

Физико-математические науки

СОВРЕМЕННЫЕ GRID – ТЕХНОЛОГИИ

Абрамовский В.А.
 Новгородский государственный
 университет им. Ярослава Мудрого,
 Великий Новгород

Россия участвует в крупнейшем международном проекте по физике высоких энергий «Большой адронный коллайдер» (LHC, ЦЕРН, Швейцария). Начало работы ускорителя (получение первых данных) планируется на конец 2007 г.

Первая фаза вычислительного проекта GRID для LHC была одобрена на Совете ЦЕРН. Сегодня важно не только иметь доступ к информации, но и распределенным образом обрабатывать ее. Четыре гигантских детектора этого ускорителя – ALICE, ATLAS, CMS и LHCb – будут накапливать больше чем 10 миллионов гигабайт данных в течение каждого года о событиях, происходящих при столкновении частиц. Это эквивалентно содержанию, примерно, 20 миллионов компьютерных компакт – дисков.

Ведущие российские исследовательские центры различных ведомств ведут работы по проекту LHC, как по созданию самого ускорителя, так и всех четырех его детекторов. На выполнение соответствующих заказов по проекту LHC задействованы десятки российских заводов. В целом вклад России в проект LHC можно оценить в размере 5%.

Научно-исследовательские центры РФ, участвующие в проекте «Большой адронный коллайдер», разрабатывают два основных направления развития GRID технологий:

- интенсивные операции с базами данных, *data intensive GRID* (проект DataGRID);

- вычислительный *computational GRID* (проект EuroGRID), в котором создаваемая инфраструктура нацелена на достижение максимальной скорости расчетов за счет глобализации распределения (распараллеливания) вычислений.

Концепция **GRID** (название по аналогии с электрическими сетями - *electric power grid*) предполагает создание компьютерной инфраструктуры нового типа, обеспечивающей *глобальную интеграцию информационных и вычислительных ресурсов* на основе управляющего и оптимизирующего программного обеспечения (middleware) нового поколения. Для достижения этой цели создается набор *стандартизированных служб* для обеспечения надежного, совместимого, дешевого и всепроникающего доступа к географически распределенным высокотехнологичным информационным и вычислительным ресурсам – отдельным компьютерам, кластерам и суперкомпьютерным центрам, хранилищам информации, сетям, научному инструментарию и т.д.

Важнейшим является междисциплинарный характер GRID. Имеется довольно много общего в вычислительных потребностях различных областей научных исследований – развиваемые технологии применяются в физике высоких энергий, космофизике, микробиологии, экологии, метеорологии, различных инженерных приложениях (например, в самолето-

строении). Схожие проблемы наблюдаются и в других областях. Например, NASA реализует для своих нужд сеть высокопроизводительных компьютеров, роботизированных устройств массовой памяти, высокоскоростных каналов связи, научных инструментов и продвинутых интерфейсов для пользователя под названием Information Power Grid.

В настоящее время кроме LHC идет подготовка нескольких научных экспериментов нового поколения – эксперименты с использованием интерферометров для регистрации гравитационных волн бинарных пульсаров, новых сверхсвезд и иных экзотических объектов (эксперимент LIGO), а также автоматизированная цифровая космическая съемка с очень высоким разрешением (более 1012 пикселей), которая позволит значительно развить систематическое изучение звезд, галактик и крупномасштабных космических структур (эксперимент SDSS) для создания подробного каталога астрономических данных. Все эти эксперименты рассчитаны на длительный период и предполагают накопление и последующую обработку массивов данных.

Требования к вычислительным и архивным ресурсам для этих экспериментов различны. Процессорные затраты, необходимые для LIGO составляют несколько петафлопов. Объемы данных у LHC будут значительно больше, чем у LIGO, а у LIGO – значительно больше, чем у SDSS.

Среди основных направлений использования *computational GRID* на данный момент можно выделить:

- распределенные супервычисления, решение очень крупных задач, требующих огромных процессорных ресурсов, памяти и т.д.;

- «высокопоточные» вычисления (High - Throughput Computing), позволяющие организовать эффективное использование ресурсов для небольших задач, утилизируя временно простаивающие компьютерные ресурсы;

- вычисления «по требованию» (On-Demand Computing), крупные разовые расчеты;

- вычисления с привлечением больших объемов распределенных данных (Data-Intensive Computing), например, в метеорологии, астрономии, физике высоких энергий;

- коллективные вычисления (Collaborative Computing).

К общим характеристикам потребностей, которые делают оправданной организацию вычислительных архитектур типа *Data intensive GRID*, можно отнести следующие:

- большие объемы данных, распределенных по различным научным центрам, странам и континентам;

- участие большого количества специалистов в обработке данных из разных институтов и университетов;

- информация, которую следует проанализировать, имеет сложную структуру;

- алгоритмы обработки информации имеют нетривиальный характер (объем программ составляет миллионы строк текста);

- наконец, масштабируемость базового программного обеспечения (фактически, всего того, что лежит ниже прикладного уровня), которое должно устойчиво работать как на настольной машине, так и на суперкомпьютере.

Как прогнозируется, эволюционные изменения в полупроводниковых технологиях и архитектуре микропроцессоров приведут в ближайшие пять лет к десятикратному увеличению вычислительных мощностей. Уже сегодня возможности рядовых пользователей, подключенных к цифровым каналам связи с предоставлением комплексных услуг, сравнимы с теми возможностями, которыми обладали суперкомпьютерные центры 10-15 лет назад.

Технологическое основание для создания Grid – инфраструктур дают уже существующие волоконно-оптические сети, высокопроизводительные процессоры, параллельные архитектуры, протоколы связи, математическое обеспечение распределенных структур, механизмы обеспечения безопасности.

В НовГУ создана лаборатория “GRID-ТЕХНОЛОГИИ В СОВРЕМЕННОЙ ФИЗИКЕ”.

Ее задачами являются:

1. Подготовка научных сотрудников, программистов и инженеров по направлениям:

- обработка и анализ экспериментальных данных с ускорителя LHC;

- работа по созданию программного обеспечения для проекта DATAGRID,

- работа по созданию программного обеспечения для моделирования физических процессов;

- работа по проекту распределенных вычислений EuroGRID.

2. Проведение работ:

- создание и развитие российского сегмента DATAGRID;

- моделирование экспериментов на детекторах ALICE, ATLAS, CMS;

- обработка экспериментальных данных с этих детекторов, а также с детектора TOTEM.

3. Создание программного обеспечения для DATAGRID и EuroGRID.

4. Создание программного обеспечения для моделирования физических процессов взаимодействия адронов и ядер при сверхвысоких энергиях.

5. Создание программного обеспечения для триггеров редких процессов в рассеянии адронов и ядер.

На основе этих конкретных задач лаборатория также может готовить специалистов по использованию технологии GRID в других областях науки и техники, в частности, в экологии, экономике, энергетике, машиностроении, медицине, биологии.

Новгородский государственный университет включен в сеть Grid. Вычислительные ресурсы доступны для использования всеми участниками сети.

Для этого были осуществлены следующие работы:

1. Произведена установка и настройка

вычислительного элемента сети GRID и сопутствующих сервисов, а именно:

- Computing Element – система управления вычислительными ресурсами, распределением заданий, аутентификацией хостов и пользователей сети,

- Storage Element – система хранения исходных экспериментальных данных и данных, полученных в результате обработки,

- Monitoring BOX – распределенная система мониторинга отдельных хостов сети GRID и сети в целом,

- Сервисы: SSH, Firewall, VPN, Hosts Autoupdate (система обновления ПО с помощью apt-get).

2. Все имеющиеся хосты зарегистрированы в сети GRID, для них получены соответствующие OpenSSL – сертификаты в Regional Certification Authority.

3. Получен сертификат пользователя, который был зарегистрирован в виртуальной организации RDIG (Russian Data Intensive Grid).

4. Локально устранены мелкие недоработки системы автоматического развертывания сайтов (совокупности сервисов в рамках одного вычислительного центра), связанных с тем, что ПО для сети GRID находится еще в стадии разработки и предварительного тестирования.

Работа представлена на научную конференцию «Новые технологии и современные системы автоматизации», Тунис, 12-19 июня 2005 г., поступила в редакцию 29.04.2005г.

ВЕРОЯТНОСТНЫЕ ИГРЫ НА МЕДИАНУ

Афанасьев В.В., Суворова М.А.

Ярославский государственный педагогический университет имени К.Д.Ушинского, Ярославль

Изучение теории вероятностей через рассмотрение различных азартных игр вызывает интерес у студентов и учащихся. В основе одного из таких подходов лежит нахождение числовых характеристик положения случайных величин [1.С.81-86]. Большая часть задач сводится к вычислению моды или математического ожидания, тем более, что они удобны для аналитических преобразований. А вот задачи, в которых выбор стратегии зависит от нахождения медианы, в литературе встречается крайне редко. Напомним, что медианой дискретной случайной величины $X = \{x_i\}$ ($P\{X = x_i\} = p_i$) называется такое

значение x_k , что $\sum_{i=1}^k p_i \geq \frac{1}{2}$ и $\sum_{i=k}^n p_i \geq \frac{1}{2}$.

В работе предлагается система задач, инициированных одной идеей, и её обобщение. Такое изложение может являться и иллюстрацией идеи развивающего обучения Д.Б. Эльконина – В.В. Давыдова, в которой утверждается: «для того, чтобы прийти к какому-либо обобщению при таком подходе, необходимо решить достаточно большое количество задач, постепенно выделяя «общие» для всех задач черты.