

Библиотеки в «облаках»: практические аспекты

Доклад посвящён использованию grid и облачных технологий в библиотеках российских регионов. Даны базовые определения grid и облачных вычислений и некоторые сравнительные характеристики. Распространённые приёмы работы приведены в соответствии с моделями услуг облачных вычислений. Выделены их достоинства и недостатки. Представлены практические рекомендации по работе в каждой из этих моделей.

Ключевые слова: облачные технологии, облачные вычисления, grid-системы, платформа как услуга, программное обеспечение как услуга, инфраструктура как сервис.

Развитие телекоммуникационной и информационной инфраструктур, стремительное совершенствование компьютерной техники, включая персональные устройства, и экспоненциальный рост объёмов доступной цифровой информации обусловили проникновение новых технологий в сферу традиционной библиотечной автоматизации. Проблема неразрешимости ряда задач, стоящих перед библиотечными специалистами, в рамках традиционных систем автоматизации библиотек (САБ) отчётливо проявилась несколько лет назад [1]. С течением времени проблемы только обострились.

Растущие объёмы данных и количество обращений, потребность в новых сервисах и т.д. обусловили необходимость применения новейших компьютерных технологий в библиотеках. В частности, всё чаще упоминаются облачные вычисления применительно к библиотекам. Причём в публикациях под облачными вычислениями понимают достаточно широкий круг технологий: от традиционных *web, ftp, e-mail* [2] до достаточно новых *grid*, виртуализации и распределённых систем [3]. Попробуем разобраться в особенностях этих технологий, их достоинствах и недостатках, а также в практических аспектах работы с ними.

На начальных этапах библиотечной автоматизации одним из базовых требований была 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 распределённой, гетерогенной среде, которая поддерживает коллекции пользователей и ресурсов (виртуальных организаций) в традиционных административных и организационных доменах (реальных организаций)» [4, 5]. Основопологающим принципом является распределённость данных, управление которыми, поддержку которых 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 распределённых 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 были достаточно обременительны для хозяев данных. Требовались вычислительная техника и программное обеспечение (ПО), квалифицированный ИТ-персонал. С ростом объёмов данных проявлялись дополнительные проблемы, связанные с их сохранностью, архивированием и защитой от вирусных и хакерских атак.

Grid – это не единственные системы, где присутствовала *виртуализация*, в приведённом 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 веб-сервисов привело 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 веб-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 веб-сервисы, а в облачных технологиях используются *grid*-решения. Это отмечено в ряде публикаций, в том числе и посвящённых сравнительному анализу обеих технологий [6, 7].

При разделении *grid* и облачных систем, по-видимому, имеет смысл придерживаться того, что *grid* – распределённая 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 определённого 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 её 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; ПО использует собственное. Потребитель при этом не управляет сетями, серверами, операционными системами (ОС) и системами хранения данных (базовой инфраструктурой «облака»), но осуществляет контроль над развёрнутыми приложениями и, возможно, некоторыми параметрами конфигурации среды хостинга. То есть на вычислительной технике провайдера устанавливаются САБ, веб-сервер и сайт библиотеки, вспомогательное ПО. В российских библиотеках для доступа к такому «облаку» чаще всего используется удалённый рабочий стол.

Достоинства такой облачной среды:

Доступность – подключиться можно из любой точки мира, где есть Интернет;

Гибкость – неограниченность вычислительных ресурсов за счёт виртуализации, лёгкость масштабирования и администрирования;

Надёжность – 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 неё проблему хранилища данных. Провайдер берёт на себя обязанность хранения, архивации и восстановления данных после сбоев, причём это касается не только каталогов и коллекций, но и ПО. В повседневной работе создаётся ежедневная копия данных, что обеспечивает высокую надёжность модели.

Однако стоит упомянуть и о недостатках:

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

На современном этапе использование такой модели в библиотеках носит достаточно ограниченный характер. Она работает в тестовом режиме либо в каком-либо проекте, либо в корпоративном коммунальном «облаке». Поэтому пользователям бывает достаточно трудно определить реальную цену такой работы, хотя бы гипотетически. По-видимому, для оценки предполагаемых финансовых затрат можно рекомендовать таблицу цен крупных вендоров Интернета, которые приведены в некоторых публикациях (см. табл.) [9].

Расценки наиболее крупных облачных вендоров (в долларах США)

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

SaaS (Software as a Service – программное обеспечение как услуга). При таком способе работы библиотека не покупает специализированное ПО (например, САБ) полностью или частично, что позволяет ей 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.

Такая 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 собственного ПО, но существенно ниже, чем стоимость его приобретения. Расширенный набор услуг, включающий адаптацию и доработку ПО под

нужды конкретной библиотеки, предоставляется за дополнительную плату. Достоинства такого варианта:

1. Те же, что и в предыдущей модели;
2. Возможность сэкономить на покупке вычислительной техники, ПО и IT-персонале;
3. Качество сопровождения, как правило, существенно выше за счёт 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 счёт 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 – около 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 не быть.

IaaS (Infrastructure as a Service инфраструктура) как сервис. Такую модель «облака» связывают преимущественно с набором услуг крупных вендоров в Интернете, 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 бренды Интернета есть, по сути, облачные технологии – социальные сети, онлайн-информационные ресурсы, индексы научного цитирования, агрегация информационных ресурсов и т.д. 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 услуги бесплатными и ещё более привлекательными для пользователей. Так, 70% наших клиентов САБ ИРБИС используют для обмена информацией электронную почту mail.ru, а не собственные почтовые серверы – их просто нет на всех уровнях 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 Интернета хорошо известны. Но есть и недостатки, отражённые во многих публикациях [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 Интернетом. Однако в наше время это не такой и большой недостаток, особенно при использовании мобильного соединения;

3. ПО и его кастомизация – пользователь имеет ограничения в работе с ПО, иногда весьма существенные, и не всегда имеет возможность настроить его под свои цели; 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 не рекомендуют хранить наиболее ценные для компании документы на публичном «облаке», так как в настоящее время нет технологии, которая бы гарантировала стопроцентную конфиденциальность хранимых данных;
5. Обеспечение надёжности и безопасности – если вы потеряли информацию, хранимую в «облаке», то вы её утратили 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 определённая 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 работе 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 вначале 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 веб-услуг.

Список источников

1. **Ковязина Е. В.** Перспективы развития автоматизации библиотек / Е. В. Ковязина // Науч. и техн. б-ки. – 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. – Режим доступа: <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. – Режим доступа: <http://www.ijidt.com/index.php/ijidt/article/viewFile/3.4.9/pdf>
4. **Интероперабельность** в облачных вычислениях [Текст] / Е. Е. Журавлев [и др.] // Журн. радиоэлектроники. – 2013. – № 9. – С. 1–63. – Режим доступа: <http://razinkin.16mb.com/publications/clouds>

5. **Хрусталеv Е. Ю.** Методический подход к проектированию сервисов упрощенной интеграции распределенных ИТ-ресурсов / Е. Ю. Хрусталеv, А. А. Чумичкин // Информ. 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. – Режим доступа: <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. – С. 188–194. – Режим доступа: http://scientific-journals.org/journalofsystemsandsoftware/archive/vol2no5/vol2no5_4.pdf

8. **Ильин В. А.** Больше данных, хороших и разных! / В. А. Ильин, В. Е. Велихов // В мире науки. – 2014. – № 2. – С. 38–44.

9. **Новиков И.** Облачные вычисления: на пороге перемен [Электронный ресурс] / И. Новиков // PC Magazine/RE. – 2011. – № 4. – Режим доступа: <http://www.pcmag.ru/solutions/detail.php?ID=44441>

10. **Емельянов И.** Миф о дешевизне облачных решений [Электронный ресурс] / И. Емельянов // Компьютерра. – 2013. – № 10. – Режим доступа: <http://www.computerra.ru/cio/5574>