

Прототип системы распределенной обработки данных радиолокационного космического мониторинга

В.М. Котов, М.А. Минеев, А.В. Яковлев

Объединенный институт ядерных исследований

В статье рассмотрены технологические решения интеграции программного обеспечения удаленного мониторинга экспериментов на БАК и инструментального программного комплекса NEST ESA, а также создания на их основе прототипа системы распределенной обработки данных радиолокационного космического мониторинга. Описаны архитектура и алгоритмы работы интегрированного комплекса, методы организации и управления распределенными потоками данных. Предлагаемые решения основаны на программном обеспечении с открытым кодом и общих стандартах, включая промежуточное программное обеспечение, программные комплексы BEST, BEAM, библиотеку NetCDF и другие.

1. Введение

Сбор, предварительная обработка и анализ данных, полученных при дистанционном зондировании Земли (ДЗЗ) с помощью космических радиолокаторов с синтезированной апертурой (РСА) (Synthetic Aperture Radar - SAR), связаны с накоплением и обработкой информации, объемы которой в сотни тысяч раз превышают средний уровень потребностей и достигают десятков Tb в год.

Европейское Космическое Агентство (European Space Agency (ESA)) начиная с 90-х годов прошлого столетия предоставляет пользователям данные космических аппаратов, имеющих РСА и к текущему моменту собран большой архив данных.

Традиционные подходы к разработке инфраструктуры и обработки таких объемов данных (получивших название Большие данные) не годятся, необходимы новые подходы к обработке и анализу данных для современных РСА, в том числе и реализация интерактивного взаимодействия в режиме удаленного доступа для многоуровневой, распределенной географически вычислительной системе обработки данных больших объемов.

Аналогичная проблема стояла и в области экспериментальной физики высоких энергий еще в 90-е годы прошлого столетия при создании системы сбора и обработки данных экспериментов на Большом адронном коллайдере (БАК).

Опыт разработки и реализации системы сбора и обработки данных экспериментов на БАК может быть использован при создании систем обработки данных в системах ДЗЗ, и в частности для данных, полученных с помощью космических радиолокаторов с синтезированной апертурой.

2. Технологии, используемые при создании системы распределенной обработки данных ДЗЗ

Next ESA SAR Toolbox

Необходимость первичной обработки радиолокационных данных определяется особенностью форматов и структурой данных SAR, требующих предварительной обработки для выполнения в последующем полного цикла обработки информации с SAR. Кроме того, структура радарных данных позволяет проводить эффективную предварительную обработку на уровне первичной обработки радиолокационных изображений, оперируя изображением в целом как образом, для быстрого опознания и анализа объектов и их характеристик.

ЕКА в октябре 2007 г. заключило договор с канадской фирмой Array Systems Computing на разработку инструментального программного обеспечения и с 2010 г. предоставляет комплект программного обеспечения с открытым исходным кодом NEST (Next ESA SAR Toolbox),

включающего в себя функциональность всех предыдущих версий инструментального ПО BEST, BEAM и др.

По условиям технического задания ЕКА инструментарий NEST предназначен для помощи в подготовке элементов системы обработки данных SAR, но не является SAR процессором или готовой системой обработки SAR данных в полном технологическом цикле.

Вместе с тем, статус пакета NEST как программы с открытым исходным кодом позволяет использовать его в составе облачной платформы для разработки приложений радиолокационного космического мониторинга.

В настоящий момент в рамках реализованного прототипа системы распределенной обработки данных ДЗЗ используются операторы из пакета Graph Processing Tools (GPF Operator) ПО NEST. Они применяются в качестве алгоритмов обработки SAR данных. Такое использование указанных операторов позволяют выполнять обработку SAR данных на удаленных вычислительных хостах.

Кроме того, в дальнейшем предполагается использовать полный пакет ПО NEST в качестве отдельного клиентского ПО для доступа к системе обработки данных ДЗЗ. Данный вариант предполагает сохранение ПО NEST в качестве отдельного самостоятельного программного комплекса, способного реализовывать все стандартные возможности, заложенные в него при разработке. При этом ПО NEST получит дополнительный инструментарий, который позволит проводить обработку SAR данных на удаленных вычислительных ресурсах непосредственно из данного программного комплекса.

G-POD (ESA Grid Processing on Demand)

Также, начиная с 2002 г. под руководством ЕКА успешно реализуется проект G-POD (ESA Grid Processing on Demand).

Проект G-POD является общим операционным окружением, основанном на технологиях Грид, где конкретные приложения обработки данных могут быть легко подключены к системе. Сочетая высокую производительность и значительные вычислительные ресурсы, управляемые при помощи Грид технологий, G-POD обеспечивает необходимую гибкость для создания приложений виртуальной среды с быстрой доступностью к данным, вычислительным ресурсам и результатам.

Web-портал G-POD представляет собой гибкую, безопасную, общую и одновременно и распределенную платформу, на которой пользователь может легко управлять всеми своими задачами. В ходе всех стадий от создания новой задачи и до публикации результатов, через стадии отбора данных и мониторинга выполнения задач, пользователя окружает доступный отовсюду дружелюбный и интуитивно понятный интерфейс.

Архитектура G-POD представлена на рис. 1.

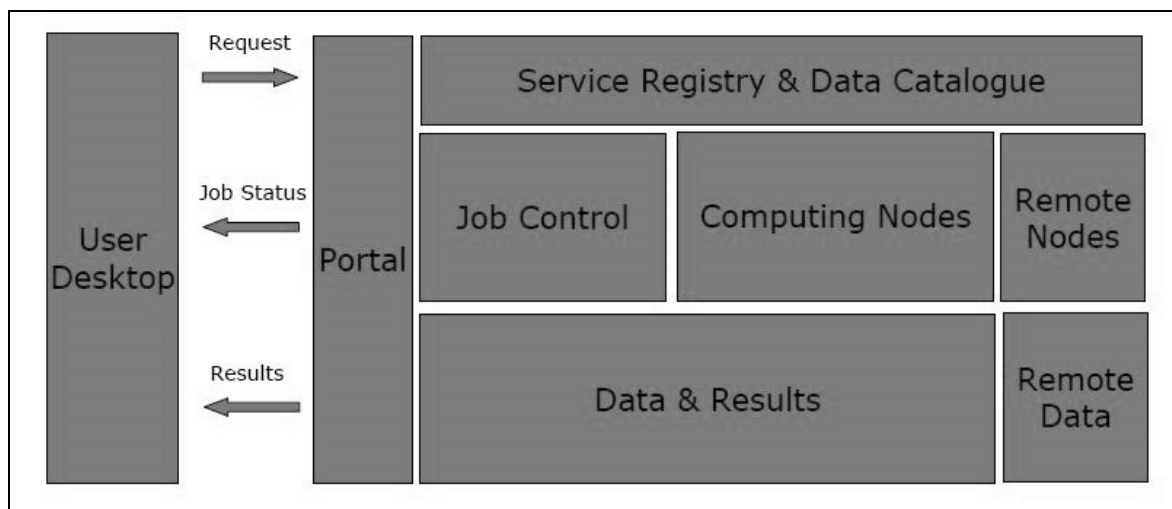


Рис. 1. Архитектура G-POD

Идеология и архитектура G-POD предполагает использование ресурсов локального вычислительного кластера, а при необходимости использование облачных вычислительных ресурсов

Концепция проекта G-POD, а также современные парадигмы ЕКА при разработке ПО для обработки данных космического мониторинга и специфические особенности обработки радарных данных были использованы при разработке общей модели и архитектуры предлагаемой системы распределенной обработки данных ДЗЗ.

ATLAS TDAQ

Учитывая опыт разработки и реализации системы сбора и обработки данных экспериментов на БАК, было принято решение использовать набор компонентов из Системы сбора и обработки данных в реальном времени эксперимента ATLAS БАК (ПО ATLAS TDAQ Online Software и High Level Trigger) в качестве базового ядра предложенной системы распределенной обработки данных радиолокационного космического мониторинга. Кроме существующих сервисов TDAQ были разработаны дополнительные. При этом новые сервисы построены таким образом, что бы полностью использовать все существующие возможности TDAQ и High Level Trigger, являясь фактически надстройкой над существующими сервисами, либо дополнением этих сервисов.

3. Архитектура прототипа системы распределенной обработки данных ДЗЗ

Для проверки работоспособности разработанной архитектурной модели системы распределенной обработки данных ДЗЗ был создан прототип, в котором реализован минимальный набор функций.

В настоящий момент прототип представляет собой классическую архитектуру и состоит из трех основных компонентов - пользовательский интерфейс, центральный сервер, представленный в виде набора управляющих сервисов и исполняемые модули.

Архитектура прототипа на рис. 2.

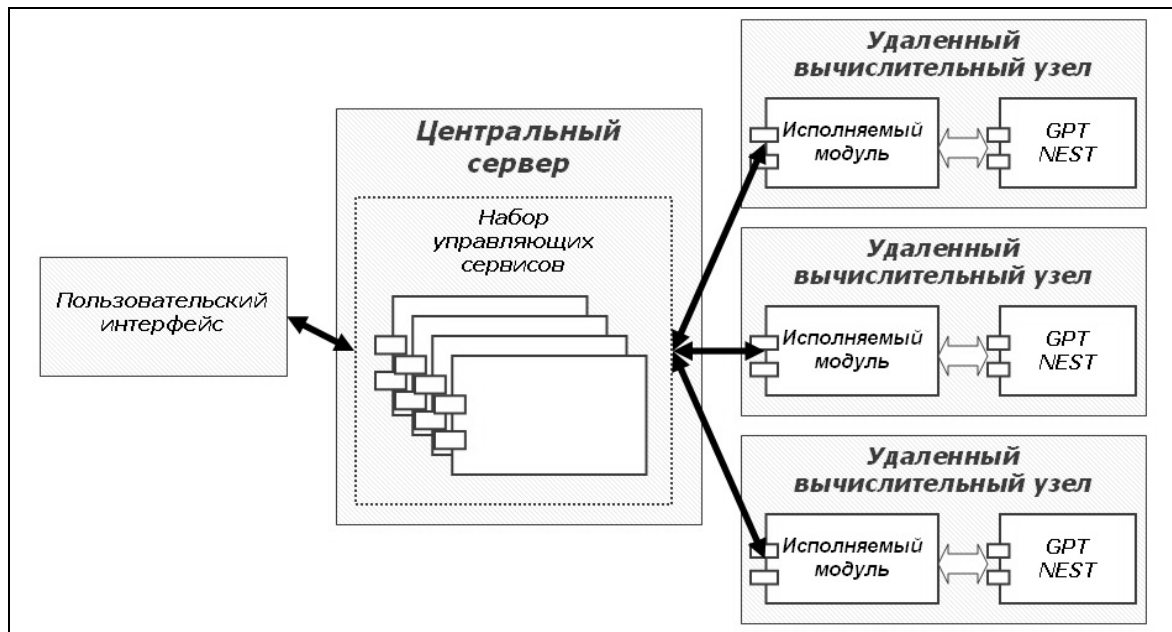


Рис. 2. Архитектура прототипа системы распределенной обработки данных ДЗЗ

3.1 Пользовательский интерфейс

Пользовательский интерфейс применяется для взаимодействия пользователя с остальными компонентами серверной части прототипа. Через пользовательский интерфейс пользователь составляет задания, посылает задание центральному серверу, наблюдает за ходом обработки, получает от центрального сервера результаты выполнения задания.

В настоящий момент пользовательский интерфейс реализован в виде отдельного клиентского приложения. В дальнейшем предполагается разработать Web-портал, который будет являться для удаленного пользователя центральной точкой доступа ко всем ресурсам системы обработки данных ДЗЗ.

3.2 Центральный сервер

Центральный сервер прототипа, взаимодействует как с пользовательским интерфейсом, так и с исполняемыми модулями, установленными на удаленных вычислительных узлах. Полученное от пользователя задание центральный сервер пересылает исполняемому модулю, установленному на конкретном удаленном вычислительном узле. Центральный сервер периодически опрашивает все исполняемые модули, проверяя их текущий статус. После окончания выполнения задания центральный сервер получает от исполняемого модуля полученный результат, и по запросу пользователя пересылает это результат ему.

В настоящий момент центральный сервер прототипа, представляет собой минимальный набор управляющих сервисов и компонентов, необходимых для реализации системы распределенной обработки данных.

В частности реализованы:

- авторизация и аутентификация пользователей и разделение доступа;
- очередь заданий,
- операционный мониторинг состояния удаленных вычислительных узлов.

Также предполагается разработать отдельный пользовательский интерфейс для администратора системы.

3.3 Исполняемый модуль

Исполняемый модуль устанавливается на вычислительные узлы, на которых будет производиться обработка данных ДЗЗ. Исполняемый модуль получает от центрального сервера задание, поднимает всю необходимую инфраструктуру Graph Processing Tools ПО NEST и передает ему задание на выполнение. По итогам выполнения задания исполняемый модуль передает полученный результат центральному управляющему сервису по запросу последнего.

4. Заключение

Создание прототипа полноценной системы распределенной обработки данных ДЗЗ на основе инструментального пакета NEST ESA и системы удаленного доступа СУДРВ ОИЯИ (ATLAS TDAQ) позволит впоследствии создать развитую платформу для разработки систем предварительной обработки радарных данных космических локаторов с синтезированной апертурой высокого разрешения и отладки их в условиях реальной работы в большой географически распределенной вычислительной Грид-системе обработки данных экспериментов на БАК..

Литература

1. Костюк Е.А., Веремчук Ю.А., Денисов П.В. Перспективные технологии обработки космической радиолокационной информации в НКПОР Оператора КС ДЗЗ // V Международная конференция «Космическая съемка — на пике высоких технологий».
2. Mapelli L. Spanning from Data Acquisition to GRID - Today and a view of tomorrow. // XXIII International Symposium on Nuclear Electronics & Computing NEC'2011.
3. В.В. Кореньков, В.М. Котов, Н.А. Русакович, А.В. Яковлев. Система удаленного доступа реального времени (СУДРВ), как композитный сервис распределенной ГРИД-системы обработки данных экспериментов на Большом Адронном Коллайдере (БАК) // Параллельные вычислительные технологии (ПаВТ'2010): Труды международной научной конференции (Уфа, 29 марта – 2 апреля 2010 г.) [Электронный ресурс] – Челябинск: Издательский центр ЮУрГУ, 2010. – с. 668.

4. Software Architecture Document (SAD) for the Next ESA SAR Toolbox (NEST) (ARR-NEST-RS07-016); http://www.array.ca/nest/Software_Architecture_Document_v2.0.pdf
5. В.В. Кореньков, В.М. Котов, Н.А. Русакович, А.В. Яковлев. Модель и технология интеграции online-сервисов эксперимента ATLAS на Большом Адронном Коллайдере (БАК) и сервисов ГРИД-инфраструктуры // Параллельные вычислительные технологии (ПаВТ'2011): труды международной научной конференции (Москва, 28 марта – 1 апреля 2011 г.) [Электронный ресурс] – Челябинск: Издательский центр ЮУрГУ, 2011. - с. 516–521
6. ATLAS High-Level Trigger, Data Acquisition and Controls. Technical Design Report (ATLAS TDR-016); <http://atlas-proj-hltdaqdcs-tdr.web.cern.ch/atlas-proj-hltdaqdcs-tdr/tdr-v1-r4/PDF/TDR.pdf>
7. В.В. Кореньков, В.М. Котов, М.А. Минеев, Н.А. Русакович, А.В. Яковлев. Система удаленного доступа к Грид-инфраструктуре экспериментов на БАК как инструментальная платформа PaaS для разработки геоприложений космического мониторинга в системах дистанционного зондирования Земли // Параллельные вычислительные технологии (ПаВТ'2012): труды международной научной конференции (Новосибирск, 26 – 30 марта 2012 г.) [Электронный ресурс] - Челябинск: Издательский центр ЮУрГУ, 2012. - с. 531–536
8. В.В. Кореньков, В.М. Котов, Н.А. Русакович, А.В. Яковлев. Создание облачной платформы уровня Tier3 в ГРИД-инфраструктуре экспериментов на ЛНС для разработки приложений радиолокационного космического мониторинга // Распределенные вычисления и ГРИД-технологии в образовании и науке (GRID'2011): труды 5-й международной научной конференции (Дубна, 16 – 21 июля 2012 г.) – Дубна: ОИЯИ, 2012. - с. 349–354