

**Processamento Analítico de Dados em Larga Escala**  
**Lista de Referências**  
**Profa. Dra. Cristina Dutra de Aguiar**

**Conceitos Relacionados a Business Intelligence**

W. Grossmann and S. Rinderle-Ma. *Fundamentals of Business Intelligence*. Springer, 2015.

M. Golfarelli, S. Rizzi, and I. Cella. *Beyond Data Warehousing: What's Next in Business Intelligence?* In *Proceedings of the 7<sup>th</sup> International Workshop on Data Warehousing and OLAP*, p. 1-6, 2004.

**Conceitos Relacionados a Data Warehousing**

R. Kimball and M. Ross. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. Wiley, 2nd edition, 2002.

W. H. Inmon. *Building the Data Warehouse*. Wiley, 4th edition, 2005.

C. T. Salley and E.F. Codd. *Providing OLAP to User-Analysts: An IT Mandate*. White paper, 1998.

S. Chaudhuri and U. Dayal. *An Overview of Data Warehousing and OLAP Technology*. *SIGMOD Record*, v. 26, n.1, p. 65–74, 1997.

A. Vaisman and E. Zimányi. *Data Warehouse Systems: Design and Implementation*. Springer, 2014.

**Conceitos Relacionados a Data Lake**

J. Couto, O. Borges, D. Ruiz, S. Marczak, and R. Prikladnicki. *A Mapping Study About Data Lakes: An improved Definition and Possible Architectures*. In *Proceedings of the 31st International Conference on Software Engineering and Knowledge Engineering*, p. 453-458, 2019.

C. Mathis. *Data lakes*. *Datenbank-Spektrum*, v. 17, n. 3, p. 289–293, 2017.

## **Conceitos Relacionados a Big Data**

M. Chen, S. Mao and Y. Liu, Big Data: A Survey. Mobile Network Applications, v. 19, p. 171-209, 2014.

X. L. Dong, and D. Srivastava. Big Data Integration. Proc. VLDB Endowment, v. 6, n. 11, p. 1188-1189, 2013.

S. Sharma, and V. Mangat. Technology and Trends to Handle Big Data: Survey. In: Proc. 5th International Conference on Advanced Computing & Communication Technologies, p. 266-271, 2015.

## **Conceitos Relacionados à Computação em Nuvem**

K. Hwang, J. Dongarra and G. Fox. Distributed and Cloud Computing: From Parallel Processing to the Internet of Things. Burlington, Massachusetts, USA: Morgan Kaufmann, 2013.

N. Sadashiv and S. M. D. Kumar. Cluster, Grid and Cloud Computing: A Detailed Comparison. In: Proceedings of the 6th International Conference on Computer Science Education, 2011. p. 477-482.

M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica and M. Zaharia. Above the clouds: A Berkeley View of Cloud Computing. Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, 2009.

P. Mell, T. Grance. The NIST Definition of Cloud Computing. Computer Security Division, Information Technology Laboratory, National Institution of Standards and Technology, 2011.

## **Conceitos Relacionados ao Processamento Paralelo e Distribuído e ao Modelo de Programação Map Reduce**

K. Shvachko, H. Kuang, S. Radia, and R. Chansler. The Hadoop Distributed File. In Proceedings of the IEEE 26th Symposium on Mass Storage Systems and Technologies, p. 1–10, 2015.

Apache Hadoop. Disponível em: <https://hadoop.apache.org>.

Apache Spark. Disponível em: <https://spark.apache.org/>.

J. Dean and S. Ghemawat. MapReduce: Simplified Data Processing on Large Clusters. Communications of the ACM, v. 51, n. 1, p. 107-113, 2008.

P. Derbeko, S. Dolev, E. Gudes and S. Sharma. Security and Privacy Aspects in MapReduce on Clouds: A survey. Computer Science Review, v. 20, p. 1-28, 2016.

M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker and I. Stoica. Spark: Cluster Computing with Working Sets. In: 2nd USENIX Workshop on Hot Topics in Cloud Computing. Berkeley, CA, USA: USENIX Association, 2010.

Morais, T. S. Survey on Frameworks for Distributed Computing: Hadoop, Spark and Storm. In: Proceedings of the 10th Doctoral Symposium in Informatics Engineering. Porto, Portugal: Faculdade de Engenharia da Universidade do Porto, 2015. p. 95-105.

### **Conceitos Relacionados ao Modelo Relacional e à Linguagem SQL**

R. Elmasri and S. B. Navathe. Sistemas de Banco de Dados. Pearson, 7ª edição, 2019.

A. Silverschatz, H. F. Korth and S. Sudarshan. Sistema de Banco de Dados. LTC, 7ª edição, 2020.

### **Conceitos Relacionados a Spark SQL e pyspark**

Damji, J. S.; Wenig, B.; Das, T.; Lee, D. (2020) Learning Spark: Lightning-Fast Data Analytics. Second edition. O'Reilly.

M. Armbrust, R. S. Xin, C. Lian, Y. Huai, D. Liu, J. K. Bradley, X. Meng, T. Kaftan, M. J. Franklin, A. Ghodsi and M. Zaharia. Spark SQL: Relational Data Processing in Spark. In: ACM SIGMOD International Conference on Management of Data, 2015. p. 1383–1394.

Apache Spark SQL. Disponível em:

<https://spark.apache.org/docs/latest/api/python/reference/pyspark.sql.html>