

«УНИВЕРСИТЕТСКИЙ КЛАСТЕР» - ИНФРАСТРУКТУРА ИССЛЕДОВАНИЙ, РАЗРАБОТОК И ОБРАЗОВАНИЯ В ОБЛАСТИ ПАРАЛЛЕЛЬНЫХ И РАСПРЕДЕЛЕННЫХ ВЫЧИСЛЕНИЙ

А.И. Аветисян, С.С. Гайсарян, О.И. Самоваров

В настоящее время параллельные и распределенные вычисления – это одна из ключевых современных технологий, применение которой требует обширных знаний, поэтому ее внедрение и успех от использования в индустрии требует тесной интеграции с образованием. Одним из эффективных способов такой интеграции является создание единой инфраструктуры исследований и разработок в области параллельных и распределенных вычислений, а так же создание сообщества пользователей и разработчиков соответствующих технологий.

С этой целью в конце 2008 года ИСП РАН совместно с компанией Hewlett-Packard и Межведомственным Суперкомпьютерным Центром РАН при поддержке Министерства связи и коммуникаций РФ, а также Федерального агентства по делам молодежи в сотрудничестве с компанией «Синтерра» учредили программу «Университетский кластер» [1].

Одним из перспективных направлений использования распределенных вычислительных инфраструктур являются «облачные вычисления» (cloud computing)[2], которые из концепции предоставления пользователям масштабируемых ресурсов в виде Интернет сервиса, сегодня стали трендом во многом определяющим развитие информационных технологий, в том числе и развитие аппаратных архитектур и платформ. В настоящее время проектами по развитию «облачных» инфраструктур занимаются такие компании как Google (сервисы Google Mail и Google Apps - предложения ряда стандартных офисных приложений в режиме он-лайн), Microsoft (сервис Microsoft Office Live и платформа Windows Azure), Amazon, (платформа Amazon Elastic Compute Cloud) .и многие другие.

ИСП РАН, МСЦ РАН и РИЦ «Курчатовский институт» совместно стали одним из семи «центров компетенции» (Center of Excellence, COE) международного сообщества по исследованиям в области «облачных вычислений» Open Cirrus [3],[4]. Сообщество Open Cirrus организовано компаниями Hewlett-Packard, Intel и Yahoo! при участии Национального правительственного научного фонда США (National Science Foundation, NSF), Университета Иллинойса в Урбане-Шампэйн (University of Illinois at Urbana-Champaign, UIUC), Министерство развития компьютерных коммуникаций Сингапура (Infocomm Development Authority, IDA), Технологического института в Карлсруэ (Karlsruhe Institute of Technology, KIT). Целью Open Cirrus является создание распределенной тестовой вычислительной лаборатории и открытой программной платформы для поддержки разработчиков, как прикладных, так и системных программных средств в области «облачных вычислений». На каждой площадке Open Cirrus организована инфраструктура «облачных вычислений», которая включает кластерные системы с числом ядер от 1000 до 4000, аппаратные компоненты с поддержкой функций DCMI Data Center Management Interface [4], Node Manager (NM), средства аппаратной поддержки виртуализации и др. Стек программных средств входящих в состав инфраструктуры Open Cirrus построен на базе свободно-распространяемого программного обеспечения с открытым исходным кодом и включает в себя: Apache Tashi[5], Apache Hadoop[6], Pig[7] и др. Платформа Open Cirrus ориентирована на предоставление доступа исследователям в области «облачных вычислений» к следующим сервисам: PRS (Physical Resource Set) – Сервис, обеспечивающий непосредственный доступ исследователя к аппаратуре (вычислителям, устройствам хранения и коммуникаций и др. с уровнем доступа root; Cluster management service – Сервис по управлению ресурсами. В настоящее время реализовано два подхода по управлению ресурсами. Первый – это Cells as a Service (CaaS), когда распределенные ресурсы могут быть сгруппированы в виртуальные ячейки и использованы по желанию исследователя. Второй – это использование системы управления кластерами Tashi, для массивной обработки больших распределенных баз данных; Application framework services – Сервис для поддержки разработки и выполнения приложений пользователей. Для реализации и выполнения приложений предлагается использование Hadoop, Pig и MPI.

На первом этапе реализации программы «Университетский кластер» создана и развивается вычислительная инфраструктура, как тестовый полигон для параллельных и распределенных вычислений, проводятся мероприятия, целью которых является оказание поддержки участникам программы в использовании данной инфраструктуры для решения их собственных задач. Так же проводятся работы по созданию «центра компетенции» в рамках международного сообщества Open Cirrus.

Сегодня в программе «Университетский кластер» принимают участие 35 университетов, которые расположены в 21 городе Российской Федерации. За 2009 год в рамках программы «Университетский кластер» на базе 12 университетов из состава участников программы развернута распределенная вычислительная инфраструктура, в состав которой так же входят 3 ресурсных центра: МСЦ РАН, Таганрогский Технологический институт Южного Федерального Университета, Вятский государственный университет.

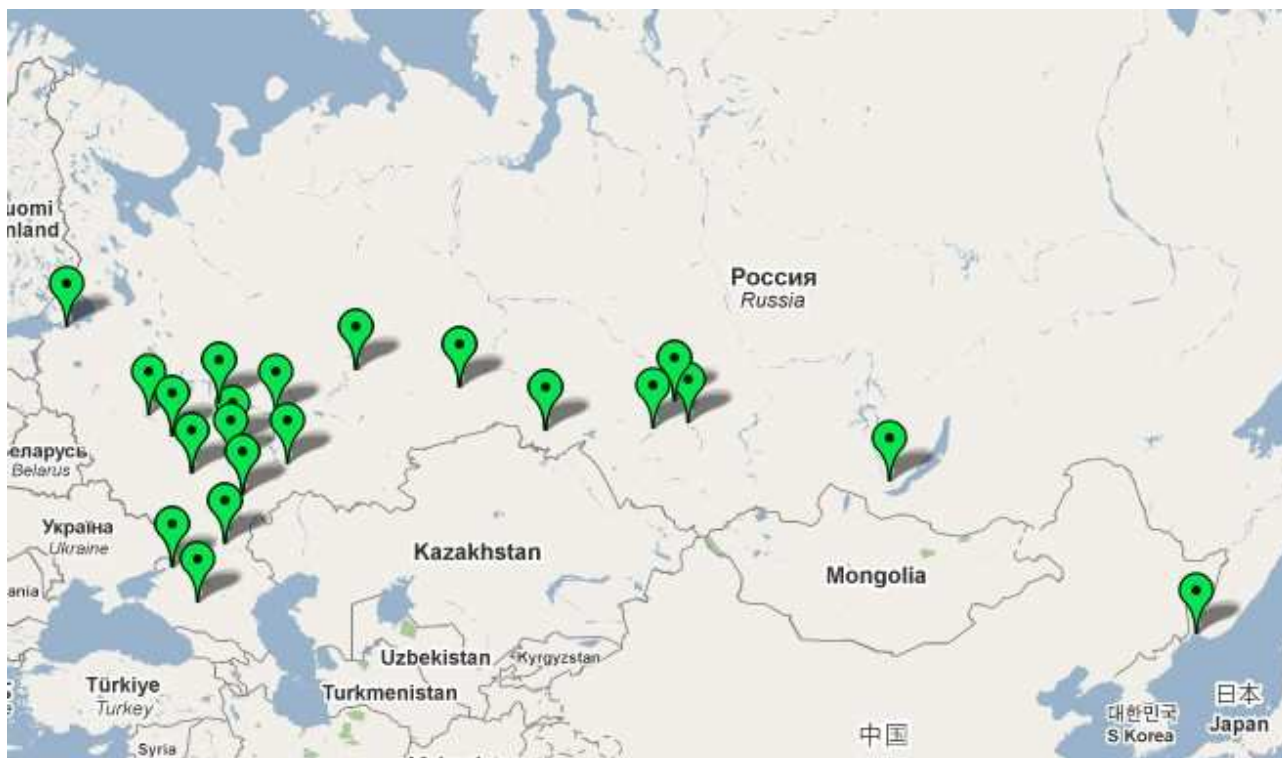


Рисунок 1. География программы «Университетский кластер».

Университетам на базе которых построена распределенная инфраструктура компания Hewlett-Packard предоставила оборудование для 12 кластерных вычислительных систем начального уровня, которое включает блейд-шасси c3000, два сдвоенных блейд-сервера половинной высоты DL 2x220c G5 в качестве вычислительных узлов, что в сумме дает 32 вычислительных ядра Intel Xeon 5554 2.33Гц, 32 Гб оперативной памяти, 480 Гб локального дискового пространства. Функции управляющего узла выполняет блейд-сервер половинной высоты DL 260c G5. В состав оборудования включено устройство Onboard Administrator, которое реализует функции безагентного управления и мониторинга всех компонент кластера – узлов, сетевых коммутаторов, блоков питания, вентиляторов.

Компания «Синтерра» объединяет участников программы, а также ресурсные центры специальной выделенной оптоволоконной сетью с пропускной способностью 10 Мбит/сек, с потенциальной возможностью расширения канала до 10Гбит/сек.

ИСП РАН является координатором программы «Университетский кластер», решает технологические задачи связанные с установкой системного программного обеспечения на кластерные системы и развертывания на их основе и основе ресурсных центров распределенной вычислительной инфраструктуры, обеспечивает информационную и техническую поддержку (включая консультации по реализации собственных проектов) ее участников. На все кластерные системы установлен и настроен необходимый системный программный стек на базе свободно-распространяемого программного обеспечения с открытым исходным кодом - операционной системы Linux CentOS и пакета OSCAR. Пакет OSCAR включает все необходимые утилиты для инсталляции и модификации кластерной системы, утилиты обеспечивающие централизованное управление и администрирование, систему пакетной обработки Torque и менеджер ресурсов MAUI, средства мониторинга Ganglia и Jobmonarch. Дополнительно на кластерные системы устанавливаются средства резервного копирования и средства поддержки пользователей кластерных систем Bugzilla и Mediawiki. Распределенная вычислительная инфраструктура развертывается с использованием пакета Globus Toolkit 4.1. и мета-планировщика Gridway. Все кластерные системы прошли обязательное тестирование на целостность и корректность работы системного окружения, а также проведены комплексные тесты с использованием пакетов HPL, PMB, SPECMPI 2007.

Для участников программы «Университетский кластер» были организованы 3 учебные группы, в составе которых в период с 27-29 апреля, 4-6 мая, 11-13 мая на базе Центра компетенции по параллельным и распределенным вычислениям они смогли пройти тренинг курс по работе на кластерных вычислительных системах. В рамках тренинг курса рассматривались вопросы, связанные с работой в операционной системе Linux, архитектурой кластерных вычислительных систем, программного стека среднего уровня на базе пакета OSCAR. Участники познакомились с архитектурой, оборудованием, программным обеспечением, предлагаемым в рамках программы «Университетский кластер», был проведен практикум по работе на одной из кластерных систем. Участники посетили технопарк компании Hewlett-Packard и Межведомственный Суперкомпьютерный Центр РАН.

Для участников программы «Университетский кластер» на базе центра компетенции по параллельным и распределенным вычислениям ИСП РАН организована служба технической поддержки, через которую они могут обратиться с вопросом или за консультационной помощью связанной с функционированием предоставленных в рамках программы кластерных вычислительных систем, функционированием созданной в рамках программы распределенной вычислительной инфраструктуры, установки и настройки системного и прикладного программного обеспечения, реализации собственных проектов с использованием технологий параллельных и распределенных вычислений и др.

Кроме того, участники программы смогут подавать заявки с предложениями по использованию в своих проектах как Российского, так и международного сегментов Open Cirrus и смогут получить бесплатно как ресурсы, так и техническую поддержку со стороны этого сообщества.

ЛИТЕРАТУРА:

1. Программа «Университетский кластер». <http://www.unicluster.ru/>
2. David Chappell. A Short Introduction to Cloud Platforms
3. Распределенная лаборатория исследований в области «облачных вычислений» Open Cirrus <https://opencirrus.org/>
4. Roy Campbell, Indranil Gupta, Mike Heath, Steve Ko. Open Cirrus Cloud Computing Testbed: Federated data centers Source Systems and Services research.
5. Data Center Manageability Interface - Optimized for Data Center Servers http://www.intel.com/technology/product/DCMI/index.htm?iid=tech_biz+dcmi
6. Tash. Система управления кластерами. <http://incubator.apache.org/tashi/>
7. Pig. <http://hadoop.apache.org/pig/>
8. Hadoop <http://hadoop.apache.org/core/>