ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ
БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ
Физический

институт
имени
П.Н.Лебедева
Российской академии наук
Ф И А Н

ПРЕПРИНТ

8

Е.А. ИСАЕВ, В.В. КОРНИЛОВ, П.А. ТАРАСОВ,
В.А. САМОДУРОВ, М.В. ШАЦКАЯ

ПЕРЕДАЧА, ХРАНЕНИЕ И ОБРАБОТКА БОЛЬШИХ ОБЪЕМОВ АСТРОНОМИЧЕСКИХ ДАННЫХ

Москва 2014

УДК 004.73

ПЕРЕДАЧА, ХРАНЕНИЕ И ОБРАБОТКА БОЛЬШИХ ОБЪЕМОВ АСТРОНОМИЧЕСКИХ ДАННЫХ

Исаев Е.А.[1,2,3,4], Корнилов В.В.[2,3], Тарасов П.А.[5], Самодуров В.А.[1,2,3], Шацкая М.В.[6]

(1) – Пущинская Радиоастрономическая обсерватория АКЦ ФИАН, Россия
(2) – Национальный исследовательский университет «Высшая школа экономики», Россия
(3) – Институт математических проблем биологии РАН
(4) – ООО "ИТЭК", Россия
(5) – ООО «НПК-ИНФОРМ»
(6) – Астрокосмический центр ФИАН

Аннотация

В настоящее время в астрономии и астрофизике наблюдается значительный рост объёмов экспериментальных данных. В данной работе рассматриваются крупные астрономические проекты с точки зрения передачи, хранения и обработки больших научных данных. Рассмотрена актуальность этих проблем в настоящее время и в будущем.

Currently in astronomy and astrophysics has seen significant growth in the experimental data. This paper discusses the major astronomical projects in terms of communication, storage and processing of big scientific data. We consider the relevance of these issues now and in the future.

Ключевые слова: радиоастрономия, большие объёмы научных данных, высокоскоростная передача данных, оптоволоконные сети.

Keywords: radioastronomy, large amounts of scientific data, high-speed data transfer, fiber optic network.

Введение

В настоящее время практически во всех областях науки наблюдается стремительный рост объёмов данных, получаемых в ходе научных наблюдений

[1]. Огромный прогресс в области информационных технологий, микро- и наноэлектроники, приводит к созданию экспериментальных установок, генерирующих объемы данных, достигающие сотен терабайт и петабайт в самых различных сферах человеческой деятельности, в том числе и в астрономических и радиоастрономических наблюдениях. Полученные в ходе эксперимента данные надо уметь хранить, обрабатывать, передавать и анализировать, получая из этих данных новое знание. Учитывая колоссальные объёмы этих данных и скорость их прироста, каждая из данных задач становиться достаточно сложной для эффективного решения.

Для успешной работы исследователи должны иметь удаленный доступ к большим объемам данных. Исследователи должны получить возможность фильтровать данные, поступающие из отдаленных источников в реальном масштабе времени, и отбирать лишь небольшую долю этих данных. Проблема здесь связана с получением доступа к нужной информации, размещенной в определенном месте, в нужное время. С другой стороны, возникает проблема эффективного управления экспериментальной установкой с удаленного рабочего места исследователя. Еще одна особенность современных научных экспериментов – это сочетание распределенного хранилища данных с необходимостью удаленного доступа к высокопроизводительным вычислительным комплексам для анализа этих данных и получения результатов эксперимента.

Таким образом, в современном мире мы сталкиваемся с необходимостью решения проблемы резкого увеличения передаваемых объемов информации в локальных и региональных сетях передачи данных, что в ряде случаев уже приводит к исчерпанию имеющихся ресурсов, а реальные прогнозы потребностей указывают на продолжение роста информационных потоков в десятки и сотни раз [2].

Особо актуально решение проблемы работы с большими объемами данных в современной астрономии. Приборы для астрономических наблюдений позволяют получать данные с всё более высоким разрешением со все больших участков неба (вплоть до обзоров всего неба), наблюдения астрономических объектов ведутся не только в видимом свете, как раньше, а во всем диапазоне электромагнитного спектра, при этом единственное наблюдение, которое длиться от нескольких секунд до нескольких минут, дает от нескольких мегабайт до нескольких гигабайт информации. Исследователи заинтересованы в долгосрочном хранении полученных архивов для возможности последующих

исследований: так как данные астрономических наблюдений привязываются к конкретным объектам, то их необходимо хранить пока эти объекты существуют; а поскольку времена эволюции астрономических объектов очень велики, то астрономические данные могут храниться бесконечно долго. Кроме того, астрономические данные в большинстве своем не имеют ограничений приватности или коммерческой тайны, и поэтому научное сообщество естественным образом заинтересовано в общедоступности полученных данных, планирование новых задач, исследований и экспериментов строится на анализе текущих результатов и сравнении их с архивными, что в целом накладывает дополнительные требования к возможности оперативного удалённого доступа к данным.

1. Российские научные астрономические проекты, работающие с большими объемами данных.

Рассмотрим некоторые крупнейшие российские и зарубежные научные астрономические проекты, оперирующие огромными массивами информации.

Пущинская радиоастрономическая обсерватория Астрокосмического Центра "Федеральное государственное учреждение науки Физического Института им. Лебедева РАН" (ПРАО) является одной из крупнейших астрофизических обсерваторий в России и располагает тремя радиотелескопами, каждый из которых является одним из лучших инструментов в своей нише [3]:

- радиотелескоп РТ-22, параболический рефлектор, главное зеркало которого имеет диаметр 22 м;
- диапазонный крестообразный радиотелескоп ДКР-1000;
- большая синфазная антенная (БСА).

Суточный поток данных от всех радиотелескопов ПРАО составляет от 10 до 100 Гигабайт (потенциально до 500 Гбт [4]). Локальная вычислительная сеть (ЛВС) ПРАО была создана в 1995 году, и с тех пор было проведено несколько модернизаций ЛВС, в результате чего пропускная способность увеличилась с 1 Мбит/сек до 1 Гбит/сек [5], а в некоторых местах и до 10 Гбит/с. Пропускной способности данной сети достаточно для того, чтобы снимать и обрабатывать данные трех уникальных радиоастрономических комплексов.

Радиотелескоп РТ-22 (рис. 1) в настоящее время используется в качестве станции слежения для проекта «Радиострон» Астрокосмического центра Физического института им. П.Н. Лебедева (АКЦ ФИАН). Указанная

космическая радиообсерватория работает как гигантский интерферометр с базой между спутником, который был создан в научно-производственном объединении (НПО) им. Лавочкина [6] и смонтирован на космическом аппарате «Спектр-Р», и системой наземных радиотелескопов. Приемные станции находятся в США (Грин Бэнк), в Пущино под Москвой и в Австралии (Тидбинбилла).

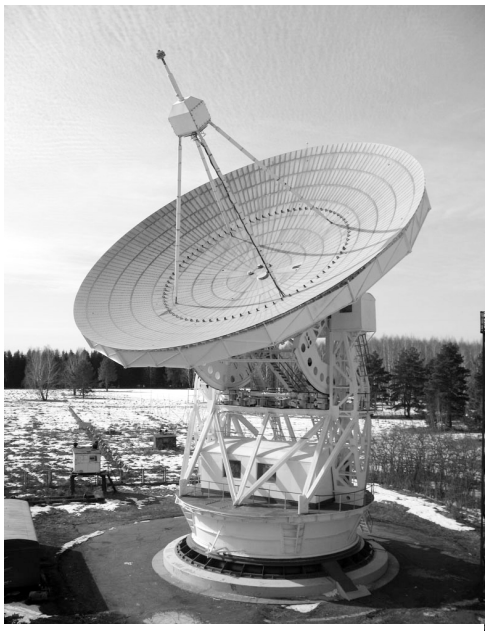
Рисунок 1. Радиотелескоп РТ-22.

Проект «Радиоастрон» рассчитан на систематические исследования таких необычных небесных объектов, как сверхмассивные чёрные дыры в ядрах далёких и близких галактик, черные дыры звёздных масс в нашей галактике, нейтронные (а возможно и кварковые) звёзды, области образования звёзд и планетных систем в нашей галактике и в ядрах других галактик, облака межзвёздной плазмы и гравитационного поля Земли, микроквазары, пульсары, космические мазеры и другие радиоисточники [7].

Суточный поток данных от данного радиоастрономического комплекса составляет чуть более 1,28 Тб. Для работы данного проекта используется канал связи пропускной способностью в 1 Гб/с, который связывает между собой станцию слежения и буферный дата-центр, расположенные на территории ПРАО, а также центр обработки научной информации (ЦОНИ) Астрокосмического центра (АКЦ) ФИАН в Москве. Размер буферного дата-центра ПРАО составляет 24 Тб. Система хранения ЦОНИ состоит из основной системы хранения на 80 Тб и системы резервного копирования на 32 Тб. Сам вычислительный кластер ЦОНИ имеет производительность 430 Гфлоп/с, а внутри него между его серверами пропускная способность составляет 10 Гб/с [8]. В связи с возрастающим интересом в мире к данному проекту и с последующим возрастанием количества получаемых от проекта данных для организации хранения поступающей от Радиоастрона информации до июля 2014 года будет организовано новое серверное хранилище данных размером до 800 ТБ с доступом в оперативном режиме и 2000 Тб с доступом в отсроченном режиме [9].

2. Зарубежные астрономические проекты, работающие с большими объемами данных.

Теперь проведем обзор зарубежных научных проектов, для которых актуальна высокоскоростная передача данных.

Одним из крупнейших зарубежных проектов, регулярные наблюдения на котором проводятся с декабря 2012 года, но который еще не запущен на полную мощность, является массив радиотелескопов LOFAR (LOw Frequency ARray – «низкочастотная антенная решётка») [10]. Данный проект разработан университетом ASTRON в Гронингене (Нидерланды), и служит для исследования низкочастотного радиоизлучения в поисках первых звезд и галактик, потенциальных сигналов внеземного разума, а также изучения черных дыр и пульсаров, представляет собой радиотелескоп с центром в Нидерландах и со станциями в Германии, Швеции, Франции и Великобритании, соединенных вместе при помощи оптоволоконных линий связи (рис. 2) [11].

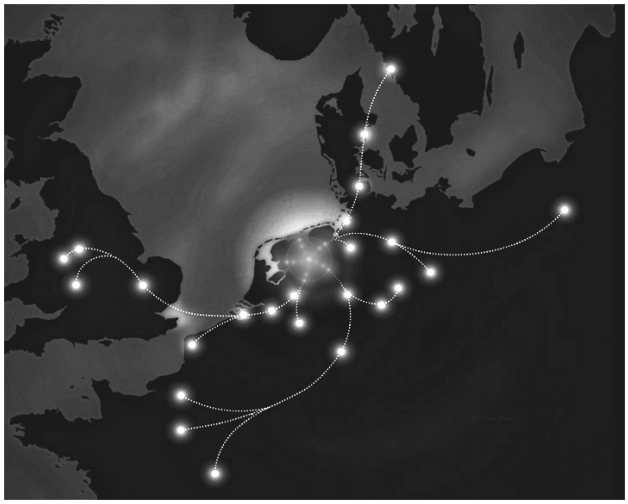

Рисунок 2. Сеть радиотелескопа LOFAR

Для работы данного массива телескопов в штатном режиме от системы кабелей, соединяющей станции, необходима пропускная способность от 2 до 20 Гб/с. Всего посредством данных станций будет объединено около 20 000 радиоантенн. Сигналы со всех этих станций будут объединяться при помощи суперкомпьютера, превращая массив телескопов в самый сложный радиотелескоп в мире (до полного запуска проекта SKA) с разрешением эквивалентным телескопу, составляющему 1000 км в диаметре. Благодаря такому колоссальному разрешению LOFAR за раз сможет исследовать огромные участки неба и работать одновременно над несколькими научными проектами. Главный компьютер проекта «Blue Gene/L», один из самых скоростных в мире, уже работает в университете в Гронингене. Его скорость в 27 терафлопов достаточна для преобразования данных, непрерывно поступающих от станций со скоростью около 500 Гбит в секунду, в

изображения в режиме реального времени. Объем памяти в 1 петабайт позволяет хранить данные для последующей обработки сигналов.

Самым глобальным астрономическим проектом, который разрабатывается в настоящее время, является The Square Kilometre Array (SKA) [12]. Одна из частей данного радиотелескопа, она называется ASKAP- Australian Square Kilometre Array Pathfinder («Австралийский поисковый телескоп площадью в квадратный километр»), уже запущена в режиме наблюдений в октябре 2012 года в Западной Австралии [13]. Система ASKAP состоит из 36 антенн диаметром 12 м (малый размер антенн позволяет быстро переводить с одной точки на небе на другую) площадью 4000 кв. м, работающих в диапазоне 700 – 1800 МГц (с полосой приема 300 МГц) и формирует 36 лучей, покрывающих 30 квадратных градусов. ASKAP называют сейчас «самым быстрым обзорным радиотелескопом в мире». Радиоинтерферометрическая система строит изображения практически в режиме реального времени при помощи мощных компьютеров мощностью в 2 петафлопа, потоки данных составляют 70 Тб/с [14].

В целом ядро радиотелескопа SKA будет состоять из нескольких тысяч антенн, при этом будет использоваться технология, позволяющая объединить приемные площади отдельных радиотелескопов в одну общую площадь размером в один квадратный километр. Часть антенн будут расположены в Западной Австралии и Новой Зеландии, часть - в Южной Африке. Выход на полную мощность сбора данных данным проектом планируется к 2024 году.

Рисунок 3. 3D схема проекта The Square Kilometre Array.

Утверждается, что с помощью SKA можно будет на расстоянии в 50 световых лет уловить излучающие сигналы мощностью, сравнимой с сигналом обычных радаров, используемых в аэропортах. Особое место в структуре SKA занимает обработка данных. Ожидается, что телескоп SKA после выхода на

полную мощность будет генерировать более 1 экзабайта информации в день, что сравнимо с объёмом всего существующего на данный момент интернет-трафика. В сутки будет требоваться сохранять до 1 петабайта сжатых данных. Огромные технологические проблемы содержит в себе и обработка этого сверхбольшого потока данных. Для достижения запланированных параметров, станции SKA должны быть связаны широкополосными оптоволоконными линиями связи со скоростью передачи 160 Гбит в секунду, а мощность центрального компьютера должна быть порядка 100 петафлопов.

Единственной технологией, которая способна удовлетворить растущие потребности по передаче научных данных, является волоконно-оптическая связь, использующая в качестве носителя информационного сигнала электромагнитное излучение оптического диапазона, а в качестве направляющих систем — волоконно-оптические кабели. Пропускная способность волоконно-оптических линий многократно превышает пропускную способность всех других систем связи. Каналы этих сетей уже сегодня способны обеспечить пропускную способность исчисляемую десятками гигабит в секунду, ведутся разработки и испытания каналов с пропускной способностью исчисляемой терабитами в секунду. Самая быстрая скорость передачи данных по оптоволокну на данный момент достигнута японскими компаниями Nippon Telegraph и Telephone Corporation (NTT), работавшими совместно еще с тремя партнерскими организациями, компанией Fujikura Ltd., университетом Хоккайдо и Датским техническим университетом (Technical University of Denmark, DTU). Их эксперимент в сентябре 2012 года продемонстрировал рекордную скорость передачи информации по одному оптоволоконному кабелю. Во время испытаний новой линии связи специалистами была зарегистрирована скорость передачи данных 1 петабит в секунду по оптоволоконному кабелю с 12 световодными каналами и длиной 52.4 километра. Это на порядки больше, чем показатель кабелей, находящихся сегодня в коммерческой эксплуатации [15]. Но уже сегодня в коммерческую эксплуатацию вводятся сети с большой пропускной способностью. Например, компания «Т8», отечественный лидер по разработке и внедрению магистральных сетей DWDM, объявила о достижении скорости передачи в 1 Тбит/с в одном пролете на 500,4 км с канальной скоростью 100 Гбит/с. Для передачи 10 каналов по 100 Гбит/с были использованы усилители с удаленной накачкой [16].

Стоит отметить и недавно установленный рекорд скорости передачи данных из космоса с помощью новой системы лазерной связи. Этот

эксперимент осуществлялся National Aeronautics and Space Administration (NASA) и осуществлялся с помощью космического зонда Lunar Atmosphere and Dust Environment Explorer (LADEE, см. рис. 4).

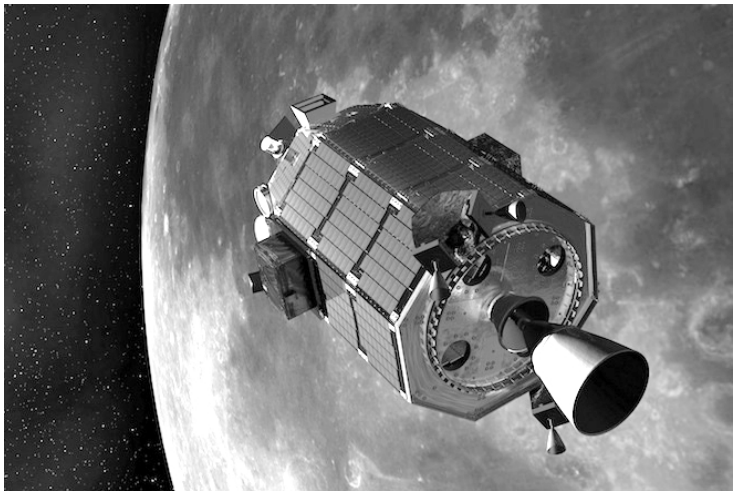

Рисунок 4. Космический зонд НАСА LADEE

С помощью лазерного луча удалось передать данные со скоростью 622 Мб/с (у «Радиоастрона» – 128 Мб/с). Данные передавались от станции в Нью-Мехико до зонда LADEE – его расстояние до поверхности Луны в настоящий момент составляет около 235 км. Основная задача эксперимента заключалась в подтверждении работоспособности новой технологии передачи данных. NASA пытается найти альтернативу радиоволнам, планируя в будущем высокоскоростную передачу трехмерных изображений, а также фото и видеоданных высокой четкости из космоса [17].

Помимо быстрой передачи данных немаловажной научной проблемой является передача информации на большие расстояния (например, для рассмотренных проектов LOFAR и SKA). Китайская компания ZTE и оператор China Telecom установили мировой рекорд дальности передачи данных по оптоволокну без промежуточных усилителей и преобразователей сигнала, говорится в сообщении компании. Передача на скорости 1 Тбит/с продолжалась 24 часа на расстояние 3,2 тысячи километров. Данные передавались в реальном времени [18].

Наконец, еще один амбициозный астрономический проект будет запущен в 2020 г – это Большой обзорный телескоп (Large Synoptic Survey Telescope, LSST, широкоугольный обзорный телескоп-рефлектор с апертурой 6.68 м), предназначенный для непрерывной обзорной съёмки всей доступной области неба [19]. На LSST предполагается каждые трое суток получать очередной фотографический обзор всего доступного для наблюдений неба в оптическом диапазоне. Архитектура LSST способна обеспечить очень широкое поле зрения: его диаметр – 3,5 градуса, а площадь – 9,6 квадратного градуса. Цифровая фотокамера с матрицей 3,2 гигапикселя (глубина цвета 16 бит) будет делать 15-секундные экспозиции каждые 20 секунд [20], соответственно каждые 20 секунд

будет добавляться очередные 6,4 ГБ данных (за ночь – 15 Тбт, в год до 6.8 Пбт). С учетом технического обслуживания, плохой погоды и т.д., фотокамера, как предполагается, реально будет производить около 200 000 фотографий (1,28 петабайта в несжатом виде) за год. Управление и эффективный интеллектуальный анализ столь огромного количества данных на выходе телескопа, как ожидается, будет наиболее технически сложной частью проекта. Первоначальные требования к вычислительному центру оцениваются в 60 терафлопс вычислительной мощности и 200 петабайт для хранения данных.

2. Вычислительные системы и системы хранения, работающие с большими объемами данных.

Далее рассмотрим, что именно представляют из себя вычислительные системы, предлагаемые ведущими производителями компьютерной техники для обработки больших объемов данных, и предоставляющие как возможности хранения больших объемов данных, в том числе распределенные, так и средства аналитики и параллельной обработки данных в реальном масштабе времени.

В статье, посвящённой анализу платформ обработки «больших данных» [21], рассматривается концепция идеальной платформы, полноценно решающей обработки и хранения «больших данных». Такое решение, по мнению автора, «должно обеспечивать возможность работы с данными всех типов и в произвольных форматах; иметь средства визуализации, обнаружения и представления в удобном для восприятия и поиска виде; включать инструменты аналитики, как в традиционном пакетном режиме, так и в режиме реального времени; предоставлять средства поддержки традиционных хранилищ данных и системы обработки потоков различных сведений без потерь времени на промежуточное хранение». Соглашаясь в целом с этими требованиями, отметим, что вероятно, компьютерные системы, которые мы приводим далее, не в полной мере отвечают всем заявленным характеристикам, но это лучшие решения для работы с «большими данными» из существующих на сегодняшний день. Специализированные компьютеры, предназначенные для аналитической работы с «большими данными», сегодня создают IBM, HP, SAP, Oracle, Teradata, Microsoft, EMC и другие мировые вендоры. Такие устройства в общих чертах представляют собой массивы хранения данных, снабженных дополнительными функциональными уровнями предварительной обработки и представления

информации.Корпорация Oracle после слияния с Sun разработала целую линейку таких продуктов. Впервые были созданы готовые вычислительные комплексы на основе специализированного программного и типового аппаратного обеспечения, оптимизированные для решения задач по обработке и хранению данных. Сначала появилась Exadata Database Machine – машина баз данных, предназначенная для кардинального повышения производительности работы баз данных. Следующей была выпущена Exalogic Elastic Cloud, оптимизированная для достижения максимальной скорости работы приложений и развертывания частных и публичных облаков. А недавно в продажу поступила «машина больших данных» Oracle Big Data Appliance, предназначенная для обработки огромных массивов неструктурированных данных. Важно, что во всех этих продуктах в единое целое объединены серверы, устройства хранения, сетевое оборудование и программное обеспечение. Они не требуют предварительной настройки и поставляются как полностью готовый к эксплуатации и простой в управлении программно-аппаратный комплекс. По сути,- это своего рода «строительные блоки» для создания центров обработки данных нового поколения – инфраструктуры, необходимой для развития облачных услуг [22].

Компания SGI объявила о выпуске в продажу суперкомпьютера SGI UV 2 с общим количеством вычислительных ядер до 4096, объемом когерентной основной памяти до 64 терабайт и объёмом общей памяти до восьми петабайт [23]. При пиковой скорости ввода-вывода до 4 терабайт в секунду и когерентной общей памяти, которая может работать в 1000 раз быстрее флэш-памяти, все эти функциональные возможности делают SGI UV самой мощной системой in-memory (концепция проведения вычислений в памяти) для всех задач, требующих обработки больших массивов данных. Таким образом, это идеальная платформа для ускорения инноваций в сферах поддержки принятия решений, геномики и биологических наук, химии и обработки материалов, физики, интегративной системотехники, национальной безопасности, проектирования изделий и в других областях, требующих обработки больших объемов данных. Центр геномного анализа (The Genome Analysis Centre, TGAC) и Центр биологических наук (Centre Biological Sciences, CBS) Датского технического университета, которые пользуются широким признанием каждый в своей области – секвенирования нового поколения и метагеномики соответственно – одними из первых начнут использовать SGI UV 2 в данных областях исследований.

Не осталась в стороне и корпорация IBM, объединившая все свои адекватные для этой сферы решения в единую платформу. В ее состав вошли: Netezza – специализированный программно-аппаратный комплекс на базе IBM x-Series, поставляемый с предустановленной одноименной СУБД и предназначенный для построения аналитических приложений и хранилищ данных объемом свыше 1 Пбайт; nfoSphere BigInsights – решение по анализу и обработке неструктурированных данных на основе технологий Hadoop; InfoSphere Streams и Vivisimo – средство анализа потоковой информации и комплексной обработки больших объемов неструктурированных данных [24]. IBM недавно также объявила о создании чипа, ускоряющего скорость интернет-соединения до 200-400 гигабит в секунду [25].

Компания Hitachi Data Systems предлагает два специализированных программно-аппаратных комплекса для решения задач построения систем хранения для Больших Данных: Hitachi Content Platform (HCP) — платформа для хранения контента, предназначенная для хранения и управления большими объемами неструктурированных данных и Hitachi Network Attached Storage (HNAS) — решение для обеспечения файлового доступа к данным, которое позволяет хранить и управлять большим количеством файлов [26]. HCP представляет собой программно-аппаратный комплекс, состоящий из узлов хранения на базе серверов x86 и внешней системы хранения общей емкостью до 40 Пбайт. Функционал платформы позволяет решать широкий спектр задач для хранения информации, обеспечения безопасности и доступности содержимого, а также создавать облачные хранилища и территориально распределенные файловые репозитории. Объектный подход к хранению контента, возможности по индексации больших объемов данных позволяют HCP работать с «большими данными» наиболее эффективно. Система хранения данных Hitachi Network Attached Storage – это интегрированное решение для работы непосредственно с локальной вычислительной сетью организации. Пользователи могут использовать HNAS для хранения своих документов и программ, а приложения (Microsoft Exchange Server, Microsoft SQL Server, Microsoft SharePoint, Oracle и др.) для хранения данных. Возможности динамического выделения пространства и иерархического хранения данных позволяют эффективно использовать дисковое пространство. Для пользователей это даёт ощутимый эффект при хранении и доступе к информации, а также при резервном копировании и восстановлении данных.

И в заключение данного краткого обзора представим специализированный компьютер корпорации EMC – Greenplum HD Data Computing Appliance. Новая модель обладает способностью получать информацию из «облаков» и работать с гигантскими массивами данных – эти качества обеспечивает встроенная в DCA поддержка интегрированной среды Hadoop [27]. В продукте EMC Greenplum HD DCA собраны вместе три главные технологии аналитической обработки больших объемов структурированных и неструктурированных данных: свободно распространяемая среда Apache Hadoop, СУБД EMC Greenplum Database 4.0 и аппаратная платформа EMC Greenplum HD DCA. От других спецмашин Greenplum HD DCA отличается интеграцией Hadoop с СУБД Greenplum, характеризуемой возможностью масштабирования до петабайт, эластичностью с точки зрения используемой аппаратной основы (серверы, СХД) и применимых аналитических методов. Стандартная модель DCA выпускается в трех модификациях: GP10 Quarter Rack, GP100 Half Rack и GP1000 Full Rack. В каждой из них есть два основных сервера (Master Servers) и 4, 8 или 16 рабочих серверов сегментов (Segment Servers) с общим числом процессорных ядер 48, 96 или 192 и с памятью 192, 384 или 768 Гбайт. В данной модели устанавливается 48, 96 или 192 дисков HDD SAS с некомпрессированной емкостью 9, 18 или 36 Тбайт и компрессированной емкостью 36, 72 или 144 Тбайт. Другая модель, High Capacity DCA (GP10C Quarter Rack, GP100C Half Rack и GP1000C Full Rack), ориентирована на большие объемы данных, но меньшую оперативность, поэтому отличается дисками : в ней устанавливаются 48, 96 или 192 дисков HDD SATA с некомпрессированной емкостью 31, 62 или 124 Тбайт и компрессированной емкостью 124, 248 или 496 Тбайт.

Обращает внимание тенденция на преимущественное использование в компьютерных платформах для работы с «большими данными» систем хранения данных, напрямую присоединённых к вычислительным узлам (DAS, Direct-attached storage – устройства хранения данных, непосредственно подключаемые к серверу или рабочей станции без помощи сети хранения данных). Это могут быть и твердотельные дисковые системы (SSD) и массивы традиционных HDD дисков, подключаемых с использованием различных интерфейсов. Возвращение моды на использование DAS-решений, после практически полного их вытеснения сетевыми решениями для хранения данных классов NAS (Network Attached Storage, сетевая система хранения данных) и SAN (Storage Area Network, сети хранения данных) связывают именно с необходимость работы с большими данными [28]. Неожиданный возврат к DAS стимулирован

распространением приложений, основанных на распределенных базах данных, подобных Hadoop, где узлы кластера, поддерживающего такое приложение, могут подключаться к дисковому массиву посредством различных компьютерных интерфейсов (SATA, SAS, SCSI или Fibre Channel), но в любом случае напрямую, а не по сети. В общем случае архитектуры хранения SAN и NAS, позволяющие разделять данные или неиспользуемые ресурсы с другими серверами в сети, являются относительно более медленными, сложными и дорогими. Эти качества несовместимы с требованиями к системам анализа «больших данных», приоритетными для которых являются производительность системы, удобство инфраструктуры и низкая стоимость. Необходимость обеспечения скорости анализа данных работы в режиме реального времени требует минимизации задержек передачи данных везде, где это возможно. Отсюда приоритет систем с обработкой данных в памяти.

В качестве альтернативы облачным технологиям для решения задач обработки больших объемов научных данных предлагается использование высокопроизводительных локальных кластеров научно-исследовательских центров и технология GRID [29]. Последняя представляет из себя концепцию, подразумевающую совместное использование научно-исследовательскими организациями своих вычислительных мощностей для интенсивных операций с научными данными. По своей сути, это разновидность распределённых вычислений, в которой вычислительные ресурсы различного типа объединяются вместе единой инфраструктурой для выполнения ресурсоемких заданий. Технология GRID успешно применяется для решения научных задач, требующих значительных вычислительных ресурсов. Преимуществом распределённых вычислений является то, что в качестве отдельных узлов GRID-системы могут использоваться даже обычные неспециализированные компьютеры. Таким образом, теоретически можно получить те же вычислительные мощности, что и на суперкомпьютерах, но с гораздо меньшей стоимостью. К сожалению, данная технология не подходит, когда возникает необходимость передачи для обработки на удалённый ресурс большого объёма информации из-за возможно недостаточной скорости передачи данных по имеющимся компьютерным сетям. Тем не менее, GRID-технология успешно применяется, например, для моделирования и обработки данных в экспериментах на уже упомянутом в данной статье Большом адронном коллайдере. Распределённая вычислительная система, предназначенная для обработки данных, получаемых с LHC, имеет иерархическую структуру. На

верхнем уровне расположен собственно компьютерный центр CERN, который, несмотря на его мощность, располагает лишь 20% требуемых вычислительных ресурсов. Поэтому остальные данные распределяются для хранения и обработки между компьютерными центрами по всему миру, в том числе и российскими вычислительными центрами [30].

ИМПБ РАН также активно участвует в GRID–инфраструктуре, являясь членом консорциума РДИГ (Российский грид для интенсивных операций с данными - Russian Data Intensive Grid, RDIG). В рамках этой организации свои вычислительные мощности для интенсивных операций с научными данными совместно используют целый ряд российских научно-исследовательских организаций в Москве, Санкт-Петербурге, Новгороде и подмосковных научных центрах. Консорциум РДИГ в свою очередь входит в структуру EGEE (Enabling Grids for E-sciencE, "Развёртывание гридов для развития е-науки") в качестве региональной федерации для обеспечения полномасштабного участия России в этом проекте. EGEE – это крупнейшая в мире грид-инфраструктура для выполнения задач в области многих дисциплин. В неё входят свыше 120 организаций. Они образуют надёжную и способную к расширению систему компьютерных ресурсов, доступных исследовательскому сообществу Европы и всего мира. Сейчас в ней участвуют 250 сайтов в 48 странах и более 68 тыс. компьютерных устройств; с ними могут работать круглосуточно 7 дней в неделю около 8 тыс. пользователей [31].

Теперь рассмотрим наиболее традиционное из решений, используемое в настоящее время для работы с «большими данными», а именно локальные компьютерные кластеры научно-исследовательских центров. Традиционно кластером принято называть несколько связанных между собой высокоскоростными каналами связи компьютеров, используемых как единый вычислительный ресурс. Вычислительные кластеры позволяют существенно уменьшить время расчетов, по сравнению с одиночным компьютером, разбивая задание на параллельно выполняющиеся ветки, которые обмениваются данными по связывающей узлы кластера сети. При этом имеется возможность построения относительно высокопроизводительных комплексов из обыкновенных недорогих компьютеров на основе бесплатного программного обеспечения и простых сетевых технологий. Именно кластерные вычислительные системы в последнее время лидируют в рейтинге наиболее высокопроизводительных компьютерных систем TOP500; это можно отследить по публикуемым раз в полгода последним выпускам TOP500. Для определенности разберем данные

для конца 2012 года, выпуск №40 [32] в сравнении с предыдущим. Самым высокопроизводительным суперкомпьютером в мире, достигнувшим в тесте Linpack производительности в 17.59 Petaflop/s, признан построенный в Национальной лаборатории Оук-Ридж (США) компанией Cray кластер Titan [33]. Кластер включает в себя 18688 16-ядерных процессоров Opteron 2.200GHz и столько же вычислительных акселераторов на базе GPU NVIDIA Tesla K20x. Самый производительный из отечественных кластеров Lomonosov (МГУ им. М.В. Ломоносова) [34] за полгода переместился с 22 на 26 место в рейтинге. Всего в Top500 конца 2012 года вошло 8 отечественных суперкомпьютеров, что на 3 больше, чем в предыдущей редакции рейтинга [35]. Одновременно был опубликован пятый выпуск альтернативного рейтинга кластерных систем Graph 500 [36], ориентированного на оценку производительности суперкомпьютерных платформ, предназначенных прежде всего для решения задач по обработке больших массивов данных. В отличие от теста Linpack, который демонстрирует в основном вычислительные возможности суперкомпьютеров, не отражая скорость обработки массивов данных, рейтинг Graph 500 нацелен прежде всего на оценку производительности обработки экстремально больших объемов данных в таких областях применения высокопроизводительных систем, как информационная безопасность (криптография), медицинская информатика, биоинформатика, социальные и нейронные сети. Первые позиции в этом рейтинге занимают суперкластеры Ливерморской национальной лаборатории им. Э. Лоуренса (Калифорния, США) – 65536 вычислительных узлов, 1048576 процессорных ядер и Аргоннской национальной лаборатории (Иллинойс, США) – 32768 вычислительных узлов, 524288 ядер [1].

Еще одна научная проблема,- это хранение большого объема получаемых экспериментальных данных. Хранение и первоначальная обработка (например, фильтрация) огромного объема данных невозможно в быстрой энергозависимой памяти DRAM. Поэтому компания IBM разработала память Storage Class Memory (SCM), гибридный вариант между памятью DRAM и технологией жёсткого диска. Данная память не является энергозависимой, предоставляет намного более быстрый, чем жесткий диск или флеш-память, доступ к данным, легко масштабируется и имеет более низкую стоимость за единицу хранения информации, чем жесткие диски. Единицей хранения информации для SCM является значение сопротивления в химических соединениях типа халькогенидов. Данное сопротивление меняется с изменением структуры материала с кристаллической на аморфную под воздействием нагрева, причем

значений сопротивления может быть несколько, что способствует более лучшему масштабированию, чем в памяти DRAM [37].

Для обычного хранения без обработки огромного объема архивных данных наиболее оптимальным с точки зрения соотношения цена/скорость доступа являются комплексы ленточных библиотек. Например, ленточная библиотека компании IBM System Storage TS3500, лидирующее в отрасли решение по интеграции ленточных накопителей, поддерживает IBM System Storage Tape Library Connector (коннектор-шлюз), который обеспечивает возможность соединить до 15 библиотек в библиотечный комплекс с совокупной емкостью 900 ПБ со скоростью доступа 250 Мб/с к картриджу [38].

Заключение

Таким образом, приведённые примеры научных астрономических проектов свидетельствуют о том, что в настоящее время для эффективного получения нового знания в условиях быстрого роста количества данных, производимых современными научными установками, необходимо обеспечить хранение и обработку огромного объема данных, а также наличие эффективно функционирующих современных высокопроизводительных компьютерных сетей глобального уровня, позволяющих гибко управлять крупномасштабными потоками данных и предоставляющих удаленный доступ исследователей к сложному научному оборудованию и вычислительным ресурсам в реальном масштабе времени. В связи этим необходимы принципиально новые технологические решения, позволяющие существенно увеличить скорость передачи данных с помощью уже существующих коммуникационных сетей, обеспечение хранения больших объемов данных с возможностью доступа к ним в режиме реального времени, а также строительство новых каналов связи, которые позволят справиться с передачей сверхбольших объёмов данных с учётом прогнозируемых темпов их роста.

Данная работа частично поддержана грантом РФФИ 14-07-00870а.

СПИСОК ЛИТЕРАТУРЫ

[1] Исаев Е.А., Корнилов В.В. Проблема обработки и хранения больших объемов научных данных и подходы к ее решению. Математическая биология и биоинформатика. 2013. Т. 8. № 1. С. 49–65.

[2] Исаев Е.А., Корнилов В.В., Тарасов П.А. Научные компьютерные сети – проблемы и успехи в организации обмена большими объемами научных данных. Математическая биология и биоинформатика. 2013. Т. 8. № 1. С. 161–181.
[3] Радиотелескопы ПРАО. Пущинская радиоастрономическая обсерватория. http://www.prao.ru/radiotelescopes/telescopes.php .
[4] Думский Д.В., Исаев Е.А., Китаева М.А., Пугачев В.Д., Самодуров В.А. Системы передачи и хранения научных данных ПРАО АКЦ ФИАН. Препринт №5, Москва, ФИАН, 2014.
[5] Думский Д.В., Исаев Е.А., Самодуров В.А. Локальная вычислительная сеть Пущинского научного центра. Препринт №23, Москва, ФИАН, 2012.
[6] Сайт проекта Радиоастрон: http://www.asc.rssi.ru/radioastron/index.html .
[7] Сайт Федерального космического агенства: http://www.federalspace.ru/185/
[8] Шацкая М.В., Гирин И.А., Исаев Е.А., Лихачев С.Ф., Пимаков А.С., Селиверстов С.И., Федоров Н.А. : Организация центра обработки научной информации для радиоинтерферометрических проектов. Космические исследования, 2012. Т. 50. № 4. С. 346–350.
[9] М.В. Шацкая, А.А. Абрамов, И.А. Гирин, Е.А. Исаев, В.И. Костенко, С.Ф. Лихачев, А.С. Пимаков, С.И. Селиверстов, Н.А. Федоров: Центр обработки научной информации проекта Радиоастрон: два года работы. Всероссийская Астрономическая Конференция «Многоликая Вселенная», тез. докл., 2013 г., Санкт-Петербург, с. 277.
[10] Официальный сайт проекта LOFAR: http://www.lofar.org/ .
[11] Официальный сайт Нидерландского института радиоастрономии: http://www.astron.nl/
[12] Официальный сайт проекта SKA: http://www.skatelescope.org/
[13] Официальный сайт проекта ASKAP:
http://www.atnf.csiro.au/projects/askap/index.html
[14] ASKAP Technologies: Digital Systems
http://www.atnf.csiro.au/projects/askap/digital_systems.html
[15] World Record One Petabit per Second Fiber Transmission over 50-km. Официальный сайт компании NTT Group, 2012.
http://www.ntt.co.jp/news2012/1209e/120920a.html
[16] Официальный сайт компании Т8: http://t8.ru/?p=4449
[17] Официальный сайт НАСА.
http://solarsystem.nasa.gov/missions/profile.cfm?MCode=LADEE&Display=ReadMore

[18] Официальный сайт корпорации ZTE.
http://wwwen.zte.com.cn/en/press_center/news/201310/t20131028_410952.html
[19] Официальный сайт проекта LSST: http://www.lsst.org/lsst/
[20] LSST Basic Configuration : http://www.lsst.org/lsst/science/survey_requirements
[21] Черняк Л. Платформы для Больших Данных. Открытые системы. 2012. № 07.
[22] Артемов С. Big Data: новые возможности для растущего бизнеса. «Инфосистемы Джет». http://www.jet.su
[23] Announcing the New SGI UV: The Big Brain Computer. Business Wire: http://www.businesswire.com/news/home/20120618005340/en
[24] Выходцев А. Платформа для Больших Данных. Открытые системы. 2012. № 06.
[25] Официальный сайт компании IBM:
http://www-03.ibm.com/press/us/en/pressrelease/43171.wss
[26] Яхина И. Хранилище для Больших Данных. Открытые системы. 2012. № 07.
[27] Серов Д. Машины для аналитиков. Открытые системы. 2011. № 04.
[28] Черняк Л. Большие данные возрождают DAS. Computerworld Россия. 2011. № 14.
[29] Eric E.S., Linderman M.D., Sorenson J., Lee L., Nolan G.P. Computational solutions to large-scale data management and analysis. Nat Rev Genet. 2010. V. 11. P. 647–657.
[30] Essers L. Фильтр секретов мироздания. Computerworld Россия. 2011. № 18.
[31] Проект EGEE-RDIG. URL: http://www.egee-rdig.ru
[32] TOP500 List of the world's top supercomputers. November 2012. URL: http://www.top500.org/lists/2012/11/ (дата обращения: 10.02.2013).
[33] URL: http://www.olcf.ornl.gov/titan/
[34] URL: http://parallel.ru/cluster/lomonosov.html
[35] The OpenNet Project.
URL: http://www.opennet.ru/opennews/art.shtml?num=35358
[36] The Graph 500 List. URL: http://www.graph500.org/
[37] Официальный сайт подразделения IBM Research. URL: http://researcher.watson.ibm.com/researcher/files/us-gwburr/Almaden_SCM_overview_Jan2013.pdf
[38] Официальный сайт компании IBM. URL:
http://www-03.ibm.com/systems/ru/storage/tape/ts3500/

Подписано в печать 27.02.2014 г.
Формат 60x84/16. Заказ № 12. Тираж 140 экз. П.л 1,25.
Отпечатано в РИИС ФИАН с оригинал-макета заказчика
119991 Москва, Ленинский проспект, 53. Тел. 499 783 3640