
CURSO: Ciência de Dados e Inteligência Artificial – 1º semestre de 2023

DISCIPLINA: **Modelagem Informacional**

PROFESSOR(ES): Júlio César Chaves

CARGA HORÁRIA: 60h

PRÉ-REQUISITO: Banco de Dados

PLANO DE ENSINO

1. Ementa

Formatos de Dados. Modelagem Dimensional e Modelos de Dados. Dados Estruturados e Dados não Estruturados. Extração, Transformação e Carga (ETL e ELT). Dados em Streaming. Projetos de Data Warehouse e Data Lake. Sistemas e Melhores Práticas para Gestão de Big Data.

2. Objetivos da disciplina

Esta disciplina tem como objetivo geral capacitar o aluno para organizar demandas informacionais, de forma a intermediar através da modelagem, a concepção de soluções computacionais. Especificamente, o objetivo é realizar a exposição dos seguintes conteúdos:

Modelo Conceitual e Modelo Relacional, já visto em Bancos de Dados (revisão);

MIR (Modelagem Informacional de Requisitos), para auxiliar na tarefa de elicitação de requisitos;

Modelo Dimensional;

Dados estruturados vs. dados não-estruturados;

ETL vs. ELT;

Dados em Streaming;

Gestão Básica de Projetos;

Projetos de Data Warehouse;

Projetos de Data Lake;

Governança de Dados.

3. Procedimentos de ensino (metodologia)

A exposição dos conteúdos irá obedecer a estrutura macro da disciplina, que segue propositalmente a cronologia aproximada do surgimento dos conceitos no mercado: 1) Modelo Relacional, 2) Modelo Dimensional, 3) Dados não-estruturados, 4) Data Lake, e finalmente a 5) Governança de Dados.

Além da estrutura macro, sempre que possível será aplicada uma estrutura “micro” em cada aula, que é permeada pelos conceitos principais da Matriz de Zachman (Data Model Patterns: a metadata map), da seguinte forma: 1) Análise de Requisitos, 2) Modelo Conceitual, 3) Modelo Lógico, 4) Modelo Físico, e 5) Implementação.

Apesar da existência de diversas tecnologias, o foco maior será na capacidade do aluno pensar e modelar os problemas, sendo capaz de implementar os modelos, algo que o aluno poderá tirar proveito em qualquer tecnologia.

As ferramentas principais da disciplina serão:

O servidor de laboratório da EMAP, além de outros recursos na nuvem.

ERDPLUS (<https://erdplus.com/>) – Modelagem em diagramas ER (Conceitual), relacional, e Dimensional.

Visual Paradigm (online.visual-paradigm.com) – Modelagem de diversos tipos de diagramas.

MS SQL Server – Banco de dados para modelo relacional e dimensional.

HADOOP – Ponto de partida para um Data Lake. Tentativa de créditos de nuvem para uso dos alunos, conhecendo uma plataforma de Data Lake disponibilizada pelo serviço de nuvem.

Tableau – Ferramenta de visualização de dados.

Gremlin (Gremlify.com) – Plataforma para manipular bases de dados de grafos, como exemplo de dados não-estruturados.

4. Conteúdo programático detalhado

Datas	Tópico	Atividades
13/2/23	Estrutura da disciplina e fundamentos iniciais.	Fórum.
15/2/23	MIR – Modelagem Informacional de Requisitos.	Definição dos grupos de trabalho. Iniciar minimundo UNIVW.
20/2	Carnaval	
22/2	Carnaval	
27/2/23	Correção exercício da aula anterior.	
1º/3/23	Conceitos básicos de projetos.	Elaboração de um plano de trabalho em equipe a partir do minimundo resolvido (UNIVW). Apresentação na próxima aula.
6/3/23	Apresentação em grupo do plano e da execução do projeto UNIVW.	Apresentação do UNIVW em grupo .

Datas	Tópico	Atividades
8/3/23	DW – conceitos	Sorteio dos temas do trabalho para a A1.
13/3/23	DW – modelo dimensional	
15/3/23	DW – modelo dimensional	
20/3/23	DW – ETL, OLAP, Ferramentas	
22/3/23	DW – ETL, OLAP, Ferramentas	Questionário
3/4/23	Nuvem - conceitos	
5/4/23	Nuvem - conceitos	Laboratório
10/4/23	Espaço para revisão de dúvidas para a apresentação dos trabalhos.	
12/4/23	Avaliação – peso 4	Apresentação de trabalho em grupo.
17/4	Semana A1	Divulgação
19/4	Semana A1	Em 28/4
24/4/23	DL – Big Data	
26/4/23	DL – Data Lake vs DW	
1º/5/23	São José (Dia do Trabalho)	
3/5/23	DL – Data Lake vs DW	
8/5/23	DL – arquiteturas e projetos	Questionário
10/5/23	DL – rodando um map-reduce (lab individual AWS)	Avisar sobre entrega individual: desenho de arquitetura de uma startup na nuvem. Tema livre.
15/5/23	Gestão e governança de dados	Avisar sobre oficina de dados.
17/5/23	Gestão e governança de dados (trabalho em grupo durante a aula)	Trabalho em grupo : apresentação de estratégia de implantação de governança de dados no seu minimundo.
22/5/23	Apresentações sorteadas de arquitetura de startup na nuvem.	Prazo de entrega e sorteio para apresentação individual da arquitetura em nuvem.
24/5/23	Gestão e governança de dados (LGPD – anonimização)	
29/5/23	Espaço para oficina de dados	
31/5/23	Espaço para oficina de dados	Prazo de entrega individual da oficina de dados.
5/6/23	Espaço para revisão de dúvidas para a apresentação dos trabalhos.	Apresentação externa.

Datas	Tópico	Atividades
7/6/23	Jordan Vinícius de Oliveira (DPO – Data Protection Officer da FGV)	Apresentação externa.
12/6/23	Bergson Lopes (blrdata.com.br) sobre Gestão e Governança de Dados	Apresentação externa.
14/6/23	Avaliação – peso 4	Apresentação de trabalho em grupo.
19/6	Semana A2	
21/6	Semana A2	
26/6	Semana A2	
28/6/23		
3/7/23	Divulgação A2	
5/7/23		
10/7	Semana AS	Divulgação
12/7	Semana AS	Em 14/4
14/7/23	Encerramento do semestre	

5. Procedimentos de avaliação

A avaliação será feita através da participação na entrega dos exercícios e trabalhos, da seguinte forma: Trabalho 1 (4 pontos) + Exercícios (6 pontos). A recuperação contará unicamente com a Prova, que será um minimundo completo.

Os trabalhos serão em torno de 7 minimundos, a serem sorteados para grupos de 3. A apresentação será por sorteio. A nota poderá ser diferente entre os membros se for identificada a não participação de alguém. Todo o material apresentado deverá ser entregue, juntamente com uma descrição de contribuição de cada membro do grupo.

6. Bibliografia Obrigatória

Inmom, W. H. Data Lake Architecture: Designing the Data Lake and Avoiding the Garbage Dump.

Hay, David. Data Model Patterns: a metadata map. Morgan Kaufmann, 2006.

Sharma, Ben. Architecting Data Lakes: Data Management Architectures for Advanced Business Use Cases <<http://www.oreilly.com/data/free/architecting-data-lakes.csp>>

7. Bibliografia Complementar

Holden Karau. Learning Spark: Lightning-Fast Big Data Analysis

Vermeulen, Andreas François. Practical Data Science: A Guide to Building the Technology Stack for Turning Data Lakes into Business Assets

Silverston, Len. The Data Model Resource Book, Vol. 1: A Library of Universal Data Models for all Enterprises. Wiley, 2001.

Ambler, Scott; Pramod J. Sadalage. Refactoring Databases: Evolutionary Database Design. Addison-Wesley Professional. 2006.

Gamma, Erich; Richard Helm, Ralph Johnson, John Vlissides. Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley Professional. 2016

8. Minicurrículo do(s) Professor(s)

Concluiu a graduação em tecnologia em informática na UNIVERCIDADE/RJ (2006). Possui mestrado (2011) e doutorado (2018) pela COPPE/UFRJ sob o programa de engenharia civil, tendo como foco de pesquisa a mobilidade urbana a partir de dados georreferenciados, tais como dados providos por operadoras de telefonia móvel, do IBGE e do setor de transportes. As pesquisas buscam enriquecer, a partir da análise abrangente dos dados, as metodologias tradicionais de inferência de mobilidade. Atualmente coordena a recém-criada área de governança de infraestrutura para pesquisas na Fundação Getúlio Vargas.

Formação acadêmica/titulação

2012 - 2018

Doutorado em Engenharia Civil.

Universidade Federal do Rio de Janeiro, UFRJ, Brasil.

Título: ESTIMATIVA DE MATRIZES DE ORIGEM-DESTINO A PARTIR DE REGISTROS DE CHAMADAS DE TELEFONIA CELULAR, Ano de obtenção: 2018.

Orientador: Alexandre Gonçalves Evsukoff.

Coorientador: Moacyr Alvim Horta Barbosa da Silva.

Palavras-chave: mobilidade; urbana; cdr; celular; matriz; origem.

Grande área: Ciências Exatas e da Terra

Setores de atividade: Transporte terrestre; Transporte aéreo; Obras de infra-estrutura.

2009 - 2011

Mestrado em Engenharia Civil (Conceito CAPES 6).

Universidade Federal do Rio de Janeiro, UFRJ, Brasil.

Título: AVALIAÇÃO DE SEGMENTAÇÕES URBANAS ATRAVÉS DE INTELIGÊNCIA COMPUTACIONAL, Ano de Obtenção: 2011.

Orientador: Alexandre Gonçalves Evsukoff.

Grande área: Ciências Exatas e da Terra.

Atuação Profissional

FGV - Fundação Getúlio Vargas

2019 - Atual

Vínculo: Colaborador, Enquadramento Funcional: Governança de Infraestrutura para Pesquisas, Carga horária: 44, Regime: Dedicação exclusiva.

2010 - 2018

Vínculo: Colaborador, Enquadramento Funcional: DBA Líder, Carga horária: 44, Regime: Dedicação exclusiva.

2007 - 2010

Vínculo: Outro, Enquadramento Funcional: Administrador de Banco de Dados. Carga horária: 44, Regime: Dedicação exclusiva.

VARIGLOG

2005 – 2007

Vínculo: Outro, Enquadramento Funcional: Administrador de Banco de Dados. Carga horária: 44, Regime: Dedicação exclusiva.

CLARO

2003 – 2005

Vínculo: Outro, Enquadramento Funcional: Administrador de Banco de Dados. Carga horária: 44, Regime: Dedicação exclusiva.

DANA/ATH - Albarus/GKN

1994 – 2002

Vínculo: Outro, Enquadramento Funcional: Aprendiz SENAI em Mecânica Geral, Atendente de Service Desk, Aprendiz de Administrador de Banco de Dados. Carga horária: 44, Regime: Dedicção exclusiva.

Áreas de atuação

1. Grande área: Ciências Exatas e da Terra / Área: Ciência da Computação / Subárea: Sistemas de Computação.

Idiomas

Espanhol

Compreende Razoavelmente, Fala Pouco, Lê Razoavelmente, Escreve Pouco.

Inglês

Compreende Bem, Fala Razoavelmente, Lê Bem, Escreve Razoavelmente.

Português

Compreende Bem, Fala Bem, Lê Bem, Escreve Bem.

Produções

Chaves, J. C., A.H.B. da Silva, M., & Evsukoff, A. G. (2021). Gravity Model Parameter Variation during a Long-Term Study using Mobile Phone Data in the Rio de Janeiro Metropolitan Area. Case Studies on Transport Policy. <https://doi.org/10.1016/J.CSTP.2021.12.017>

BARBOZA, MATHEUS H. C. ; ALENCAR, RICARDO DE S. ; CHAVES, JULIO C. ; SILVA, MOACYR A. H. B. ; ORRICO, ROMULO D. ; EVSUKOFF, ALEXANDRE G. . Identifying Human Mobility Patterns in the Rio de Janeiro Metropolitan Area using Call Detail Records. TRANSPORTATION RESEARCH RECORD, v. 1, p. 036119812097765, 2020.

Citações:1

LENORMAND, MAXIME ; SAMANIEGO, HORACIO ; Chaves, Júlio César ; DA FONSECA VIEIRA, VINÍCIUS ; DA SILVA, MOACYR ALVIM HORTA BARBOSA ; EVSUKOFF, ALEXANDRE GONÇALVES . Entropy as a Measure of Attractiveness and Socioeconomic Complexity in Rio de Janeiro Metropolitan Area. Entropy, v. 22, p. 368, 2020.

Citações:3

Apresentações de Trabalho

1.

Chaves, Júlio César; ALMEIDA, M. S. G. ; BACHA, M. N. ; AYALA, K. ; FERREIRA, I. ; VINHAS, F. . I ENCONTRO DE GESTÃO DE DADOS DE PESQUISA EM HUMANIDADES E CIÊNCIAS SOCIAIS. 2019. (Apresentação de Trabalho/Outra).

2.
CHAVES, J. C.. Arquiteturas de TI como suporte de processamento para pesquisas. 2019. (Apresentação de Trabalho/Seminário).

Eventos

1.
OPEN SCIENCE/OPEN DATA. 2018. (Outra).
2.
Workshop on Urban Computing and Society. 2018. (Outra).

9. Link para o Currículo Lattes

<http://lattes.cnpq.br/6558177209348272>