

Leitura Complementar

Disciplina: Interações entre *Big Data* e *Cloud Computing*

Autor: Aimar Martins Lopes



Prezado aluno, selecionamos as referências abaixo visando o aprofundamento das temáticas estudadas na disciplina e a complementação dos seus estudos. Para conferir as indicações, acesse a nossa biblioteca virtual: <<https://biblioteca-virtual.com/>> e boa leitura!

► Tema 01 - ***Big Data***: fundamentos, infraestrutura e interfaces ***Big Data***, mais volume, velocidade e variedade

O artigo sugerido apresenta os conceitos do Big Data e questiona quais problemas ele pode resolver. Embora haja muitos dados circulando no mundo, desde dados históricos, mídia social, sensor de IoT entre outros, o autor afirma que Big Data não é sobre dados, e sim sobre o seu significado. O conceito envolve um ecossistema de tecnologia que inclui coletar dados brutos de várias fontes, armazenar para uso posterior e gerar significado.

PERRY, J., S. **What is Big Data? More than volume, velocity and variety...** Publicado em: 22 mai. 2017. Disponível em: <https://developer.ibm.com/dwblog/2017/what-is-big-data-insight/>. Acesso em: 3 jun. 2019.

Entendendo mais de Big Data

O artigo indicado traz uma revisão de conceitos que envolvem o Big Data, faz uma revisão das tecnologias envolvidas, tais como virtualização, processamento paralelo, Hadoop, *cloud computing*, IoT, etc. Apresenta a cadeia de valor com início na geração de dados, seguida de armazenamento e análise. Discute os possíveis desafios envolvendo o Big Data e aborda as aplicações que tratam as redes sociais, inteligência coletiva e outras. É sugerida especialmente a leitura do item 5 “Big Data Analysis” em que alguns métodos de análise são apresentados.

CHEN, M.; MAO, S.; LIU, Y. Big Data: A Survey. **Mobile Netw Appl**, v. 19, p. 171-209, 2014. Disponível em: <http://www2.egr.uh.edu/~zhan2/ECE6111/class/BigDataSurvey2014.pdf>. Acesso em: 19 jun. 2019.

► Tema 02 - Gerenciamento de Big Data: coleta e processamento de dados

Um overview do Apache Hadoop

A Apache Software Foundation desenvolve e mantém vários softwares de código aberto, o Hadoop é um deles. Por meio da navegação pelo site, recomendo a leitura da seção “Overview” para explorar a configuração de um nó e de um cluster, conhecer comandos e sistema de arquivo. É sugerida especialmente a leitura da seção “HDFS” para entendimento do sistema de arquivo distribuído do Hadoop e a seção do Mapreduce, a estrutura de software tolerante a falha para desenvolvimento de aplicativo que trata grandes quantidades de dados em processamento paralelo em milhares de clusters de hardware.

HADOOP. **Apache Hadoop 3.1.2**. Disponível em: <https://hadoop.apache.org/docs/r3.1.2/index.html>. Acesso em: 24 mai. 2019.

HDFS – Hadoop Distributed File System

Saber como funciona o sistema de armazenamento distribuído em massa do Hadoop é importante e desafiante. O artigo mostra uma visão geral do HDFS, como ele trabalha para armazenar e recuperar os dados, explica por que é tolerante a falhas, etc. Investiga as interfaces com aplicativos e mostra os vários objetivos do HDFS. Descreve sua arquitetura e as relações entre os nós *name node* e *data node*, bem como a replicação e a integridade dos dados.

HANSON, J. J. An introduction to the Hadoop Distributed File System. Explore HDFS framework and subsystems. IBM, 2011. Disponível em: <https://www.ibm.com/developerworks/library/wa-introhdhfs/index.html>. Acesso em: 19 jun. 2019.

► Tema 03 - Aplicação da ciência de dados no gerenciamento empresarial


Impacto do Big Data

Este artigo apresenta a inteligência e a análise de negócios, ou seja, o BI&A, como importante área de estudo para profissionais e organizações e reflete os impactos do Big Data nos problemas organizacionais. O objetivo é mostrar os impactos do Big Data, uma estrutura que mostra sua evolução, seus aplicativos e áreas de pesquisa emergentes em BI&A. São apresentadas suas características e capacidades. Em seguida, são apresentadas as oportunidades associadas a pesquisa e educação em BI&A. Finalmente, é apresentada uma proposta de estrutura de pesquisa em BI&A. É sugerida especial atenção ao item “BI&A Applications: From Big Data to Big Impact”.

CHEN, H.; CHIANG, R. H.; STOREY, V. C. Business intelligence and analytics: from Big Data to big impact. MIS Quarterly, v. 36, n. 4, p. 1.165-1.188, dez. 2012. Disponível em: <https://pdfs.semanticscholar.org/f5fe/b79e04b2e7b61d17a6df79a44faf358e60cd.pdf>. Acesso em: 19 jun. 2019.

O valor dos dados na criação de *insights* organizacionais

Este artigo apresenta como os líderes do mundo questionam o valor obtido dos dados. O leitor encontra nesse artigo algum valor de determinados recursos para ajudar a compreender o caminho dos dados. Nesse sentido, é apresentado como os líderes de organizações de todo o mundo se perguntam se estão obtendo valor total dos dados de suas organizações. Considera-se que as tecnologias coletam enorme volume de dados, mas



muitas organizações ainda não sabem como aproveitar e utilizar para criar valor competitivo. É sugerida uma leitura crítica na intenção de descobrir um caminho para uso dos dados.

LAVALLE, S.; LESSER, E.; SHOCKLEY, R.; HOPKINS, M. S.; KRUSCHWITZ, N.

Big Data, Analytics and the Path from Insights to Value. MIT Sloan Management Review, Magazine: v. 52, n. 2, p. 21-32, inverno 2011. Disponível em: <https://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to-value/>. Acesso em: 19 jun. 2019.

Tema 04 - Big Data Analytics

Introdução ao *machine learning*

O livro aborda o conceito de *machine learning*, com fundamentos, álgebra, rede neural, estatística e algoritmos. É sugerida a leitura do capítulo 1 “Preliminaries”, em que o autor discorre sobre o conceito, as funções de entrada e saída e alguns exemplos de aplicação. O capítulo 4 “Neural Networks” também é interessante para entender como os algoritmos tentam simular o pensamento humano.

NILSSON, N. J. Introduction to machine learning. Robotics Laboratory Department of Computer Science Stanford University Stanford, CA. 1998, Cap. 4, p. 35-62. Disponível em: <https://ai.stanford.edu/~nilsson/MLBOOK.pdf>. Acesso em: 19 jun. 2019.

A indústria automotiva na revolução 4.0

Neste artigo, Bernard Marr apresenta de forma objetiva como a Jaguar Land Rover modifica a indústria automobilística. Apresenta exemplo de inteligência artificial e carro autônomo. A Jaguar tem a tradição de inovação e pretende manter essa posição realizando investimentos para incluir a AI e *machine learning* nos seus automóveis. Recomendo atenção para a narração que muda o conceito da direção do automóvel.

MARR, B. How Jaguar Land Rover Is Getting Ready For The 4th Industrial Revolution: AI & Autonomous Cars. **Forbes**, 2018. Disponível em: <https://www.forbes.com/sites/bernardmarr/2018/10/26/how-jaguar-land-rover-is-getting-ready-for-the-4th-industrial-revolution-autonomous-vehicles/#6a19deba3a5e>. Acesso em 9 abr. 2019.

DL31-v11n4a2017-4

Tema 05 - Algoritmos de aprendizado de máquina para minerar os dados

Clusterização como método para interpretar dados

A análise de dados que agrupa elementos com características comuns é muito utilizada em Big Data para extrair conhecimento de conjuntos de dados. Essa análise é chamada de clusterização, sua aplicação é facilitada pela disponibilidade de armazenamento, processamento e alto fluxo de dados. Este artigo apresenta um método para ajudar a compreensão de cluster com a intenção de melhorar a compreensão e interpretação desses agrupamentos. Recomendo atenção no item 2, “Aprendizado de máquina não supervisionado”, pois algoritmos que tratam esses tipos de dados são muito utilizados.

METZ, J.; MONARD, M. C. Clustering hierárquico: uma metodologia para auxiliar na interpretação dos clusters. In: XXV Congresso da Sociedade Brasileira de Computação. **Anais...** São Leopoldo-RS: Sociedade Brasileira de Computação, 22 a 29 de julho de 2005. Disponível em: <http://www.lbd.dcc.ufmg.br/colecoes/enia/2005/061.pdf>. Acesso em: 24 jun. 2019.

Fundamentos de algoritmo

O texto trata vários conceitos envolvidos com algoritmo de *machine learning*, fundamentos, estrutura e forma de atuação. Destina-se a pesquisar e compreender como funcionam. É sugerida especialmente a leitura que trata de algoritmo supervisionado e rede neural.

BROWNLEE, J. **A Tour of Machine Learning Algorithms**. Publicado em: 25 nov. 2013. Disponível em: <https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>. Acesso em: 24 jun. 2019.

► Tema 06 - *Cloud computing* e Big Data

A relação entre computação em nuvem e o Big Data

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. O objetivo é verificar o aumento do Big Data na computação em nuvem. São revisadas as características e a classificação de Big Data no ambiente *cloud* e, finalmente, o armazenamento e o Hadoop.

HASHEM, I. A. T.; YAQOOB, I.; ANUAR, N. B.; MOKHTAR, S.; GANI, A.; KHAN, S. U. The rise of “Big Data” on cloud computing: review and open research issues. *Information Systems*, v. 47, p. 98-115, 2015. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.665.1326&rep=rep1&type=pdf>. Acesso em: 24 jun. 2019.

Exemplo de um serviço em *cloud computing*

A AWS é uma grande fornecedora mundial de *cloud computing*. O serviço Amazon Elastic Compute Cloud (Amazon EC2) é um dos serviços ofertados, 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.

Amazon EC2. **Capacidade computacional segura e redimensionável na nuvem.** Lance aplicativos quando necessário, sem compromissos antecipados. Disponível em: <https://aws.amazon.com/pt/ec2/>. Acesso em: 24 abr. 2019.

► Tema 07 - Estruturas de programação em nuvem

Composição da computação em nuvem

Este artigo apresenta a tendência do crescimento dos serviços em rede e da concorrência entre os fornecedores, como também a busca da qualidade dos serviços fornecidos. É apresentado o processo difícil e complicado para selecionar os serviços disponíveis na nuvem. A estrutura do artigo é dividida em quatro grupos: resolução de problemas e identificação de parâmetros de qualidade de serviço; objetivos pretendidos e ambientes em desenvolvimento; resultados benéficos; e estatísticas. É sugerida especial atenção ao item 3 *"Cloud computing"*.

JULA, A.; SUNDARARAJAN, E.; OTHMAN, Z. Cloud computing service composition: A systematic literature review. **Expert Systems with Applications**, v. 41, p. 3.809-3.824, 2014. Disponível em: <https://romisatriawahono.net/lecture/rm/survey/network%20security/Jula%20-%20Cloud%20Computing%20-%202014.pdf>. Acesso em: 24 jun. 2019.

Conhecendo o MS Azure

O MS Azure, como o IBM cloud ou AWS entre outros, é um grande fornecedor de serviço em nuvem. Leia com atenção as características técnicas e se aprofunde nas ferramentas ofertadas. Confira as justificativas que indicam o MS Azure, consulte casos de uso de organizações que adotaram o serviço e, finalmente, estude a comparação do MS Azure com o AWS cloud.

MICROSOFT. **Microsoft Azure**. Disponível em: <https://azure.microsoft.com/pt-br/overview/what-is-azure/>. Acesso em: 29 abr. 2019.



Bons estudos!