

Среда управления задачами KIAM_JOB_CONTROL и ее применение в облачных и GRID вычислениях

Д. В. Пузырьков^{1,a}, В. О. Подрыга^{1,b}, С. В. Поляков^{1,c},
М. В. Якововский^{1,d}

¹Институт прикладной математики им. М.В. Келдыша Российской академии наук,
Россия, 125047, г. Москва, Миусская пл., д. 4

E-mail: ^a dpuzyrkov@gmail.com, ^b pvictoria@list.ru, ^c polyakov@imamod.ru, ^d lira@imamod.ru

В настоящее время большой интерес представляют научные задачи, связанные с исследованием сложных технологических процессов в реальных условиях. Решение таких задач производится методами математического моделирования и требует использования высокопроизводительных вычислительных систем различного вида, например, GRID-систем и НРС-кластеров. Внешний удаленный доступ пользователей к подобным вычислительным ресурсам обычно существенно ограничен. Это ограничение ставит перед пользователями задачу получения расчетных данных в процессе вычислений, а также перераспределения расчетов между различными доступными вычислительными системами. В данной работе представляется прототип облачной среды KIAM Job_Control, ориентированный на проведение массовых вычислительных экспериментов с помощью высокопроизводительных компьютерных систем. Среда KIAM Job_Control представляет собой многопользовательский облачный сервис, который состоит из серверной части, веб-интерфейса и клиентского приложения. Среда позволяет выполнять длительные расчеты на множестве суперкомпьютеров путем автоматической пересылки приложений и расчетных данных. Основными функциями среды являются регистрация вычислительных ресурсов, доступ к этим ресурсам, запуск приложений, их мониторинг и управление, сохранение контрольных точек, сборка рассчитанных данных, перемещение контрольной точки на другой супервычислитель, перезапуск, интерактивное взаимодействие с приложением во время вычислений. Среда KIAM Job_Control написана на языке Python с использованием специальных скриптов для организации внутрисерверных операций. Благодаря модульной архитектуре среда KIAM Job_Control может использоваться на широком классе расчетных задач с множеством различных протоколов доступа к вычислительным ресурсам, таким как SSH и HTTP. Тестирование среды проводилось на примере решения актуальных задач нанотехнологии методами молекулярной динамики. В подобных расчетах появление сверхбольших массивов данных является обычным явлением. Эффективное управление данными в этом случае становится одной из важных частей общего алгоритма. Использование среды KIAM Job_Control подтвердило этот факт и позволило существенно ускорить процесс вычислений, проводившихся на нескольких вычислительных кластерах.

Ключевые слова: Облачные технологии, Управление задачами, Управление ресурсами, Планирование параллельных задач

Работа выполнена при поддержке Российского фонда фундаментальных исследований (проекты 15-07-06082-а, 15-29-07090-офи_м).

© 2016 Дмитрий Валерьевич Пузырьков, Виктория Олеговна Подрыга, Сергей Владимирович Поляков, Михаил Владимирович Якововский

Введение

В настоящее время основным способом исследования математических моделей и проверки их качественных и количественных показателей служит вычислительный эксперимент. Используемые в его рамках математические модели и их компьютерные реализации могут оперировать огромным числом параметров и входных данных. Необходимость поддержки решения подобных задач привела к массовому использованию различного рода суперкомпьютеров (НПС-кластеры, GRID-системы и др.).

Очевидно, что способы и полнота доступа к различным суперкомпьютерным системам могут существенно различаться. В локальных сетях небольших организаций реализуется полномасштабный прямой и/или удаленный доступ ко всем вычислителям. В случае же использования больших вычислительных систем коллективного доступа возможности рядового пользователя, наоборот, сильно ограничены. В таких системах пользователям предоставляются учетные данные и квота на дисковое пространство и процессорное время. В результате, пользователю, выполняющему масштабный расчет, не уместающийся на ресурсе по какой-либо квоте, приходится самостоятельно следить за использованием квот и в ручном режиме перезапускать задачи и сохранять рассчитанные данные по сети на собственный компьютер. При наличии у пользователя доступа к нескольким удаленным вычислительным системам ему приходится перемещать расчетные данные с одного ресурса на другой и перезапускать расчеты также в ручном режиме.

Данная проблема побудила авторов к созданию среды, позволяющей объединить все доступные пользователю вычислительные ресурсы в одной управляющей платформе. Такая организация доступа похожа на классический GRID, однако вместо распараллеливания решения одной задачи, она призвана поддерживать параллельное выполнение множества задач с учетом необходимых для них ресурсов, а так же следить за состоянием задач и обеспечивать их миграцию с ресурсов, исчерпавших квоту, на ресурсы, имеющие свободные мощности.

В данной работе представлен прототип облачной среды KIAM Job_Control, позволяющей проводить масштабные суперкомпьютерные расчеты с учетом указанной выше функциональной схемы. Среда является многопользовательским облачным сервисом, состоящим из серверной части, веб-интерфейса и клиентского приложения. KIAM Job_Control дает возможность выполнять пользовательские задачи на множестве суперкомпьютеров путем автоматической пересылки задач и их расчетных данных с одного ресурса на другой. Представленная среда является частью программного комплекса KIAM Multilogin [2], нацеленного на решение многомасштабных задач нанотехнологии на высокопроизводительных вычислительных системах.

Постановка задачи

В результате исследования потребностей, возникающих у пользователей суперкомпьютерных ресурсов, были сформированы следующие особенности, которыми должна обладать среда, описанная выше.

Во-первых, среда должна предоставлять функционал управления вычислительными ресурсами, доступными пользователю:

- позволять пользователю вести базу данных всех доступных ему вычислительных ресурсов, их учетных данных и возможностей;

- автоматически определять тип ресурса, доступную квоту по дисковому пространству и по процессорному времени, а так же позволять самостоятельно задавать необходимые данные и время доступности системы, используя календарь;
- предоставлять единую точку доступа ко всем зарегистрированным в ней ресурсам.

Во-вторых, в среде должен быть реализован функционал управления пользовательскими задачами:

- регистрация задач пользовательского типа;
- указание вычислительных ресурсов, доступных для выполнения каждой конкретной задачи;
- развертка пользовательских задач на указанных ресурсах, компиляция из исходных кодов при необходимости, или использование предоставленных пользователем бинарных файлов или скриптов;
- запуск и остановка пользовательских задач.

В-третьих, необходим функционал мониторинга:

- отслеживание статуса выполняющихся задач;
- отслеживание оставшихся квот и общей доступности ресурсов;
- получение стандартного вывода задачи и взаимодействие с ее стандартным вводом.

В-четвертых, требуется механизм миграции выполняющихся задач с ресурсов, исчерпавших квоту, на другие доступные ресурсы:

- подготовка задачи на свободном ресурсе по истечению квоты на активном ресурсе;
- учет времени, необходимого на переброску расчетных данных задачи, для минимизирования времени простоя;
- функционал для сокращения времени пересылки задач, позволяющий по возможности обходиться без управляющего сервера при миграции исходных данных.

Также от среды требуется отсутствие необходимости установки управляющей части на используемые вычислители, так как у пользователя может не оказаться прав соответствующего уровня доступа.

Обзор существующих решений

В результате поиска подходящих под описанные требования систем были найдены следующие программные комплексы.

Metacluster [5] — мощная система управления множеством вычислительных кластеров, предоставляющая единую точку доступа к ним. Система имеет планировщик задач, позволяющий в автоматическом режиме определять наиболее эффективный ресурс для пользовательского задания. Однако при этом требуется установка управляющей части на используемые вычислители, что не позволяет с его помощью связать произвольные суперкомпьютеры в одной системе.

FlyElephant [4] — платформа, которая предоставляет ученым и инженерам готовую инфраструктуру для проведения высокопроизводительных вычислений. Платформа позволяет объединять множество вычислительных ресурсов, однако является коммерческой разработкой и имеет закрытый исходный код.

В итоге, ни одна из указанных, а также других ранее использовавшихся платформ не проектировалась с целью поддержания широкомасштабных вычислений с использованием гетерогенного множества вычислительных систем и перераспределения вычислительных задач между ними в процессе длительных вычислений. Поэтому в данной работе была предпринята попытка спроектировать и реализовать среду, полностью учитывающую вышеуказанные требования.

Архитектура KIAM Job_control

Предлагаемая среда KIAM Job_control построена на клиент-серверной архитектуре. В состав среды KIAM Job_Control входят:

1. серверное приложение kiamjobctl, которое предоставляет доступ к различным вычислительным ресурсам, осуществляет запуск задач и мониторинга их выполнения;
2. CLI интерфейс kiamjobctlm — основной способ взаимодействия со средой, который предоставляет командный интерфейс для настройки среды, регистрации задач и управления ресурсами;
3. веб-интерфейс, предоставляемый приложением kiamjobctlweb, который позволяет пользователям работать со средой в браузере, удобно отображать статистику и управление задачами.

Заметим, что серверное приложение kiamjobctl можно установить как на общий сервер в режиме многопользовательского доступа, так и на локальную рабочую станцию в однопользовательском режиме. Все управление, настройки и пользовательские действия выполняются через HTTP REST API. Такая схема позволяет легко использовать в качестве клиента к этой системе как веб-интерфейс, так и CLI (Command Line Interface — интерфейс командной строки), установленный локально.

Схема взаимосвязей и взаимодействия компонент показана на рис. 1.

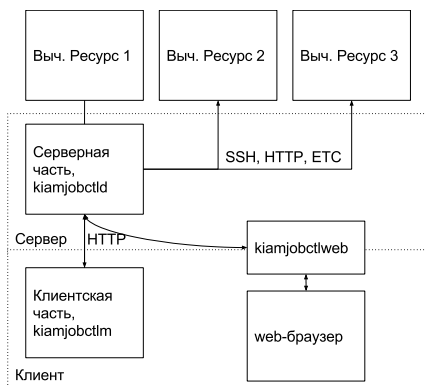


Рис. 1. Компоненты среды KIAM Job_Control и их связи.

Доступ и управление ресурсами, описание пользовательских задач

Доступ к ресурсам. На низком уровне доступ к вычислительным ресурсам происходит посредством модулей, написанных на языке Python и поддерживающих необходимый API, состоящий из необходимых методов, таких как разворачивание задачи, сборка, запуск, мониторинг и другие. В случае с Linux-ресурсами, доступными по SSH, это модуль, использующий библиотеку Paramiko и разворачивающий на целевой системе специальные bash-скрипты, которые выполняют компиляцию, запуск и автономный мониторинг. Через эти же API среда KIAM Job_control получает информацию о возможностях конкретного вычислителя. На данный момент KIAM Job_control поддерживает только вычислительные ресурсы с доступом по SSH.

Описание пользовательских задач. Для того чтобы пользователь мог запускать какие-либо вычисления на доступных ресурсах, он должен описать задачу и ее особенности. Структура описания пользовательской задачи представляет собой директорию, в которой находятся исходные коды приложения или бинарный файл, все необходимые для задачи исходные данные, а также текстовый файл с описанием задачи на языке YAML. В таком файле пользователю нужно переопределить установленные по умолчанию параметры задачи, а также указать систему сборки (на данный момент это GNU Make и пользовательский build-скрипт). После того, как это описание будет готово, пользователю нужно отправить архив этой директории на сервер, где его проверят на правильность заданных значений и внесут в базу задач. После этого пользователь может запускать развертывание задачи на конкретном вычислительном ресурсе или указать ресурсы, на которых нужно постоянно выполнять данную задачу, мигрируя ее по истечению заданной квоты.

Управление задачами на ресурсе. На данный момент поддерживается только один способ управления работающей задачей — использование системы управления кластером SLURM или стандартной системы управления процессами Linux. В будущем планируется дополнить среду управления специальной библиотекой, которая позволит давать команды на останов, сохранение данных и другие через файловый интерфейс приложения. Приложение kiamjobtld по истечению таймаута проверяет квоты и возможности узла, делая на их основе вывод, стоит ли останавливать задачу. По истечению заданной (или определенной автоматически) квоты среда инициирует останов задачи, и при необходимости мигрирует задачу на другой вычислительный ресурс. Таким же образом реализован механизм обратных вызовов: если развернутые скрипты имеют http доступ к управляющему серверу, они до истечения таймаута могут инициировать какое-либо действие, например, перекачку рассчитанных данных в удаленное хранилище.

Практическое применение

Тестирование прототипа среды KIAM Job_Control было проведено на реальной программе из комплекса GIMM_NANO [1]. Целями вычислений являлись масштабное молекулярно-динамическое моделирование, а так же анализ его результатов и их визуализация [3]. Задача представляла собой моделирование установления термодинамического равновесия в микросистеме азот-никель при различных температурах газа. В качестве вычислительных ресурсов использовались доступные на тот момент кластеры МВС-10П (МЦ РАН), "Ангара-К1" (АО "НИЦЭВТ") и ИММ6 (ИПМ им. М.В. Келдыша РАН). Для визуализации и анализа расчета требовалось рассчитать и хранить около 3 ТБ данных на один расчетный вариант. При этом имелись довольно жесткие квоты по процессорному времени и дисковому пространству: МВС-10П была доступна 24 часа в неделю с 0.5 ТБ дисковой квоты; "Ангара-К1" был доступен 94 часа в неделю с 10 ТБ дисковой квоты; ИММ6 был

доступен 168 часов в неделю с дисковой квотой в 60 ТБ. При этом визуализировать данные можно было только на ИММ6, потому что только на данном кластере находится специальное программное обеспечение для визуализации. Использование среды KIAM Job_Control позволило выполнить эту задачу с минимальным вмешательством пользователя для миграции данных с вычислителя на вычислитель, в автоматическом режиме перезапуская задачу на доступном вычислителе и сохраняя данные в хранилище ИММ6.

Заключение

В работе представлен прототип облачной среды KIAM Job_control, целью которого является проведение масштабных вычислительных экспериментов с использованием высокопроизводительных систем. Прототип позволяет объединить доступные пользователю вычислители различного типа и автоматически перераспределять приложения и данные между ними. Благодаря модульной архитектуре среда KIAM Job_Control поддерживает узлы любой внутренней организации и со множеством различных протоколов доступа к вычислительным ресурсам. Тестирование текущей реализации среды KIAM Job_Control проводилось на примере решения актуальных задач нанотехнологии методами молекулярной динамики, для которых характерны сверхбольшие объемы расчетных данных. Использование среды KIAM Job_Control подтвердило важность эффективного управления данными в подобных задачах и позволило существенно ускорить процесс вычислений, проводившихся на нескольких вычислительных системах. В будущем планируется доработка планировщика задач, которая позволит учитывать время перекачки данных с упреждением для сокращения времени простоя ресурсов.

Список литературы

- Bondarenko A. A., Kononov E. M., Kosolapov O. A., Polyakov S. V., Yakobovskiy M. V.* Programmnyi kompleks GIMM_NANO [The GIMM_NANO software suite]. International Supercomputer Conference “Scientific service in Internet network: all sides of parallelism”(September 23-28, 2013, Novorossiysk). - Moscow: Press MSU, 2013. - P. 333 - 337 (in Russian)
- Polyakov S. V., Vyrodov A. V., Puzyrkov D. V., Yakobovskiy M. V.* Oblachniy servis dlya resheniya mnogomoshstbnish zadach nanotekhnologii na superkompiuternih sistemah [Cloud service for multiscale nanotechnology computations on supercomputers] // Proceedings of ISP RAS. - 2015. - Vol. 27, No. 6. - P. 409 - 420. (in Russian)
- Puzyrkov D. V., Podryga V. O., Polyakov S. V.* Raspredeleonnaya vizualizacia v zadachah modelirovaniya termodinamicheskogo ravnovesiya v microsistemah gaz-metall metodami molekulyarnoy dinamiki [Parallel visualization of the Molecular Dinamic simulation in the gas-metal microsystem] // Parallel Computing Technologies (PCT'2016): Proceedings of the International Scientific Conference (28 March - 1 April 2016, Arkhangelsk). - Chelyabinsk: South Ural State University Publishing Center. - P. 284 - 297. (in Russian)
- FlyElephant [Electronic resource]: <https://flyelephant.net/>
- Metacluster [Electronic resource]: <http://cluster.software.unn.ru/MC/Management.aspx>

KIAM_JOB_CONTROL task management environment and its application to cloud and GRID computing

D. V. Puzyrkov^{1,a}, V. O. Podryga^{1,b}, S. V. Polyakov^{1,c}, M. V. Iakobovskii^{1,d}

¹Keldysh Institute of Applied Mathematics,
4 Miusskaya sq., Moscow, 125047, Russia

E-mail: ^a dpuzyrkov@gmail.com, ^b pvictoria@list.ru, ^c polyakov@imamod.ru, ^d lira@imamod.ru

Now the scientific tasks related to a research of complex technological processes in actual practice are of great interest. The solution of such tasks is made by methods of mathematical modeling and requires use of high-performance computing systems of different type, for example, GRID systems and HPC clusters. Remote user access to such computing resources is usually quite limited. This limitation poses the problem of data acquisition during the process of calculation, as well as the redistribution of tasks between different available computer systems. This paper presents a prototype of a cloud service KIAM Job_Control, focused on holding the mass computational experiments on the different high-performance computer systems. The KIAM Job_Control is a multiuser cloud environment, which consists of the server side, the web interface and client application. This environment allows to perform long calculations on a set of supercomputers by automatically forwarding the applications and computational data. Main tasks of this service are computational resource registration and access, application deploying and starting, management and monitoring of resources, gathering computational data, relocation of data-save-point to another resource and interactive communication with the application. KIAM Job_Control service is written in Python using special scripts for the organization of internal resource operations. Due to the modular architecture it can be used on a wide class of computational problems and computational resources with different access protocols such as SSH and HTTP. Testing was performed on the example of solution of actual problems of nanotechnology using molecular dynamics. In these calculations an appearance of ultra-large data sets is a common phenomenon. Effective data management in this case becomes one of the important parts of the overall algorithm. Use of the KIAM Job_Control service confirmed this fact and helped to speed up the calculation process performed using several computer clusters.

Keywords: Cloud computing, Task management, Resource management, Scheduling parallel tasks

The work was supported by Russian Foundation for Basic Research (projects 15-07-06082-a, 15-29-07090-ofi_m).

© 2016 Dmitry V. Puzyrkov, Viktoria O. Podryga, Sergey V. Polyakov, Mikhail V. Iakobovskii