

Mapeamento das Publicações Acadêmico-Científicas sobre *Cloud Computing*

Débora Cabral Nazário^{1,2}, Paulo Silva¹, Mário Antonio Ribeiro Dantas¹

¹Engenharia e Gestão do Conhecimento (EGC)
Universidade Federal de Santa Catarina (UFSC)
Florianópolis – SC – Brasil

²Departamento de Ciência da Computação (DCC)
Universidade do Estado de Santa Catarina (UDESC)
Joinville – SC – Brasil

debora.nazario@gmail.com, paulosilvafurb@gmail.com, mario@inf.ufsc.br

Abstract. *Conceptual aspects of cloud computing are presented in this article and a bibliometric survey is developed about this subject. Different issues of cloud computing discussed in some of the cited articles from the bibliometric survey are commented in this article. The results leads to an increasing research subject, with considerable amount of different issues already published showing some challenges and aspects for further publications.*

Resumo. *Este artigo aborda aspectos conceituais de cloud computing e desenvolve uma pesquisa bibliométrica sobre o tema. São comentados os diferentes assuntos de cloud computing discutidos entre alguns dos artigos mais citados, dentro da pesquisa bibliométrica realizada. Os resultados apontam para um tema de pesquisa em ascensão, com uma grande quantidade de diferentes assuntos já publicados sobre o tema cloud computing e destacam alguns desafios e aspectos válidos para publicações futuras.*

1. Introdução

O NIST – *National Institute of Standards and Technology* conceitua *cloud computing* como um modelo de fornecimento de acesso à rede sob demanda com compartilhamento de recursos computacionais configuráveis que podem ser rapidamente alocados e liberados com o mínimo de interação com o provedor [Zhang, Cheng, Boutaba, 2010].

Alguns aspetos relacionados com *cloud computing* são o baixo investimento inicial e custo de operação, a alta escalabilidade fornecida aos clientes e a redução de riscos ao negócio, devido à alta disponibilidade do ambiente de *cloud*. Além destes aspectos existem também tecnologias distintas que se relacionam intimamente com o *cloud computing*, como: *grid computing*, *utility computing*, virtualização e *autonomic computing*.

Nos últimos anos muito se tem publicado sobre *cloud computing*, inclusive abordando diferentes conceitos, características e aspectos tecnológicos físicos e lógicos. Estas publicações evoluíram ao longo do tempo formando uma base de conhecimento sobre o estudo de *cloud computing* no meio científico. Este artigo visa recuperar e

analisar esta base de conhecimento sobre publicações de *cloud computing* e apresentar um panorama bibliométrico sobre o tema.

O artigo está organizado da seguinte maneira, a segunda seção apresenta aspectos conceituais sobre *cloud computing*, descrevendo alguns artigos publicados sobre o assunto, a terceira seção apresenta uma fundamentação relacionada com bibliometria, os procedimentos metodológicos da bibliometria realizada sobre *cloud computing* são descritos na seção 4. A seção 5 apresenta os resultados da pesquisa bibliométrica sobre *cloud computing* e a seção 6 comenta algumas considerações finais.

2. Cloud Computing: aspectos conceituais

Höfer e Karagiannis (2011) descrevem uma arquitetura de *cloud computing* dividida em quatro camadas: *hardware*, infraestrutura, plataforma e aplicação. Nesta arquitetura os serviços oferecidos pelo *cloud* podem estar organizados de três maneiras diferentes, conforme Figura 1.

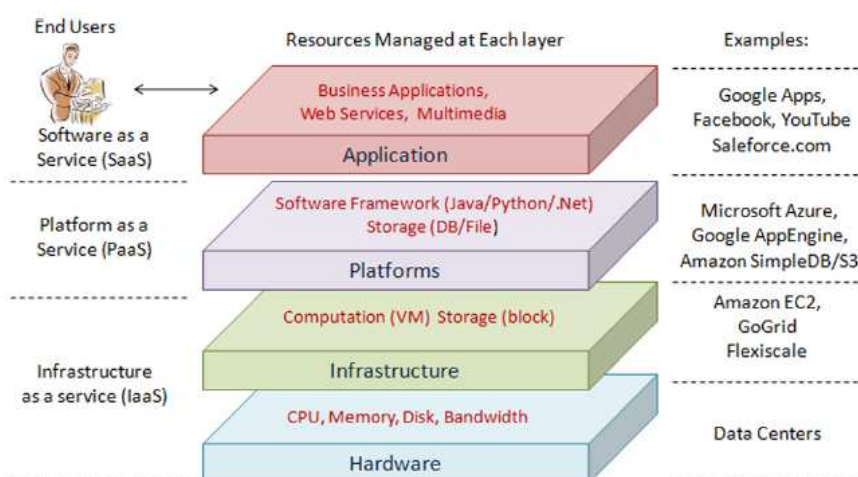


Figura 1. Arquitetura de *Cloud Computing*. Fonte: Höfer e Karagiannis (2011)

A primeira forma de organização dos serviços de *cloud computing*, IaaS – infraestrutura como serviço, consiste no fornecimento apenas de virtualização de recursos de armazenamento e processamento. Uma característica importante desta forma de serviço é a associação dinâmica de recursos físicos compartilhados entre diferentes clientes. Exemplos deste tipo de serviços são a *Amazon EC*, *GoGrid* e *GigaSpaces*.

A segunda forma de organização dos serviços de *cloud computing*, PaaS – plataforma como serviço, fornece aos clientes mais do que infraestrutura, nesta organização já estão disponíveis sistemas operacionais e aplicações básicas padrões como: Java, python, .net, etc. Exemplos deste tipo de organização são: *MS-Azure*, *Google AppEngine* e *Amazon SimpleDB*.

Ainda segundo a arquitetura de Höfer e Karagiannis (2011) a terceira forma de organização dos serviços de *cloud computing* é a SaaS – software como serviço. Nesta organização são oferecidas aos clientes uma estrutura completa com *hardware*,

infraestrutura, plataforma e aplicações. Exemplos desta organização são o *google docs*, o *CRM Salesforce.com* e o *facebook*.

Outros artigos com significativa relevância científica (segundo estudo bibliométrico) têm discutido diferentes aspectos do *cloud computing* nos últimos anos. Buyya et al (2009) comenta que os avanços científicos realizados nas últimas décadas colocarão a computação em uma importante posição de utilidade global. Para tanto, um grande número de paradigmas de computação tem sido propostos, dentre estes um dos mais importantes é o *cloud computing*. Neste contexto, o artigo comenta que um aspecto nas arquiteturas de *cloud* é a orientação ao mercado, ou seja, os modelos de *cloud* devem ser desenvolvidos com foco na demanda dos clientes para tanto o SLA – acordo de nível de serviços é um elemento fundamental. Outro elemento destacado por Buyya et al (2009) é o uso da virtualização nos ambiente de *cloud* para fornecimento da escalabilidade e disponibilidade da qual estes ambientes necessitam.

Haynes (2008) apresenta uma abordagem mais corporativa sobre o assunto de *cloud computing*. O autor comenta o fenômeno da migração de dados e aplicativos dos computadores dos usuários para os grandes *data centers* de *cloud computing* e discute o impacto de tal fenômeno para os desenvolvedores de software, gerentes de TI e fornecedores de *hardware*. Para Haynes, esta é a segunda onda de *cloud computing* na computação, pois 50 anos atrás algo semelhante aconteceu com a migração dos dados dos usuários para os grandes *mainframes*, porém uma diferença importante nesta segunda onda é que os usuários (clientes do *cloud*) possuem uma grande liberdade e flexibilidade no gerenciamento de seus dados e aplicativos, fato que não ocorria na era dos *mainframes*. Entre as várias características do *cloud computing*, Haynes destaca a importância da escalabilidade, pois caso o *software* necessite de mais recursos para desenvolver suas tarefas ele não interrompe seu funcionamento por falta de recursos, mas sim os recursos são disponibilizados ao *software* pelo *cloud*. Questões que ainda estão em aberto sobre o *cloud*, segundo Haynes, são a privacidade dos dados e a segurança e confiabilidade dos ambientes de *cloud*.

Vaquero et al (2009) foca especificamente no conceito de *cloud computing*. O artigo apresenta um estudo que consolida mais de 20 diferentes definições de *cloud* e gera um novo conceito com base neste estudo e nas principais características discutidas pela comunidade científica sobre o assunto. O conceito final apresentado pelo artigo é de que *cloud computing* é um grande conjunto ou repositório de recursos virtuais que apresentam facilidade de acesso e uso. Estes recursos devem ser dinâmicos e configuráveis e devem se ajustar a diferentes escalas de necessidades permitindo a otimização dos recursos compartilhados. Este conjunto ou repositório de recursos normalmente é contratado através de um SLA e o pagamento é feito por uso sob demanda.

Langmead et al (2009) apresenta uma ferramenta de *cloud computing* – *Crossbow* – que executa sobre o *middleware Hadoop* e é capaz de processar grandes quantidades de dados (relacionadas com genoma humano) em tempo viável a um custo consideravelmente baixo em relação a outras soluções existentes.

Um aspecto bastante distinto da aplicação e evolução do *cloud computing* é apresentado por Robinson (2009). Seu artigo discute a geração de lixo eletrônico e o impacto de tal tipo de lixo para o meio ambiente. O artigo cita o *cloud computing*, entre

outras tecnologias, como possível redutor da produção de lixo eletrônico e conseqüentemente benéfico para o meio ambiente. Isto ocorre porque o *cloud* reduz a quantidade de dispositivos eletrônicos junto ao usuário e concentra a maior parte dos dispositivos eletrônicos em grandes aglomerados, onde seria mais fácil aplicar a destinação correta aos equipamentos eletrônicos obsoletos.

Nos artigos comentados acima foi possível observar uma pequena amostra da amplitude de aspectos que são publicados cientificamente sobre o tema *cloud computing*. A amplitude de assuntos e a evolução destes durante o tempo motivam o desenvolvimento de uma pesquisa bibliométrica sobre o assunto.

3. Técnicas Bibliométricas

A Bibliometria é uma ferramenta estatística que permite mapear e gerar diferentes indicadores de tratamento e gestão da informação e do conhecimento, especialmente em sistemas de informação e de comunicação científicos e tecnológicos, e de produtividade, necessários ao planejamento, avaliação e gestão da ciência e da tecnologia, de uma determinada comunidade científica ou país. É um instrumento quantitativo que permite minimizar a subjetividade inerente à indexação e recuperação das informações, produzindo conhecimento, em determinada área de assunto. A Bibliometria contribui para tomadas de decisão na gestão da informação e do conhecimento, uma vez que auxilia na organização e sistematização de informações científicas e tecnológicas [Guedes e Borschiver, 2005].

De acordo com Santos, Maldonado e Santos (2011), o princípio básico subjacente às abordagens bibliométricas é que a comunicação científica é um dos principais resultados da pesquisa, constituindo um aspecto central da ciência. Embora as publicações não sejam o único indicador da atividade científica, são certamente elementos muito importantes do processo de troca de conhecimento. Para isso, as técnicas bibliométricas utilizam *corpus* de literatura para realizar suas análises e também se preocupam com o aspecto da temporalidade. Padrões de comportamento do campo de estudo podem ser visualizados ao longo do tempo. Na prática, os estudos bibliométricos baseiam-se em indicadores e em técnicas de visualização de informação da correlação dos elementos bibliográficos disponíveis nas bases de dados.

A bibliometria foi escolhida por ser um método planejado que possibilita coletar, selecionar e analisar criticamente os estudos. No presente trabalho utilizam-se as técnicas bibliométricas para realizar um mapeamento das publicações acadêmico-científicas sobre o tema *Cloud Computing*.

As fontes de um estudo de bibliometria são, portanto, publicações provenientes de estudos originais disponíveis em um banco de dados. A base de dados utilizada é a *Web of Science* considerada pela comunidade científica uma boa fonte para estudos bibliométricos. Nesta base encontra-se conteúdo científico multidisciplinar datado desde 1900, estes são relevantes tanto pela quantidade como pela qualidade dos periódicos científicos indexados.

4. Procedimentos Metodológicos

Este artigo tem natureza exploratória de caráter descritivo [Vergara, 2003], com a utilização de técnicas bibliométricas. O processo de desenvolvimento deste estudo

consistiu em três etapas: coleta de dados, análise de dados e síntese e representação dos dados. Os passos seguidos são descritos a seguir.

Como passo inicial tem-se a escolha da base de dados, devido ao seu reconhecimento científico e à acessibilidade aos pesquisadores foi escolhida a base *Web of Science*, incluindo as seguintes sub-bases:

- *Science Citation Index Expanded (SCI-EXPANDED): 1945-present;*
- *Social Sciences Citation Index (SSCI): 1956-present;*
- *Arts & Humanities Citation Index (A&HCI): 1975-present;*
- *Conference Proceedings Citation Index- Science (CPCI-S): 1990-present;*
- *Conference Proceedings Citation Index- Social Science & Humanities (CPCI-SSH): 1990-present;*
- *Book Citation Index-Science (BKCI-S): 2003-present;*
- *Book Citation Index- Social Science & Humanities (BKCI-SSH): 2003-present.*

Foi usado o período de busca disponível na base “*All Years*”.

Após a escolha da base de dados foram estabelecidos os critérios e realizadas as buscas. O termo utilizado na busca foi “*cloud computing*” no campo denominado *Topic* que compreende o título, as palavras-chave, e o resumo das publicações. O resultado da busca foi o conjunto com 1.304 publicações, sendo realizada no dia 25 de novembro de 2011. Esse conjunto de publicações foi utilizado para algumas análises. Para isso, o conjunto dos dados foi importado como arquivo de texto (.txt) para o software *HistCite*. O *Histcite* possibilita a organização dos dados bibliográficos provenientes da base *Web of Science*, fornecendo uma análise dos dados mais completa. As análises foram feitas a partir da elaboração de tabelas, gráficos e figuras para representações bibliográficas.

Na próxima seção apresentam-se os resultados mais relevantes sob o ponto de vista da abstração e do alto nível de agregação dos dados.

5. Resultados

Os resultados gerais do conjunto de dados analisados estão representados na Tabela 1, estes resultados foram retirados do *software HistCite*.

Tabela 1. Resultados gerais

Critérios	Quantidade
Publicações	1.304
Autores	3.396
Instituições	674
Países	58
Fontes de publicação	652
Referências citadas	14.645
Palavras-chave	2.192

Em relação à distribuição dos tipos de publicações identificou-se um predomínio de publicações em Anais de Eventos com 776 ocorrências (59,1%), seguido por artigos em periódicos científicos com 483 ocorrências (36,8%), outros tipos são menos significativos como capítulo de livro, editorial, revisão, livro, etc.

Em relação à frequência das publicações por ano, na Figura 2 é apresentada a distribuição desde o ano de 2007, com as duas primeiras publicações na *Web of Science*. Desde então o número de publicações vem aumentando, chegando a 507 publicações no ano de 2010. Em 2011 foram publicados 357 trabalhos até o momento (ano incompleto) e já constam 13 trabalhos no ano de 2012.

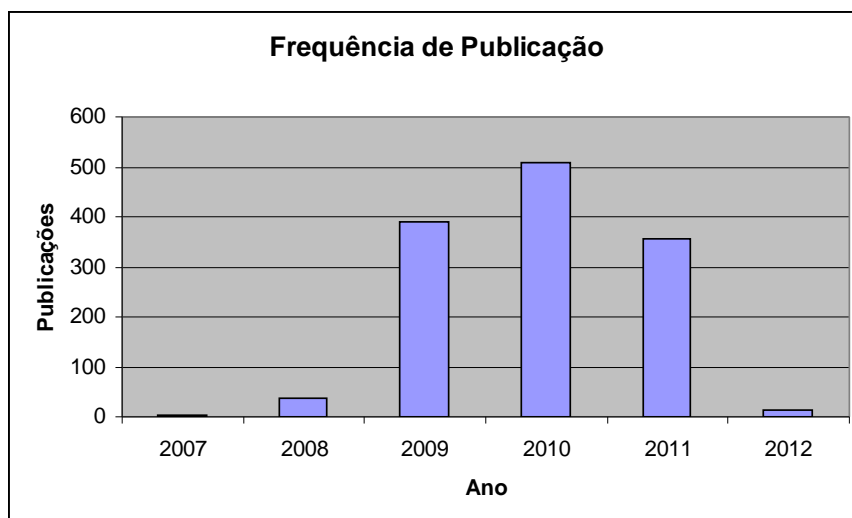


Figura 2. Frequência de Publicações por Ano

Com relação aos países dos autores, Os Estados Unidos lideram com a quantidade de 399 publicações, seguido pela China com 239 trabalhos e Alemanha com 93. O Brasil está em 21º com 10 trabalhos publicados.

As áreas na base *Web of Science* com mais publicações nos quais os trabalhos foram indexados são: Ciências de Computação (75%), seguida da área de engenharia (30%) e telecomunicações (10%).

As cinco fontes com maiores quantidades de trabalhos publicados são: *Cloud Computing, proceedings* (51 trabalhos), *IEEE Internet Computing* (23), *Computer* (20), *Cloud Computing: principles, systems and applications* (18) e *Future generation computer systems – the international journal of grid computing and escience* (18). Apesar de uma concentração um pouco maior na primeira fonte, se observa uma dispersão da quantidade de trabalhos publicados nas demais fontes.

O indicador TGCS (*Total Global Citation Score* – Escore Total Global de Citações) refere-se à quantidade total de vezes que os trabalhos foram citados por outros trabalhos da base ISI/WoS como um todo. Com base neste indicador, percebe-se que as fontes com mais citações são: *Future Generation Computer Systems – The International Journal of Grid Computing – Theory Methods and Applications* (TGCS = 135), *Communications of the ACM* (108), *IEEE Internet Computing* (70), *Computer Communication Review* (58) e *Genome Biology* (56).

A seguir, na Tabela 2 estão relacionados os dez trabalhos mais citados de acordo com o GCS (*Global Citation Score* – Escore Global de Citações) que se refere à quantidade de vezes que os trabalhos foram citados por outros trabalhos da base ISI/WoS como um todo.

Tabela 2. Publicações com mais citações.

Autor	Título	Fonte Publicação / Ano	GCS
Buyya R, Yeo CS, Venugopal S, Broberg J, Brandic I	<i>Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility</i>	<i>Future Generation Computer Systems – The International Journal of Grid Computing – Theory Methods and Applications / 2009</i>	107
Hayes B	<i>Cloud computing</i>	<i>Communications of the ACM / 2008</i>	60
Vaquero LM, Rodero-Merino L, Caceres J, Lindner M	<i>A Break in the Clouds: Towards a Cloud Definition</i>	<i>Computer Communication Review / 2009</i>	58
Armbrust M, Fox A, Griffith R, Joseph AD, Katz R, et al.	<i>A View of Cloud Computing</i>	<i>Communications of the ACM / 2010</i>	30
Li H, Homer N	<i>A survey of sequence alignment algorithms for next-generation sequencing RID D-9344-2011</i>	<i>Briefings in Bioinformatics / 2010</i>	27
Langmead B, Schatz MC, Lin J, Pop M, Salzberg SL	<i>Searching for SNPs with cloud computing</i>	<i>Genome Biology / 2009</i>	25
Robinson BH	<i>E-waste: An assessment of global production and environmental impacts</i>	<i>Science of the Total Environment / 2009</i>	22
Nurmi D, Wolski R, Grzegorzczak C, Obertelli G, Soman S, et al.	<i>The Eucalyptus Open-source Cloud-computing System</i>	<i>CCGRID: 2009 9 TH IEEE International Symposium on Cluster Computing and the Grid / 2009</i>	21
Buyya R, Yeo CS, Venugopal S	<i>Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities</i>	<i>HPCC 2008: 10 TH IEEE International Conference on High Performance Computing and Communications, Proceedings / 2008</i>	19
Sotomayor B, Montero RS, Llorente IM, Foster I	<i>Virtual Infrastructure Management in Private and Hybrid Clouds</i>	<i>IEEE Internet Computing / 2009</i>	18

Na Figura 3 são representadas as 100 palavras-chave com maior frequência de uso no total dos 1.304 trabalhos analisados. Foram retiradas as palavras *Cloud* e *Computing*, que são as mais utilizadas. Algumas das palavras-chave com maior frequência de uso são *Based*, *Data*, *Service(s)*, *Management*, *Clouds*, *Virtual*, *System*, *Applications*, *Security*, *Grid*.



Figura 3. Representação da frequência das palavras-chave.

Das 14.645 referências citadas por este conjunto de trabalhos, destacam-se os autores Armbrust M., Buyya R., Barham P., Weiss A., Foster I., Dean J., Vaquero L.M., Ghemawat S., Hayes B., Jain R.

Após a realização desta análise, é possível ter uma visão geral do tema pesquisado. Com a técnica de bibliometria pode-se partir para uma análise mais aprofundada dos trabalhos. Uma maneira possível seria reduzindo o conjunto de dados, através da escolha de algum critério (como filtro). Por exemplo, pelo tipo de publicação (ex. periódicos), mais antigos, mais recentes, mais citados, tema mais específico (ex. AND “green”, que resultam 26 publicações), ou outro critério de acordo com o objetivo do estudo.

Com um conjunto menor de trabalhos é possível uma análise mais aprofundada pela leitura de todos os resumos e/ou textos completos. Com isso, é possível identificar os desafios ao longo do período de análise, traçar um “mapa” dos temas tratados, perceber as lacunas para pesquisa. Ainda pode-se utilizar mais de uma base de dados, com alguma ferramenta apropriada que elimine os trabalhos repetidos, ampliando assim o espaço de busca.

6. Considerações Finais

Esta pesquisa bibliométrica forneceu uma visão sistêmica sobre o tema pesquisado *cloud computing* através de alguns indicadores estatísticos. Este é um tema de pesquisa em ascensão, com muitas publicações a partir 2009, principalmente em conferências.

A seção 2 comentou os temas apresentados pelos artigos que estão nas posições 1, 2, 3, 6 e 7 entre os artigos mais citados pela pesquisa bibliométrica. Nestes artigos foi possível observar a amplitude de diferentes aspectos sobre o mesmo tema – *cloud computing*. Houve artigos tratando de aspectos conceituais, corporativos, práticos, aplicados e ambientais, como na discussão sobre a relação entre lixo eletrônico e *cloud computing*.

Com base nos trabalhos mais citados (Tabela 2), percebe-se que há bastante espaço para pesquisa em tópicos como infraestrutura, comunicação, segurança e aplicações. Diversos são os desafios encontrados, como confiabilidade dos ambientes, segurança dos dados dos usuários, otimização de recursos de energia, virtualização de plataformas, entre outros.

Como trabalho futuro sugere-se uma análise aprofundada de um conjunto restrito de publicações, extraindo outras informações relevantes. Para uma maior contribuição, como extensão deste trabalho também se sugere a aplicação de técnicas de revisão sistemática de literatura.

Ainda como sugestão de trabalho futuro, pode-se utilizar técnicas bibliométricas para identificar os "temas quentes" na área da computação. Como por exemplo, detectando os temas que estão em ascensão ou descensão, grupos de pesquisa podem utilizar ferramentas desse tipo para auxílio em seus estudos.

Referências

- Buyya R., Yeo C.S., Venugopal S., Broberg J., Brandic I. (2009). "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility". *Future Generation Computer Systems – The International Journal of Grid Computing – Theory Methods and Applications*. 25 (6). pp. 599-616.
- Guedes, Vânia; Borschiver, Suzana. (2005) "Bibliometria: uma ferramenta estatística para a gestão da informação e do conhecimento, em sistemas de informação, de comunicação e de avaliação científica e tecnológica". In *Proceedings CINFORM - Encontro Nacional de Ciência da Informação VI*, Salvador - Bahia.
- Hayes B. (2008). "Cloud computing". *Communications of the ACM*. 51 (7). pp. 9-11.
- Höfer, C.N. and Karagiannis, G. (2011) "Cloud computing services: taxonomy and comparison". *Journal of Internet Services and Applications*, 2 (2). pp. 81-94.
- Langmead B., Schatz M.C., Lin J., Pop M., Salzberg S.L. (2009). "Searching for SNPs with cloud computing". *Genome Biology*. 10 (11).
- Robinson B.H. (2009). "E-waste: An assessment of global production and environmental impacts". *Science of the Total Environment*. 408 (2). pp. 183-191.

- Santos, J. L. S. S; Maldonado, M. U.; Santos, R. N. M. (2011) “Mapeamento das Publicações Acadêmico-Científicas sobre Memória Organizacional”. XXXV Encontro da ANPAD. Rio de Janeiro.
- Vaquero L.M., Rodero-Merino L., Caceres J., Lindner M. (2009). “A Break in the Clouds: Towards a Cloud Definition”. *Computer Communication Review*. 39 (1). pp. 50-55.
- Vergara, S. C. (2003) “Projetos e relatórios de pesquisa em administração”. São Paulo: Atlas.
- Zhang, Qi; Cheng, Lu; Boutaba, Raouf. (2010) “Cloud computing: state-of-the-art and research challenges”. *Journal of Internet Services and Applications* Volume 1, Number 1, 7-18.