

## Библиотеки в «облаках»: практические аспекты Libraries in «clouds»: Practical aspects

*Е. В. Ковязина*

*Институт вычислительного моделирования СО РАН,  
Красноярск, Россия*

*Elena Kovyazina*

*Institute of Computational Modeling  
of the Russian Academy of Sciences Siberian Branch,  
Krasnoyarsk, Russia*

Доклад посвящен использованию grid и облачных технологий в библиотеках российских регионов. Приведены базовые определения grid и облачных вычислений, даны некоторые сравнительные характеристики. Распространенные приемы работы приведены в соответствии с моделями услуг облачных вычислений. Выделены их достоинства и недостатки. Уделено внимание практическим рекомендациям при работе в каждой из этих моделей.

The report focuses on the use of grid and cloud technology in the libraries of the Russian regions. Basic definitions of grid and cloud computing and their some comparative characteristics are given. Common methods of work are described in accordance with the models of cloud computing services. Their advantages and disadvantages are highlighted. Practical recommendations for each of the models are offered.

Развитие телекоммуникационной и информационной инфраструктур, стремительное совершенствование компьютерной техники, включая персональные устройства, с одной стороны, и экспоненциальный рост объемов доступной цифровой информации, с другой, обусловили проникновение новых технологий в сферу традиционной библиотечной автоматизации. Проблема неразрешимости ряда задач, стоящих перед библиотечными специалистами, в рамках традиционных систем автоматизации библиотек (САБ) отчетливо проявилась несколько лет назад [1]. С течением времени проблемы только обострились. Растущие объемы данных, количество обращений, потребность в новых сервисах и т.п. обусловили необходимость применения новейших компьютерных технологий в библиотеках. В частности, все чаще упоминается об облачных вычислениях по отношению к библиотекам. Причем, в публикациях под облачными вычислениями понимают достаточно широкий круг технологий: от традиционных web, ftp, e-mail [2] до достаточно новых grid, виртуализации и распределенных систем [3]. Попробуем разобраться в особенностях технологий, их достоинствах и недостатках, а также практических аспектах работы с ними.

На начальных этапах библиотечной автоматизации одним из базовых требований являлась поддержка сетевых протоколов Z39.50. Корпоративные библиотечные системы, построенные на их основе, представляют собой пример grid вычислений. Обратимся к определению: «*grid (grid computing) – система, которая связана с интеграцией, виртуализацией и управлением услугами и ресурсами в распределенной, гетерогенной среде, которая поддерживает коллекции пользователей и ресурсов (виртуальных организаций) в традиционных административных и организационных доменах (реальных организаций)*» [4-5]. основополагающим принципом является распределенность данных, управление, поддержку и доступ к которым обеспечивает множество организаций-хозяев на основе оговоренных стандартов и правил. Пользователю обеспечивается доступ к единому виртуальному ресурсу, который он использует как реальный. По отношению к библиотекам, например, систему распределенных библиотечных каталогов можно считать гетерогенной средой лишь условно, имея в виду интеграцию различных САБ. В отдельных случаях, обеспечивался и доступ к истинно гетерогенным данным, например, таким как музейные экспонаты. Поддержка и обслуживание подобных систем была достаточно обременительна для хозяев данных. Требовалось наличие вычислительной техники и программного обеспечения, квалифицированного ИТ-персонала. С ростом объемов данных проявлялись дополнительные проблемы, связанные с их сохранностью, архивированием и защитой от вирусных и хакерских атак.

Grid не единственные системы, где присутствовала *виртуализация*, в приведенном случае на уровне данных. Достаточно распространены виртуальные машины на персональном компьютере или сервере, предназначенные для разделения операционной среды и приложений. Например, LINUX на Windows-машинах, и наоборот, или несколько виртуальных машин для различных пользователей на сервере. Проникновение виртуализации в Интернет на фоне бурного развития web-сервисов привели к широкому распространению концепции облачных вычислений. Общепринятое определение таково: «**Облачные вычисления** (*cloud computing*) это модель предоставления повсеместного и удобного сетевого доступа по требованию к общему пулу конфигурируемых вычислительных ресурсов (например, сетей, серверов, систем хранения, приложений и сервисов), которые могут быть оперативно предоставлены и освобождены с минимальными эксплуатационными затратами и/или обращениями к провайдеру услуг» [4-5]. С точки зрения истории развития облако – это симбиоз виртуализации и web-сервисов. Образно говоря, облако – это стена между пользователями и провайдерами, которая скрывает то, что происходит на стороне провайдера, предоставляя пользователю только необходимый ему набор услуг.

В общем случае разница между grid и облачными вычислениями достаточно условна, так как налицо взаимопроникновение технологий. Grid-системы зачастую реализуются через web-сервисы, а в облачных технологиях используются grid решения. Это отмечено в ряде публикаций, в том числе и посвященных сравнительному анализу этих двух технологий [6-7]. При разделении grid и облачных систем, по-видимому, имеет смысл придерживаться того, что grid – распределенная система, частью которой может быть и ваша библиотека, облако – система централизованная, имеющая единого стороннего хозяина, как правило, провайдера услуг. В силу этого grid обычно некоммерческое объединение организаций, связанных общим проектом, а облако – коммерческая структура вполне определенного владельца. Только самые крупные и финансово независимые организации могут позволить себе владеть собственным сколько-нибудь большим облаком [8].

Для огромного числа малых и средних организаций внешнее облако представляется весьма желанной и экономной структурой. Можно не приобретать сложную и дорогостоящую вычислительную технику, не оплачивать квалифицированный персонал для ее обслуживания. Предоставляемые внешними провайдерами услуги, как правило, дешевле, чем совокупные траты на покупку и содержание парка вычислительной техники. Как следствие, библиотеки сравнительно легко отказываются от GRID-систем и дрейфуют в облако, не всегда ясно осознавая последствия и трезво оценивая перспективы такого перехода.

Попробуем выделить варианты жизни библиотек в облаке в терминах моделей услуг облачных технологий, а затем оценить их с точки зрения достоинств и недостатков облачных систем.

**PaaS (Platform as a Service – платформа как услуга).** Потребитель (библиотека) арендует у провайдера (возможно, у соседней более крупной библиотеки) вычислительную технику и хранилища данных. Программное обеспечение используется собственное. Потребитель при этом не управляет сетями, серверами, операционными системами и системами хранения данных (базовой инфраструктурой облака), но осуществляет контроль над развернутыми приложениями и, возможно, некоторыми параметрами конфигурации среды хостинга. То есть на вычислительной технике провайдера устанавливается САБ, web-сервер и сайт библиотеки, вспомогательное программное обеспечение. В российских библиотеках для доступа к такому облаку чаще всего используется удаленный рабочий стол. Достоинства такой облачной среды:

1. Доступность – подключиться можно из любой точки мира, где есть Интернет;
2. Гибкость – неограниченность вычислительных ресурсов за счет виртуализации, легкость масштабирования и администрирования;
3. Надежность – как правило, такие облака находятся в специально оборудованных Центрах обработки данных (ЦОД), которые имеют резервные источники питания, охрану, профессиональных работников, регулярное резервирование данных, высокую пропускную способность Интернет-каналов, высокую устойчивость к вирусным и хакерским атакам.

В такой модели библиотека-пользователь решает, пожалуй, самую дорогостоящую для нее проблему – проблему хранилища данных. Провайдер берет на себя обязанность хранения, архивации и восстановления данных после сбоев, причем это касается не только каталогов и коллекций,

но и ПО. В повседневной работе создается, как правило, ежедневная копия данных, что обеспечивает высокую надежность работы в такой модели.

Однако, стоит упомянуть и о недостатках:

1. Требуется постоянное соединение с сетью. На случай аварии может возникнуть необходимость иметь резервную копию всего комплекса.
2. В силу урезанных прав администрирования ОС, пользователь имеет ограничения в ПО, и не всегда ему предоставлена возможность настроить его под свои собственные цели. Зачастую пользователь ограничен также и в возможности исследования эксплуатационных характеристик работающего ПО и лишен права оптимизации его работы.

На современном этапе использование данной модели в библиотеках носит достаточно ограниченный характер. Она либо работает в тестовом режиме в каком-либо проекте, либо в корпоративном коммунальном облаке. Поэтому пользователям бывает достаточно затруднительно определить реальную цену такой работы, хотя бы гипотетически. По-видимому, можно рекомендовать для оценки предполагаемых финансовых затрат таблицу цен крупных вендоров Интернет, приводимых в некоторых публикациях (таб.1 [9]).

*Таблица 1. Расценки наиболее крупных облачных вендоров*

	Единица измерения	Amazon	Google	Microsoft
Объем хранящихся данных	Гбайт/мес.	0,10 долл.	0,15 долл.	0,15 долл.
Количество транзакций к данным в хранилище	на каждый 10 тыс обращений	0,10 долл.		0,10 долл.
Исходящий трафик	Гбайт	0,10–0,17 долл.	0,12 долл.	0,15 долл.
Входящий трафик	Гбайт	0,10 долл.	0,10 долл.	0,10 долл.
Процессорное время	за 1 час работы	0,10–1,20 долл.	0,10 долл.	0,12 долл.

**SaaS (Software as a Service – программное обеспечение как услуга).** При таком способе работы библиотека не покупает специализированное программное обеспечение, например, САБ, совсем, либо частично, что позволяет ей на этом сэкономить. Это достаточно известный и наиболее часто используемый метод работы – аутсорсинг. Провайдер, владеющий ПО, держит на своем сервере данные, к которым пользователи подключаются с помощью тонкого или толстого клиента. Эта модель часто используется при построении сводных каталогов и региональных ведомственных библиотечных систем. Плата за аренду ПО в этой модели обычно выше, чем плата за стандартное сопровождение своего собственного ПО, но существенно ниже, чем его покупка. Расширенный набор услуг, включающий адаптацию и доработку ПО под нужды конкретной библиотеки, предоставляется за дополнительную плату. Достоинства:

1. Те же, что и в предыдущей модели.
2. Возможность сэкономить на покупке вычислительной техники, ПО и IT-персонале.
3. Качество сопровождения в этой модели, как правило, существенно выше за счет высокой квалификации персонала и использования дополнительных услуг, таких как CRM и/или бесплатный справочный телефон.

Однако, перечисленные достоинства при ближайшем рассмотрении можно считать недостатками:

1. В случае прекращения оплаты аутсорсинга ваши данные остаются у провайдера и могут быть переданы вам, в лучшем случае, в виде выгруженного из системы файла или группы файлов.
2. Высокое качество сопровождения, достигаемое за счет дополнительных услуг, должно окупаться, поэтому все дополнительные услуги предоставляются за отдельную плату.
3. Существуют исследования, оценивающие экономию средств при аутсорсинге ПО [10]. В соответствии с представленными в них данными, если в первый год использования ПО в этой модели экономия средств порядка 70% (аренда вместо покупки), то уже на второй год она падает до 47% (аренда вместо сопровождения). А если пользователям требуется адаптация системы, то экономии может и вовсе не случиться.

**IaaS (Infrastructure as a Service инфраструктура как сервис).** Такую модель облака связывают преимущественно с набором услуг крупных вендоров в Интернет, хотя она существует в зачаточном состоянии внутри представленной выше модели SaaS в виде CRM. Библиотеки широко используют эту инфраструктуру, являясь своего рода экспертами в этой области. Все широко известные брэнды Интернет есть по сути облачные технологии. Социальные сети, онлайн-овые информационные ресурсы, индексы научного цитирования, агрегация информационных ресурсов и т.д. – все это облачные технологии. Мы не знаем, как они устроены, из чего складывается ценообразование платных систем, но все мы охотно пользуемся ими. Часть услуг, предоставляемых этими системами, оплачивают рекламодатели, что делает эти услуги бесплатными и еще более привлекательными для пользователей. Так 70% наших клиентов САБ ИРБИС используют для обмена информацией электронную почту mail.ru, а не собственные почтовые сервера – их просто нет на всех уровнях их организационной структуры, включая администрацию. Предоставляют вендоры и услуги двух предыдущих моделей PaaS и SaaS, но они не так распространены среди российских пользователей. Достоинства инфраструктуры облачных систем Интернет хорошо известны. Но есть и недостатки, отраженные во множестве публикаций [4, 5, 9]:

1. Отсутствие интероперабельности – нет набора универсальных стандартов и интерфейсов, что увеличивает зависимость от поставщика (попробуйте, например, переместить всю вашу почту из mail.ru на yandex.ru).
2. Постоянное соединение с сетью – для получения доступа к услугам облака необходимо постоянное соединение с сетью Интернет. Однако в наше время это не такой и большой недостаток, особенно при использовании мобильного соединения.
3. Программное обеспечение и его кастомизация – пользователь ПО имеет ограничения в используемом ПО, иногда весьма существенные, и не всегда имеет возможность настроить его под свои собственные цели; также некоторые поставщики ограничивают выбор языковых пакетов.
4. Сохранение конфиденциальности – конфиденциальность данных, хранимых на публичных облаках, в настоящее время вызывает много споров, но в большинстве случаев эксперты сходятся в том, что не рекомендуется хранить наиболее ценные для компании документы на публичном облаке, так как в настоящее время нет технологии, которая бы гарантировала 100%ную конфиденциальность хранимых данных.
5. Обеспечение надёжности и безопасности – что касается надёжности хранимой информации, то с уверенностью можно сказать одно, что если вы потеряли информацию, хранимую в облаке, то вы ее потеряли навсегда; что же до безопасности – облако само по себе является достаточно надёжной системой, однако при проникновении в него злоумышленник получает доступ ко всему огромному хранилищу данных.

Кроме того, существует определенная настороженность пользователей к отчуждаемым данным – данные в облачном хранилище воспринимаются как не вполне свои. А отсутствие интероперабельности приводит к многократному дублированию работы по созданию данных. Наиболее яркий пример дублирования – многочисленные индексы научного цитирования.

**Выводы.** Облачные технологии становятся повседневной реальностью в практике работы библиотек. Однако, их внедрение требует ясного понимания целей и перспектив такого шага. Не стоит забывать, что облачные вычисления – коммерческие технологии, нацеленные на извлечение прибыли, даже если на первых парах вам предлагают воспользоваться ими бесплатно. Необходима всесторонняя и продуманная оценка финансовых затрат и рисков, например, в случае нестабильного финансирования. Будьте готовы к тому, что недостающее финансирование будет компенсировано навязчивой рекламой в процессе работы. Следует также понимать, что полномасштабное использование облачных технологий в российских библиотеках находится в зачаточном состоянии в силу финансовых и технологических причин. Однако, есть надежда, что большинство проблем будут решены, появятся отечественные разработки, облака, говорящие по-русски и реализующие весь спектр web-услуг.

## Литература

1. Ковязина Е.В. Перспективы развития автоматизации библиотек [Текст] / Е.В.Ковязина // Научные и технические библиотеки. – 2011. – №2 – с. 89–92.
2. Sanchati, R. Cloud Computing in Digital and University Libraries [Текст] [Электронный ресурс] / R. Sanchati, G. Kulkarni // Global Journal of Computer Science and Technology. – 2011. – Vol. XI, Iss. XII, ver. 1.0. – с. 37–41. – URL: <http://computerresearch.org/stpr/index.php/gjst/article/viewFile/860/765>.
3. Kaushik, A. Application of Cloud Computing in Libraries [Текст] [Электронный ресурс] / A. Kaushik, A. Kumar // International Journal of Information Dissemination and Technology. – 2013. – Vol. 3(4). – с. 270–273. – URL: <http://www.ijidt.com/index.php/ijidt/article/viewFile/3.4.9/pdf>.
4. Интероперабельность в облачных вычислениях [Текст] / Е.Е.Журавлев [и др.] // Журнал радиоэлектроники. – 2013. – №9. – с. 1–63. – URL: <http://razinkin.16mb.com/publications/clouds>.
5. Хрусталеv, Е.Ю. Методический подход к проектированию сервисов упрощенной интеграции распределенных ИТ-ресурсов [Текст] / Е.Ю.Хрусталеv, А.А.Чумичкин / Информационные ресурсы России. – 2012. – №3. – с. 2–6.
6. Myerson, J. V. Cloud computing versus grid computing Cloud computing versus grid computing Cloud Computing versus grid computing [Текст] [Электронный ресурс] / J.M. Myerson. – IBM, 2009. – URL: <http://www.ibm.com/developerworks/library/wa-cloudgrid/wa-cloudgrid-pdf.pdf>.
7. Hashemi, S.M. Cloud Computing Vs. Grid Computing [Текст] [Электронный ресурс] / S.M.Hashemi, A.K. Bardsiri // ARPJ Journal of Systems and Software. – 2012. – Vol. 2, № 5. – с. 188–194. – URL: [http://scientific-journals.org/journalofsystemsandssoftware/archive/vol2no5/vol2no5\\_4.pdf](http://scientific-journals.org/journalofsystemsandssoftware/archive/vol2no5/vol2no5_4.pdf).
8. Ильин, В.А. Больше данных, хороших и разных! [Текст] / В.А.Ильин, В.Е.Велихов // В мире науки. – 2014. – №2. – с. 38–44.
10. Новиков И. Облачные вычисления: на пороге перемен [Текст] [Электронный ресурс] / И.Новиков // PC Magazine/RE. – 2011. – №4. – URL: <http://www.pcmag.ru/solutions/detail.php?ID=44441>.
11. Емельянов И. Миф о дешевизне облачных решений [Текст] [Электронный ресурс] / И.Емельянов // Компьютерра. – 2013. – №10. – URL: <http://www.computerra.ru/cio/5574>.