DOI: 10.17117/ns.2015.01.089

Поступила (Received): 24.01.2015

<http://ucom.ru/doc/ns.2015.01.089.pdf>

Kafedra_pi@rshu.ru

**Истомин Е.П., Колбина О.Н.,
Степанов С.Ю., Сидоренко А.Ю.
Механизмы моделей и систем интеграции
разнородных геопроостранственных данных**

**Istomin E.P., Kolbin O.N., Stepanov S.Yu., Sidorenko A.Yu.
Mechanisms models and systems of integration
of heterogeneous geospatial data**

В статье представлено описание технологий интеграции гетерогенных источников данных. Рассматриваются подходы к организации систем управления разнородными геоданными. Приведена возможность использования промежуточных источников данных и формирования основы единого хранилища пространственной информации

The article presents a description of the technology of integration of heterogeneous data sources. Examines approaches to systems management of heterogeneous GEODATA. Given the possibility of using interim data sources and form the basis of a single repository of spatial information

Ключевые слова: база данных, модель

Key words: database, model

Истомин Евгений Петрович

Доктор технических наук, профессор, декан
Российский государственный
гидрометеорологический университет
г. Санкт-Петербург, Малоохтинский проспект, 98

Istomin Eugene Petrovich

Doctor of Engineering Sciences, Professor, Dean
Russian state hydrometeorological university
Saint-Petersburg, Malookhtinsky ave., 98

Колбина Ольга Николаевна

Старший преподаватель
Российский государственный
гидрометеорологический университет
г. Санкт-Петербург, Малоохтинский проспект, 98

Kolbina Olga Nikolaevna

Senior lecturer
Russian state hydrometeorological university
Saint-Petersburg, Malookhtinsky ave., 98

Степанов Сергей Юрьевич

Старший преподаватель
Российский государственный
гидрометеорологический университет
г. Санкт-Петербург, Малоохтинский проспект, 98

Stepanov Sergey Yurievich

Senior lecturer
Russian state hydrometeorological university
Saint-Petersburg, Malookhtinsky ave., 98

Сидоренко Артем Юсупович

Старший преподаватель
Российский государственный
гидрометеорологический университет
г. Санкт-Петербург, Малоохтинский проспект, 98

Sidorenko Artem Yusupovich

Senior lecturer
Russian state hydrometeorological university
Saint-Petersburg, Malookhtinsky ave., 98

Введение (Introduction)

Одной из основных задач технологии консолидации разнородных данных является обработка и последующий комплексный анализ, что позволяет получить качественно новые знания об объекте исследования в интересах принятия управленческих решений и прогнозирования различного рода ситуаций. Такая информационная среда может объединять многие типы информационных ресурсов: геоинформационные, электронные, цифровые и другие ресурсы.

В задачи исследования ставится рассмотрение некоторых моделей обработки гетерогенных данных [1, 2]. В список моделей и систем обработки гетерогенных данных для использования в прикладных ГИС вошли следующие системы:

1. Модель интеграции гетерогенных источников данных на основе рекурсивной декомпозиции.
2. Технологическая платформа для распределенных гетерогенных систем сбора и обработки данных.

Материалы и методы (Materials and Methods)

Для достижения цели исследования вышеописанных прикладных Геосистем, в работе были использованы такие методы, как: анализ, моделирование, классификация и аналогия.

Стадии и этапы исследования:

1. Проведение анализа исследуемой модели интеграции гетерогенных источников данных.
2. Исследование процесса консолидации данных.
3. Декомпозиция уровней модели.
4. Анализ внутренней инфраструктуры геомодели.
5. Проведение анализа исследуемой технологической платформы для распределенных гетерогенных систем.
6. Декомпозиция уровней модели.
7. Распределение нагрузки и уровня доступа к данным.
8. Подведение результатов исследования.

Результаты (Results)

В основе первой рассматриваемой модели интеграции данных, заложено поддержание унифицированного представления разнородных источников данных, так же управление ограничениями целостности, управление выполнением операций манипулирования данными и запросов, согласование данных из разных источников, возможность расширения и настройки на новые источники данных. Исследуемый подход основан на рекурсивной декомпозиции источников данных, при которой каждый источник информации последовательно разбивается на элементы данных, причем на каждом уровне рекурсивной вложенности данные и их описания представляются единообразно. Представленная модель позволяет осуществлять консолидацию различных источников данных на любом уровне посредством задания связей между произвольными элементами, ограничений целостности и допустимых операций. Проектная модель

представления источников данных, которая позволяет осуществлять многоуровневую интеграцию разнородных данных в единое геоинформационное пространство, обеспечивает поддержку ограничений целостности на любом уровне интеграции источников данных, а также поддержку структурных и ассоциативных связей между источниками данных на любом уровне интеграции. Обеспечивается возможность динамического изменения схемы данных, а также расширяемость системы за счет новых моделей данных [3].

Модель соединяет гетерогенные данные в единую систему, с общей схемой для реализации единого механизма управления данными, независимого от моделей данных отдельных источников. В основе модели данных – рекурсивное представление источников данных. Предлагаемый способ представления позволяет моделировать произвольно сложные источники данных, сводя их к все более простым элементам. Каждый источник данных представляется в виде набора пар (атрибут, источник данных), который разделяет данный источник на составные источники. Данные текущего уровня помещаются в атрибуты пар. Нижний уровень вложенности соответствует неделимому элементу данных. Промежуточные уровни источников данных представляют собой метаданные. Такой подход к моделированию обработки данных и метаданных осуществляется единообразно. Каждому источнику данных ставится в соответствие набор связей с другими источниками (любого уровня), набор ограничений целостности и набор допустимых операций.

Архитектура системы интеграции геоданных является многослойной, в ней можно выделить три уровня (рис. 1 – архитектура системы интеграции данных) [11]:

- уровень данных (I);
- уровень унифицированного представления данных (II);
- уровень логического представления данных (III).

На нижнем уровне системы располагаются источники данных и механизмы (драйверы) доступа к ним.

Рис. 1. Архитектура системы интеграции данных

Гетерогенные данные могут представлять собой объектно-ориентированные, объектно-реляционные базы данных, реляционные, сетевые, иерархические, электронные таблицы, XML-файлы, и др. Для доступа к разнородным данным могут использоваться такие механизмы, как ODBC, OLE DB, ADO.NET, либо иные специфические драйверы, которые поддерживают интерфейс выполнения манипуляций над источником данных, которые от него требуются в данной системе.

На уровне внутренней инфраструктуры реализуется физическое представление схемы интеграции ИД, средства по поддержанию схемы в согласованном состоянии и выполнения базовых операций. Кроме этого, на данном уровне находится компонент, отвечающий за добавление в схему новых ИД. Схема данных представляет собой внутреннюю рекурсивную структуру моделей источников данных в соответствии с описанной выше математической моделью.

Менеджер транзакций осуществляет выполнение глобальных транзакций в системе. Перед выполнением транзакции в объекте сохраняется информация о состоянии источников гетерогенных данных, которые будет охватывать транзакция, до ее начала. Это необходимо для осуществления процедуры возврата в случае, если будет выполнено условие отката транзакции. На глобальном уровне транзакции выполняются последовательно и не могут пересекаться.

Во время выполнения операций глобальных транзакций, а также периодически с некоторым интервалом времени происходит процесс контроля менеджером ограничений целостности. Выполнение глобальных транзакций может состоять из произвольных операций манипулирования источниками данных, что может привести к рассогласованию данных в силу нарушения установленных ограничений целостности. Данный компонент осуществляет проверку после выполнения каждой базовой операции над источником данных на наличие нарушений ограничений целостности [4]. В случае, если нарушение было зафиксировано, то осуществляется возврат операции и, по возможности, всей транзакции. Периодическая проверка выполнения ограничений целостности необходима на случай, если в разнородных данных произошли локальные транзакции под управлением самих источников, что привело к нарушению ограничений целостности, установленных в данной схеме. При обнаружении такого нарушения пользователю сообщается об этом и работа с этим источником данных становится невозможной до тех пор, как согласованность данных на глобальном уровне не будет восстановлена.

Менеджер выборки данных отвечает за соединение источников данных и представление данных в виде, отвечающем потребностям пользователя. На вход компонент получает запросы, сформулированные в терминах сущностной схемы, а на выходе выдает результат выборки данных из разнородных данных.

Менеджер импорта осуществляет импорт в систему схем данных определенного типа. В основе работы компонента лежат алгоритмы декомпозиции источников данных. Расширение спектра поддерживаемых системой моделей данных требует добавления средств интерпретации данных моделей в этот компонент.

Внешнее представление – это логическое представление интегрированной схемы, отражающее сущности и взаимосвязи конкретных предметных

областей. Схема сущностей обеспечивает работу с системой в терминах предметной области. Компонент обеспечивает высокоуровневое представление источников данных, позволяющее решать задачи предметной области, а не тратить ресурсы на интерпретацию схемы интеграции. Построение подобной схемы сущностей, по сути, означает определение отображения схемы источников данных на предметно-ориентированную схему сущностей. Менеджер схем сущностей позволяет создавать и редактировать новые схемы сущностей и устанавливать отображение между источниками данных и сущностями.

Рассмотренная модель консолидации разнородных данных, позволяет осуществлять многоуровневую интеграцию гетерогенных источников данных в единое информационное пространство, так же обеспечивает поддержку ограничений целостности на любом уровне интеграции источников данных; поддержку структурных и ассоциативных связей между источниками данных на любом уровне интеграции.

Преимущества модели:

- интеграция гетерогенных данных на любом уровне декомпозиции;
- динамическое изменение схемы данных;
- расширяемость системы за счет новых моделей данных;
- универсальное представление связей, ограничений целостности и операций над данными.

В основе следующей модели рассматривается технологическая платформа для распределенных гетерогенных систем сбора и обработки данных. В данной работе проводится анализ разработки платформы для систем сбора данных и принципов ее построения.

Отсутствие в требованиях разрабатываемой системы конкретных источников данных позволяет говорить именно о разнородной технологической платформе. Под разнородностью понимается возможность взаимодействия с различными по количеству и качеству источниками данных, например, датчики на измерительных станциях выдают данные в разных форматах.

В основе технологической платформы лежит система сбора гидрологических данных, на примере которой предпринимается попытка выделения общих черт таких систем и основных проблем при их реализации.

Требования к платформе:

1. Организация сбора гидрометеорологических и гидрохимических данных из различных источников.

- Датчики, сбор с которых ведется в реальном времени.
- Архивы, синхронизация с которыми осуществляется по запросу пользователя.
- Ручной ввод данных, сбор которых не может быть организован автоматически.

2. Организация единого распределенного хранилища данных.

3. Организация многопользовательского удалённого доступа к собранным данным. Потребителями данных могут быть как средства визуализации и мониторинга в реальном режиме времени, так и различного рода ПО для проведения расчетов.

Общая структура системы представлена на рис. 2 – архитектура системы. Структура системы несколько уточняет классическую трехуровневую систему управления и сбора данных и содержит уровень объектов мониторинга, уровень сбора данных, уровень хранения и подготовки данных, сервисный уровень для поддержания клиентских предложений.

Рис. 2. Архитектура системы

Распределенное хранилище подразумевает распределение нагрузки между дублирующими друг друга службами. Алгоритм получения данных и распределения нагрузки представлен в виде диаграммы последовательности UML и показан на рис. 3 – распределение нагрузки и разграничение доступа.

Пользователь направляет запрос службе авторизации и распределения нагрузки. В случае успешной авторизации он получает адрес службы обработки данных и маркер, по которому остальные службы смогут его идентифицировать. Одновременно с этим служба обработки данных и разграничения доступа имеющая условно наименьшее количество активных соединений, получает тот

же маркер и список прав доступа пользователя и запускает таймер для проверки активности пользователя.

Рис. 3. Распределение нагрузки и разграничение доступа

Далее клиент формирует запрос на данные и передает его по полученному адресу вместе со своим маркером. В случае наличия у службы маркера переданного маркера, клиент считается авторизованным и таймер его активности сбрасывается. В случае отсутствия запросов с данным маркером в течение некоторого времени, сеанс считается завершенным и маркер удаляется, а служба авторизации уведомляется о количестве активных клиентов.

При получении запроса с определенным маркером служба распределения нагрузки проверяет список прав доступа, запрашивает данные из БД и формирует ответ в соответствии с запрашиваемым форматом данных и правами пользователя. Далее запросы обрабатываются до тех пор, пока маркер пользователя не будет удален.

Развертывание и регистрация в системе дополнительных служб обработки данных и распределения доступа позволит масштабировать систему в целом. Каждая из таких служб может находиться на выделенном сервере [5].

Следует отметить, что равномерное распределение клиентов по службам не всегда эффективно. Такое алгоритм распределения нагрузки уместен только в том случае, если все службы имеют одинаковую производительность. Что справедливо только в одном случае – все службы запущены на машинах с одинаковой производительностью (или на одной машине) и имеют одинаковую конфигурацию. В остальных же случаях нужно вести речь об условном весе каждого клиента и условном весе службы. Вес клиента характеризует сложность и частоту отправляемых им запросов, а вес службы характеризует ее производительность относительно некоторой эталонной службы [6]. Так же будет полезным реализация механизма динамического формирования веса в процессе взаимодействия с ним. Что позволит наиболее эффективно распределять нагрузку между службами.

Обсуждение и заключение (Discussion and Conclusion)

Достоинствами такой платформы является обеспечение требований масштабируемости и надежности на уровне архитектуры и модели вычислений, также независимость от предметной области и конкретного вида данных до определенного уровня абстракции. Говоря о недостатках модели следует указать недостаточную проработанность механизма взаимосвязи системы с различными видами датчиков на низком уровне.

Описанные выше технологии реализации консолидации гетерогенной информации из различных источников требуют получения ответов на следующие вопросы:

- как может быть решена проблема идентификации данных об одном и том же объекте в различных источниках данных;
- каким образом можно справиться с разнообразием структур представления данных в различных источниках.

В задачах устойчивого развития территорий, необходимо обрабатывать различные данные: аэрокосмические, данные наземных измерений, счетчики и другую разноформатную информацию. Это накладывает существенные особенности на все процессы обработки данных и применения прикладных инструментов, используемых для этого [7]. Часть проблем отражена в ряде публикаций [8, 9, 10].

Предложенные в данной работе общие структуры моделей интегрирования гетерогенных данных могут быть взяты за основу при создании подобных моделей в других областях применения пространственных данных.

Список используемых источников:

1. Колбина О.Н., Зоринова Е.М., Истомин Е.П., Степанов С.Ю. Обзор технологических методов доступа к базам данных распределенных ГИС // Информационные технологии и системы: управление, экономика, транспорт, право. Вып. 3(14). СПб: Андреевский издательский дом, 2014. 116 с.
2. Степанов С.Ю., Гришин Н.М., Колбина О.Н., Соловьева А.Д. ГИС. Обзор программного обеспечения // Информационные технологии и системы: управление, экономика, транспорт, право. Вып. 3(14). СПб: Андреевский издательский дом, 2014. 116 с.
3. Управление федеральной службы государственной регистрации, кадастра и картографии – информационный портал.
URL: http://www.to33.rosreestr.ru/upload/to33/files/Концепция_ситуационного_центра_ABT

4. Жижимов О.Л., Молородов Ю.И., Пестунов И.А., Смирнов В.В., Федотов А.М. Интеграция разнородных данных в задачах исследования природных экосистем // Вестник НГУ. 2011. Т. 9. Вып. 1.
5. Кашников А.В. Интеграция гетерогенных источников данных на основе рекурсивной декомпозиции // Information Technologies & Knowledge. Vol. 5. № 3. 2011.
6. Ильин Е.В. Технологическая платформа для распределенных гетерогенных систем сбора и обработки данных // Студент и научно-технический прогресс: Информационные технологии. Новосиб. гос. ун-т. Новосибирск, 2012.
7. Степанов С.Ю. Сравнительный анализ открытых геоинформационных систем // Информационные технологии системы: управление, экономика, транспорт, право. Вып. 10. СПб: Андреевский издательский дом, 2013. 156 с.
8. Котиков П.Е. Варианты построения темпоральных баз данных в геоинформационных системах // Научный аспект. № 4. 2014. Самара: Аспект, 2014. №4. С. 118-120.
9. Котиков П.Е., Муратова О.Н. Существующие сложности перехода к применению пространств данных в геоинформационных системах // Актуальные проблемы технических наук. Уфа: Аэтерна, 2015. С. 42-44.
10. Котиков П.Е. Некоторые решения технологических проблем хранения пространственных данных ГИС // Научный аспект. № 2. 2015. Самара: Аспект, 2014. №2. Т. 2. С. 212-214.
11. Котиков П.Е., Муратова О.Н. Обзор вариантов архитектурных решений построения распределенных и параллельных баз данных для геоинформационных систем // Закономерности и тенденции развития науки. Уфа: Аэтерна, 2015. С. 52-54.

References:

1. Kolbina O.N., Zharinova E.M., Istomin Y.P., Stepanov S.Yu. Review of technological methods of access to databases distributed GIS. Information technology and systems: management, economy, transport, right. Vol. 3(14). SPb: Andriivskiy, 2014. 116 p.
2. Stepanov S.Y., Grishin N.M. Kolbin O.N., Soloveva A.D. GIS. Software review. Information technology and systems: management, economy, transport, right. Vol. 3(14). SPb: Andriivskiy, 2014. 116 p.
3. The Federal service of state registration, cadastre and cartography and information portal. URL: http://www.to33.rosreestr.ru/upload/to33/files/Концепция_ситуационного_центра_ABT
4. Zhizhimov O.L., Molorodov I., Pestunov I.A., Smirnov V.V., Fedotov A.M. Integration of heterogeneous data in study of natural ecosystems. Vestnik NSU. 2011. Vol. 9.
5. Kashnikov A.V. Integration of heterogeneous data sources based on a recursive decomposition. Information technologies and knowledge. Vol. 5. Issue 3. 2011.
6. Ilyin E.V. Technology platform for distributed heterogeneous systems of data collection and processing. Student and scientific-technical progress: Information technologies. Novosibirsk. state University-Novosibirsk, 2012.
7. Stepanov S.Y. Comparative analysis of open geographic information systems. Information technology systems: management, economy, transport, right. Vol. 10. St. Petersburg: Andrew Izdatelsky house. 2013.156 p.
8. Kotikov P.E. Options for the design of temporal databases in geoinformacionnyh systems. Scientific aspect №4. 2014. Samara: Aspect, 2014. № 4. P. 118-120.
9. Kotikov P.E., Muratova O.N. The Existing difficulties of the transition to the use of data spaces in geographic information systems. Actual problems of technical Sciences. Ufa: Aeterna, 2015. P. 42-44.
10. Kotikov P.E. Some of the technological problems of spatial data storage GIS. Scientific aspect. № 2. 2015. Samara: Aspect, 2014. № 2. Vol. 2. P. 212-214.
11. Kotikov P.E., Muratova O.N. Overview of architectural solutions for building distributed and parallel data bases for geographic information systems. Regularities and tendencies of development of science. Ufa: Aeterna, 2015. P. 52-54.