

Grid технологии в системах мониторинга окружающей среды

Н.Н. КуССуль¹, Е.А. Лупян², А.Ю. Шелестов¹, Л. Глухи³, П. Копп⁴

¹*Институт космических исследований НАНУ-НКАУ
03680 Киев, Украина, проспект Глушкова 40, корп. 4/1
E-mail: inform@ikd.kiev.ua*

²*Институт космических исследований РАН
117997 Москва, ул. Профсоюзная, 84/32
E-mail: evgeny@d902.iki.rssi.ru*

³*Institute of Informatics, Slovak Academy of Sciences
Dubravska cesta 9, 84507 Bratislava, Slovakia
E-mail: hluchy.ui@savba.sk*

⁴*Centre National d'Etudes Spatiales (CNES)
18 avenue Edouard Belin, F-31401 Toulouse Cedex 9, France
E-mail: paul.kopp@cnes.fr*

В работе рассматриваются вопросы интеграции систем спутникового мониторинга. Выделены два возможных способа взаимодействия подобных систем: на уровне обмена данными и на уровне выполнения задач. Сформулированы требования к реализации, а также проанализированы преимущества и недостатки каждого из подходов.

Введение

Современный этап развития систем экологического мониторинга, прогнозирования чрезвычайных ситуаций и поддержки принятия соответствующих решений характеризуется глобализацией поставленных задач и активным использованием данных из разных источников, в первую очередь, спутниковых наблюдений. На решение таких задач не только в региональном, но и в глобальном масштабе, направлены международные многочисленные программы и инициативы, в том числе GEOSS, GMES и INSPIRE. Эти инициативы предполагают совместное использование продуктов и интеграцию региональных и национальных систем спутникового мониторинга. В данной статье анализируются возможные пути интеграции систем мониторинга. Детально рассматриваются проблемы, возникающие в процессе интеграции региональных или национальных систем, а также обсуждаются возможные пути их решения. Подавляющее большинство предлагаемых технологий реально апробировано в процессе создания международных систем спутникового мониторинга.

Подходы к интеграции систем

Возможны два подхода к интеграции систем мониторинга: на уровне обмена данными результатами их обработки либо на уровне совместного решения задач в рамках общей инфраструктуры. Первый подход гораздо проще в реализации. Для его воплощения требуется лишь обеспечение стандартизованного обмена данными между приложениями, наличие общего каталога метаданных, а также общего программного и Web-интерфейса. Такой подход применяется для интеграции систем спутникового мониторинга Института космических исследований НАНУ-НКАУ (ИКИ НАНУ-НКАУ) и Института космических исследований РАН (ИКИ РАН). Для обмена данными между системами используется протокол WMS, а для реализации интерфейса пользователя — шаблон на основе толстого клиента и программное обеспечение OpenLayers [1]. Реализация взаимодействия систем мониторинга ИКИ НАНУ-НКАУ и ИКИ РАН будет рассмотрена ниже.

Второй способ интеграции систем более сложный. Он предполагает взаимодействие систем на уровне решения задач, т.е. запуска приложений, использования отдельных моделей, совмест-

ного использования данных и вычислительных ресурсов. Помимо стандартизации обмена данными и наличия общего программного и пользовательского интерфейса в данном случае необходимо обеспечить единую политику безопасности, общую вычислительную инфраструктуру, согласованное планирование и запуск задач и мониторинг нагрузки. Такой подход возможен только на основе интеграции отдельных Grid-систем. Получаемую в результате инфраструктуру можно назвать Inter-Grid.

Предпосылки создания Inter-Grid системы

На данный момент в мире сложилась ситуация, когда разные космические агентства и организации, связанные с обработкой космических данных либо уже имеют собственные Grid-инфраструктуры (ESA, NASA, JAXA), либо находятся в процессе их создания (CNES, CNR, НАНУ и НКАУ). Активное развитие Grid-подхода в данной области обусловлено распределенностью космических данных. Кроме того, создание продуктов обработки космических данных разных уровней требует использования моделей, интеграции данных разной природы, и, как следствие, больших вычислительных мощностей, которые не всегда имеются в каждой отдельной организации [2].

Создание интегрированной Grid-инфраструктуры поддерживается международными инициативами, в том числе GEOSS и GMES. В области наук о Земле Grid-технологии активно развиваются в рамках EGEE (например, проект DEGREE, <http://eu-degree.eu/>). Инициирован ряд международных проектов, направленных на объединение Grid-систем. Среди них особо следует отметить проект Wide Area Grid (WAG), поддерживаемый рабочей группой по информационным системам и сервисам WGISS комитета CEOS, в котором активное участие принимает и НКАУ [3].

В Украине имеются все предпосылки для активного включения в международные интеграционные процессы. НКАУ и НАНУ инициировали совместную программу по созданию украинского сегмента GEOSS (система GEO-UA), информационная инфраструктура которой базируется на Grid-технологии (UASpaceGrid). Разрабатываемая система спутникового мониторинга основывается на международных стандартах обмена данными и допускает естественное объединение с другими стандартизованными системами [4]. Организован Grid-сегмент Национальной академии наук Украины (НАНУ), включающий в себя вычислительные мощности ведущих институтов НАНУ, в частности, Института теоретической физики им. Боголюбова НАНУ, Института кибернетики НАНУ (ИК НАНУ), Института космических исследований НАНУ-НКАУ (ИКИ НАНУ-НКАУ) и др. С 2007 года Украина является членом программы EGEE, что облегчает взаимодействие с другими Grid-системами в рамках этой программы.

В рамках реализации проекта WAG и нескольких других международных проектов в настоящее время создается Inter-Grid инфраструктура, объединяющая ресурсы Украины, Китая и ESA для совместного решения задач мониторинга наводнений на основе интеграции данных разной природы и усвоения их в модели.

Таким образом, приобретает актуальность задача изучения и разработки средств взаимодействия Grid-систем, которые в общем случае могут быть реализованными на разных платформах (Globus Toolkit, gLite, NorduGrid, Alien и т.д.).

Задачи обеспечения взаимодействия Grid-систем и способы их решения

При разработке средств взаимодействия Grid-платформ можно выделить ряд задач, решение которых позволит реализовать большую часть функциональности Grid и использовать высокоуровневые сервисы на ее основе с учетом специфики задач спутникового мониторинга. К таким задачам относятся следующие:

- обеспечение взаимодействия между системами безопасности разных Grid-платформ;
- реализация надежной передачи файлов между Grid-платформами;
- реализации высокоуровневого доступа к геопространственной информации.

— запуск и мониторинг задач на ресурсах разных Grid-платформ;

Для отработки методов решения вышеупомянутых задач было выбрано две Grid-платформы: Globus Toolkit v4 и gLite v3, поскольку абсолютное большинство других платформ в той или иной степени совместимы с ними.

Обеспечения взаимодействия систем безопасности разных Grid- платформ

На сегодняшний день системы безопасности всех распространенных Grid- платформ используют принципы инфраструктуры открытых ключей (Public Key Infrastructure, PKI) [5]. Наиболее распространенной реализацией является использование стандарта сертификатов X.509 [6], который поддерживается практически всеми существующими платформами для разработки ПО.

Для начала работы инфраструктуре открытых ключей новый пользователь может создать запрос на сертификат, который состоит из двух частей: собственно запроса с информацией о пользователе (имя, организация, подразделение и прочее) и закрытого ключа, недоступного никому, кроме пользователя. После создания открытая часть запроса передается в сертификационный центр (Certificate Authority, CA), чтобы убедиться в том, что пользовательские данные в запросе являются правильными. В случае корректности представленного запроса сертификационный центр подписывает его, используя механизм цифровой подписи. Результатом этой процедуры является сертификат, который передается пользователю по открытым каналам связи.

Вместе с закрытой частью ключа сертификат может использоваться для аутентификации, цифровой подписи и шифрования. Также существует возможность создания на основе пары «сертификат/закрытый ключ» новых сертификатов и ключей с ограниченным сроком действия, которые могут использоваться для временного делегирования полномочий пользователя другому пользователю или сервису.

Для того чтобы пользователь с сертификатом, подписанным определенным CA, мог получить доступ к некоторому Grid-узлу, между системой безопасности этого узла и CA пользователя должны быть установлены отношения доверия (trust). Это означает, что администрация этого Grid-узла (Grid-системы) убеждена в корректности процедур проверки лица пользователя, которые используются в данном CA.

Таким образом, существует несколько возможных путей для обеспечения базового взаимодействия между пользователями и узлами Grid-сегментов, которые используют разные Grid-платформы,

— Создать собственные CA, которые будут обслуживать потребности ассоциированных Grid-сегментов и установить отношения доверия между этими CA;

— Направить запросы сертификатов к общедоступным CA, предназначенным для обслуживания потребностей определенного региона или проекта. Сертификаты таких CA находятся в открытом доступе в соответствующих репозиториях. Примером такого репозитория может выступать сайт <http://www.eugridpma.org> проекта European Policy Management Authority for Grid Authentication. Этот репозиторий содержит сертификаты сертификационных центров, которые обслуживают потребности разнообразных Grid-проектов, которые реализовываются в Европе и других частях мира;

— Использовать комбинированный подход, при котором одни Grid-сегменты создают собственные CA, а другие используют открытые CA, политика которых отвечает направлению деятельности или региона расположения этих сегментов.

При налаживании взаимодействия систем безопасности Grid- сегментов UA Space Grid и RSGS Grid (Китай) были отработаны подходы 2 и 3 с использованием со стороны UA Space Grid как сертификатов собственного CA, созданного на основе ПО TinyCA, так и с использованием сертификатов, предоставленных Russian Data Intensive Grid CA (RDIG, <http://ca.grid.kiae.ru/RDIG/>), который является членом EUGridPMA и политика которого допускает

выдачу сертификатов научным организациям Украины.

Следует отметить, что Grid-платформы Globus Toolkit v4 и gLite v3 совместимы по форматам используемых сертификатов, но лишь частично совместимы по форматам описаний сертификационных политик, определяющих, каким именно пользователям разрешен доступ к ресурсам системы. Для преодоления этой несовместности необходимо использовать два разных способа указания идентификатора CA в файле описания политики.

Реализация надежной передачи файлов между Grid-платформами

Общепринятым средством передачи файлов между Grid-платформами является протокол GridFTP (<http://www-fp.mcs.anl.gov/dsl/GridFTP-Protocol-RFC-Draft.pdf>), представляющий собой расширение стандартного Internet-протокола передачи файлов FTP. Среди основных отличий GridFTP от FTP отметим следующие:

- интеграция с системой информационной безопасности Grid;
- возможность параллельной передачи данных несколькими потоками;
- возможность параллельной передачи данных от/к нескольким клиентам;

Некоторые Grid-платформы предоставляют более развитые средства передачи файлов. Так, например, Globus Toolkit v4 предоставляет сервис Reliable File Transfer, который, используя GridFTP на нижнем уровне, позволяет выполнять передачу файлов без вмешательства пользователя и с использованием делегирования полномочий. Однако интеграцию Grid-платформ следует проводить на нижнем уровне сервисов, которые в том или другом виде доступны для всех платформ.

Среди недостатков ПО GridFTP следует отметить сложность его развертывания и настройки в сетевой среде сложной структуры. Значительную проблему составляет использование в некоторых сетях механизмов трансляции адресов (Network Address Translation, NAT). Для преодоления этого препятствия необходимо вносить изменения как в настройках маршрутизации компьютерной сети, так и в настройках сервера GridFTP.

Grid-платформа gLite v3 содержит два сервера GridFTP с разными механизмами авторизации:

- GridFTP-сервер с использованием механизма авторизации Virtual Organization Membership System (VOMS) (<http://edg-wp2.web.cern.ch/edg-wp2/security/voms/voms.html>);
- GridFTP-сервер с использованием механизма авторизации Grid Mapfile, сопоставляющий пользователей Grid, которые идентифицируются посредством имени их сертификата, с локальными пользователями системы.

Эти серверы могут работать одновременно при условии использования ими разных TCP-портов. Для взаимодействия с платформой Globus Toolkit можно использовать оба подхода, однако применение сервера с VOMS-авторизацией требует использования такого типа авторизации от всех клиентов, что может быть очень сильным ограничением. Сервер с авторизацией через Grid Mapfile не накладывает таких ограничений и может свободно использоваться с другими системами авторизации.

Для отработки механизмов передачи файлов между UA Space Grid и RSGS Grid (Китай) был использован вариант с авторизацией через Grid Mapfile. Выполненные тесты показали успешную передачу файлов в обоих направлениях и с разными ролями узлов (сервер или клиент).

Высокоуровневый доступ к геопространственной информации

Главной особенностью Grid-систем спутникового мониторинга является использование геопространственных данных разного пространственного и временного разрешения. Поэтому высокоуровневый доступ к геопространственной информации является важнейшей задачей Inter-Grid систем, связанных с обработкой спутниковой информации. Эта функциональность может быть реализована двумя способами: через сервис WSRF и через контейнер OGSA-DAI.

Схема организации доступа к геопространственным данным через WSRF-сервис показана на рис. 1. Преимуществом такого подхода является простота реализации базовой функциональности

(при наличии соответствующих программных средств), а также простота развертывания. Однако дополнительная функциональность — обеспечение безопасности, индексирование и т.п. — должна быть реализована вручную. Кроме того, при таком подходе сложно обеспечить интеграцию с другим ПО обработки данных.

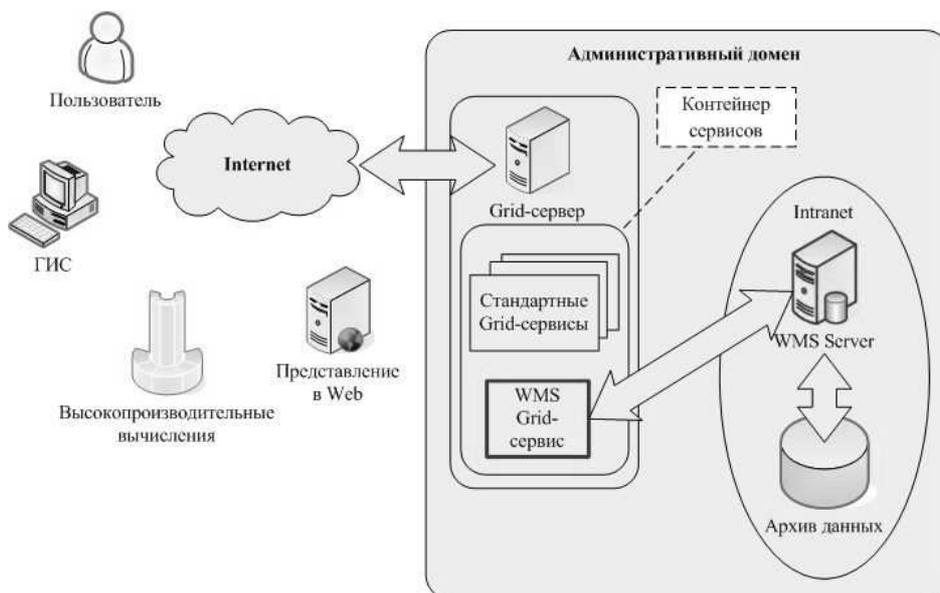


Рис. 1. Доступ к геопространственным данным через сервис WSRF

Второй подход к организации доступа к высокоуровневым данным состоит в использовании контейнера OGSA-DAI (рис. 2).

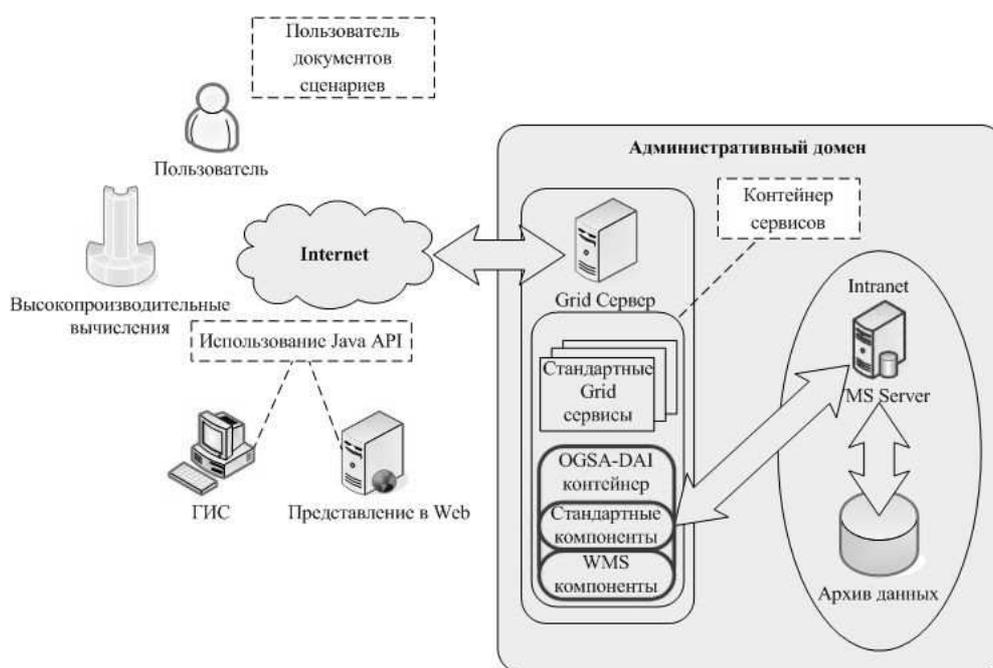


Рис. 2. Доступ к геопространственным данным через контейнер OGSA-DAI

При таком подходе большая часть проблем реализации решается автоматически, в том числе вопросы безопасности, надежность передачи данных между различными источниками. Инфраструктура является легко расширяемой и совместимой с другими системами обработки данных. Однако обработка исключений и добавление новых функций требует гораздо больше серьезных

усилий и навыков, чем при первом подходе. Кроме того, в этом случае требуется установка дополнительного программного обеспечения.

Запуск и мониторинг задач на ресурсах разных Grid-платформ

Представление задач к выполнению на вычислительных ресурсах является одной из важнейших задач Grid-систем, поскольку они часто рассматриваются как универсальное средство для доступа к высокопроизводительным системам. К сожалению, интерфейсы представления задач довольно сильно отличаются между разными Grid-платформами и задача их унификации требует значительных усилий.

В ходе исследования возможностей для представления задач на разных Grid-платформах было выявлено два возможных варианта реализации:

1. Использование Grid-портала на основе каркаса, который поддерживает разные Grid-платформы. Примерами таких каркасов являются GridSphere и P-GRADE Portal (рис. 3);

2. Использование высокоуровневого Grid-планировщика, который поддерживает разные Grid-платформы и предоставляет стандартный программный интерфейс (рис. 4).

Использование Grid-портала обладает преимуществом сравнительно легкого развертывания, настройки и управления. Недостатком порталного решения является предоставление только интерфейс уровня пользователя, но не программного интерфейса, который можно использовать для построения высокоуровневых сервисов.

Напротив, подход с использованием Grid-планировщика (в качестве которых рассматривались планировщики GridWay [7] и Karajan [8]) позволяет получить некоторый стандартный программный интерфейс, который может быть использован для программирования взаимодействия с этим планировщиком. ПО GridWay предоставляет внешние программные интерфейсы GRAM4 и DRMAA, а Karajan — специальный WSRF-сервис, который может быть использован для выполнения не только отдельных задач, а и их последовательностей (workflow). Недостатками подхода с использованием планировщиков являются отсутствие простого интерфейса пользователя и сложность развертывания и администрирования такой системы.

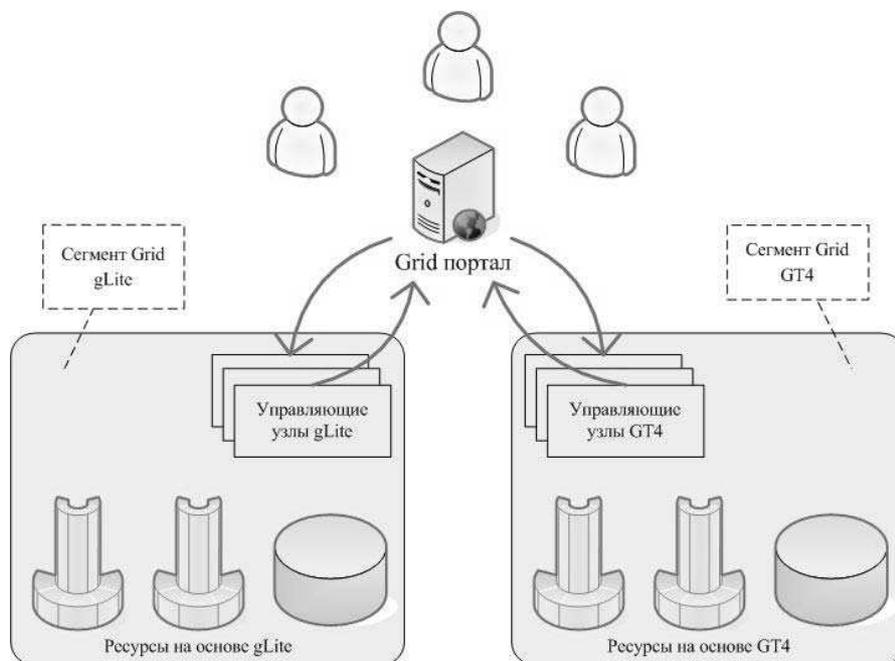


Рис. 3. Интеграция с помощью Grid-портала

Наиболее функциональным и удобным является комбинированный подход, который использует планировщик для предоставления стандартизированного интерфейса к вычислительным ре-

сурсам системы и Grid-портал, взаимодействующий с планировщиком через эти интерфейсы, для предоставления удобного Web-интерфейса пользователя.

Следует отметить, что эти же подходы могут использоваться и для обеспечения мониторинга и управления задачами, выполняемыми в разнородной Grid-среде, поскольку мониторинг и управление являются функциями, тесно связанными с представлением задач к выполнению.

Для отработки механизмов представления задач к выполнению был развернут Grid-портал (<http://gridportal.ikd.kiev.ua:8080/gridsphere>), предоставляющий доступ к ресурсам вычислительных и архивных ресурсов ИКД НАНУ-НКАУ, ИК НАНУ и RSGS CAS. Пример экрана портала приведен на рис. 5.

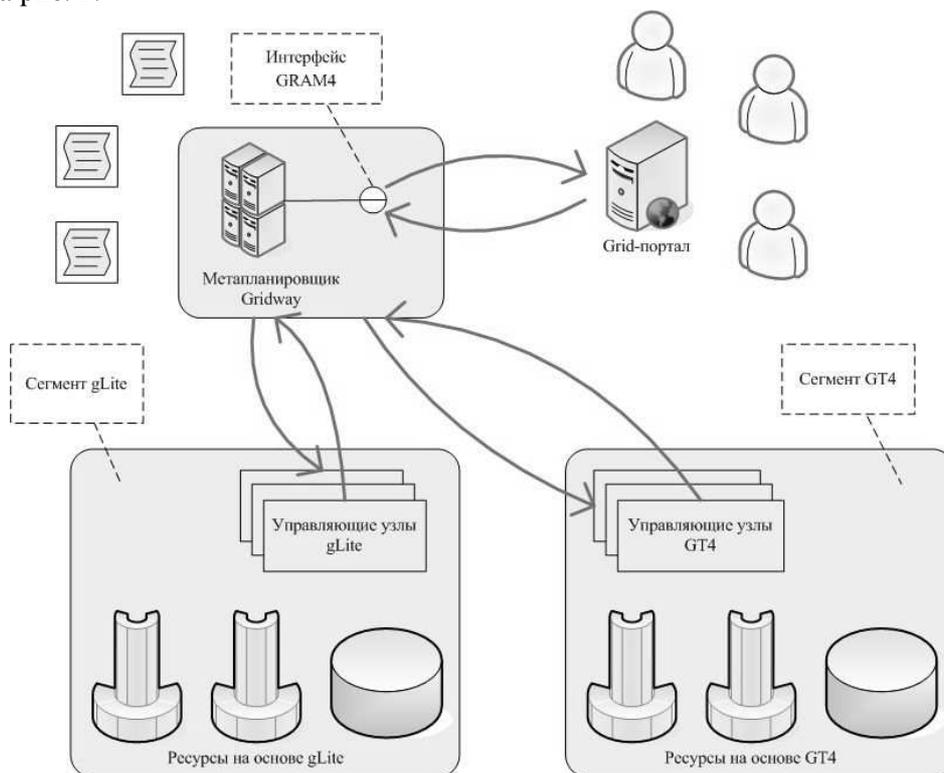


Рис. 4. Интеграция с помощью высокоуровневого планировщика

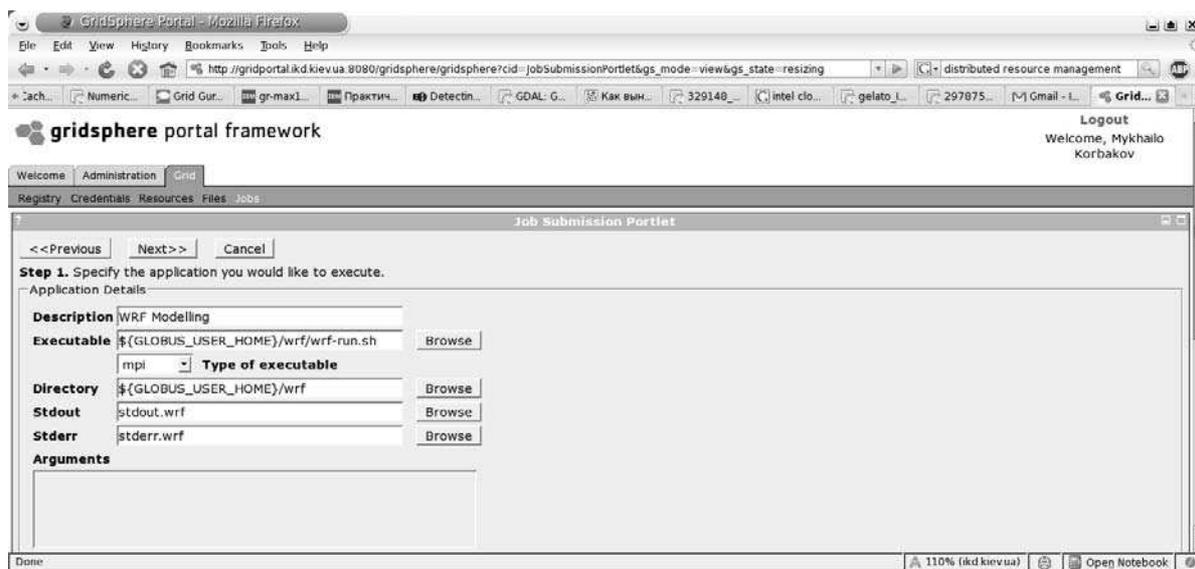


Рис. 5. Интерфейс представления задачи в Inter-Grid с помощью портала

Дальнейшие работы планируется развивать в направлении использования комбинированного подхода, который будет объединять преимущества как Grid-портала, так и Grid-планировщика.

Пример интеграции систем мониторинга на уровне данных

Первый подход к интеграции систем спутникового мониторинга — на уровне обмена данными — апробирован при отработке взаимодействия между системами ИКИ НАНУ-НКАУ и ИКИ РАН на примере задач сельскохозяйственного мониторинга.

Российская система спутникового мониторинга на основе стандарта WMS предоставляет данные по индексу NDVI, пожарам и базовую картографию. В домене ИКИ НАНУ-НКАУ развернуты 2 WMS-сервера, обеспечивающие продукты MODIS (точные данные) и модельные данные по температуре земной поверхности, а также векторные данные наземных метеоизмерений.

Интеграция данных реализована по шаблону толстого клиента на базе ПО OpenLayers с поддержкой технологии AJAX. Схема организации взаимодействия показана на рис. 6.

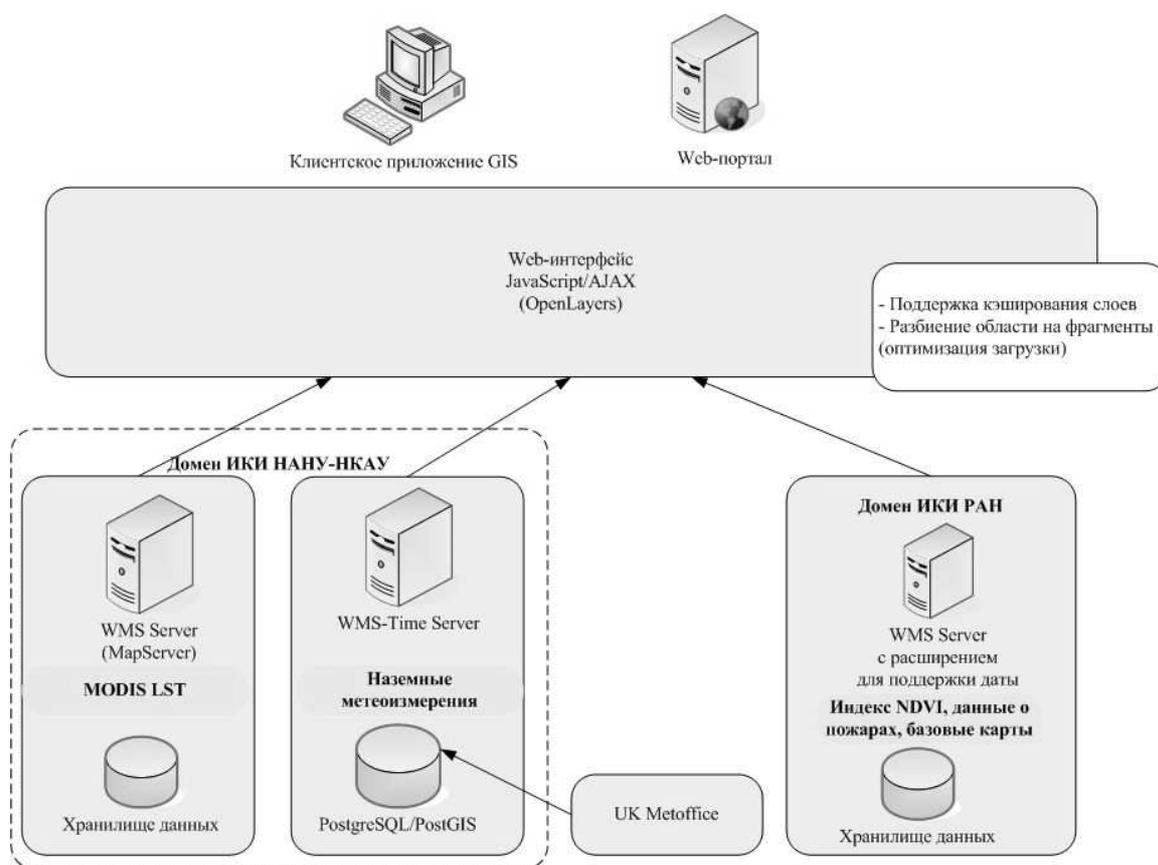


Рис. 6. Схема интеграции систем мониторинга на уровне данных

Выбранная технология обладает следующими преимуществами: обеспечивается поддержка кэширования слоев и оптимизация загрузки изображений за счет разбиения всей отображаемой области на фрагменты. Это существенно ускоряет отображение результатов в интерфейсе пользователя. Дополнительным преимуществом такой технологии является отсутствие необходимости в серверной части, обеспечивающей интеграцию данных. Вся функциональность по извлечению данных из разных систем мониторинга и их отображению реализуется на стороне клиента.

Пример интерфейса пользователя показан на Рис. 7, где одновременно отображены данные обеих систем мониторинга: карта индексов NDVI из системы мониторинга ИКИ РАН и карта температуры земной поверхности с сервера ИКИ НАНУ-НКАУ.

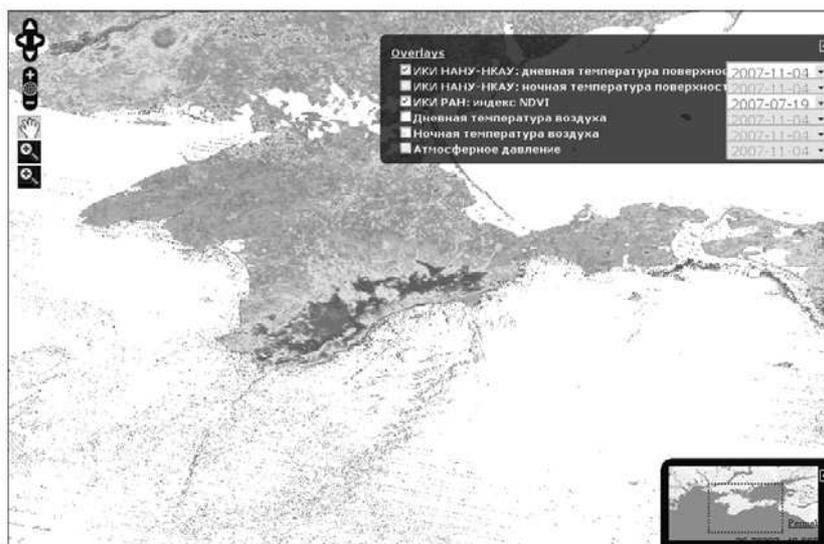


Рис. 7. Интерфейс пользователя на основе OpenLayers

Пример интеграции систем мониторинга в рамках Inter-Grid

Второй подход к интеграции систем мониторинга — на уровне выполнения задач — был использован для разработки сегмента Wide Area Grid (WAG) в рамках проекта, инициированного Французским космическим агентством CNES, а также проекта CAT-1. Целью этих проектов является создание Grid- системы, объединяющей ресурсы космических агентств и других организаций разных стран с целью решения задач GEOSS и GMES. В разработанном сегменте объединены ресурсы ИКИ НАНУ-НКАУ, ИК НАНУ и Remote Sensing Ground Station of CAS (China) [9].

Текущее состояние Inter-Grid системы, разрабатываемой в рамках проектов CAT-1 и WAG, представлено на рис. 8. Все организации, представленные в данном сегменте, предоставляют свои вычислительные ресурсы для решения задач мониторинга окружающей среды в контексте GEOSS [10]. На данный момент в Inter-Grid-системе выполняются следующие задачи [11]: моделирование метеорологических параметров с помощью численной модели WRF; картографирование площадей затопленных территорий с помощью радиолокационных данных спутников ERS и ENVISAT.

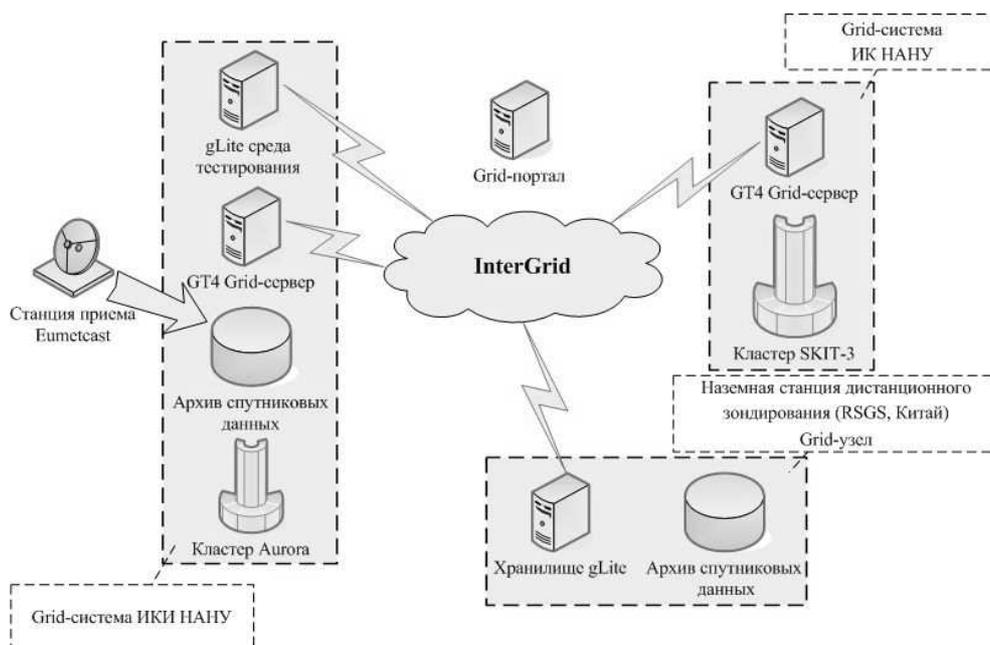


Рис. 8. Текущая InterGrid инфраструктура, создаваемая в рамках проектов CAT-1 и WAG

Кроме вычислительных ресурсов, ИКИ НАНУ-НКАУ и RSGS of CAS имеют возможность представлять в сегменте свои архивы данных ДЗЗ, которые могут быть прозрачно использованы для решения тематических задач.

Выводы

В данной работе рассмотрены вопросы интеграции систем спутникового мониторинга. Выделены два возможных способа взаимодействия подобных систем: на уровне обмена данными и на уровне выполнения задач. Сформулированы требования к реализации, а также проанализированы преимущества и недостатки каждого из подходов.

Более глубокую интеграцию систем обеспечивает их объединение на уровне выполнения задач, т.е. в рамках Inter-Grid инфраструктуры. Выделены основные проблемы, возникающие в процессе совместного использования Grid-систем на разных платформах, и проанализированы возможные пути их решения.

Практически все предлагаемые технические решения успешно апробированы при объединении систем мониторинга ИКИ НАНУ-НКАУ и ИКИ РАН на уровне данных, а также при построении Inter-Grid системы в рамках проектов Wide Area Grid и CAT-1, объединяющего ресурсы Украины, Европейского космического агентства и RSGS CAS (Китай). Созданный сегмент Inter-Grid можно рассматривать в качестве прототипа информационной инфраструктуры системы GEOSS.

Работа выполнена при поддержке грантов INTAS-CNES-NSAU "Data Fusion Grid Infrastructure" (Ref. Nr 06-100024-9154) и Европейского космического агентства (ЕКА) Category-1 "Wide Area Grid Testbed for Flood Monitoring using Spaceborne SAR and Optical Data" (№4181).

Литература

1. *Shelestov A., Kravchenko O., Ilin M.* Distributed visualization systems in remote sensing data processing GRID // International Journal "Information Technologies and Knowledge". 2008. Volume 2 (in print).
2. *Fusco L., Goncalves P., Linford J., Fulcoli M., Terracina A., D'Acunzo G.* Putting Earth-Observation on the Grid // ESA Bulletin. 2003. 114. P. 86-91.
3. *Куссуль Н.Н., Лупян Е.А., Шелестов А.Ю. и др.* Grid технологии в системах мониторинга окружающей среды // Пятая Юбилейная Открытая Всероссийская конференция «Современные проблемы дистанционного зондирования Земли из космоса» Москва, ИКИ РАН, 12-16 ноября 2007 г.
4. *Kussul N., Shelestov A., Korbakov M., et al.* Grid Infrastructure for Satellite Data Processing in Ukraine // International Journal "Information Technologies and Knowledge". 2008. Volume 2.
5. *Ed Gerck*, Overview of Certification Systems: x.509, CA, PGP and SKIP, The Black Hat Briefings — 1999, <http://www.securitytechnet.com/resource/rsc-center/presentation/black/vegas99/certover.pdf>
6. ITU Recommendations – X.509 : Information technology - Open Systems Interconnection - The Directory: Public-key and attribute certificate frameworks, 2005.
7. *Vázquez-Poletti J.L., Huedo E., Montero R.S., Llorente I.M.* A Comparison Between two Grid Scheduling Philosophies: EGEE WMS and GridWay // Multiagent and Grid Systems, Vol. 3, № 4, 2007.
8. *Laszewski G., Hategan M., Kodeboyina D.* Java CoG Kit Workflow // Workflows for e-Science Scientific Workflows for Grids, 2007.
9. *Shelestov A.* Workflow Modelling in Grid System for Satellite Data Processing// International Journal on Information Theory and Applications. 2008. Vol. 15.
10. *Kussul N., Hluchy L., Kopp P., Lupian E.* Grid Infrastructure for Data Fusion Problems// Сборник тезисов Седьмой Украинской конференции по космическим исследованиям. НЦУИКС, Евпатория, Украина. 2007. С. 206.
11. *Kopp P., Petiteville I., Shelestov A., Li G.* Wide Area Grid (WAG) // Сборник тезисов Седьмой Украинской конференции по космическим исследованиям. НЦУИКС, Евпатория, Украина. 2007. С. 209.