

SCC5911 - Procedência dos Dados e Data Warehousing

OBJETIVOS

Estudo de conceitos sobre procedência dos dados, com ênfase na investigação de aspectos relacionados à procedência dos dados em processos de integração. Estudo de conceitos sobre ambientes de *data warehousing*, com ênfase nas características básicas desses ambientes, na modelagem multidimensional dos dados, no uso de visões materializadas e índices. Estudo de ambientes de *data warehousing* não convencionais, incluindo *data warehousing* espacial, *data warehousing* espaço-temporal e *data warehousing* de imagens.

CONTEÚDO (EMENTA)

1. Procedência dos Dados em Processos de Integração

- 1.1 Conceitos básicos de integração de dados: definições, integração em nível de esquema, integração em nível de esquema, ambiguidade na identificação de entidades e resolução de conflitos de valores de atributos
- 1.2 Conceitos básicos de procedência dos dados: definições e características básicas
- 1.3 Quatro aspectos da procedência: o que armazenar, como coletar, como armazenar e como consultar
- 1.4 Uso da procedência dos dados na reuplicação de decisões de integração

2. Data Warehousing

- 2.1 Introdução aos conceitos de processamento analítico dos dados: definição de *data warehouse*, *data warehousing* e OLAP (*on-line analytical processing*), diferenças entre o processamento analítico e o processamento transacional, características e organização dos dados do *data warehouse*, arquitetura do ambiente de *data warehousing*
- 2.2 Modelagem multidimensional dos dados: aspectos estáticos e dinâmicos, sistemas ROLAP (*relational OLAP*), exemplos práticos.
- 2.3 Visões materializadas: conceito e características, vantagens e desvantagens, uso em consultas OLAP.
- 2.4 Indexação em ambientes de *data warehousing*: índice bitmap de junção
- 2.5 Ambientes de *data warehousing* não-convencionais: *data warehouse* espacial, *data warehouse* espaço-temporal, *data warehouse* de imagem

BIBLIOGRAFIA

Fundamentais:

- 1) KIMBALL, R., ROSS, M. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, 2nd edition. John Wiley and Sons, Inc., 2002.
- 2) CIFERRI, C., CIFERRI, R., GÓMEZ, L., SCHNEIDER, M., VAISMAN, A., ZIMÁNYI, E. Cube algebra: A generic user-centric model and query language for OLAP cubes. *Journal of Data Warehousing and Mining*, v. 9, n. 2, p. 39-65, 2013.
- 3) TAN, W.-C. Research problems in data provenance. *IEEE Data Engineering Bulletin*, v. 27, n. 4, p. 45-52, 2004.
- 4) FREIRE, J.; KOOP, D.; SANTOS, E.; SILVA, C. T. Provenance for computational tasks: A survey. *IEEE Computing in Science & Engineering*, v. 10, n. 3, p. 11-21, 2008.

Complementares (se houver):

- 1) BLEIHOLDER, J.; NAUMANN, F. Data fusion. *ACM Computing Surveys*, v. 41, n. 1, p. 1:1-1:41, 2009
- 2) BUNEMAN, P., KHANNA, S., TAN, W. C. Why and where: A characterization of data provenance. In *ICDT '01: Proceedings of the 8th International Conference on Database Theory*, pages 316–330. 2001.
- 3) CIFERRI, C. D. A. Distribuição dos Dados em Ambientes de Data Warehousing: O Sistema WebD2W e Algoritmos Voltados à Fragmentação Horizontal dos Dados. Tese de Doutorado, Universidade Federal de Pernambuco (UFPE), 2002. (In Portuguese)
- 4) GAEDE, V.; GUNTHER, O. Multidimensional access methods. *Journal of Molecular Biology (JMB)*, v. 30, p. 170–231, 1998.
- 5) GOLFARELLI, M.; RIZZI, S.; IURIS, C. Beyond Data Warehousing: What's Next in Business Intelligence? In *Proceedings of the 7th ACM International Workshop on Data Warehousing and OLAP*, p. 1-6, 2004.
- 6) HARINARAYAN, V.; RAJARAMAN, A.; ULLMAN, J. D. Implementing data cubes efficiently. *ACM SIGMOD Record*, v. 25, p. 205-216, 1996.
- 7) O'NEIL, P. E.; GRAEFE, G. Multi-table joins through bitmapped join indices. *ACM SIGMOD Record*, v. 24, n. 3, p. 8-11, 1995.
- 8) TEIXEIRA, J. W.; ANNIBAL, L.P. ; FELIPE, J.C. ; CIFERRI, R. R. ; CIFERRI, C. D. A. A Similarity-Based Data Warehousing Environment for Medical Images. *Computers in Biology and Medicine*, v. 66, p.190-208, 2015.
- 9) TOMAZELA, B.; HARA, C. S.; CIFERRI, R. R.; CIFERRI, C. D. A. Empowering Integration Processes with Data Provenance. *Data & Knowledge Engineering*, v. 86, p. 102-123, 2013.