

Грид-технологии в науке

Рассмотренные вопросы:

Актуальность и основные шаги создания системы грид-образования.

Типы грид-систем

Сервисы GRID

Архитектура GRID

Базовые грид-технологии для организации учебного процесса

Грид-технологии и образование в Украине

Проект ИПСА НТУУ “КПИ” на 2011 год

Суперкомпьютерный вычислительный комплекс
Института кибернетики им. В. М. Глушкова НАН Украины

World Community Grid

Актуальность и основные шаги создания системы грид-образования

В настоящее время происходит стремительное развитие технологий грид с целью создания согласованной, открытой и стандартизированной технологической среды, обеспечивающей гибкое, безопасное, скоординированное совместное использование вычислительных ресурсов глобальной сети для решения сложных и ресурсоемких задач в важнейших областях современной науки и техники.

Формально авторами концепции грид считаются Ян Фостер из Арагонской национальной лаборатории Чикагского университета и Карл Кессельман из Института информатики Университета Южной Калифорнии. Именно они в 1998 году впервые предложили термин грид-компьютинг для обозначения универсальной программно-аппаратной инфраструктуры, объединяющей компьютеры и суперкомпьютеры в территориально-распределенную информационно-вычислительную систему. По их определению, ставшему уже классическим, «Грид - согласованная, открытая и стандартизованная среда, которая обеспечивает гибкое, безопасное, скоординированное разделение ресурсов в рамках виртуальной организации».

Грид – географически распределенная инфраструктура, объединяющая множество ресурсов разных типов (процессоры, долговременная и оперативная память, хранилища и базы данных, сети), доступ к которым пользователь может получить из любой точки, независимо от места их расположения. Грид предполагает коллективный разделяемый режим доступа к ресурсам и к связанным с ними услугам в рамках глобально распределенных виртуальных организаций, состоящих из предприятий и отдельных специалистов, совместно использующих общие ресурсы. В каждой виртуальной организации имеется своя собственная политика поведения ее участников, которые должны соблюдать установленные правила. Виртуальная организация может образовываться динамически и иметь ограниченное время существования.

В основе грид лежат программные технологии, использующие новые стандарты и протоколы совместно с известными сетевыми и интернет-протоколами.

Потенциал технологий грид уже сейчас оценивается очень высоко: он имеет стратегический характер, и в близкой перспективе грид должен стать вычислительным инструментарием для развития высоких технологий в различных сферах человеческой деятельности, подобно тому, как подобным инструментарием стали персональный компьютер и интернет.

Создание распределенных грид-приложений относится к наукоемким задачам и является существенно более сложным процессом по сравнению с созданием обычных последовательных программных систем. В связи с этим весьма актуальным становится задача развертывания системы подготовки высокопрофессиональных кадров в области распределенных вычислений и грид-технологий в рамках существующей системы высшего профессионального образования.

Различают следующие классы грид:

- Computational Grid - грид ориентированный на распределенные вычисления;
- Data Grid - грид ориентированный на обработку больших потоков данных;
- Informational Grid - грид ориентированный на интеграцию крупных распределенных хранилищ (OGSA-DAI), в подобных архитектурах используется централизованный реестр, хранящий метаданные всех сервисов и распределенных хранилищ;
- Hybrid Grid - грид сочетающий в себе как Computational/Data Grid так и Informational Grid;
- Semantic Grid - это любой из описанных типов грид-архитектур в котором описывается семантика ресурсов (интерфейсы, характеристики производительности, особенности безопасности).

Типы грид-систем

В настоящее время выделяют три основных типа грид-систем:

- Добровольные гриды — гриды на основе использования добровольно предоставляемого свободного ресурса персональных компьютеров;
- Научные гриды — хорошо распараллеливаемые приложения программируются специальным образом (например, с использованием Globus Toolkit);
- Гриды на основе выделения вычислительных ресурсов по требованию — обычные коммерческие приложения работают на виртуальном компьютере, который, в свою очередь, состоит из нескольких физических компьютеров, объединённых с помощью грид-технологий.

Сервисы GRID

Распределенная вычислительная среда должна обладать набором сервисов, обеспечивающих контролируемое выполнение прикладных программ авторизованных пользователей. Специфика сервисов GRID определяется, прежде всего, характером самой вычислительной среды: GRID представляет собой динамическое образование. Для нормального функционирования распределенной системы необходимо обеспечить:

- идентификацию выполняемой программы

Для обеспечения контроля выполнения прикладной программы она должна быть снабжена уникальным номером (идентификатором). Присвоение такого идентификатора и контроль его уникальности должны выполняться специальным сервисом GRID;

- авторизацию пользователя

Сервис авторизации пользователя должен обеспечить уникальность идентификатора пользователя на всех вычислительных узлах, а также определять относительный приоритет пользователя, необходимый для нормальной работы службы распределения вычислительных ресурсов между прикладными программами;

- поиск ресурсов

Результаты работы этого сервиса определяют размеры доступного в настоящий момент времени вычислительного ресурса;

- описание ресурсов

Сервис осуществляет единое описание разнородных ресурсов, находящихся в текущий момент времени в составе GRID;

- резервирование ресурсов

Данный сервис должен обеспечить «захват» свободных ресурсов для размещения на нем прикладной программы;

- выполнение распределённых алгоритмов.

Этот сервис обеспечивает выполнение параллельных программ.

- доступ к удалённым данным.

Сервис обеспечивает работу распределенных баз данных

- распределение ресурсов.

Осуществляет разделение существующего ресурса между прикладными программами;

- обнаружение неполадок

Сервис определяет работоспособность включенных в вычислительный процесс узлов.

Архитектура GRID

При описании различных уровней архитектуры GRID, будем следовать модели “песочных часов”. Как показано на рисунке 1, горловина песочных часов соответствует небольшому количеству базовых абстракций и протоколов (таких как TCP и HTTP в Internet), которые необходимы для связи высокоуровневых служб с низкоуровневыми.

Детализируя рисунок 1, архитектуру GRID можно представить в виде иерархической структуры (рисунок 2), состоящей из нескольких уровней. На каждом из представленных уровней существуют свои сервисы, взаимодействующие посредством определенных протоколов.



Рисунок 1 - Модель песочных часов

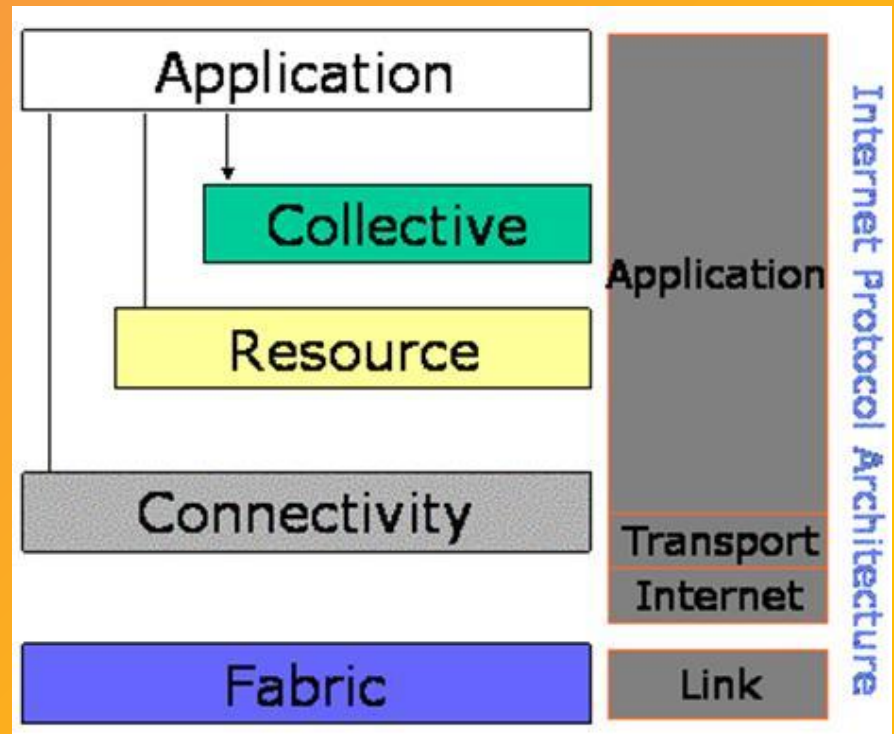


Рисунок 2 - Многоуровневая Архитектура GRID

Многоуровневая Архитектура GRID

Уровень **Fabric** предоставляет ресурсы, совместный доступ к которым обеспечивается через протоколы GRID.

Уровень **Connectivity** определяет базовые коммуникационные и идентификационные протоколы, требуемые для проведения специфичных для GRID операций (транзакций). Коммуникационные протоколы позволяют осуществлять обмен данными между ресурсами уровня Fabric.

Уровень **Resource** базируется на коммуникационном и авторизационном протоколах уровня Connectivity. Он определяет протоколы для: осуществления безопасного обмена информацией; создания учетных записей пользователей.

На уровне **Collective** сгруппированы протоколы и сервисы, которые не связаны с каким-либо конкретным ресурсом, являются более глобальными по природе и обеспечивают коллективное взаимодействие ресурсов. Примерами сервисов уровня Collective, служат: сервисы распределения и планирование ресурсов; сервис загрузки и совместных работ.

Области применения грид

Применение грид может дать новое качество решения следующих классов задач:

- массовая обработка потоков данных большого объема;
- многопараметрический анализ данных;
- моделирование на удаленных суперкомпьютерах;
- реалистичная визуализация больших наборов данных;
- сложные бизнес-приложения с большими объемами вычислений.

Грид-технологии уже активно применяются как государственными организациями управления, обороны, сферы коммунальных услуг, так и частными компаниями, например, финансовыми и энергетическими. Область применения грид сейчас охватывает ядерную физику, защиту окружающей среды, предсказание погоды и моделирование климатических изменений, численное моделирование в машиностроении и авиастроении, биологическое моделирование, фармацевтику.

Базовые грид-технологии для организации учебного процесса

Важным вопросом создания образовательной системы в области грид и распределенного компьютеринга является выбор грид-технологий для учебного процесса и оснащение ими грид-инфраструктуры. Одно из требований к оснащению гридами учебной базы - это обеспечение возможности для учащихся работать с гридами разных классов, в частности, как с вычислительными, так и информационными.

Первый подход, получивший название сервисного грида (Service Grid), предполагает развертывание распределенной сервис-ориентированной инфраструктуры, обеспечивающей унифицированный удаленный доступ к выделенным ресурсам уровня кластеров или суперкомпьютеров. Поставщиками ресурсов в подобных системах являются достаточно крупные организации, обладающие ресурсами указанного уровня. Включаемые в сервисный грид ресурсы являются гомогенными, то есть функционируют под управлением одной версии ОС и предоставляют одинаковое окружение для запускаемых заданий. Число пользователей сервисных гридов гораздо больше числа поставщиков ресурсов. При этом каждый пользователь может использовать ресурсы грида для запуска своих приложений.

Примерами сервисных гридов являются EGEE, NorduGrid, TeraGrid. Базовым промежуточным ПО подобных систем служат технологии Globus Toolkit, gLite, ARC, UNICORE. Недостатком сервисных гридов является высокая сложность установки и администрирования указанного ПО, что ограничивает круг потенциальных поставщиков ресурсов.

Второй подход, так называемый грид рабочих станций (Desktop Grid), предполагает использование ресурсов большого количества простаивающих персональных компьютеров, подключенных к сети. Поставщиками ресурсов в подобных системах являются рядовые пользователи. Подключаемые в грид ресурсы рабочих станций являются гетерогенными по своей архитектуре и программному обеспечению. При этом данные ресурсы, как правило, доступны не постоянно, а только в моменты их простоя. Поэтому, в отличие от сервисных гридов, состав ресурсов грида рабочих станций является гораздо более динамичным. В подобных системах число поставщиков ресурсов обычно гораздо больше числа пользователей, использующих ресурсы грида для запуска приложений. Примерами технологий для организации грида рабочих станций являются BOINC, Condor, XtremWeb. В отличие от технологий сервисных гридов, данные технологии позволяют легко и быстро подключать к системе новые ресурсы.

Грид-технологии и образование в Украине

Кабинет министров одобрил концепцию Государственной целевой научно-технической программы внедрения и применения грид-технологий на 2009-2013 годы. По словам министра экономики Украины Богдана Данилишина, концепция разработана Минэкономки с целью решения вопроса отсутствия в Украине механизма для обработки больших информационных массивов в разных сферах социально-экономической деятельности. «В ведущих странах мира такая проблема решается именно с помощью грид-технологий... Создание подобных систем радикально улучшит эффективность использования совокупных компьютерных ресурсов страны, предоставит принципиально новые возможности для решения сложных задач и проблем науки, здравоохранения, экономики, образования». В рамках программы к 2012 году планируется создание на базе Национальной академии наук Украины национального грид-центра. Еще шесть региональных грид-центров будут построены в 2010-2012 годах. Из 300 млн грн. общего финансирования программы.

В рамках начатого проекта главная цель Украины — включение UGrid в общую грид-инфраструктуру Европы и обеспечения постоянного функционирования ее как полноценной операционной и функциональной составляющей этой структуры. Украина получает возможность сотрудничать со странами Европейского Союза над созданием и использованием грид-технологий — для обеспечения обмена научными данными и организации их коллективного использования, а в ближайшие несколько лет преодолеть отставание от европейских стран и войти в Европейское исследовательское пространство (ERA) полноправным и квалифицированным партнером.

Проект выполняет команда из 10 различных украинских организаций (2-х академических, 6-ти образовательных и 2-х промышленных), которую возглавляет Институт Системного Анализа (ИПСА) Национального Технического Университета Украины “ Киевский политехнический Институт”. Научным руководителем проекта является Михаил Захарович Згуровский, ректор НТУУ” КПИ”, один из инициаторов внедрения Grid технологий в Украине.

Исполнителями проекта, кроме НТУУ”КПИ”, являются Институт проблем моделирования в энергетике имени Г.Е. Пухова НАНУ (ИПМЭ), Харьковский национальный университет радиоэлектроники (ХНУРЭ), Львовский национальный технический университет „Львовская политехника” (НУЛП), Запорожский национальный технический университет (ЗНТУ), Донецкий национальный технический университет (ДОННТУ), Днепропетровский национальный горный университет (ДНГУ), государственное предприятие ” Львовский научно-исследовательский радиотехнический институт”(ЛНИРИ) и предприятие ЮСТАР.

Самым слабым звеном украинской грид-инфраструктуры является даже не столько недостаток каналов связи или незначительные вычислительные мощности научных центров, а критическое отсутствие специалистов, способных поддерживать эти технологии.

Проект ИПСА НТУУ “КПИ” на 2011 год

В 2011 году по тематике проекта выполнено следующее:

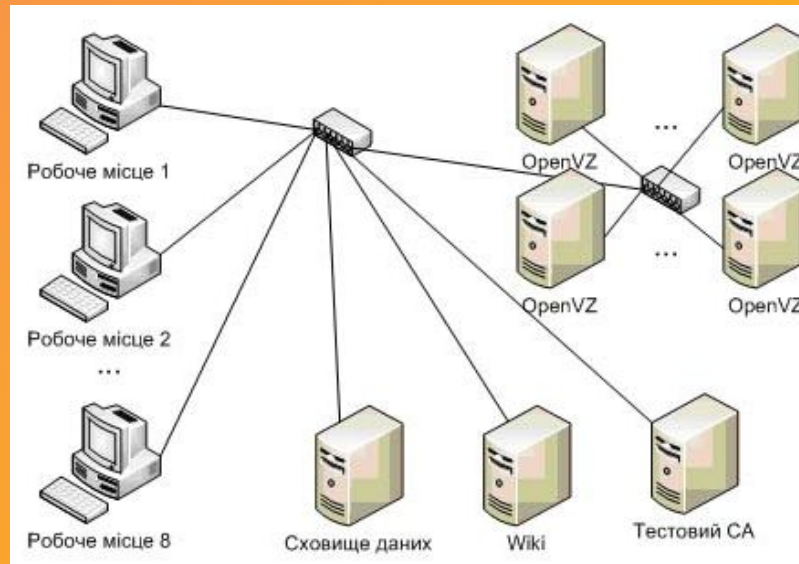
1. Предложена архитектура междисциплинарного комплекса оптимального математического моделирования GridALLTED, базирующаяся на грид-сервисах (разработанных с использованием SOAP, WSDL та UDDI).

2. Обоснована возможность создания автоматизированного метода построения схемных макромоделей компонентов, описываемых системами обычных линейных и нелинейных дифференциальных уравнений первого порядка.

3. Усовершенствован метод построения схемных макромоделей компонентов, которые описываются системами дифференциальных уравнений в частных производных, отличающийся от существующего учетом: конечности разрядной сетки компьютера; большой разницы в значениях элементов матриц C , G , L (до 30 порядков).

4. Доказана работоспособность, точность и эффективность предложенных методов по результатам моделирования в грид-среде многих типовых объектов.

Структура учебной грид-системы



Тестовый СА - для выдачи сертификатов пользователей и узлов, а также для подтверждения их подлинности;

Wiki - сервер с документацией;

Хранилище данных – для хранения репозитория пакетов, необходимых для установки сервисов ППС gLite, предоставление ресурсов хранения задачам грид.

Системы виртуализации OpenVZ – для работы виртуальных машин с сервисами ППС gLite.

Рабочие места - для доступа к ресурсам грид на базе ППС gLite.

Суперкомпьютерный вычислительный комплекс Института кибернетики им. В. М. Глушкова НАН Украины

Технические характеристики суперкомпьютера СКИТ

Количество кластеров: 4;

Количество узлов: 195;

Пиковая производительность кластера – 33,715 Тфлопс*.

Реальная производительность – 15,786 Тфлопс*.

Количество ядерных процессоров в узле - 2.

Оперативная память – 1,5 Тбайта.

Система хранения данных - обслуживает все кластеры, типа RAID5, глобальная файловая система общим объемом 160 Тбайт.

Общая потребляемая мощность - 103 кВт от сети 380 В.

Сеть передачи данных между узлами Infiniband FDR 56 Гбит / с.

флопс*- единица, используемая для измерения производительности [компьютеров](#) флопс*- единица, используемая для измерения производительности компьютеров, показывающая, сколько операций с [плавающей запятой](#) флопс*- единица, используемая для измерения производительности компьютеров, показывающая, сколько операций с плавающей запятой в [секунду](#) выполняет данная вычислительная система.

Комплекс СКИТ



World Community Grid

World Community Grid – это глобальное сообщество пользователей ПК, которые предоставляют неиспользуемое время своих вычислительных систем для реализации глобальных исследовательских инициатив (борьбе со СПИДом и онкологическими заболеваниями, моделированию и прогнозированию природных явлений и др.).

Глобальный гуманитарный проект World Community Grid запущен IBM в 2004 году при поддержке крупнейших исследовательских центров мира. ИТ-ресурсы World Community Grid позволяют проанализировать за один день такое количество данных, обработка которых на обычном компьютере заняла бы около 130 лет.

На сегодняшний день по всему миру в сети WCG насчитывается 604 494 пользователей, задействовавших 2 175 094 ПК. Общее время вычислений – 661 296 лет. Для участия в проектах достаточно зарегистрироваться [на сайте](#) и загрузить бесплатную программу.

В инфраструктуру World Community Grid могут быть подключены компьютеры, работающие на ОС Windows, Mac и Linux. Присоединившись к сети, пользователь получает на свой компьютер отдельное вычислительное задание (проект разделяется специалистами IBM на миллионы подзадач). Выполнив задание, ПК передает результаты вычислений на сервер и получает новое задание. Вычисления выполняются лишь тогда, когда компьютер не задействован. Результаты вычислений, отправленные на сервер, автоматически соединяются с результатами других заданий, и формируют общий результат.