ORIGINAL PAPER

Perspectives on cloud computing: interviews with five leading scientists from the cloud community

Gordon Blair • Fabio Kon • Walfredo Cirne • Dejan Milojicic • Raghu Ramakrishnan • Dan Reed • Dilma Silva

Received: 2 May 2011 / Accepted: 5 May 2011 / Published online: 3 June 2011 © The Brazilian Computer Society 2011

1 Introduction

Cloud computing is currently one of the major topics in distributed systems, with large numbers of papers being written on the topic, with major players in the industry releasing a range of software platforms offering novel Internet-based services and, most importantly, evidence of real impact on end user communities in terms of approaches to provisioning software services. Cloud computing though is at a formative stage, with a lot of hype surrounding the area, and this makes it difficult to see the true contribution and impact of the topic.

Cloud computing is a central topic for the Journal of Internet Services and Applications (JISA) and indeed the most downloaded paper from the first year of JISA is concerned with the state-of-the-art and research challenges related to

G. Blair Lancaster University, Lancaster, UK

F. Kon (⊠) University of São Paulo, São Paulo, Brazil e-mail: fabio.kon@ime.usp.br

W. Cirne Google, Inc., Palo Alto, USA

D. Milojicic HP Labs, Palo Alto, USA

R. Ramakrishnan Yahoo! Research, Santa Clara, USA

D. Reed Microsoft Research, Redmond, USA

D. Silva IBM Research, Yorktown Heights, USA cloud computing [1]. The Editors-in-Chief, Fabio Kon and Gordon Blair, therefore felt it was timely to seek clarification on the key issues around cloud computing and hence invited five leading scientists from industrial organizations central to cloud computing to answer a series of questions on the topic.

The five scientists taking part are:

- Walfredo Cirne, from Google's infrastructure group in California, USA
- Dejan Milojicic, Senior Researcher and Director of the Open Cirrus Cloud Computing testbed at HP Labs
- Raghu Ramakrishnan, Chief Scientist for Search and Cloud Platforms at Yahoo!
- Dan Reed, Microsoft's Corporate Vice President for Technology Strategy and Policy and Extreme Computing
- Dilma Silva, researcher at the IBM T.J. Watson Research Center. in New York

We are very pleased to have such major names from industry to comment on cloud computing and thank them very much in taking the time to offer such thoughtful insights into the area.

The interviews cover three areas:

- The first two questions focus on the essential nature of cloud computing in terms of what is fundamentally new about the area and what impact this is having on potential end users, for example in industry and commerce
- The subsequent six questions focus on open issues related to cloud computing including challenges relating to interoperability and addressing non-functional concerns
- The remaining two questions focus on the future in terms of perceived significant changes and also associated research problems (the latter for example may be of interest to PhD students working in this area and seeking clarity on what are important research topics to address)



One notable observation from the answers provided is that cloud computing is very real and having a major impact on the industry, for example in providing support for startups and encouraging innovation in this important sector. We are also struck by the consistency in terms of the definition of cloud computing showing that industry is moving towards a common view of the subject.

We hope you find the question and answer sessions that follow to be insightful and helpful to you in your work. This paper also acts as a taster for a follow-up JISA special issue on cloud computing, which will come out in a few months.

2 The interviews

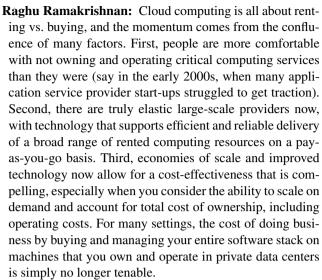
(1) There is a lot of hype around cloud computing. Is there something fundamentally new about cloud computing or is it just another marketing term from the computing industry?

Walfredo Cirne: Cloud computing is really an evolutionary step on how we use computers. About 10 years ago, most of the action on systems was on grid and utility computing. Grid computing focused on how services running on multiple administrative domains could collaborate, whereas utility computing aimed to let people acquire ondemand resources to run their applications. It turns out that grid is just too hard, and utility is insufficient.

Cloud computing enables people to gather on-demand resources and services that are used as platform for their applications, but makes no promises they can come from multiple administrative domains. Combining multiple administrative domains as a platform is extremely hard (particularly due to security) and does not add much value (except in some niche cases, especially in high-performance computing).

That said, of course there is hype on cloud computing. It has been the case for all important (r)evolutions in computer science. :-) But cloud computing does go beyond the hype. It provides a practical way to decouple application development from the platform (both raw resources and high-level services) needed to run the application.

Dejan Milojicic: If you go back to WWW, neither HTML nor HTTP were fundamentally new, but they still changed the world once deployed and universally used. New things in Cloud computing had fundamental predecessors, in particular, (a) elasticity by design; (b) self-service; (c) multitenancy; and (d) pay peruse. However, previous incarnations (Grids, clusters) did not have maturity of applications delivered as services nor the business needs and adoption. In summary, it is not about the "fundamentally new things", but about the right timing and adoption of technologies. Other examples include Java, UNIX, Internet, etc. Each one of them have predecessors, yet they made significantly higher impact.



Technically, this puts a fresh and renewed focus on many concepts, like auto-tuning, high availability, notions of consistency, horizontal scaling, etc., that have long been recognized as important challenges for large-scale systems. The growth of cloud computing and the central role these ideas play has led to an explosion of new work, and new forums (such as the ACM Symposium on Cloud Computing on the research side, and several on the industrial/trade-show side). While no one of these ideas is "new", the collective emphasis is certainly unprecedented and healthy.

Dan Reed: What's unique is how the ratios of capacity, capability and cost have changed given advances in networking, storage and computing. The shift in these ratios has enabled a dramatic improvement in the economics of massive-scale data centers, creating the opportunity for cloud computing to evolve. Further, new business models have arisen to create new types of services that run in the cloud. For example, free email and photo storage, social networks such as Facebook and twitter are all cloud services. The public cloud makes it possible for anybody to create a new cloud service business at a very low cost. It's the rare combination of the "right" technology and the "right" economics that makes this a truly disruptive trend.

Dilma Silva: Yes, beyond the buzz there is a new delivery platform for computing services that takes our industry one step further towards the (old) idea of utility computing, offering significant improvements on how computational resources can be made available. Cloud computing leverages the consolidation and software/hardware separation introduced by virtualization and enhances its flexibility by providing elasticity of resources, allowing the platform running our applications to grow and shrink as their needs change in response to application demand fluctuation or resource availability variations.

Another exciting aspect of cloud computing is that it makes it feasible for smaller organizations and research groups



that lack the large capital investment necessary to build data centers and deploy a large number of machines to, instead, budget for paying for those machines for limited periods of time. This ability to throw significant computing resources at a problem at the right time and scale can help our society to attack novel problems as we continue to explore ways in which information technology can contribute to a better world. We may see a "democratization" of high-end computing.

(2) Can you provide an example or examples where cloud computing has made a real impact on the field?

Walfredo Cirne: Start-up launching. Almost no new start-up currently launches on their own computers. They use Amazon EC2/S3, Google App Engine, and the such. This reduces their time-to-market and lower their deployment cost.

Dejan Milojicic:

- Startups can bootstrap their business dramatically easier, quicker, and cheaper than before. They can also scale up and down at ease and at fraction of cost compared to the past offerings.
- Researchers can scale up their experiments as they develop the solutions. Start with small, scale up till they hit the scalability or robustness limits, then debug/redesign, scale up, etc.
- Consolidation of computing and data (as a consequence of Cloud Computing and virtualization in particular) simplifies design, scales better, improves security, reduces costs, etc.
- 4. Consolidation of IT as a consequence of Cloud computing simplifies management, reduces cost, enables easier regulatory compliance, etc.

Raghu Ramakrishnan: There are numerous examples of small start-up companies that used the Amazon cloud to begin with, and remarkably, were able to use AWS as they grew explosively. I think this is transformative. As another example, while Yahoo!'s cloud services are not available to external developers (except in certain high-level forms such as YQL), virtually every Yahoo! site and service today relies upon the suite of Yahoo! cloud services, e.g., the Hadoop map-reduce system, the PNUTS key-value store, and our goal is to increase this shift to where we eliminate the use of siloed implementations of basic services altogether. The benefits that we reap in increased agility (ease of developing new applications and evolving existing ones) and reduced costs are basically the reasons why an increasing number of enterprises are adopting or considering private clouds. Again, as private clouds become ubiquitous, this represents a fundamental change in how we see computing.

Dan Reed: Cloud computing has changed the way we think about "big" data. New, massively parallel algorithms for

gleaning insights from distributed data collections are being created all the time. Storage models like no-SQL databases have come into close study for this reason. We now have to think about data such as graphs with billions of nodes and edges in statistical terms rather than combinatorial ones. Massive-scale data centers are built from components which can fail and when deployed in such large numbers, failure is constant. This requires rethinking the design programming models and runtime systems that make it easy to build a scalable application that tolerate failure. Advances in the computer science are also driving change in cloud computing. New, low power architectures are helping us design "greener" cloud data centers. New machine learning techniques are being incorporated into massive-scale cloud services such as language translation and augmented reality.

Dilma Silva: For certain application domains, innovation can get to the market much quicker with cloud computing. There are many examples of popular mobile phone apps or social network plugins that went "viral", going from dozens of users to several hundreds of thousands in a matter of days. With the traditional model of extending a distributed system by buying, unpacking, cabling, and finally installing the new nodes it would be very challenging to grow from a couple of servers to many hundreds in a matter of hours. By using a cloud platform, applications were able to grow in scale at an unprecedented rate.

(3) Once again we find ourselves in a position in distributed systems where we have a range of competing platforms with limited interoperability between them. Is this sustainable and can you see ways forward to resolve the resultant problems?

Walfredo Cirne: No, this is not sustainable in the long run, but is unavoidable for now. All previous attempts to come up with standards before having running solutions have failed. This should sort itself out with a few more years, where dominant solutions will be clear, and then strong standards will naturally emerge.

Dejan Milojicic: When networking started there were a few networks first, with different protocols, transports, management. Then they grew even more, till they evolved into Internet. Today there are a number of incompatible Cloud providers offering Cloud services at different levels of abstractions. Just like networking in the past resulted into a coherent and omnipresent Internet, I expect that eventually we will have a common widely adopted and ubiquitous Cloud. It takes time, effort, research, and money:-). And people like you asking the right kind of questions.

Raghu Ramakrishnan: Ultimately, the only real solution is standardization, in terms of both APIs and the form-factor of the underlying services.



Dan Reed: In some cases we can already see consensus emerging. For example, cloud data storage concepts like blobs, distributed tables and queues have very similar REST APIs from one cloud vendor to the next. In others, the rate of change is very high and new approaches are still being explored.

Dilma Silva: I believe that we're still at an initial phase of cloud deployment where the focus has been on identifying the right workloads to migrate to the cloud. Once enterprises start to rely more heavily on cloud services, I suspect their requirements in terms of interoperability and support for migrating between vendors will grow, and our industry will respond. One way I can see us addressing these problems is at the Platform-as-a-Service (PaaS) level, with the PaaS providers focusing on services interfaces that can be transparently deployed on top of the competing platforms and integrated to solutions using other PaaS offerings.

I also see a great opportunity for distributed systems researchers and practitioners to get their great ideas out in a manner that simplifies the process of getting a large number of people to try out new software or web services and provide insights for faster refinement and tuning.

(4) Another potential problem in cloud computing is supporting the development of potentially complex applications with demanding non-functional requirements, for example privacy or real-time requirements? Do you see us being in a position to meet such requirements in the near future?

Walfredo Cirne: These are hard problems. Privacy and security, in particular, have a very important non-technical aspect to it. Which technical argument are you going to give to the CIA to make them comfortable with storing their database in your cloud?

We will keep improving the cloud abilities and capabilities, but there will always exist some application running inhouse, due to technical and non-technical reasons.

Dejan Milojicic: Once the business demands it, technology will meet these requirements. Fundamentally, we have already solved some of these problems at lower level of scale, though, we will have to reapply these solutions in the context of Cloud.

Raghu Ramakrishnan: I think that the state-of-the-art in cloud computing technology is at a point where we see genuine benefits and where scalable, reliable services are broadly available. However, much of this technology stack is based on less than optimal adaptation of approaches originally designed for a non-cloud setting, and many issues are not adequately addressed; I expect that foundational work over the next several years will be needed to realize the full potential of rented computing services. This comment applies to functional as well as non-functional requirements. With respect to the specific issues you named, privacy and

real-time, we already deal with scenarios that involve privacy considerations and that leverage real-time data at Yahoo! So, de-facto, our cloud-reliant applications deal with these issues, and this is probably the case in many other applications as well. However, we have not abstracted and supported the capabilities needed to address a broad range of privacy and real-time requirements in a general way, and I expect this to be an evolutionary process as we understand the needs better with more experience of such applications.

Dan Reed: People are building cloud services that address these requirements every day. However, we do need more research on building programming models that make it easy to build these applications. Also academic computer scientists are only now starting to introduce this subject into university curricula. More work is needed to turn this craft into a discipline that can be refined and taught.

Dilma Silva: I think we will get there reasonably soon in terms of achieving the advancements in fine-grain resource and data management that may be required by many classes of non-functional requirements. It may be that first we will have specialized cloud offerings targeting certain requirements in the context of specific application domains, and then later building blocks for this support will be incorporated in the generic cloud platforms.

(5) Public, private, or hybrid?

Walfredo Cirne: The whole point in cloud computing is for the application developers not to worry about putting together and maintaining the platform needed to run their applications.

So, I can only see the motivation of a private cloud for (i) a very large organization that finds cost advantages in running a private cloud for its many application development teams, and decides that this does not distract them from their core mission, or (ii) an organization very worried about privacy, security and so on.

For case (ii), by the way, a hybrid cloud can make very good sense. You use in-house platforms for the critical part of your application, and leverage the advantages of public clouds wherever possible.

Dejan Milojicic: Yes:-). Some will be only public, e.g. start-ups, they will be leveraging elasticity to overcome financial limitations. Large corporation will flirt with public Cloud but will continue to run the core business on the private Cloud, just like large banks still use mainframes, but it is the smaller new banks that use Cloud. Then there will be middle of the road companies that will use hybrid Clouds.

Raghu Ramakrishnan: I think all of the above. Public will be the choice for many small and medium renters; private will be appropriate for many large enterprises; and even the latter will end up using public clouds, if only as reserve capacity.

Dan Reed: Yes, all of the above.



Dilma Silva: All the above. Certain domains will benefit from the economy-of-scale effects provided by public clouds, while other domains will favor private clouds as a way of leveraging cloud benefits (such as consolidation and rapid deployment) without incurring any risk associated with data or computation exposure. For many scenarios, hybrid may the right answer, allowing for the flexibility of dynamically extending private resources with virtual machines hosted on public clouds.

(6) Open source, open core, or proprietary?

Walfredo Cirne: As open as possible. Particularly important are open interfaces for the resources and services offered by the cloud.

Dejan Milojicic: Yes:-). This is orthogonal question to Cloud. Open Core may be a new model here, but the jury is still out and opinions are divided on whether it is a new model or not. There are benefits to openness and there are advantages of proprietary. The former enables more effective development through a type of crowd-sourcing the latter enables competitive differentiation.

Raghu Ramakrishnan: Yahoo! is a big proponent of open source. We've been the major contributor to Hadoop. We've also open sourced a range of other software, including Yahoo! Traffic Server and Zookeeper. I think this goes a long way towards enabling us to converge towards common standards and service suites that are interoperable. I also think this has huge benefits in allowing academia and smaller companies to become full participants in designing and developing the next generation of cloud computing systems.

Dan Reed: We are seeing great innovation in all of these areas.

Dilma Silva: I don't think that cloud computing changes the usual playing forces around these three models in any significant way. In the Infrastructure-as-a-Service level, it seems we will continue to have a few open-source or opencore offerings, with the leaders adoption proprietary solutions. But it's probably too early to know, let's see how efforts such as OpenStack will evolve.

(7) RESTful or not?

Walfredo Cirne: As with all design patterns, depends on the application. If your application naturally lends itself to it, a REST design is a great choice. If not, forcing it into a REST design will result in pain.

Dejan Milojicic: Yes:-). RESTful advanced state-of-theart in terms of scalability, simplicity, performance, etc. It always appealed to me as a sound design choice. At the same time, there might be other cases where its design choices do not best meet the requirements, e.g. using different underlying protocols from HTTP, for dedicated networks, such as for high-performance computing, sensor networks, etc. So, the key question in my mind is what do you target.

Dan Reed: We like to be SOAPy and RESTed. Both are used and will continue to be used.

Dilma Silva: In my personal experience using the cloud, REST APIs are a must for productivity.

(8) Do you see any dramatic changes to the cloud landscape in the near future, and if so, what will these be?

Walfredo Cirne: Not dramatic changes. I see improvement, consolidation, and standardization.

Dejan Milojicic:

- Federation of Clouds—there are already multiple Cloud providers today and many more players are entering Cloud business. One way to resolve this complexity for users is to enable federation of Clouds, where Cloud providers can compete, and one can transfer from one Cloud to another to avoid the lock-in.
- Standardization will be required to accomplish federation, but it may still be somewhat early. Standards need to be established at the right time, if they are too early, they are not adopted; too late and de facto standards take place (which is not necessarily bad).
- Mobile access will be major driver of Cloud services. As smartphones and tablets already surpassed PCs/laptops, the new types of Cloud services delivering and generating unstructured data will change the landscape of Clouds.

Dan Reed: Things are changing very quickly, with lots more to come. There will be new services, new hardware, new programming paradigms and runtimes, new operating system features and new applications.

Dilma Silva: It seems to me we will be seeing more offerings targeting enterprise workloads, enabled by advancements on security and service-level agreements. We may see more standardization of Infrastructure-as-a-Service APIs and more support for 'big data' workloads.

(9) On a related note, for a graduate student starting a PhD, what would you say are the key fundamental challenges of cloud computing that should be addressed by new research in the field?

Walfredo Cirne: Privacy and security is obviously a fruitful area. Cloud provisioning is also very much in its infancy. How can the cloud provider pack the most applications into its data center, yet fulfilling all the SLAs it has signed? In particular, which SLAs should a cloud provider be offering, and how should they be priced?

Dejan Milojicic:

 You mentioned earlier SLAs—there are little guarantees about QoS of Cloud services. New Cloud services will have higher requirements in terms of real-time, performance, availability, etc.



- Interconnects—data is still transferred using Fedex, hopefully optics and new generation of networks will alleviate this limitation.
- Data-intensive computing—the new types of unstructured data will require new types of services analyzing and managing these vast amounts of data that will be increasingly generated. Storing these data, accessing and managing will be key challenge.

Raghu Ramakrishnan: High availability, data management with richer capabilities and semantics, auto-tuning, elastic scalability, security, multi-tenancy, ... In general, if you pick a key systems problem from the past two decades, and add some of the twists peculiar to cloud systems, I'd say there is probably a thesis in there somewhere.

Dan Reed: Cloud computing challenges almost every aspect of our research agenda. As I noted at the outset, when the ratios of the capacities and capabilities change, the possible answers change. We are now building cloud infrastructures bigger than the entire Internet was just a few years ago. This is bring profound change to programming models, multivariate SLA languages, reliability/performance, low power hardware, scalable networking, and requiring new ways to analyze, visualize and extract knowledge from massive data.

Dilma Silva: Security continues to present significant challenges, with a cloud platform in some cases magnifying the usual web vulnerabilities. But things get to be more interesting when approaches for security and privacy are explored in the context of other parts of the cloud ecosystem such as image management (with the challenges introduced by the rapid proliferation of images) or virtualization support (for example, current work on detecting that an unpatched library or application is going to be executed).

There are many opportunities to revisit resource management mechanisms and policies in the context of supporting quality-of-service guarantees for cloud workloads. Novel optimizations in the networking and storage arena may be possible if one focuses on supporting usage patterns and requirements introduced by cloud environments. For all these, cloud economics may add new constraints to the old problems.

In recent years system management received much more attention from academia, with interesting results appearing in top operating systems conferences. Cloud computing motivates more work in the area, for example searching for novel approaches for problem determination and hands-off system maintenance that could bring considerable savings for cloud providers.

One exciting source of problems is the sheer scale introduced by cloud platforms. A single application can involve thousands of nodes and the placement of virtual machines and storage implemented by the underlying cloud platform may introduce locality challenges. These of course constitute well-known territory in distributed systems, but the fact that they are common place in a cloud environment and several instances of search workloads co-exist introduce new challenges and opportunities.

Programming models and software engineering practices targeting cloud environment are still emerging. For some application domains, a map-reduce approach is an effective way of using cloud resources. But some business intelligence and analytics applications that operate on large data sets do not map well to existing scale-out paradigms.

3 Short bios and photos



Dan Reed is Microsoft's Corporate Vice President for Technology Strategy and Policy and Extreme Computing. In this role, he drives Microsoft's long-term vision for technology innovations and the company's associated policy engagement with governments and institutions around the world. Before

joining Microsoft, Dr Reed held a number of strategic positions, including Head of the Department of Computer Science and Director of the National Center for Supercomputing Applications at the University of Illinois at Urbana - Champaign (UIUC), Chancellor's Eminent Professor at the University of North Carolina (UNC) at Chapel Hill and Founding Director of UNC's Renaissance Computing Institute (RENCI). In addition to his pioneering career in technology, Dr Reed has also been deeply involved in policy initiatives related to science, technology and innovation. He served as a member of the US President's Council of Advisors on Science and Technology (PCAST) and chair of the computational science subcommittee of the President's Information Technology Advisory Committee (PITAC). Dr Reed received his PhD in computer science from Purdue University.





Dr Dejan Milojicic is a senior researcher and director of Open Cirrus Cloud Computing testbed at HP Labs. He has worked in the areas of operating systems, distributed systems, and service management for more than 25 years. He has been the program chair of the IEEE Agent Systems and Applications

Symposium (ASA/MA'99) and of the first USENIX Workshop on Industrial Experiences with System Software (WIESS'2000). Dr Milojicic published in many journals and conferences. He is an inaugural editor in chief of IEEE Computing Now, a front end to all IEEE publications, and on editorial board of IEEE Internet Computing. He has been engaged in various standardization bodies, such as OMG and Global Grid Forum. He is an ACM distinguished engineer, IEEE Fellow and member of USENIX. He received his BSc and MSc from University of Belgrade and his PhD from University of Kaiserslautern. Prior to HP Labs, Dejan worked at Institute "Mihajlo Pupin", Belgrade and at OSF Research Institute, Cambridge, MA.



Raghu Ramakrishnan is Chief Scientist for Search and Cloud Platforms at Yahoo!, and is a Yahoo! Fellow, heading the Web Information Management research group. His work in database systems, with a focus on data mining, query optimization, and web-scale data management, has influenced query optimization in

commercial database systems and the design of window functions in SQL:1999. His paper on the Birch lustering algorithm received the SIGMOD 10-Year Test-of-Time award, and he has written the widely-used text "Database Management Systems" (with Johannes Gehrke). His current research interests are in cloud computing, content optimization, and the development of a "web of concepts" that indexes all information on the web in semantically rich terms. Ramakrishnan has received several awards, including the ACM SIGKDD Innovations Award, the ACM SIGMOD Contributions Award, a Distinguished Alumnus Award from IIT Madras, a Packard Foundation Fellowship in Science and Engineering, and an NSF Presidential Young Investigator Award. He is a Fellow of the ACM and IEEE.

Ramakrishnan is on the Board of Directors of ACM SIGKDD, and is a past Chair of ACM SIGMOD and member of the Board of Trustees of the VLDB Endowment. He

was Professor of Computer Sciences at the University of Wisconsin – Madison, and was founder and CTO of QUIQ, a company that pioneered crowd-sourcing, specifically question-answering communities, powering Ask Jeeves' AnswerPoint as well as customer-support for companies such as Compaq.



Walfredo Cirne is with Google's infrastructure group in California, USA, where he works on cluster management. He is on leave of his faculty position at the Computer Science Department of the Universidade Federal de Campina Grande, in Brazil. Dr Cirne holds a PhD from the University of California San

Diego, in the USA. Since 1997, his research focuses on Distributed Systems and Resource Management, having led the OurGrid project from 2001 to 2006. Further information and the publications of Dr Cirne can be found at http://walfredo.dsc.ufcg.edu.br/index_en.html.



Dilma da Silva is a researcher at the IBM T.J. Watson Research Center, in New York. She manages the Advanced Operating Systems group and is also Principal Investigator in the Exascale Collaboratory in Ireland. She received her PhD in Computer Science from Georgia Tech in 1997. Prior to joining IBM, she was an

Assistant Professor at University of Sao Paulo, Brazil. Her research in operating systems addresses the need for scalable and customizable system software. She has published more than 60 technical papers. Dilma is a member of the board of CRA-W (Computer Research Association's Committee on the Status of Women in Computing Research) and a co-founder of the Latinas in Computing group. More information is available at www.research.ibm.com/people/d/dilma.

References

 Zhang Q, Cheng L, Boutaba R (2010) Cloud computing: state-ofthe-art and research challenges. J Internet Serv Appl 1(1):7–18

