

УДК 004.75

СОЗДАНИЕ СИСТЕМЫ РАСПРЕДЕЛЁННЫХ ВЫЧИСЛЕНИЙ

М. И. Гарапов

Камчатский государственный университет имени Витуса Беринга, 683032,
г. Петропавловск-Камчатский, ул. Пограничная, 4
E-mail: as2021@mail.ru

В данной статье рассмотрена проблема использования систем распределённых вычислений для решения глобальных проектов имитационного моделирования и математических вычислений. Так же на историческом примере рассмотрено использование системы распределённых вычислений BOINC на практике. Выявлены проблемы практической реализации проектов распределённых вычислений и пути их решения. Создан сервер распределённых вычислений и тестовое задание к нему. Реализована система сетевых служб и взаимодействия пользователей проекта с базой данных.

Ключевые слова: распределённые вычисления, BOINC, математическое моделирование, Debian, Boinc-server, информационная система

© Гарапов М. И., 2018

MSC 90-04

CREATION OF THE DISTRIBUTED COMPUTING SYSTEM

M. I. Garapov

Vitus Bering Kamchatka State University, 683031, Petropavlovsk-Kamchatsky,
Pogranichnaya st., 4, Russia
E-mail: as2021@mail.ru

In this article, the problem of using distributed computing systems for solving global projects of simulation simulation and mathematical calculations is considered. Also, in the historical example, the use of BOINC distributed computing in practice is considered. The problems of practical implementation of distributed computing projects and ways to solve them are identified. A distributed computing server and a test task to it have been created. Implemented a system of network services and interaction of project users with the database.

Key words: distributed computing, BOINC, mathematical modeling, Debian, Boinc-server, information system.

© Garapov M. A., 2018

Введение в добровольные распределенные вычисления

Очевидно, что с появлением вычислительной техники и тотальным совершенствованием средств передачи информации – лицо науки изменилось. Сотрудничество университетов в реальном времени, передача результатов опытных исследований, онлайн-конференции и информационные ресурсы существенно сократили сроки, как разработки различных проектов, так и проверки экспериментальных данных. Но развитие информационных технологий, а так же мощности вычислительной техники принесло в мир науки и другое новшество – компьютерные методы фильтрации и расшифровки информации, а так же методы математического и имитационного моделирования. Они прочно укрепились в разных научных сферах, таких как физика, астрономия, механика, химия, биология, экономика, социология, метеорология, других науках. А так же и в прикладных задачах различных областей: в радиоэлектронике и машиностроении. Росли возможности и вместе с возможностями росли и масштабы научных изысканий.

В 1994 год в Обсерватории Аресибо в Пуэрто-Рико загружена задачей по расшифровке сигналов радиотелескопов [3]. Узкополосный радиосигнал, с периодически повторяющимися элементами - уже сильный довод в пользу существования внеземной жизни, которая использует радиосвязь [1]. Однако такая задача требует использование и серьёзных вычислительных мощностей – для расшифровки радиосигнала. Такой проект требовал специализированные суперкомпьютеры, что сильно снижало его доступность. Тогда был сделан первый шаг к становлению распределённых вычислений инструментом науки - Дэвид Геди, ученый университета в Беркли предложил использовать для вычислений кластер из множества простых компьютеров обычных пользователей с подключением к сети интернет. Так, уже в 1999 году стартовал проект SETI@home. Проект добровольных распределённых вычислений, он давал обычным людям, имеющим компьютер с подключением к сети интернет жертвовать свои вычислительные мощности на благо науки, а так же минимизировать время простоя ЭВМ. Это был первый в мире проект такого рода и он отлично себя зарекомендовал – количество добровольцев, желающих принять участие в проекте, увеличивалось из года в год. Выводы, полученные из успеха проекта были сделаны. Так, в 2005 году стартовал проект BOINC - Berkeley Open Infrastructure for Network Computing [2]. Это некоммерческое межплатформенное ПО для организации распределённых вычислений которое стало мощным толчком для появления других проектов добровольных научных вычислений, требующих большие вычислительные мощности.

Общая информация о платформе BOINC

В настоящий момент очень сложно переоценить роль проектов распределённых вычислений BOINC. Благодаря открытости платформы многие научные институты и университеты создают BOINC проекты и получают так необходимые для их реализации вычислительные мощности. Под стать поставленным задачам, требующим сложные имитационные и вычислительные модели. Ведь теперь чтобы стать участником одного или нескольких проектов достаточно просто скачать BOINC-клиент и зарегистрироваться в понравившемся проекте. Например:

- Acoustics @ home - это проект добровольных вычислений на основе BOINC, предназначенный для решения обратных задач в подводной акустике.
- Gerasim @ home - исследований по дискретной математике и логическому управлению. Текущая цель: тестирование и сравнение эвристических методов для получения разделов параллельных алгоритмов в САД-системе для проектирования логических систем управления.
- LHC@Home — проект CERN по моделированию поведения пучка заряженных частиц при различных параметрах воздействия на них управляющих магнитов Большой адронный коллайдера.
- Rosetta@home — вычисление 3-мерной структуры белков из их аминокислотных последовательностей.
- Asteroids@home — проект обрабатывающий поток фотометрических данных для исследования траекторий движения и размеров приближающихся к Земле астероидов.

Нельзя не отметить, что большинство существующих проектов являются международными [4]. Однако, оценить вклад той или иной страны в BOINC-проекты всё же можно (рис. 1).

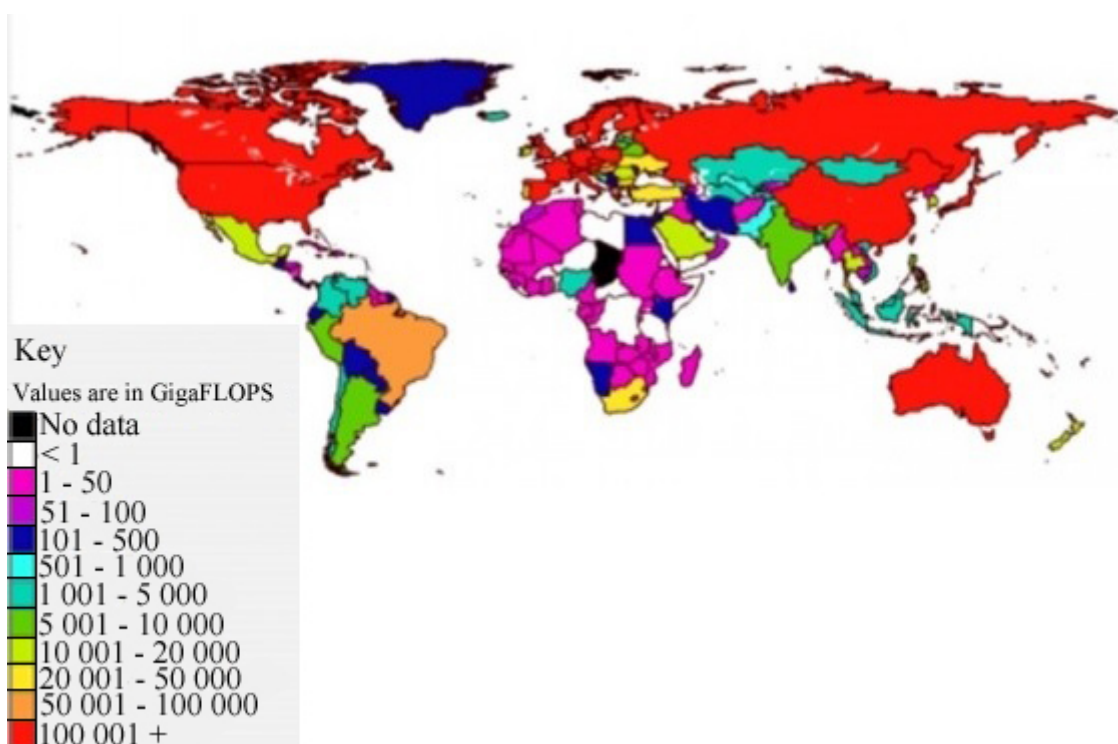


Рис. 1. Суммарная производительность узлов сети на платформе BOINC в мире.

Подобная карта наглядно демонстрирует, что даже при дефиците государственной поддержки науки – вклад её граждан в мировой прогресс может быть заметен и даже существенен.

Структура VOINC проектов

Ввиду того, что платформа VOINC распространяется по открытой лицензии LGPL найти общие и технические сведения об этой платформе может любой желающий. И если рассматривать VOINC изнутри, то первое, на что следует обратить внимание – это схема взаимодействия VOINC-клиентов и самого сервера, который генерирует задания и контролирует их выполнение.

Взаимодействие клиентской и серверной части в рамках одного проекта представлены далее (рис. 2).

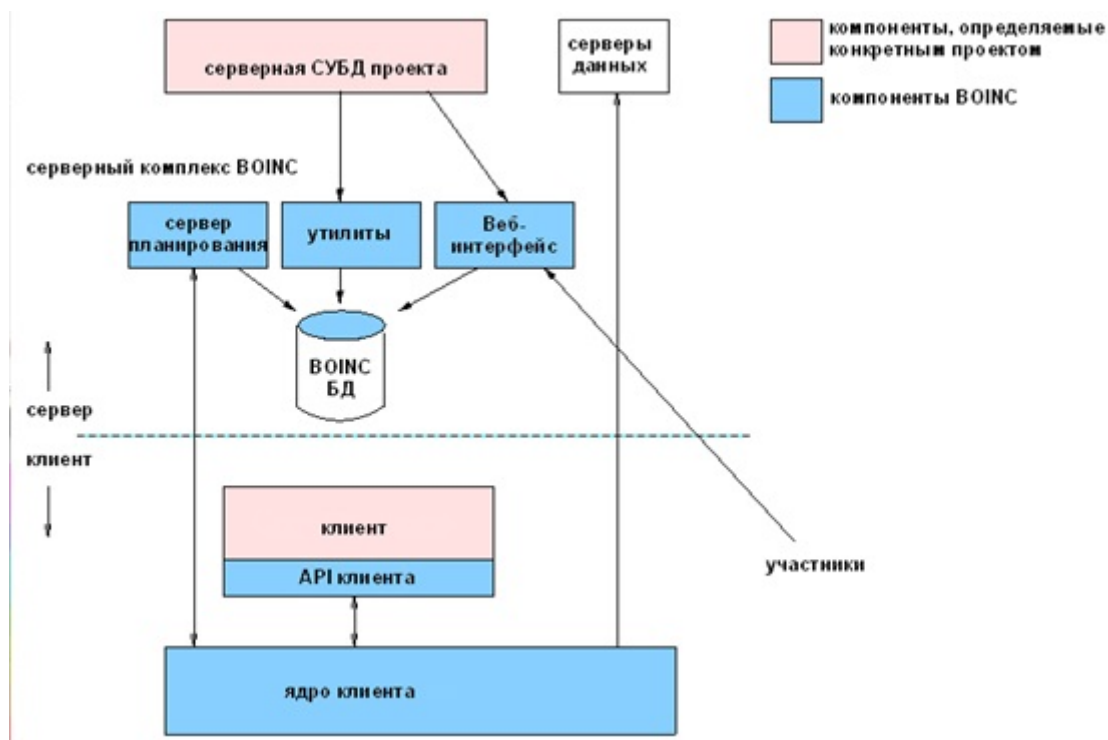


Рис. 2. Схема взаимодействия клиентской и серверной части

Модель накопления данных VOINC основана на файлах. Файлами являются входные и выходные данные приложений и запускаемые программы.

Ядро VOINC клиента передает файлы серверам данных, управляемыми проектом и получает их от них же, используя HTTP.

Для развертывания сервера требуется операционная система Unix/Linux, а также необходимы внешние вспомогательные программы. СУБД MySQL, Web-сервер Apache, интерпретаторы и стандартные библиотеки языков python и php. Объяснение такому разнообразию необходимых в том, что Web-сервер – это составная часть сервера VOINC, а база данных несет на себе огромную функциональность.

Создание сервера распределённых вычислений

Первая задача, которая встаёт при создании сервера - это установка на операционную систему всех дополнительных программных пакетов, то есть решение внешних зависимостей для ПО сервера VOINC. Далее приведён список пакетов, требуемых для запуска сервера (рис. 3).

	Server	Client	Manager	Applications	Graphic apps
⇒ git is needed to obtain the BOINC source code	X	X	X	X	X
⇒ GNU tools: make 3.79+, m4 1.4+, libtool 1.5+, autoconf 2.58+, automake 1.8+, GCC 3.0.4+	X	X	X	X	X
⇒ pkg-config 0.15+	X	X	X	X	X
Python 2.2+ with ⇒ MySQLdb module 0.9.2+ (see installation instructions) and xml module	X				
MySQL 4.0.9 or higher (with mysql-dev(el), and mysql-client)	X				
sqlite 3.1 or higher (packages sqlite-dev(el) and sqlite)			X		
libnotify-dev			X		
Apache with mod_ssl and PHP5+	X				
⇒ PHP5 with ⇒ cli support and the ⇒ GD and MySQL modules (packages php5-cli and php5-gd)	X				
⇒ OpenSSL version 0.9.8+	X	X			
libcurl-dev(el)	X				

Рис. 3. Требуемое для работы сервера ПО

Очевидно, что создание BOINC-сервера является комплексной задачей, реализация которой требует не только существенных затрат времени и определённых знаний, а так же и актуальной справочной информации, охватывающей все структурные единицы проекта. Однако ресурсы от разработчиков платформы часто и во все не содержат инструкций к решению ошибок а сторонние ресурсы часто разрознены и несут в себе устаревшую информацию. Для решения поставленной задачи в условиях недостатка данных требуется инструмент, позволяющий собрать и систематизировать данные из внешних источников, а так же дополнять информацией о выполненных действиях для решения ошибок возникающих как при работе с приложением boinc-сервера, так и с вспомогательным ПО.

В данном случае была использована технология Mind-Mapping [5]. Руководствуясь данной технологией была создана информационная система в свободно-распространяемом приложении Free Mind (рис. 4).

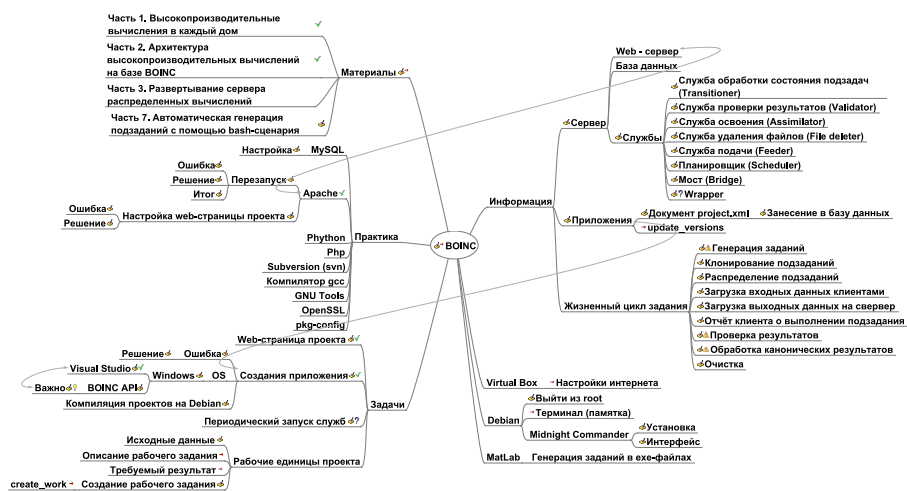


Рис. 4. Интеллект-карта проекта

Впоследствии данная информационная система может быть использована как для создания общей справочной системы по реализации и других BOINC-проектов, так и напрямую, для решения частных задач и ошибок в ходе реализации и расширения данного проекта.

После установки серверного приложения и удовлетворения его внешних зависимостей к ПО с устранением ошибок взаимодействие с сервером осуществляется с помощью системы скриптов и взаимодействием с базой данных MySQL.

Запуск сервера распределённых вычислений

Заключительными задачами реализации сервера распределённых вычислений являются следующие:

- Создание web-сервиса.
- Генерация заданий для разных клиентских операционных систем.

Создание web-сервиса осуществляется с помощью HTTP-сервера Apache2 и скриптовых команд BOINC-Server. В итоге выполнения команд генерируется редактируемая web-страница *.php по данному шаблону (рис. 5).

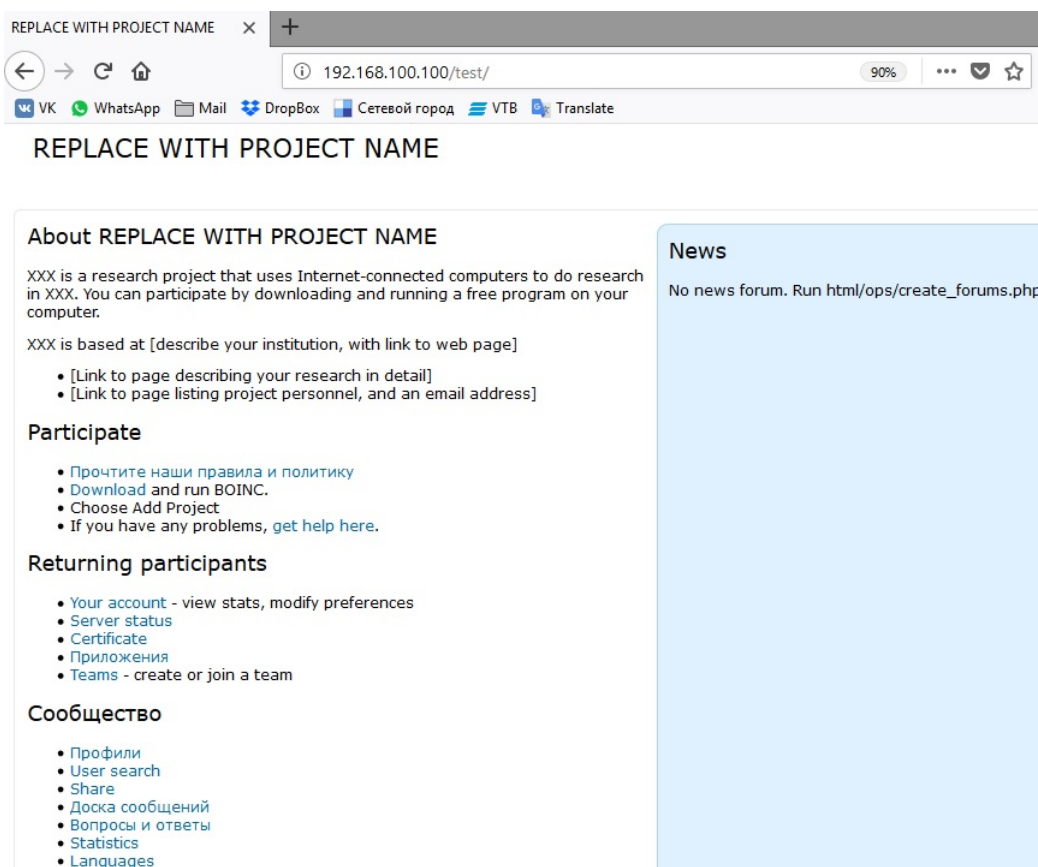
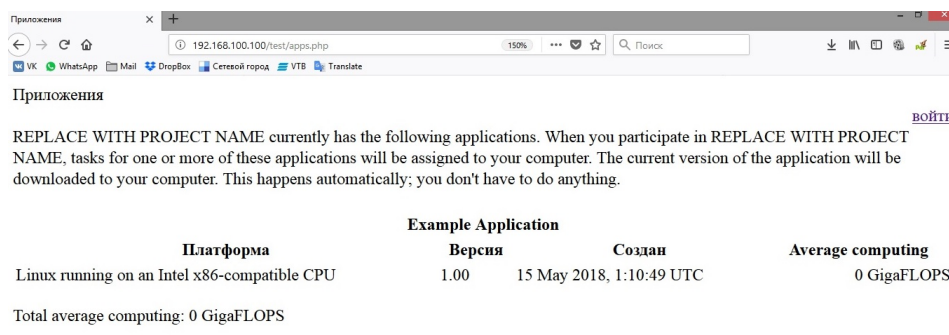


Рис. 5. Главная web-страница проекта

Для того, чтобы сервер начал распределение заданий проекта пользователю необходимо выполнить последовательность действий:

- Скомпилировать для целевой ОС пользователя файл задание, используя при этом boinc-api. Для программирования заданий проектов используется язык C++.
- Разместить задание в соответствующей папке проекта по пути со строго определёнными названиями папок (название задания, версия, целевая платформа, файлы задания).
- Добавить информацию о задании в базу данных. Для этого используется вспомогательный файл boinc-сервера - 'project.xml', а также скрипты: 'xadd' и 'update versions'.

Примечательно, что задания проекта могут быть доступны для выполнения у широкого спектра пользователей (существует множество клиентов BOINC под разные операционные системы). Для реализации возможности использовать их в сети распределённых вычислений необходимо скомпилировать задание под целевую ОС (Android, MacOS, Windows, Linux и т.д.). Наличие заданий для конкретной ОС так же можно просмотреть через web-сервис (рис. 6).



Приложения

REPLACE WITH PROJECT NAME currently has the following applications. When you participate in REPLACE WITH PROJECT NAME, tasks for one or more of these applications will be assigned to your computer. The current version of the application will be downloaded to your computer. This happens automatically; you don't have to do anything.

Платформа	Example Application	Версия	Создан	Average computing
Linux running on an Intel x86-compatible CPU		1.00	15 May 2018, 1:10:49 UTC	0 GigaFLOPS

Total average computing: 0 GigaFLOPS

Рис. 6. Доступные задания

Проверить работу сервера можно уже с помощью установленного boinc-менеджера, который должен присутствовать на клиентских компьютерах. Для этого необходимо ввести в строке подключения адрес создаваемого проекта (рис. 7, А).

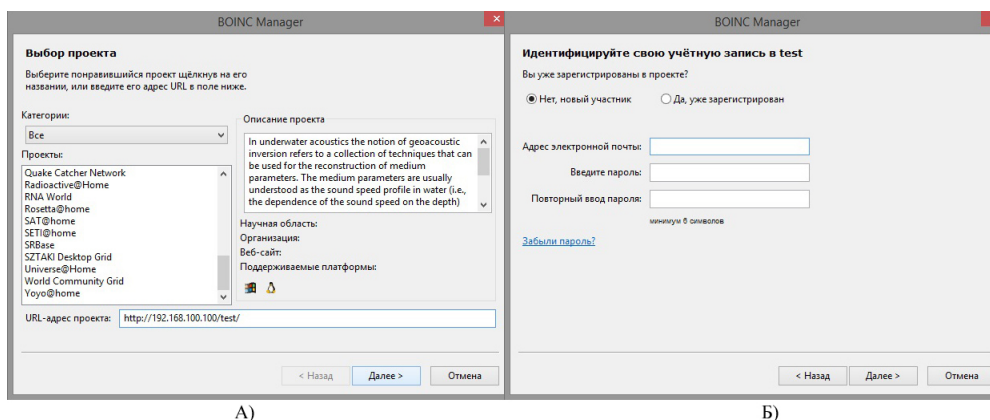


Рис. 7. Подключение к проекту

Далее будет доступно окно регистрации пользователя в проекте: информация о учётной записи, а так же основной информации о компьютере будет внесена

в базу данных MySQL (рис. 7, Б) После этого пользователь примет участие в распределении доступных заданий. Необходимые файлы на пользовательских компьютерах, а так же результаты выполнения программ-заданий скачиваются и передаются на сервер автоматически.

Заключение

Платформа BOINC является мощным инструментом распределённых математических вычислений и её внедрение и использование способно ускорить или вовсе сделать возможными широкий спектр научных изысканий в разных областях, требующих больших вычислительных мощностей. Создание же исчерпывающей справочной системы с разбором всевозможных ошибок, очевидно, актуально и способно подтолкнуть академическую среду к активному участию в научных изысканиях.

Список литературы

- [1] Интернет-ресурс проекта SETI@home, setiathome.berkeley.edu. [*Internet-resurs proekta SETI@home, setiathome.berkeley.edu*].
- [2] Попов С., “Весь мир как суперкомпьютер”, *Троицкий вариант – Наука*, **110**:16 (2012), 7. [Povov S., “Ves'mir kak superkomp'yuter”, *Troickij variant – Nauka*, **110**:16 (2012), 7].
- [3] Осипов Ю. С., *Большая российская энциклопедия*, **35**, 2004. [Osipov YU. S., *Большая российская энциклопедия*, **35**, 2004].
- [4] Heien E., Anderson D., “Computing Low Latency Batches with Unreliable Workers in Volunteer Computing Environments”, *Journal of Grid Computing*, **7** (2009), 501.
- [5] Осипов Ю. С., *Эффективная обработка информации–Mind Mapping для студентов и профессионалов*. Т. 35, 2004. [Osipov YU. S., *Effektivnaya obrabotka informacii–Mind Mapping dlya studentov i professionalov*. V. 35, 2004].

Список литературы (ГОСТ)

- [1] Интернет-ресурс проекта SETI@home setiathome.berkeley.edu
- [2] Попов С. Весь мир как суперкомпьютер // Троицкий вариант – Наука. 2012. Т. 110. №. 16. Р. 7
- [3] Осипов Ю. С. Большая российская энциклопедия // Арсибо, радиоастрономическая обсерватория. 2004. Т. 35
- [4] Heien E., Anderson D. Computing Low Latency Batches with Unreliable Workers in Volunteer Computing Environments // Journal of Grid Computing. 2009. vol 7. Р. 501
- [5] Осипов Ю. С. Эффективная обработка информации–Mind Mapping для студентов и профессионалов. 2004. Т. 35

Для цитирования: Гарапов М. И. Создание системы распределенных вычислений // *Вестник КРАУНЦ. Физ.-мат. науки*. 2018. № 2(22). С. 45-52. DOI: 10.18454/2079-6641-2018-22-2-45-52

For citation: Garapov M. I. Creation of the distributed computing system, *Vestnik KRAUNC. Fiz.-mat. nauki*. 2018, **22**: 2, 45-52. DOI: 10.18454/2079-6641-2018-22-2-45-52