. Локализация и независимость баз данных в таких условиях ставятся под сомнение, так как любой зарегистрированный пользователь должен иметь доступ к чтению, изменению и сохранению информации.

Для распределенных баз данных, как и для централизованных, применим аппарат по управлению модификацией глобальных транзакций. В нем глобальные транзакции и метки изменения данных во времени управляются контролем допуска [3].

Обеспечение целостности и согласованности данных в распределенных информационных системах при пересылке сообщения между внешним приложением и внутренней базой данных производится за счет конкурентной транзакции. Конкурентная передача данных не решает проблему гонок данных и, как следствие, повышается вероятность появления распределенных взаимоблокировок, которые могут возникнуть при одновременном изменении базы данных несколькими пользователями, и включения конкурентных механизмов контроля.

В теории управления транзакциями существуют два свойства, характеризующих правильность выполнения транзакций, – сериализуемость [4] и надежность [5]. Если используется программная среда без сбоев, то применимо свойство сериализуемости, в ином случае – надежность.

Поскольку критерий надежности применим в средах со сбоями, где может произойти потеря данных, важно при описании транзакций в системе использовать свойства согласованности и целостности данных. В системе управления данными (СУБД), в отличие от распределенной географической информационной системы (РГИС), обеспечение надежности характеризуется необходимостью выбора между сохранением локальной независимости и использованием механизма эффективного глобального восстановления. Для создания транзакций в РГИС с отказами могут быть применимы два типа протоколов. Первый тип протокола (протокол атомарной фиксации) позволяет производить завершение глобальной транзакции, имеющей одинаковые состояния во всех узлах системы. Второй тип протокола (протокол восстановления) позволяет определить действия в системе в случае наступления сбоев.

При использовании программной среды без учета сбоев можно рассмотреть свойство сериализуемости, иначе применяется свойство надежности.

Глобальная сериализуемость может быть достигнута в гетерогенных базах данных, если в каждой отдельной базе данных параллельные транзакции изолированы или порядок изоляции субтранзакций одинаков.

Обеспечение глобальной сериализуемости является достаточно сложной задачей ввиду того, что помимо прямых конфликтов, существуют еще и косвенные конфликты между глобальными транзакциями (рис. 1).

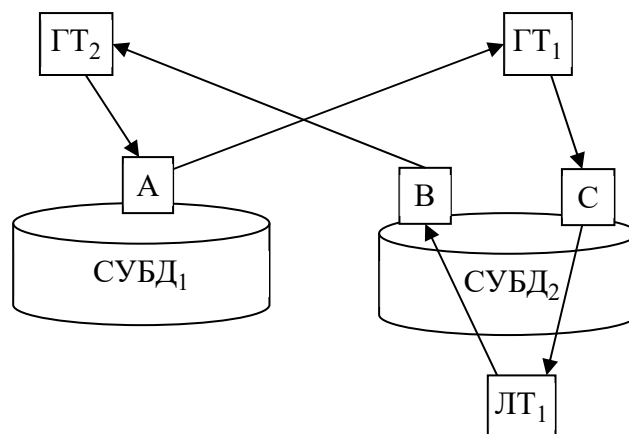


Рис. 1. Схема возможного взаимодействия двух локальных баз данных в РГИС

В РГИС выполняются три основные транзакции: одна локальная транзакция (ЛТ₁) и две глобальных транзакции (ГТ₁ и ГТ₂). На рис. 1 имеется одна субтранзакция, принадлежащая двум глобальным транзакциям в двух локальных СУБД. В СУБД₁ в первую очередь выполняется операция чтения с объекта данных А первой глобальной транзакцией (ГТ₁), после чего выполняется операция записи глобальной транзакцией (ГТ₂) [6]. Примером записи данных операций может служить формула

$$D_{ГТ1}(A)G_{ГТ2}(A), \quad (1)$$

где $D_{ГТ1}(A)$ – функция прочтения информации из объекта данных А;

$G_{ГТ2}(A)$ – функция записи в объект данных А.

Локальные операции транзакций в СУБД₁ показывают наличие прямого конфликта между двумя глобальными транзакциями. Зависимость ГТ₂ от ГТ₁ в этой системе управления баз данных показывает сериализационный порядок их выполнения [7].

Во второй СУБД транзакции ГТ₁ и ГТ₂ могут извлечь различные данные. К примеру, С записывается с помощью глобальной транзакции ГТ₁, а чуть позже В читается транзакцией ГТ₂ [7]. Записать локальные операции транзакций в СУБД₂ можно с помощью формулы

$$D_{ЛТ_1}(C) G_{ГТ_1}(C) D_{ГТ_2}(B) G_{ЛТ_2}(B), \quad (2)$$

где $D_{ГТ_2}(B)$ – функция прочтения при второй глобальной транзакции информации из объекта данных В;

$G_{ГТ_1}(C)$ – функция записи при первой глобальной транзакции в объект данных С;

$D_{ЛТ_1}(C)$ – функция прочтения при первой локальной транзакции информации из объекта данных С;

$G_{ЛТ_2}(B)$ – функция записи при второй локальной транзакции в объект данных В.

ГТ₁ и ГТ₂ относительно СУБД₂ не имеют прямого конфликта, но при этом конфликтуют друг с другом. Этот косвенный конфликт вызывается выполнением ЛТ₁. Зависимость ГТ₁ от ЛТ₁ формирует порядок сериализации транзакций в СУБД₂ (рис. 2). При этом ЛТ₁ имеет зависимость от выполнения ГТ₂.

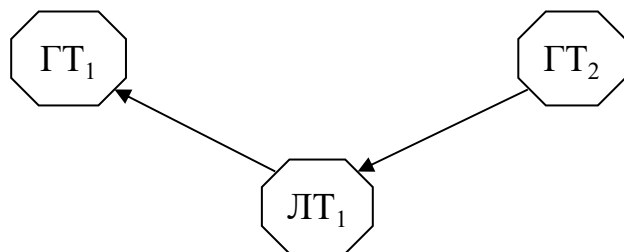


Рис. 2. Косвенный конфликт транзакций в СУБД₂

Ввиду автономности базы данных в РГИС информация о косвенных конфликтах в такой системе недоступна. Это обуславливается тем, что система может контролировать только глобальные транзакции. Поэтому регистрируются только прямые конфликты, как приведено в вышеописанном примере. Основной проблемой поддержания глобальной сериализуемости в РГИС является предупреждение косвенных конфликтов, а также образование глобальных тупиков.

При встрече множества разнородных баз данных и, соответственно, множества гло-

бальных транзакций, есть большая вероятность образования глобальных тупиков [7]. Во избежание возникновения таких ситуаций могут быть применены сценарии ожидания, а именно когда одна глобальная транзакция не начинается, пока не завершится предыдущая:

$$\begin{aligned} & ГТ_{i+1} \text{ ожидает } ГТ_i \text{ в } ГБД_i, \text{ где } 1 < i < n, \\ & \text{а } ГТ_1 \text{ ожидает } ГТ_n \text{ в } ГБД_n, \end{aligned}$$

где ГТ – глобальная транзакция;

ГБД – гетерогенная база данных;

i – порядковый номер глобальных транзакции и версии гетерогенной базы данных после завершения транзакции;

n – конечный номер глобальной транзакции и версии гетерогенной базы данных после завершения транзакции.

Еще одним сценарием развития событий может служить закрепление субтранзакции глобальной транзакцией при условии, что все дополнительные процессы передачи данных выполняются и готовы к формированию единого пакета. При обратном условии такая транзакция занимает ресурсы локальной базы данных.

Возникновение тупиков характеризуется отсутствием обмена информацией между локальными базами данных при завершении транзакций, что необходимо учесть и попытаться устранить при создании распределенной географической информационной системы управления принятием решения.

Рассмотрим пример (рис. 3), где в РГИС над объектами данных А, В, С, Д выполняются две глобальные транзакции (ГТ₁, ГТ₂) и две локальные транзакции (ЛТ₁, ЛТ₂). ГТ₁ выполняет чтение из объекта данных А, а ГТ₂ – чтение из объекта данных С. ЛТ₁ выполняет операцию чтения объекта В и допускает к чтению объекта А, после чего ЛТ₁ переходит в режим ожидания до тех пор, пока ГТ₁ не разблокирует объект А. Во второй СУБД происходят аналогичные операции на локальном уровне и ЛТ₁ переходит в режим ожидания разблокировки объекта С глобальной транзакцией ГТ₂. Следовательно ГТ₁ и ГТ₂ допускают выполнение операций, находящихся в режиме ожидания, и возникает вероятность глобального тупика.

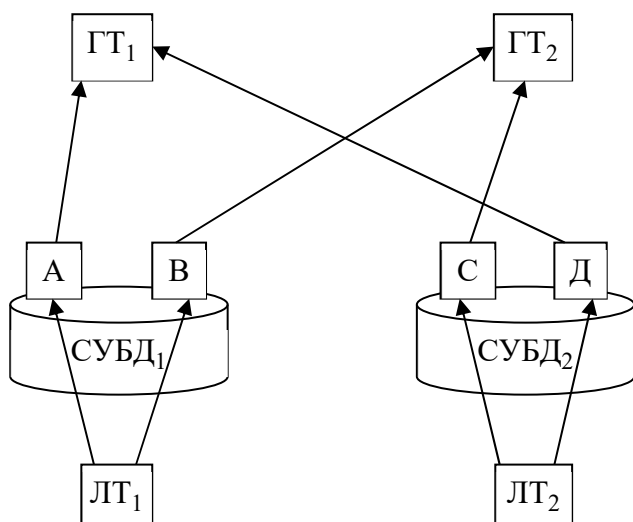


Рис. 3. Схема выполнения транзакций с вероятностью возникновения тупика

Журналирование атомарных транзакций обеспечивает бесперебойную и надежную работу РГИС [8]. Механизм фиксации выполненных транзакций в соответствующий файл состоит из модифицированных протоколов, которые фиксируют субтранзакции ГТ в случае, если сама ГТ не зафиксирована. Но в случае отката ГТ при использовании журналирования появляется необходимость удаления записей, уже закрепленных субтранзакцией.

Перечислим условия для бесперебойного управления транзакциями в РГИС:

- 1) все РБД являются локально-атомарными;
- 2) очередность выполнения субтранзакций наследуется из очередности выполнения глобальных транзакций в СУБД;
- 3) должна выполняться только одна субтранзакция ГТ в одной РБД;
- 4) РГИС, основанная на РБД, должна быть устойчива к появлению глобальных тупиков.

Независимость включаемых в РГИС баз данных должна быть одним из основных условий реализации распределенных гетерогенных систем. Достичь атомарности баз данных возможно с учетом применения глобальной сериализуемости, которая предусматривает наличие специального управляющего протокола, реализуемого при выполнении глобальных транзакций. Применение протокола атомарной фиксации предусматривает наличие беступиковых историй локальных

некаскадных транзакций и локальную сериализуемость.

Использование описанного протокола должно сопровождаться определением модели управления РГИС, позволяющей объединить разрозненные БД в единое информационное пространство (рис. 4).

В модели представлено наличие нескольких серверов разноформатных баз данных с локальным использованием конкретными клиентами, объединенных в единое пространство за счет внедрения в структуру информационной системы специализированного модуля «менеджер транзакций». Управление процессом обмена информацией происходит за счет генератора временных меток, который упорядочивает выполнение транзакций, пометая каждую транзакцию временной меткой, а объект, над которым будет проходить операция, – индикатором транзакции, и менеджера транзакций, который позволяет связать между собой серверы и структурировать передачу транзакций. Данная модель хороша тем, что при ее использовании нет необходимости распознавать тупики.

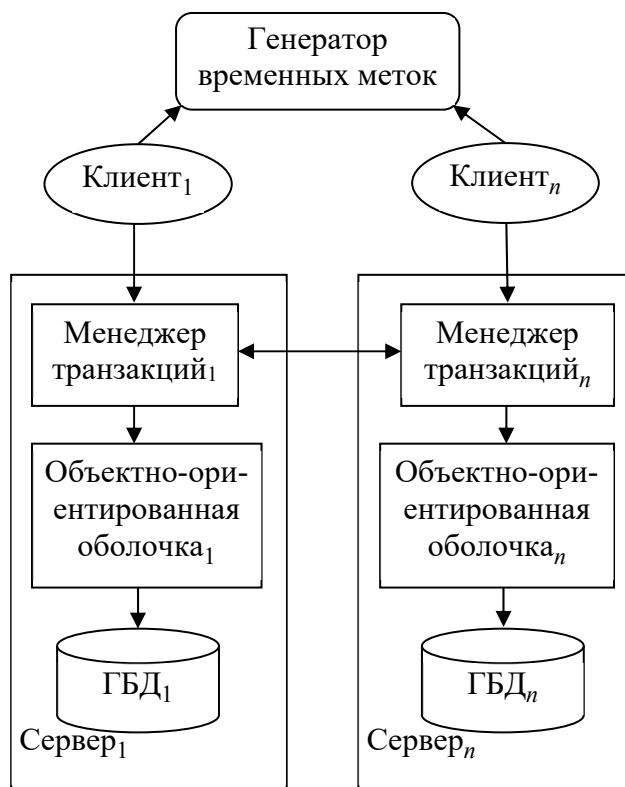


Рис. 4. Структурная модель объединения отдельных БД в единое информационное пространство

Определение механизма обработки разнородных данных

Применение многоуровневой модели для создания сложных программных систем позволяет обеспечить равномерное распределение нагрузки, надежность и дальнейшую масштабируемость геоинформационной системы [9, 10]. Самой оптимальной моделью

для РГИС является модель, которая состоит из трех уровней «данные – обработка – интерфейс». Трехуровневое представление позволяет исключить зависимость системы от количества программных модулей на каждом уровне.

Потоковая модель прохождения информации получена из структурной модели (см. рис. 4) и представлена на рис. 5.

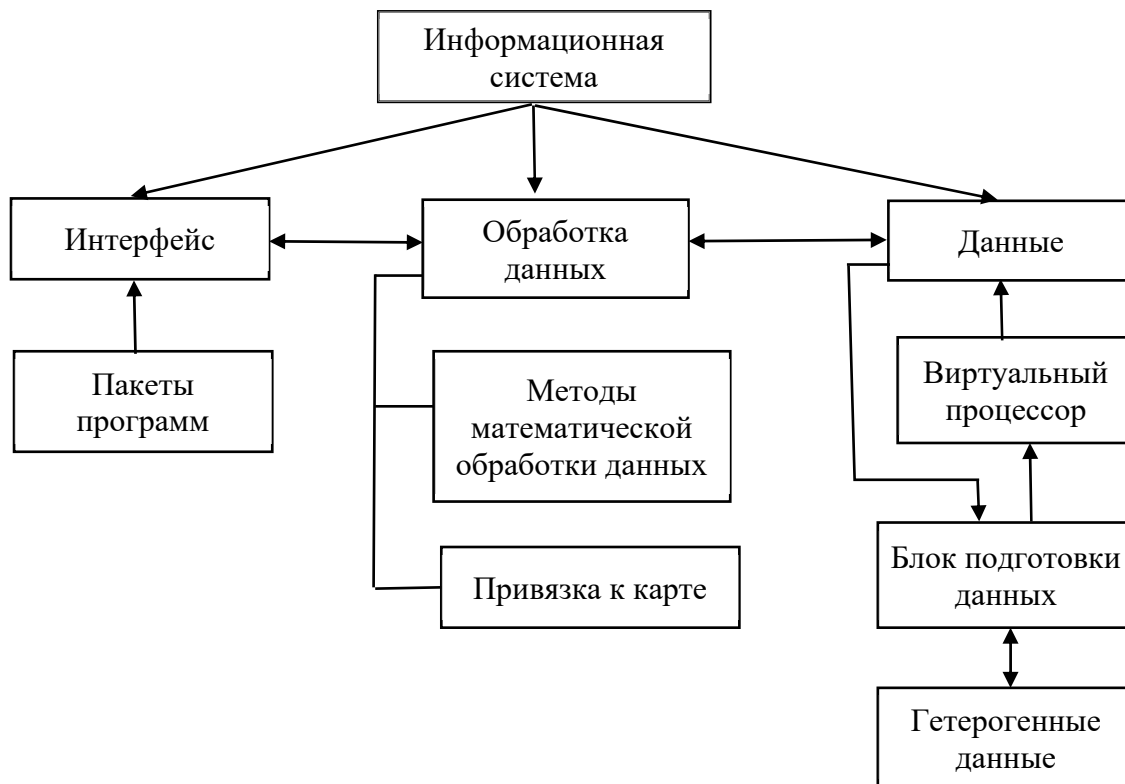


Рис. 5. Потоковая модель прохождения информации в РГИС

Для обеспечения надежного и быстрого доступа к данным введем понятие предварительной (предпроцессорной) обработки разнородных данных.

Такая обработка включает в себя виртуальный процессор данных и блок подготовки данных.

Виртуальный процессор представляется в виде программно-технологического модуля. Виртуальный процессор данных позволяет обеспечить одинаковый доступ к разнообразным данным системы, при этом скрывая размещение данных, их структуру и форматы, что позволяет не преобразовывать исходные данные под формат системы и использовать их локально в исходном виде [11]. Сбор дан-

ных и их первичная обработка производится блоком подготовки данных по запросу пользователя из системы.

Использование виртуального процессора предполагает наличие определенных состояний преобразуемых данных, а именно:

- фактические состояния. Информация о функционировании организационных подсистем объекта управления;
- планируемые и нормативные состояния. Входные документы, результаты принятия управленческих решений [12, 13].

Работа в рассматриваемой РГИС на основе РБД с использованием виртуального процессора происходит по следующему сценарию (рис. 6).



Рис. 6. Алгоритм использования виртуального процессора данных

Начало алгоритма соизмерено с моментом обращения пользователя к ГИС для решения конкретной задачи, тогда и формируется запрос для получения данных из различных источников.

Затем проводится проверка источников на согласованность [14, 15].

При успешной проверке на согласованность предоставляется доступ к базе согласованных состояний. После пройденной проверки формируются базы плановых, нормативных или вероятностных состояний.

Если проверка на согласованность не пройдена, формируются базы несогласованных состояний. В таких базах хранятся данные из разнородных локальных баз данных [16].

Все базы состояний в механизме предпроцессорной обработки данных объединяются единой интегрированной базой данных.

Диаграмма зависимости подсистем РГИС с учетом описанных выше механизмов виртуального процессора данных имеет вид, представленный на рис. 7.

Использование в структуре РГИС при поддержке принятия управленческих задач предпроцессора позволяет координировать и обеспечивать взаимодействие отдельных подзадач. Блок подготовки данных, входящий в состав виртуального процессора, при необходимости может расширяться. При этом все дополнения будут наследовать программные протоколы уже существующих блоков [17, 18].

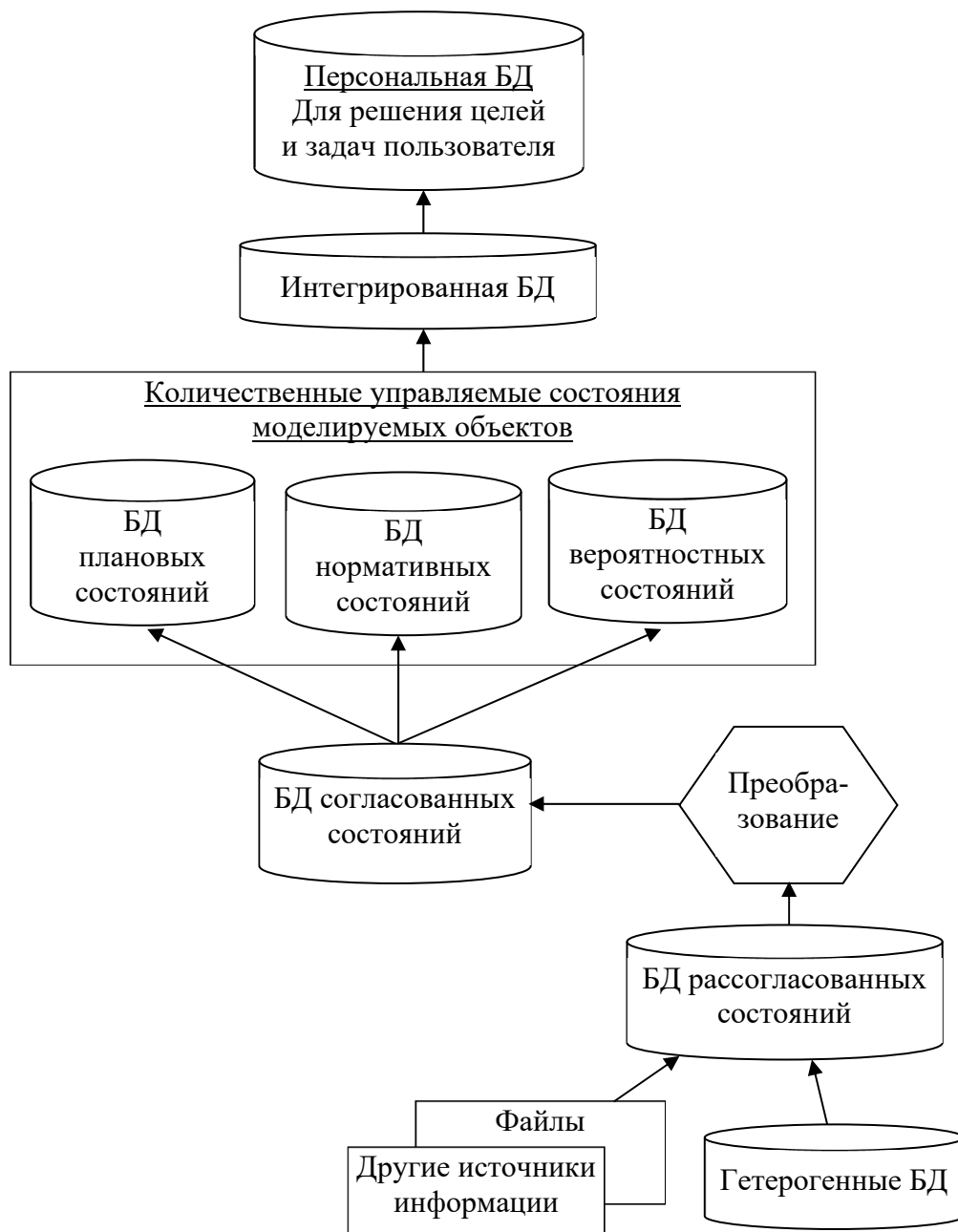


Рис. 7. Диаграмма зависимости подсистем РГИС

Заключение

Зачастую для принятия решения ЛПР должен учесть большое количество факторов, критериев и информации, которая может быть ему недоступна. Доступ к информации бывает ограничен по многим причинам, в том числе в связи с ее нахождением и использованием другими структурными подразделениями или организациями, иногда находящимися на значительном расстоянии друг от друга. Объединение данных из уже сформированных баз данных возможно, если создать распределенную геоинформационную систему управления, однако наличие большого количества источников и их разнородность могут привести к конфликтам между глобальными транзакциями [19]. Решением проблемы может служить создание дополнительного модуля при разработке геоинформационной системы принятия управленческих решений, который будет предварительно обрабатывать данные и преобразовывать их к необходимому виду и формату, что приве-

рованных баз данных возможно, если создать распределенную геоинформационную систему управления, однако наличие большого количества источников и их разнородность могут привести к конфликтам между глобальными транзакциями [19]. Решением проблемы может служить создание дополнительного модуля при разработке геоинформационной системы принятия управленческих решений, который будет предварительно обрабатывать данные и преобразовывать их к необходимому виду и формату, что приве-

дет к повышению автономности баз данных [20]. Такой модуль был применен при реализации РГИС оценки риска экстремальных явлений в Республике Венесуэле, свидетельство о государственной регистрации баз данных № 2018620801 от 04.06.2018 г. [21]. Применение механизма предпроцессорной обработки разнородных данных в геоинформаци-

онных системах поддержки принятия решения направлено на использование максимального количества необходимой информации, вне зависимости от ее территориального расположения, формата и структуры хранения данных, что позволяет более качественно принимать решения по управлению процессом или объектом.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Куракина Н. И., Минина А. А. Система поддержки принятия решений по управлению водными объектами с использованием ГИС [Электронный ресурс] : журнал ArcReview. – 2008. – № 1 (44). – Режим доступа: https://www.esri-cis.ru/news/arcreview/list.php?SECTION_ID=37.
2. Бурковский В. Л., Дорофеев А. Н., Семьин С. В. Моделирование и алгоритмизация управления гетерогенными базами данных в распределенных информационных системах. – Воронеж : Гос. техн. университет, 2003. – 136 с.
3. Коголовский М. Р. Энциклопедия технологий баз данных. – М. : Финансы и статистика, 2002. – 800 с.
4. Bernstein P. A., Hadzilacos V., Goodman N. Concurrency Control and Recovery in Database Systems, Addison-Wesley Series // Computer Science. – Addison-Wesley, Uneted States of America, 1987.
5. Hadzilagos V. A. Theory of Reliabiliti in Database Systems // Juornal of the ACM. – 1988. – Vol. 35, № 1. – P. 121–145.
6. Дорофеев А. Н., Бурковский В. Л. Анализ сериализуемости глобальных транзакций в распределенной системе мульт-БД // Современные проблемы информатизации в непромышленной сфере и экономике : сб. науч. тр. VI Международной электронной науч. конф. – Воронеж, 2001. – С. 89–90.
7. Колбина О. Н. Разработка геоинформационной системы оценки параметров климатических условий на основе распределенных гетерогенных баз данных : дисс. ... канд. техн. наук. – СПб., 2014. – 150 с.
8. Дорофеев А. Н., Бурковский В. Л. Применение протокола атомарной фиксации в системе мульт-БД // Промышленная автоматика : межвузовский сборник научных трудов. – Воронеж, 2001. – С. 30.
9. Истомин Е. П., Колбина О. Н., Степанов С. Ю. Методика проектирования геоинформационной системы управления территориями Заполярья на основе распределенных гетерогенных баз данных // Ученые записки РГГМУ. – 2015. – Вып. 39. – С. 221–228.
10. Истомин Е. П., Новиков В. В., Сидоренко А. Ю., Колбина О. Н., Степанов С. Ю. Сложная информационная система прогнозирования рисков с применением фильтра Калмана – Бьюси // Ученые записки РГГМУ. – 2014. – Вып. 36. – С. 183–188.
11. Истомин Е. П. Сетевые методы и модели распределения автоматизированных систем : дисс. ... д-ра техн. наук. – СПб., 1998. – 263 с.
12. Истомин Е. П., Колбина О. Н., Сидоренко А. Ю., Петров Я. А., Степанов С. Ю. Математическая модель обработки пространственно-распределенных разнородных геоданных для принятия управленческих решений по прокладке оптимальных маршрутов следования судов в Арктике // Естественные и технические науки. – 2019. – № 4 (130). – С. 130–133.
13. Истомин Е. П., Кирсанов С. А., Соколов А. Г., Колбина О. Н. Феномен геоинформационного управления и принципы его реализации // Вестник СПбГУ. – 2014. – Сер. 7, вып. 4. – С. 180–188.
14. Гарсиа Эскалона Х. А., Истомин Е. П., Колбина О. Н. Перспективы развития инфраструктуры пространственных данных с использованием современных технологий // Ученые записки РГГМУ. – 2018. – № 50. – С. 130–136.
15. Степанов С. Ю., Колбина О. Н. Методика проектирования геоинформационной системы для поддержки принятия управленческих решений на основе использования пространственнораспределенной разнородной информации // Информационные технологии и системы: управление, экономика, транспорт, право. – 2016. – № 2 (18). – С. 16–21.
16. Бугаков П. Ю., Колесников А. А. Анализ функциональных возможностей офисных приложений для визуализации и оценки геоданных // Вестник СГУГиТ. – 2019 – Т. 24, № 4. – С. 104–119.

17. А. В. Флегонтов, Г. Б. Воронов, В. Н. Смирнов, Г. А. Задубина. Картографическое обеспечение наземных навигационных систем // Вестник СГУГиТ. – 2019. – Т. 24, № 3 – С. 106–118.
18. Вагизов М. Р., Навалихин С. В., Баенгуев Б. А. Разработка геоинформационной системы благоустройства зеленых насаждений общего пользования г. Санкт-Петербурга // Фундаментальные исследования. – 2017. – № 11-1. – С. 35–40.
19. Новиков В. В., Истомин Е. П., Соколов А. Г. Концептуальная модель геоинформационного управления природно-техническими системами и территориями // Петербургский экономический журнал. – 2019. – № 2. – С. 55–63.
20. Zisman A., Kramer J. Towards Interoperability in Heterogeneous Database Systems. Imperial college Research Report No. Doc 95/11, December, 1995.
21. Истомин Е. П., Колбина О. Н., Гарсия Эскалона Х. А. База данных оценки риска экстремальных явления в Венесуэле : Патент России № 2018620801 от 04.06.2018 г.

Получено 09.06.2020

© О. Н. Колбина, Е. П. Истомин, Н. В. Яготинцева, М. Р. Вагизов, 2021

APPLICATION OF THE PRE-PROCESSOR MECHANISM FOR PROCESSING HETEROGENIOUS DATA IN DECISION-MAKING SUPPORT GEOINFORMATION SYSTEMS

Olga N. Kolbina

Russian State Hydrometeorological University, 79, Voronezhskaya St., St. Petersburg, 192007, Russia, Ph. D., Associate Professor, Department of Applied Informatics, phone: (812)633-01-86, e-mail: olya_kolbina@mail.ru

Evgeny P. Istomin

Russian State Hydrometeorological University, 79, Voronezhskaya St., St. Petersburg, 192007, Russia, D. Sc., Professor, Head of the Department of Applied Informatics, phone: (812)633-01-86, e-mail: biom@bk.ru

Natalya V. Yagotinceva

Russian State Hydrometeorological University, 79, Voronezhskaya St., St. Petersburg, 192007, Russia, Ph. D., Associate Professor, Department of Applied Informatics, phone: (812)633-01-86, e-mail: yagotintceva@yandex.ru

Marsel V. Vagizov

St. Petersburg State Forest Technical University, 5, Institutskiy Pereulok St., St. Petersburg, 194021, Russia, Ph. D., Associate Professor, Department of Information Systems and Technologies, phone: (812)670-93-29, e-mail: bars-tatarin@yandex.ru

The problem of local databases autonomy in distributed information systems is widely known and has been solved for a long time by specialists, but there is still no unambiguous and universally recognized approach to solving this problem. The process of working with distributed databases when creating a geographic information system for decision support is not an exception, where the inviolability of the structure and completeness of local databases should remain the primary criterion for the system's operation. The purpose of the study is to increase the autonomy of local databases in geoinformation decision support systems based on the use of preprocessing processing of heterogeneous data. This article proposes approaches and specific solutions for the physical and logical representation of the geoinformation system structure by means of the implementation of the preprocessing mechanism, which will allow you to combine and transform data even before they are displayed in the application. Thus, a person making a decision, when working in a geoinformational decision support system, will have full access to up-to-date information with the ability to record the results obtained by saving a virtual database created at his request.

Keywords: databases, transactions, distributed geographic information systems, heterogeneous data, preprocessing, data management systems, virtual processor, decision maker, information space, global deadlocks

REFERENCES

1. Kurakina, N. I., & Minina, A. A. (2008). Decision support system for the management of water bodies using GIS. *ArcReview*, 1(44). Retrieved from https://www.esri-cis.ru/news/arcreview/list.php?SECTION_ID=37 [in Russian].
2. Burkovsky, V. L., Dorofeev, A. N., & Semynin S. V. (2003). *Modelirovanie i algoritimizatsiya upravleniya geterogennymi bazami dannykh v raspredelennykh informatsionnykh sistemakh [Modeling and algorithmization of managing heterogeneous databases in distributed information systems]*. Voronezh: State Technical University Publ., 136 p. [in Russian].
3. Kogalovsky, M. R. (2002). *Entsiklopediya tekhnologiy baz dannykh [Encyclopedia of database technologies]*. Moscow: "Finance and Statistics" Publ., 800 p. [in Russian].
4. Bernstein, P. A., Hadzilacos, V., & Goodman, N. (1987). Concurrency Control and Recovery in Database Systems, Addison-Wesley Series. In *Computer Science*. Addison-Wesley, Uneted States of America.
5. Hadzilagos, V. A. (1988). Theory of Reliabiliti in Database Systems. *Journal of the ACM*, 35(1), 121–145.
6. Dorofeev, A. N., & Burkovsky, V. L. (2001). Analysis of serializability of global transactions in a distributed multi-database system. In *Sbornik nauchnykh trudov VI Mezhdunarodnoy elektronnoy nauchnoy konferentsii: Sovremennye problemy informatizatsii v nepromyshlennoy sfere i ekonomike [Proceedings of VI International Electronic Scientific Conferenced: Modern Problems of Informatization in the Non-Industrial Sphere and Economy]* (pp. 89–90). Voronezh [in Russian].
7. Kolbina, O. N. (2014). Development of a geographic information system for assessing the parameters of climatic conditions on the basis of distributed heterogeneous databases. *Candidate's thesis*. Saint Petersburg, 150 p. [in Russian].
8. Dorofeev, A. N., & Burkovsky, V. L. (2001). Application of the atomic fixation protocol in a multi-database system. In *Mezhvuzovskiy sbornik nauchnykh trudov: Promyshlennaya avtomatika [Interuniversity Collection of Scientific Papers: Industrial Automation]* (P. 30). Voronezh [in Russian].
9. Istomin, E. P., Kolbina, O. N., & Stepanov, S. Yu. (2015). The design methodology of the geographic information system for managing the territories of the Arctic based on distributed heterogeneous databases. *Uchenye zapiski RGGMU [Scientific Notes of the Russian State Hydrometeorological University]*, 39, 221–228 [in Russian].
10. Istomin, E. P., Novikov, V. V., Sidorenko, A. Yu., Kolbina, O. N., & Stepanov, S. Yu. (2014).
11. A complex information system for risk prediction using the Kalman – Bucy filter. *Uchenye zapiski RGGMU [Scientific Notes of the Russian State Hydrometeorological University]*, 36, 183–188 [in Russian].
12. Istomin, E. P. (1998). Network Methods and Distribution Models of Automated Systems. *Doctor's thesis*. Saint Petersburg, 263 p. [In Russian].
13. Istomin, E. P., Kolbina, O. N., Sidorenko, A. Yu., Petrov, Ya. A., & Stepanov, S. Yu. (2019). A mathematical model for processing spatially distributed heterogeneous geodata for making managerial decisions on laying optimal shipping routes for vessels in the Arctic. *Estestvennye i tekhnicheskie nauki [Natural and Technical Sciences]*, 4(130), 130–133 [in Russian].
13. Istomin, E. P., Kirsanov, S. A., Sokolov, A. G., & Kolbina, O. N. (2014). The phenomenon of geographic information management and the principles of its implementation. *Vestnik SPbGU [Bulletin of St. Petersburg State University]*, 7(4), 180–188 [in Russian].
14. Garcia Escalona, H. A., Istomin, E. P., & Kolbina, O. N. (2018). Prospects for the development of spatial data infrastructure using data using modern technologies. *Uchenye zapiski RGGMU [Scientific Notes of the Russian State Hydrometeorological University]*, 50, 130–136 [in Russian].
15. Stepanov, S. Yu., & Kolbina, O. N. (2016). The methodology of designing a geographic information system to support the adoption of managerial decisions based on the use of spatially distributed heterogeneous information. *Informatsionnye tekhnologii i sistemy: upravlenie, ekonomika, transport, pravo [Information Technologies and Systems: Management, Economics, Transport, Law]*, 2(18), 16–21 [in Russian].
16. Bugakov, P. Yu., & Kolesnikov, A. A. (2019). Analysis of the functionality of office applications for visualization and assessment of geodata. *Vestnik SGUGiT [Vestnik SSUGT]*, 24(4), 104–119 [in Russian].
17. Flegontov, A. V., Voronov, G. B., Smirnov, V. N., & Zadubina, G. A. (2019). Cartographic support of ground-based navigation systems. *Vestnik SGUGiT [Vestnik SSUGT]*, 24(3), 106–118 [in Russian].

18. Vagizov, M. R., Navalikhin, S. V., & Baenguev, B. A. (2017). Development of a geoinformation system for the improvement of public green spaces in St. Petersburg. *Fundamental'nye issledovaniya [Basic Research]*, 11-1, 35–40 [in Russian].
19. Novikov, V. V., Istomin, E. P., & Sokolov, A. G. (2019). The conceptual model of geoinformation management of natural-technical systems and territories. *Peterburgskiy ekonomicheskij zhurnal [Petersburg Economic Journal]*, 2, 55–63 [in Russian].
20. Zisman, A., & Kramer, J. (1995). Towards Interoperability in Heterogeneous Database Systems. Imperial college Research Report No. Doc 95/11, December, 1995.
21. Istomin, E. P., Kolbina, O. N., & Garcia Escalona, H. A. Database of risk assessment of extreme events in Venezuela. Patent of Russia No. 2018620801 dated June 04, 2018.

Received 09.06.2020

© O. N. Kolbina, E. P. Istomin, N. V. Yagotinceva, M. V. Vagizov, 2021