

Система удаленного доступа к Грид-инфраструктуре экспериментов на БАК как инструментальная платформа PaaS для разработки геоприложений космического мониторинга в системах дистанционного зондирования Земли

В.В. Кореньков, В.М. Котов, М.А. Минеев, Н.А. Русакович, А.В. Яковлев

Объединенный институт ядерных исследований

Решение проблемы эффективного и устойчивого функционирования распределённых систем дистанционного зондирования Земли (ДЗЗ), динамика сопровождения и дальнейшего их развития делает невозможным решить эту задачу в “статике” один раз и навсегда. Распределённость системы, изменение условий (числа узлов и каналов связи, их характеристик) существенно влияют на правила и алгоритмы управления системы. Проблема оптимизации распределения ресурсов, обеспечения устойчивости и эффективности системы определяет необходимость разработки инструментальной системы PaaS, как составляющей программного обеспечения в современных системах обработки для отслеживания изменений характеристик системы и требований пользователей на всём жизненном цикле системы.

1. Введение

Сбор, предварительная обработка и анализ данных, полученных при дистанционном зондировании Земли (ДЗЗ) с помощью космических радиолокаторов с синтезированной апертурой (РСА), 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 (Envisat, Radarsat, TerraSar, Cosmo-Skymed), и уже собран большой архив данных.

В настоящее время и в России в рамках Федеральной космической программы ведется разработка космических аппаратов («Метеор-М №3», «Аркон-2М»), оснащенных многофункциональной РСА с активной фазированной антенной решеткой, характеристики которой соответствуют характеристикам современных РСА [1].

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 формировании программы исследований крупнейшего проекта современности в области фундаментальной науки: созданию Большого адронного коллайдера (БАК) в Европейском центре ядерных исследований (ЦЕРН). В результате многолетней работы большого коллектива в ЦЕРНе была разработана, отлажена, запущена и успешно работает второй год в режиме реальных измерений на пучках ускорителя БАК система сбора и обработки данных эксперимента ATLAS, объединяющая в единую систему все компоненты обработки, управления и передачи Больших данных (это передача и фильтрация в реальном времени 100Gb/s и offline обработка и анализ экспериментальных данных объемом в десятки Pb в год) [2].

ОИЯИ является участником программы исследований на БАК. Важной составляющей такого участия в обработке и анализе данных, получаемых в экспериментах на БАК является создание в ОИЯИ системы удалённого доступа реального времени (СУДРВ) и интеграция ее в глобальную сервис-ориентированную архитектуру Грид-системы сбора и обработки данных экспериментов на БАК. [3]

2. Концепция

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.

На сегодня не существует технологии, обеспечивающей необходимую функциональность и эффективность обработки таких потоков и объемов экспериментальных данных в одной системе. Поэтому система сбора и обработки эксперимента ATLAS была разделена на два уровня: online - сбор и предварительная обработка и offline - полная обработка и анализ.

Offline обработка требует необычно больших вычислительных ресурсов (примерно 100000 самых мощных на сегодня процессоров или 150 Kcores) и была реализована в Грид - системе (рис.1).

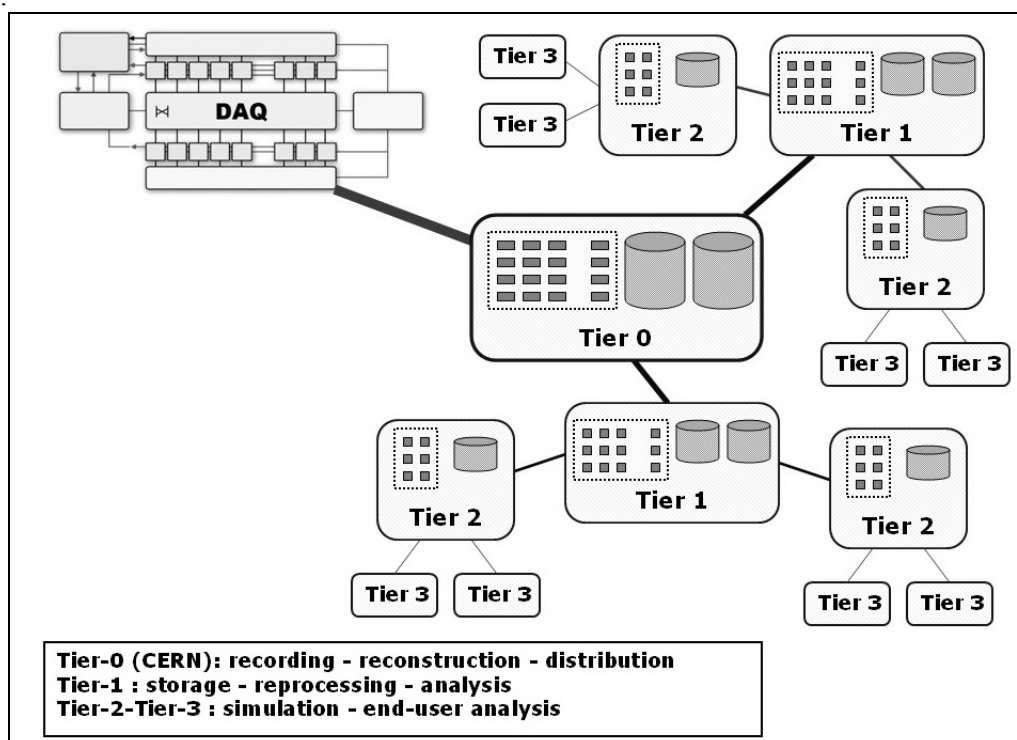


Рис. 1. Система Грид обработки

Показанная в левом верхнем углу на рис. 1. система DAQ и есть online составляющая системы по сбору и предварительной обработке входных данных.

Современная архитектура наземной географически распределенной системы обработки данных SAR (Synthetic Aperture Radar), предлагаемая Европейским космическим агентством, также имеет online уровень обработки входных данных.

Необходимость первичной обработки радиолокационных данных определяется особенно форматом 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.

Радарные данные имеют целый ряд особенностей: сложность обработки из-за геометрических искажений, а также непростая интерпретация изображений. Состав функций предварительной обработки включает следующие возможности обработки данных: фокусировка, корегистрация, удаление спекл-шумов, извлечение характеристик (включая когерентность), геокодирование, радиометрическую калибровку и нормализацию, составление мозаики и классификации.

ESA в октябре 2007 г. заключило договор с канадской фирмой Array Systems Computing на разработку инструментального программного обеспечения и с 2010 г. поставляет комплект программного обеспечения с открытым исходным кодом NEST (Next ESA SAR Toolbox) [4], включающего в себя функциональность всех предыдущих версий инструментального ПО BEST, BEAM и др. Состав пакета NEST, компоненты и потоки данных приведены на рис. 2.

По условиям технического задания ESA 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.

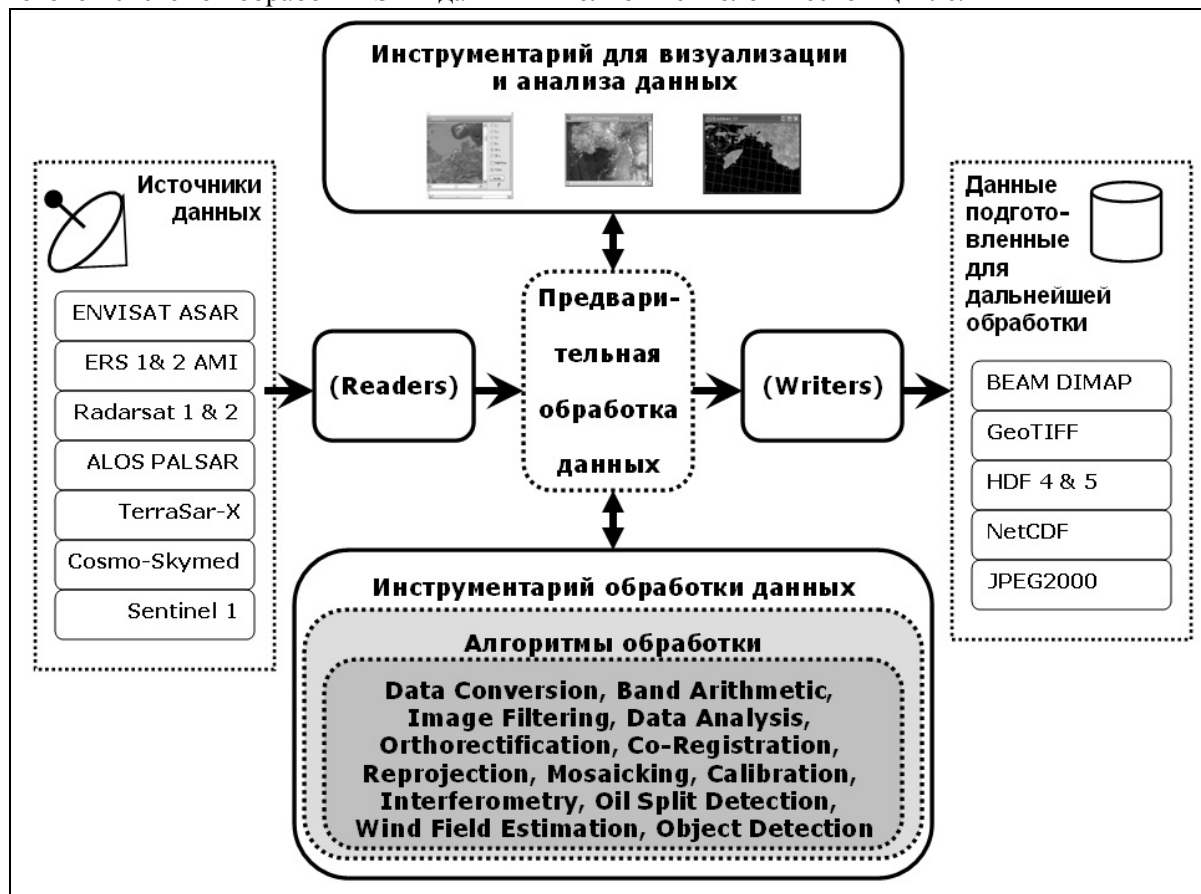


Рис. 2. Компоненты и потоки данных программного комплекса NEST

Вместе с тем, статус пакета NEST как программы с открытым исходным кодом позволяет использовать его в составе PaaS (Platform as a Service) для разработки, тестирования, развертывания и поддержки полномасштабной системы сбора и обработки данных с космических локаторов РСА, реализующей полный технологический цикл обработки радарных данных. Предполагается реализовать набор инструментов для отображения и частичной обработки радарных снимков в среде ArcGIS.

В качестве основы для разработки прототипа подобной PaaS предлагается использовать систему удаленного доступа реального времени (СУДРВ) ОИЯИ. СУДРВ представляет собой композитный сервис Грид-системы и является фрагментом общей системы обработки данных эксперимента ATLAS БАК [5]. Согласно плану развития информационной инфраструктуры экспериментов БАК ЦЕРН и в соответствии с концепцией «облачных вычислений», где все есть Сервис (XaaS), СУДРВ ОИЯИ будет применяться как PaaS для дальнейшего развития системы обработки данных эксперимента ATLAS на БАК.

Кроме того, в соответствии с концепцией «Открытой инновационной лаборатории» эксперимента ATLAS-LAB (ATLAB), на рабочем совещании ЦЕРН-ОИЯИ «Brainstorming workshop on applications from ATLAS using EU-funding for R&D-upgrades» в г. Дубне 24.10.2010 г по обсуждению доклада ОИЯИ «Real Time remote access system for ATLAS» было поддержано предложение о возможности прикладного использования СУДРВ ОИЯИ в области космического мониторинга, проводимых при поддержке ESA совместно с ЦЕРН.

Структура программного обеспечения СУДРВ и NEST использует объектно-ориентированный подход проектирования и соответствует стандарту PSS05 ESA. Объединение NEST и СУДРВ в единую платформу обеспечит интеграцию NEST в общую систему Грид-обработки данных экспериментов БАК, а значит и возможность отладки в последующем и offline режима обработки данных космического мониторинга в географически распределенной вычислительной системе Грид-обработки и партнерство с участием ESA и ЦЕРН.

Следует отметить, что кроме инструментария NEST, отражающего специфику обработки радарных данных, предлагаемая платформа PaaS будет обеспечивать также доступ к сервисам сбора и обработки данных, необходимых для функционирования NEST в составе СУДРВ.

В качестве базового ядра предлагаемой системы удаленного доступа для сбора и обработки космической радиолокационной информации предполагается использовать набор компонентов из Системы сбора и обработки данных в реальном времени эксперимента ATLAS LHC (ATLAS TDAQ) [6].

Базовое программное обеспечение имеет два различных логических представления: представление в виде подсистем и представление в виде уровней ПО.

Разделение на подсистемы произведено в соответствии с основными функциональными возможностями, которые базовое программное обеспечение предоставляет внешним пользователям. Каждая подсистема разделена в свою очередь на пакеты, которые отвечают за отдельные аспекты, соответствующие функциональным возможностям. Каждый пакет реализует функциональные возможности, определенные в API, и предоставляет интерфейс для всех типов пользователей. Основные подсистемы базового программного обеспечения:

Конфигурация - Пакеты подсистемы Конфигурация содержат описание системной конфигурации и обеспечивают доступ для записи информации в конфигурационную базу данных о ходе работы во время сбора и обработки данных..

Управление - содержит пакеты для управления всеми процессами в системе, обеспечивает инициализацию и выключение системы, передачу управляющих команд для запуска и останова процессов внутри системы.

Мониторинг - содержит классы для поддержки мониторинга в системе. Классы системы Мониторинг сообщают о возникающих ошибках, для опубликования информации о состоянии систем, для передачи гистограмм, данных в процессе их сбора.

Система разделена на 3 уровня:

Уровень интерфейсов - это классы API предоставляющие интерфейсы для пользователей базового программного обеспечения.

Уровень сервисов – содержит сервисы и интерфейсы необходимые для их вызова.

Уровень промежуточного программного обеспечения – это пакеты для внутреннего использования другими пакетами. Пакеты промежуточного программного обеспечения не принадлежат какой-либо подсистеме.

Большая часть компонентов базового программного обеспечения является сервисами, которые используются другими системами. Взаимосвязь этих пакетов и подсистем показана на рис 3.

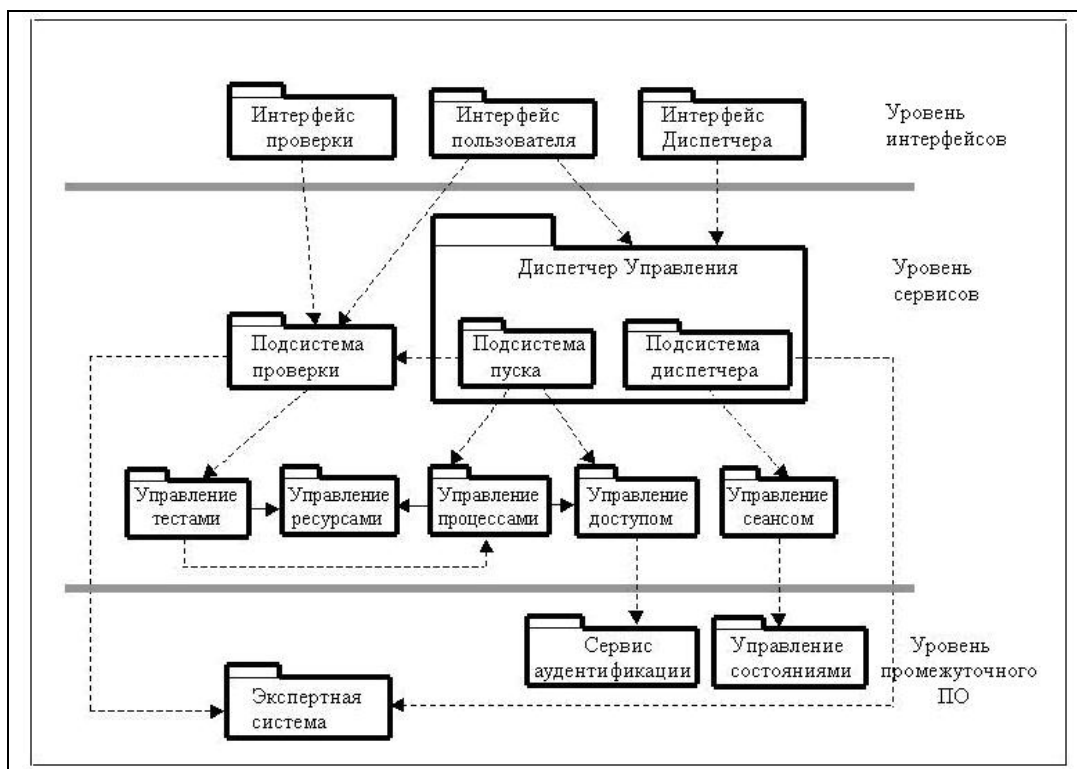


Рис. 3. Компоненты базового программного обеспечения и отношения между компонентами каждого уровня

В состав компонентов программного обеспечения PaaS входят также подсистемы: проверки управления тестами, сервис аутентификации и набор сервисов по интеграции с Грид-системой обработки данных экспериментов на БАК.

3. Заключение

Интеграция инструментального пакета NEST ESA и системы удаленного доступа СУДРВ ОИЯИ позволит создать развитую платформу (PaaS) для разработки прототипов (макетов) 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 конференции «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>