

ГРИД-СЕРВИС РЕШЕНИЯ ЗАДАЧ СТАТИСТИЧЕСКОГО ПРОГНОЗИРОВАНИЯ PREDICTOR

Лавренюк С.И., Перевозчикова О. Л., Тульчинский В.Г.

Институт кибернетики им. В.М.Глушкова НАН Украины, пр. Глушкова, 40, Киев, Украина

dep145@gmail.com

Аннотация. *Рассмотрен пример создания программного обеспечения для грид-сетей в виде прикладных сервисов, основанных на компонентной технологии. Так компоненты грид-сервиса Predictor, который описан в данной статье, используются не только в грид-сети, но и локальными пользователями вычислительного кластера СКИТ-3. Кроме решения прикладных задач пользователей – статистическое прогнозирование, грид-сервис успешно решает системные задачи по прогнозированию состояния грид-узлов и очередей задач: локальной, грид-очереди.*

Ключевые слова

Predictor, статистика, кластер, грид, грид-узел, прогнозирование, сервис.

1 Введение

В основном грид-сети построены на базе существующих вычислительных кластеров. Так, Украинский Академический грид (УАГ) [1] состоит из вычислительных кластеров и систем хранения данных, принадлежащих организациям Национальной Академии Наук Украины. И дальше в Украине идет развитие грид-вычислений.

Эволюция грид-технологии привела к возникновению архитектуры Open Grid Services Architecture [3]. В ней грид-система предоставляет обширный набор служб, которые виртуальные организации могут объединять различными способами. Архитектура OGSA, созданная на концепциях и технологиях, разработанных специалистами в области Grid и Web-услуг [4], определяет единообразную семантику предоставления услуг, стандартные механизмы для создания, именованя и обнаружения экземпляров грид-сервисов, обеспечивает прозрачность местонахождения и связывание различных протоколов и поддерживает интеграцию с базовыми механизмами нижележащих платформ.

Можно сказать, что ключевым понятием многих грид-систем является грид-сервис или грид-служба. Сервис – программный компонент, к которому можно удаленно обратиться посредством компьютерной сети, и который предоставляет функциональность запрашивающей стороне.

Преимущественно грид-сервисы выполняют функции обеспечения работы грид-сети. Но наряду с этим, необходимы грид-сервисы для решения конкретных прикладных задач, например, статистического прогнозирования [6]. В предложенном решении удалось совместить возможности решения прикладных задач, а также системных задач по оптимизации работы грид-сети и кластеров.

2 Предложенное решение

Грид-сервис предоставляет пользователю программный интерфейс, который можно представить в терминах объектно-ориентированной парадигмы. Методы, определенные в программном интерфейсе, можно удаленно вызвать через сеть, соединяющую грид-узлы, например, Интернет.

Обычно для связи между грид-узлами и грид-сервисами используется Интернет. Интернет и грид связаны между собой, но это – разные технологии. Как глобальная система сетей Интернет соединяет множество компьютеров и других сетей и позволяет им взаимодействовать. Веб – это способ доступа к информации, находящейся на удаленном, но включенном в Интернет компьютере. Грид – способ совместного использования вычислительных ресурсов и ресурсов хранения данных, распределенных по разным, территориально

удаленным компьютерам большой мощности.

Ключевым понятием грид-системы на основе специализированного программного обеспечения Globus Toolkit (GT) является грид-сервис. GT является надстройкой (расширением) операционной системы вычислительных узлов Grid-системы и реализует основные сервисы грид-системы [5].

Пакет Globus Toolkit (GT) от сообщества Globus Alliance был одним из первых программных продуктов для построения грид-сетей. В настоящее время на базе GT создано несколько проектов middleware, использующих часть функциональности GT. Например, gLite разработан в рамках проекта EGEE или NorduGrid и успешно используется в УАГ.

Исходя из сервис-ориентированности многих грид-сетей, на наш взгляд, целесообразно программы пользователей создавать как полнофункциональные грид-сервисы, но с удобным и простым способом управления ими, например, через паспорт задания. Поскольку сегодня грид-сети начинают активно применять не только для решения узко научных задач, но и прикладных задач пользователей, не являющихся специалистами по грид или кластерным вычислениям, необходимо строить прикладные грид-сервисы, ориентированные на конечных пользователей.

Прикладной грид-сервис решает определенную прикладную задачу, при этом имеет простой интерфейс (способ взаимодействия) для пользователя (модель прикладного грид-сервиса см. на рис. 1).

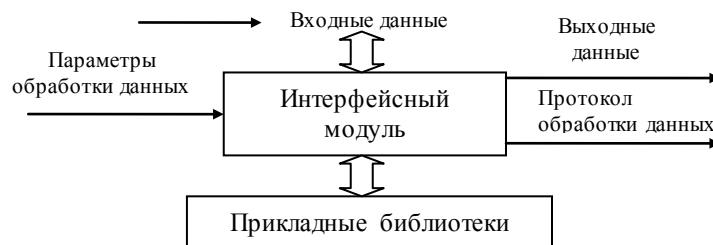


Рис.1. Модель прикладного грид-сервиса

При таком подходе пользователь создает источники данных (например, файлы) с входными данными и паспорт задания, куда заносит параметры обработки данных, в описании грид-задания указывает название грид-сервиса как Runtime Environment и направляет задачу на выполнение в грид-сеть.

По завершении задания пользователь получает выходные данные (результат) и протокол обработки данных, в котором с определённой степенью детализации указано, насколько успешно выполнена обработка. Согласно предложенному принципу, пользователь сконцентрирован только на подготовке исходных данных и паспорта задания и не занимается поиском необходимой программы, изучением ее механизма ее работы и т.п.

По такому принципу, в рамках выполнения тематического плана по созданию Национального Грида разработаны грид-сервис PREDICTOR и библиотека решения задач статистического анализа эконометрических (временных) рядов. Это программное обеспечение установлено на узле ресурсного центра УАГ ICYB SCIT-3, работающего на кластере СКИТ-3 [6] Института кибернетики им. В.М. Глушкова НАН Украины.

Как один из мощных кластеров СКИТ-3 в УАГ предоставляет пользователям оригинальное интеллектуальное программное обеспечение, например, библиотека численных алгоритмов libinpar, библиотека libarrpx для построения разных способов аппроксимации функций одной или нескольких переменных, библиотека predictor.

Назначение грид-сервиса Predictor – статистическое прогнозирование в бизнес-расчетах, например, с целью анализа объемов предполагаемых продаж, выбора ассортимента товаров и услуг по спрогнозированным ценам, удовлетворения потребностей в численности персонала. Ключевой момент - максимальная рекомендательная способность сервиса Predictor применительно к надежности прогнозов.

По определению невозможно априори оценить точность прогнозов, поэтому в специальном режиме *Тест*, предполагая неизвестными последние значения входной последовательности данных, делается попытка их спрогнозировать выбранным методом, а затем полученные прогнозы сравниваются с реальными значениями. Если они сильно отличаются, нецелесообразно строить прогноз этим методом с указанными параметрами. Иначе прогнозирование продолжается в режиме *Прогноз* с использованием всех значений исходных данных, поэтому результаты отличаются от режима *Тест* и получаются более точными.

Согласно рис. 1 грид-сервис PREDICTOR состоит из интерфейсного модуля Predictor и одноименной библиотеки. В библиотеке собраны более 26 методов статистического прогнозирования, а также функции интегральной оценки, допускающие автоматический выбор лучших методов прогнозирования наборов

данных [7], см. подробнее авторское свидетельство [8].

Библиотека Predictor предназначена для работы на вычислительном кластере с использованием MPI или в грид-сети. Библиотека Predictor дает пользователю возможность использовать готовые приложения из пакета для быстрого прогнозирования или создания своего приложения, которое будет использовать функции библиотеки Predictor, используя мощь кластера СКИТ-3 в параллельном режиме MPI.

Интерфейсный модуль состоит из функций обработки паспорта задания и из интерфейсного класса, в котором реализованы основные функции взаимодействия элементов библиотеки, и взаимодействие с пользователем.

Грид-сервис PREDICTOR может работать как полностью в автономном и автоматическом режиме, так и в режиме, когда пользователь может вызывать только определенные блоки обработки данных.

Для использования грид-сервиса PREDICTOR необходимо создать паспорт задания для интерфейсной программы. Паспорт задания для PREDICTOR описывается в текстовом файле:

```
[PREDICTOR]
[Files]
FileInput=y_in.txt;
FileOutput=y_out.txt;
FileRes=y_res.txt;
[MakeForecast]
SeasonNum=1;
HorizontNum=3;
NumBestRes=4;
...
- секция файлы
- файл с входными данными
- файл с результатом построения прогноза
- файл с результатом подбора методов прогнозирования
- секция параметры прогнозирования
- к-во сезонов
- глубина (горизонт) прогноза
- к-во методов с лучшим результатом прогноза
```

Паспорт задания для грид-сети с установленным на узле грид-сервисом PREDICTOR выглядит так:

```
&(executable="run_predictor")
(arguments="zadanie.ini")
(inputfiles= ("zadanie.ini" "zadanie.ini") ("y_in.txt" "y_in.txt"))
(outputfiles= ("y_out.txt" "y_out.txt") ("y_res.txt" "y_res.txt"))
(jobname=prognoz)
(stdout=stdout.txt)
(stderr=stderr.txt)
(cpuTime="240")
(count=16)
(runtimeEnvironment="PREDICTOR")
```

Из паспорта задания видно, что пользователь не сконцентрирован на вопросах, какое программное обеспечение установлено и в какой операционной системе. Основная задача пользователя – предоставление входных данных. Обработка данных выполняется на любом грид-узле, соответствующем параметрам задания, например, как показано на рис. 2. Грид-сервис можно установить на любом грид-узле.

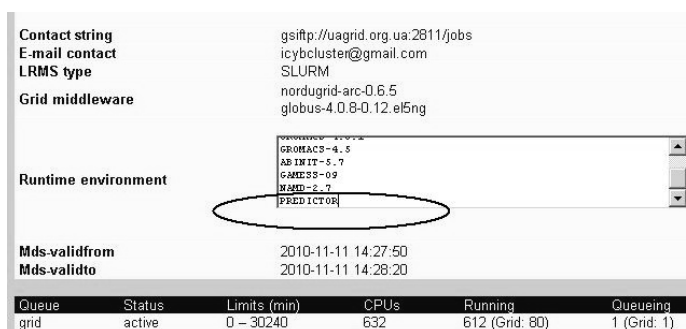


Рис.2. Список пакетов грид-узла СКИТ-3.

Задача статистического прогнозирования актуальна для многих предприятий. Если по какой-то причине пользователи не работают в грид-сети, можно использовать сервис PREDICTOR для пользователей кластера СКИТ-3 через традиционный доступ. Пользователи кластера СКИТ-3, разрабатывающие свои приложения для выполнения на кластере или в грид-сети, могут использовать библиотеку libpredictor.

Поскольку грид-вычисления еще развиваются и претерпевают изменения, то грид-сервисы целесообразно рассматривать как объекты тщательного изучения, чтобы пользователи получали удобные сервисы для решения конкретных задач.

Грид-сервіс PREDICTOR успішно застосовується для рішення системних задач, наприклад: прогнозування стану локальної черги задач кластерного комплексу; а при використанні даних про роботу ґрид-сетей, за певний період [9], прогнозує стан ґрид-вузлів, ґрид-черги [10]. На рис. 3 показано приклад прогнозування середнього за день числа виконаних задач на кластері СКІТ-3 (ось Y). Приведені дані за липень 2010 г. [9]. В проведеному експерименті взяті дані за перші 27 днів місяця (ось X), обрані 3 найкращі методи, які описують заданий часовий ряд і побудований прогноз на наступні три дні. Помилковість становила 5%.

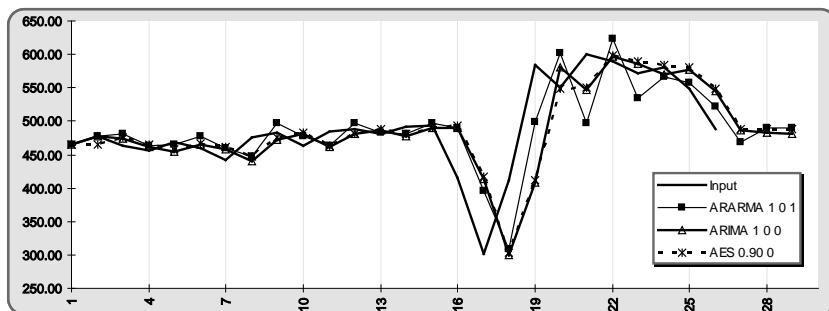


Рис.3. Прогнозування середньоденної завантаженості СКІТ-3.

3 Висновок

Все більше користувачів кластерів переходять від традиційного доступу до кластеру до роботи в ґрид-сеті. Так, за 2010 рік завантаженість СКІТ-3 ґрид-задачами становила 5.21% від загального числа виконаних задач. А за 6 місяців 2011 року кількість ґрид-задач на СКІТ-3 вже становила 11.6% (див. таблицю 1).

Таблиця 1. Список організацій, що найбільше використали ресурси СКІТ-3

№	Назва організації	% завантаження СКІТ-3	
		За 2010 г.	За 6 місяців 2011 г.
1.	Інститут клітинної біології та генної інженерії НАНУ	37,05	38,6
2.	Служби безпеки України	20,77	29,4
3.	Інститут кібернетики ім. В.М.Глушкова НАНУ	17,22	5,6
4.	Інститут фізики конденсованих систем НАНУ	6,33	<1
5.	Український національний ґрид	5,21	11,6
6.	Фізико-техн. ін-т низьких температур ім. Б.І.Веркіна НАНУ	5,21	3,9
7.	Інститут молекулярної біології та генетики НАНУ	2,00	<1
8.	Інститут математических машин та систем НАНУ	1,90	3,4
9.	Інститут гідромеханіки НАНУ	1,87	<1
10.	Ужгородський національний університет	1,53	<1
11.	Інститут фізики композитних матеріалів НАНУ	<1	3,9
12.	Запорізький ін-т гуманітарного та муніципального управління	<1	1,7

Становлять востребованими ґрид-сервіси, які вирішують прикладні задачі, а також розширюють можливість системного програмного забезпечення для оптимізації роботи користувачів. Тому планується продовжити проектування та побудову зручних ґрид-сервісів для користувачів та встановити їх на кластері СКІТ-3.

Роботи по описаним ґрид-сервісам проведені в 2010-2011 гг. по проекту №69-87 «Інтелектуалізація інформаційних технологій кластерних вичислень в ґрид-середі» згідно з Державною цільовою науково-технічною програмою «Внедрення та використання ґрид-технологій на 2009-2013 роки».

Література

- [1] Український Академічний Ґрид. - <http://grid.bitp.kiev.ua/>
- [2] I. Foster et al. The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration. *TR Glous Project*. – <http://www.globus.org/research/papers/ogsa.pdf>
- [3] Christensen E. et al. Web Services Description Language 1.1, *W3C Note*. – <http://www.w3.org/TR/wsd/>
- [4] Marco Aldinucci, Marco Danelutto, Giorgio Zoppi, Peter Kilpatrick. Advances in autonomic components & services. *From Grids to Service and Pervasive Computing*, Springer, August 2008.
- [5] Лавренко С.І., Бандура А.Ю., Горенко С.А. Порівняння та вибір ґрид-інфраструктури // *Розподілені комп'ютерні системи. Т.І. Збірн. праць конф. РКС-2010. - Київ, НТУУ «КПІ», 6-8 квітня 2010. -200 с. - С-32-35*
- [6] СКІТ-3 - <http://icybcluster.org.ua/>
- [7] Інтелектуальні пакети статистичного прогнозування /О.Л.Перевозчикова, І.Н. Пшонковська, Т.К. Терзян, В.Г. Тульчинський та др. // *Упр. системи та машини. - 1997. - №6. - С. 56-67.*

[8] Лавренюк С.И., Перевозчикова О.Л., Тульчинский В.Г., Харченко А.В. Комп'ютерна програма «Програмний пакет PREDICTOR для автоматичної побудови числових прогнозів» - *Свідоцтво про реєстрацію авторського права на твір №33934 від 30.06.2010.*

[9] Лавренюк С.И., Лавренюк А.Н., Грипич Ю.А. Построение базы данных мониторинга состояния узлов grid-инфраструктуры. *Мат. 6-й междунар. конф. ТАAPSD' –2009* Киев, 09-10 декабря 2009 г., - С. 124-131.

[10] Лавренюк С. И., Перевозчикова О. Л. Определение оптимального метода прогноза загрузки кластерных ресурсов и грид-узлов // *Кибернетика и системный анализ.* — 2011. — № 2. — С. 159-172.