

Определение понятия грид: фиксируем точку зрения экспертов

Вариант 1.0 от 26 июня 2006 года

Хайнц Штокингер
Швейцарский Институт биоинформатики (Vital-IT)
CH-1015 Лозанна, Швейцария

Defining the Grid: A Snapshot on the Current View

Draft 1.0, 26 June 2006

Heinz Stockinger
Swiss Institute of Bioinformatics (Vital-IT)
CH-1015 Lausanne, Switzerland
Heinz.Stockinger@isb-sib.ch

Основной материал от:

Greg Astfalk, Malcolm Atkinson, Miguel Bote-Lorenzo, Rajkumar Buyya, Lorenzo Cerutti, Walfredo Cirne, Brian Coghlan, Jose Cunha, Andrea Domenici, Flavia Donno, Dietmar Erwin, Laurent Falquet, Stephen Flinter, Ian Foster, Geoffrey Fox, Fabrizio Gagliardi, Wolfgang Gentzsch, Andrew Hanushevsky, Emir Imamagic, Fotis Karayannis, Daniel S. Katz, Dieter Kranzlmüller, Domenico Laforenza, Erwin Laure, Max Lemke, Rodrigo Fernandes de Mello, Miron Livny, Gabriel Mateescu, Rodrigo Mello, André Merzky, Reagan Moore, John Morrison, Maria S. Perez, Ron Perrot, Jean-Marc Pierson, Thierry Priol, Jean Salzemann, Dave Snelling, Michela Taufer, Domenico Talia, Sathish Vadhiyar, Frank van Lingen, Gregor von Laszewsk

Резюме

Термин "грид" был введен в начале 1998 года публикацией книги "Грид. Новая инфраструктура компьютеринга". С того времени произошли большие технологические изменения, как в аппаратных средствах, так и в программном обеспечении. Самым важным изменением является, по-видимому, широкое употребление Web-служб. Хотя в последнем десятилетии базовая идея грида не претерпела существенных изменений, у многих людей различные представления о том, чем является конкретная реализация грида на самом деле. В настоящей статье мы представляем обзор мнений большого числа специалистов в области грид-компьютинга, которых мы попросили высказаться об их понимании этого вопроса.

"Вычислительные гриды эквивалентны гридам электроснабжения" [1, 12]

"Web-службы дают возможность цвести тысяче цветов.

Грид дает нам возможность организовать посев и выращивание растений так, чтобы повысить качество сбора урожая." [MA]

1. Введение

Идея грид-компьютинга начала витать в воздухе значительно раньше появления книги [1], изданной Яном Фостером (Ian Foster) и Карлом Кессельманом (Carl Kesselman). Однако, выпуск этой книги обозначил начало новой эры в компьютеринге и возникновение нового поля исследования: грид-компьютинга. Первоначальные понятия и определения базировались на сравнении вычислительного грида с гридом электроснабжения [1]. И действительно, это слово даже содержится в названиях таких европейских компаний как Австрийский грид электроснабжения, Швейцарский грид (Austrian Power Grid, swissgrid) и т.д., которые являются названиями национальных гридов. Это первоначальное представление Ян Фостер расширил за счет следующего контрольного списка атрибутов [2], получившего общее признание:

- 1) *координирует ресурсы, не подчиняющиеся централизованному контролю...*
- 2) *...используя стандартные, открытые протоколы и интерфейсы общего назначения ...*
- 3) *...чтобы предоставить возможность работы с нетривиальными свойствами служб.*

Совсем недавно Ян Фостер и Стив Тьюк (Steve Tuecke) дали четкое описание того, что они подразумевают под понятиями грид и *ориентированная на службы архитектура (service-oriented architecture)* [3]. Там же они дают четкие определения *компьютинга услуг (utility computing)* и *компьютинга по требованию (on-demand computing)* и их отличия от грид-компьютинга. Определения грида других авторов можно найти в [8, 9, 10].

Вообще-то, в компьютерной науке и инженерии программного обеспечения определения не всегда являются такими строгими, как определения в физике и математике. Как результат этого "дефицита определений" у многих занимающихся гридом ученых и работающих с грид-технологиями специалистов сложились различные точки зрения на то, что такое грид. Самые часто встречающиеся расхождения касаются определений аппаратных средств. (Для кого-то локальный кластер под управлением системы промежуточного программного обеспечения является гридом, в то время, как другие полагают, что к этому необходимо добавить наличие пространственно-распределенной сетевой связи.) Другие главные расхождения относятся к программному обеспечению: что на самом деле превращает пакет программного обеспечения в "грид-обеспечение"? Становится ли грид-обеспечением некое промежуточное программное обеспечение, если оно использует присущие гриду методы обеспечения безопасности и т.д.? Большинство из нас думали об аналогичных вопросах (и обсуждали их), но так и не находили убедительного ответа.

Благодаря последним достижениям в технологиях Web- и грид-служб часто бывает непонятно, где провести разграничительную линию между Web-службами и грид-службами [4]. И особенно нас интересует текущая точка зрения пользователей на то, из чего состоит грид. Поэтому в начале 2006 года мы провели исследование этого вопроса, для чего пригласили специалистов и попросили их высказаться на эту тему. В настоящей статье мы рассказываем о мнениях многих людей из всемирного грид-сообщества, пытаюсь поместить в центр внимания основные особенности грида. Представление о текущей точке зрения дает возможность сформировать мнение о том, как исследователи

грида представляют себе понятия грида и как эти понятия могут быть использованы в других областях науки.

2. Подготовка исследования

Весной 2006 года мы начали исследование, в ходе которого мы связались с более чем 170 специалистами по гриду во всем мире и попросили изложить нам их точку зрения на определение понятия грид. Главная предпосылка нашего исследования: мы не должны оказывать какого-либо влияния на содержание ответов, то есть мы отказались от метода задавания вопросов или определений, с которыми нужно было соглашаться или не соглашаться. Люди, с которыми мы связывались, должны были иметь максимум свободы при формулировке своих определений. Главным лозунгом было следующее:

Попытайтесь определить важнейшие аспекты, которым нужно следовать при построении конкретного грида, что является безусловным, каковы границы с распределенным компьютерингом, Интернет-компьютерингом и т.д.

Кроме этого, мы попросили дать сжатые ответы максимум на одной странице. На это обращение нам ответили более 40 человек, в этой статье вы найдете тщательно обработанное резюме этих ответов. Мы сознаем, что предоставленная нами свобода связана с трудностями суммирования всех высказанных мнений. Впрочем, это отражает действительность, так как у большинства специалистов различные точки зрения по этому вопросу. Располагая пулом ответов, мы их классифицируем в соответствии с небольшим числом категорий, являющихся характерными для вычислительных гридов.

3. Результаты исследования

Одним из главных интересов исследования было выяснить, есть ли у людей более или менее общее понимание грида, или насколько велико расхождение во мнениях. Результат, безусловно, не беспристрастен в том смысле, что мы спрашивали главным образом тех специалистов, которые сейчас в этой области работают активно. Но это имеет и свою положительную сторону, так как мы получили сконцентрированное мнение профессионалов, а не то, что думает рядовой гражданин.

К тому же не существует безусловного способа оценки полученных ответов. Мы воспользовались методом, при котором на первом проходе фиксируются все главные ключевые слова, используемые при описании грида. Не удивительно, что для описания общего представления и главных характеристик было использовано много похожих слов и фраз. Поэтому на втором проходе был использован классификационный метод категоризации всех ответов в соответствии с:

- 1) общим представлением о гриде, т.е. на каких базовых идеях основывается грид-компьютеринг и какие основные цели нужно реализовать;
- 2) отличием по отношению к другим разделам компьютеринга, таким, как распределенный компьютеринг, Интернет- и Web-компьютеринг;
- 3) характеристиками грида, т.е. что "делает" систему гридом? Иногда для описания понятия удобнее привести список характеристик и особенностей.

В следующих подразделах мы описываем результаты исследования на основе полученных нами ответов. Часто мы непосредственно цитируем опрошиваемых,

сопровождая цитаты их инициалами. Подробности об этом представленном инициалами человеке можно найти в разделе 5. Иногда люди соглашаются с некоторыми определениями форума GGF [6] или проекта CoreGRID [11]. В этом случае прямое цитирование не используется.

Иногда в ответах приводятся более подробные рассмотрения, кто-то описывает отдельные части грида, описывают характеристики и какие имеются отличия в отношении традиционных подходов. Здесь много пересечений и почти никаких противоречащих высказываний. Мы считаем это главным посылом статьи: результаты опроса грид-сообщества хорошо согласуются с тем, чем является грид на самом деле.

3.1 Концептуальное описание грида

3.1.1 Обзор

Общее представление о гриде, приведенное в [1], не изменилось, но появилось несколько дополнительных уточнений аналогичных тем, которые приведены ниже. Например, "в общем представлении о гриде существует различие между: (а) **грид-подходом** (парадигмой), выражающим общее понятие и идею грида для расширения представления о нем у высокоразвитых международных научных и ориентированных на бизнес коллабораций и (b) **физической реализацией** производственного грида (**production grid**) на базе доступных ресурсов для расширения представления о гриде у высокоразвитых международных научных и ориентированных на бизнес коллабораций" [GvL].

Инфраструктура грида должна предоставлять набор технических возможностей, таких, как [3]:"

- **Моделирование ресурсов.** Описывает имеющиеся ресурсы, их возможности, взаимоотношение между ними в целях более эффективного их обнаружения, настройки, а также повышения качества управления обслуживанием.
- **Мониторинг и уведомление.** Предоставляет возможности просмотра состояния ресурсов и уведомляет приложение и службы управления инфраструктурой об изменении состояния. Протоколирование важных событий (**logging**) и переходов ресурсов в новое состояние также необходимо для поддержания функций учета и аудита.
- **Аллокация.** Обеспечивает гарантированное качество обслуживания для всего множества ресурсов на весь период их использования приложением. Это достигается в процессе переговоров о требуемом уровне обслуживания и гарантированном обеспечении наличия подходящих ресурсов посредством некоторой формы резервирования - в сущности, в виде динамического создания соглашения об уровне обслуживания.
- **Настройка, управление жизненным циклом и прекращение действия.** Дает возможность автоматического конфигурирования выделенного ресурса для прикладного использования, управляет ресурсом все время решения текущей задачи, приводит ресурс в первоначальное состояние для дальнейшего использования.
- **Учет и аудит.** Прослеживает использование разделяемых ресурсов и предоставляет механизмы пересылки тарифов членам коллектива пользователей и выставляет счета за использование ресурсов приложениями и пользователями".

- В добавление к этому, безопасность как очень важный аспект [GA].

"Мы можем рассматривать грид как комбинацию **распределенных, совместно используемых систем большой пропускной способности**, для эффективного **разделения и распределенной координации ресурсов**, принадлежащих **различным административным доменам [MP]**". Вообще, каждый грид предоставляет "инфраструктуру для распределения компьютерной мощности. Предполагается, что она предоставит исследователям (пользователям) **единую точку входа** для запуска заданий" [LF]. Проще говоря, грид означает "распределенный компьютеринг на множестве административных доменов" [DS]. Иногда грид также называют "средой программного обеспечения" [GA], которая **интегрирует, виртуализирует** распределенные ресурсы (программное обеспечение и аппаратные средства) и **управляет** ими. С другой точки зрения, каждый грид это "**система управления ресурсами очень большого объема**" [AD].

Важно отметить, что любой грид может быть построен с помощью различных технологий. Это, как правило, означает, что не существует такой вещи как "типовая грид-технология". "Web-службы это всего лишь один механизм (из обширного списка возможных механизмов), который можно использовать для построения планируемого грида" [AH]. Это еще раз подчеркивается в следующем утверждении: "Очень важно не упускать из вида, что **грид-службы это 'текущий подход'**, так как для построения грид-инфраструктуры вполне возможно обратиться к другим технологиям, не связанным со службами (например, компоненты программного обеспечения)" [MB]. "Следовательно, множественность технологий желательна и их совместное использование может оказаться необходимым в гетерогенном гриде" [JM]. Отсюда следует, что другие технологии, не основывающиеся на Web-службах, могут и будут применяться при построении гридов.

Часть специалистов рассматривает грид "скорее как **концепцию** или **движение**, а не как систему" [ML], которая объединяет людей. Характер этого стимулирующего фактора "скорее **социологический** или **культурный**, чем технический" [ML]. В том же направлении звучит следующее мнение "Я считаю, что следует полностью развязать определение грида, которое скорее является необходимо определяемым с точки зрения пользователя принципом, от технической реализации архитектуры, протоколов, служб и технологии. Следует согласиться, что определение грида скорее могло бы быть определением набора свойств. Если перед нами некая не идентифицированная система, в которой наличествуют все эти свойства, то про эту систему можно сказать, что она грид" [JS].

В заключение мы также приведем уже общепринятые определения от GGF и the CoreGRID Network of Excellence, также прозвучавшие в нашем исследовании.

GGF [6]:

"Система, связанная с функциями интеграции, виртуализации и управления службами и ресурсами в распределенной, гетерогенной среде, которая поддерживает совокупность пользователей и ресурсов (виртуальные организации) на совокупности традиционных административных и организационных доменов (фактических организаций)".

CoreGRID (предоставил [TP] от имени исполнительного комитета CoreGRID):

"Полностью распределенная, динамически реконфигурируемая, масштабируемая и автономная инфраструктура, обеспечивающая независимый от места хранения, проникающий, надежный, безопасный и эффективный доступ к открытому для поиска множеству служб, обладающих свойствами инкапсуляции и виртуализации ресурсов (вычислительных средств, хранилищ, инструментов, данных и т.д.) в целях генерации знаний".

3.1.2 Классификация

Часто пытаются классифицировать различные "типы гридов" в соответствии с их главными функциональными возможностями, но с такой классификацией соглашаются не всегда. Впрочем, наша задача попытаться ознакомить читателей с основными идеями. В принципе, большинство различают чистые **вычислительные гриды** и более развитые **гриды данных**. Впрочем, существуют также и другие классификации такие, как [DS]:

- 1) **"Коллаборативные гриды:** Эти гриды обслуживают много организаций (институций) и отдельных пользователей, защищенные домены; используют различные протоколы, механизмы поиска и т.д.". Важные аспекты:
 - "Широко распределенные, виртуальные организации (VO).
 - Соглашения об уровне обслуживания и коммерческие партнерства.
 - Бизнес-модель: увеличение конечного дохода.
- 2) **Корпоративные гриды:** В своем большинстве эти гриды технически также сложны, как и гриды из пункта 1). Они реализуют полный жизненный цикл развертывания, обеспечения, управления и прекращения действия служб точно так же, как и коллаборативные гриды. Однако множественные домены либо отсутствуют, либо сильно интегрированы по крайней мере на уровне политик. Это производственные гриды главных центров данных. Важные аспекты.
 - Виртуализация корпоративных ресурсов и приложений.
 - Объединение и централизация управления.
 - Бизнес-модель: уменьшить общую стоимость эксплуатации.

"Для корпорации безопасность и аудит имеют даже еще большую важность" [GA].

- 3) **Кластерные гриды:** Предназначенные для компьютеринга высокой производительности/пропускной способности, эти гриды являются большей частью средами планирования рабочей нагрузки. В отличие от предыдущих, они скорее статичны, чем динамичны. Службы кластерных гридов по своей природе являются либо службами общего назначения, например, служба загрузки заданий, либо оказывают все время одну и ту же услугу. Как правило, эти гриды не поддерживают полный цикл жизни службы". [DS]. Впрочем, сами кластеры (если они не связаны с другими кластерами) обычно не называют гридами.

Другой способ категоризации гридов это использовать их "географическое распределение, организационную форму и права собственности на ресурсы".[GM] Таким образом, мы можем различать **кластерные гриды**, **"университетские" гриды**, **корпоративные гриды** и **глобальные гриды**. "Кластерный грид (называемый также гридом подразделения) содержит ресурсы, сосредоточенные на одном сайте в рамках одной организации, он принадлежит одному собственнику. Университетский грид отличается от кластерного тем, что его ресурсы принадлежат нескольким собственникам. В отличие от университетских гридов, корпоративные гриды содержат ресурсы,

находящиеся на многих сайтах. И, наконец, глобальные гриды содержат ресурсы от многих организаций" [GM]. Коллаборативные гриды иногда называют также "Защищенными гридами (Beyond Firewall Grids)" [DT]. Еще один способ именования различных гридов: "Интра-грид, Экстра-грид и Интер-грид" [DT].

Иногда кластеры и гриды используются в одном и том же контексте как синонимы, но большинство опрошенных специалистов четко разграничивают эти понятия: "Ключевым отличием кластеров от гридов является в первую очередь то, как производится управление ресурсами. В случае кластеров распределение ресурсов осуществляется централизованным менеджером ресурсов и часто много узлов кооперативно работает вместе как единый унифицированный ресурс. В случае гридов у каждого узла свой собственный менеджер ресурсов, и узел не стремится обеспечить возможность своей интеграции в одну систему" [RB]. "Узлы в гридах 'автономны', тогда как узлы в кластерах автономными не являются. Отсюда гриды используют децентрализованное управление ресурсами, а кластеры централизованы и поэтому выглядят как одна система" [RB]. Другое отличие в том, что "грид образован различными административными доменами, чьи ресурсы управляются динамическими виртуальными организациями" [MP].

3.1.3 Аппаратные средства или программное обеспечение

Физическая реализация конкретного грида зависит от аппаратных средств и компонент программного обеспечения. Тогда как в отношении аппаратных средств никаких конкретных особенностей не зафиксировано (иногда важной составляющей грида считают широкомасштабные региональные вычислительные сетевые связи [4]), этого нельзя сказать про программное обеспечение. Например, "ключевым отличием грида от остального распределенного компьютеринга является использование **промежуточного программного обеспечения грида**" [DK]. Впрочем, по вопросу определения промежуточного программного обеспечения не всегда имеется единодушное мнение, и поэтому многие предлагают "избегать определений, связанных с ресурсами и с уровнем промежуточного программного обеспечения" [AM].

3.1.4 Базовые службы

По крайней мере, с появлением открытой архитектуры грид-служб (Open Grid Service Architecture, OGSA) стало ясно, что с помощью или посредством грида можно воспользоваться практически любой традиционной "службой" (в смысле ориентированной на службы архитектуры). Самыми фундаментальными службами являются следующие: выбор ресурсов, планирование, безопасное выполнение, управление данными, целостность данных и приватность, аутентификация и восстановление после сбоев.

3.2 Сходства и различия с другими областями компьютеринга

В конце 1990 годов грид-компьютеринг возник как новый раздел компьютерной науки, хотя стандартные методы и протоколы были взяты из смежных областей, таких, как Интернет, распределенный компьютеринг или "сообщество" баз данных. Впрочем, грид-компьютеринг нельзя обсуждать в отрыве от "традиционных" областей, с которыми он имеет много пересечений. Тем не менее, значительная часть опрашиваемых нами специалистов проводит строгое различие между грид-компьютерингом, распределенным

компьютингом и интернет-компьютингом, что является, по-видимому, наиболее противоречивой частью нашего исследования. Рассмотрим это чуть подробнее.

3.2.1 Грид или распределенный компьютер

Некоторые рассматривают грид как "общую" [DT], некоторые как "специальную" форму распределенного компьютеринга [4], в то время как другие думают, что отличительной чертой грида является его сложность, проявляющаяся различно и характеризующаяся масштабируемостью и прозрачностью:

- **Масштабируемость:**
 - "Границы с распределенным компьютерингом могут быть определены теми точками, где двусторонний контекст начинает заменяться N-сторонним контекстом. Под контекстом мы понимаем, прежде всего, безопасность, архитектурные модели и модели программирования" [BC].
 - Число вовлеченных организаций: "главными отличиями являются (потенциальные) межорганизационные характеристики и слабая взаимозависимость между действующими партнерами (либо службами, либо институциями)." [JP].
- **Прозрачность:** кроме этого, грид должен быть "платформенным агностиком" [SF] и уметь распоряжаться гетерогенными ресурсами (как аппаратными средствами, так и программным обеспечением).

Впрочем, мы располагаем мнениями противоположного характера, такими, что не существует "границы" между распределенным и грид-компьютингом, и что они скорее "дополняют друг друга и являются частью друг друга [MT]".

Подводя итог, самыми важными характеристиками, которые могут выявить "искomое" различие между гридами и распределенным компьютерингом являются различия, приведенные в разделе 3.3.

3.2.2 Грид или Интернет (Web)

Грид-сообщество приняло и подвергло дальнейшему развитию много технологий Интернета и Web-служб. Например, грид-служба это Web-служба с дополнительными возможностями. Впрочем, нет общего согласия в том, где проходит граница (если она вообще имеется) между Интернет и грид-компьютингом. Некоторые рассуждают так: "если в Интернете (Web'e) обмен сообщениями происходит между двумя точками, то грид обеспечивает более высокий уровень абстракции" [DKr]. Более того, представление об этом идет даже дальше и расширяет понятие Интернета до грида. "Аналогично построению нынешней Всемирной паутины как глобальной информационной платформы, мы занимаемся построением **Всемирного грида** с целью создания глобальной платформы для сотрудничества, объединяя компьютеры и память, приложения и данные, эксперименты, инструментарий, сенсоры и другие цифровые устройства" [WG].

Используя метафору, приведенную нами на первой странице, [MA] проводит следующее различие: "Я считаю **web-службы** подсистемой компьютеринга, **независимо разрабатываемой и независимо развертываемой на гетерогенных платформах**. Они являются независимо управляемыми службами, они могут состоять из других служб, например, при взаимодействии в процессе выполнения. **Для web-служб не обязательно должна существовать некая априорная структура и согласованность реализации** для облегчения создания такой композиции при удовлетворении стандартов, аналогичных

стандартам WSI. Нет никаких априорных договоренностей, разрешающих распределенное WS-управление. В случае грида имеющиеся службы разрабатываются и устанавливаются независимо друг от друга на различных гетерогенных платформах. Впрочем, разработчики грида предпочитают отказаться от **независимости** в определенной мере; вместо этого от служб требуется согласование с общепринятыми стандартами более высокого уровня, которые реализуют виртуальную гомогенность. Эта предпочитаемая согласованность нужна для того, чтобы легче устанавливать программное обеспечение и службы на всем гриде и легче реализовывать совместную работу служб, предлагаемых гридом". Как правило, грид обеспечивает коммуникационный слой, позволяющий службам общаться друг с другом, что дает возможность приводить аргументацию, аналогичную вышеприведенной: "...отделяет гриды от Web'a, который является (большим) множеством независимых служб" [AD].

Утверждение, с которым соглашаются легче, это то, что грид можно рассматривать как продолжение и развитие Интернета. "Следовательно, к миру грида могут быть применимы некоторые базовые правила - интеграция гетерогенных ресурсов может быть достигнута при использовании стандартизованных протоколов и служб. Протоколы Интернета предоставляют хорошую базу для связывания ресурсов. Однако, необходим более широкий набор стандартов для того, чтобы реализовывать более сложные функциональности, такие, как выполнение заданий, управление данными, операции по обеспечению безопасности и т.д." [EI]. Репрезентативный аргумент, подчеркивающий, что грид расширяет Интернет: "Службы Интернета не являются отдельной областью для исследований, а частью исследований грида" [MT].

3.2.3 Грид, кластеры и системы P2P

В области грида все больше и больше используется технология P2P. Приводим следующий набор характеристик, который поможет нам отличать друг от друга гриды, кластеры и системы P2P [RB]:

Характеристика	Кластер	Грид	P2P
популяция	commodity PC компьютеры	профессиональные компьютеры	сетевые концы (настольные PC)
собственники	один	многие	многие
обнаружение	службы членства	централизованный индекс и децентрализованная информация	децентрализовано
Пользовательское управление	централизовано	децентрализовано	децентрализовано
Управление ресурсами	централизованное	распределенное	распределенное
распределение/ планирование	централизованное	децентрализовано	децентрализовано
Интероперабельность	на базе VIA	более развитое (например, WSRF)	стандартов нет
образ одной системы	да	нет	нет

масштабируемость	100	1000	миллионы
возможности	гарантированы	переменны, но высоки	переменны
пропускная способность	средняя	высокая	очень высокая
скорость (латентность, полоса пропускания)	низкая, высокая	высокая, низкая	высокая, низкая

3.3 Характеристики грида

Обычно гриды описывают набором характеристик. Доминантными характеристиками, с которыми, как правило, соглашаются все, являются следующие:

- Сотрудничество
- Объединение
- Виртуализация
- Ориентированность на службы
- Гетерогенность
- Децентрализованное управление
- Стандартизация и интероперабельность
- Прозрачность доступа
- Масштабируемость
- Возможности реконфигурирования
- Безопасность

В дополнение к этому мы можем добавить еще несколько важных тем и аспектов:

- Поддержка приложения
- Модель компьютеринга
- Модели лицензирования
- Процедуры и политики
- Возможности аудита

Сотрудничество

Аспект грида, с которым все соглашаются, это разделение ресурсов методами распределенного компьютеринга. Более того, грид "перекрывает **множество административных доменов** бесшовно" [AM]. Разделение ресурсов доходит даже до уровня, определяемого как "коллаборативные гриды" [GF]. Более того, очень важно, что коллаборация обеспечивает положительную синергетику среди пользователей и провайдеров служб. "При корректной и продуманной реализации мы получим неожиданные синергетические преимущества, которые иначе были бы для нас недоступны" [JM]. И, наконец, ресурсы должны быть "разделены справедливым образом" [FvL].

Объединение

Каждый грид больше суммы всех своих частей: "Грид объединяет в одно целое много ресурсов и поэтому дает нам объединение **мощностей индивидуальных ресурсов** в виртуальный ресурс с мощностями более высокого уровня. Возможности индивидуальных ресурсов сохраняются. Как следствие, с глобальной точки зрения, грид позволяет выполнять большие приложения быстрее (возможности объединения), а с локальной точки зрения грид позволяет запускать новые приложения" [GM]. Объединение используется также для "увеличения производительности, повышения качества обслуживания, лучшей утилизации и более удобного доступа к данным" [FD]. Наконец, ресурсы могут (или должны) добавляться динамически или статически [DE].

Виртуализация

Грид-службы часто предоставляются с определенным интерфейсом, который скрывает сложность нижележащих ресурсов. По-другому это называется виртуализацией, предоставляющей абстрактный "слой" между клиентом и ресурсами [GA]. Следовательно, грид дает "возможность виртуализировать сумму частей в своеобразную широкомасштабную модель программирования" [BC]. Виртуализация охватывает как данные (плоские файлы, базы данных и т.д.), так и компьютерные ресурсы [WG]. Список подлежащих виртуализации ресурсов может быть достаточно большим, как, например, следующий [RM]:

- Грид как виртуализация рабочего потока – использование вычислительных грид-служб для выполнения процессов и управления ими на поле многих платформ компьютеринга.
- Гриды данных как виртуализация данных - управление разделяемыми наборами данных независимо от удаленной системы хранения, на которой находятся данные.
- Семантический грид, как виртуализация информации - возможность делать логические выводы на основе обработанных атрибутов из нескольких независимых информационных репозитариев.

"В основе виртуализации лежит способность управлять правилами именования, информацией о состоянии, методами доступа и удаленными операциями независимо от удаленного ресурса. Среда грида требует" [RM]:

- Виртуализации пространства имен, таких логических имен для ресурсов, пользователей, файлов и метаданных, которые не зависят от пространств имен, используемых на удаленном ресурсе.
- Виртуализация доверительности, способность управлять аутентификацией и авторизацией независимо от того, что данный ресурс является удаленным.
- Виртуализация ограничений, способность управления доступом независимо от того, что данный ресурс является удаленным
- Виртуализация доступа, способность портировать механизм произвольного доступа наверх промежуточного программного обеспечения грида. Для гридов данных это является способностью поддерживать доступ к множеству загружаемых библиотек (Windows, Perl, Python, C), Java, цифровых библиотек

(DSpace, Fedora, OAI-PMH), акторам (actor)¹ рабочих потоков (Kepler), Web-браузерам и т.д.

- Виртуализация сети, способность управлять функциями транспорта при наличии сетевых устройств, типа сетевой защиты, выравнивателей загрузки, частных виртуальных сетей. Это, как правило, требует множественности протоколов для поддержки инициированных клиентом, а не сервером операций ввода/вывода, групповых, а не однофайловых операций;

- Управление латентностью, способность минимизировать количество сообщений, пересылаемых через широкомасштабные региональные вычислительные сети. Сюда входит выполнение процедур на удаленном ресурсе, когда сложность (отношение выполняемых операций к передаваемым байтам) достаточно мала. Стандартный случай это фильтрация данных или пересылка части файлов (**sub-setting**);

- Объединение, способность интероперабельности на множественных грид-средах. Это требует умения совместно использовать пространство логических имен и аутентифицировать в стиле Shibboleth². Грид требует наличия механизмов доверительности, позволяющих подтвердить аутентичность индивидуума из своего "домашнего" грида.

Ориентация на службы

Грид предоставляет службы, следуя принципам ориентированной на службы архитектуры. В самом широком смысле "все **широкомасштабные наборы служб** можно рассматривать как гриды" [GF].

Гетерогенность

Грид, как типовой объект, состоит из "гетерогенных вычислительных ресурсов" [RM], т.е. имеется многообразие различных компонент аппаратных устройств и программного обеспечения с различными характеристиками латентности и эффективности.

Децентрализованное управление

Мы уже познакомились с этими характеристиками в трех пунктах контрольного списка Яна Фостера, но здесь снова говорим об этом, так как в ответах нашего исследования он упоминался несколько раз. Другими словами, "компоненты находятся под **управлением многих сущностей**. Ключевая трудность в гридах заключается именно в том, что нет единого 'собственника' всей системы" [WC], т.е. ресурсы находятся "в собственности разных владельцев" [JM]. "Одним из требований грида является использование **механизмов распределенного управления**" [MP].

Стандартизация и интероперабельность

Грид "борется за **определение стандартных интерфейсов для служб**, которым необходимо **взаимодействовать** друг с другом, чтобы создать общую распределенную

¹ См. Глоссарий Web-служб на сайте <http://www.gridclub.ru>

² Проект комитета Middleware Architecture Committee for Education, под эгидой которого создается Internet2. Shibboleth представляет собой полную систему аутентификации и контроля доступа для ресурсов в Web. (примечание пер.)

инфраструктуру, обеспечивающую выполнение задач пользователя, и предоставление утилит пользовательского уровня [FD]. "Грид-системы, реализующие некоторый стандарт, должны взаимодействовать с гридами, которые придерживаются того же самого стандарта" [DE].

"Совершенно очевидно, что гриду необходим повышенный уровень **интеграции отличающихся технологий** и повышенные требования к соглашениям по **стандартизации служб**. Успех в построении грида очень сильно зависит от этих аспектов" [JS]. Более того, грид должен обеспечить унифицированный доступ к гетерогенным ресурсам на основе виртуализации [GM].

Вот еще более строгое высказывание о стандартах и интероперабельности: "Любой грид, не базирующийся на стандартах, расточителен. Если Вы сосредоточитесь на грид-службах, которые не являются интероперабельными, то увидите, что сам принцип просто не работает и ничего из обещанного предоставляться не будет. Первое правило в гридах или для нас в HP это 'безжалостная стандартизация'" [GA].

Прозрачность доступа

Грид "должен разрешать **обращаться к вычислительной инфраструктуре пользователям**, не обладающим глубокими знаниями **базовой архитектуры** или сетевой топологией" [SF]. Это иногда считают "*самым отличительным аспектом* грид-компьютинга, иначе говоря, это уровень **предоставляемой конечному пользователю прозрачности** на пути виртуализации ресурсов" [JC].

Масштабируемость

Если даже реализации и инфраструктуры гридов иногда не решают "новую проблему", часто сам объем данных, ресурсов и пользователей дополнительно влияет на сложность грида. Это также выражается по-другому, Грид должен быть "нетривиальным в смысле тех задач, которые пользователь не мог решить прежде" [SV].

Реконфигурируемость

Грид должен быть "динамически реконфигурируемым" так, как это специфицировано в определении от CoreGRID.

Безопасность

Безопасный доступ к ресурсам - существенная черта грида. Поэтому "у **авторизованных пользователей** и приложений ограниченный набор операций (даже вообще никакой)" [JM], которые они могут выполнять с помощью служб. В сущности, безопасность грида это один из первостепенных моментов, с которыми реальные пользователи грида имеют дело, и поэтому она неотъемлема для любой системы программного обеспечения грида, которая охватывает много административных доменов.

Поддержка приложений

Вообще говоря, грид мог бы поддерживать большое разнообразие различных приложений. "**Приложения должны также считаться частью грида** и вся среда грида

(где под средой я понимаю аппаратные средства, промежуточное программное обеспечение и приложения) **должна быть управляема данными**. Это в особенности касается ее возможности **реагировать на происходящие в системе изменения** и на поведение приложения, фиксируемое приложением и системными данными" [MT].

Модель компьтинга

Как правило, грид поддерживает "различные вычислительные модели (например, пакетную, интерактивную, распределенную и параллельный компьтинг...)" [AD].

Модель лицензирования

Так как гриды зародились в академическом сообществе, наличествует "общая тенденция подчеркивать производство **программного обеспечения открытого кода**" [FK], этого также придерживаются и различные компании, занимающиеся разработкой грида.

Поведение и политики

Пользователи грида и поставщики служб взаимодействуют друг с другом почти так же как на открытом рынке, где требуется соблюдать определенные правила поведения. Поэтому "поведение и политики" [FG] должны присутствовать при обеспечении (координированного) разделения ресурсов.

3.4 Обсуждение

Хотя грид-сообщество полно энтузиазма, не все считают, что высокие цели, декларируемые в академических определениях грида, получают на сегодняшний день удовлетворительную грид-реализацию. В какой-то мере это с очевидностью следует из полученных нами ответов. В следующем разделе отражается положение дел в этом вопросе на настоящий момент и как к этому относится IT-сообщество.

Положение дел и тенденция

"Самой большой опасностью для грид-компьтинга является то, что его можно было бы **рассматривать как сегодняшнюю секс-технологию**, которая будет быстро заменена завтрашней секс-технологией" [SF]. Пользователям грида и грид-технологам пора начинать указывать на использующие грид приложения (полученные с их помощью результаты) и на новые, невыполнимые без грида приложения. Аналогичное мнение звучит так: "Современные грид-реализации все еще весьма далеки от того, какими они вырисовывались вначале, и далеки от широкого распространения" [EI].

Грид и IT-сообщество в широком смысле

В последнем обзоре рынка мы анализировали главным образом рынок европейского IT и нас интересовало то, как грид-технологии применяются или могут быть применены к бизнесу и/или к среде коммерческого IT [6]. Главный вывод состоит в том, что много компаний используют технологии распределенного компьтинга, но всё еще не готовы принять модель грид-компьтинга. Отсюда встает вопрос, почему грид еще не широко распространился в коммерческом мире? Наш обзор ставит новый вопрос:

"Есть ли нечто, что мейнстрим корпоративного ИТ может с пользой для себя извлечь из грида, или это прибежище для 'яйцеголовых', проводящих ядерное моделирование, протеиновые эксперименты или бог знает что еще? Как сможет воспользоваться банковский ИТ-менеджер или служащий страховой компании грид-технологиями для решения своих коммерческих или технических проблем?" [SF].

4 Заключение

Представленное исследование является одной из первых попыток получить всеобъемлющее впечатление от того, что сообщество грид-пользователей думает об определении понятия грид. Предварительно мы проинтервьюировали ряд компаний об их взгляде на использование грида в среде бизнеса [7] и обнаружили, что мнения ИТ-лидеров на производстве в отношении грид-компьютинга достаточно разнообразны. Поэтому было очень интересно узнать, что само грид-сообщество думает по этому поводу. Один из интересных результатов - внутри сообщества занимающихся гридом едва ли имеется какое-то значительное расхождение во взглядах. Можно было бы поспорить о том, что проведенное исследование не дает много новой информации, если пользователи грида соглашаются относительно всех главных пунктов. Но главное то, что есть общее понимание о гриде (представление о нем), хотя в последние годы и произошло много технологических изменений и продвижений вперед.

5 Соавторы

InIT	Name	Organization	Country
AD	Andrea Domenici	University of Pisa	Italy
AH	Andrew Hanushevsky	Stanford Linear Accelerator Center	USA
AM	André Merzky	University of Amsterdam	The Netherlands
BC	Brian Coghlan	Trinity College Dublin	Ireland
DE	Dietmar Erwin	Research Centre Jülich	Germany
DK	Dan Katz	Louisiana State University / JPL	USA
DKr	Dieter Kranzlmüller	University of Linz / CERN	Austria / Switzerland
DL	Domenico Laforenza	CNR Pisa	Italy
DS	Dave Snelling	Fujitsu	UK
DT	Domenico Talia	University of Calabria	Italy
EI	Emir Imamagic	University of Zagreb	Croatia
EL	Erwin Laure	CERN	Switzerland
FD	Flavia Donno	CERN	Switzerland
FG	Fabrizio Gagliardi	Microsoft	Switzerland
FK	Fotis Karayannis	GRNet	Greece
FvL	Frank van Lingen	California Institute of Technology	USA
GA	Greg Astfalk	HP	USA
GF	Geoffrey Fox	Indiana University	USA
GM	Gabriel Mateescu	National Research Council Canada	Canada
GvL	Gregor von Laszewski	Argonne National Lab	USA

InIT	Name	Organization	Country
IF	Ian Foster	Argonne National Lab / U. Chicago	USA
JC	Jose Cunha	University of Lisbon	Portugal
JM	John Morrison	University College Cork	Ireland
JP	Jean-Marc Pierson	INSA Lyon	France
JS	Jean Salzemann	CNRS Clermont-Ferrant	France
LC	Lorenzo Cerutti	Swiss Institute of Bioinformatics	Switzerland
LF	Laurent Falquet	Swiss Institute of Bioinformatics	Switzerland
MA	Malcolm Atkinson	National e-Science Centre	UK
MB	Miguel Bote-Lorenzo	University of Valladolid	Spain
ML	Max Lemke	European Commission	Belgium
ML	Miron Livny	University of Wisconsin	USA
MP	Maria S. Perez	Technical University of Madrid	Spain
MT	Michela Taufer	University of Texas, al Paso	USA
RB	Rajkumar Buyya	University of Melbourne	Australia
RM	Reagan Moore	San Diego Supercomputing Center	USA
RM	Rodrigo Ferandes de Mello	University of São Paulo	Brasil
RP	Ron Perrot	Queen's University Belfast	UK
SF	Stephen Flinter	Science Foundation Ireland	Ireland
SV	Sathish Vadhiyar	Indian Institute of Science	India
TP	Thierry Priol	CoreGRID	France
WC	Walfredo Cirne	Federal University of Campina Grande	Brasil
WG	Wolfgang Gentzsch	D-Grid Initiative	Germany

Благодарности

Автора поддержал Европейский союз проектом EMBRACE Grid, который финансирует Европейская комиссия в рамках программы FP6; общее название тем "Науки о жизни, геномиках и биотехнология для здоровья", номер контракта LUNG-CT-2004-512092.

Ссылки

- [1] Ian Foster, Carl Kesselman. The Grid. Blueprint for a new computing infrastructure. Morgan Kaufman, 1998.
- [2] Ian Foster. What is the Grid? A Three Point Checklist.
<http://www-fp.mcs.anl.gov/~foster/Articles/WhatIsTheGrid.pdf>, July 20, 2002.
- [3] Ian Foster, Steve Tuecke. Enterprise Distributed Computing, ACM Queue, Vol. 3, No. 6 - July/August 2005.
- [4] Heinz Stockinger. Grid Computing: A Critical Discussion on Business Applicability, IEEE DS Online, 7(6), art. no. 0606-o6002, June 2006.
- [5] Oracle Grid Index, http://www.oracle.com/corporate/press/2005_apr/emeagridindex2.html, 2005.

-
- [6] <http://www.ggf.org/documents/GFD.44.pdf>
- [7] Erich Schikuta, Flavia Donno, Heinz Stockinger, Elisabeth Vinek, Helmut Wanek, Thomas Weishäupl, Christoph Witzany. Business In the Grid: Project Results, 1st Austrian Grid Symposium, OCG Verlag, Hagenberg, Austria, December 1-2, 2005.
- [8] Ian Foster, Carl Kesselman, Steve Tuecke, The Anatomy of the Grid: Enabling Scalable Virtual Organizations, International Journal of Supercomputing Applications, 15(3):200-222, 2001.
- [9] Miguel L. Bote-Lorenzo, Yannis A. Dimitriadis· Eduardo Gómez-Sánchez. Grid Characteristics and Uses: A Grid Definition. First European Across Grids Conference, Santiago de Compostela, Spain, February 13-14, 2004.
- [10] Andrew Grimshaw. What is a Grid, Grid Today, 1(26), 2002.
- [11] CoreGRID Network of Excellence, <http://www.coregrid.org>, June 2006.
- [12] Madhu Chetty and Rajkumar Buyya, Weaving Computational Grids: How Analogous Are They with Electrical Grids?, Computing in Science and Engineering, ISSN 1521-9615, Volume 4, Issue 4, Pages: 61-71, IEEE Computer Society Press and American Institute of Physics, USA, July-August 2002.