

Композитные вычислительные облака для мобильных платформ

Д.Ж. Ахмед-Заки¹, А.Ж. Акжалова², Б.А. Кумалаков¹

¹Казахский национальный университет имени аль-Фараби, г. Алматы, РК

²Казахстанско-Британский технический университет, г. Алматы, РК

e-mail: darhan_a@mail.ru, b.kumalakov@gmail.com

Аннотация

Целью данной статьи является определение отличительных свойств облачных вычислительных платформ для мобильных устройств и предложение метода автоматизированного формирования вычислительного облака с использованием SLA соглашений. В статье рассмотрены основные концептуальные характеристики мобильных устройств, технологические и коммуникационные предпосылки, а также основные требования к содержанию системных сообщений.

В условиях постоянно развивающегося рынка компьютерных услуг встает вопрос об ограничениях, связанных с возможностями отдельно взятых вычислительных машин. На данный момент существуют несколько основных подходов к решению данной проблемы, самым перспективным из которых является организация облачных вычислительных систем. Основная идея данного подхода заключается в (1) повышении эффективности использования ресурсов за счет динамического наращивания и снижения производительной мощности отдельно взятой системы; (2) возможности закупки дополнительных вычислительных мощностей без вмешательства в существующую физическую архитектуру системы.

Однако реализованные на данный момент облачные платформы ориентированы на мощные серверные машины, в то время как динамично развивающийся рынок мобильных устройств остается относительно свободной нишей. Основные игроки рынка, такие как корпорации Google и Microsoft, предоставляют возможность потребления облачных услуг посредством мобильных телефонов. Так, по данным официального сайта Microsoft мобильная платформа Windows Azure Toolkit for Windows Phone 7 позволяет использовать вычислительные и программные ресурсы облачной инфраструктуры при условии, что само облако создано на платформе Microsoft Azure [1]. Таким образом, при написании мобильных приложений программист имеет возможность использовать сервисы облака, такие как аутентификация, авторизация и т.д., путем обмена HTTP сообщениями между сотовым телефоном и интерфейсом облака.

Основной соперник Microsoft корпорация Google также предоставила возможность потреблять некоторые услуги Google App с сотового телефона при использовании ее операционной системы Android [2].

Однако отсутствие интеграции мобильного устройства с самой ИТ инфраструктурой в обоих случаях позволяет определить мобильное устройство как клиентское, в то время как интеграция с облачной инфраструктурой подразумевает не только потребление, но также и предоставление собственных ресурсов облаку.

В данной статье авторы предлагают альтернативный подход к облачным вычислениям на мобильных устройствах. В основе авторского подхода находятся принципы

всеобщего разделения ресурсов всеми участниками облачной инфраструктуры и возможности участия в ней любых вычислительных устройств.

Для разработки авторского подхода необходимо выявить принципиальные требования и особенности облачных платформ для мобильных устройств; псевдо-сценарий анализа необходимости подключения и отключения вычислительных устройств; сформировать схему процесса подключения и отключения устройств.

Рассмотрим сущность понятия «облачные вычисления». Основной отличительной чертой данной технологии являются автоматическое регулирование объема потребляемых вычислительных ресурсов, быстрота приспособления к требованиям пользователя и мультитенантность.

Под автоматическим регулированием объема потребляемых вычислительных ресурсов понимается «своевременное наращивание и уменьшение мощностей в зависимости от потребностей процессов, расположенных в облаке» [3]. Таким образом достигается экономическая выгода от использования только тех ресурсов, которые реально необходимы для решения задач. По данным исследования, проведенного корпорацией VMWare, коэффициент полезного действия (КПД) серверного оборудования составляет лишь 10-30%, в то время как КПД рабочих станции всего 5% [4]. Иными словами, бесполезно расходуется до 95% потребляемых ресурсов, в том числе электроэнергия, людские и интеллектуальные ресурсы, затраты на аренду и содержание помещения и т.д. Вычислительные облака же предоставляют возможность отключать незадействованные машины и подключать выключенные, по мере востребованности ресурсов во время работы программ.

Быстрота приспособления облачных систем означает способность при необходимости изменять виртуальную ИТ инфраструктуру в кратчайшие сроки. Данный функционал основан на принципе создания, изменения и утилизации виртуальных машин с предустановленным и сконфигурированным программным обеспечением по запросу пользователя. Основной принцип в основе данного подхода – разделение виртуального (логического) и физического уровня ИТ. То есть, логическая структура компьютерной сети не должна зависеть от физических машин.

Мультитенантность – способность пользователей облачной инфраструктуры иметь доступ ко всем данным и приложениям в пределах облака [5]. Данная концепция предоставляет преимущества распределенного хранения данных и распределенной обработки процессов с возможностью одновременного использования нескольких стандартов и видов платформ [5, 8].

Названные характеристики в той или иной степени присущи всем облачным платформам в виде набора облачных услуг, называемых «все как услуга». Проанализировав несколько источников [5, 6, 7], мы пришли к следующему заключению: классические услуги облачных систем, в общем обозначаемые как «все как услуга», обозначают динамическое и автоматизированное предоставление вычислительных ресурсов во время выполнения задач облачной системой. Различие между типами услуг заключается в уровне контроля, предоставляемого пользователю, который варьируется от управления настройками приложения до управления виртуальной сетевой архитектурой.

Сами же облачные системы, в зависимости от уровня общественного доступа к ним, делятся на три основных типа – общественные, частные и гибридные.

Частное облако – облачная инфраструктура, доступ к которой имеет лишь замкнутый

круг пользователей [9]. Чаще всего данный вид вычислительного облака располагается в пределах одной организации или одного здания, то есть имеет географический или социальный характер. Среди достоинств данного подхода можно выделить защищенность и полный контроль группы пользователей над ИТ инфраструктурой. Недостатки же включают невозможность потребления дополнительных ресурсов извне организации и необходимость инвестирования в оборудование, а также обучение сотрудников.

Общественное облако предоставляет доступ к ресурсам всем пользователям, подключенным к его компьютерной сети [9]. На данный момент существуют риски хищения данных при использовании общественных облаков по той причине, что получив доступ к облаку, хакер имеет доступ абсолютно ко всей информации. С другой стороны, общественное облако чаще всего обслуживается сторонней организацией (вендор) и его использование не требует дополнительных инвестиции в оборудование, обучение персонала, затрат на аренду помещений и т.д. Более того, в данном сценарии реализуема возможность закупки вычислительных услуг сторонних облачных систем, таких как Google App Engine или Amazon Elastic Compute Cloud [10, 11].

Третий вид облачной инфраструктуры - гибридное облако является средним между общественным и частным облаками. Круг пользователей данной инфраструктуры жестко ограничен по определенному признаку, но при этом предоставляется возможность обмениваться ресурсами с иными облаками. Ключевым свойством данной технологии является четкое распределение прав доступа пользователей в пределах взаимодействующих облаков.

В разрезе мобильных устройств облачные платформы имеют некоторые ограничения:

Во-первых, облачные вычислительные инфраструктуры предоставляют возможность экономить на сокращении потребляемых компьютерами ресурсов. Мобильные же устройства, в силу своей компактности, не отличаются высокой производительностью, а наоборот часто нуждаются в дополнительных мощностях. Таким образом, для мобильных устройств более актуально потребление вычислительных мощностей иных устройств, чем сбережение собственных ресурсов.

Во-вторых, мобильные вычислительные устройства не имеют постоянного географического расположения. Таким образом, классическая инфраструктура управления группой вычислительных машин, состоящая из главных и управляемых узлов, нуждается в пересмотре.

В результате, облачная платформа для мобильных устройств должна быть основана на:

- использовании высокоскоростных беспроводных технологий обмена данными;
- способности каждого мобильного устройства принимать роль подчиненного и управляющего узла в зависимости от ситуации;
- способности каждого устройства сохранять работоспособность при объединении ресурсов с облаком.

Безусловно, при проектировании облачных платформ для мобильных устройств необходимо учесть скорости существующих беспроводных интерфейсов обмена данными, при этом уделяя внимание их распространенности и перспективности в реалиях Республики Казахстан. На данный момент наиболее развитыми в Казахстане технологиями межтелефонной передачи данных являются GSM, 3G, Bluetooth и передача по средствам инфракрасных портов (ИК). По данным различных источников наиболее медленной

технологией передачи данных является GSM (скорость передачи 10 Кб/с [12]), за ней идет Bluetooth (2,1 Мб/с) [13] в то время как скорость передачи данных 3G составляет 3,2 Мб/с [14]. Наиболее быстрым методом передачи данных с мобильных устройств является ИК порт, способный транслировать данные на скорости до 1 Гб/с [15]. Однако данная технология является не столь распространенной по той причине, что для использования ИК портов мобильные устройства должны не только быть в зоне видимости друг от друга, но и оставаться в статичной позиции относительно друг друга.

Вторым немаловажным критерием мобильной облачной платформы является концептуальный отход от выделения определенной машины на роль управляющей в инфраструктуре вычислительного облака. Классический подход, изображенный на Рисунке 1, предполагает установку программного обеспечения управляющих звеньев на отдельные машины, специально сконфигурированные для данной деятельности.



Рис. 1: Классическая облачная инфраструктура

Машины в мобильном облаке принимают любую роль в зависимости от того, кем запущен виртуальный сервис и какую смысловую нагрузку он в себе несет. Рассмотрим данный принцип на примере. Допустим, что мобильному устройству «А» необходимо выполнить ресурсоемкую задачу и оно находится в радиусе действия иных мобильных устройств, готовых поделиться ресурсами. Устройство «А» посылает запрос о формировании вычислительного облака другим устройствам. Устройства, готовые поделиться частью вычислительных ресурсов, отправляют ответ, в котором содержится информация об источнике и объеме ресурсов, которые он может предоставить в данный момент. Далее устройство, пославшее запрос, берет на себя роль управляющего звена, то есть получает право распределять задачи и нагрузку, регулировать последовательность и порядок обмена данными между остальными устройствами в облаке. Происходит процесс формирования общего соглашения устройств о предоставлении ресурсов, иерархии устройств в облаке и формировании путей доступа к ресурсам. С данной целью создается SLA соглашение – файл, содержащий все ограничения, реквизиты вычислительных устройств и иные параметры, регулирующие взаимодействия сторон. В отличие от классических облачных систем, в данном сценарии отсутствует свод логических ограничений QoS. Однако его функции (определение требований к скорости соединения,

пропускной способности соединений и т.д.) учтены запрашивающим устройством во время формирования SLA, основываясь на ограничениях, заданных процессом, для которого формируется облако.

Иными словами, в данной модели облачные вычисления носят недолговременный характер, то есть облако формируется для выполнения определенного процесса. В данном подходе отсутствуют долговременные требования QoS, так как ограничения диктуются выполняемым процессом.

Для определения необходимости формирования вычислительного облака необходимо ввести понятия «достаточности» и «равновесия» ресурсов. На Рисунке 2 изображен график спроса на ресурсы приложениями, расположенными в вычислительной инфраструктуре (D), в течение некоторого времени.

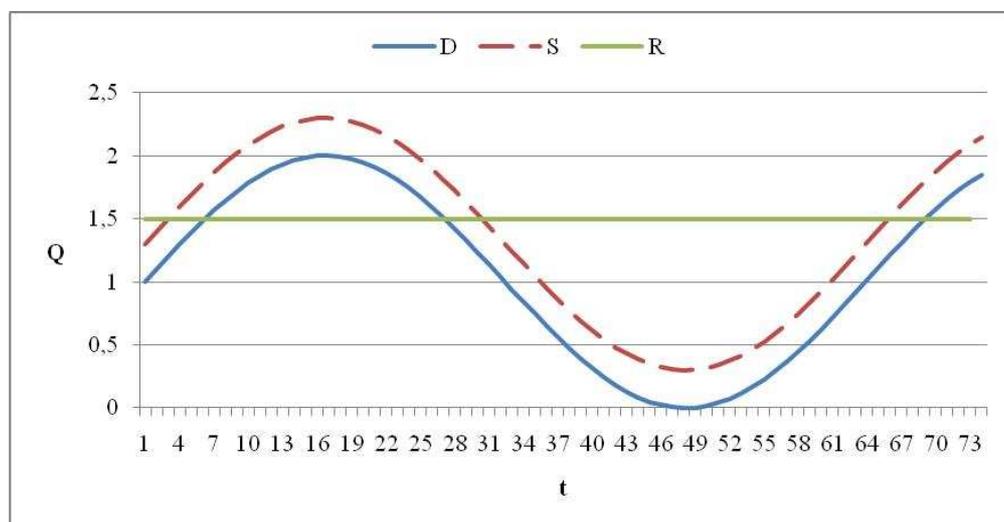


Рис. 2: Спрос на ресурсы в вычислительной инфраструктуре

Достаточность ресурсов в данном контексте означает способность вычислительной системы предоставить объем ресурсов, необходимый для полноценного функционирования приложений, расположенных в ней. В данном случае ресурсы достаточны во всех точках, где спрос на ресурсы ниже уровня производительности системы ($D < R$). В ином случае ($D > R$), она должна заблаговременно попытаться сформировать облако. С целью выполнения данного вида прогнозирования, мы вводим понятие равновесия ресурсов (S), которое может быть логически описано в виде:

$$(\neg(S \gg D) \mid S \geq D \wedge S = D + k)$$

где, k – фактор надежности системы. Иными словами, точка равновесия – это верхний предел ожидаемой вариации спроса на вычислительные ресурсы, расчет которого основан на статистических данных. Необходимость формирования вычислительного облака должна определяться приближением или пересечением лимита ресурсов вычислительной системы равновесием ($S \geq R$).

Определив возрастающую потребность в вычислительных ресурсах, устройство или облако нуждается в подключении дополнительных устройств к его архитектуре. Схема

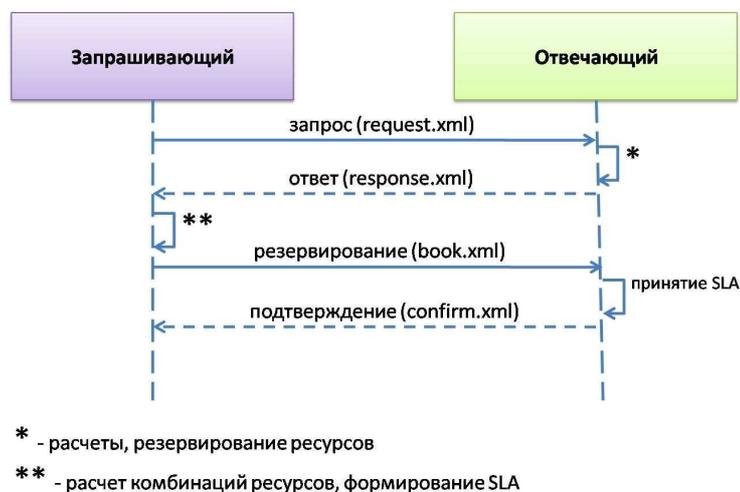


Рис. 3: Процесс подключения устройства к облачной инфраструктуре

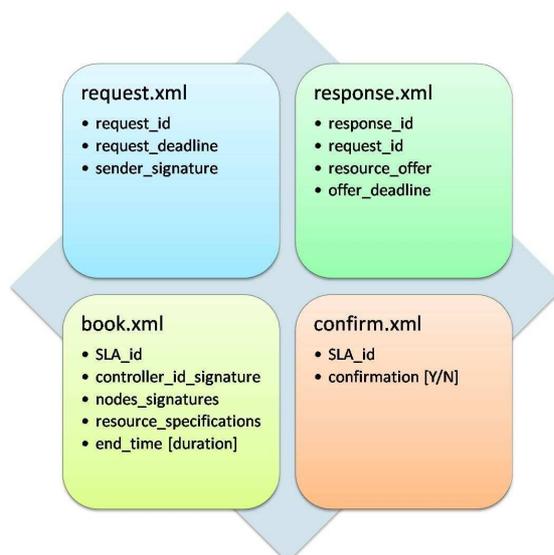


Рис. 4: Структура «.xml» файлов

процесса подключения устройств к облачной инфраструктуре изображена на Рисунке 3. Центральную роль в данном процессе занимает SLA соглашение – файл с расширением «.xml», который содержит спецификацию ресурсов и сервисов, доступных участникам облака. Он формируется управляющим звеном на основании данных, полученных в ходе обмена сообщениями, описанными в Рисунках 3 и 4.

Из предложенной структуры файлов видно, что сигнатура каждого устройства используется не только для обозначения устройства, но и для определения его роли в структуре вычислительного облака (например, «root_signature» обозначает управляющее звено). Спецификации выделяемых устройствами ресурсов даются в виде описаний ссылок на сервисы и процессы.

Немаловажно, что выделяемые устройствами ресурсы должны быть отделены от основного резерва, таким образом обеспечивая работоспособность самого устройства по мере участия в облачной инфраструктуре.

После того, как SLA соглашение сформировано, управляющее звено распределяет процессы согласно спецификации ресурсов. Важно отметить, что данная модель предусматривает расформирование вычислительного облака по окончании процесса, для выполнения которого оно было сформировано. Данное решение обусловлено тем, что мобильные облака, по определению, приведенному в начале статьи, формируются для выполнения определенной задачи.

В данной статье представлена модель облачной платформы для мобильных устройств, таких как сотовые телефоны и PDA устройства. Она позволяет интегрировать устройства в инфраструктуру облака, тем самым позволяя не только потреблять, но и обмениваться вычислительными ресурсами с учетом их особенностей.

Список литературы

- [1] *Codeplex*. <http://watoolkitwp7.codeplex.com/>, по состоянию на: 17/06/2011
- [2] *Google*. <http://code.google.com/intl/ru-RU/android/c2dm/index.html#arch?>, по состоянию на: 17/06/2011
- [3] *Dubey A., Wagle. D.* Delivering Software as a Service. – The McKinsey Quarterly, May 2007, 1-12 p.
- [4] *VMWare*. Addressing Desktop Challenges. – VMWare, 2008, URL: http://www.vmware.com/pdf/vdi_solution.pdf, по состоянию на: 11/01/2011
- [5] *Федоров А., Мартынов Д.* Microsoft Azure: Облачная платформа Microsoft. – Москва: Microsoft, 2010.
- [6] *Buyya R., Yeo C.S., Venugopal S., Broberg J., Brandic I.* Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility – Future generation computer systems, 2008
- [7] *Desprez F., Kramer-Fuhrman O., Yahyapour R.* Cloud computing: Introduction o the Special Theme. – ERCIM News (83), 2010, 12-13 p.
- [8] *Ohshima, H., Oyama, S., Tanaka, K.* Cloud as Virtual Databases: Bridging Private Databases and Web Services. – DASFAA 2010, Part I, LNCS 5981, 2010, 491– 497 p.
- [9] *Srinivasa R., Nageswara R., Ekusuma K.* Cloud computing: An overview. - Journal of Theoretical and Applied Information Technology Vol.9 (1), 2009, URL: <http://www.jatit.org/volumes/research-papers/Vol9No1/10Vol9No1.pdf>, по состоянию на: 25/04/2011
- [10] *Amazon*. Amazon Elastic Compute Cloud. – 2011, URL: <http://aws.amazon.com/ec2/>, по состоянию на: 20/04/2011
- [11] *Google*. Google App Engine. URL: <http://code.google.com>, 2011, по состоянию на: 20/04/2011

- [12] ROADCOM Ltd. Протоколы сотовой связи. – 1999, URL: <http://www.protocols.ru/files/Protocols/Cellular.pdf>, по состоянию на: 20/04/2011
- [13] *Нечай О.* Bluetooth в мобильном телефоне. – 2006, URL: <http://www.computerra.ru/terralab/peripheral/270061/>, по состоянию на: 20/04/2011
- [14] Wikipedia. 3G. - 2011, URL: <http://ru.wikipedia.org/wiki/3G>, по состоянию на: 20/04/2011
- [15] *Nemo K.* Новая беспроводная технология Giga-IR. – 2008, URL: <http://www.htpc.ru/news/31097>, по состоянию на: 20/04/2011

Akhmed-Zaki D.Zh., Akzhalova A.Zh., Kumalakov B.A., Composite computing clouds on mobile platforms, The Bulletin of KazNU, ser. math., mech., inf. 2011, №3(70), 49 – 56

The aim of the article is to define unique features of the mobile devices' cloud computing platform and offer method of automatic mobile cloud organization using service level agreements (SLA). The article contains an overview of main conceptual differences of mobile devices, technological and communication prerequisites and general structure of the system messages.

Д.Ж. Ахмед-Заки, Ә.Ж. Ақжалова, Б.А. Құмалақов, Мобильді платформаларда композитті бұлтты есептеу, ҚазҰУ хабаршысы, мат., мех., инф. сериясы 2011, №3(70), 49 – 56

Ұсынылып тұрған жұмыстың мақсаты - икемді құрылғылар үшін бұлтты есептеу кешендерінің ерекшеліктерін анықтап, SLA келісімшарт негізінде есептеу бұлттарының автоматты түрде қалыптасуының үлгісін ұсыну. Мақалада икемді құрылғылардың тұжырымдамалық негіздері, бұлтты есептеуге керекті технологиялық және байланыс алғышарттары мен жүйелік хабарламаларға қойылатын негізгі талаптар айтылып кеткен.