

**Особености на използването на
виртуализирани разпределени компютърни ресурси**

Кирил Боянов, Димитър Тодоров, Христо Турлаков

Институт по информационни и комуникационни технологии – БАН
ул. "Акад.Г.Бончев" бл.25 А, 1113 София

Abstract: The paper describes some Cloud computing architecture problems related with its realization and exploitation. These problems are from economical, security and performance domain. Its goals are better understanding of the Cloud computing properties, advantages and trends and how universal it could be in deployment of existing and future applications.

Keywords: Information and Communications Technologies, Cloud computing, Grid Computing, Distributed Computing, Computer Systems, Telecommunications, Networks

1. Въведение

Характерно за настоящия етап от развитието на информационните технологии е наличието на множество мрежово добре свързани изчислителни ресурси, които се използват на практика самостоятелно и изключително неефективно – с натовареност често под 2-3% [1]. Това води до естествената идея за централизиране и рационализиране на тяхната експлоатация чрез виртуалната им интеграция и осигуряване на достъп на потребителите до компютърни услуги, без те да се интересуват какви реални ресурси са ангажирани с тяхното изпълнение.

Формата за реализация на тази идея днес е използването на концепцията на *Cloud computing*. Същността на тази концепция за организиране и използване на компютърна среда [2] се свежда най-общо до **ползване на виртуални, разпределени и динамично предоставяни компютърни ресурси**, което означава че потребителят ползва **виртуална компютърна среда**, която е:

- достъпна чрез добре дефиниран и установен потребителски интерфейс,
- реализирана с помощта на разпределени компютърни структури и
- предоставяна динамично в зависимост от текущите потребителски нужди.

Използването на cloud computing (виртуална компютърна среда) е може би най-голямата промяна в парадигмата на използването на Информационно-Комуникационните Технологии след въвеждането на режима на времеделене през 1960-те години [1].

2. Варианти за ползване

Реализацията на концепцията за ползване на *Cloud computing* се осъществява най-често в някоя от следните форми:

- SaaS (software-as-a-service) – предоставяне на софтуерни приложения, които са достъпни от множество потребители като Интернет услуга.
- PaaS (Platform-as-a-service) – вариант на SaaS, при който се предоставя платформа за разработка на приложения като софтуерна услуга през Интернет.
- IaaS (Infrastructure-as-a-service) – предоставяне на компютърни ресурси като памет и виртуални сървери, които потребителите могат да ползват в обема според техните текущи изисквания.

От гледна точка на начина на ползване на *cloud computing* (*виртуалната компютърна среда*) могат да се определят и понятията *публична, частна и хибридна среда* (*public cloud, private cloud и hybrid cloud*).

При частния *cloud computing* има съвпадение между собственика и потребителя на виртуалната среда. Тъй като това е една и съща организация става възможно оптимизиране на средата към конкретните особености на използването, например постигане на по-специални изисквания за надеждност, информационна сигурност и др. Недостатък е, че не може да се постигне оптимално натоварване на средата и са възможни претоварвания или неефективно използване.

Под публичен *cloud computing* се има предвид среда, при която собственикът и потребителят са различни и която е достъпна без ограничения през Интернет. Предимствата са, че е налице възможност за оптимално натоварване, тъй като са възможни пазарни отношения с потребителите. Проблем е липсата в повечето случаи на контрол върху поведението на потребителите и трудното търсене на отговорност в случай на некоректно поведение. Това води до повишен риск за сигурността на средата и отблъскване на сериозни клиенти с изисквания в тази посока.

При хибридният *cloud computing* нещата са смесени. Собственикът и основен потребител на ресурсите предоставя достъп и на външни потребители, с което оптимизира натоварването на средата, но същевременно повишава и рисковете за пробив в сигурността.

3. Икономически особености

Процесът на интеграция на локално обособени компютърни ресурси въз основа на тяхната мрежова свързаност с цел виртуалното им използване от множество потребители е изключително мащабен и сложен. Икономическите му параметри обаче дават възможност да се реализират редица предимства спрямо съществуващите практики.

Собствениците на компютърни ресурси имат възможността да реализират значителна икономия от мащаба, като ги концентрират в много големи изчислителни центрове. Както се вижда от Таблица 1 [9], съотношението на цените на някои основни ресурси, каквито са мрежовата свързаност и запомнящата среда в среден и в голям изчислителен център, е между 5 и 7 пъти. Същото се отнася и до разходите за тяхното администриране. За да се реализира подобна икономия, е необходимо

концентрирането на ресурсите в огромни изчислителни центрове, за които инвестицията е в размер от няколкостотин милиона до около 1 милиард долара [11]. Тъй като консумацията на енергия от подобен изчислителен център също е огромна и е над няколко десетки мегавата, изборът на местоположението на такъв център в район с евтина електроенергия е източник на допълнително намаление на разходите.

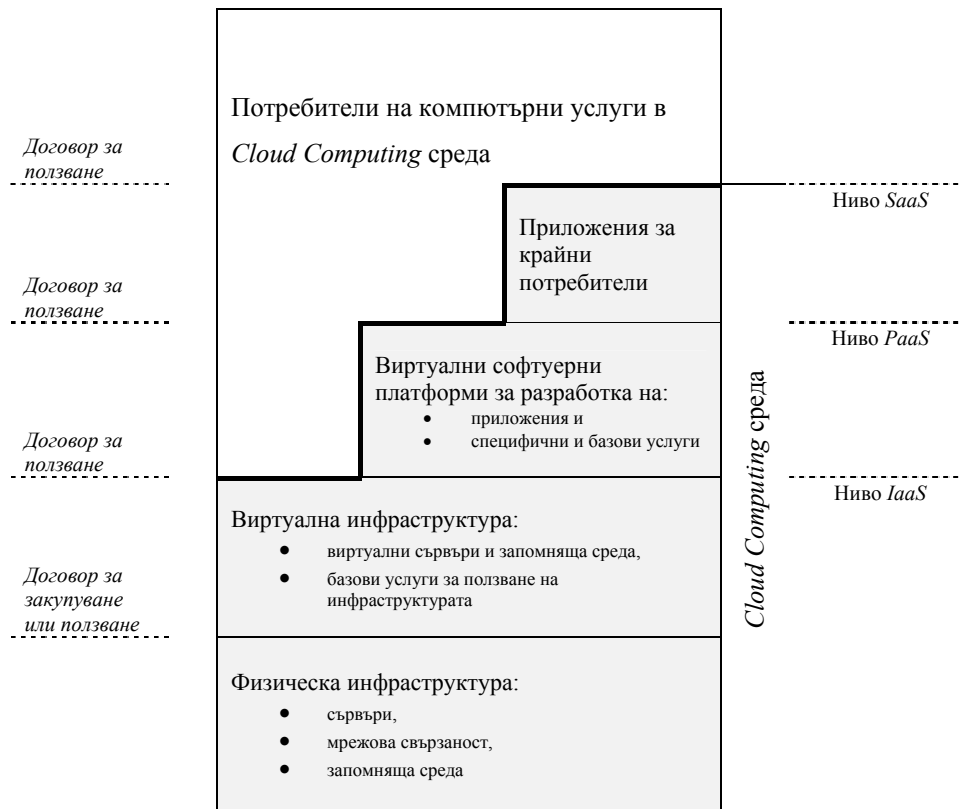
Таблица 1. Влияние на мащаба върху цените на компютърните ресурси.

Вид на ресурса	Средно голям изчислителен център (~1000 сървъра)	Много голям изчислителен център (~50 000 сървъра)	Съотношение
Цена на мрежова свързаност за 1Mbit/s на месец	\$95,00	\$13,00	7.1
Цена на дискова памет за 1Gbyte на месец	\$2,20	\$0,40	5.7
Брой на сървърите, администрирани от един човек	~140	>1000	7.1

Важна особеност на виртуалното използване на компютърни ресурси е, че благодарение на виртуализацията, на масовостта на потребителите и на стремежа да се реализира икономия от мащаба чрез концентриране на ресурсите в големи изчислителни центрове се създават благоприятни условия да няма съвпадение между собствениците и потребителите на компютърните ресурси. В резултат на това разделение става възможно необходимите начални капиталови разходи да се осигуряват от собствениците на ресурсите, които ги предоставят на потребителите срещу наем без да се интересуват в подробности за приложенията, които ще се изпълняват. Съществената разлика от досега съществуващия масов модел за ползване на компютърни услуги е, че става възможно независимо оптимизиране на бизнес моделите за осигуряването на компютърна инфраструктура и тези за нейното ползване. Резултатът за потребителите се свежда до **трансформиране на капиталовите разходи в оперативни**.

Идеята за отделянето на собствениците на компютърни ресурси от техните потребители не е нова, но с технология от типа на Cloud computing тя може да се реализира в мащаби, които позволяват **компютърната инфраструктура да стане стока, продавана на дребно** с ценови модели, напомнящи тези на продажбата на електроенергия, газ и др. и да има много ценната за потребителите масова достъпност, ясно и стабилно ценообразуване и установени характеристики.

Вариантите на ползване на Cloud Computing среда са обобщени на Фиг. 1. От фигурата се вижда, че ползването на такава среда може да стане най-грубо на три нива: SaaS, PaaS и IaaS. На всяко от нивата IaaS и PaaS е възможно да има междинни доставчици на услуги за крайни клиенти на ниво SaaS, които също се явяват и потребители на Cloud Computing среда.



Фиг.1: Използване на *Cloud Computing* среда

Заплащането на виртуалните компютърни ресурси от страна на клиентите въз основа ползването, независимо дали те са доставчици на услуги или крайни потребители, води и до много ценната допълнителна възможност количеството на ресурса да може бързо, лесно и на малки стъпки да се адаптира в зависимост от текущите им изисквания. Тази особеност, наречена **еластичност** [9], означава на практика прехвърляне на риска за точното планиране на необходимите ресурси върху техния доставчик и създаване на илюзия за потребителя, че има достъп до компютърни ресурси с неограничен обем.

Представа за значението на еластичността се обуславя от факта, че необходимостта от компютърни ресурси при предлагането на дадена компютърна услуга естествено минава през различни фази, които са с различни изисквания към техния състав и качество. Фирмите, които създават и внедряват дадена услуга в *Cloud computing* среда, са много улеснени при стартирането на нови инициативи, тъй като рискът и загубите при неуспех са на много по-ниско ниво.

Друг аргумент за важността на еластичността е, че над половината инсталирани в САЩ сървъри са от категорията на малките до средни кълстери [12] и се

характеризират с неравномерна натовареност на ресурсите, която в един по-дълъг период е доста ниска, например под 10%. Именно потребителите, които най-често са и собственици на подобни кълстери са и най-вероятните бъдещи потребители на Cloud computing, тъй като при тях има силен икономически стимул да преминат от закупуване към наемане на ресурс, с което да намалят риска при промяна на бизнес средата и в даден момент да се окажат ресурсно необезпечени или пък да са принудени да поемат инфраструктурни разходи, които впоследствие да се окажат ненужни. Допълнителен аргумент за такова очакване е, че инфраструктурата осигуряваща функционирането на малките кълстери обикновено осигурява значително по-ниско качество от тази на големите центрове.

Собственикът на изчислителните ресурси е в състояние да поеме риска да осигури желаната еластичност поради факта, че отклоненията в общото натоварване при огромен брой потребители са сравнително малки и то може да се прогнозира по-достоверно и да се осигури при ниско ниво на резервираност.

Широкото използване на *виртуализацията* само по себе си също води до икономическа изгода, тъй като се създава една стабилна виртуална архитектура на използваните ресурси, която позволява по-висока степен на унификация и на рационализация на разработката на нови приложения и която води до допълнително намаляване на необходимото начално финансиране при внедряване на нови услуги през cloud computing.

Постигането на горните икономически предимства и преминаването към ползване на Cloud Computing среда е свързано с промени в моделите за разработка на програмни следства, които следва да:

- постигат желаната производителност, като се възползват в максимална степен от паралелизма, а не разчитат на възможностите на конкретен процесор;
- са създадени за изпълнение върху виртуални машини, т.е. да не се ограничават от конкретните особености на дадена компютърна архитектура за постигане на своите цели включително и по-висока производителност;
- са реализирани с идеята, че техните потребители ще плащат за ползването на софтуерния продукт, а няма да го закупуват.

Реализирането на тези промени е свързано с разходи, които следва да се имат предвид при преминаването на дадена фирма към предлагане на продукти в Cloud Computing среда. Подобни необходими промени могат да се формулират и за производителите на хардуерни средства, които следва да са проектирани и реализирани с идеята, че:

- ще се конфигурират в кълстери и бъдещето им търсене все повече ще се преориентира от отделни устройства към обеми с над 10 и повече шкафа;
- все по-съществена става енергийната ефективност, в това число и ефективното реализиране на икономични „спящи” режими при малки натоварвания на устройствата, което е пряко свързано с търсенето на икономическа ефективност от високата плътност на тяхното разполагане и намаляване на разходите за охлаждане;
- значение придобива и ефективността на хардуерните средства при реализацията на виртуални машини и ефективността на тяхната свързаност;
- вероятно в нивата на реализиране на запомнящата среда ще стане належащо въвеждането на допълнително ниво от флаш памети [9], което

би довело до намаляване на консумацията на енергия и до повишаване на производителността на запомнящата среда.

За икономическата ефективност на приложенията при ползване на Cloud Computing среда съществено значение имат съотношенията и тенденциите при цените на свързаността на компютърните ресурси, цената на изчисленията и цената на запомнящата среда. Както е показано в [10] и потвърдено в [9], спадането на цените на свързаността изостава спрямо това на изчисленията и на запомнящата среда и препоръката е, че в разпределени системи, каквато е и Cloud Computing, големите информационни масиви следва да се разполагат в близост до мястото, където ще се обработват. Това винаги е било препоръчително за ефективното ползване на информационните технологии от началото на тяхното развитие и очевидно ще продължи да бъде до момента, когато евентуално ще настъпи нов срив в цените на комуникациите.

4. Проблеми със сигурността

Сигурността е един от сериозните и все още нерешени проблеми за ползването на Cloud Computing по начин, позволяващ реално развързване на собствеността върху ресурсите и доставчиците на компютърни услуги. Наличието на този проблем може да се отрази върху темповете на бъдещото разпространение на тази технология и на постигането на възможните икономически изгоди.

В момента основният вариант на осигуряване на сигурността на крайните клиенти е ползването на Web базиран интерфейс през ниво SSL (Secure Sockets Layer) и той позволява постигането на масовост на крайното потребление на ниска цена [8]. Той обаче изобщо не решава проблема за постигане на сигурност от гледна точка на доставчиците на услуги и на техните потребители при обработката и съхранението на информацията на компютърни ресурси, които са под контрола на техния собственик. Това означава, че всеки доставчик на услуги или техен потребител, който обработва или съхранява конфиденциална информация би разчитал на Cloud Computing среда само ако притежава и компютърните ресурси, което е силно ограничаващо за разгръщане на пълния потенциал на тази технология. За съжаление в момента големите доставчици на глобални услуги заобикалят проблемите с осигуряване на сигурността на своята информация, като инвестират и в огромни изчислителни центрове вместо да търсят външни доставчици за своите потребности. Това решение неминуемо намалява възможностите за осигуряване на оптималното натоварване на компютърните им ресурси.

Част от проблемите по управлението на сигурността в Cloud Computing среда според [15] се свеждат до:

- Управление на достъпа на потребителите – в това число и управление на привилегировани потребители, на „федерирана” идентичност (federated identity – идентифициране на потребителя през няколко различни системи за управление на достъпа), осигуряване на възможност за влизане в системата през едно място (Single sign-on), управление на правата на достъп на потребителите (User provisioning), и др.;
- Осигуряване на защита срещу неототоризирано изтичане на информация (Data Leak Protection);

- Управление и поддръжане на дневници за активността на потребителите и на Cloud Computing средата (Log Management).

Освен гореизброените проблеми по сигурността трябва да се добавят и проблеми по надеждността (резервиране на ресурсите, гарантиране на нивото на производителност, управление на възстановяването на информацията при аварийни ситуации и други подобни). Допълнително сериозно усложняване на ползването е липсата на опит в договорното гарантиране на нивото на услугите в тази нова тристранна схема, включваща потребител – доставчик на услуга – доставчик на изчислителен ресурс. Не бива да се пропуска и факта, че решаването на проблемите по управление на сигурността в Cloud Computing средата са свързани и с разработването на нови стандарти, за което е нужно и технологично време.

5. Проблеми с производителността

Основните проблеми с производителността на Cloud Computing средата до голяма степен са в тясна връзка с качествата на нейната свързаност, която следва да се разглежда от две страни:

- Свързаност на потребителите с ресурсите на Cloud Computing средата;
- Свързаност в границите на самата Cloud Computing среда.

В първия случай свързаността на потребителя се превръща в тясно място, което може силно да затрудни постигането на желаната производителност, ако неговата обработка на информацията изисква прехвърляне на големи информационни масиви към и от Cloud Computing средата – например като входни данни или като краен резултат. В частност това би се получило, ако потребителят притежава големи обеми от данни, планира да ги съхранява локално и ползва изчислителния ресурс на Cloud Computing за тяхната обработка. Възможно решение е потребителят да прехвърли ангажимента за съхранението на тези данни на Cloud Computing средата, но това означава да приеме и риска за тяхната сигурност. За доставчика на Cloud Computing това означава, че трябва да осигури разположението на информационните масиви на приложението в близост до мястото, където ще се обработват.

Във втория случай става дума за влиянието на вътрешната свързаност на една разпределена изчислителна система върху нейната ефективност. В частност в една Cloud Computing среда проблем с вътрешната свързаност ще има, ако обработката се нуждае от интензивен информационен обмен между процесорните елементи, които я изпълняват. Това е проблем, който е част от по-общия въпрос, възможно ли е една Cloud Computing среда да изпълнява ефективно суперкомпютърни приложения. Отговорът не е еднозначен и е възможно в бъдеще на ниво *IaaS* да се предлагат и процесорни елементи с осигурена високоскоростна свързаност, както е в сега съществуващите GRID структури.

6. Заключение

Всеобщо е мнението, че разпространението на технология от типа на Cloud Computing е неизбежно и необратимо. Това разпространение завинаги ще промени използването, покупката и продажбата на софтуера и на останалите ресурси свързани с информационните технологии. Налагането на концепцията за Cloud Computing е

свързано пряко и с всички известни последствия от глобализацията, които се наблюдават в другите сфери на бизнеса. В частност неизбежно е окрупняването на доставчиците на компютърни ресурси, което в момента се загатва с реализацията на огромните изчислителни центрове, а така също неизбежно е и по-нататъшното доминиране на глобални доставчици на услуги. Тези процеси, както всяка друга концентрация и глобализация, ще доведат до повишаване на качеството и намаляване на цената на услугите, базирани на информационните технологии, но така също и до по-трудно навлизане на нови играчи на този пазар, което ще го направи по-малко динамичен. Това може да се приеме, че е една от тенденциите при развитие на информационните технологии.

Литература

- [1] Mache Creeger, "Cloud Computing: An Overview", Originally published in Queue vol. 7, no. 5, ACM New York, NY, USA, June 12, 2009, Retrieved from <http://portal.acm.org/citation.cfm?id=1554608>
- [2] Боянов К., Д. Тодоров, Хр. Турлаков, „Икономически особености на виртуализираните разпределени компютърни ресурси”, сп. Автоматика и Информатика, бр.2, 2010 г., стр.29-33, София, ISSN 0861-7562.
- [3] Ian Foster and Steven Tuecke, "Describing the Elephant: The Different Faces of IT as Service", Originally published in Queue vol. 3, no. 6, ACM New York, NY, USA, August 18, 2005, Retrieved from <http://portal.acm.org/citation.cfm?id=1080874>
- [4] "Cloud computing" article from Wikipedia, Retrieved from http://en.wikipedia.org/wiki/Cloud_computing
- [5] Ian Foster, "What's faster – a supercomputer or EC2?", August 05, 2009, Retrieved from <http://ianfoster.typepad.com/blog/2009/08/whats-faster-a-supercomputer-or-ec2.html>
- [6] Steve Campbell, "Timesharing 2.0", November 03, 2009, HPCwire.com, Retrieved from http://www.hpcwire.com/specialfeatures/cloud_computing/features/Time-sharing-20-66169142.html?viewAll=y
- [7] Wolfgang Gentzsch, "Grids or Clouds for HPC?", November 02, 2009, HPCwire.com, Retrieved from http://www.hpcwire.com/specialfeatures/cloud_computing/features/Grids-or-Clouds-for-HPC-67796917.html?viewAll=y
- [8] Ian Foster, Yong Zhao, Ioan Raicu, Shiyong Lu, "Cloud Computing and Grid Computing 360-Degree Compared", 31 Dec 2008, arXiv:0901.0131v1 [cs.DC], Retrieved from <http://arxiv.org/abs/0901.0131>
- [9] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, Matei Zaharia, "Above the Clouds: A Berkeley View of Cloud Computing" - Technical Report No. UCB/EECS-2009-28, 10 February 2009, Electrical Engineering and Computer Sciences University of California at Berkeley - Reliable Adaptive

- Distributed Systems Laboratory, Retrieved from <http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf>
- [10] Jim Gray, "Distributed Computing Economics" Technical Report MSR-TR-2003-24, March 2003, Microsoft Research - Microsoft Corporation, Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.150.1335&rep=rep1&type=pdf>
 - [11] Rich Miller, „Apple Planning \$1 Billion iDataCenter”, May 26th, 2009, Data Center Knowledge, Retrieved from <http://www.datacenterknowledge.com/archives/2009/05/26/apple-planning-1-billion-idatacenter/>
 - [12] Edward Walker, "The Real Cost of a CPU Hour," Computer, vol. 42, no. 4, pp. 35-41, Apr. 2009, doi:10.1109/MC.2009.135
 - [13] Iordanov, I. and V. Georgiev, Design and Modeling of Cloud Application Architecture, Proceedings of 5th International Scientific Conference on Computer Science, Sofia, Bulgaria, 4th – 5th November, 2009. pp 142. – 147.
 - [14] Iordanov, I. and V. Georgiev, Architecting Process for High-level Social-oriented Resource Allocation Cloud, Proceedings of 5th International Scientific Conference on Computer Science, Sofia, Bulgaria, 4th – 5th November, 2009. pp 148. – 153.
 - [15] Jeleu, R. and V. Georgiev, Service Functionality for Cloud Systems, Proceedings of 3rd Int Conference on Information Systems and Grid technologies, Sofia, Bulgaria, 28th – 29th May, 2009. pp 168 – 182.
 - [16] Matthew Gardiner „Identity as Security Glue for the Cloud”, 27 July, 2010, CA Inc., Retrieved from http://www.bitpipe.com/detail/RES/1280261072_739.html