

К.т.н. М.В. Михелев, К.В. Кузнецов,  
к.т.н., доц. В.М. Михелев (НИУ «БелГУ»)

M.V. Mikhelev, K.V. Kuznetsov, V.M. Mikhelev

**ПОСТРОЕНИЕ ИНФОРМАЦИОННОЙ ИНФРАСТРУКТУРЫ  
ВЫСОКОПРОИЗВОДИТЕЛЬНОЙ КОМПЬЮТЕРНОЙ СИСТЕМЫ  
ДЛЯ РЕАЛИЗАЦИИ ОБЛАЧНЫХ ВЫЧИСЛЕНИЙ**

**METHOD OF CONSTRUCTION INFORMATION  
INFRASTRUCTURE HIGH-PERFORMANCE COMPUTER SYSTEM  
FOR THE IMPLEMENTATION OF CLOUD COMPUTING**

*Проведено исследование возможности предоставления пользователям вычислительных ресурсов как услуги в сети Интернет. Оценена целесообразность переноса существующих приложений в облачную среду. Предложен способ построения информационной инфраструктуры для реализации облачных вычислений на факультете КНИТ НИУ БелГУ. Показаны преимущества облачной инфраструктуры в области масштабирования приложений.*

*A study on the possibilities of providing users with computing resources as services on the Internet. Evaluated the usefulness of porting existing applications in a cloud environment. A way to build the information infrastructure for implementing cloud computing at the Faculty of KNIT NIL BelGU. Showing the benefits of cloud infrastructure to scale applications.*

*Ключевые слова: вычислительные ресурсы, облачные вычисления, информационная инфраструктура, Web-приложение.*

*Key word: computing resources, cloud computing, information infrastructure, Web application, parallel computing.*

**Введение**

Эффективность и популярность любого крупного вычислительного комплекса для решения научных и инженерных задач определяется многими факторами. К этим факторам, в первую очередь, относятся широта спектра решаемых задач, приемлемая точность в декларируемом диапазоне определяющих параметров, надёжность и безотказность функционирования. Но наиболее важным фактором является, интерфейс вычислительного комплекса (API), который определяет необходимую простоту пользования этим ре-

сурсом, ясность и однозначность составления вычислительного задания. Качество интерфейса «клиент – ресурс» особенно важно при коммерциализации программных разработок. В настоящее время использование больших вычислительных мощностей затруднено, в первую очередь, из-за необходимости больших финансовых вложений и наличия квалифицированного персонала, способного обслуживать вычислительный комплекс.

Эти проблемы практически полностью снимает применение современного направления информационных технологий – CloudComputing («Облачные вычисления»). Сущность CloudComputing заключается в переносе обработки данных с персональных компьютеров и рабочих станций на серверы Всемирной Сети или вычислительных центров. Таким образом потенциальный пользователь становится не покупателем вычислительных программ и комплексов, а их арендатором, которому предоставляются разнообразные услуги.

### **Теоретические основы**

Облачные вычисления включают в себя следующие технологии [4,5]:

- технология PaaS («Platform as a Service», «Платформа как услуга»);
- технология IaaS («Infrastructure as a Service», «Инфраструктура как услуга»),
- технология SaaS («Software as a Service», «Программное обеспечение как услуга»)

Основой для облаков, или облачных вычислений является именно IAAS, т.е. платформа, которая предоставляет возможности компьютерной инфраструктуры в форме виртуализации. IaaS состоит из трех основных компонентов.

- аппаратные средства (серверы, системы хранения данных, клиентские системы, сетевое оборудование);
- операционные системы и системное ПО (средства виртуализации, автоматизации, основные средства управления ресурсами),
- связующее ПО (например, для управления системами)

Наша задача создать платформу (IAAS), которая будет предоставлять по запросу клиента необходимое количество динамиче-

ских ресурсов (вычислительных и хранилища), виртуальных серверов, сетевой инфраструктуры, удаленных рабочих мест на основе концепции облачных вычислений. При этом платформа должна позволять максимально оптимизировать использование арендуемых мощностей.

Разрабатываемая платформа должна осуществлять взаимодействие между аппаратными ресурсами таким образом, чтобы для клиента вся группа аппаратных ресурсов (серверов, хранилищ и т.д.) выглядела как единый виртуальный сервер. Клиент также должен иметь возможность прозрачно и с высокой гибкостью менять объемы потребляемых ресурсов в случае изменения своих потребностей (увеличивать/уменьшать мощность сервера).

Для создания облачной платформы воспользуемся открытой программной платформой Eucalyptus [3]. Eucalyptus является программной платформой для реализации частных облачных вычислений на компьютерных кластерах. В настоящее время эта платформа предоставляет интерфейс для взаимодействия, совместимый с сервисами Amazon [1], к тому же платформа имеет модульную структуру и может поддерживать множество разных интерфейсов одновременно. К основным преимуществам этой платформы можно отнести [2]:

- Совместимость с AmazonWebServices API.
- Установка и развертывание от источника или DEB и RPM пакетов.
- Безопасная связь между внутренними процессами через SOAP и WS-Security.
- Поддержка Linux и Windows, виртуальных машин (VM).
- Поддержка нескольких кластеров как одного облака
- Гибкие группы доступа и безопасности.
- Пользователи и группы управления.
- Бухгалтерская отчетность.
- Настраиваемые политики планирования и соглашений об уровне обслуживания

Eucalyptus реализует схему IAAS (Infrastructure-as-a-Service), а именно, “нижний” уровень облака, позволяющий поднимать виртуальные машины по запросу вышестоящих приложений. Возмож-

ности Eucalyptus во многом позаимствованы из Amazon EC2 [1], поэтому Eucalyptus с некоторыми ограничениями и допущениями их можно рассматривать как открытую альтернативу сервисам Amazon. Иерархическая структура (рис. 1). Eucalyptus состоит из четырех уровней [2].

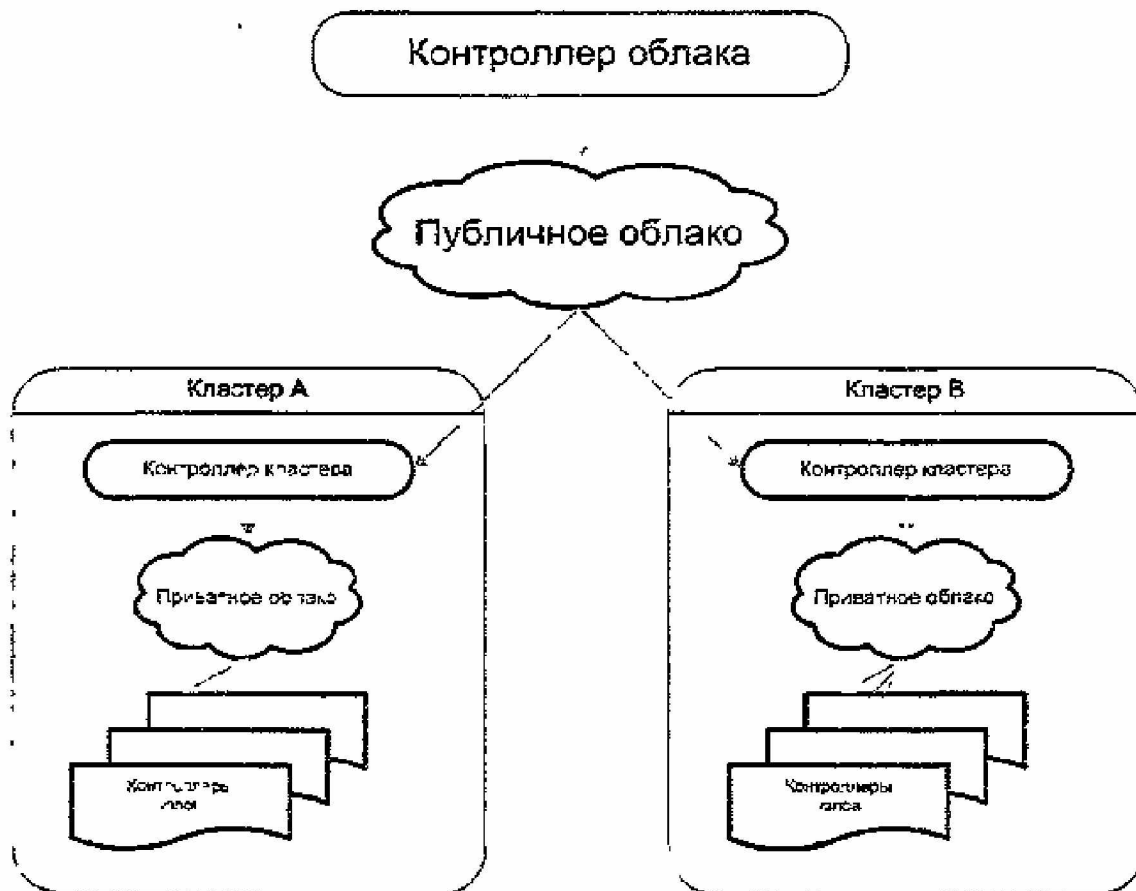


Рис. 1  
Иерархическая структура Eucalyptus

Контроллер узла (NodeController, NC) управляет запуском, работой и остановом виртуальных машин на подведомственном узле. Узел – это машина с работающим гипервизором (например, Xen), который осуществляет работу виртуальных машин (instances в терминологии Eucalyptus).

Контроллер кластера (ClusterController, CC) управляет подведомственными контроллерами узлов (NC): собирает информацию о загрузке узлов и принимает решения, на каких узлах будет выполнен запуск виртуальных машин.

Контроллер хранилища (StorageController) – место для хране-

ния образов (image) виртуальных машин. В качестве хранилища выступает Walrus – сервис, подобный Amazon S3.

Контроллер облака (CloudController, CLC) представляет собой точку входа. Со стороны конечного пользователя (или вышестоящего приложения) поступают запросы на запуск виртуальных машин. Со стороны контроллеров кластеров поступают данные о загруженности узлов облака.

### **Способ построения облака**

В разрабатываемом вычислительном комплексе, для развёртывания собственного «облака» на факультете КНИТ НИУ БелГУ рекомендуется выделить машины со следующей конфигурацией:

1. Сервер выполняющий функции контроллера облака и сетевого хранилища – Fujitsu ETERNUS JX40. Масштабируемость данной системы может достигать уровня в 72 ТБ – это максимальное значение вместительности данного устройства. Хранилище данных ETERNUS JX40 поддерживает накопители с интерфейсом SAS. Подключается установка напрямую к серверу, выполняющему сложные задания, где всегда требуется поддержка надежного хранилища. Система хранения данных Fujitsu ETERNUS JX40 способна вмещать до 144 носителей SAS, также устанавливаются диски с интерфейсом SATA. В настоящее время в системе установлено 12 накопителей общим объемом более 4 ТБ.

2. Сервер, выполняющий функции контроллера кластера A-Fujitso RX600 S5. Характеристики сервера представлены в таблице (табл. 1).

Таблица 1

**Характеристика сервера Fujitso RX600 S5**

| Характеристика               | Значение                      |
|------------------------------|-------------------------------|
| Чипсет                       | Intel 7500                    |
| Семейство процессора         | QuadCore Intel Xeon 7520      |
| Частота процессора           | 1.86ГГц                       |
| Количество процессоров       | 4                             |
| Количество ядер на процессор | 4                             |
| Объем ОЗУ                    | 32Гб                          |
| Объем HDD                    | 300Гб                         |
| Сеть                         | 4xGigabitEthernet (1000 Mbps) |

## АЛГОРИТМЫ И ПРОГРАММЫ

3. Кластерная система А включает в себя три серверных ЭВМ в качестве вычислительных узлов (узел А), характеристики такого узла представлены в табл. 2.

Таблица 2

### Технические характеристики вычислительного узла кластера А

| Характеристика              | Значение                    |
|-----------------------------|-----------------------------|
| Семейство процессора        | QuadCore Intel Xeon         |
| Частота процессора          | 1.6ГГц                      |
| Количество процессоров      | 2                           |
| Количество ядерна процессор | 4                           |
| Объем ОЗУ                   | 4Гб                         |
| Объем HDD                   | 300Гб                       |
| Сеть                        | GigabitEthernet (1000 Mbps) |

Суммарные технические характеристики кластерной системы приведены в табл. 3 (к характеристикам узлов прибавлено четыре процессора управляющей виртуальной машины и 4 Гб ОЗУ).

Таблица 3

### Суммарные технические характеристики кластера А

| Характеристика         | Значение                    |
|------------------------|-----------------------------|
| Семейство процессора   | QuadCore Intel Xeon         |
| Частота процессора     | 1.6ГГц                      |
| Количество процессоров | 10                          |
| Количество ядер        | 28                          |
| Объем ОЗУ              | 16Гб                        |
| Объем HDD              | 900Гб                       |
| Сеть                   | GigabitEthernet (1000 Mbps) |

4. Кластер В состоит из трех вычислительных серверных систем.

Технические характеристики узла кластера В приведены в табл. 4.

Таблица 4

**Технические характеристики вычислительного узла кластера В**

| Характеристика              | Значение                    |
|-----------------------------|-----------------------------|
| Семейство процессора        | QuadCore Intel XeonE5620    |
| Частота процессора          | 2.4ГГц                      |
| Количество процессоров      | 1                           |
| Количество ядерна процессор | 4                           |
| Объем ОЗУ                   | 20Гб                        |
| Объем HDD                   | 250Гб                       |
| Количество графических плат | 4 / 4/ 2                    |
| Сеть                        | GigabitEthernet (1000 Mbps) |

В составе вычислительных узлов имеются также графические ускорители NVidiaTeslaC2070

Суммарные технические характеристики кластера В приведены в табл. 5.

Таблица 5

**Технические характеристики кластера В**

| Характеристика              | Значение                    |
|-----------------------------|-----------------------------|
| Количество процессоров      | 3                           |
| Количество ядер             | 12                          |
| Объем ОЗУ                   | 60Гб                        |
| Объем HDD                   | 750Гб                       |
| Количество графических плат | 10                          |
| Объем видеопамяти           | 60Гб                        |
| Сеть                        | GigabitEthernet (1000 Mbps) |

5. Высокопроизводительный вычислительный комплекс (ВВК), выполняющий функции кластера С.

Комплекс состоит из: вычислителя, системной сети, управляющей сети, служебных серверов. Пиковая производительность ВВК не менее 26,7 TFLOPS.

В итоге получена платформа (рис.2), обеспечивающая реализацию облачных вычислений на высокопроизводительной гибридной системе.

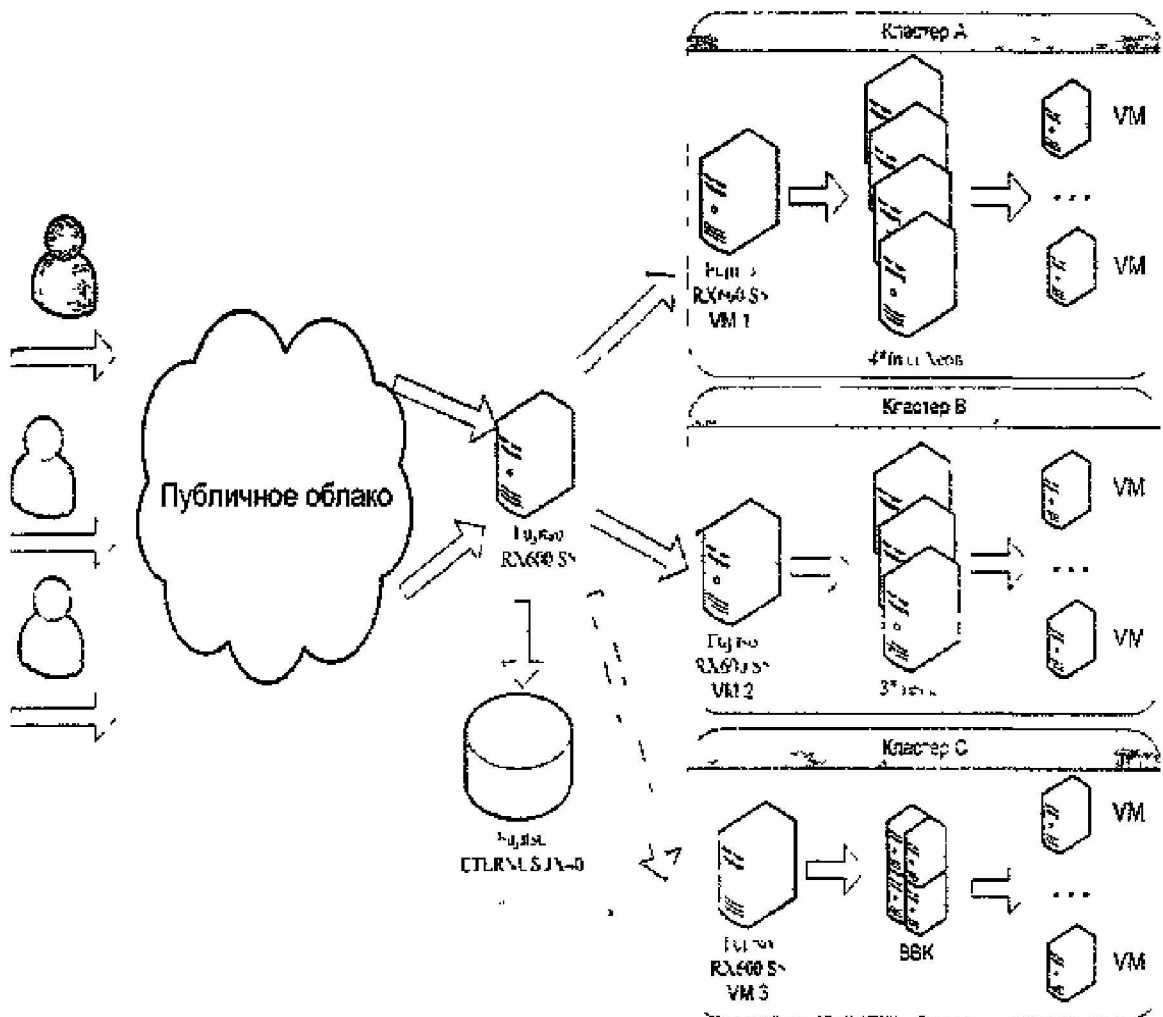


Рис. 2

Платформа, обеспечивающая реализацию облачных вычислений

**Выводы**

С использованием программной платформы Eucalyptus создана «облачная» инфраструктура. Созданное решение может применяться для предоставления как внутренних, так и публичных сервисов. Eucalyptus не самая простая в использовании система, однако, настроив один раз облако, получаем в распоряжение свой собственный Amazon, который может работать достаточно длительные периоды (несколько месяцев), не требуя вмешательства.

В результате этого будет создан центр облачных вычислений в НИУ БелГУ, который позволит обеспечить решение вычислительно трудоемких научно-технических задач по различным направлениям.



Литература

1. Amazon Web Services home page. <http://aws.amazon.com/>.
2. Программная платформа eucalyptus <http://www.eucalyptus.com/>.
3. Eucalyptus Public Cloud (EPC) <http://eucalyptus.cs.ucsb.edu/wiki/EucalyptusPublicCloud/>.
4. Фингар П. Dot.Cloud: облачные вычисления - бизнес-платформа XXI века: Акваринная книга. Б.м., 2011. 253 с.
5. Риз Дж. Облачные вычисления: СПб., БХВ-Петербург, 2011 288 с.

*Статья поступила 09 12 2011*

**К.т.н., доц. В.М. Михелев, К.В. Кузнецов, Д.А. Торопчин**  
(НИУ «БелГУ»)

**V.M. Mikhelev, K.V. Kuznetsov, D.A. Toropschin**

**ПАРАЛЛЕЛЬНЫЕ ВЫЧИСЛЕНИЯ  
ПРИ РАСПОЗНАВАНИИ НЕЧЕТКИХ ТЕКСТОВЫХ  
ДУБЛИКАТОВ НА ОСНОВЕ ТЕХНОЛОГИИ CUDA**

**PARALLEL COMPUTING OF FUZZY RECOGNITION  
OF DUPLICATE TEXT BASED OF TECHNOLOGY CUDA**

*Проведено исследование нескольких основных методов определения текстовых дубликатов, основанных на построении частотных словарей документов или коллекций и алгоритмов семейства шинглов. Предложен новый алгоритм распознавания нечетких текстовых документов, позволяющий использовать параллельные вычисления на основе технологии CUDA.*

*A study of several basic methods for measuring text duplicates based on building of frequency dictionary documents or collections and algorithms family of shingles. A new algorithm of fuzzy recognition text documents allows you to use parallel computing based on CUDA technology.*

*Ключевые слова: параллельное программирование, текстовый дубликат, метод шинглов, технология CUDA.*

*Key words: parallel computing, duplicate text, computer recognition method of shingles, cuda technology.*

**Введение**

Рост объемов информации в последние годы обуславливает