

# Banco de Dados em Nuvem: Conceitos, Características, Gerenciamento e Desafios

Darlan Florêncio de Arruda e José Almir Freire de Moura Júnior  
Faculdade de Ciência e Tecnologia de Caruaru - Universidade de Pernambuco (UPE) - Caruaru - PE- Brasil

Luís Alexandre Tavares Silva  
Faculdade de Ciência e Tecnologia de Caruaru - Universidade de Pernambuco (UPE) - Caruaru - PE- Brasil

**Resumo** — Computação em Nuvem surge como um novo paradigma na implantação de aplicações em que os recursos computacionais são fornecidos como um serviço através de uma conexão de rede. Esses serviços são disponibilizados em forma de software, infraestruturas e plataformas. Com a utilização desses serviços usuários estão disponibilizando informações, muitas vezes confidenciais em servidores remotos desconhecidos geograficamente, ou seja, os bancos de dados em nuvem. Este artigo apresenta um visão geral desse tipo de banco de dados, assim como características e vantagens, além de tratar dos desafios encontrados no gerenciamento dos dados armazenados nesse ambiente.

**Palavras-chaves** — Banco de Dados em Nuvem, Desafios, Gerenciamento de Dados.

## I. INTRODUÇÃO

Com o avanço tecnológico, a idéia de ter, armazenar e distribuir informações de forma rápida vem chamando a atenção de muitos usuários, fazendo com que aumente a busca por esse tipo de serviço, principalmente em grandes corporações. Diante disso, a *Cloud Computing* surge como um novo paradigma na implantação de aplicações em que os recursos computacionais são fornecidos como um serviço através de uma conexão de rede. Esses serviços são disponibilizados em forma de *software*, infraestruturas e plataformas, onde usuários podem adquiri-los sob demanda [15].

Mas, como toda nova tecnologia, a *Cloud Computing* também preocupa os usuários quanto a diversos pontos relacionados ao gerenciamento e segurança dos dados que estarão armazenados na nuvem. Sistemas de gerenciamento de dados são fortes candidatos para implantação em nuvem. Pois com o uso dos serviços disponibilizados, há também o uso de uma grande quantidade de dados que precisa ser armazenada.

Sistemas de gerenciamento de dados em nuvem possuem custo reduzido e maior facilidade de acesso aos dados, entretanto existem desafios relacionados à consistência e segurança dos dados, que são importantes em ambientes *Cloud computing* e deste modo precisam ser tratados cautelosamente Sousa et al [15]. Este artigo busca enfatizar os principais desafios enfrentados no gerenciamento de informações que são armazenadas em banco de dados em nuvem.

## II. METODOLOGIA

Para Silva [14] a metodologia científica é o conjunto de processos e operações mentais que se deve empregar nas investigações. É a linha de raciocínio adotado no processo da pesquisa. Ou seja, para que uma pesquisa seja efetuada é necessário um conjunto de procedimentos intelectuais e técnicos. Uma pesquisa pode ser classificada por diversas características, que podem ser, por exemplo, quantos aos fins e quanto aos meios.

Quanto aos fins, a presente pesquisa pode ser considerada descritiva, pois de acordo com Gil [11] este tipo de pesquisa adota como objetivo primordial as descrições das características de uma determinada população ou determinado fenômeno ou o estabelecimento de relações entre variáveis.

Quanto aos meios, esse trabalho possui características de pesquisa bibliográfica. Segundo Silva [14], pesquisa bibliográfica é elaborada a partir de material já publicado, constituído principalmente de livros, artigos de periódicos e atualmente com material disponibilizado na internet.

## III. TRABALHOS RELACIONADOS

Algumas publicações encontradas na literatura tratam de gerenciamento de dados na nuvem, bem como suas limitações, desafios e as oportunidades que surgem como agentes impulsionadores para a solução desses desafios. Abadi [1], por exemplo, apresenta uma discussão sobre as limitações e possibilidades de implantação de técnicas de gerenciamento de dados em plataformas de *Cloud Computing* emergentes, como a *Amazon Web Services*, por exemplo. No referido trabalho também são apresentadas algumas características que um Sistema de

Gerenciamento de Dados em Nuvem deve possuir quando se é projetado para o armazenamento de dados em grande escala. Wei et al. [16], por sua vez, enfatiza a importância da escalabilidade e da alta disponibilidade em serviços de dados em nuvem. Entretanto para garantir essas características ele cita que ocorrem alguns problemas como consistência de dados e segurança dos dados, que devem ser resolvidos, porém no referido artigo não são explanadas soluções para a resolução dos mesmos.

Em seu trabalho, Sousa et al. [15] aborda os principais conceitos de computação em nuvem com ênfase para os desafios no gerenciamento de dados nesse tipo de ambiente, entre os quais são citados escalabilidade e consistência, qualidade dos serviços de dados, e segurança dos serviços de dados.

Além de explicar problemas similares, os trabalhos supracitados têm em comum o fato de apontarem desafios sem, todavia fazerem propostas efetivas de soluções para a resolução dos mesmos. A seção seguinte traz uma visão geral do conceito de Computação em Nuvem.

#### IV. COMPUTAÇÃO EM NUVEM

Segundo Nist [13], Computação em nuvem é um modelo que possibilita acesso, de modo conveniente e sob demanda, a um conjunto de recursos computacionais configuráveis (por exemplo, redes, servidores, armazenamento, aplicações e serviços) que podem ser rapidamente adquiridos e liberados com mínimo esforço gerencial ou interação com o provedor de serviços. Este modelo de nuvem promove disponibilidade e é composto por cinco características essenciais, três Modelos de Serviços (SaaS, PaaS e IaaS), e quatro Modelos de Implantação (privada, pública, comunitária e híbrida)1.

Diante disso, a Computação em Nuvem apresenta-se como uma maneira bastante eficiente de maximizar e flexibilizar os recursos computacionais, diante da multiplicidade de serviços prestados pelos provedores de Computação em Nuvem, tais como, armazenamento de dados, desenvolvimento de aplicativos personalizados e gestão de infraestrutura, por exemplo.

Cada parte desta infraestrutura é provida como um serviço e, estes são normalmente alocados em centros de dados, utilizando *hardware* compartilhado para computação e armazenamento Buyya et al [4]. Na seção a seguir, serão explanadas as principais características dos Banco de Dados em Nuvem.

#### V. BANCO DE DADOS EM NUVEM (BDN)

Banco de Dados em Nuvem (BDN) [1] estão começando a ser utilizados e têm o potencial de atrair clientes de diversos setores do mercado, desde pequenas empresas com o objetivo de reduzir o custo total, por meio da utilização de infraestrutura e sistemas de terceiros, até grandes empresas que buscam soluções para gerenciar milhares de máquinas e permitir o atendimento de um aumento inesperado de tráfego.

Segundo Cottman [6], há uma palavra que resume bem a origem dos BDN: volume. Sendo que esse volume se torna um grande desafio na concepção de um BDN, visto que ainda existem problemas de escalabilidade nos ambientes de *Cloud Computing*. A evolução da Internet tem gerado grandes volumes de dados (terabytes, petabytes, e assim por diante) que surpreendem até mesmo a mente de um profissional de TI, em contrapartida os bancos de dados relacionais convencionais não se mostraram suficientemente preparados para lidarem da melhor forma possível com um quantitativo de dados dessa magnitude. Não é por acaso que quando se

pergunta como portais bem estabelecidos como o Google, o Facebook e o Twitter poderiam armazenar, processar e arquivar essa vasta quantidade de dados, a resposta é uma só: através de BDN.

Apesar do uso cada vez mais generalizado do termo BDN, não foi encontrada na literatura uma definição formal para o mesmo. Não obstante, em linhas gerais pode-se defini-lo como uma coleção de dados inter-relacionados que estão armazenados na web, e que podem ser gerenciados e manipulados através de Sistemas de Gerenciamento de Dados em Nuvem (SGDN), softwares estes que segundo Cottman [6] são especialmente concebidos para atingir o processamento massivamente paralelo – na ordem das centenas de milhares de tarefas simultâneas –, utilizando os recursos distribuídos de computação em grid.

Para Barros [3], conexões de internet mais rápidas favorecem a adoção de BDN. A mesma possibilitará aos sistemas estarem hospedados na web. Como um banco de dados totalmente online, operações como backups, restores, pesquisas, inserções e deleções poderão ser realizadas facilmente por intermédio de qualquer browser.

Barros [3] aponta ainda um crescimento da procura por BDN tanto a nível acadêmico quanto profissional.

Uma pesquisa realizada pela Embarcadero Technologies [10] reforça esta tese. Segunda a pesquisa - realizada com 1230 profissionais da área de banco de dados - com o objetivo de recolher informações sobre as tendências de banco de dados, desafios, principais iniciativas e ferramentas atuais que estão sendo utilizadas, pôde-se verificar que a tecnologia de BDN será a que mais afetará o setor de banco de dados.

33,6% dos entrevistados citaram BDN como a tecnologia que exercerá maior impacto na comunidade. Também foram citadas outras tecnologias como virtualização, discos de estado sólido e tuning visual, conforme ilustrado na figura 1.



Fig.1. Tecnologias impactantes na comunidade de banco de dados (Fonte: Embarcadero, 2010)

Tecnologias que envolvem BDN possuem um alto grau de escalabilidade, fazendo com que as aplicações cresçam de acordo com a sua demanda de utilização. Esse crescimento se dá uniformemente - a escalabilidade em BDN implica diretamente no desempenho do sistema utilizado. Deste modo, pode-se dizer que a utilização de BDN possui algumas vantagens em relação aos Bancos de Dados convencionais.

Dentre essas podem-se destacar a flexibilidade/escalabilidade, a tecnologia de ponta utilizada, o acesso à informação em qualquer lugar (informação

distribuída) e o custo. A seção seguinte aponta a importância do Gerenciamento de Dados em Nuvem, bem como algumas de suas principais características.

## VI. GERENCIAMENTO DE DADOS EM NUVEM

O gerenciamento de dados é um fator muito importante dentro do contexto de computação em nuvem, uma vez que a segurança desses dados é fator crucial em ambientes de *Cloud Computing* e deve ser tratado com certa atenção. Outro fator relevante é que os Sistemas de Gerenciamento de Banco de Dados (SGBDs) relacionais não possuem escalabilidade quando existe uma grande quantidade de dados armazenados por eles.

Assim, aspectos de armazenamento de dados, processamento de consultas e controle transacional têm sido flexibilizados por algumas abordagens para garantir a escalabilidade, mas ainda não existem soluções que combinem estes aspectos de forma a melhorar o desempenho sem comprometer a consistência dos dados [1].

A escalabilidade do sistema deve ser transparente para os usuários, podendo estes armazenar seus dados na nuvem sem a necessidade de saber a localização dos dados ou a forma de acesso [15].

A escalabilidade do sistema ocorre quando o mesmo possui a capacidade de crescer conforme a demanda de uso. Já a disponibilidade do serviço possibilita aos usuários acesso aos dados quando e onde quiserem ou precisarem, os Sistemas de Gerenciamento de Dados em Nuvem (SGDN) devem dispor de uma alta disponibilidade, visto que o meio de comunicação entre o usuário e o sistema é a internet, e que pode ocorrer atrasos e indisponibilidade do sistema.

Outro item relevante no contexto de gerenciamento de dados em nuvem é a consistência de dados, onde todos os nós devem ter a mesma visão dos dados ao mesmo tempo, fazendo assim, com que todos os dados armazenados sejam iguais para todos os nós do sistema, por exemplo, se um determinado dado for atualizado, então ele deverá ser atualizado para todos os nós do sistema, mantendo assim os dados íntegros, coerentes, consistentes.

Existem diversas ferramentas que viabilizam o gerenciamento de dados em nuvem, algumas das mais relevantes são apresentadas a seguir.

## VII. SISTEMAS DE GERENCIAMENTO DE DADOS EM NUVEM (SGDN)

Em linhas gerais, um Sistema de Gerenciamento de Dados em Nuvem (SBDN) pode ser definido como um ou mais programas que possibilita(m) a criação, manipulação e gerenciamento de Bancos de Dados em Nuvem.

Segundo Sousa [15] a infraestrutura de SGDN possui várias vantagens para os usuários, como: previsibilidade e custos mais baixos, proporcional à qualidade do serviço (QoS) e cargas de trabalho reais, complexidade técnica reduzida, graças a interfaces de acesso unificado e a delegação de *tuning* e administração de SGBDs e a elasticidade<sup>5</sup> e escalabilidade, proporcionando a percepção de recursos quase infinitos.

Existem diversos sistemas para o gerenciamento de dados em nuvem. Em uma enquête realizada pela Dzone[9] puderam-se verificar quais são os sistemas de gerenciamento de dados em nuvem que estão sendo utilizados ou despertam interesse para adoção. O resultado dessa enquête pode ser visto na figura 2

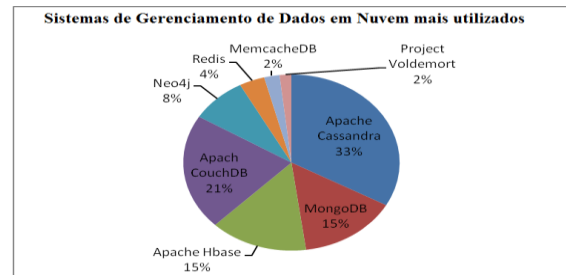


Fig. 2. SGBDs em Nuvem mais utilizados (Fonte: Dzone, 2010)

Estes dados são referentes a um universo de 736 pessoas que responderam a pesquisa, não representando a mesma proporção entre o uso dos SGDN. Conforme evidencia a figura 2, dentre os SGDN mais comuns destacam-se o Cassandra, o CouchDB, o MongoDB e o Microsoft SQL Azure. A seguir são apresentadas de forma resumida algumas das principais características desses sistemas.

### A. CASSANDRA

O Apache Cassandra [5] é um sistema de armazenamento distribuído para o gerenciamento de grandes quantidades de dados espalhados por centenas de máquinas. Ele pode funcionar em hardware de baixo custo e lida com alta taxa de escrita sem sacrificar a eficiência na leitura, é um sistema tolerante a falhas, fazendo com que aumente sua confiabilidade. Recentemente ganhou popularidade por ser utilizado, entre outros, pelo Twitter e pelo Facebook.

### B. COUCHDB

O Apache CouchDB [7] é um SGDN orientado a documento<sup>6</sup> que possui uma série de características que torna sua utilização viável em servidores que possuem hardware de baixo desempenho e utiliza técnicas de armazenamento e controle de concorrência baseadas na estrutura do documento.

### C. MONGODB

O MongoDB [12] é um SGDN escalável e de alto desempenho, de código aberto, orientado a documento, escrito em C++ e que não possui transações ou junções.

### D. MICROSOFT SQL AZURE

O Microsoft SQL Azure [2] é um SGDN composto por um conjunto de serviços para o armazenamento e processamento de dados em nuvem. Juntamente com o Windows Azure Storage formam o SGDN da Microsoft. O SQL Azure implementa alta disponibilidade, tolerância a falhas e o

conceito de multi-inquilino. O SQL Azure foi projetado com base em tecnologia de SGBD relacional do SQL Server onde seu principal componente é o SQL Azure Database.

## VIII. ANÁLISE COMPARATIVA

O quadro I mostra os resultados de uma análise comparativa - realizada pelos autores do presente artigo - entre os SGBD explorados no trabalho. Os dados presentes na análise são oriundos de pesquisas realizadas nas páginas oficiais dos sistemas supracitados. No comparativo foram levadas em consideração algumas características essenciais para esses ambientes, como distribuição, modelo de dados, armazenamento de dados, escalabilidade, entre outras.

QUADRO I. COMPARATIVO ENTRE SISTEMAS DE GERENCIAMENTO DE DADOS EM NUVEM

Características	Apache Cassandra	Couch DB	Mongo DB	Microsoft SQL Azure
Distribuição dos dados	Distribuído	Distribuído	Distribuído	Distribuído
Modelo dos dados	Orientado a coluna	Orientado a documento	Orientado a documento	Relacional
Armazenamento dos dados	Índices	Árvore B+	Índices	Tabelas
Custo	Proprietário	Open Source	Open Source	Proprietário
Escrito em	JAVA	Erlang OTP	C++	-
Escalabilidade	Forte	Eventual	-	Eventual
Quem Usa	Facebook, e Twitter, Digg	Meebo e BBC	Engine Yard	-

Pelo quadro acima nota-se nitidamente que a escalabilidade ainda não atingiu um grau satisfatório em pelo menos metade dos SGBD analisados, apontando a escalabilidade como um desafio importante a ser superado na área de BDN.

Por outro lado, algumas outras características relevantes não puderam estar presentes no quadro 1 tais quais *segurança* e *qualidade de serviço*, visto que ainda não se mostraram suficientemente maduras no contexto de gerenciamento de BDN, e por isso também se apresentam como desafios relevantes para a referida área.

A seção seguinte aponta quais seriam os principais desafios para o gerenciamento de BDN nos dias atuais.

## IX. DESAFIOS PARA O GERENCIAMENTO DE BANCO DE DADOS EM NUVEM

A Computação em Nuvem proporciona diversas vantagens à aqueles que a utiliza, porém mediante a tantas vantagens, surgem diversos desafios que devem ser superados nesse tipo de ambiente, entre os quais podem-se destacar a *segurança da informação*, a *escalabilidade* e a *consistência de dados*, *qualidade do serviço de dados*, entre outros.

### A. SEGURANÇA

Para Dillon [8] a segurança tem tido um papel muito importante no impedimento do desenvolvimento da Computação em Nuvem. Questões de segurança tradicionais

já bastante conhecidas e tratadas em sistemas distribuídos, como controle de acesso, confidencialidade, integridade de dados, tolerância a falhas, também fazem parte desta nova tecnologia.

Por utilizar a internet como meio para a disponibilização dos serviços, isto se torna um tanto complexo, diante da diversidade de recursos que são utilizados, como por exemplo, sistemas operacionais distintos, *softwares*, domínio de redes, políticas de segurança da informação. A percepção de que a Nuvem é um aglomerado de informações pode caracterizá-la como sendo um alvo propício a ataques.

Ameaças como estas podem afetar diretamente os pilares da segurança da informação: disponibilidade, confidencialidade, integridade e não-repúdio, e conseqüentemente comprometer toda a Nuvem. A segurança da informação deve ser tratada como peça-chave no desenvolvimento de BDN, uma vez que proporciona a autenticidade, integridade e confidencialidade ao sistema.

### B. ESCALABILIDADE E CONSISTÊNCIA DE DADOS

As soluções em nuvem focam muito em escalabilidade e com isso oferecem uma fraca consistência dos dados armazenados. Os SGBD devem ser altamente escaláveis, pois não se pode definir uma quantidade de dados a serem armazenados, e a escalabilidade proporcionará o crescimento do sistema de acordo com a demanda de uso, porém não garante a consistência dos dados por eles armazenados.

A maioria das soluções de BDN utiliza a consistência eventual, que é um tipo de consistência fraca. Este tipo de consistência não permite a construção de uma ampla gama de aplicações, tais como serviços de pagamento e leilões *online*, que não podem trabalhar com dados inconsistentes [16].

Dessa maneira, a consistência dos dados torna-se assunto crucial a ser analisado e tratado quando se for projetar aplicações e soluções de BDN, sendo visto como um desafio que deve ser superado para que sejam projetados SGBD mais eficientes quanto ao seu uso, e sobretudo com maior grau de integridade e correção dos dados.

### C. QUALIDADE DE SERVIÇOS DE DADOS

A qualidade do serviço é característica essencial em ambientes de *Cloud computing*, proporcionando aos usuários algumas garantias como disponibilidade e desempenho. A disponibilidade de serviços permite aos usuários acessar e utilizar a nuvem onde e quando desejarem.

Dessa forma, os ambientes de Computação em Nuvem devem prover alta disponibilidade, que pode acontecer através de garantias de *QoS (Quality of Service)*. Os sistemas em nuvem apesar de possuir algumas limitações em termos de segurança e de rede, devem fornecer desempenho elevado, e ser muito flexíveis para se adequar a uma grande quantidade de requisições.

O desempenho passa a ser um desafio considerável em SGBD, pois como alguns são de acesso público não há como dimensionar a quantidade de requisições realizadas, aumentando a dificuldade em fazer estimativas e obter garantias de *QoS*.

Deste modo, o desempenho se apresenta como mais um grande e importante desafio a ser superado para que esses sistemas possam ser executados em sua amplitude.

## X. CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho abordou os principais aspectos e características dos Bancos de Dados em Nuvem, enfatizando os desafios que são encontrados no gerenciamento dos dados que são armazenados em ambientes deste tipo.

Embora essa abordagem traga diversas vantagens para quem a usa, ainda existem diversos desafios que precisam ser estudados e superados. Dentre os mesmos destacam-se a *segurança*, a *consistência dos dados*, a *escalabilidade* e as garantias de *qualidade do serviço*, que são essenciais para o bom funcionamento de um sistema que faça uso de um BDN. Neste cenário, é importante fazer uma análise das vantagens oferecidas por essa tecnologia e as ameaças que podem ser geradas quando se migra para esse tipo de ambiente.

Como sugestão de trabalhos futuros destaca-se a necessidade de ampliar o estudo através de análises mais profundas sobre segurança da informação em BDN, também verificar e expandir métodos que garantam a consistência dos dados armazenados mantendo a qualidade do serviço, além de garantir a escalabilidade do sistema, visto que esta última se trata de uma das principais características e necessidades de aplicações em nuvem.

Além de analisar outros desafios que por hora não são os mais importantes para o desenvolvimento da área, mas, que possam contribuir para a qualidade dos Sistemas de Gerenciamento de Dados em Nuvem.

## REFERÊNCIAS

- [1] Abadi, D. J. *Data management in the cloud: Limitations and opportunities*. IEEE Data Eng. Bull., 32:3–12, 2009.
- [2] Azure. *Microsoft Azure*. Disponível em: <http://www.microsoft.com/azure/>. Acessado em: 05 de janeiro de 2011.
- [3] Barros, L. E. B. *Banco de Dados em Nuvem*. Disponível em: <http://pesquompile.wikidot.com/banco-de-dados-em-nuvem>. Acessado em: 10 de Fevereiro de 2011.
- [4] Buyya, R.; YEO C. S.; Venugopal, S. *Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities*. Grid Computing and Distributed Systems (GRIDS) Laboratory Department of Computer Science and Software Engineering. The University of Melbourne, Australia, 2008.
- [5] Cassandra. *Cassandra* Disponível em: <http://cassandra.apache.org/>. Acessado em: 05 de Janeiro de 2011.
- [6] Cottman, G. *Why Cloud Databases*. Disponível em: [http://wiki.toadforcloud.com/index.php/Why\\_cloud\\_databases](http://wiki.toadforcloud.com/index.php/Why_cloud_databases). Acessado em: 12 de Fevereiro de 2011.
- [7] Couchdb. *The CouchDB Project*. Disponível em: <http://couchdb.apache.org>. Acessado em: 05 de Janeiro de 2011.
- [8] Dillon, Tharam.; Chang, Chen. *Cloud Computing: Issues and Challenges*. 24th IEEE International Conference on Advanced Information Networking and Applications, 2010.
- [9] Dzone. *NoSQL DZone Poll Results*. Disponível em: <http://java.dzone.com/articles/nosql-dzone-poll-results>. Acessado em: 05 de Janeiro de 2011.
- [10] Embarcadero. *Database Trends Survey Report*. Disponível em: <http://www.embarcadero.com/images/dm/technical-papers/database-survey-report.pdf>. Acessado em: 18 de dezembro de 2010.

- [11] Gil, Antonio C. *Metodos e técnicas de pesquisa social*. 5.ed. São Paulo: Atlas, 1999.
- [12] MongoDB. *Mongo DB*. Disponível em: <http://www.mongodb.org/>. Acessado em: 05 de Janeiro de 2011.
- [13] Nist. *The NIST Definition of Cloud Computing*. 2009. Disponível em: <http://csrc.nist.gov/groups/SNS/cloud-computing/>. Acessado em: 23 de novembro 2010.
- [14] Silva, Edna Lúcia. *Metodologia da pesquisa e Elaboração de dissertação*. 3º Ed. Florianópolis: 2001.
- [15] Sousa, Flávio R. C.; Moreira, Leonardo O.; Macêdo, José Antônio de; MACHADO, Javam C. *Gerenciamento de Dados em Nuvem: Conceitos, Sistemas e Desafios*. Disponível em: [http://www.es.ufc.br/~flavio/files/Gerenciamento\\_Dados\\_Nuvem.pdf](http://www.es.ufc.br/~flavio/files/Gerenciamento_Dados_Nuvem.pdf). Acessado em 20 de dezembro de 2010.
- [16] Wei, Z., Pierre, G., Chi, C.-H. (2009). *Scalable transactions for web applications in the cloud*. Disponível em: [http://www.globule.org/publi/STWAC\\_europar2009.pdf](http://www.globule.org/publi/STWAC_europar2009.pdf). Acessado em 23: de dezembro de 2010