

Моделирование грид и облачных сервисов как средство повышения эффективности их разработки

© В.В. Кореньков

© Д.И. Пряхина

korenkov@jinr.ru

© А.В. Нечаевский

© В.В. Трофимов

Объединенный институт ядерных исследований,
Дубна

nechav@jinr.ru

© Г.А. Ососков

© А.В. Ужинский

ososkov@jinr.ru

Аннотация

После введения, показывающего актуальность сетей распределенных вычислений, основанной на грид-облачных структурах в эпоху Больших Данных, описана новая система моделирования грид и облачных сервисов, разрабатываемая в ЛИТ ОИЯИ, ориентированная на повышения эффективности разработки путем учета качества работы уже функционирующей системы при проектировании ее дальнейшего развития за счет объединения самой 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 использована вне рамок физического эксперимента, например, как хранилище информации общего доступа типа электронной библиотеки.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 14-07-00215.

1 Введение

Впечатляющий прогресс компьютерных технологий, программных средств и взрывное развитие глобального информационного пространства, возникшего с появлением Интернета, объединившего между собой компьютерные сети во всемирную систему передачи информации с помощью информационно-вычислительных ресурсов, — все это ознаменовало вступление человечества в новую эру Больших Данных (Big Data).

Все окружающие нас процессы в технической, научной и, особенно, в социальной сферах

постоянно обрушивают на нас непрерывный поток информации, идущий из наших компьютеров и мобильных телефонов, передач различных масс медиа, регулярно перебиваемых назойливой рекламой, различных сенсорных устройств, GPS навигаторов и множества других источников, среди которых, в частности, по оценке Д. Акста, автора книги о будущем электронных библиотек, ожидается, что годовое поступление всех мировых источников цифровой информации: оцифрованных книг, фильмов, оптических и магнитных записей потребует порядка полутора петабайт хранения [1].

Говоря о «больших данных», надо понимать, что это не просто «очень много данных» и, кроме объема, следует учитывать и другие их характеристики. Еще в 2001 г. Мета Групп [2] ввела в качестве определяющих характеристик для больших данных так называемые «*три V*»:

1. **объем** (*volume*, в смысле величины физического объема),

2. **скорость** (*velocity* в смыслах как скорости прироста, так и необходимости высокоскоростной обработки и получения результатов),

3. **многообразие** (*variety*, в смысле возможности одновременной обработки различных типов структурированных и неструктурированных данных

Однако, когда общий поток данных растет экспоненциально, удваиваясь каждый год, за счет революционных технологических изменений, к 2014 г. даже эту «3V» модель предлагают расширить, добавляя новые и новые «V», включая Validity (обоснованность, применимость), Veracity (достоверность), Value (ценность, полезность), и Visibility (обозримость, способность к визуализации) и т.д. [3, 4].

Часто Большие Данные определяют проще, как такие, которые слишком велики и сложны, чтобы их можно было эффективно запомнить, передать и проанализировать стандартными средствами доступных баз данных и иных имеющихся систем хранения, передачи и обработки. В то же время эти Большие Данные, непрерывно поступающие из множества вышеперечисленных источников, должны быть доступны для поисковых систем, проанализированы в центрах бизнеса, производства, медицины, правоохранения, обороны, науки и просто индивидуумов, которые их могут затребовать.

Заметим, что влиянию результатов подобного анализа теперь подвержены все мы, в том числе и люди, никак не связанные с компьютером и интернетом. Дело в том, что информация о любом нашем обращении в полицию, финансовые налоговые, медицинские учреждения, использование банковских карт и бонусов сетевых магазинов оседает в соответствующих базах данных и может быть оттуда извлечена по заказу компетентных органов, а также различных поисковых сетевых служб или, наконец, недобросовестными хакерами. Много примеров подобного использования Больших Данных можно найти в популярном ролике А. Сербанта [5]. Обстоятельный обзор революционных изменений, которые вносят Большие Данные в современное общество дан в бестселлере В. Майер-Шенбергера и К. Кукьера [6].

Однако чаще в качестве одного из наиболее впечатляющих примеров больших данных приводят потоки экспериментальных данных физики высоких энергий, поступающие с Большого Адронного Коллайдера (БАК) в ЦЕРНе [7]. За время первого запуска БАК к 2012 г. четыре экспериментальные установки на нем ALICE, ATLAS, CMS и LHCb выдавали каждую секунду один петабайт (10^{15} байт) данных. Запомнить такое количество данных невозможно ни на какой из современных вычислительных систем. Поэтому после

сверхбыстрой сложной электронной преобработки, оставившей только одно полезное физическое событие из 10 тысяч, выполнялся их анализ в ЦЕРНовском компьютерном центре обработки из многих тысяч процессоров. После этого анализа оставался только 1% событий, возможно содержащих искомый физический феномен. Но даже после такого радикального сокращения потока экспериментальных данных в миллион раз для этих четырех больших экспериментов требовалось хранить поступающие в год 25 петабайт данных в специальных роботизированных ленточных хранилищах, т.к. копии этих данных подлежат передаче в сотни физических центров в 36 странах мира для более тщательного анализа на сотнях тысяч компьютеров, объединенных во Всемирную сеть распределенных вычислений — Worldwide LHC Computing Grid (WLCG). Ежедневно в WLCG обрабатываются полтора миллиона заданий, на что одному даже самому мощному современному компьютеру потребовалось бы 600 лет.

Сравнительная диаграмма на рисунке 1 по общим объемам перерабатываемых в 2012 г. данных в социальных сетях, поисковых системах, разных отраслях бизнеса, медицины, климатических прогнозов и БАК наглядно показывает, что исследования в ЦЕРНе идут в условиях Больших Данных.

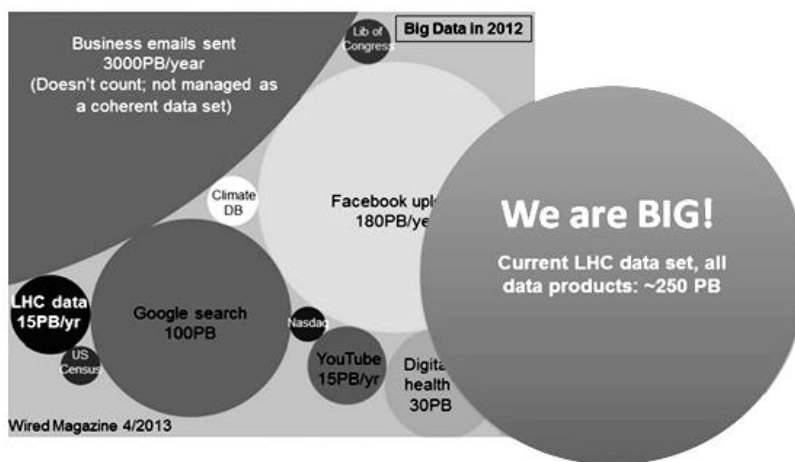


Рис. 1. Сравнительная диаграмма потоков данных в 2012 г. для разных приложений (бизнес, поисковые системы, БАК, медицина, климатические прогнозы) [8]

Известный специалист в области хранения информации Джим Грэй предсказал, что нарастающий поток научной информации должен неизбежно преобразовать практику науки, и назвал это изменение «четвертой парадигмой», в дополнение к трем предыдущим научным парадигмам – экспериментальной, теоретической и вычислительной [9].

На торжестве 4 июля 2012 г. по поводу получения ЦЕРНом нобелевской премии за открытие бозона Хиггса директор ЦЕРНа Рольф Хойер прямо назвал грид-технологии одним из трех столпов успеха (наряду с ускорителем ЛHC и физическими установками). Этот успех также

подтверждает, что ЦЕРН, войдя в эру Больших Данных, эффективно преодолевает проблемы четвертой парадигмы, что является одним из примеров (наряду с созданием в ЦЕРНе WWW-всемирной паутины), когда разработки в области физики частиц начинают влиять на исследования в других научных областях.

Следует отметить, что в 2015 г. предстоит вторичный запуск БАК после его существенной модернизации, когда поток данных возрастет в 2,5 раза при удвоении времени на их обработку. В своих планах на такое развитие компьютеринга в ЦЕРНе после 2015 г., которое смогло бы обеспечить потенциально новую физику, помимо значительного

увеличения вычислительных мощностей и ресурсов хранения данных, совершенствования программных средств анализа и моделирования, активного использования распределенных параллельных вычислений, предлагается повысить эффективность распределенных систем вычислений на базе WLCG путем синтеза грид и облачных технологий [8].

Примером уже имеющейся технологии, реализующей подобный синтез работы с Большими Данными является система PanDA (Production and Distributed Analysis – обработка данных и распределенный анализ) эксперимента ATLAS на LHC. Сегодня PanDA рассматривается как возможная система для российского мегапроекта НИКА в Объединенном институте ядерных исследований (г. Дубна) [10].

Как было показано в недавно защищенной в ОИЯИ диссертации Н.А. Кутовского, включение в грид с его жесткой структурой, ориентированной на интеграцию уже существующих процессорных и программных ресурсов, облачных структур, более гибких за счет использования виртуальных кластеров из виртуальных вычислителей, позволяет сократить время решения широкого круга задач в области физики высоких энергий и повысить эффективность использования ресурсов [11].

Следует подчеркнуть, что в силу своей сложности и высокой стоимости разработка таких сложнейших грид-облачных систем сбора, передачи и распределенной обработки сверхбольших объемов информации требует больших предварительных исследований по выбору оптимальной их структуры с учетом стоимости и предполагаемых ресурсов и загрузки. Подобные исследования должны основываться на тщательном моделировании как потока заданий с учетом их типов и статистических данных о распределении времени их поступления и требуемых компьютерных ресурсов для их выполнения, так и состава моделируемой грид-структуры.

Настоящая работа посвящена разрабатываемой в ЛИТ ОИЯИ новой системе моделирования грид и облачных сервисов, ориентированной на повышения эффективности их разработки путем учета качества работы уже функционирующей системы в прогнозах на ее дальнейшее развитие. Это предлагается сделать за счет объединения самой программы моделирования с системой мониторинга реального (или модельного) грид-облачного сервиса через специальную базу данных, осуществляющую сбор и статистический анализ по вычислению распределений данных мониторинга, используемых затем для динамической коррекции параметров моделирования. Приведен пример применения программы для моделирования достаточно общей облачной структуры, которая может быть использована вне рамок физического эксперимента, например, как хранилище информации общего доступа типа электронной библиотеки.

2 Принципы моделирования грид и облачных инфраструктур

В силу очевидной необходимости предварительного моделирования грид-систем были написаны несколько программ для их моделирования, обзор которых можно найти в работе [12], где после сравнительного анализа этих программ был сделан выбор в пользу программы грид-моделирования GridSim [13]. GridSim – это библиотека классов, построенная на стандартной библиотеке SimJava, с помощью которой можно моделировать поток дискретных событий во времени.

В предыдущей работе авторов [14] описана программа моделирования, основанная на расширении классов GridSim и объединении их в программу, которая моделирует обработку потока заданий проектируемой грид-структурой, и алгоритмов планирования потока заданий ALEA [15]. 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 эту грид-структуру. В качестве результатов вычисляются временные оценки искомых параметров потока заданий.

Как уже было отмечено выше, постоянное развитие современных грид-систем требует непрерывных корректировок большинства параметров моделирования. Это необходимо для прогнозирования поведения системы при значительных ее изменениях с учетом статистики эксплуатации системы, получаемой на основе имеющихся программных средств ее мониторинга.

3 Мониторинг грид-систем

Система мониторинга – это набор программных и аппаратных средств для анализа и контроля состояния некоторой системы распределенных вычислений (см. например [16]). Система мониторинга и учета ресурсов (СМУР) предназначена для отслеживания текущего состояния ресурсов, заданий и других объектов в грид-системе. Инструментарий СМУР должен предоставлять как статическую, так и динамическую информацию о функционировании грид-системы (примером динамической информации может служить состояние очередей на вычислительном кластере), а также результаты статистического анализа этой информации. Среди основных задач мониторинга отметим следующие:

- Непрерывное наблюдение за состоянием грид-сервисов, как базовых (общих для всей инфраструктуры), так и относящихся к отдельным ресурсным центрам;
- Получение информации о вычислительных ресурсах (количество вычислительных узлов для

выполнения задач, архитектура вычислительной системы, установленное программное обеспечение, доступные специализированные программные пакеты) и о потребленном процессорном времени;

- Данные о доступе виртуальных организаций к ресурсам и использовании ими квот на вычислительные ресурсы;

- Мониторинг выполнения вычислительных заданий и задач (запуск, изменение состояния, коды завершения и т.п.).

Среди параметров мониторинга, необходимых для последующего моделирования, наиболее существенными являлись следующие:

- 1) число задач (симуляция, анализ, реконструкция) поступающих в систему;
- 2) объем используемой оперативной памяти;
- 3) использованное процессорное время;
- 4) число обработанных событий;
- 5) время расчета задачи;
- 6) объем используемых данных.

Примеры диаграмм, показывающих динамику изменения параметров мониторинга показаны на рисунках 2 и 3.

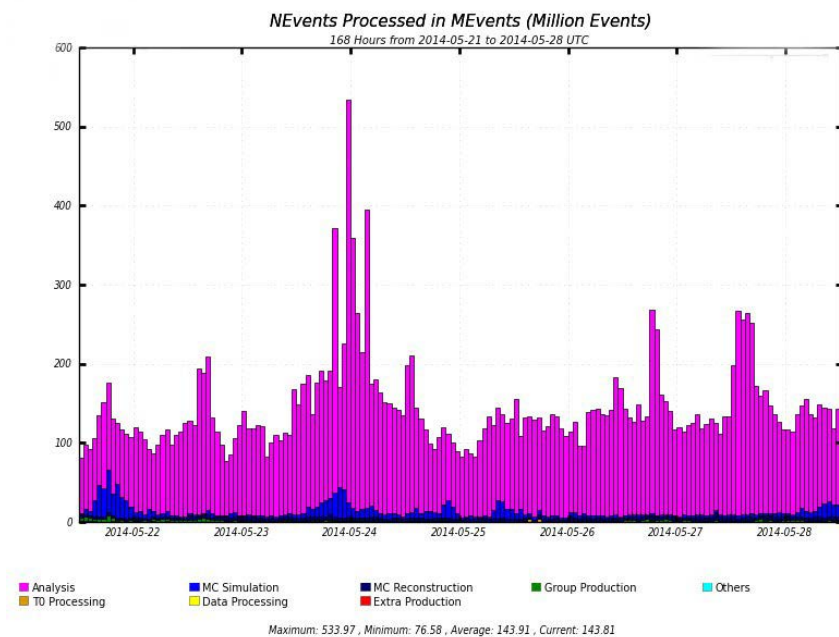


Рис. 2. Динамика распределения числа физических событий (в млн) по группам решаемых задач (анализ, симуляция, обработка)

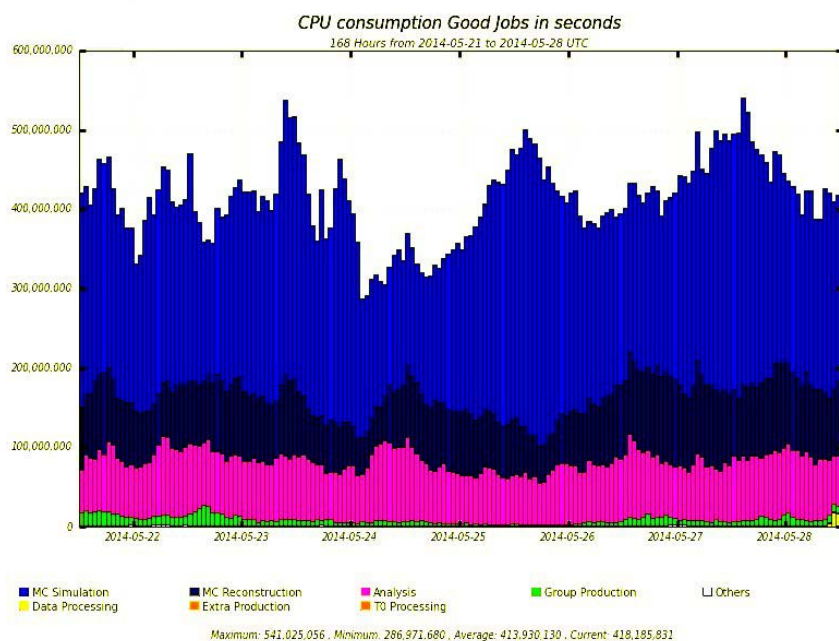


Рис. 3. Динамика распределения загрузки процессоров (число выполняемых заданий в секунду) по разным типам решаемых задач

4 Схема программы моделирования SyMSim

В настоящей работе на основе данных мониторинга одной из грид-систем WLCG [7], сохраняемых в специально разработанной базе данных, через веб-интерфейс выполняется их статистический анализ, результаты которого позволяют затем генерировать адекватный поток заданий для изменения параметров моделирования.

Схема программы SyMSim (Synthesis of Monitoring and Simulation – Синтез Мониторинга и Моделирования), реализующей идею синтеза процессов мониторинга и моделирования, представлена на рисунке 4. Данные мониторинга реальной грид-системы поступают в базу данных следующим образом: задание отправляется на сервер (1), далее сайт-пилот запрашивает задание на обработку (2), сервер отправляет задание на исполнение (3), информация о выполнении задания поступает в базу данных StatDB (4). На основе

данных мониторинга пользователь задает входные параметры модели (5) и потока заданий (6), далее модель обрабатывает задания (7,8). Результаты работы модели доступны пользователю для дальнейшего анализа.

При разработке программы SyMSim удалось расширить ее сферу применения на ставшие весьма актуальными в последнее время системы облачных вычислений.

Для иллюстрации возможностей разработанной программы SyMSim по имитационному моделированию облачных вычислений ниже приведен пример ее применения для оптимизации простой облачной структуры. Эта структура мыслилась, как предназначенная для обработки данных физического эксперимента, но в принципе в такую схему укладываются и другие структуры связанные с хранением и обновлением больших массивов цифровой информации, в том числе и возможный вариант реализации электронной библиотеки.

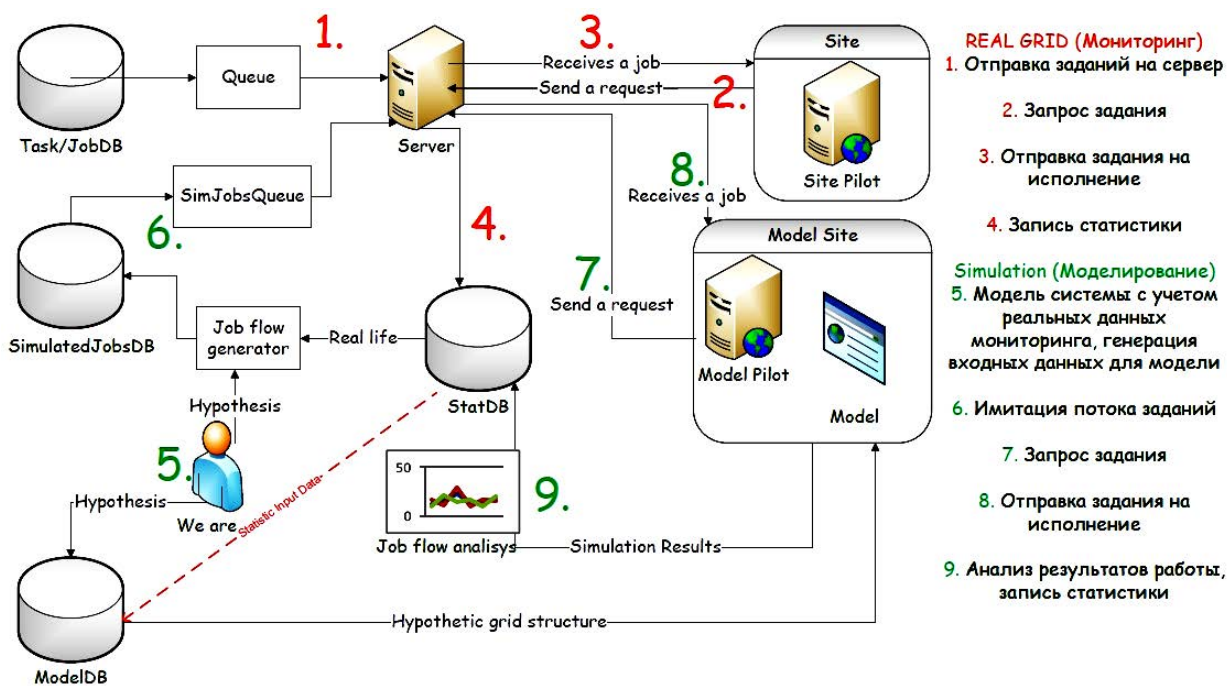


Рис. 4. Схема программы SyMSim моделирования системы распределенных вычислений с учетом данных мониторинга этой системы

5 Постановка задачи и результаты моделирования

Рассматривается модель реализации облачной структуры, предназначенной для хранения данных в роботизированной библиотеке с тысячами кассет с магнитными лентами, из загрузчиков-драйвов которых робот автоматически достаёт требуемые ленты и устанавливает в одно или несколько устройств чтения-записи. Достижения современных

технологий позволяют записать до 6,25 терабайт на один ленточный картридж, причем современные стандарты записи на магнитные ленты под углом поддерживают деление ее на разделы подобно дискам, что позволяет обращаться к ним, как к традиционному каталогу. Фотография такого хранилища с роботом-загрузчиком представлена на рисунке 5.

Проектируемая структура состоит из ленточного робота, массива ленточных картриджей и кластера процессоров. Стоимость драйва – 5 условных единиц, процессора – 3 единицы.



Рис. 5. Фото хранилища с роботом-загрузчиком

В качестве критерия оценки выбирается время прохождения тестового потока из 99 заданий. Надо найти оптимальное соотношение количества процессоров и количества драйвов при ограниченном бюджете в 100 условных единиц.

Результаты моделирования

1. Определение степени загрузки кластера

Загрузка кластера $W = T_{100}/T_a$, где T_{100} – процессорное время выполнения пакета T_a – астрономическое время.

На рисунке 6 показана степень загрузки кластера в зависимости от количества процессоров и драйвов. Мы

видим, что при большом количестве процессоров загрузка кластера падает, поскольку процессоры простаивают в ожидании монтирования кассет с данными на драйвы. Следовательно, надо выбирать оптимальное соотношение.

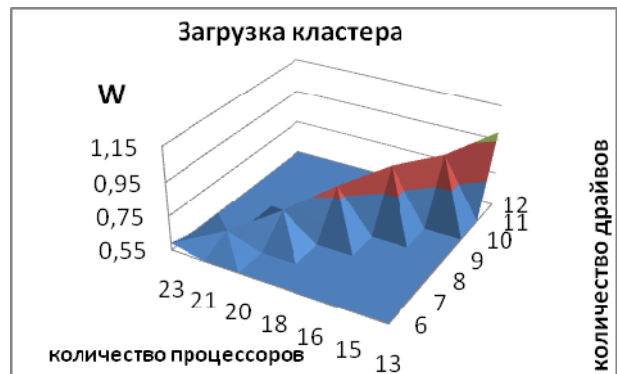


Рис. 6. Степень загрузки кластера в зависимости от количества процессоров и драйвов

2. Время выполнения пакета заданий в зависимости от количества процессоров и драйвов

На рисунке 7 стрелкой показан оптимум по числу вычислительных процессоров в кластере и дорогих драйвов. Таким образом, конфигурация, обеспечивающая минимальное время исполнения должна состоять из 18 вычислительных процессоров и 9 драйвов-загрузчиков.

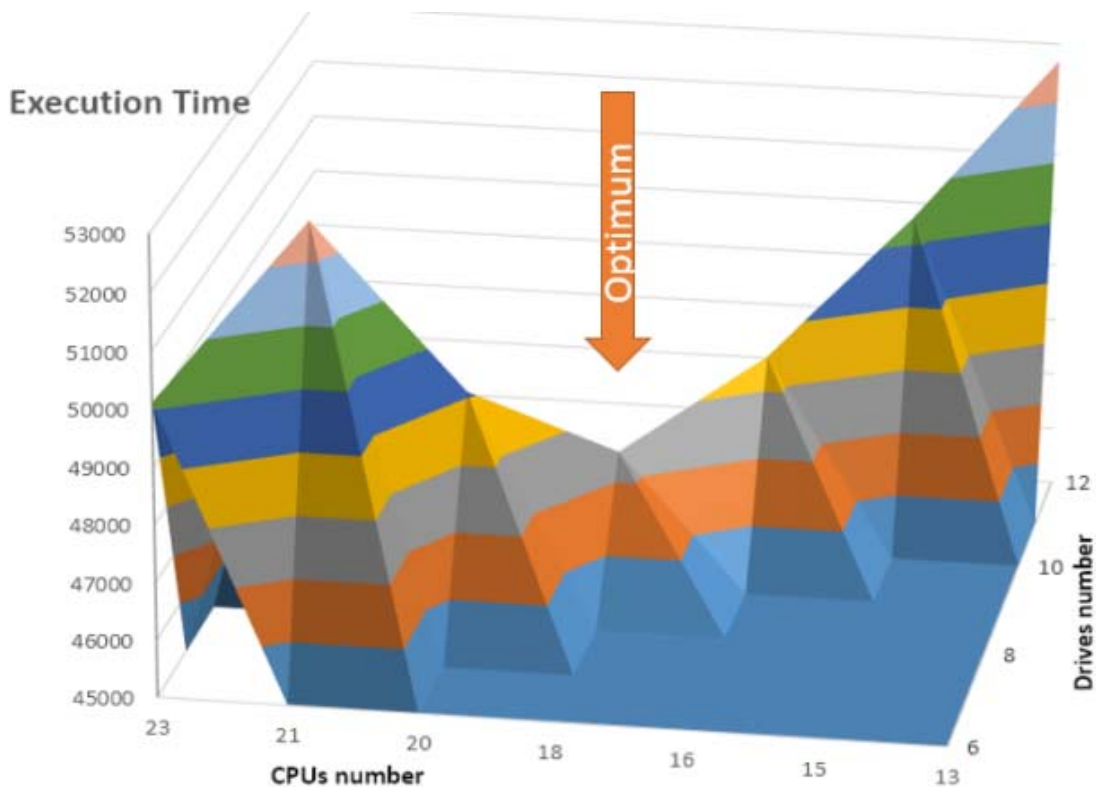


Рис. 7. Время выполнения пакета заданий в зависимости от количества процессоров и драйвов

Заключение

Предложенный подход к моделированию и анализу вычислительных грид-облачных структур в экспериментальной физике высоких энергий основан на учете данных их мониторинга, используемых затем для динамической коррекции параметров моделирования. 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. В частности программу можно применять для проектирования и последующего развития хранилищ информации общего доступа, не ограниченных областью физического эксперимента.

Литература

- [1] Akst, D. (2003). The Digital Library: Its Future Has Arrived. *Carnegie Reporter*, 2(3), 4–8.
- [2] Doug Laney Meta Group: <http://blogs.gartner.com/douglaney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
- [3] Rob Livingstone Advisory: <http://roblivingstone.com/2013/06/big-data-or-black-hole/>
- [4] Timo Elliott: 7 Definitions of Big Data You Should Know About <http://timoelliott.com/blog/2013/07/7-definitions-of-big-data-you-should-know-about.html>
- [5] А. Себрант. Что такое Big data: <http://www.slideshare.net/yandex/big-data-30799013>
- [6] В. Майер-Шенбергер, К. Кукьер. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. Издательство: Манн, Иванов и Фербер, 2014.
- [7] Компьютинг в ЦЕРНе: <http://home.web.cern.ch/about/computing>
- [8] José M. Hernández (CIEMAT, Madrid) On behalf of the Spanish LHC Computing community, Perspectives on LHC Computing, Jornadas CPAN 2013, Santiago de Compostela.
- [9] Jim Gray et al, Scientific Data Management in the Coming Decade, *SIGMOD Record*, Vol. 34, No. 4, Dec. 2005.
- [10] А. Климентов, А. Ваняшин, В. Кореньков. За большими данными следит ПАНДА // *Суперкомпьютеры*. – 2013. – №15. – С. 56.
- [11] Кутовский Н.А. Развитие методов построения грид-сред и систем облачных вычислений для задач физики высоких энергий: дис. ... канд. физ-мат. наук. – Дубна: ОИЯИ, 2014.
- [12] А.В. Нечаевский, В.В. Кореньков. Пакеты моделирования DataGrid // *Системный анализ в науке и образовании: электронный журн.* – 2009. – № 1.
- [13] GridSim web-portal: <http://www.gridbus.org/gridsim/>
- [14] В.В. Кореньков, А.В. Нечаевский, В.В. Трофимов. Разработка имитационной модели сбора и обработки данных экспериментов на ускорительном комплексе НИКА // *Информационные технологии и вычислительные системы*. – 2013. – № 4. – С. 37–44.
- [15] D. Klusacek, L. Matyska, and H. Rudova. Alea – Grid scheduling simulation environment // *7th International Conference on Parallel Processing and Applied Mathematics (PPAM 2007)*. – Springer, 2008. – Vol. 4967 of LNCS. – P. 1029–1038.
- [16] Грид в ОИЯИ, Мониторинг и аккаунтинг. Веб-портал: http://grid.jinr.ru/?page_id=77

Simulation of Grid and Cloud Services as the Means of the Efficiency Improvement of Their Development

Vladimir V. Korenkov, Andrey V. Nechaevskiy, Gennadiy A. Ososkov, Dariya I. Pryahina, Vladimir V. Trofimov, Aleksandr V. Uzhinskiy

After the introduction, showing the relevance of the distributed computing networks based on grid/cloud structures at the Big Data era, a new grid and cloud services simulation system are described. This system is developed at the LIT JINR Dubna and focused on 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 this purpose the simulation program is combined with real monitoring system of the grid/cloud service through a special database. An example of the program usage to simulate a sufficiently general cloud structure, which can be used, e.g., as a repository for information sharing, such as the digital library, is given.