



Interações entre Big data e cloud computing



Cloud Computing e Big Data

Computação em nuvem (*Cloud Computing*)
e seus modelos

Bloco 1

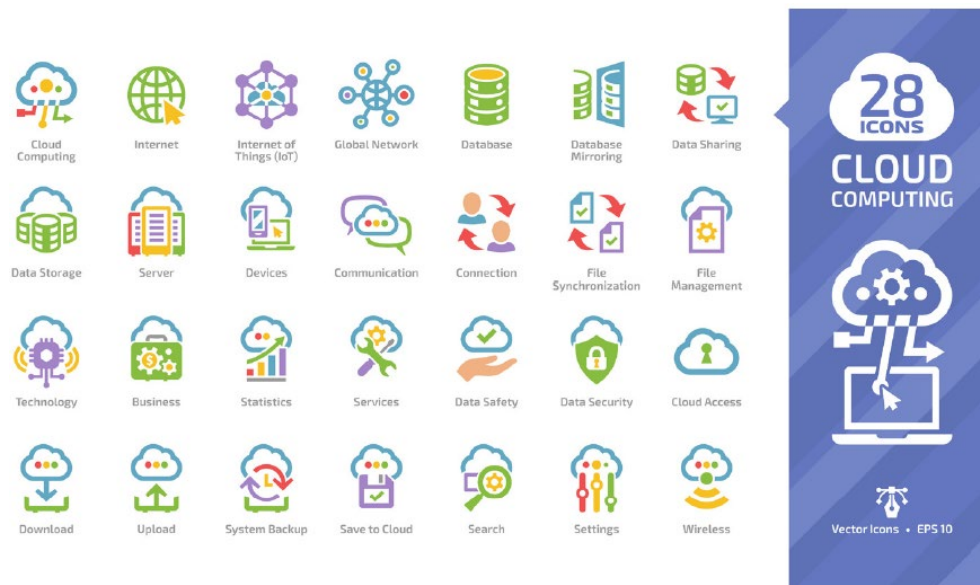
Deivid Sardinha



➤ Computação em nuvem (*Cloud Computing*)

- O termo computação em nuvem é utilizado por profissionais e usuários finais para tudo que pode ser conectado dinamicamente de forma onipresente e conveniente à *web service* (serviços de internet).

Figura 1 – Representação de solução



Fonte: Yuriy Bucharskiy/iStock.com.

➤ Modelos de Serviço (SaaS)

- ***Software as a Service***: fornecimento de software (aplicações) e dados como serviço na internet. Podem ser acessados de qualquer lugar e não há risco de perda de informação se seu computador quebrar ou acabar a energia.



➤ Modelos de Serviço (PaaS)

- ***Platform as a Service:*** fornecimento de plataforma como serviço. Estruturas de desenvolvimento de aplicações são disponibilizadas para desenvolvedores construírem novas aplicações. Não há necessidade da compra de hardware e software; o ambiente disponibilizado possui as ferramentas necessárias para o ciclo de vida da aplicação, do desenvolvimento até a entrega.



➤ Modelos de Serviço (IaaS)

- ***Infrastructure as a Service***: os fornecedores, chamados de provedores, disponibilizam recursos computacionais, tais como servidores, armazenamento, software, rede e espaços físicos, que são pagos de acordo com a demanda.



► Plataformas *Big Data* na nuvem

- Google Cloud Services.
- MS Azure.
- AWS S3.



Relação e desafios do *Big Data* e da computação na nuvem

Deivid Sardinha





Desafios do *Big Data*

- *Big Data* também precisa romper barreiras como:
 - **Demanda cada vez maior.**
 - **Armazenamento.**
 - **Processamento em alta escala.**



► Disponibilidade e escalabilidade

- **Disponibilidade dos recursos na nuvem:**
 - O desafio está no crescimento da demanda no futuro.
- **Escalabilidade do armazenamento:**
 - A escalabilidade se refere à capacidade de armazenar o volume crescente de dados de maneira confiável.



► Integridade e privacidade

- **Integridade de dados:**
 - A integridade dos dados se refere à perda, modificação, acesso e exatidão.
- **Escalabilidade do armazenamento:**
 - Ataques que podem violar dados ou retirar os sistemas do ar são preocupantes. Para tanto, os provedores de serviços na nuvem instalam sistemas com tolerância a falhas.





Qualidade e variedade

- **Qualidade dos dados coletados:**
 - O *Big Data* colhe dados de várias fontes, o problema é que muitas não são conhecidas nem confiáveis e não há maneiras de verificação na maioria delas.
- **Variedade dos dados:**
 - Com a expansão, a dificuldade de armazenar e tratar dados não estruturados e semiestruturados aumenta.



➤ Preparação, análise e segurança de *Big Data*

- **Preparação de dados:**
 - Após coletar e reunir os dados, é preciso fazer a preparação para que estejam no mesmo formato estrutural.
- **Análise de *Big Data*:**
 - São necessários profissionais capacitados, comprometidos e de confiança para elaborar os modelos de análise, bem como ferramentas modernas.
- **Segurança de *Big Data*:**
 - Os serviços de computação em nuvem precisam de criptografia, algoritmos para gerenciar as chaves de segurança.



Teoria em Prática

Bloco 3

Deivid Sardinha



➤ Reflita sobre a seguinte situação

Observe esse caso de mineração de dados por meio de um serviço de computação em nuvem, que foi realizada em uma base de dados do Twitter na nuvem com o uso dos serviços da AWS. Computação em nuvem usada para analisar grandes quantidades de dados no Twitter. O algoritmo Page Rank foi utilizado para obter rankings de usuários da base do Twitter. Os serviços de infraestrutura, armazenamento e processamento de nuvem da AWS foram usados. A hospedagem de todos os cálculos relacionados também. O processo computacional foi realizado em duas fases: na primeira (fase de rastreamento), todos os dados foram recuperados das bases do Twitter; na segunda, o processamento feito pelo algoritmo Page Rank analisava os dados coletados. Durante a fase de rastreamento, foi observado a geração de 50 milhões de nós e 1,8 bilhão de bordas, que significava aproximadamente dois terços da base de usuários do Twitter. Diante do volume, esse é um exemplo de solução relativamente barata para aquisição e análise de dados com o uso de uma infraestrutura de nuvem (HASHEM *et al.*, 2014).

Page Rank é um algoritmo para classificar e posicionar sites de acordo com a busca na internet. Qualifica de acordo com a quantidade e qualidade da busca.

Realize uma leitura sobre o algoritmo Page Rank para melhor compreensão de seu funcionamento.



Dicas do(a) Professor(a)

Bloco 4

Deivid Sardinha





Leitura Fundamental

Prezado aluno, as indicações a seguir podem estar disponíveis em algum dos parceiros da nossa Biblioteca Virtual (faça o login através do seu AVA). Algumas indicações também podem estar disponíveis em sites acadêmicos como o Scielo, repositórios de instituições públicas, órgãos públicos, anais de eventos científicos ou periódicos científicos, acessíveis pela internet.

Isso não significa que o protagonismo da sua jornada de autodesenvolvimento deva mudar de foco. Reconhecemos que você é a autoridade máxima da sua própria vida e deve, portanto, assumir uma postura autônoma nos estudos e na construção da sua carreira profissional.

Por isso, te convidamos a explorar todas as possibilidades da nossa Biblioteca Virtual e além! Sucesso!





Indicação de leitura 1

Este artigo apresenta as relações da computação em nuvem e o *Big Data*. O *Big Data* está em constante crescimento e uma grande infraestrutura computacional para processamento e análise de dados é necessária. Como o *Cloud Computing* é um conjunto de tecnologia capaz de executar e armazenar esses dados, no artigo são apresentadas as relações dessas tecnologias. O *cloud* elimina a necessidade de hardware nas dependências da organização e permite escalabilidade, por isso é recomendável para *Big Data*.

Referência:

HASHEM, I. A. T. *et al.* The rise of “Big Data” on cloud computing: review and open research issues. **Information Systems**, v. 47, p. 98-115, 2015.





Indicação de leitura 2

A AWS é uma grande fornecedora mundial de *Cloud Computing*. O serviço Amazon Elastic Compute Cloud (Amazon EC2) é um dos serviços ofertados, e refere-se a processamento escalável em *cloud* de acordo com a demanda do cliente, além de controle dos recursos envolvidos. O próprio cliente pode inicializar e dimensionar a capacidade contratada rapidamente. E para encerrar, o Amazon EC2 fornece ferramentas para os desenvolvedores criar aplicativos resistentes a falhas e isolá-los de cenários comuns de falhas.

Referência:

AWS. Amazon EC2. Capacidade de computação segura e redimensionável para praticamente qualquer workload. **Amazon**, São Paulo, [s.d.].





Dica do(a) Professor(a)

O Apache Hadoop é um *framework* (estrutura) que executa aplicativos em grandes conjuntos de hardware (*clusters*). A estrutura é composta de aplicativos que trabalham de forma transparente e confiável o processamento de dados. Possui uma técnica computacional chamada Mapear/ Reduzir (MapReduce). No processamento, o aplicativo é dividido em pequenos fragmentos de código e executado na nuvem em qualquer parte do *cluster* de hardware. Os dados são armazenados em um sistema de arquivos distribuídos (HDFS), tanto o armazenamento quanto o processamento são realizados de forma distribuída e paralela. O *framework* também é projetado para ser tolerante a falha, ou seja, a falha é corrigida automaticamente (APACHE HADOOP, [s.d.]).





Referências

AWS. Amazon EC2. Capacidade de computação segura e redimensionável para praticamente qualquer workload. **Amazon**, São Paulo, [s.d.]. Disponível em: <https://aws.amazon.com/pt/ec2/>. Acesso em: 24 abr. 2019.

APACHE HADOOP. [s.d.]. Disponível em: https://wiki.apache.org/hadoop#Apache_Hadoop. Acesso em: 23 abr. 2019.

AWS. O que é a computação em nuvem?. **Amazon**, São Paulo, [s.d.]. Disponível em: <https://aws.amazon.com/pt/what-is-cloud-computing/>. Acesso em: 22 abr. 2019.

DHABHAI, A.; GUPTA, Y. K. A Study of Big Data in Cloud Environment with their Related Challenges. **IJESC**, Índia, v. 6, n. 8., 2016.

HASHEM, I. A. T. *et al.* The rise of “Big Data” on cloud computing: review and open research issues. **Information Systems**, v. 47, p. 98-115, 2015. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0306437914001288>. Acesso em: 21 jun. 2022.



Bons estudos!

