

Programa de Ensino

1) Identificação

Disciplina: INE5643 - Data Warehouse
Carga horária: 72 horas-aula Teóricas: 48 Práticas: 24
Período: 2º semestre de 2020 até a presente data

2) Cursos

- Ciências da Computação (208)
- Engenharia, áreas Elétrica e Mecânica, habilitação Controle e Automação (220)
- Sistemas de Informação (238)

3) Requisitos

- Ciências da Computação (208)
 - INE5417 - Engenharia de Software I
- Sistemas de Informação (238)
 - INE5616 - Bancos de Dados II

4) Ementa

Conceitos básicos. Os Ecossistemas de Informação. O Ciclo de Vida do DW: Planejamento e Administração, Levantamento de Requisitos, Modelagem Dimensional, Projeto Físico, o Back-Room e o Front-Room, Metadados, Arquiteturas, Implementação, Segurança e Internet. Implantação, Suporte e Treinamento.

5) Objetivos

Geral: Fornecer conhecimentos teóricos e práticos sobre data warehouses e tendências recentes como data lakes e ciência de dados, habilitando o aluno a aplicar os conhecimentos obtidos na concepção, construção e operação de Data Warehouses, data marts e data lakes para análise de informação.

Específicos:

- Assimilar os principais conceitos acerca da tecnologia de Data Warehouse (DW);
- Compreender as principais metodologias de desenvolvimento de DW;
- Conhecer e saber aplicar a modelagem dimensional (esquemas estrela e floco de neve);
- Aprender a implementar processos de Extração, Transformação e Carga de Dados (ETL/ELT).
- Saber efetuar análises de dados com Processamento Analítico OnLine (OLAP).
- Compreender o estado da arte e as tendências atuais em data warehouses, data lakes e análise de dados, utilizando inclusive algumas técnicas e ferramentas da ciência de dados.

6) Conteúdo Programático

- 6.1) Introdução [8 horas-aula]
 - Dos dados a informação.
- 6.2) A Evolução dos SADs [4 horas-aula]
 - Histórico
 - SGDBs e Sistemas transacionais

- 6.3) Ecossistemas de Informação [8 horas-aula]
 - Business Intelligence
 - Business Management
 - Business Operacional
- 6.4) O Ciclo de Vida do DW [16 horas-aula]
 - Planejamento do DW
 - Escopo do projeto
 - Plano do projeto
 - Administração
 - Definição de requisitos
 - Questionários para entrevistas e sessões coordenadas
- 6.5) Modelagem Dimensional [8 horas-aula]
 - Passos para construção do esquema estrela
 - Técnicas básicas
 - Técnicas avançadas
- 6.6) Back Room (ETL) [4 horas-aula]
 - Extração, Transformação e Carga
 - Área de estagiamento
 - Ferramentas open source
- 6.7) Front Room (OLAP) [8 horas-aula]
 - Rolap, Molap, Holap, Dolap e Wolap
 - Ferramentas e seu uso
- 6.8) Projeto Final [16 horas-aula]

7) Bibliografia Básica

- KIMBALL, Ralph. The data warehouse lifecycle toolkit: expert methods for designing, developing, and deploying data warehouses. 2nd edition. New York: John Wiley & Sons, 2011. 672p. ISBN 978-1-118-07956-0

8) Bibliografia Complementar

- INMON, William H. Building the data warehouse. 3rd ed New York: J. Wiley, c2002. 412p. ISBN 0471081302 (1 exemplar na BU)
- KIMBALL, Ralph. The data warehouse lifecycle toolkit: expert methods for designing, developing, and deploying data warehouses. New York: John Wiley & Sons, c1998. 771p. ISBN 0471255475 (4 exemplares na BU)
- KIMBALL, R. & ROSS, M. The Data Warehouse Toolkit. The Complete Guide to Dimensional Modeling. Second Edition, John Wiley & sons, Inc., New York, 2002.
- GOLFARELLI, M. & RIZZI, S. Data Warehouse Design: Modern Principles and Methodologies (1st. ed.). McGraw-Hill, Inc., USA, 2009.
- MOSS, L. T. & ATRE, S. Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications. Boston, MA - Addison-Wesley Information Technology Series, 2003. ISBN: 0201784203
- MACHADO, F. N. Tecnologia e projeto de Data Warehouse: uma visão multidimensional. São Paulo: Érica, 2004.
- TURBAN, E.; SHARDA, R.; ARONSON, J.; KING, D. BUSINESS INTELLIGENCE: um enfoque gerencial para a inteligência do negócio. Porto Alegre: Bookman, 2009.
- SHARMA, B. Architecting Data Lakes, 2nd Edition, O'Reilly Media, Inc. 2018. ISBN: 9781492032991
- GORELIK, A. The Enterprise Big Data Lake, O'Reilly Media, Inc., 2019. ISBN: 9781491931554.
- LAURENT, A.; LAURENT, D.; MADERA, C. Data Lakes, Volume 2. Wiley 2020. Print ISBN:9781786305855 Online ISBN:9781119720430.