

ПРОЕКТ ЕВРОПЕЙСКОГО КОСМИЧЕСКОГО АГЕНТСТВА: СОСТАВЛЕНИЕ ТРЕХМЕРНОЙ КАРТЫ НАШЕЙ ГАЛАКТИКИ С ПОМОЩЬЮ АВТОМАТИЧЕСКОГО КОСМИЧЕСКОГО АППАРАТА GAIA И ТЕХНОЛОГИИ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ INTERSYSTEMS CACHÉ

Краткое описание

Европейское космическое агентство (European Space Agency, ESA) выбрало технологию InterSystems Caché®, как самую передовую в области управления базами данных, для разработки астрометрического решения (Astrometric Global Iterative Solution, AGIS), которое будет использоваться при анализе данных, полученных научно-исследовательским комплексом GAIA.

Целью запуска этого аппарата является создание точной карты Галактики, включающей около миллиарда астрономических объектов. Во время работы станции на орбите данное астрометрическое решение будет в итеративном режиме повышать точность всех выполняемых пространственных измерений, в конечном счете добиваясь уровня точности порядка 20 микросекунд.

В рамках проекта выдвинуты чрезвычайно жесткие требования к обработке данных, поэтому было решено создать пилотную версию системы. В ее разработке пригласили участвовать корпорацию InterSystems. Для доказательства верности выбранной концепции требовалось за 24 часа вставить 5 миллиардов Java-объектов по 600 байт каждый в базу данных InterSystems Caché. Работая под управлением Red Hat Enterprise Linux 5.5 на базе 8-ядерного 64-битного процессора Intel, СУБД Caché успешно справилась с этой задачей за 12 часов 18 минут, то есть загружая в среднем 112 000 объектов в секунду.

ПРОЕКТ ЕВРОПЕЙСКОГО КОСМИЧЕСКОГО АГЕНТСТВА: СОСТАВЛЕНИЕ ТРЕХМЕРНОЙ КАРТЫ НАШЕЙ ГАЛАКТИКИ С ПОМОЩЬЮ АВТОМАТИЧЕСКОГО КОСМИЧЕСКОГО АППАРАТА GAIA И ТЕХНОЛОГИИ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ INTERSYSTEMS CACHE

Вступление

Проекты, предусматривающие запуск космических аппаратов, долгосрочны. Обычно такие проекты реализуются в течение 15-20 лет, и для их успешного осуществления требуются надежные технологии обработки данных, хранения и манипулирования данными, обеспечивающие эффективную работу на протяжении длительного времени. Эти технологии также должны обеспечивать оперативный анализ критически важных данных для быстрого выполнения необходимых корректировок, если таковые потребуются.

Миссия GAIA считается на сегодняшний день самым крупномасштабным проектом в области астрономических исследований. Европейское астрономическое агентство выбрало СУБД InterSystems Caché® для поддержки работы проекта в части обработки исследовательских данных, которые будут получены во время полета автоматического космического аппарата GAIA.

Корпорация InterSystems и Европейский космический центр астрономии (European Space Astronomy Center, ESAC) с 2008 года проводят совместную работу по определению преимуществ, которые сможет обеспечить СУБД InterSystems Caché в контексте требований миссии GAIA к обработке данных. Кроме того, разрабатывается экономичная архитектура вычислительной системы, наиболее полно соответствующая жестким процессинговым требованиям в рамках этого крупномасштабного проекта.

ПРОЕКТ ЕВРОПЕЙСКОГО КОСМИЧЕСКОГО АГЕНТСТВА: СОСТАВЛЕНИЕ ТРЕХМЕРНОЙ КАРТЫ НАШЕЙ ГАЛАКТИКИ С ПОМОЩЬЮ АВТОМАТИЧЕСКОГО КОСМИЧЕСКОГО АППАРАТА GAIA И ТЕХНОЛОГИИ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ INTERSYSTEMS CACHE

Миссия GAIA

Запуск автоматического космического аппарата GAIA планируется осуществить с помощью ракеты-носителя «Союз-Фрегат» в 2012 году с космодрома, расположенного во французской Гвиане. За первые месяцы спутник будет выведен на орбиту во второй точке Лагранжа (L2) системы Земля-Солнце на расстоянии около 1,5 миллиона километров от Земли, а в течение следующих 5 лет будет проводить пространственные измерения. Цель — создание трехмерной карты нашей Галактики.

Спутник массой более 2 тонн оснащен двумя астрометрическими телескопами, обзорающими одновременно два участка небесной сферы для определения точных координат светил, в фокальной плоскости третьего телескопа установлен спектрограф для определения лучевых скоростей звезд. Кроме того, несколькими ПЗС-фотометрами предполагается проведение многоцветной фотометрии. В ходе полета космического аппарата GAIA будут выполнены наблюдения и измерения примерно для миллиарда астрономических объектов, каждый из которых будет исследоваться около 100 раз.

Помимо астрометрических и фотометрических исследований каждой звезды, GAIA измерит спектры около 150 миллионов источников излучения. Ожидается, что уровень точности астрометрических данных в итоговом каталоге составит порядка 20 микросекунд. Чтобы добиться такой точности, необходимо исключительно сложное программное обеспечение.

Все программное обеспечение для обработки данных в рамках миссии GAIA написано на языке Java, включая пакет программ для обработки результатов астрометрических исследований — Astrometric Global Iterative Solution (AGIS). Это специализированное программное решение способно в итерационном режиме повышать точность всех пространственных измерений. Поскольку GAIA, свободно вращаясь вокруг своей оси, будет проводить наблюдения, результаты которых будут соотноситься только с результатами других, выполненных этим же аппаратом наблюдений, то при обработке данных потребуются учитывать пространственное положение космического аппарата, характеристики его орбиты и скорость движения. После этого вся полученная системой информация может быть приведена в соответствие с данными Международной системы астрономических координат (International Celestial Reference System, ICRS). Эта задача (объем которой составит 10%–50% от всего объема обрабатываемых в рамках проекта данных) будет выполнена с помощью специализированного решения AGIS на базе технологии InterSystems Caché.

Научные цели миссии GAIA многочисленны, но в целом они направлены на выяснение структуры нашей Галактики и истории ее формирования.

Технические задачи и требования

Ожидается, что аппарат GAIA сможет пронаблюдать около 1 миллиарда звезд, а с учетом того, что по каждому объекту за 5 лет наблюдений будет получено около 100 оценок координат и скоростей движения, GAIA выполнит около 100 миллиардов измерений. Часть данных (10%–50%) будет использована для создания глобальной инерциальной системы координат (global reference frame), наиболее точной из существующих, с помощью AGIS. После калибровки данных к этой системе координат будут привязываться измеренные пространственные координаты и скорости остальных объектов, которые войдут в результирующую базу данных GAIA. На рис. 1 показана обобщенная схема задач, относящихся к потокам данных между основной базой данных (Main Database) и БД AGIS Caché.

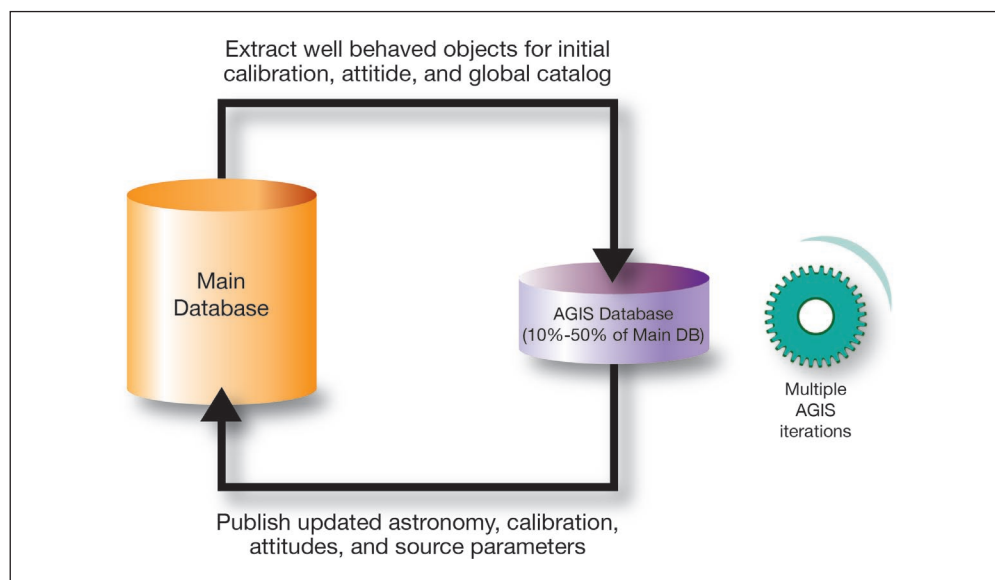


Рис. 1. Обобщенная схема задач по извлечению и обработке данных с использованием AGIS

Согласно проведенной оценке, БД AGIS Caché будет содержать данные о приблизительно 100 000 000 космических объектах (по результатам около 10 000 000 000 измерений). Объем этих данных ориентировочно составит 20 терабайт. Однако недавно было высказано предположение, что БД AGIS могла бы содержать данные об объектах количеством до 500 000 000 (по результатам 50 000 000 000 измерений), что увеличивает объем БД до 100 Тб. Эти данные необходимо поместить в базу данных за 7 дней, поэтому их обработка может начаться немедленно.

Ожидается, что после помещения данных в AGIS потребуется около 40 итераций для полной калибровки и корректировки данных, а все итерации должны быть выполнены за 120 дней. По завершении корректировки данные из AGIS переносятся в основную базу данных, и начинается следующий цикл. Такая итерационная обработка данных будет продолжаться в течение всего срока миссии GAIA. Более того, весь процесс обработки полученных данных носит итеративный характер: уточнение информации о пространственном положении звезд, полученной из AGIS, даст возможность получать более точные результаты по другим процессам, таким как фотометрические измерения светил. Эти данные в свою очередь будут использованы для создания следующего, более совершенного решения AGIS.

Загрузка данных в БД AGIS Caché

Объектная модель данных AGIS состоит из набора Java-классов. В частности, в AGIS каждое астрономическое наблюдение считается относящимся к отдельному «астро-элементарному» объекту (AstroElementary Object). Как показано на рис. 2, объект типа «AstroElementary» обладает различными свойствами (в основном принадлежащими к типу данных «Long» по стандарту IEEE) и имеет объем около 600 байт.

Кроме того, БД AGIS содержит несколько вспомогательных индексов, которые создаются на этапе вставки данных в БД. Применение этих индексов облегчает выполнение запросов во время обработки данных в AGIS, а также обеспечивает быстрое составление специальных отчетов.

С использованием InterSystems Caché, вместе с интерфейсом Caché eXTreme для Java-объектов, Java-программы AGIS будут загружать 100 Тб данных, сгенерированных научно-исследовательским комплексом GAIA как 50 000 000 000 отдельных объектов типа «AstroElementary». Вставку этих данных, наряду с созданием вспомогательных индексов, требуется завершить за 5 дней, чтобы добиться показателя скорости загрузки, равного 115 000 объектов типа «AstroElementary» в секунду.

```
class AstroElementary {
    long transitTimes[];
    long transitTimeErrors[];
    long HEALPIXID;
    long HTMID;
    double etaObs[];
    double zeta.Res;
    double[] sourceParam;
    double[] eta.Res;
    double zeta;
    double zetaError;
    float flux;
    float fluxError;
    float bg;
    float bgError;
    long id;
    long telescope;
    long codRow;
    short pixelColumns[];
    long detTime;
    long detTimeError;
    int typeFlag;
    long sourceId;
}
```

Рис. 2. Объект типа «AstroElementary»

Пробная оценка скорости загрузки данных

В рамках пробной оценки скорости загрузки данных в БД корпорация InterSystems и Европейский космический центр астрономии совместно со специалистами из компании NetApp разработали испытательную модель для загрузки 5 000 000 000 объектов типа «AstroElementary», что составляет около 10% от совокупного объема данных, которые, как ожидается, будут находиться в БД AGIS после завершения миссии GAIA. Первоначально предполагалось, что эти данные будут вставлены в БД за 12 часов. Однако вследствие того, что при проведении этой оценки применялось не совсем адекватное аппаратное обеспечение, Европейский космический центр астрономии решил, что пробные испытания будут считаться успешными, если эти данные будут инкорпорированы в БД за 24 часа.

В таблице 1 приведены обобщенные сведения о тестовой системе, которая применялась для проведения упомянутой пробной оценки.

Сервер	One 8-core Intel-based system
Операционная система	Red Hat Enterprise Linux 5.5 (2.6.18-194.el5), 64-bit
Оперативная память	32GB RAM → 11GB выделено для СУБД Caché (global buffers)
Файловая система	ext3
Система хранения	NetApp FAS3160 with 176 x 1 TB SATA disks@7200 RPM
Сеть	10 GigE, single-port, single-channel, Jumbo Frames enabled
Интерфейс и скорость доступа к дисковому массиву	iSCSI over 10 GigE
Версия СУБД Caché	2010.2

Таблица 1. Общая информация об архитектуре тестовой системы

С использованием технологии Caché eXTreme для Java модуль Test Harness оказался способен вставить 5 000 000 000 объектов типа «AstroElementary» за 12 часов 18 минут, то есть со средней скоростью загрузки, равной 112 000 объектов в секунду.

Число вставленных объектов типа «AstroElementary»	5 000 000 000
Общее время загрузки (Total run time)	44 616 с (~12,5 ч.)
Запланированное время загрузки (Target (allotted) run time)	86 400 с (24 ч.)
Средняя скорость загрузки (Average ingestion rate)	112 000 объектов/с

Таблица 2. Сводка результатов пробной оценки скорости загрузки данных

Результаты пробного испытания были восприняты как исключительно успешные в особенности потому, что все данные были вставлены в БД на 50% быстрее запланированного 24-часового промежутка времени при номинальной системной конфигурации.

Будущие испытания, возможно, будут проводиться с применением многочисленных параллельно работающих программ для вставки данных, что позволит еще увеличить скорость загрузки и сократить необходимое для ее выполнения время.

Заключение

В проекте пробной оценки скорости загрузки данных в БД, осуществленном Европейским космическим агентством и корпорацией InterSystems, дискретные астрометрические данные были вставлены в БД InterSystems Caché со средней скоростью 112 000 объектов в секунду. Вся загрузка в рамках этого теста была завершена за 12 часов 18 минут, т. е. затрачено времени вдвое меньше выделенных 24 часов для вставки данных в БД с использованием номинального тестового аппаратного обеспечения. Этот результат стал еще одним подтверждением обоснованности выбора СУБД Caché для выполнения требований к обработке данных в реализации крупнейшего астрометрического проекта — составления карты нашей Галактики по результатам космической миссии GAIA.

Информация о СУБД Caché

InterSystems Caché® — высокопроизводительная объектная система управления базами данных с поддержкой SQL и возможностью прямого доступа к многомерным данным. СУБД Caché применяется во многих странах мира как платформа для создания передовых приложений для учреждений государственного сектора, здравоохранения, поставщиков финансовых услуг, предприятий энергетики, телекоммуникаций, розничной торговли и других вертикальных рынков.

Информация о Caché eXTreme для Java

Caché eXTreme для Java — новая инновационная технология в арсенале InterSystems Caché. Технология предоставляет Java-приложениям, которым требуется обеспечивать высокую производительность при обработке потока данных, возможность высокоскоростного доступа к InterSystems Caché. Взаимодействие Java и Caché происходит через технологию JNI, что позволяет практически без задержек сохранять данные и производить исключительно быстрый поиск требуемых данных.

Более подробная информация приведена на веб-сайте по адресу: InterSystems.com/java

Информация о Европейском космическом агентстве

Европейское космическое агентство (European Space Agency, ESA) можно назвать окном из Европы в космос. Главная цель этой организации — придать конкретные очертания процессу дальнейшего развития области космических исследований и обеспечить получение практической выгоды жителями Европы и всего мира от инвестиций в космические проекты. Задача агентства — составить европейскую космическую программу и осуществить ее. Инициативы этой организации направлены на то, чтобы мы смогли больше узнать о Земле, об околоземном пространстве, нашей Солнечной системе и о Вселенной, а также ее усилия ориентированы на разработку основанных на применении космических спутников технологий и услуг, которые способствовали бы успешной деятельности европейских предприятий. Европейское космическое агентство тесно сотрудничает с аналогичными неевропейскими организациями.

Агентство имеет представительства в ряде стран Европы, каждое из которых отвечает за решение поставленных перед ним задач. Европейский космический центр астрономии (European Space Astronomy Center, ESAC) входит в структуру Европейского космического агентства и осуществляет научно-исследовательскую деятельность. Этот центр расположен в окрестностях Мадрида в Испании. Европейский космический центр астрономии — то место, где инициируются действия, связанные с работой космических телескопов, и где хранятся все научные данные, которые собираются по результатам проведенных наблюдений и которые доступны для широкого ознакомления.

Информация о корпорации InterSystems

Корпорация InterSystems — один из ведущих в мире поставщиков ИТ-платформ. Корпорация имеет головной офис в Кембридже, штат Массачусетс, США, и зарубежные представительства в 23 странах, включая Россию. InterSystems предоставляет инновационные технологические решения, применяющиеся для создания передовых программных приложений. InterSystems Caché® — высокопроизводительная объектная система управления базами данных, обеспечивающая высокую скорость разработки приложений. InterSystems Ensemble® — платформа для интеграции и разработки корпоративных приложений. InterSystems HealthShare® — платформа для обмена медицинской информацией, позволяющая быстро создавать национальные и региональные системы ведения «электронной медицинской карты». InterSystems DeepSee® — платформа для разработки решений бизнес-аналитики.

Дополнительную информацию можно получить на сайте корпорации по адресу: InterSystems.com

InterSystems
Москва, 123610
Краснопресненская наб., 12, ЦМТ-2
Тел.: +7 495 967-00-88
Факс: +7 495 967-00-18
InterSystems.ru

INTERSYSTEMS
Advanced technologies
for breakthrough applications