Research Challenges for Cloud Computing Economics

Darko Andročec

Faculty of Organization and Informatics

University of Zagreb

Pavlinska 2, 42000 Varaždin, Croatia

`darko.androcec@foi.hr`

Abstract. Cloud Computing is a new paradigm of computing infrastructure provision which promises to achieve a vision of computer utilities. The most active research topic in Cloud Computing is its economic aspect. In this article we give overview of existing literature on Cloud Computing economics (pricing of Cloud services, costs, benefits and risk of Clouds, ROI and cost/benefits models) and propose some new research challenges. Some of the most interesting future topics are a complete cost-benefit analysis methodology development, using simulations to identify tangible cost reduction, sustainability of current prices of Cloud services and system administration cost in a Cloud environment.

Keywords. Cloud Computing, research challenges, cost, benefit, risk, pricing

1 Introduction

Cloud Computing is a new paradigm for the provision of computing infrastructure, platform or software as a service. In a modern society basic services (water, electricity, gas, etc.) are provided such that everyone can have easy access to them. Cloud Computing promises reliable services delivered through new virtualized data centers and it achieves a 21st century vision of computer utilities [2]. Today we have high cost of computing and we need highly specialized labor to keep it running well. Cloud Computing address these needs with [4] on-demand access, elasticity, pay-per-use, connectivity, resource pooling and abstracted infrastructure. For consumers, Cloud Computing is primarily a new business paradigm [14].

The most active research topic in Cloud Computing is its economic aspect. The main question is whether or not your organization should use Cloud Computing. This depends on a number of factors, e.g. [17]: cost/benefit ratio, speed of delivery, used capacity, sensitivity of date, organization's corporate and IT structure. The second problem is pricing of Cloud Computing services. In this article we give overview of existing literature on Cloud Computing economics and propose some new research challenges.

This paper proceeds as follows. Firstly, in Section 2, we present an overview of the Cloud Computing paradigm based on current literature. Section 3 shows review of Cloud Computing economics. In the next section we analyze actual and future research challenges and issues in Cloud Computing economics. Finally, our conclusions are presented in the last section.

2 Cloud Computing paradigm

There are many Cloud Computing definitions. Linthicum writes in [9] that the most comprehensive definition of Cloud Computing provided NIST Information Technology Laboratory: "Cloud computing is a pay-per-use model for enabling available, convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, services) that can be rapidly provisioned and released with minimal management effort or service provider interaction"[9]. Authors of the paper [16] gave an integrative definition of Cloud Computing: „Clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services). These resources can be dynamically reconfigured to adjust to a variable load (scale),

allowing also for an optimum resource utilization. This pool of resources is typically exploited by a pay-per-use model in which guarantees are offered by the Infrastructure Provider by means of customized SLAs. "[16]. They conclude that minimum definition of Cloud Computing must contain scalability, pay-per-use utility model and virtualization. "Gartner defines Cloud Computing as a style of computing where massively scalable IT-related capabilities are provided as a service using Internet technologies to multiple external customers". [17]

In the technical report [1] authors conclude there are three new aspects in Cloud Computing from a hardware point of view: the illusion of infinite computing resources available on demand, the elimination of an up-front commitment by Cloud users and the ability to pay for use of computing resources as needed. They also show ten obstacles to and opportunities for growth of Cloud Computing, e.g. one of the obstacles is availability of service, and opportunities for this example are use of multiple Cloud providers and use of elasticity to prevent DDOS. Other obstacles mentioned in [1] are data lock-in, data confidentiality and auditability, data transfer bottlenecks, performance unpredictability, scalable storage, bugs in large distributed systems, scaling quickly, reputation fate sharing and software licensing. In the book [17] authors define benefits and limitations of Cloud Computing. Benefits are scalability, simplicity, knowledgeable vendors, more internal resources (IT department is freed up to work on important business-related tasks), increased security if you use reputable cloud vendors. Some of Clouds limitations are: work with sensitive information, privacy concerns with a third party, regulatory issues. Author in [9] lists following benefits of Cloud Computing: cost, better mix and match services, innovative, expandability, speed to implementation, environment friendly. In the same book [9] author also give some drawbacks of this new paradigm: security, cloud provider control of IT infrastructure, proprietary nature of the paradigm, compliance issues.

Article [19] considers functional aspects of Cloud Computing and why is it distinct from other computing paradigms. Clouds have services to access hardware, software and data resources in a transparent way. Type of services are [19]: Hardware as a Service (HaaS), Sofware as a Service (SaaS), Data as a Service (DaaS) and Infrastructure as a Service (IaaS). Linthicum [9] gives 11 major types of services: Storage-as-a-service, Database-as-a-service, Information-as-a-service, Process-as-a-service, Application-as-a-service, Platform-as-a-service, Integration-as-a-service, Security-as-a-service, Management/governance-as-a-service, Testing-as-a-service, Infrastructure-as-a-service. The Cloud Computing is different from other paradigms in the following aspects [19]: user-centric interfaces, on-demand service provisioning, QoS guaranteed offer,

autonomous system, scalability and flexibility. Technologies behind Clouds are: distributed storage system, Cloud programming models, virtualization, service flow and workflow orchestration, service oriented architecture and Web 2.0.

Topology of Cloud Computing systems is made up of clients, the datacenter and distributed servers. Cloud Computing also has several deployment models [9]: private cloud (infrastructure is owned or leased by and operated only for a single organization), community cloud (infrastructure is shared by several organizations), public cloud (the infrastructure is owned by an organization selling cloud services to the general public), hybrid cloud (composition of two or more before mentioned cloud deployment models).

3 Cloud Computing economics

After review of Cloud Computing economics literature, we assign articles to three main categories (pricing of Cloud services, costs, benefits and risks of Clouds, ROI and cost/benefits models) which are described in the following subsections.

3.1 Pricing of Cloud services

In the paper [11] a dynamic pricing mechanism for the allocation of shared resources is proposed and its performance is evaluated. The economic properties of this pricing scheme are formally proved using the mechanism design framework.

Authors of the article [15] propose a new resource pricing and allocation policy in Cloud Computing. In it users can predict future resource prices. Authors used game theory to solve the multi-user equilibrium allocation problem. They integrated budget and deadline conditions and solved the Cloud service price prediction problem with incomplete knowledge. They used simulation framework Cloudsim to make experiments.

Yeo et al. [21] proposed an autonomic metered pricing for a utility computing service. Today Cloud providers charge users using a simple pricing scheme, fixed prices based on resource types. Authors of [21] advocated charging variable prices and providing guaranteed quality of service through the use of advanced reservations that are guarantees of access to a computing resource at a particular time in the future for a particular duration. They said that charging fixed prices is not fair to both the provider and users in Cloud Computing environment. Providers could maximize revenue by differentiating the value of computing services provided to different types of users.

3.2 Cost, benefits and risks of Clouds

In the book [17] authors defined several types of Cloud Computing benefits: operational benefits, economic benefits and staffing benefits. Operational benefits are benefits to the way a organization operate and some of the most important are: reduced cost, increased storage, automation (provider updates application), flexibility, better mobility, better use of IT staff. Some economic benefits to consider are: fewer needed staff members, Cloud suppliers can purchase hardware much cheaper, pay as you go, faster time to market. Staffing benefits in SaaS include benefits for the consumer and benefits for the provider [17]. The consumer has following benefits: no software installation or maintenance, shorter deployment time, worldwide availability, SLA adherence, the vendor ensures that application is constantly improved. For the provider benefits are: operating environment owning, predictable revenue stream, small and regular upgrades, customer relationship management.

Rosenthal et al. [14] examined how the biomedical informatics community can take advantage of Cloud Computing. They concluded that for applications that are not I/O sensitive and do not demand a fully mature environment Cloud Computing can sometimes provide major improvements. The goal of their paper is to help decision makers at biomedical laboratories to understand how to asses adequacy of Cloud services. Three major cost drivers of biomedical informatics systems are system administration, idle capacity, and power usage and facilities [14]. Authors also identified some Cloud Computing qualitative benefits: less to manage, scalability, superior resiliency, homogeneity, fewer issues to negotiate with institutional authorities. Some risks are also shown: risks due to hackers, multi-tenancy risks, protections at virtual machine boundaries, nontechnical outsourcing risks.

In [5] authors elaborated on the concepts, benefit and risk of Software-as-a-service (SaaS) and Platform-as-a-service (PaaS) for mobile operators. Benefits for platform owner in SaaS are economies of scale, predictable revenues, shorter sales cycle, expanded customer base and shorter development lifecycle. Benefit for the same stakeholders in PaaS is that revenues are based on hosted applications' usage. End-users in SaaS models have little initial investment, can eliminate software management activities, have lower total cost of ownership of IT resources and stable, reliable and flexible experience. In PaaS they have access to innovative software. From developer perspectives PaaS model has many benefits: single environment, faster time-to-market, reduced IT infrastructure provisioning, interoperability with other applications. Authors of the paper [5] also listed risks in SaaS and PaaS models. In SaaS there are risks for platform owners (up-front infrastructure investment, diverse new skills, managing network of suppliers) and for end-users (exposing and losing business-critical data, lock-in, high switching costs, less tailoring software). PaaS model also has many risks for different stakeholders: platform adoption, unavailable applications, long learning curve, lock-in, restricted to available APIs, closed platform, lack of interoperability etc. In SaaS and PaaS there are performance, security and scalability risks.

Armbrust et al. [1] believe that although the economic appeal of Cloud Computing is often described as converting capital expenses to operating expenses, more accurate benefit to the client is told in the phrase "pay as you go". The absence of up-front capital expense allows capital to be redirected to core business investment and there is lower risk of underutilization and saturation (benefits of elasticity and risk transfer). Misestimating workload is then shifted from the service operator to Cloud service provider.

The paper [13] discusses the use of Cloud Computing in government and its tangible and intangible risks. Tangible risks specific to government use of Cloud services are: access (private data must be secured, unwanted or outside access request must be denied, all access privileges must be capable of customer auditing), availability (Cloud vendor outages, determination of the priority of users on the specific Cloud, denial of service attacks, availability of vendor itself), infrastructure (risks and cost in migrating information to Cloud infrastructure, there are no standards about Cloud interoperability and compatibility, issue of compliance regulations), integrity (accuracy of managed information). Intangible risks are due to law and policy issues.

The key advantages of Cloud Computing described in [10] are: lower cost of entry for smaller firms, immediate access to hardware resources with no upfront user's capital investments, lower IT barriers to innovation, easy scaling of services, new classes of application and services. A SWOT analysis is also shown in [10]. Main Cloud Computing strengths are the ability to scale up services, the ability to effectively use time-distributed computing resources, reduced infrastructure costs, energy savings, reduced upgrades and maintenance costs. Some of weaknesses are the loss of physical control of the data and insufficient uptime for mission-critical applications for large organizations. Significant opportunities of Cloud Computing lie in its potential to help developing countries, mashups, energy efficiency, equipment recycling, new innovative services. The biggest threats to Cloud Computing adoption are lack of standards, security, regulation at the local, national and international level.

3.3 ROI and cost/benefits models

Linthicum [9] presented many dimensions of the Cloud Computing value: ongoing operational cost reduction, capital preserving, value of upsizing and

downsizing on demand, shifting the risk, agility and reuse. In the same book he also proposed the cost/benefit methodology that is composed of eight steps: understand the existing issues, assign costs, model "as is", model "to be", define value points, define hard benefits, define soft benefits, create final business case.

Authors of the article [3] proposed an economic model for a cloud cache suitable for the querying service of large scientific datasets. It is based on a cost model that takes into account network bandwidth, disk space and CPU time.

Metrics and formulas for the calculation of Cloud Computing TCO (Total Cost of Ownership – cost spent to build and operate a Cloud) are presented in [8]. Authors defined calculation model which contains server, software, network, support and maintenance, power, cooling, facilities and real-estate cost. Another cost is Cloud Utilization Cost which embodies the cost caused by the Cloud utility and reflects sensitively the dynamic utility. They also develop a web calculation tool which provides a way to analyze the effect of different metrics on the final cost.

The paper [20] examined the economics of Cloud Computing charging from the perspective of a supercomputing resource provider. The competitiveness of author's computing center with Cloud Computing resources was evaluated. Authors of [20] used Amazon EC2 charging model and applied it to their current supercomputing resources to test cost effectiveness of being a Cloud provider. They concluded that charging for computational time may be appropriate and charging for data traffic not. When they extended the analysis to their future new cluster they concluded that their resource will be competitive with current and anticipated Cloud provider rates.

Misra and Mondal [12] described ROI model with two steps. First step is identification of a company's suitability for the adoption of Cloud Computing using suitability index calculation. They made mathematical formulas using relevant factors and its assigned credits (weights) according to their relative importance. Some of the key factors shown in [12] are: size of the IT resources (number of servers, size of customer base, annual revenue from IT offering, number of countries IT is spread across), the utilization pattern of the resources (average usage, peak usage, amount of data handling), sensitivity of the company's data, criticality of work done by the organization. On doing the calculation a numerical value would be obtained, and if the value for a particular organization is below the lower limit, company is not suitable for Cloud services adoption. In other cases, companies start calculating ROI using generic model presented in [12]. Various intangible Cloud benefits have also been included in this model.

The article [7] compared the performance and monetary cost-benefits of Cloud Computing for desktop grid applications. In it authors examined performance measurements and monetary expenses of desktop grids (e.g. SETI@home) and the Amazon EC2 Cloud service. They also considered hybrid approaches where a volunteer computing server is hosted on Amazon EC2 and concluded that savings due to cheaper start-up and lower monthly costs ranges between 40-95% depending on various resource usage.

In [18] it is described modeling tool for quantitative comparison of leasing CPU time and using a local server cluster. Author of this article wrote that NPV (net present values) models used for equipment lease or buy decisions don't account for a CPU nonmonetary performance depreciation. He proposes a concept of the time value of CPU and gives equations for a purchase, lease and purchase-upgrade case. In addition, two case studies are mentioned. In this model price volatility in the online CPU market was not considered. Model described in [18] is accurate if barriers of entry into the market for hirer were low and technology lock-in does not exist. Today these two conditions are not fully met.

4 Research challenges

Marston et al. [10] suggested their research agenda of Cloud Computing. One of the six categories is Cloud Computing economics. Research topics will include pricing strategies for Cloud Computing providers and the role of enablers effect on Cloud service provider economic value. These authors also said that development of a methodology to assess Clouds risk is also very interesting and promising research topic.

Authors of paper [6] discussed some of the Cloud research challenges from an organizational perspective. They saw that the vast majority of current work on Cloud costs is simplistic from an enterprise perspective. The economic issues around application migration and existing procurement policies are not considered. Further work is also required to examine the true costs of using Cloud Computing services in a specific organization with help of tools and techniques.

We will also propose some interesting future research topics about Cloud Computing economics. One interesting topic is a development of complete cost-benefit analysis methodology for assessing tangible and intangible costs, benefits and risk of Cloud Computing migration. This methodology can use business process model simulations to identify tangible cost reduction. Another question is whether current prices of Cloud services are realistic and sustainable in future. A potential research theme is using Cloud Computing to enable virtual organization. Costs of system administration after move to a Cloud Computing service are not yet investigated. Many tasks of system administration are just as complex as with own server instances.

5 Conclusions

Cloud Computing is a new business and computing paradigm for the provision of infrastructure, platform or software as a service. We firstly review the most active research field of Cloud Computing: its economic aspect. We classify current relevant work into three main categories: pricing of Cloud services, cost, benefits and risks of Cloud Computing, ROI and cost/benefits models. Several relevant articles in each category are listed and briefly described, so we can see state-of-the-art of Cloud Computing economics. Thereafter other author's research agendas of Cloud Computing are shown. Our next challenge was to suggest some new interesting research themes. In this paper we propose following topics: a complete cost-benefit analysis methodology development, using simulations to identify tangible cost reduction, sustainability of current prices of Cloud services and system administration cost in a Cloud environment.

References

[1] Armbrust M., Fox A., Griffith R, Joseph A. D., Katz R.H., Konwinski A., Lee G., Patterson D.A., Rabkin D.A., Stoica I., Zaharia M.: Above the Clouds: A Berkeley View of Cloud Computing available at http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf, Accessed: 15th March 2011.

[2] 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, Amsterdam, Netherlands, 2009, pp. 599-616

[3] Dash D., Kantere V., Ailamaki A.: An Economic Model for Self-Tuned Cloud Caching, ICDE '09 Proceedings of the 2009 IEEE International Conference on Data Engineering, Washington, USA, 2009, pp. 1687-1693

[4] Durkee D.: Why Cloud Computing Will Never Be Free, Communications of the ACM, pp. 62-69

[5] Goncalves V., Ballon P.: Adding value to the network: Mobile operators' experiments with Software-as-a-Service and Platform-as-a-Service models, Telematics and Informatics, New York, USA, 2011, pp. 12-21.

[6] Khajeh-Hosseini A., Sommerville I., Sriram I.: Research Challenges for Enterprise Cloud Computing, available at http://arxiv.org/ftp/arxiv/papers/1001/1001.3257.pdf, Accessed: 24th March 2011.

[7] Kondo D., Javadi B., Malecot P., Cappelo F., Anderson D.P.: Cost-Benefit Analysis of Cloud Computing versus Desktop Grids, IPDPS '09 Proceedings of the 2009 IEEE International Symposium on Parallel&Distributed Processing, Washington, USA, 2009, pp. 1-12

[8] Li X., Li Y., Liu T., Qiu T., Wang F.: The Method and Tool of Cost Analysis for Cloud Computing, CLOUD '09 Proceedings of the 2009 IEEE International Conference on Cloud Computing, Washington, USA, 2009, pp. 93-100

[9] Linthicum D.S.: Cloud Computing and SOA Convergence in Your Enterprise, Addison-Wesley, New York, USA, 2009.

[10] Marston S., Li Z., Bandyopadhyay S., Zhang A., Ghalsasi A.: Cloud Computing – The business perspective, Decision Support Systems, 2011, pp. 176-189

[11] Mihailescu M., Teo Y.M.: On Economic and Computational-Efficient Resource Pricing in Large Distributed Systems, CCGRID '10 Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, Washington, USA, 2010, pp. 838-843

[12] Misra S.C., Mondal A.: Identification of a company's suitability for the adoption of cloud computing and modelling its corresponding Return on Investment, Mathematical and Computer Modelling, Amsterdam, The Netherlands, 2011, pp. 504-521

[13] Paquette S., Jaeger P.T., Wilson S.C., Identifying the security risks associated with governmental use of cloud computing, Government Information Quarterly, 2010, pp. 245-253

[14] Rosenthal A., Mork P., Li M.H., Stanford J., Koester D., Reynolds P.: Cloud computing: A new business paradigm for biomedical information sharing, Journal of Biomedical Informatics, Amsterdam, The Netherlands, 2010, pp. 342-353

[15] Teng F., Magoules F.: Resource Pricing and Equilibrium Allocation Policy in Cloud Computing, CIT '10 Proceedings of the 2010 10th IEEE International Conference on Computer and

Information Technology, Washington, USA, 2010, pp. 195-202

[16] Vaquero L. M., Rodero-Merino L., Caceres J., Lindner M.: A Break in the Clouds: Towards a Cloud Definition, ACM SIGCOMM Computer Communication Review, New York, USA, 2009, pp. 50-55.

[17] Velte A.T., Velte T.J., Elsenpeter R.: Cloud Computing: A Practical Approach, McGraw-Hill, New York, USA, 2010.

[18] Walker E.: The Real Cost of a CPU Hour, Computer, 2009, pp. 35-41

[19] Wang L., von Laszewski G., Younge A., He X., Kunze M., Tao J., Fu C.: Cloud Computing: a Perspective Study, New Generation Computing, Tokyo, Japan, 2010, pp. 137-146.

[20] Woitaszek M., Tufo H.M.: Developing a Cloud Computing Charging Model for High-Performance Computing Resources, CIT '10 Proceedings of the 2010 10th IEEE International Conference on Computer and Information Technology, Washington, USA, 2010, pp. 210-217

[21] Yeo C.S., Venugopal S., Chu X., Buyya R.: Autonomic metered pricing for a utility computing service, Future Generation Computer Systems, Amsterdam, The Netherlands, 2010, pp. 1368-1380