## Практические аспекты работы библиотек в «облаках»

***Libraries in “clouds” in practice***

# Ковязина Елена Васильевна, научный сотрудник, кандидат технических наук

*Федеральное государственное бюджетное учреждение науки Институт вычислительного моделирования Сибирского отделения Российской академии наук*

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 их этих моделей.

Проникновение облачных вычислений в библиотеки обусловлено растущими объемами данных, количеством обращений, потребностью в новых сервисах для читателей и еще целым рядом задач. Однако само толкование облачных вычислений в применении к библиотекам достаточно далеко от однозначности. Даже сам термин «облачные вычисления», как прямой перевод cloud computing, трудно воспринимается сотрудниками библиотек, для которых «вычисления» - это что-то из области математики. Предпочтительнее для легкости понимания употреблять термин «облачные технологии», хотя точное его определение ещё более затруднительно. В статьях отечественных и зарубежных авторов под облачными технологиями в библиотеках понимают достаточно широкий круг информационных технологий - от традиционных http, ftp, e-mail [1] до grid, виртуализации и распределенных систем [2]. Попытаемся разобраться в деталях технологий, достоинствах и недостатках их применения, а также практических аспектах работы с ними.

На заре библиотечной автоматизации одним из базовых требований к корпоративным проектам являлась поддержка сетевых протоколов Z39.50, а системы, построенные на их основе, представляют собой пример grid-вычислений. Обратимся к определению: «GRID – 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)» [3-4]. Основополагающим принципом является распределенное хранение данных, поддержку и доступ к которым обеспечивает организация-хозяин ресурсов на основе оговоренных стандартов и взаимно согласованных внутри системы правил. Пользователю обеспечивается доступ к данным через единый пользовательский интерфейс, с помощью которого он (пользователь) получал некоторый виртуальный информационный ресурс, воспринимаемый как единая база данных. Систему распределенных библиотечных каталогов можно считать гетерогенной средой лишь условно, имея в виду только лишь различие систем автоматизации библиотек (САБ) в организациях. В отдельных случаях, обеспечивался доступ к истинно гетерогенным данным, таким как музейные коллекции. Поддержка подобных систем была достаточно обременительна для хозяев ресурсов. Она требовала наличия парка вычислительной техники и программного обеспечения, содержания квалифицированного персонала. С ростом объемов данных проявлялись дополнительные проблемы, связанные с сохранностью данных, их архивированием и защитой от внешних атак [5].

Grid не единственные системы, где мы сталкиваемся с виртуализацией. Широко известный пример – виртуальные машины на локальном компьютере, предназначенные для разделения операционной среды и приложений. Например, LINUX на Windows-машинах или несколько виртуальных машин на мощном сервере, предназначенных для работы различных пользователей в режиме тонкого клиента. Проникновение виртуализации в Интернет на фоне стремительного развития web-сервисов привели к возникновению концепции облачных вычислений. Имеется много определений этого понятия, самое распространенное из которых гласит: «Облачные вычисления 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» [6-7]. Исторически 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 облаке используются grid-решения. Это отмечают множество публикаций, в том числе и посвященных сравнительному анализу технологий [7-8]. При разделении grid и облачных систем, по-видимому имеет смысл руководствоваться главным отличием: grid – распределенная система, самостоятельной частью которой может быть и ваша библиотека, облако – система централизованная, работающая по непубличным правилам и имеющая стороннего хозяина, как правило, провайдера услуг. По мнению некоторых авторов публикаций облако – это коммерческий вариант grid [8]. Только очень крупные и богатые организации могут позволить себе владеть собственным сколько-нибудь большим облаком. Тем не менее, для организаций работа в облаке представляется весьма желанной и экономной. Можно не держать сложную и дорогостоящую вычислительную технику, а также квалифицированный персонал для ее обслуживания. Предоставляемые провайдерами услуги, как правило, дешевле, чем совокупные траты на покупку и содержание парка вычислительной техники. Библиотеки легко отказываются от grid-систем и переходят в облако, не всегда трезво осознавая последствия и ясно оценивая перспективы.

Попробуем выделить области сегодняшнего использования облачных технологий в библиотеках в терминах моделей услуг облачных вычислений, а затем оценить их с точки зрения достоинств и недостатков облачных систем.

**PaaS (платформа как услуга)**. Использование вычислительной техники (платформы и хранилища). 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 России сейчас для доступа к такому облаку используется чаще всего удаленный рабочий стол. Каковы преимущества такого облака:

1. Доступность – подключиться можно из любой точки мира, где есть Интернет;
2. Гибкость – неограниченность вычислительных ресурсов за счет виртуализации;
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 этой ситуации библиотека решает, пожалуй, самую дорогостоящую для нее проблему – проблему хранилища данных. Большинство малых и средних библиотек просто не имеют необходимых денежных средств для покупки вычислительной техники, обеспечивающей решение этой задачи. Провайдер берет на себя обязанность хранения, архивации и восстановления данных после сбоев, причем это касается не только каталогов и коллекций, но и прикладного программного обеспечения (ПО). Не лишена такая модель и недостатков:

1. Требуется постоянное соединение с сетью. На случай аварии необходимо держать резервную копию всего комплекса.
2. В силу урезанных прав администрирования операционной системы (ОС) пользователь имеет ограничения в используемом ПО и не всегда имеет возможность настроить его под свои собственные цели. Зачастую пользователь ограничен также в возможности исследования некоторых характеристик работающего ПО и лишен права оптимизации его работы.

Использование данного метода работы перспективно, широко используется в зарубежных библиотеках, но в России в настоящее время достаточно ограничено. В библиотеках, которые работают в таком режиме, это либо работа в тестовом режиме в рамках исследовательского проекта, либо корпоративное коммунальное облако в локальной сети организации. Оба этих варианта не предполагают четко определенной платы за услуги, поэтому пользователям достаточно затруднительно определить реальную цену такой работы в случае, если вдруг придется перейти на взаимные платежи. Любопытно, что аналогичные вопросы возникают и у зарубежных библиотек. Может быть, поэтому авторы некоторых публикаций рекомендуют для оценки реальных финансовых затрат использовать цены крупных вендоров Интернет (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. |

Если мы попытаемсяпосчитать стоимость услуг по этим расценкам, то получим весьма немалые деньги, которые, конечно, существенно меньше, чем траты на покупку и обслуживание оборудования для систем хранения и архивации данных. Однако и не так малы, чтобы такие траты могла себе позволить небольшая библиотека без значительной финансовой поддержки со стороны. Учитывая то, что с ростом объемов электронных ресурсов, аренда сторонних хранилищ становится насущной необходимостью каждой библиотеки, проблемы ее финансирования требуется решать на государственном уровне.

**SaaS (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 корпоративных библиотечных систем. Стоимость аренды программного обеспечения в этой модели обычно выше, чем плата за стандартное сопровождение собственного ПО, но существенно ниже, чем его покупка. Ежегодная оплата составляет приблизительно от 30 % до 50 % стоимости покупки ПО, в то время как сопровождение стоит примерно 11-15 %. Расширенный набор услуг, включающий адаптацию и доработку ПО для нужд конкретной библиотеки, предоставляется за дополнительную плату. Достоинства:

1. Те же, что и в предыдущей модели.
2. Возможность сэкономить не только на покупке вычислительной техники, но и 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 (центр дистанционного обслуживания) и бесплатный круглосуточный телефон.

Однако, перечисленные достоинства имеют свою оборотную сторону, составляющую недостатки модели:

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.
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 таком режиме экономия средств составляет примерно 70 %, то уже на третий год она падает до 47 %. А если пользователям требуется адаптация системы, то экономии может и вовсе не случиться.

**IaaS (инфраструктура как сервис).** Такая модель облака, хотя и существует в зачаточном состоянии внутри представленной выше модели SaaS в виде CRM, связывают преимущественно с набором услуг крупных вендоров в Интернет. Библиотеки широко используют эту инфраструктуру, являясь своего рода экспертами в этой области. Все широко известные брэнды Интернет есть по сути облачные технологии. Социальные сети, онлайновые 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 этими системами. Часть услуг, предоставляемых этими системами, оплачивают рекламодатели, что делает эти услуги еще более привлекательными для пользователей. Так 70% библиотек-пользователей САБ ИРБИС в Красноярске 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 пользователей. Достоинства инфраструктуры облачных систем Интернет хорошо известны. Но есть и недостатки, отраженные во множестве публикаций [4, 5, 9]:

* отсутствие интероперабельности – нет универсальных стандартов и интерфейсов, что увеличивает зависимость от поставщика (попробуйте, например, перенести вашу почту из mail.ru на google.com или ваш корпоративный почтовый сервер);
* постоянное соединение с сетью - для получения доступа к услугам облака необходимо постоянное соединение с сетью Интернет;
* программное обеспечение и его кастомизация – есть ограничения по ПО, которое можно разворачивать в облаках и предоставлять пользователю. Пользователь ПО имеет ограничения в используемом ПО и зачастую не имеет возможности настроить его под свои собственные цели. Также некоторые поставщики ограничивают выбор языковых пакетов;
* сохранение конфиденциальности – конфиденциальность данных, хранимых на публичных облаках, в настоящее 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;
* обеспечение надёжности и безопасности. В отношении надежности хранения информации, можно с уверенностью сказать, что если вы потеряли информацию, хранимую в облаке, то вы ее потеряли навсегда, что же до безопасности – облако само по себе является достаточно надежной системой, однако при проникновении в него злоумышленник получает доступ к огромному хранилищу данных.

Литература

1. 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. - c. 37-41. - URL: <http://computerresearch.org/stpr/index.php/gjcst/article/viewFile/860/765>.
2. 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). – c. 270-273. – URL: http://www.ijidt.com/index.php/ijidt/article/viewFile/3.4.9/pdf.
3. Интероперабельность в облачных вычислениях [Текст] / Е. Е. Журавлев [и др.] // Журнал радиоэлектроники. – 2013. - № 9. – c. 1-63. – URL: <http://razinkin.16mb.com/publications/clouds>.
4. Хрусталев, Е. Ю. Методический подход к проектированию сервисов упрощенной интеграции распределенных IT-ресурсов [Текст] / Е. Ю. Хрусталев, А. А. Чумичкин / Информационные ресурсы России. – 2012. – № 3. – с. 2-6.
5. Ковязина, Е. В. Перспективы развития автоматизации библиотек [Текст} / Е. В. Ковязина // Научные и технические библиотеки. – 2011. - № 2 – с. 89-92.
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 // ARPN Journal of Systems and Software. – 2012. - Vol. 2, № 5. – c. 188-194. – URL: <http://scientific-journals.org/journalofsystemsandsoftware/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.
9. 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>.
10. Емельянов И. Миф о дешевизне облачных решений [Текст] [Электронный ресурс] / И. Емельянов // Компьютерра. – 2013. - № 10. – URL: <http://www.computerra.ru/cio/5574>.