**КОМПЬЮТЕРНЫЕ ИССЛЕДОВАНИЯ
И МОДЕЛИРОВАНИЕ 2015 Т. 7 № 3 С. 691−698**

Ки&М

СЕКЦИОННЫЕ ДОКЛАДЫ

УДК: 004.414.23, 519.876.5

Синтез процессов моделирования и мониторинга для развития систем хранения и обработки больших массивов данных в физических экспериментах

В. В Кореньков[a], А. В. Нечаевский, Г. А. Ососков, Д. И. Пряхина, В. В. Трофимов, А. В. Ужинский

Лаборатория информационных технологий, Объединенный институт ядерных исследований, Россия, 141980, Московская обл., г. Дубна, ул. Жолио-Кюри, д. 6

E-mail: [a]symsim@jinr.ru

Получено 10 февраля 2015 г.

Представлена новая система моделирования грид и облачных сервисов, ориентированная на повышение эффективности их развития путем учета качества работы уже функционирующей системы. Результаты достигаются за счет объединения программы моделирования с системой мониторинга реального (или модельного) грид-облачного сервиса через специальную базу данных. Приведен пример применения программы для моделирования достаточно общей облачной структуры, которая может быть также использована и вне рамок физического эксперимента.

Ключевые слова: имитационное моделирование, грид, облака, хранение данных, оптимизация, мониторинг

Synthesis of the simulation and monitoring processes for the development of big data storage and processing facilities in physical experiments

V. V. Korenkov, A. V. Nechaevskiy, G. A. Ososkov, D. I. Pryahina, V. V. Trofimov, A. V. Uzhinskiy

Joint institute for nuclear researches, Laboratory of Information Technologies, 6 Joliot-Curie st., Moscow reg., Dubna, 141980, Russia

The paper presents a new grid and cloud services simulation system. This system is developed in LIT JINR, Dubna, and it is aimed at improving the efficiency of the grid-cloud systems development by using work quality indicators of some real system to design and predict its evolution. For these purpose, simulation program is combined with real monitoring system of the grid-cloud service through a special database. The paper provides an example of the program usage to simulate a sufficiently general cloud structure, which can be used for more common purposes.

Keywords: grid computing, cloud computing, data storage, monitoring, optimization, simulation

Работа выполнена при поддержке Гранта РФФИ № 14-07-00215.

Citation: *Computer Research and Modeling*, 2015, vol. 7, no. 3, pp. 691–698 (Russian).

Введение

В различных областях деятельности существует множество вычислительных систем различного масштаба для обработки информации. Наибольший интерес представляют системы, обрабатывающие сверхбольшие объемы данных. В качестве примера можно привести WLCG — грид-систему распределенной обработки данных Большого адронного коллайдера (БАК). По доступной статистике, объем информации, сохраненный и обработанный в четырех экспериментах БАК на протяжении первого прогона (RAN 1), составил сотни петабайт.

В 2015–2020 гг. на экспериментах БАК ожидается увеличение объема данных и неизбежный переход к грид-облачным комплексам. Это необходимо для потенциально новой физики, но сталкивается с новыми серьезными требованиями к компьютингу БАК, а именно:

- значительное увеличение вычислительных мощностей и сетевых ресурсов хранения данных;
- необходимость доступа к данным из грид и облаков;
- активное использование распределенных параллельных вычислений;
- совершенствование кодов программ анализа и моделирования.

Столь быстрое развитие распределенных вычислительных систем требует непрерывного моделирования всех процессов хранения, передачи и анализа данных.

В настоящее время при проектировании грид-систем используется подход, когда задача создания модели и формулировки рекомендаций по построению выполняется однократно при проектировании системы. В предыдущей работе авторов [Кореньков и др., 2013] описана программа моделирования, основанная на использовании языка GridSim [GridSim..., 2012] и алгоритмов планирования потока заданий ALEA [Klusacek et al., 2008]. Для запуска программы требуется задать состав и топологию центров обработки моделируемой грид-структуры, а также распределение ресурсов между заданиями. После этого программа выполняет имитационное моделирование процессов прохождения сгенерированного набора заданий через грид-структуру. В качестве результатов вычисляются временные оценки искомых параметров потока заданий.

Однако эксперименты продолжаются годами и десятилетиями, одновременно с эксплуатацией системы происходит ее развитие, не только качественное, но и количественное. При эволюции WLCG произошло качественное изменение систем хранения информации, а вместо планируемых трех уровней обработки данных появилось четыре. Таким образом, даже при значительных усилиях, вложенных на этапе проектирования в понимание конфигурации систем и их количественных характеристик, невозможно развивать систему без дополнительных исследований. Разработчики и эксплуатирующие организации сталкиваются с проблемой прогнозирования поведения системы после проведения планируемых модификаций.

Моделирование системы позволяет ответить на ряд вопросов. При создании распределенной системы требуется принять решения по архитектуре инфраструктуры, количеству ресурсных центров, объему требуемых ресурсов. Кроме того, необходимо обеспечить достаточную пропускную способность, решить проблемы сохранности данных (устойчивость к повреждениям и удалениям) на протяжении всего жизненного цикла проекта, обеспечить распределение ресурсов между различными группами пользователей, выбрать алгоритмы обработки и запуска задач и многое другое.

Таким образом, требуется создание методологии и программного окружения, позволяющего моделировать системы на постоянной основе, прогнозировать поведение системы при значительных изменениях.

Объединив моделирование и мониторинг в рамках одного программного пакета, можно добиться существенного снижения эксплуатационных затрат и вложений в увеличение мощности с целью сохранения скорости получения результата экспериментов при постоянном повышении потока данных.

Выбор средств моделирования

Говоря о том, какую технологию моделирования применить, следует учесть, что возможность применения аналитических моделей для рассматриваемых задач ограничена по следующим соображениям. Существует несколько подходов при аналитическом моделировании грид- и облачных систем, которые можно сгруппировать в два типа:

– система рассматривается как многоканальная система массового обслуживания с состояниями, управляемыми марковским процессом, с ограничениями на распределения входных потоков и на дисциплины обслуживания, вызванными теоретическими предпосылками;

– система рассматривается как динамическая стохастическая сеть, описываемая системами уравнений, позволяющими учитывать как маршрутизацию, так и распределение ресурсов в сети, причем изучению подлежат равновесные и неравновесные состояния сети [Попков, 2003].

Оба подхода выдают результат моделирования, как правило, в виде асимптотических распределений и в силу ограниченных теоретических предпосылок не могут быть применены для моделирования конкретных сложных компьютерных сетей многоуровневой архитектуры с реальными распределениями входных потоков заданий, сложной многоприоритетной дисциплиной их обслуживания и динамическим распределением ресурсов. Поэтому мы считаем правильным использовать имитационное моделирование.

На сегодняшний день существуют различные программные инструменты имитационного моделирования грид-систем и облаков [Nechaevskiy, Korenkov, 2009; Кореньков, Муравьев, Нечаевский, 2014]. Например, GridSim — библиотека классов, предназначенных для построения модели грид-системы. Она, в свою очередь, построена на стандартной библиотеке SimJava, с помощью которой можно моделировать поток дискретных событий во времени. Однако моделирование облачных вычислительных центров этой системой не предусмотрено.

Облачные вычислительные центры могут быть определены как тип параллельных и распределенных систем, состоящих из набора взаимосвязанных и виртуальных компьютеров, которые предоставлены динамически как один или несколько объединенных вычислительных ресурсов на основе соглашения об уровне обслуживания через договор между провайдером сервиса и потребителем. Для моделирования облачных инфраструктур существуют различные программные продукты, например CloudSim, iCanCloud, CReST (см. обзор в [Кореньков, Муравьев, Нечаевский, 2014]). Эти программные пакеты позволяют создавать модели облачных систем с определенной функциональностью и конфигурацией. Готовая модель запускается на прогон с модельным потоком заданий, в результате чего системы моделирования предоставляют статистическую информацию по наиболее важным характеристикам: время выполнения задач, жизненный цикл виртуальных машин, использование ресурсов. Эти системы моделирования ориентированы на моделирование определенного уровня облака. Функциональность CloudSim позволяет наиболее подробно моделировать уровни SaaS и IaaS. Для анализа работы уровней PaaS и SaaS облачной инфраструктуры можно использовать iCanCloud. Разработку дата-центра с минимальными затратами электроэнергии и эффективным охлаждением можно реализовать в CReST, который подробно моделирует PaaS-уровень. Однако представленные системы моделирования рассчитаны на решение своих узкоспециализированных задач и не обладают набором функций для полноценного моделирования облачных вычислительных центров для хранения и обработки данных физических экспериментов.

Предлагаемое нами программное решение основано на расширении классов GridSim и их объединении в программу, которая моделирует обработку потока заданий грид-облачной структурой, обладающей заданными ресурсами и дисциплиной их резервирования и использования. В последнее время выдвигается идея интеграции в грид-инфраструктуры центров, построенных по принципу облачных вычислений, а также реализации служб грид на оборудовании «облачных» центров. Поэтому методы и средства, которые разрабатываются в рамках про-

екта, допускают моделирование объединения в грид-инфраструктуру центров, имеющих облачную архитектуру.

Описание подхода к моделированию

Постоянное развитие современных грид-систем требует непрерывных корректировок большинства параметров моделирования. Это необходимо для прогнозирования поведения системы при значительных ее изменениях. Для корректировки параметров предлагается использовать статистику эксплуатации системы, получаемую на основе имеющихся программных средств ее мониторинга.

В связи с этим возникают две проблемы:

1) обеспечение совпадения исходных данных для модели с реальными;

2) проверка адекватности моделирования, т. е. доказательство того, что моделирование произведено корректно и поведение модели не отличается от поведения реальной системы.

Наш подход состоит в следующем.

1. Если речь идет о модернизации существующей установки обработки, то использовать подходящие накопленные данные. К примеру, в проекте WLCG имеются как глобальные, так и специализированные под конкретные эксперименты системы мониторинга и аккаунтинга. При этом результаты моделирования обработки потока заданий должны совпадать в пределах погрешности с результатами мониторинга прохождения того же потока заданий в системе.

2. Для новых установок эта проблема разрешается выдвижением гипотез о типах потоков входной информации, их параметрах, и процедурах их обработки с последующим моделированием как самих входных потоков, так и процессов их обработки. Такие гипотезы можно сформулировать на основании данных мониторинга подобных систем (оценивая интенсивность и основные характеристики потоков заданий и файлов). Обработка результатов моделирования заключается в анализе распределения времени событий, которые генерируются при обработке входного потока данных. Затем эти распределения сравниваются с результатами, полученными из мониторинга существующей системы.

Таким образом, модель должна рассматриваться как неотъемлемая часть системы обработки данных, а данные мониторинга — как входные для моделирования. Это позволит принимать более обоснованные проектные решения при развитии системы.

Предлагаемый нами подход состоит в интеграции средств мониторинга процессов прохождения задач и передачи данных с возможностями имитационного моделирования. В рамках этой концепции техническое решение проверяется на модели прежде, чем обсуждается его практическая реализация. В идеале процесс принятия решений по развитию вычислительной установки должен выглядеть следующим образом: данные мониторинга реальной грид-системы поступают в базу данных, далее на основе данных мониторинга пользователь задает входные параметры модели и потока заданий, модель обрабатывает задания и возвращает пользователю результат для дальнейшего анализа. Центральным компонентом такой процедуры принятия решения является имитационная модель вычислительной структуры, в которую в качестве входной поступает информация, накопленная в ходе мониторинга существующей установки и модифицированная в соответствии с представлениями о том, как она будет меняться.

Для реализации модели потребовались существенные изменения GridSim:

- введены классы, описывающие специфическое для облачных центров хранилище информации;
- входной поток заданий формируется через базу данных;
- принцип обмена данными изменен с симуляции пакетов на симуляцию передачи файла;
- обработка результатов моделирования вынесена за рамки программного пакета.

Для иллюстрации возможностей разработанной программы SyMSim (Synthesis of Monitorng and Simulation) ниже приведен пример ее применения для оптимизации простой облачной структуры.

Пример использования программы SyMSim

Объектами моделирования являются вычислительные установки, предназначенные для обработки информации объемом до десятков петабайтов в год, который производят ускорители заряженных частиц, например LHC-CMS [CMS detector, 2015], LHC-Atlas [ATLAS detector, 2015] и находящиеся в процессе создания или проектирования FAIR-PANDA [PANDA..., 2015], BES-III [BESIII..., 2015], NICA-MPD [Сисакян, Сорин, 2011]. Как показал многолетний опыт работы центров разных уровней для распределенных вычислений и хранения данных, объединенных в систему WLCG, единственный способ хранения объемов информации, производимых такими детекторами, является использование роботизированных библиотек. Данные затем обрабатываются на фермах, включающих тысячи процессоров. Предполагается, что моделируемая структура предназначена для обработки данных физического эксперимента, но другие структуры, связанные с хранением и обновлением больших массивов цифровой информации, также могут быть смоделированы.

Итак, рассматривается модель реализации облачной структуры, предназначенной для хранения данных в роботизированной библиотеке с тысячами кассет с магнитными лентами, из загрузчиков-драйвов которых робот автоматически достает требуемые ленты и устанавливает в одно или несколько устройств чтения–записи. Схема прохождения задания через систему моделирования SyMSim представлена на рис. 1. Задание начинает выполняться, если есть свободный слот-процессор и все файлы доступны на дисковом хранилище облака. Если файл хранится в роботизированной библиотеке, задание резервирует слот, но выполнение задерживается до момента его загрузки на диск. Процесс перемещения файла из библиотеки в дисковое хранилище включает в себя операцию помещения ленточного картриджа на драйв, которую выполняет рука робота, монтирования файловой системы картриджа на драйве и записи файла на диск.

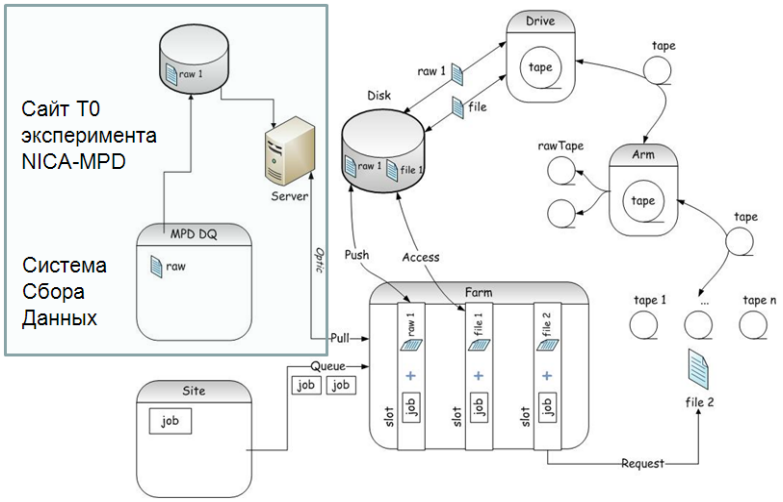

Рис. 1. Схема прохождения заданий через SyMSim

Проектируемая структура состоит из ленточного робота IBM 3500 [IBM..., 2015], массива ленточных картриджей (220 лент) и кластера из 100 абстрактных процессоров. Было взято 9 драйвов LTO6 для пула из 150 лент для работы с файлами и 2 драйва LTO6 для пула из 70 лент для записи файлов с «сырыми» данными (RAW), технические параметры которых соответствуют реальным. Дисковый пул T1 — 590 ГБ. Канал связи — 10 Гб/с. Имитация заключалась в моделировании прохождения 1000 заданий по этой структуре. Поток заданий генерируется на основе распределений, полученных при статистическом анализе данных, доступных для эксперимента Atlas.

Рассмотрим на этом примере возможности модели.

Моделирование нагрузки на процессоры показано на рис. 2, а ее равномерность во времени — на рис. 3. Равномерность загрузки зависит от равномерности поступления заданий, настроек очередей, системы вторичной памяти и т. д. Все астрономическое время выполнения пакета разбито на одинаковые интервалы.

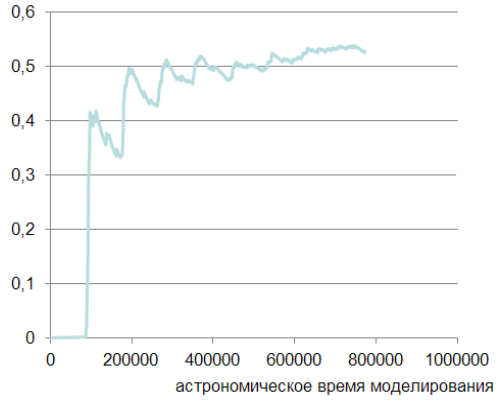

Рис. 2. Отношение затраченного процессорного времени к максимально возможному затраченному к этому моменту

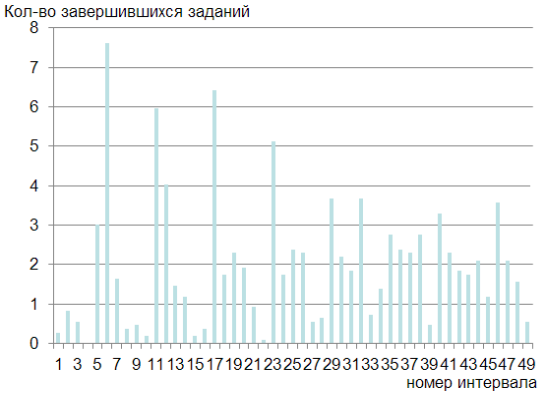

Рис. 3. Количество заданий, завершившихся во временном интервале

Размер дискового буфера. Одно из ограничений модели: одно задание может требовать только один файл. Однако разные задания могут требовать один из файлов, которые уже подгружены. Алгоритм сборщика мусора заимствован из dCache [dCache..., 2015]. Возникает вопрос, хватит ли нам буфера. На рис. 4 мы видим, что буфер используется от 60 % до 80 %.

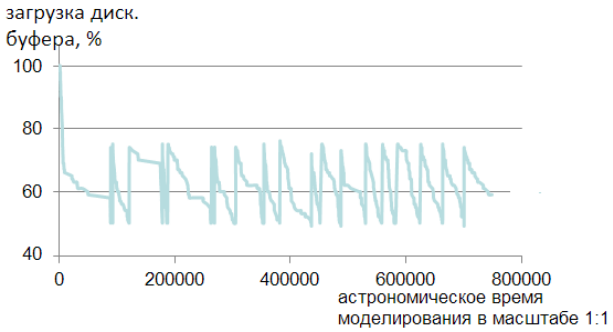

Рис. 4. Процент загрузки дискового буфера

Нагрузка на руку робота показана на рис. 5. Нагрузка определяется следующим образом: исходя из среднего времени движения руки 6 с, вычисляется максимальное количество движений за временной промежуток. Загрузка — отношение количества движений при моделировании к максимально возможному количеству движений. Оказалось, что рука робота будет загружена не более чем на 4 %. Причем вначале нагрузка на руку возрастает, потому что идет массовая загрузка файлов, которые требуются для выполнения задач, а потом нагрузка снижается, потому что часть файлов уже есть в буфере.

Таким образом, для данной интенсивности потока задач мощность вычислительных узлов достаточна, если SLA допускает прохождение данного пакета за $0.8 \cdot 10^6$ с. Рука робота загружена слабо, т. е. можно использовать шкафы высокой плотности. Диски объемом 0,5 ТБ достаточны для поставленных задач.

Результаты моделирования по критерию минимального времени прохождения задания при достаточно высокой загрузке процессоров могут служить обоснованием выбора конфигурации облачного кластера и аргументом в пользу покупки или отклонения более дорогого оборудова-

ния, хотя не следует забывать, что на выбор конфигурации также влияют и другие соображения: надежность, перспективы развития, величина резерва и т. д.

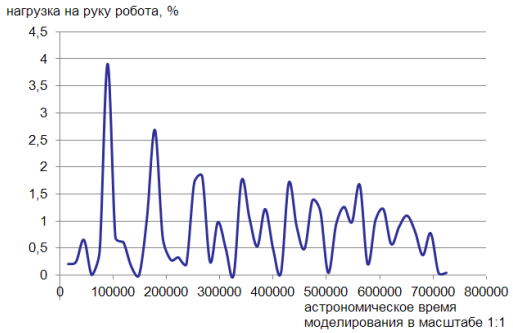

Рис. 5. Нагрузка на руку робота в процентах от максимальной

Заключение

Разработка грид-облачных систем сбора, передачи и распределенной обработки информации требует тщательного моделирования, эффективность которого зависит от наличия динамических данных о качестве работы уже функционирующей инфраструктуры. Авторами разработана система моделирования облачных вычислительных центров SyMSim. Предложенный подход к моделированию и анализу вычислительных грид-облачных структур основан на учете данных их мониторинга, используемых затем для динамической коррекции параметров моделирования. Новизна подхода в моделировании состоит в соединении моделирования и мониторинга в рамках одного проекта.

В силу общности своей реализации разработанная программа моделирования SyMSim может быть также применена для решения более широкого класса задач проектирования виртуальных центров обработки и хранения больших массивов данных, не ограниченных областью физического эксперимента.

Список литературы

Кореньков В. В., Муравьев А. Н., Нечаевский А. В. Пакеты моделирования облачных инфраструктур // Системный анализ в науке и образовании. — 2014. — Вып. 2. Дубна.

Кореньков В. В., Нечаевский А. В., Трофимов В. В. Разработка имитационной модели сбора и обработки данных экспериментов на ускорительном комплексе НИКА // Информационные технологии и вычислительные системы. — 2013. — No. 4. — С. 37–44.

Попков Ю. С. Макросистемы и grid-технологии: моделирование динамических стохастических сетей // Проблемы управления. — 2003. — № 3.

Сисакян А. Н., Сорин А. С. Многоцелевой Детектор – MPD для изучения столкновений тяжелых ионов на ускорителе NICA (Концептуальный дизайн-проект), версия 1.4. [электронный ресурс] — 2011. — URL: http://nica.jinr.ru/files/CDR_MPD/MPD_CDR_ru.pdf (дата обращения: 02.02.2015).

ATLAS detector [электронный ресурс] // CERN, Switzerland. — 2014. — URL: http://home.web.cern.ch/about/experiments/atlas (дата обращения: 19.01.2015).

BESIII — веб-портал проекта [электронный ресурс] // Beijing, China. — 2014. — URL: http://bes3.ihep.ac.cn/ (дата обращения: 12.01.2015).

CMS detector [электронный ресурс] // CERN, Switzerland. — 2014. — URL: http://home.web.cern.ch/about/experiments/cms (дата обращения: 17.01.2015).

dCache — веб-портал проекта [электронный ресурс] // URL: http://www.dcache.org (дата обращения: 26.01.2015).

GridSim: A Grid Simulation Toolkit For Resource Modelling And Application Scheduling For Parallel And Distributed Computing [электронный ресурс] // The University of Melbourne, Australia. — 2015. — URL: http://www.gridbus.org/gridsim (дата обращения: 06.01.2015).

IBM System Storage TS3500 Tape Library [электронный ресурс] // IBM. — 2014. — URL: http://www.ibm.com/ru/servers/storage/tape/ts3500 (дата обращения: 16.01.2015).

Klusacek D., Matyska L., and Rudova H. Alea — Grid scheduling simulation environment // In 7th International Conference on Parallel Processing and Applied Mathematics (PPAM 2007). Vol. 4967 of LNCS, pages 1029–1038. Springer, 2008.

Nechaevskiy A. V., *Korenkov V. V.* DataGrid simulation packages // System Analysis in Science and Education (Online), ISSN: 2071-9612, Issue 1, 2009.

PANDA — веб-портал проекта [электронный ресурс] // Darmstadt, Germany. — 2014. — URL: http://www-panda.gsi.de/ (дата обращения: 15.01.2015).