

## СЛАБОСТРУКТУРИРОВАННЫЕ БАЗЫ ДАННЫХ В ГИБРИДНОЙ ОБЛАЧНОЙ ИНФРАСТРУКТУРЕ

Плужник Е.В.<sup>1</sup>, Никульчев Е.В.<sup>1</sup>

<sup>1</sup>НОУ ВПО «Московский технологический институт «ВТУ», Москва, Россия (117292, г. Москва, ул. Кедрова, д. 8, кор. 2), e-mail: e\_nikulchev@mti.edu.ru

Статья посвящена исследованию вопросов построения и эффективности использования облачных технологий для построения информационных систем научно-образовательного назначения, работающих со слабоструктурированными базами данных. Определены основные принципы функционирования автоматизированных систем управления облачными ресурсами. Особенности предложенной архитектуры является использование порталных технологий, автоматическое управление ресурсами и гибридная облачная инфраструктура. Для формирования системы управления рабочими потоками запросов к системе научно-образовательного содержания проведены экспериментальные исследования запросов к гибридной базе данных с XML-данными. Приведены результаты экспериментов, показывающие эффективность использования запросов к облачным сервисам. Показано, что для сложных запросов к слабоструктурированным данным время передачи данных из внешнего облака сравнимо с поиском на локальных серверах. На основе опыта разработки систем приведена технология формирования информационных систем, ориентированных на использование облачных сервисов.

Ключевые слова: облачные вычисления, образовательные информационные системы, слабоструктурированные базы данных, гибридные облачные инфраструктуры.

## SEMISTRUCTURED DATABASE OF HYBRID CLOUD COMPUTING INFRASTRUCTURES

Pluzhnik E.V.<sup>1</sup>, Nikulchev E.V.<sup>1</sup>

<sup>1</sup> Moscow Technology Institute «VTU», Moscow, Russia (117292, Moscow, Kedrov St., 8, box. 2), e-mail: nikulchev@mail.ru

The article is devoted to research aimed at the design and efficiency of cloud technology in the field of science and education. The features of these systems is to use a semistructured databases. The experimental results are attached. The article defines the basic principles of automated control systems for cloud resources. The features of the developed architecture is the use of portal technology, automatic control resource and hybrid cloud infrastructure. To form the workflow management system queries to the system of scientific and educational content of the experimental studies of hybrid queries to a database with XML-data. The experimental results are attached. Thesis results are showing the effectiveness of queries to cloud services. For complex queries for semistructured data demonstrated that time of transmission of data from the public cloud is comparable with the time of the search query on local servers. Generalizing the experience of systems development is given technology of information systems focused on the use of cloud services.

Keywords. cloud computing, educational information systems. semistructured databases, hybrid cloud infrastructure.

### Постановка задачи

Под облачными вычислениями, как правило, понимают интернет-сервисы, предоставляемые специализированными центрами обработки данных в виде аппаратного и системного программного обеспечения [9]. В соответствии с [5] облачные вычисления – это распределенная вычислительная система, состоящая из набора взаимосвязанных виртуальных машин, позволяющая динамически предоставлять вычислительные ресурсы с определенным уровнем обслуживания.

Оба приведенных определения дают представление о современном направлении в информационных технологиях — облачных сервисах, дающих возможность размещать

программные приложения и базы данных во внешних центрах обработки данных и делать их доступными через Интернет. Разделяют три вида облачных сервисов: инфраструктура как услуга (IaaS), платформа как услуга (PaaS) и программное обеспечение как услуга (SaaS). Для поддержки приложений Amazon, HP, IBM, Google развернули облачные центры обработки данных по всему миру.

Управление информационными системами на основе полноценного использования облачной инфраструктуры предлагает решение сложных задач [6-8]:

- гарантирование качества обслуживания (Quality of Service, QoS);
- оптимизация ресурсов (снижения энергозатрат, оптимизация стоимости и пр.);
- обеспечение безопасности (как гарантирование сохранения конфиденциальности, так и сохранность данных).

В целом мировые тенденции таковы, что облачные сервисы вытесняют «классические» архитектуры информационных систем, основанные на построении и сопровождении собственных дата-центров (в настоящее время обеспечение отказоустойчивости и критического восьмичасового времени восстановления требует вложения не одной сотни миллионов рублей в серверную плюс затраты на энергообеспечение и строительные работы; последние часто во многих зданиях вообще не осуществимы). Поэтому надо быть готовым к переносу существующих систем в облака. Особенно это актуально для вузов [3], использующих современные дистанционные технологии, и компаний, оказывающих услуги по Интернету для большой пространственно-распределенной группы потребителей [9].

В последнее время появилось значительное количество работ зарубежных авторов, посвященных разработке автоматизированных систем управления динамическим распределением облачных ресурсов. Автономные системы управления облачными ресурсами объединяют в себе свойства самоконтроля, самовосстановления и оптимизации, основанные на мониторинге собственных ресурсов и вычислении настройки собственной работы [6].

Статья посвящена разработке методологии построения информационных систем академического назначения, использующих технологию гибридных облаков, основанную на опыте Московского технологического института «ВТУ». Особенностью систем научного и образовательного назначения является значительное количество слабоструктурированных данных [4].

### ***1. Архитектура управления облачными системами***

В [6] определены основные принципы функционирования автоматизированных систем управления ресурсами. Применительно к задачам построения информационных

систем в облачной инфраструктуре (компоненты архитектуры) они будут иметь вид, приведенный на рис. 1.

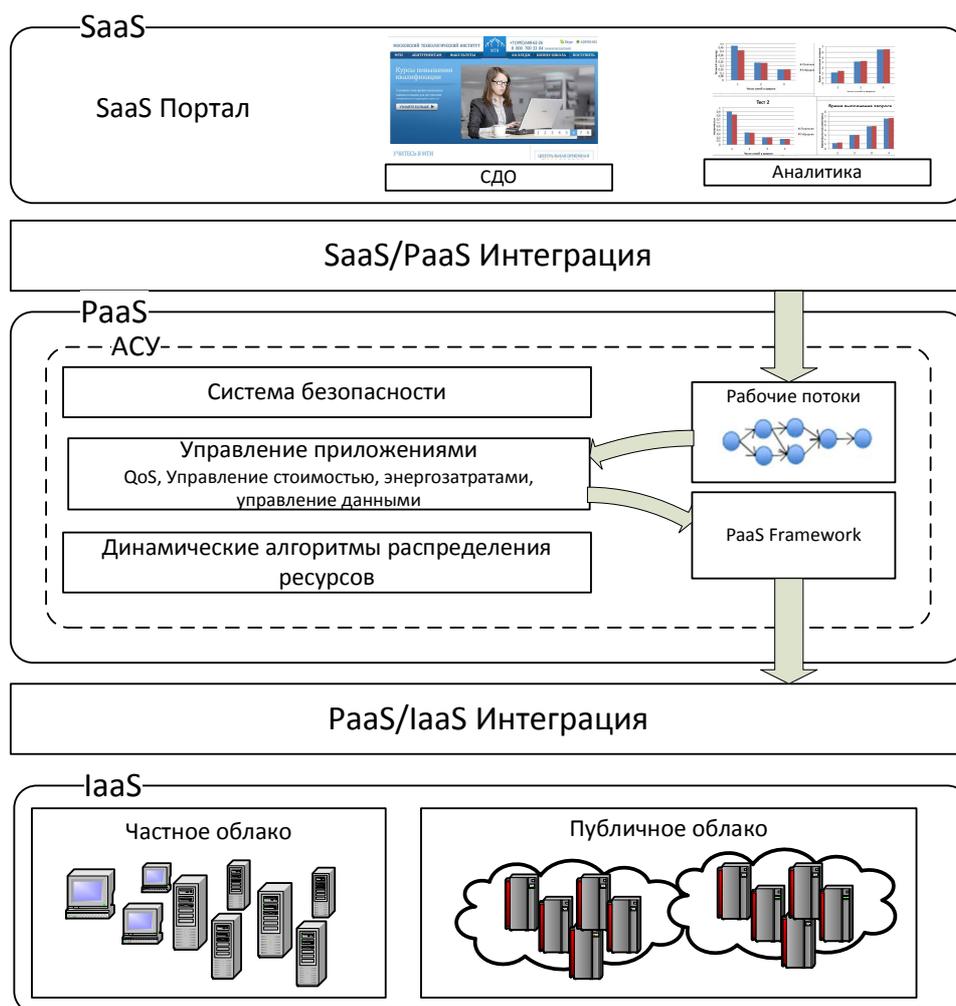


Рис. 1. Структура облачных сервисов с автоматическим управлением.

Основными компонентами архитектуры являются:

- применение SaaS-портала — позволяет реализовать персонафикацию доступа к приложениям [2];
- автономная система автоматического управления и Framework — включает реализацию принципов автономного управления, включая модули оптимизации [8], при этом QoS-приложения и планировщик добавляются в качестве пользовательских плагинов [1];
- гибридные облака в IaaS — комплексное использование частных облаков компании и публичные облачные сервисы [3].

## 2. Структура экспериментального исследования запросов

Для формирования блока управления рабочими потоками запросов к системе научно-образовательного содержания проведены экспериментальные исследования, которые

схематично можно представить в виде рис. 2. Для экспериментального исследования использовались базы данных, структура которых приведена на рис. 3а.

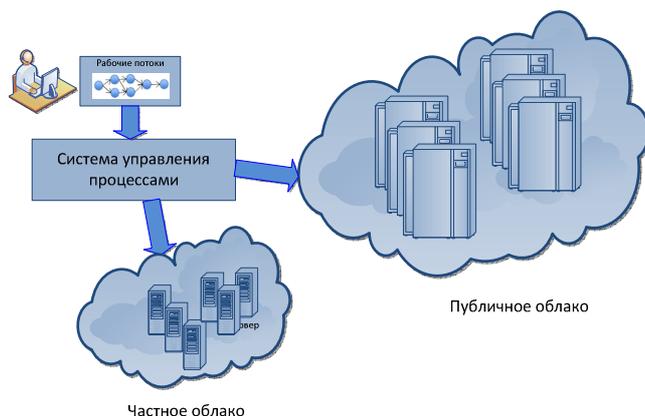


Рис. 2. Структура экспериментального построения системы управления потоками.



а) Схема данных локальной БД

б) Схема данных Гибридного хранилища

Рис. 3. Структуры данных.

В таблице Articles хранятся статьи, размер статьи от 100 Кб до 3 Мб. В таблице Authors содержатся данные авторов статей. Таблица AuthorOfArticles связывает автора со статьёй. У одной статьи может быть один главный автор и несколько соавторов. В тестовой загрузке может быть от 0 до 9 соавторов.

В локальной БД (articlesLocal) данные об авторах и статье хранятся в реляционной базе данных MS SQL Server. Занимаемая память на сервере БД (articlesLocal) 26667,25 МБ.

Структура гибридной БД приведена на рис. 2б. В гибридной БД (articlesHybrid) информация об авторах и статьях в локальной БД на MS SQL Server, а тело статьи — в

облачном хранилище Azure Storage. Занимаемая память на сервере БД (articlesHybrid) 47,08 Мб, в облаке — приблизительно 27 Гб.

Произведены тестовые поисковые запросы. Результаты двух экспериментов приведены в табл. 1, 2 и на рис. 4, 5.

Таблица 1. Результаты эксперимента 1.

| Записей в секунду |           | Число записей в запросе | Время извлечения всех статей |           | Среднее время выполнения запроса |           | Число извлеченных статей, всего |
|-------------------|-----------|-------------------------|------------------------------|-----------|----------------------------------|-----------|---------------------------------|
| Локальная         | Гибридная |                         | Локальная                    | Гибридная | Локальная                        | Гибридная |                                 |
| 0,908074          | 0,826204  | 1                       | 110,123                      | 121,035   | 1,101                            | 1,210     | 100                             |
| 0,335978          | 0,329768  | 2                       | 297,638                      | 303,243   | 2,976                            | 3,032     | 200                             |
| 0,21054           | 0,208225  | 3                       | 474,969                      | 480,248   | 4,749                            | 4,802     | 300                             |
| 0,150936          | 0,149446  | 4                       | 662,533                      | 669,136   | 6,625                            | 6,691     | 400                             |

Таблица 2. Результаты эксперимента 2.

| Записей в секунду |           | Число записей в запросе | Время извлечения всех статей |           | Среднее время выполнения запроса |           | Число извлеченных статей, всего |
|-------------------|-----------|-------------------------|------------------------------|-----------|----------------------------------|-----------|---------------------------------|
| Локальная         | Гибридная |                         | Локальная                    | Гибридная | Локальная                        | Гибридная |                                 |
| 0,556313          | 0,545451  | 1                       | 179,755                      | 183,334   | 1,7975                           | 1,833     | 100                             |
| 0,244723          | 0,242872  | 2                       | 408,625                      | 411,739   | 4,0862                           | 4,117     | 200                             |
| 0,160304          | 0,159421  | 3                       | 623,816                      | 627,270   | 6,238                            | 6,272     | 300                             |
| 0,113419          | 0,113139  | 4                       | 881,689                      | 883,8712  | 8,8168                           | 8,838     | 400                             |
| 0,091235          | 0,090954  | 5                       | 1096,066                     | 1099,451  | 10,960                           | 10,994    | 500                             |
| 0,075317          | 0,075191  | 6                       | 1327,718                     | 1329,938  | 13,277                           | 13,299    | 600                             |
| 0,062359          | 0,062254  | 7                       | 1603,613                     | 1606,330  | 16,036                           | 16,063    | 700                             |
| 0,050944          | 0,050865  | 8                       | 1962,932                     | 1965,999  | 19,629                           | 19,659    | 800                             |
| 0,04377           | 0,043714  | 9                       | 2284,675                     | 2287,614  | 22,846                           | 22,876    | 900                             |
| 0,03787           | 0,037819  | 10                      | 2640,578                     | 2644,182  | 26,405                           | 26,441    | 1000                            |

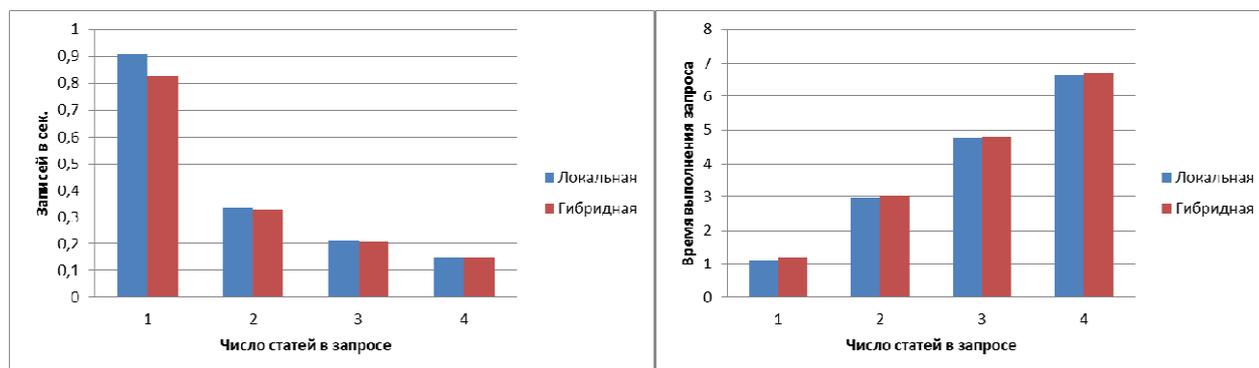


Рис. 4. Результаты эксперимента 1.

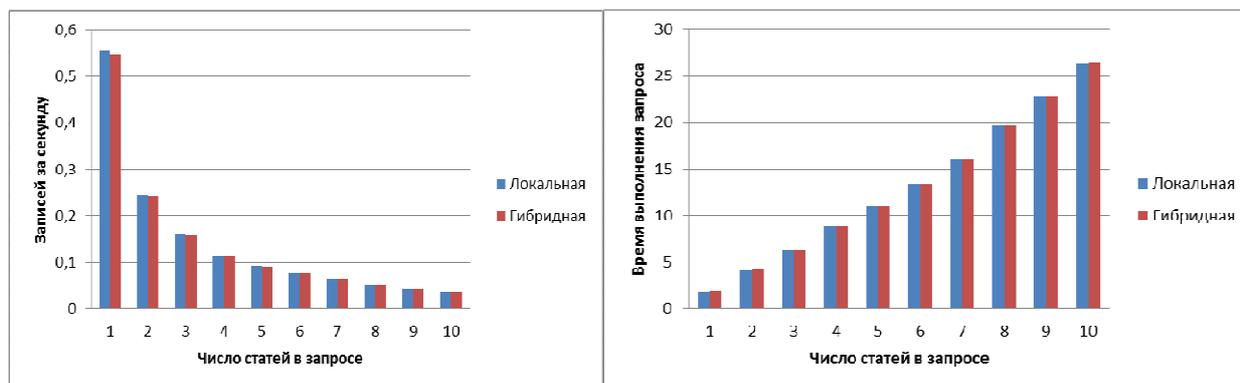


Рис. 5. Результаты эксперимента 2.

### 3. Технология построения облачных информационных систем

Проведенные исследования показали, что для больших баз данных со сложными запросами к слабоструктурированным данным имеет место отсутствие временных потерь на передачу данных в проведенном эксперименте.

Таким образом, для построения информационных систем научного и образовательного содержания со слабоструктурированными данными разработана технология, состоящая из четырех этапов.

1. Оценка общих параметров системы (максимальное количество пользователей для одновременной работы, возможность масштабирования сервисов, наличия персонафицированного доступа).
2. Оценка стоимости проекта (наличие собственных серверных мощностей, сравнение стоимости построения со стоимостью аренды сервисов).
3. Оценка времени доступа к данным, оценка производительности запросов для облачных инфраструктур.
4. Построение автоматической системы распределения ресурсов и направления запросов в распределенной базе данных.

Для решения первого этапа используются многокритериальные методы принятия решений, второй этап реализуется на основе экономико-математических методов оценки проектов; третий и четвертый - на основе методов оптимизации и эффективной оценки поисковых запросов.

Результаты использованы для построения систем в НОУ ВПО «Московский технологический институт «ВТУ».

*Работа выполнена частично при финансовой поддержке РФФИ (грант № 11-07-00772-а).*

### Список литературы

1. Никульчев Е.В. Построение модели загрузки каналов связи в сетях передачи данных на основе геометрического подхода / Е.В. Никульчев, С.В. Паяин // Известия высших учебных заведений. Проблемы полиграфии и издательского дела. — 2008. — № 6. — С. 91-95.
2. Плужник Е.В., Никульчев Е.В. Персонализованные информационные системы и слабоструктурированные базы данных в облачных технологиях [Электронный ресурс] // Всероссийская конференция: Индустриальные информационные системы. - URL: <http://conf.nsc.ru/iis2013/ru/reportview/161281> (дата обращения: 10.08.2013).
3. Плужник Е.В., Никульчев Е.В. Функционирование образовательных систем в гибридной облачной инфраструктуре // Известия вузов. Проблемы полиграфии и издательского дела. — 2013. — № 3. — С. 65–71.
4. Шокин Ю.И. Технология создания программных систем информационного обеспечения научной деятельности, работающих со слабоструктурированными документами / Ю.И. Шокин, А.М. Федотов, В.Б. Барахнин // Вычислительные технологии. — 2010. — Т. 15. — № 6. — С. 111-125.
5. Armbrust M. A view of cloud computing / M. Armbrust, A. Fox, R. Griffith etc. // Communications of the ACM. — 2010. — Vol. 53. — No. 4. — P. 50-58.
6. Babu S. Automated control in cloud computing: challenges and opportunities / S. Babu, J. Chase, S. Parekh // 1st workshop on Automated control for datacenters and clouds, 2009. — P. 13–18 (DOI: 10.1145/1555271.1555275).
7. Beloglazov A. Managing overloaded hosts for dynamic consolidation of virtual machines in cloud data centers under quality of service constraints / A. Beloglazov, R. Buyya // IEEE Transactions on Parallel and Distributed Systems. — 2013. — Vol. 24.— No. 7. — P. 1366-1379.
8. Naak S. Autonomic benchmarking for cloud infrastructures: an economic optimization model / S. Naak, M. Menzel // 1st ACM/IEEE Workshop on Autonomic Computing for Economics, 2011. — P. 27-32 (DOI: 10.1145/1998561.1998569).
9. Rittinghouse J.W. Cloud computing—implementation, management, and security / J.W. Rittinghouse, J.F. Ransome. — NY: Taylor and Francis Group, 2010.

**Рецензенты:**

Ковшов Е.Е., д.т.н., профессор, зав. кафедрой управления в технических система МГТУ «СТАНКИН», г. Москва.

Барахнин В.Б., д.т.н., доцент, с.н.с. ИВТ СО РАН, г. Новосибирск.