

Разработка методов и технологий для распределенного хранения и обработки данных дистанционных наблюдений окружающей среды

С. И. СМАГИН¹, Е. А. ЛУПЯН², А. А. СОРОКИН¹, М. А. БУРЦЕВ², С. П. КОРОЛЁВ^{1,*},
А. А. ПРОШИН², Л. С. КРАМАРЕВА³

¹Вычислительный центр ДВО РАН, Хабаровск, Россия

²Институт космических исследований РАН, Москва, Россия

³Дальневосточный Центр ФГБУ “Научно-исследовательский центр космической гидрометеорологии “Планета”, Хабаровск, Россия

*Контактный e-mail: serejk@febras.net

При исследовании и разработке современных методов и алгоритмов хранения, обработки и анализа сверхбольших объемов данных дистанционного зондирования Земли используются возможности облачных вычислительных сред. Создаются программные средства и инструменты, позволяющие формировать и поддерживать распределенные архивы данных дистанционных наблюдений и результатов их обработки, предоставляющие для пользователей (в том числе различных специализированных информационных систем) удобные сервисы для доступа и анализа данных.

В статье рассмотрены особенности решения перечисленных задач в России, дана оценка возможных технологических средств, позволяющих повысить эффективность и скорость операций с данными, приведены первые результаты работ по данному направлению.

Ключевые слова: облачные технологии, наборы научных данных, база данных, дистанционное наблюдение, информационная система, распределенная обработка и хранение данных, RESTful web-сервис, окружающая среда.

Введение

Системы дистанционного зондирования Земли являются одним из основных источников объективной информации для оценки и контроля состояния окружающей среды и рационального природопользования в различных регионах планеты. С увеличением числа специализированных космических аппаратов и их функциональных возможностей практически взрывным образом растут и объемы принимаемых инструментальных данных. Для их использования в решении крупномасштабных научных задач (например, таких как исследования вулканической активности, изучение изменения растительного покрова и т. п.) в последние годы быстрыми темпами развиваются технологии обработки и методы анализа данных дистанционных наблюдений [1–5]. Так, территориально-распределенная система ФГБУ НИЦ “Планета” ежедневно принимает с 16 за-

рубежных и отечественных спутников наблюдения Земли более 1 Тбайта инструментальных данных, формируя в том числе с использованием технологий и систем, разработанных в Институте космических исследований РАН (ИКИ РАН), порядка 350 видов информационных продуктов. При этом общий объем оперативных и долговременных архивов составляет порядка 0.6 Пбайта.

Востребованность этих данных высока, тем более в нынешней ситуации, когда меняются подходы к проведению научных работ и отмечается растущая тенденция к выполнению междисциплинарных исследований. Ученым необходим доступ к большим наборам неоднородных данных, полученным и хранящимся в разных географических точках, системах и центрах информации, в различных стандартах и форматах данных [6, 7]. Однако создание технологий и программных средств для работы с такого рода данными, ведения сверхбольших архивов данных, их распределенной обработки и использования в определенной мере запаздывает. Это снижает возможность и эффективность применения данных дистанционного зондирования Земли (ДЗЗ) результатов и их обработки в решении крупномасштабных научных задач и проблем.

В статье рассматриваются возможные методы и подходы для создания алгоритмов и технологий работы с постоянно пополняющимися распределенными архивами на примере действующей информационной инфраструктуры ИКИ РАН, НИЦ “Планета” и ВЦ ДВО РАН.

1. Формулировка проблемы

Для эффективной работы с данными дистанционного зондирования необходимо наличие специальной системы, которая решала бы технологические задачи, связанные со сбором, унификацией данных, формированием из них структурированных архивов и других тематических наборов данных, и предоставляла для исследователей реализованные в виде программных средств инструменты для их анализа.

В решении рассматриваемых задач сейчас невозможно идти традиционным путем, когда под каждую научную задачу формируются отдельные, постоянно пополняемые архивы данных из всех доступных источников и разрабатываются средства для работы с ними. Такой подход обладает рядом существенных недостатков. Необходимо развивать и поддерживать масштабные технологические сервисы под каждый, даже небольшой, научный проект. Помимо этого, требуются значительные вычислительные ресурсы для обеспечения хранения и обработки информации, а также для функционирования большого количества программного обеспечения. Поэтому все чаще разработчики обращаются к облачным технологиям [8–10], переводя решение ресурсоемких задач в виртуальную вычислительную среду, взаимодействие с которой для прикладных информационных систем организуется через гибкие RESTful web-сервисы.

В качестве примеров реализованных систем для работы с данными дистанционных наблюдений, построенных с использованием таких подходов и технологий, можно привести опыт NASA и NOAA. Для данных дистанционного зондирования, формируемых в NASA, создана система обработки в режиме, близком к реальному времени [11]. Она основана на возможностях провайдера Amazon Web Services, включающих Simple Storage Service (Amazon S3), Amazon Elastic Block Store (EBS), а также облака Amazon Elastic Cloud Compute (EC2), предоставляющих серверы данных и картографических приложений. Созданная система позволяет проводить масштабируемую параллельную обработку данных в реальном времени на большом числе виртуальных вычислительных

узлов. Компания Amazon совместно с The Weather Company, Esri, Planet OS реализовала проект Public Data Sets (<http://aws.amazon.com/public-data-sets/landsat/>), подразумевающий доступ к сверхбольшим архивам спутниковых снимков NOAA.

В последние годы стали предприниматься попытки проведения работ в этой области и в нашей стране. К ним можно отнести решения, созданные авторами статьи. Все они базируются на использовании распределенной инфраструктуры приема спутниковой информации НИЦ “Планета” и Роскосмоса. В частности, разработан ряд систем для ведения и сопровождения архивов сверхбольших данных [12–18] и их поставки в специализированные распределенные информационные сервисы, такие как “Вега” [1], Вега-ДВ [2], VSV [3], STS [5] и т. д. В качестве дополнительных источников данных сформированы наборы инструментальных данных от специализированных наземных сетей наблюдений, для которых организовано облачное хранилище на базе платформы Openstack Swift [19, 20] и созданы RESTful web-сервисы для доступа к информации из специализированных систем, таких как АИС “Сигнал” [21–23] и VSV.

Общемировой опыт и, в частности, опыт авторов показал, что перечисленные подходы являются жизнеспособными и эффективными. С опорой на них начаты исследования и разработка современных методов и алгоритмов хранения, обработки и анализа данных в облачных вычислительных средах на примере накопленных сверхбольших архивов данных ИКИ РАН, НИЦ “Планета” и ВЦ ДВО РАН с общим объемом более 1 Пбайта.

Указанная работа подразумевает создание программной платформы, которая предоставит инструментарий, позволяющий прозрачно для пользователя (или внешней информационной системы) вести поиск и получать необходимые инструментальные данные и результаты их обработки из сотен тематических серверов и хранилищ, распределенных по территории России.

2. Методы и технологии решения проблемы

Выполнение рассматриваемой работы связано с созданием технологий по следующим основным направлениям:

- ведение распределенных архивов данных ДЗЗ и результатов их обработки;
- обеспечение отказоустойчивого хранения данных в составе распределенных архивов;
- предоставление и анализ данных на основе RESTful-сервисов.

Разработанные высокопроизводительные алгоритмы и программные средства должны обеспечить решение таких ключевых задач, как:

- оптимизация обработки данных с учетом близости их размещения относительно имеющихся вычислительных ресурсов (locality);
- создание обобщенного языка запросов и сервисов для доступа к данным;
- интеграция, анализ, поиск закономерностей и обеспечение повторного использования данных (кэширование);
- разграничение доступа к данным между пользователями программной платформы.

Это, в свою очередь, позволит:

- упростить междисциплинарную интеграцию и обеспечить масштабируемое и надежное хранение научных данных с общей моделью данных и языком запросов;
- ускорить запись и чтение данных;

- проводить распределенный поиск метаданных и отслеживание происхождения данных (data provenance);
- облегчить проведение интерактивного анализа и визуального поиска закономерностей, а также повторное использование данных.

Проводимые исследования также подразумевают оценку эффективности применения современных информационных технологий для организации надежного хранения и доступа к данным относительно задач проекта. Это распределенные параллельные файловые системы (Lustre, GlusterFS, Ceph и т. п.), распределенные базы данных (RasDaMan, Apache HBase, Apache Cassandra и т. п.), а также системы облачного хранения (Openstack и т. п.).

Первым результатом работ по данному проекту можно считать разработку архитектуры системы UNISAT для ведения распределенных архивов данных ДЗЗ и результатов их обработки. Помимо этого, она реализует технологии по обмену метаданными между распределенными узлами, а также ведению единого универсального справочника по типам данных и методикам построения на их основе различных “виртуальных” информационных продуктов.

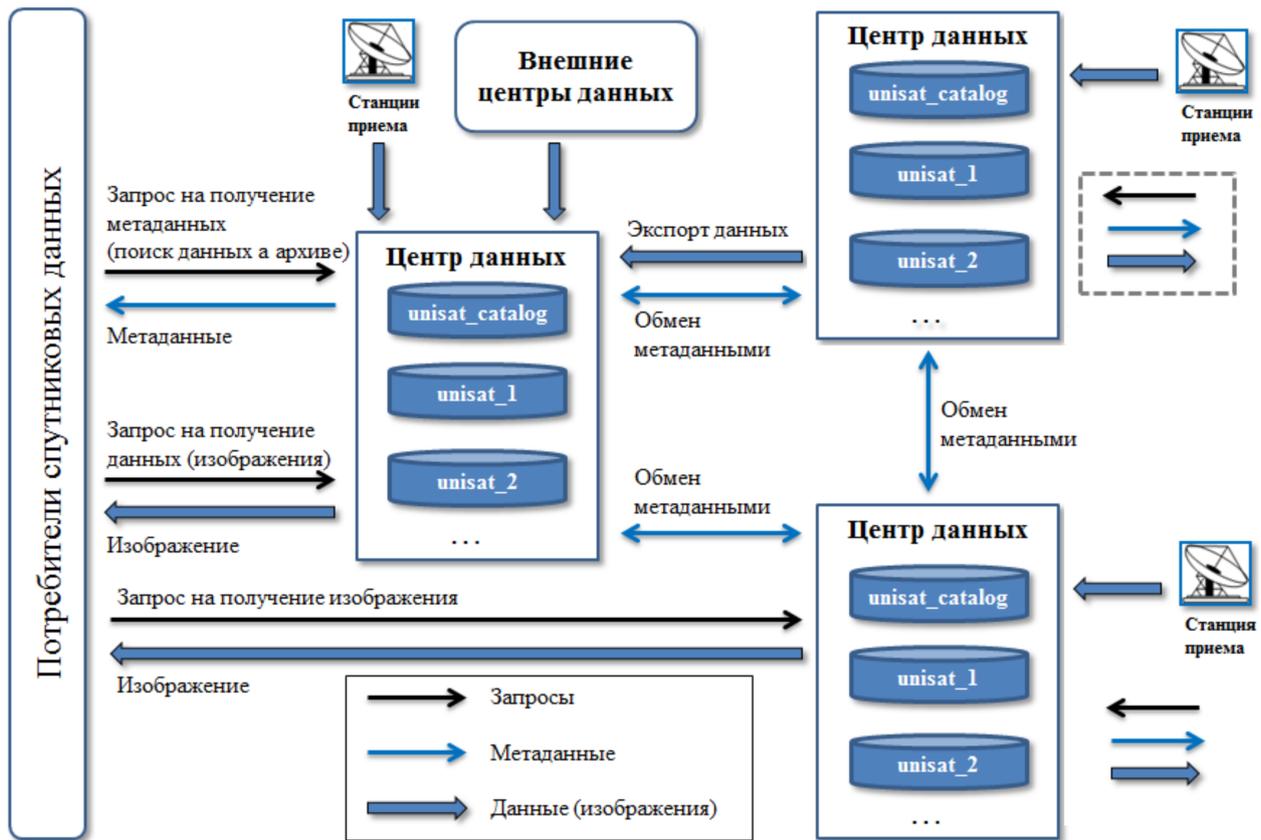
3. Система UNISAT

Новые требования к сервисам доступа и режиму работы со спутниковыми данными приводят к необходимости унификации механизмов доступа к данным разных типов. Сервисы доступа должны включать в себя не только стандартные, классические инструменты для выбора и просмотра данных, но и различный инструментарий для их онлайн-анализа. К нему можно отнести инструменты визуального сравнения данных, проведения интерактивной обработки и создания новых информационных продуктов непосредственно в системе доступа.

Необходимость реализации таких возможностей приводит к концепции так называемых “виртуальных продуктов”, т. е. продуктов, динамически формируемых на момент их запроса пользователем на основе анализа хранимых в архивах данных. Использование этой концепции позволяет существенно расширять и впоследствии изменять и наращивать диапазон доступной пользователю информации без постоянных повторных обработок и дублирования хранимых данных.

Исследования по реализации такого подхода ведутся коллективом авторов на протяжении последних нескольких лет. Первые результаты работ с виртуальными продуктами были основаны на подходе, когда для каждого типа спутниковой информации (сеансовых данных высокого, низкого и среднего разрешения, композитных данных) создавались отдельные наборы средств и технологий обработки. Это было обусловлено тем, что для каждого типа данных были реализованы свои механизмы хранения, на тот момент оптимально соответствующие специфике данных. Сегодня же при постоянном росте числа исходных инструментальных данных и типов продуктов, а также увеличении сложности механизмов доступа на первый план выходит унификация механизмов хранения и работы с разнородными данными, что позволяет существенно упростить разработку и поддержку масштабных информационных систем и сервисов для работы с данными ДЗЗ.

В рамках такой концепции авторами предложена и реализована новая архитектура системы хранения спутниковых данных UNISAT. Ее основным отличием от традиционных архитектур является возможность хранения не только самих данных, в том числе



Принципиальная схема построения распределенных архивов

совершенно разнородных, но и описания процедур и методик работы с ними. Основными компонентами системы UNISAT являются:

- универсальная БД unisat, обеспечивающая хранение спутниковых данных;
- единая справочная БД unisat_catalog, содержащая всю необходимую информацию об источниках и характеристиках данных (спутниках, приборах, спектральных каналах и т. д.), а также типах и способах получения виртуальных продуктов на основе этих данных;
- API для взаимодействия системы доступа к данным с системой хранения.

Принципиальная схема распределенных архивов данных, создаваемых на основе данной технологии, приведена на рисунке.

Создаваемые программные средства и технологии проходят апробацию на базе информационной системы «Вега — Дальний Восток» (Вега-ДВ) [2]. Она предназначена для обеспечения доступа к данным распределенной информационной системы коллективного использования данных космического дистанционного зондирования Земли (ИСКИ ДЗЗ) с целью проведения научной, образовательной и инновационной деятельности в области исследования и контроля состояния окружающей среды в регионах Дальнего Востока России.

Заключение

Успешное решение поставленных задач обеспечит расширение перечня информационных продуктов, доступных пользователям системы Вега-ДВ более чем на 0.5 Пбайт

данных, и развитие возможностей обработки и анализа данных дистанционных наблюдений на основе распределенной вычислительной инфраструктуры, сформированной на ресурсах ВЦ ДВО РАН, ИКИ РАН и территориальных центров НИЦ “Планета”. Это позволит на территории Дальнего Востока России улучшить возможности мониторинга растительного покрова и его изменений вследствие различных факторов.

Благодарности. Работа выполнена при поддержке РФФИ (грант № 15-29-07953 офи_м), Программ фундаментальных научных исследований государственных академий наук и Программы фундаментальных исследований ДВО РАН “Дальний Восток” (гранты № 15-I-4-071 и 15-I-4-072).

Список литературы / References

- [1] Лупян Е.А., Савин И.Ю., Барталев С.А., Толпин В.А., Балашов И.В., Плотников Д.Е. Спутниковый сервис мониторинга состояния растительности (“Вега”) // Совр. пробл. дистан. зондирования Земли из космоса. 2011. Т. 8, № 1. С. 190–198.
Loupian, E., Savin, I., Bartalev, S., Tolpin, V., Balashov, I., Plotnikov D. Satellite service for vegetation monitoring VEGA // Current Problems in Remote Sensing of the Earth from Space. 2011. Vol. 8, No. 1. P. 190–198. (In Russ.)
- [2] Смагин С.И., Лупян Е.А., Сорокин А.А., Бурцев М.А., Верхотуров А.Л., Гирина О.А., Ефремов В.Ю., Крамарева Л.С., Прошин А.А., Толпин В.А. Информационная система работы с данными спутниковых наблюдений региона Дальнего Востока России для проведения научных исследований в различных областях знаний // Совр. пробл. дистан. зондирования Земли из космоса. 2013. Т. 10, № 1. С. 277–291.
Smagin, S.I., Loupian, E.A., Sorokin, A.A., Bourtsev, M.A., Verkhoturov, A.L., Girina, O.A., Efremov, V.Yu., Kramareva, L.S., Proshin, A.A., Tolpin, V.A. Information system for scientific research in various fields of expertise using satellite data in the region of the Far East of Russia // Current Problems in Remote Sensing of the Earth from Space. 2013. Vol. 10, No. 1. P. 277–291. (In Russ.)
- [3] Ефремов В.Ю., Гирина О.А., Крамарева Л.С., Лупян Е.А., Маневич А.Г., Мельников Д.В., Матвеев А.М., Прошин А.А., Сорокин А.А., Флитман Е.В. Создание информационного сервиса “Дистанционный мониторинг активности вулканов Камчатки и Курил” // Совр. пробл. дистан. зондирования Земли из космоса. 2012. Т. 9, № 5. С. 155–170.
Efremov, V.Yu., Girina, O.A., Kramareva, L.S., Loupian, E.A., Manevich, A.G., Melnikov, D.V., Matveev, A.M., Proshin, A.A., Sorokin, A.A., Flitman, E.V. Creating an information service “remote monitoring of active volcanoes of Kamchatka and the Kuril Islands” // Current Problems in Remote Sensing of the Earth from Space. 2012. Vol. 9, No. 5. P. 155–170. (In Russ.)
- [4] Лупян Е.А., Милехин О.Е., Антонов В.Н., Крамарева Л.С., Бурцев М.А., Балашов И.В., Толпин В.А., Соловьев В.И. Система работы с объединенными информационными ресурсами, получаемыми на основе спутниковых данных в центрах НИЦ “Планета” // Метеорология и гидрология. 2014. № 12. С. 89–97.
Lupyan, E.A., Milekhin, O.E., Antonov, V.N., Kramareva, L.S., Burtsev, M.A., Balashov, I.V., Tolpin, V.A., Solov’ev, V.I. Problems of data organization and use within the Planeta Research Center for space hydrometeorology United Satellite Data Access System // Meteorologiya i Hydrologiya. 2014. No. 12. P. 89–97. (In Russ.)

- [5] **Лупян Е.А., Матвеев А.М., Уваров И.А., Бочарова Т.Ю., Лаврова О.Ю., Митягина М.И.** Спутниковый сервис See the Sea — инструмент для изучения процессов и явлений на поверхности океана // Совр. пробл. дистан. зондирования Земли из космоса. 2012. Т. 9, № 2. С. 251–262.
Loupiian, E.A., Matveev, A.A., Uvarov, I.A., Bocharova, T.Yu., Lavrova, O.Yu., Mityagina, M.I. The satellite service See the Sea — a tool for the study of oceanic phenomena and processes // Current Problems in Remote Sensing of the Earth from Space. 2012. Vol. 9, No. 2. P. 251–262. (In Russ.)
- [6] **Samet, R., Tural, S.** Web based real-time meteorological data analysis and mapping information system // Intern. J. of Education and Information Technologies. 2010. Vol. 4, No. 4. P. 187–196.
- [7] **Шокин Ю.И., Жижимов О.Л., Пестунов И.А., Синявский Ю.Н., Смирнов В.В.** Распределенная информационно-аналитическая система для поиска, обработки и анализа пространственных данных // Вычисл. технологии. 2007. Т. 12, спец. выпуск 3. С. 108–115.
Shokin, Yu.I., Zhizhimov, O.L., Pestunov, I.A., Sinyavskiy, Yu.N., Smirnov, V.V. Distributed information-analytical system for searching, processing and analysis of spatial data // Comput. Technologies. 2007. Vol. 12, spec. issue. 3. P. 108–115. (In Russ.)
- [8] **Antonopoulos, N., Gillam, L.** Cloud computing: Principles, systems and applications. Springer, 2010. 396 p.
- [9] **Pallickara, S., Pierce, M.** Scientific data management in the cloud: A survey of technologies, approaches and Challenges // Handbook of Cloud Computing / Eds B. Furht, A. Escalante. Springer Science+Business Media, 2010. P. 517–533.
- [10] **Zhizhin, M., Medvedev, D., Mishin, D., Poyda, A., Novikov, A.** Transparent data cube for spatiotemporal data mining and visualization // Grid and Cloud Database Management / Eds S. Fiore, G. Aloisio. Springer, 2011. P. 307–330.
- [11] **Рамарриян, Н.К.** Development, operation and evolution of EOSDIS — NASA’s major capability for managing Earth science data // CENDI/NFAIS Workshop on Repositories in Science & Technology: Preserving Access to the Record of Science. Washington, November 30, 2011. http://www.cendi.gov/activities/11_30_2011_CENDI_NFAIS_FLICC.html
- [12] **Лупян Е.А., Саворский В.П., Шокин Ю.И., Алексанин А.И., Назиров Р.Р., Недолужко И.В., Панова О.Ю.** Современные подходы и технологии организации работы с данными дистанционного зондирования Земли для решения научных задач // Совр. пробл. дистан. зондирования Земли из космоса. 2012. Т. 9, № 5. С. 21–44.
Loupiian, E.A., Savorsky, V.P., Shokin, Yu.I., Aleksanin, A.I., Nazirov, R.R., Nedolugko, I.V., Panova, O.Yu. Up-to-date approaches and technology arrangement of Earth observation data applications aimed to solve scientific tasks // Current Problems in Remote Sensing of the Earth from Space. 2012. Vol. 9, No. 5. P. 21–44. (In Russ.)
- [13] **Миклашевич С.Э., Балашов И.В., Бурцев М.А., Ефремов В.Ю., Мазуров А.А., Матвеев А.М., Прошин А.А., Радченко М.В., Флитман Е.В.** Программно-аппаратный комплекс для сбора, обработки, архивации и распространения спутниковых данных и продуктов их тематической обработки // Совр. пробл. дистан. зондирования Земли из космоса. 2012. Т. 9, № 4. С. 47–56.
Miklashevich, S.E., Balashov, I.V., Burcev, M.A., Efremov, V.Yu., Mazurov, A.A., Matveev, A.M., Proshin, A.A., Radchenko, M.V., Flitman, E.V. Complex system for the receiving, processing, archiving and distribution of satellite data and products of thematic processing // Current Problems in Remote Sensing of the Earth from Space. 2012. Vol. 9, No. 4. P. 47–56. (In Russ.)

- [14] Savorskiy, V.P., Lupyán, E.A., Kaevitser, V.I., Smirnov, M.T., Ermakov, D.M., Balashov, I.V., Bourtsev, M.A., Efremov, V.Yu., Proshin, A.A., Tolpin, V.A., Uvarov, I.A., Flitman, E.V. Distributed infrastructure enabling effective integration of earth observation information resources for collective solution of archiving, searching, processing and eo data analyzing tasks // Intern. Archives of the Photogrammetry, Remote Sensing and Spatial Inform. Sci. XXII ISPRS Congress, 25 August – 01 September 2012, Melbourne, Australia. 2012. Vol. XXXIX-B4. P. 237–242.
- [15] Лупян Е.А., Мазуров А.А., Назиров Р.Р., Прошин А.А., Флитман Е.В., Крашенинникова Ю.С. Технологии построения информационных систем дистанционного мониторинга // Совр. пробл. дистан. зондирования Земли из космоса. 2011. Т. 8, № 1. С. 26–43.
Loupian, E.A., Mazurov, A.A., Nazirov, R.R., Proshin, A.A., Flitman, E.V., Krashenninnikova, Yu.S. Technologies for building remote monitoring information systems // Current Problems in Remote Sensing of the Earth from Space. 2011. Vol. 8, No. 1. P. 26–43. (In Russ.)
- [16] Антонов А.В., Бурцев М.А., Ефремов В.Ю., Калашников А.В., Крамарева Л.С., Крашенинникова Ю.С., Лупян Е.А., Матвеев А.М., Прошин А.А., Флитман Е.В. Построение объединенного каталога распределенных архивов спутниковых данных различных центров // Совр. пробл. дистан. зондирования Земли из космоса. 2010. Т. 7, № 2. С. 84–89.
Antonov, A.V., Bourtsev, M.A., Efremov, V.Yu., Kalashnikov, A.V., Kramareva, L.S., Krashenninnikova, Yu.S., Loupian, E.A., Matveev, A.M., Proshin, A.A., Flitman, E.V. Design of the union catalogue of distributed satellite data archives of various centers // Current Problems in Remote Sensing of the Earth from Space. 2010. Vol. 7, No. 2. P. 84–89. (In Russ.)
- [17] Bourtsev, M.A., Loupian, E.A., Mazurov, A.A., Nazirov, R.R., Proshin, A.A., Flitman, E.V. Technology of automated satellite data archiving and representing systems development // Proc. of the 7th Intern. Symp. on Reducing Costs of Spacecraft Ground Systems and Operations (RCSGSO). 11–15 June 2007. Moscow. Russian Federation (ESA SP-648, July 2007), 2007. P. 3b2-1–3b2-4.
- [18] Ефремов В.Ю., Лупян Е.А., Мазуров А.А., Прошин А.А., Флитман Е.В. Технология построения автоматизированных систем хранения спутниковых данных // Совр. пробл. дистан. зондирования Земли из космоса. 2004. Т. 1, № 1. С. 437–443.
Efremov, V.Yu., Loupian, E.A., Mazurov, A.A., Proshin, A.A., Flitman, E.V. Technology of automated satellite data storage systems // Current Problems in Remote Sensing of the Earth from Space. 2004. Vol. 1, No. 1. P. 437–443. (In Russ.)
- [19] Сорокин А.А., Королев С.П., Смагин С.И., Поляков А.Н. Макет отказоустойчивой информационной системы для облачного хранения наборов научных данных // Вычисл. технологии. 2013. Т. 18, № 1. С. 87–95.
Sorokin, A.A., Korolev, S.P., Smagin, S.I., Polyakov, A.N. Layout of a resilient informational system for cloud storage of science data sets // Comput. Technologies. 2013. Vol. 18, No. 1. P. 87–95. (In Russ.)
- [20] Ханчук А.И., Сорокин А.А., Смагин С.И., Королев С.П., Макогонов С.В., Тарасов А.Г., Шестаков Н.В. Развитие информационно-телекоммуникационных систем в ДВО РАН // Информ. технологии и вычисл. системы. 2013. № 4. С. 45–57.
Khanchuk, A.I., Sorokin, A.A., Smagin, S.I., Korolev, S.P., Makogonov, S.V., Tarasov, A.G., Shestakov, N.V. Development of information and telecommunication systems in the Far Eastern Branch of RAS // Information Technologies & Computer Systems. 2013. No. 4. P. 45–57. (In Russ.)

- [21] Sorokin, A.A., Korolev, S.P., Makogonov, S.V., Shestakov, N.V. The use of cloud computing for storing and processing instrumental data from seismological observations networks // Proc. of Intern. Conf. on Computer Science and Environmental Engineering (CSEE 2015), Beijing, May 17–18, 2015. P. 546–552.
- [22] Сорокин А.А., Королев С.П., Урманов И.П., Верхотуров А.Л., Шестаков Н.В., Гирина О.А. Информационная система для работы с данными инструментальных наблюдений с целью проведения исследований и мониторинга опасных природных явлений на Дальнем Востоке России // Матер. Всерос. науч. конф. с междунар. участием “Геодинамические процессы и природные катастрофы. Опыт Нефтегорска”, Южно-Сахалинск, 26–30 мая 2015. Владивосток: Дальнаука, 2015. Т. 2. С. 443–447.
Sorokin, A.A., Korolev, S.P., Urmanov, I.P., Verkhoturov, A.L., Shestakov, N.V., Girina, O.A. Information system for work with instrumental observations data for research and monitoring of natural hazards in the Far East Russia // Proc. of the Russ. Conf. “Geodynamic Processes and Natural Disasters. Experience Neftegorsk”. Vladivostok: Dal’nauka, 2015. Vol. 2. P. 443–447. (In Russ.)
- [23] Королев С.П., Сорокин А.А., Верхотуров А.Л., Коновалов А.В., Шестаков Н.В. Автоматизированная информационная система для работы с инструментальными данными региональной сети сейсмологических наблюдений ДВО РАН // Сейсмические приборы. 2014. Т. 50, № 3. С. 30–41.
Korolev, S.P., Sorokin, A.A., Verkhoturov, A.L., Konovalov, A.V., Shestakov, N.V. Automated information system for instrument data processing of the regional seismic observation network of FEB RAS // Seismic Instruments. 2014. Vol. 50, No. 3. P. 30–41. (In Russ.)

Поступила в редакцию 11 января 2016 г.

Development of methods and technologies for distributed storage and processing of the remote sensing data for environment monitoring purposes

SMAGIN, SERGEY I.¹, LUPYAN, EVGENY A.², SOROKIN, ALEXEI A.¹, BURTSEV, MIKHAIL A.², KOROLEV, SERGEY P.^{1,*}, PROSHIN, ANDREY A.², KRAMAREVA, LIUBOW S.³

¹Computing Center FEB RAS, Khabarovsk, 680000, Russia

²Space Research Institute RAS, Moscow, 117997, Russia

³Far-Eastern Center of State Research Center for Space Hydrometeorology “Planeta”, Khabarovsk, 680000, Russia

*Corresponding author: Korolev, Sergey P., e-mail: serejk@febras.net

Purpose. In recent years an increased number of tools with improved qualitative characteristics for remote environmental observation along with developed of satellite data analysis techniques and expanded fields of their application have resulted in a sharp growth of the number and volumes of storage of satellite data archives. At the same time, the creation of technologies and software packages for storage of archives of such large data and their distributed processing and use are somewhat lagged behind. This reduces the ability and efficiency for application of remote sensing data and their processing results for the solution of research problems on a large scale. The authors

of the paper conduct studies and develop contemporary techniques and algorithms for storage, processing and analysis of super-large volumes of remote sensing data using the capabilities of cloud computing environments.

Methodology. This work implies the creation of a programming platform which provides tools that allow searching for (transparent to user or external information system) and receiving of the necessary instrumental data and their processing results from hundreds of thematic servers and repositories distributed over the territory of Russia. The implementation of this work is associated with an implementation of technologies in compliance with the following directions:

- distributed archive storage of Earth remote sensing (ERS) data and their processing results;
- providing fail-safe distributed archive data storage;
- data presentation and analysis based on RESTful services.

Originality. Development of the UNISAT system architecture for distributed archive storage of Earth remote sensing (ERS) data and their processing results can be considered as the first result of our work to implement this Project. Besides, it realizes technologies on metadata exchange between the distributed nodes, and also, a single universal directory for data type storage as well as techniques of construction of different “virtual” information products on their basis.

Findings. The paper describes the features related to the solution of above listed problems in Russia, estimation of possible technical tools which allow us to increase the efficiency and speed of operations with the data, and also, the first results of our work in this direction.

Keywords: cloud technologies, science data sets, database, remote sensing, information system, distributed data storage and processing, RESTful web-service, environment.

Acknowledgements. This research was partly supported by RFBR (grant No. 15-29-07953), FEB RAS research program “Far East” (grants No. 15-I-4-071, 15-I-4-072).

Received 11 January 2016