

**FACULDADE DE ADMINISTRAÇÃO E NEGÓCIOS DE
SERGIPE – FANESE**



**CURSO DE PÓS GRADUAÇÃO
MBA EM ADMINISTRAÇÃO DE BANCO DE DADOS**

AUGUSTO LEITE PINTO DE CARVALHO

**Uma abordagem sobre Banco de Dados como Serviço
(DbaaS): A nova tendência da Computação em Nuvem**

Aracaju – SE

2016.1

AUGUSTO LEITE PINTO DE CARVALHO*

Uma abordagem sobre Banco de Dados como Serviço (DbaaS): A nova tendência da Computação em Nuvem

Trabalho de Conclusão de Curso apresentado ao Núcleo de Pós-Graduação e Extensão – NPGE, da Faculdade de Administração de Negócios de Sergipe – FANESE, como requisito para a obtenção do título de Especialista em Administração de Banco de Dados.

Nome completo do Avaliador

Nome completo do Coordenador de Curso

Nome completo do Aluno

Aprovado (a) com média: _____

Aracaju (SE), ____ de _____ de 2016.

*Graduado em Ciências da Computação – Universidade Tiradentes
augusto_lpc@hotmail.com

SUMARIO

RESUMO	3
ABSTRACT	4
1. INTRODUÇÃO	5
2. COMPUTAÇÃO NA NUVEM	6
3. BANCO DE DADOS NA NUVEM	8
3.1. VANTAGENS	10
3.2. DESVANTAGENS	10
4. CONCLUSAO	12
BIBLIOGRAFIA	13

RESUMO

O presente artigo introduz o conceito de banco de dados como serviço (DBaaS - Database as a Service) na nuvem. Relacionando as características do cloud computing com os níveis de virtualização de banco de dados (Database Cloud) como facilidades de configuração, disponibilidade e escalabilidade. Por fim apresentar os principais desafios encontrados pelos provedores desse serviço.

PALAVRAS CHAVES: Computação em nuvem. Banco de Dados como Serviço, Banco de dados na nuvem.

ABSTRACT

This article introduces the concept of database as a service (DBaaS - Database as a Service). Relating the characteristics of cloud computing with the database virtualization levels (Database Cloud) as configuration facilities, availability and scalability. Finally present the main challenges faced by that service providers.

Key Words: Cloud computing, Cloud database, DataBase as a service.

1. INTRODUÇÃO

A virtualização tem sido tema muito abordado nas decisões estratégicas das empresas atualmente. A ideia de pagar apenas pelos serviços e recursos que se utiliza dar oportunidade principalmente para médias e pequenas empresas que passam a ter acesso a poderosas estruturas através da rede com menor custo.

Esse artigo define o conceito de computação em nuvem que vem revolucionando as infraestruturas de TI. Após uma evolução da computação distribuída, dos clusters e virtualização, surgiu o conceito de disponibilizar infraestrutura como serviço na web. Com isso muitas aplicações e softwares têm sido migrados para nuvem.

O mais recente serviço a ser migrado é o banco de dados, DBaaS – Database as a Service. Que herda muitas características do cloud computing, porém os sistemas gerenciadores de banco de dados possuem particularidades no processamento e armazenamento dos dados que se dificultam no ambiente virtual.

Dentre as características que serão apresentadas, será dada uma atenção maior para a elasticidade de recursos como CPU, memória e disco. Uma grande vantagem da virtualização que vem trazendo muitos desafios para os fornecedores dos serviços de banco de dados na nuvem.

Na prática, os serviços de cloud computing oferecidos por algumas empresas diferem do conceito de database como serviço (DBaaS) e application service provider (ASP). Ao invés do fornecedor criar, instalar e manter o serviço do banco de dados, eles mantêm apenas o hardware e disponibilizam uma série de máquinas virtuais para que seja tudo instalado e configurado pelo consumidor. [1]

Por fim, serão listadas as vantagens e desvantagens da implantação de um banco de dados na nuvem e levantar os desafios dos gerenciadores da base de dados de dados (DBMS – Database Management Softwares) com os conceitos de virtualização.

2. Computação na nuvem (Cloud Computing)

A computação em nuvem é uma tendência recente de tecnologia cujo objetivo é proporcionar serviços de Tecnologia da Informação (TI) sob demanda com pagamento baseado no uso.

O surgimento vem da evolução computacional no início dos anos 2000 marcada pelo uso de Web Services e o surgimento de arquiteturas como clusters e grades computacionais que usam o poder de instituições ao redor do mundo para a colaboração entre pesquisadores e cálculo de aplicações científicas [3,4]. Junto com os avanços das técnicas de virtualização que ao invés de dedicar um servidor para uma função específica na qual pode ultrapassar os recursos estimados em momentos de pico, são criadas máquinas virtuais (VM) que podem ser redimensionadas com as necessidades de cada aplicação.

Segundo o Instituto Nacional de Padrões e Tecnologia, o NIST (National Institute of Standards and Technology):

Computação em nuvem é uma infraestrutura com capacidades de provisionamento de recursos de forma rápida e elástica, em alguns casos automáticas medidas para aumentar e diminuir o número de recursos. Para os usuários, tais capacidades muitas vezes parecem ser ilimitadas, podendo ser realizadas em qualquer quantidade e a qualquer momento.

Existem duas divisões no conceito de elasticidade no cloud computing, uma abordagem horizontal que é possível aumentar e diminuir o número de instancias, além de migrar para diferentes nós de processamento. E a vertical que redimensiona os atributos de CPU, disco, memória ou rede, assim como alocar e liberar nós de computação [5].

A alocação elástica de recursos é importante tanto para suprir as necessidades quando o comportamento do sistema ultrapassa a configuração subestimada nos momentos de pico, como também em caso de desperdício de recursos quando mesmo em pico não atinge a média alocada. A figura a seguir ilustra esse conceito de alocação estática e elástica de recursos:

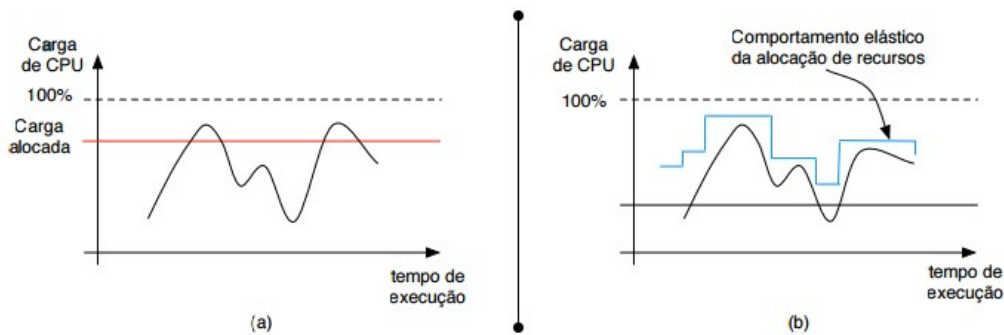


Figura 1: (a) – carga de serviço ultrapassa a alocada; (b) tratamento elástico de acordo com o comportamento do serviço [5]

Existem diversas métricas e medidas de monitoramento, como carga de CPU, tempo de execução de tarefas, memória usada pelo serviço ou ainda espaço em disco, que servem como parâmetros para lançar as ações de elasticidade.

Porém, a adição de recursos não é tão simples quanto apenas migrar o sistema para um servidor mais poderoso; e sim os recursos são obtidos através da adição de novas instancias para realizar mais tarefas. Com isso, as aplicações precisam saber trabalhar em paralelismo com as novas instancias para que se tire proveito dos recursos adicionais. Alguns provedores como o Google's App Engine que fornece além da estrutura da nuvem um grupo de API para os desenvolvedores terem facilidades para utilizar melhor da elasticidade de recursos [1].

Outros exemplos de fornecedores de Infraestrutura como serviço (IaaS) são AWS – Amazon Web Service, Windows Azure, GoGrid, AT&T's Synaptic Hosting, AppNexus, Rackspace Cloud Hosting.

Além de fornecer recursos computacionais (Processamento, disco, memória e rede), a nuvem também oferece sistemas de software como serviço (SaaS – Software as a Service). O modelo proporciona o usuário acessar as aplicações de qualquer lugar da web e sem se preocupar em manter a estrutura física e ainda reduzir os custos, pois é dispensada a aquisição de licenças. Como exemplo de SaaS pode-se destacar os serviços de Customer Relationship Management (CRM) da Salesforce e o Google Docs.

Nas próximas sessões, será abordado como o serviço de bancos de dados pode ser iniciado na nuvem; o armazenamento, disponibilidade,

segurança dos dados e os desafios encontrados pelos fornecedores desse serviço.

3. Banco de Dados na Nuvem (DataBase as a Service)

Como descrito acima, 'cloud computing' pode ser definido como um sistema distribuído e paralelo que disponibiliza serviço através da rede. O cloud database deve também facilmente gerenciar recursos com a necessidade, distribuir os dados em diferentes localidades, ser acessível de qualquer lugar e a qualquer momento, e reduzir custos.

A estrutura de banco de dados na nuvem difere dos SGBDs tradicionais pela complexidade de ter diversos nós para armazenar e consultar os dados distribuídos. Tanto que inicialmente esse tipo de serviço era utilizado apenas em sistemas com muitas consultas e poucas escritas como, por exemplo, as aplicações de BI.

As aplicações interagem com o banco de dados na nuvem utilizando uma camada de comunicação como o JDBC e as requisições são feitas por drives que garantem a privacidade dos dados (inibindo o conteúdo dos administradores do serviço). Quando uma consulta é realizada, o gerenciador (frontend Nodes) utiliza seu metadata para definir os nós envolvidos, o plano de consulta, gerenciamento de failover (tolerância à falha) e faz um controle de desempenho entre os nós [2]. A figura2 a seguir demonstra a interação dos elementos da estrutura de dados na nuvem com as requisições das aplicações clientes.

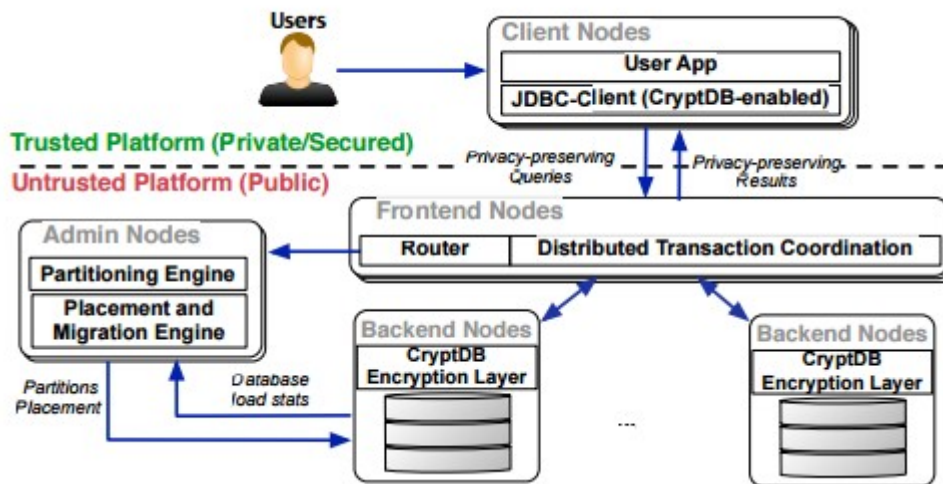


FIGURA 2: Arquitetura do Banco de Dados na nuvem.[2]

O cloud database management system (CDBMS) utiliza as informações de monitoramento do frontend para determinar a necessidade de particionamento da base, quando a mesma excede a capacidade de algum nó, migram as partições sem causar paralisação de serviço (downtime), replica os dados para disponibilidade e criptografa os dados no novo host que esta hospedando a base.

Existem diferentes provedores de banco de dados como serviço: Amazon RDS, Microsoft SQL Azure, Google AppEngine Datastore e Amazon SimpleDB. Cada provedor difere entre eles nas limitações de funcionalidades do serviço, como por exemplo, a capacidade de armazenamento (Amazon RDS disponibiliza mais de 1TB em um banco de dados, SQL Azure apenas 50GB). A portabilidade de serviço, ou seja, permitir migrar a estrutura de cloud de outro fornecedor. Opções de configuração (alguns provedores limitam os parâmetros de configuração de servidores pelo usuário). E os mecanismos de acesso ao banco de dados que podem ser através de drivers como Java Database Connectivity ou interfaces (HTTP) e protocolos como Service-Oriented Architecture (SOA) e REST (Representational State Transfer) [6].

3.1 VANTAGENS

Cloud Databases naturalmente oferecem aos usuários os mesmo benefícios que o cloud computing: um ambiente flexível, escalonável, fácil gestão, acessível de qualquer local na web e com monitoramento de utilização e desempenho.

Provedores se responsabilizam pela alta disponibilidade e manutenção do serviço como replicação, failover, backup, restore, patches e upgrades. E ainda realizam tarefas de tuning database para melhorar desempenho tirando essa responsabilidade das empresas que teriam que contratar especialistas para manter o serviço e com isso podendo dedicar mais tempo em iniciativas estratégicas.

Facilidades para as pequenas empresas que passam a ter acesso a soluções de banco de dados que apenas grandes empresas tinham. Devido a menor custo de infraestrutura e licenças.

3.2 DESVANTAGENS

As desvantagens quando se utiliza banco de dados como serviço são as limitações de configuração dos servidores que hospedam o serviço, não ter controle da localização de armazenamento e limitação de tamanho da base. Outra limitação é a necessidade de conexão com a internet, sendo que a taxa de transferência de banco de dados tradicionais é superior.

Cada fornecedor de DBaaS tem suas particularidades de armazenamento e configuração dos servidores, com isso fica uma tarefa complicada migrar de o serviço de provedor.

3.1 DESAFIOS DO CLOUD DATABASE

Nessa sessão serão relatados os principais desafios dos bancos de dados na nuvem principalmente para os modelos transacionais onde tem que garantir o conceito de ACID (atomicidade, consistência, integridade e disponibilidade) dos dados, diferente dos modelos analíticos como bases utilizadas para mineração de dados e suporte a decisão, que possuem maior frequência de leitura nos discos.

A segurança é um conceito que deve ser levado em consideração em todas as camadas da computação em nuvem. No banco de dados existe a preocupação de proteger os dados contra furtos, alterações e destruição. Não faz sentido os fornecedores de cloud database terem acesso as informações de seus clientes sem devidas permissões, porém para garantir alta disponibilidade da informação, os serviços são replicados em diferentes regiões, países e até continentes. Submetendo-se as leis e regulamentos de cada destino, por exemplo, os EUA exigem que o governo tenha acesso a todas as informações armazenadas no seu território. [2]

Como os dados hospedados nos provedores são criptografados, os gerenciadores de banco de dados na nuvem (CDBMS) devem realizar consultas e operações nesse tipo de dado com eficiência. Para isso existem algumas técnicas de criptografia para executar com SQL queries como Randomized Encryption (RND), Deterministic Encryption (DET) e os mais recentes Order-Preserving Encryption (OPE) e Homomorphic Encryption (HOM) [1].

O sistema gerenciador de cloud database (CDBMS) tem o grande desafio de processar queries que envolvem dados particionados em diferentes nós [2]. Com a necessidade estão surgindo diferentes sistemas de gestão de dados escaláveis como Bigtable do Google, PNUTS da Yahoo!, Dynamo da Amazon. Muitas pesquisas ainda estão voltadas para uma solução de gerenciamento de bancos transacionais num ambiente escalonado. [7]

5. CONCLUSÃO

As aplicações estão sendo migradas pra nuvem buscando praticidade de acesso e alta disponibilidade da informação. Além de reduzir altos custos de investimentos e dando oportunidade para pequenas e médias empresas, retira a responsabilidade de manutenção da estrutura de servidores e serviços fazendo com que as empresas se dediquem e invistam mais tempo em análises estratégicas.

É possível encontrar ótimos serviços de infraestrutura e aplicações na nuvem. Porém o conceito de cloud database ainda demonstra limitações e desafios no desempenho de transação comparado com modelos tradicionais, na elasticidade do serviço e na privacidade da informação espalhada por distâncias geográficas.

Atualmente o conceito está sendo mais utilizado nas aplicações de 'big datas', mineração de dados e BI onde são realizadas grandes cargas de dados e muita leitura.

REFERÊNCIAS

- [1] - **Daniel J. Abadi (2009)**. Data Management in the Cloud: Limitations and Opportunities, Bulletin of the IEEE Computer Society Technical Committee on Data Engineering. 32(1): 100-105. <ftp://131.107.65.22/pub/debull/A09mar/abadi.pdf>
- [2] - **Carlo Curino, Evan P. C. Jones, Raluca Ada Popa, Nirmesh Malviya(2010)**. Relational Cloud: A Database-as-aService for the Cloud. 235-240 California. Acessado em 30/05/2016. http://www.cidrdb.org/cidr2011/Papers/CIDR11_Paper33.pdf
- [3] **LUO, S.; PENG, X.; FAN, S.; ZHANG, P.** Study on computing grid distributed middleware and its application. In: . c2009. v. 3. p. 441–445.
- [4] **DA ROSA RIGHI, R.; PILLA, L. L.; CARISSIMI, A.; NAVAUX, P. A.; HEISS, H.-U.** Parallel Processing Letters, v. 20, n. 2, p. 123–144, June 2010.
- [5] **DA ROSA RIGHI**, Elasticidade em cloud computing: conceito, estado da arte e novos desafios. Revista brasileira de Computação Aplicada, v.5, n.2, p. 2 -17. Outubro 2013.
- [6] **Shehri, A. W.** Cloud Database – DATABASE AS A SERVICE. Sydney, Australia <http://airccse.org/journal/ijdms/papers/5213ijdms01.pdf>
- [7] **Agrawal, D; Abadi, E. A.; Antony S.; Das, S.** Data Management Challenge in Cloud Computing Infrastructures. http://link.springer.com/chapter/10.1007%2F978-3-642-12038-1_1#page-1
- [8] **Welch, D.** Running databases on VMs: The Time is now. <http://searchservvirtualization.techtarget.com/feature/Running-databases-on-VMs-The-time-is-now>